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

1 **Na Liu[1], Veronika Kivenson[1, 2], Xuefeng Peng[3, 4], Zhisong Cui[5], Thomas S. Lankiewicz[6, 7],**
2 **Kelsey M. Gosselin[1], Chance J. English[3, 8], Elaina M. Blair[6], Michelle A. O'Malley[6, 9], David L.**
3 **Valentine[3, 10, *]**

4 [1]Interdepartmental Graduate Program in Marine Science, University of California Santa Barbara,
5 Santa Barbara, CA, USA

6 [2]Present address: Innovative Genomics Institute, University of California Berkeley, CA, USA

7 [3]Marine Science Institute, University of California Santa Barbara, Santa Barbara, CA, USA

8 [4]Present address: School of Earth, Ocean and Environment, University of South Carolina, Columbia,
9 SC, USA

10 [5]Marine Bioresource and Environment Research Center, Key Laboratory of Marine Eco-
11 Environmental Science and Technology, First Institute of Oceanography, Ministry of Natural
12 Resources of China, Qingdao, China

13 [6]Department of Chemical Engineering, University of California, Santa Barbara, CA, USA

14 [7]Present address: Allonnia, Boston, MA, USA

15 [8]Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara,
16 CA, USA

17 [9]Biological Engineering Program, University of California, Santa Barbara, CA, USA

18 [10]Department of Earth Science, University of California Santa Barbara, Santa Barbara, CA, USA

19 *** Correspondence:**
20 Corresponding Author
21 valentine@ucsb.edu

22 **Keywords: polysaccharides, novel bacterium, *Kiritimatiellota*, CAZymes, sulfatases, nitrogen**
23 **fixation**

24 **Abstract**

25 Marine macroalgae produce abundant and diverse polysaccharides which contribute substantially to
26 the organic matter exported to the deep ocean. Microbial degradation of these polysaccharides plays
27 an important role in the turnover of macroalgal biomass. Various members of the *Planctomycetes-*
28 *Verrucomicrobia-Chlamydia* (PVC) superphylum are degraders of polysaccharides in widespread
29 anoxic environments. In this study, we isolated a novel anaerobic bacterial strain NLcol2^T from
30 microbial mats on the surface of marine sediments offshore Santa Barbara, California, USA. Based
31 on 16S rRNA gene and phylogenomic analyses, strain NLcol2^T represents a novel species within the
32 *Pontiella* genus in the *Kiritimatiellota* phylum (within the PVC superphylum). Strain NLcol2^T is able
33 to utilize various monosaccharides, disaccharides, and macroalgal polysaccharides such as agar and
34 iota-carrageenan. A near-complete genome also revealed an extensive metabolic capacity for
35 anaerobic degradation of sulfated polysaccharides, as evidenced by 202 carbohydrate-active enzymes
36 (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^{T} (= DSM 113125^{T} = MCCC 1K08672^{T}) is proposed to be the type strain of a novel species in *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 on 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^{T} 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.

1 Introduction

Marine macroalgae are important primary producers in coastal ecosystems. They sequester about 173 TgC yr^{-1} 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_2 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 etc. (4, 5). In contrast to terrestrial plants, marine polysaccharides are usually decorated by sulfate and other functional groups, which require specialized enzymes for removal, and thereby limit the range of microbes that can access and degrade these compounds (6).

Members of the PVC superphylum (named for *Plantomycetes, Verrucomicrobia, 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 to anoxic environments ranging from the intestine 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^{T}, is a halophilic saccharolytic bacterium isolated from an anoxic cyanobacterial mat from a hypersaline lake on the Kiritimati Atoll (14). *Pontiella desulfatans* F1^{T} and *Pontiella sulfatireligans* F21^{T} were isolated from Black Sea sediments and are capable of degrading sulfated polysaccharides like iota-carrageenan and fucoidan (15, 16). *Tichowtungia aerotolerans* S-5007^{T} was isolated from surface marine sediment and can grow under microaerobic conditions (17).

80 In this study, we enriched and isolated a novel anaerobic bacterial strain NLcol2^T from the marine
81 sediments offshore Santa Barbara, California, USA, which belongs to the *Kiritimatiellota* phylum.
82 We fed the strain with agar, iota-carrageenan, and fucoidan as carbon substrate to test whether it is
83 able to degrade these polysaccharides or not. Among other isolates of *Kiritimatiellota*, ammonium
84 has been identified as the nitrogen source, but nitrogen fixation has not been observed. However,
85 macroalgal polysaccharides are depleted in nutrients including nitrogen, therefore, we used nitrogen
86 gas as the sole nitrogen source to test its ability of nitrogen fixation. Strain NLcol2^T is characterized
87 by phylogenomic, morphological, chemotaxonomic, and physiological traits. We further investigated
88 its metabolic potential by analyzing CAZymes, sulfatases, and nitrogenases in the genome in detail.

89 2 Materials and Methods

90 2.1 Inoculum source, enrichment, and isolation of strain NLcol2^T

91 Strain NLcol2^T was enriched and isolated from microbial mats found on the surface of marine
92 sediments at Shane Seep (34.40616 N, 119.89047 W) within the Coal Oil Point seep field offshore
93 Santa Barbara, California, USA. Microbial mat samples were collected at 20 m depth with an in-situ
94 temperature of 15 °C in October 2017. The seep area is characterized by a large amount of
95 hydrocarbon gas emissions, microbial mat coverage, and high sulfide and alkalinity in sediment
96 porewater (18–20). The samples used for inoculum contained both microbial mats and partially
97 decomposed macroalgae (**Figure 1a**). The microbial mats were scraped off their attached surface as
98 the inoculum source. The cultures were enriched anaerobically in semi-solid agar (0.25% w/v, BD
99 Difco Agar, Granulated) in the top layer of the sulfide gradient media (**Figure 1b**) modified from
100 Kamp et al., 2006. Cultures were maintained at room temperature in the dark and were transferred
101 into fresh media every two to three weeks for a year.

102 Further isolation of strain NLcol2^T was performed by streaking on agar plates in an anaerobic
103 chamber (Coy Laboratory Products) (**Figure 1c**). The medium is the same as the top agar medium in
104 enrichment cultures, except that 1.5 % w/v agar (BD Difco Agar, Noble) was added as both gelling
105 agent and substrate and 2 mM sulfide added as reducing agent. The Petri dishes were kept in the
106 anaerobic chamber at room temperature (22 °C). Single colonies formed after three weeks and were
107 picked from agar plates. Streak plating was repeated for three more rounds to ensure the purity of the
108 culture. Pure culture was subsequently maintained in liquid media with D-galactose (1g/L) as
109 substrate at 22 °C and was transferred every other week. A full modified medium contained: 28.0 g
110 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
111 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
112 DSMZ medium 141), 0.5 mL Na-resazurin solution (0.1% w/v), 1.0 g D-galactose, 1.0 g NH_4Cl
113 (optional), 0.75 g Na_2CO_3, 0.5 g $Na_2S \cdot 9 H_2O$, 10.0 mL Wolin's Vitamin Solution (see DSMZ
114 medium 141), in 1000 mL distilled water. All ingredients except carbonate, sulfide and vitamins were
115 dissolved under N_2/CO_2 (80:20) atmosphere in Hungate tubes or serum bottles and autoclaved.
116 Carbonate was added from a sterile anoxic stock solution prepared under N_2/CO_2 (80:20)
117 atmosphere. Sulfide and vitamins were added from sterile anoxic stock solutions prepared under
118 100% N_2 gas. Purity of the isolate was checked by full-length 16S rRNA gene sequencing and
119 observation of morphology under the microscope.

120 2.2 Phylogenetic reconstruction by 16S rRNA gene

121 Full-length 16S rRNA gene of strain NLcol2^T was sequenced by GENEWIZ (Azenta Life Sciences),
122 from colonies grown on agar plates. 16S rRNA gene sequence was searched using the website tool

123 BLASTn (21) against the 16S rRNA database and compared to the sequence identity to the other four
124 isolated strains in the *Kiritimatiellota* phylum.

125 To construct a phylogenetic tree based on the 16S rRNA gene, 106 sequences over 1200 bp from the
126 *Kiritimatiellales* order in SILVA Ref NR SSU r138.1 database (released August 2020, accessed
127 November 2021) (22) were selected for alignment. The full-length 16S rRNA genes of strain
128 NLcol2^T, *Tichowtungia aerotolerans* strain S-5007^T, and two *Verrucomicrobia* (ABEA03000104,
129 AF075271 as outgroups) were also added to the alignment using SINA Aligner v1.2.11 (23). The
130 alignment was trimmed using the "gappyout" method in TrimAl v1.4 (24) to remove ambiguous ends
131 and columns with >95% gaps. All trimmed nucleotide sequences represent >50% of the 1568
132 alignment columns. A maximum-likelihood tree was constructed using RAxML v.8.2.9 (25) with
133 GTRGAMMA model of evolution. Rapid bootstrap search was stopped after 1000 replicates with
134 MRE-based criterion. The best-scoring ML tree with support values was visualized in the iTOL
135 server (26).

136 **2.3 Genome sequencing and analyses**

137 Genomic DNA was extracted from the isolate cultures using FastDNA Spin Kit for Soil (MP
138 Biomedicals, OH). Genomic DNA library preparation and sequencing were performed at the
139 University of California Davis Genome Center on Illumina HiSeq 4000 platform with 150-base pair
140 (bp) paired-end reads. Trimmomatic v.0.36 (27) and Sickle v.1.33 (28) were used to remove adapter
141 and low quality or short reads. Trimmed reads were assembled into contigs using MEGAHIT v.1.1.1
142 (29). Contigs longer than 2500 bp were kept and the trimmed reads were mapped back to those
143 contigs using Bowtie2 v.2.3.4.1 (30) and Samtools v.1.7 (31). Contigs were visualized using Anvi'o
144 v.3 interactive interface (32) and manual binning was performed based on coverage, GC content, and
145 tetranucleotide frequency signatures. Completion and redundancy for the reconstructed genome was
146 determined using CheckM v.1.0.7 (33).

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

159 For phylogenomic analyses, high-quality genomes in the *Kiritimatiellales* order from NCBI's
160 GenBank database and the Genome Taxonomy Database (GTDB) r95 were selected (accessed on Feb
161 1, 2021). *Opitutus terrae* PB90-1 from the *Verrucomicrobia* phylum was selected as the outgroup.
162 All genomes meet the GTDB quality criterion based on completeness and redundancy from CheckM:
163 completeness – 5×redundancy > 50. 120 single-copy genes were searched and aligned using GTDB-
164 Tk v1.4.0 (47). The concatenated alignment was further trimmed using TrimAl v1.4 (24) with
165 "gappyout" parameter, which results in a final alignment with 4488 amino acid columns. Maximum-
166 likelihood phylogenetic tree was calculated using RAxML v.8.2.9 (25) with PROTGAMMALG

167 model of evolution. Rapid bootstrap search was stopped after 350 replicates with MRE-based
168 criterion. The best-scoring ML tree with support values was visualized in the iTOL server (26). The
169 average nucleotide identity (ANI) and average amino acid identity (AAI) between genomes were
170 calculated using the ANI/AAI calculator (48).

171 Carbohydrate-active enzymes were predicted using dbCAN2 meta server (49). In brief, uploaded
172 protein sequences were searched against the dbCAN CAZyme domain HMM database v.7, CAZy
173 database (www.cazy.org) and PPR library using HMMER, DIAMOND, and Hotpep programs
174 respectively (49). Only genes predicted by no less than two programs were defined as CAZymes for
175 further analysis. CAZyme gene clusters were predicted by the CGC-Finder on dbCAN2 server with
176 at least one CAZyme and one transporter detected within a maximum distance of two genes (49). To
177 classify sulfatases into families and subfamilies, gene sequences with an annotated sulfatase domain
178 (PF00884) were searched and classified by the SulfAtlas database v.1.1 (50) using the BLASTp
179 program (21). Additionally, SignalP v.5 (51) was used to predict signal peptides for translocation of
180 sulfatases into the periplasmic space and outside of the cells.

181 To better understand the evolution of nitrogen fixation in the *Kiritimatiellota* phylum, reannotation
182 and phylogenetic analysis of the *nifH* gene were performed for all 52 genomes in this phylum from
183 NCBI's GenBank database (accessed on Mar 3, 2020). The same annotation pipeline described above
184 was used to keep consistency and allow better comparison. *nifH* gene sequences were aligned with
185 879 full-length *nifH* genes from the genomes of cultivated diazotrophs
186 (https://wwwzehr.pmc.ucsc.edu/Genome879/) using MUSCLE v.3.8 (52). Two light-independent
187 protochlorophyllide reductases were included as outgroups: ChlL from *Trichormus variabilis* ATCC
188 29413 (WP_011320185.1) and BchL from *Chlorobium limicola* DSM 245 (WP_012467085.1). The
189 alignment was trimmed in Jalview v.2.10.5 (53) to remove ambiguous ends and the columns with
190 >95% gaps. All trimmed amino acid sequences represent >81% of the alignment columns. A
191 maximum-likelihood tree was constructed using RAxML v.8.2.9 (25) with LG substitution model
192 plus GAMMA model of rate heterogeneity. Rapid bootstrap search was stopped after 350 replicates
193 with MRE-based criterion. The best-scoring ML tree with support values was visualized in the iTOL
194 server (26).

195 **2.4 Microscopy**

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

206 **2.5 Chemotaxonomic analysis**

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

210 Cellular fatty acids were extracted twice using a modified Folch method (54) with a chloroform:
211 methanol mixture (2:1) and tridecanoic acid as an internal standard. The samples were partitioned
212 and the organic phase containing the total lipid extract (TLE) was retained. Transesterification of the
213 TLE was performed by adding toluene and 1% sulfuric acid in methanol to the TLE after it was
214 brought to complete dryness under N_2. The acidic methanol/toluene TLE was heated at 90 °C for 90
215 minutes to produce fatty acid methyl esters (FAME). The FAMEs were extracted from the acidic
216 methanol by adding hexane and water, vortexing, centrifuging, and removing the top (hexane)
217 fraction to a new vial twice. The combined transesterified hexane extracts were dried under N_2 to a
218 final volume of 300 μL. Each extract was spiked with methyl heptadecanoate to calculate the
219 recovery of the internal standard and analyzed by gas chromatography with flame ionization
220 detection (GC-FID).

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

227 Analyses of catalase, oxidase, and API ZYM assay for semi-quantitation of enzymatic activities (e.g.
228 beta-galactosidase) were carried out by DSMZ Services, Leibniz-Institut DSMZ – Deutsche
229 Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany.

230 **2.6 Physiology**

231 Bacterial growth of strain NLcol2^T was monitored by measuring optical density (OD) of liquid
232 cultures at 600 nm wavelength. Growth at different temperature (4, 10, 14, 22, 26, 31, 37, 55 °C),
233 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)
234 conditions were determined in triplicates when growing on D-galactose with ammonium supplied.
235 Growth was tested on various substrates (1 g/L) in triplicates at optimum temperature, salinity and
236 pH conditions with ammonium supplied: D-glucose, D-galactose, D-fructose, L-fucose, L-rhamnose,
237 D-mannose, D-mannitol, meso-inositol, D-arabinose, D-xylose, D-cellobiose, lactose, sucrose,
238 maltose, xylan from corn core (TCI), starch (Sigma-Aldrich), cellulose (Sigma-Aldrich), alginic acid
239 (Acros Organics), agarose (Sigma-Aldrich), agar (BD Difco Agar, Noble), iota-carrageenan (TCI),
240 fucoidan from *Macrocystis pyrifera* (Sigma-Aldrich), commercially bought dried red algae
241 (*Porphyra* spp.), commercially bought dried brown algae (*Saccharina japonica*), and the giant kelp
242 (*Macrocystis pyrifera*) harvested from offshore Santa Barbara, California, USA.

243 To test the utilization of several nitrogen sources by strain NLcol2^T, we cultured them with sodium
244 nitrate (1 g/L), ammonium chloride (1 g/L), and without any nitrogen species supplemented in the
245 liquid media. Two sets of tubes with headspace gases of nitrogen gas or helium gas were made as
246 experimental and control group respectively. Triplicate cultures were supplied with 1 g/L D-galactose
247 as substrate and incubated at room temperature (22 °C) for 14 days. Growth was monitored by OD
248 (600 nm) measurements.

249 **2.7 Metabolite analysis from galactose fermentation**

250 To quantify the metabolic products of strain NLcol2^T from galactose fermentation, cultures were
251 grown in triplicates at room temperature (22 °C) with D-galactose as carbon source for 10 days. No
252 ammonium was added in the media and N_2 gas was served as the sole nitrogen source. Growth was

253 monitored by measuring optical density (OD) at 600 nm wavelength. 2mL of culture was subsampled
254 each day (twice a day during exponential phase) for quantification of metabolites.

255 The chromatography protocols used in this study are similar to those previously described (55, 56).
256 Galactose, acetate, succinate, and malate concentrations were measured on an Agilent Infinity 1260
257 (Agilent Technologies, Santa Clara, CA, USA) high-performance liquid chromatograph (HPLC)
258 using an Aminex HPX-87H analytical column (part no. 1250140, Bio-Rad, Hercules, CA, USA)
259 protected by, first, a 0.22 μm physical filter, followed by a Coregel USP L-17 guard cartridge
260 (Concise Separations, San Jose, CA, USA). Separations were performed at 60 °C with a flow rate of
261 0.6 mL/min and a 5 mM sulfuric acid (H_2SO_4) mobile phase. Acetate, succinate, and malate were
262 measured using a variable wavelength detector set to 210 nm, while galactose was measured using a
263 refractive index detector set to 35 °C. Samples and standards for HPLC were acidified to a
264 concentration of 5 mM H_2SO_4, incubated for 5 min at room temperature, and spun at maximum speed
265 in a tabletop centrifuge for 5 min to pellet bacterial cells. The samples were removed from above the
266 cell pellet, and 0.22 μm filtered through a polyethersulfone (PES) membrane into HPLC vials with
267 300 μL polypropylene inserts. Standard curves for each compound of interest were constructed using
268 triplicate standards of 0.1, 0.5, and 1.0 g/L. Peaks were integrated using OpenLab CDS analysis
269 software (version 2.6, Agilent Technologies).

270 Hydrogen gas production was measured on a Fisher Scientific TRACE 1300 Gas Chromatograph
271 (Thermo Fisher Scientific, Waltham, MA) using a TRACE TR- 5 GC Column (part no. 260E113P,
272 Thermo Fisher Scientific) at 30 °C, with an Instant Connect Pulsed Discharge Detector (PDD) (part
273 no. 19070014, Thermo Fisher Scientific) at 150 °C, and ultra-high purity He as a carrier gas. All
274 injections of samples and standards were 100 μL in volume. Supplier-mixed standards of 50 ppm,
275 500 ppm, and 1% hydrogen were run before and after injecting samples, and hydrogen peaks were
276 integrated using Chromeleon Chromatography Data System (CDS) Software (version 6.8, Thermo
277 Fisher Scientific). CO_2 was not considered due to the carbonate-buffered medium and N_2/CO_2
278 atmosphere.

279 A tentative fermentation balance was formulated based on the concentrations of galactose, succinate,
280 acetate, malate, and hydrogen measured above. The changes of concentrations in mmol/L were taken
281 as coefficients for these compounds. The biomass was formulated with a C:N molar ratio of 106:16
282 following the canonical Redfield ratio and the coefficient was determined by the balance of carbon.
283 Nitrogen and ammonium were also included for electron balance and as part of the biomass.

284 **2.8 Metabolite analysis from agar and i-carrageenan degradation**

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

290 Agar and carrageenan concentrations were quantified as polymeric galactose, the main sugar
291 component of the two polymers. Polymeric galactose was quantified as the difference between total
292 galactose and free galactose. To measure total galactose, 5ml of 0.22um-filtered media was acid
293 hydrolyzed to cleave glycosidic linkages and release galactose. Samples were hydrolyzed in 1M HCl
294 at 100°C for 20 hours. Following hydrolysis samples were neutralized by N_2 evaporation and diluted
295 1:1000 with ultrapure water. Galactose was quantified using high performance anion exchange

296 chromatography with pulsed amperometric detection (HPAEC-PAD) on a DIONEX ICS5000+
297 equipped with a CarboPac PA10 column using an isocratic elution of 18mM NaOH for 20 minutes
298 (57). Free galactose was measured by HPAEC-PAD before acid hydrolysis. Incubation media was
299 1:25 diluted with ultrapure water to reduce the salt concentration and quantified using the same
300 gradient program described above.

301 Acetate and succinate concentrations were measured on the Agilent Infinity 1260 HPLC using a
302 similar protocol described above in section 2.7, except using a refractive index detector.

303 2.9 Acetylene reduction assay

304 To test the nitrogenase activity of strain NLcol2^T when growing with nitrogen gas as the sole
305 nitrogen source, acetylene reduction assay was performed following Hardy et al., 1968. In short,
306 acetylene (C_2H_2) can be reduced to ethylene (C_2H_4) when nitrogenases actively fix nitrogen gas at the
307 same time. Cultures were grown on D-galactose in triplicates at 22 °C and triplicate media bottles
308 without inoculation were used as controls. 1.2 mL of acetylene was injected to all culture and control
309 bottles, which contained 80 mL of liquid and 80 mL of headspace pressurized at 150 kPa at the
310 beginning. Gas concentrations and OD_{600} were measured at 6 time points during the 18-day
311 incubation. Acetylene and ethylene concentrations were resolved on a Shimadzu 8A Gas
312 Chromatograph with a flame ionization detector (GC-FID). 1.5 mL samples and standards were
313 injected, then carried by N_2 at a flow rate of 20 mL/min through an n-octane on Res-Sil C packed
314 column (Restek, Centre County, PA, USA) set at 25 °C. 0.5% and 1.0% GASCO calibration gas
315 mixtures of acetylene and ethylene (Cal Gas Direct Incorporated, Huntington Beach, CA, USA) were
316 used for the standard curves.

317 Data Availability

318 Strain NLcol2^T has been deposited at Leibniz-Institut DSMZ (= DSM 113125^T) and Marine Culture
319 Collection of China (= MCCC 1K08672^T). The GenBank accession number for the full-length 16S
320 rRNA gene sequence of strain NLcol2^T is OQ749723, and the genome of strain NLcol2^T was
321 deposited at NCBI under the accession number JARVCO000000000.

322 3 Results and Discussion

323 3.1 Phylogenetic analyses

324 Phylogenetic placement of strain NLcol2^T was determined by comparing full-length 16S rRNA gene,
325 single-copy genes, and whole-genome similarity metrics including average nucleotide identity (ANI)
326 and average amino acid identity (AAI).

327 Strain NLcol2^T was classified within the R76-B128 clade (*Pontiellaceae* family in GTDB database)
328 of the *Verrucomicrobia* phylum under the current SILVA taxonomy (SILVA Ref NR SSU r138.1).
329 Full-length 16S rRNA gene of the isolate shares 84.1%, 88.9%, 92.9%, and 94.5% identity with the
330 four reported cultivated strains in the *Kiritimatiellota* phylum: *Kiritimatiella glycovorans* strain L21-
331 Fru-AB^T, *Tichowtungia aerotolerans* strain S-5007^T, *Pontiella sulfatireligans* strain F21^T, and
332 *Pontiella desulfatans* strain F1^T, respectively. Strain NLcol2^T is more closely related with *P.*
333 *desulfatans* and *P. sulfatireligans* than *K. glycovorans* and *T. aerotolerans*. The 16S rRNA gene
334 identities compared to *P. desulfatans* and *P. sulfatireligans* were absolutely higher than the 86.5%
335 threshold for family level, but fall on the edge of the threshold for a new genus as 94.5% (58). A
336 maximum-likelihood tree of 16S rRNA gene sequences from the *Kiritimatiellota* phylum was

337 reconstructed by RAxML (**Figure S1**). The R76-B128 clade (*Pontiellaceae* family) formed a
338 monophyletic group with MSBL3 clade (*Tichowtungiaceae* family) as the sister group, both of which
339 are in a different cluster from the *Kiritimatiellaceae* family. It is clear that strain NLcol2^T is not
340 affiliated with *K. glycovorans* within the *Kiritimatiellaceae* family nor with *T. aerotolerans* within
341 the *Tichowtungiaceae* family (MSBL3 clade), but belongs to the *Pontiellaceae* family (R76-B128
342 clade) within the *Kiritimatiellales* order as do *P. desulfatans* and *P. sulfatireligans* (15).

343 To resolve the phylogeny of strain NLcol2^T in detail, we further performed genome-level
344 phylogenetic analyses using the Genome Taxonomy DataBase toolkit (47). A concatenated
345 phylogenomic tree was reconstructed from 120 bacterial single-copy genes of genomes in the
346 *Kiritimatiellales* order (**Figure 2**). Here, strain NLcol2^T falls within the *Pontiellaceae* family with a
347 bootstrap value of 100. Additionally, the average amino acid identity (AAI) values of the genomes
348 between strain NLcol2^T and *P. desulfatans* and *P. sulfatireligans* are 69.94% and 68.51%, which are
349 slightly above the threshold of 65% for same genus (59). However, within *Pontiella* genus, it
350 represents a different group from *P. desulfatans* and *P. sulfatireligans*. Moreover, the average
351 nucleotide identity (ANI) values of the genomes between strain NLcol2^T and *P. sulfatireligans* and *P.*
352 *desulfatans* are 72.73% and 73.71% respectively, which was much lower than the 95% ANI criterion
353 for the same species (60, 61). Therefore, we propose that strain NLcol2^T represents a novel species
354 within the *Pontiella* genus according to the phylogenetic analyses above.

355 **3.2 General features of the genome**

356 The draft genome of strain NLcol2^T is 95% complete with 4% redundancy. The genome consists of
357 12 contigs (N50 is 1,265,434 bp) with a total length of 4,436,865 bp and the mean coverage is 593x.
358 DNA G+C content is 52.4 mol%. 5S, 16S and 23S rRNA genes and 50 tRNA genes were found in
359 the genome.

360 3,611 ORFs were predicted by Prodigal, among which 2,757 proteins in the genome were assigned
361 with COG (Cluster of Orthologous Groups) functional category codes. The number of genes in each
362 functional category is shown in **Figure S2**. More genes are involved in carbohydrate (260) and amino
363 acid (188) transport and metabolism than those of nucleotides (65) and lipids (61), which is similar to
364 that in *Kiritimatiella glycovorans* (12). A further detailed analysis of genes involved in macroalgal
365 polysaccharide degradation and nitrogen fixation is presented in sections 3.5 and 3.6.

366 **3.3 Morphologic and chemotaxonomic characterization of strain NLcol2^T**

367 Single colonies on agar plates were white or ivory, circular, and smooth after growing anaerobically
368 for 2 weeks at 22 °C. Bacterial cells of strain NLcol2^T have a round to ovoid shape with a size of 1
369 µm in diameter observed under microscope (**Figure 3**). Cells divide by binary fission and genes of
370 bacterial cell division complex including FtsZ family were present. No motility or flagella were
371 observed, although a full set of genes coding for flagellar assembly was present in the genome. No
372 spore formation was observed. A Gram-negative cell wall structure of outer membrane, periplasmic
373 space and cytoplasmic membrane was shown by electron microscopy (**Figure 3**). There are also
374 genes coding for proteins involved in lipopolysaccharide export and peptidoglycan synthesis in the
375 genome. Some bacteria in the PVC superphylum exhibit compartments inside the cells (62), but like
376 other strains in the *Kiritimatiellota* phylum, no compartmentalization of the cytoplasm was observed
377 in strain NLcol2^T. There were unknown inclusions or granules present inside the cells, and genes
378 involved in the synthesis and utilization of polyphosphate and glycogen were found in the genome,
379 which may serve as phosphate and energy storage materials, respectively.

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

389 Strain NLcol2^T tested negative for both catalase and oxidase, which is common in strict anaerobes
390 (**Table 1**). Beta-galactosidase was tested positive with ~ 5 nanomoles of substrate hydrolyzed in the
391 API Zym assay.

392 **3.4 Physiology of growth**

393 Strain NLcol2^T exhibited consistent growth between 10-37 °C (optimum 31 °C), with NaCl
394 concentration between 10-60 g/L (optimum 25-30 g/L), and with pH 6.0-9.0 (optimum pH 8.0) when
395 D-galactose was utilized as the substrate. It was determined as a mesophilic and neutrophilic
396 bacterium, which is similar to the other four isolated strains from the *Kiritimatiellota* phylum (**Table**
397 **1**). Growth with ammonium supplied in the medium was faster than when dependent on nitrogen
398 fixation. The doubling times are 15 h and 65 h when growing with and without ammonium
399 respectively, at room temperature (22 °C). Strain NLcol2^T was considered as obligately anaerobic,
400 being unable to grow with the presence of oxygen and even in non-reduced medium lacking sulfide
401 as the reducing agent.

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

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

414 $$Galactose \rightarrow Succinate + Acetate + Malate + Hydrogen + \{biomass\}$$

415 $$4.3\ C_6H_{12}O_6 \rightarrow 3.2\ C_4H_6O_4 + 2.4\ C_2H_4O_2 + 0.7\ C_4H_6O_5 + 0.12\ H_2 + \{biomass\}$$

416 **3.5 Anaerobic degradation of macroalgal polysaccharides**

417 **3.5.1 CAZyme analyses**

418 Microbial degradation of macroalgal polysaccharides involves complex metabolic pathways and
419 requires a large number of enzymes during the process (63–66). CAZymes, especially glycoside
420 hydrolases (GHs) and polysaccharide lyases, can break down polysaccharides into oligosaccharides

421 (67). In the genome of strain NLcol2^T, 202 genes (5.6% of predicted ORFs) were predicted to be
422 CAZymes and associated carbohydrate-binding modules (CBM) by dbCAN2 meta server (49) (**Table**
423 **S2**). Among these, 164 genes were annotated to be in the glycoside hydrolase (GH) families. GH2,
424 GH29, GH86 and GH117 are the most abundant families mainly represented by β-galactosidase, α-L-
425 fucosidase, β-agarase and α-1,3-L-neoagarooligosaccharide hydrolase. 100 GHs were predicted with
426 signal peptide sequences indicating 61% of GHs target to the cell membrane or can be exported
427 outside of the cell. Extracellular and membrane associated GHs may hydrolyze large extracellular
428 polymers that cannot otherwise enter the cell. Four porins and nine sugar transporters of the major
429 facilitator superfamily were also present in the genome which may help with the acquisition of
430 carbohydrate molecules by the cell.

431 **3.5.2 Sulfatase analyses**

432 As most marine polysaccharides are sulfated, another group of enzymes called sulfatases are needed
433 in the degradation pathway, which can cleave sulfate ester groups off the carbohydrate backbone
434 (50). It has been shown that *Kiritimatiellota* as well as PVC superphylum have large numbers of
435 copies of sulfatase genes in their genomes (15), and it is the same case in strain NLcol2^T. We found
436 165 sulfatase genes (PF00884), comprising 4.6% of predicted ORFs in the genome.

437 Sulfatases are activated via post-translational modification by other enzymes before functioning. The
438 most common one is formylglycine-generating enzyme (FGE) which transforms a cysteine or serine
439 residue into a catalytic formylglycine (68). These fGly-sulfatases are classified as type I sulfatases
440 (family S1) which contain all carbohydrate sulfatases and is the largest sulfatase family (6).
441 Sulfatases were classified into 22 subfamilies in the SulfAtlas database (50), all of which belongs to
442 family S1 fGly-sulfatases (**Table S3**). The most abundant subfamilies (>5% of total sulfatases) are
443 S1_16, S1_7, S1_15, S1_24, S1_8, S1_19, S1_17 and S1_20. Homologous sulfatases with known
444 enzymatic activities within these subfamilies include D-galactose-6-sulfate 6-O-sulfohydrolase,
445 endo-/exo-xylose-2-sulfate-2-O-sulfohydrolase, endo-/exo-galactose-4-sulfate-4-O-sulfohydrolase,
446 endo-3,6-anhydro-D-galactose-2-sulfate-2-O-sulfohydrolase, exo-fucose-2-sulfate-2-O-
447 sulfohydrolase etc., and the known substrates of these sulfatases include alpha-/iota-/kappa-
448 carrageenan, fucan, ulvan etc. (**Table S3**). These results imply that strain NLcol2^T has the potential to
449 target a vast variety of sulfated polysaccharides, similar to isolates *K. glycovorans* (12), *P.*
450 *desulfatans*, and *P. sulfatireligans* (15, 16). However, due to limited number of characterized fGly-
451 sulfatases, there are still many unknowns about the specific substrates and/or reactions catalyzed by
452 sulfatases in each subfamily (50). 128 sulfatases have the best match genes from organisms in the
453 PVC superphylum and 32 from *Bacteroidota*. 96% of sulfatases (158) were predicted to have a signal
454 peptide sequence, indicating most sulfatases could be membrane-anchored or exported outside of the
455 cell.

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

460 **3.5.3 Growth on macroalgal polysaccharides**

461 We further confirmed the ability of strain NLcol2^T to grow on different macroalgal polysaccharides
462 in live cultures. Bacterial growth was observed in anaerobic cultures with agarose, agar, and iota-
463 carrageenan, but not fucoidan. Many commercially bought algal polysaacharides are contaminated
464 with co-extracted impurities, so we took direct measurements of polysaccharides to confirm the

465 degradation of agar and iota-carrageenan by strain NLcol2^T. Agar and iota-carrageenan
466 concentrations were quantified as polymeric galactose after acid hydrolysis. Polymeric galactose of
467 agar and iota-carrageenan decreased by 88% and 91% respectively, while the fermentation products
468 of succinate and acetate increased by 88% ~ 93% along with bacterial growth (**Figure S4**). The
469 carbon recovery rates are 78% and 87% for agar and iota-carrageenan degradation, respectively. This
470 indicates that strain NLcol2^T is able to degrade agar and iota-carrageenan and their growth were
471 mainly fueled by these polysaccharides but not the impurities. This is the first strain reported with the
472 ability of utilizing agar as substrate in the *Kiritimatiellota* phylum. We further tested their growth on
473 seaweeds and cells also exhibit consistent growth on dried red algae (*Porphyra* spp.) and dried brown
474 algae (*Saccharina japonica*), but not on the giant kelp (*Macrocystis pyrifera*). Since agar, porphyran,
475 and carrageenan are all sulfated polysaccharides extracted mainly from red algae with a similar
476 structure (5, 71), it is not surprising that cells can grow on *Porphyra* spp. directly.

477 Agar is a mixture of agarose and agaropectin which is commonly used as solidifying agent for culture
478 media. Agarose is composed of alternating α-1,3 linked D-galactose and β-1,4 linked 3,6-anhydro-α-
479 L-galactose with little sulfate modification, while agaropectin is heavily modified with sulfate (72–
480 74). Carrageenan is structurally related to agarose, except the β-linked unit is D-galactose-6-sulfate
481 (73). Fucoidan is also a sulfated polysaccharide composed mainly of L-fucose units adorned with
482 sulfate esters, while minor xylose, galactose, mannose, glucuronic acid can be present too (75). Algal
483 polysaccharide degradation has been well studied in *Zobellia galactanivorans* Dsij^T, the marine
484 *Bacteroidota* model for the discovery of agarases, porphyranases, and carrageenases (63, 64, 76). We
485 found potential genes not only involved in the degradation pathways of agar and iota-carrageenan,
486 but fucoidan as well (**Figure 5; Table S4**). Homologous genes encoding for potential β-agarases, ι-
487 carrageenanases, and associated sulfatases were found in the genome of strain NLcol2^T and could be
488 involved in degrading agar and iota-carrageenan into galactose and anhydrogalactose, which then can
489 be directed to the central metabolism for energy. Potential fucoidanases were also found in the
490 genome, but this contrasts with the experimental observation that cells did not grow with fucoidan
491 from *M. pyrifera* as sole carbon source. However, bacterial growth was not supported by L-fucose
492 either (see section 3.4), indicating that strain NLcol2^T may house potential fucoidanases only to
493 remove fucose from fucoidan, but cannot further metabolize fucose and cannot gain energy from
494 fucoidan degradation to support its growth. Alternatively, these genes may not encode fucoidanases
495 to degrade fucoidan, but may encode enzymes for other purposes, for example, removing the fucose
496 cap from mucin-like molecules (77).

497 A neighborhood analysis of the genome shows that GHs and sulfatases are often located nearby
498 (within the distance of five genes), suggesting certain sulfatases and glycoside hydrolases could be
499 regulated together to degrade sulfated polysaccharide (78). In some cases, histidine kinase (PF07730,
500 PF02518), response regulator (PF00072), and TonB-dependent transporters (PF00593, PF03544) are
501 in the neighborhood too. The histidine kinase and response regulator together form a two-component
502 signal transduction system that may help bacteria sense available substrates and respond to the
503 changing environments (79). There are cases when sulfatases themselves cluster together, for
504 example 4 or 6 copies in a row. A complete pathway for assimilatory sulfate reduction is also present
505 in the genome and the cells may utilize the cleaved sulfate group for biosynthesis of reduced sulfur
506 compounds.

507 A comparative study of GHs and sulfatases in selected genomes of the *Kiritimatiellales* order
508 revealed that not all genomes harbor enzymes involved in degradation pathways of agar, iota-
509 carrageenan, and fucoidan, and some bacteria don’t have any GHs or sulfatases at all (**Figure 2;**
510 **Table S5**). However, certain genomes in the *Pontiella* genus show a relatively larger component of

511 GHs and sulfatases. This indicates that these bacteria may adopt the lifestyle of utilizing macroalgal
512 polysaccharides like agar, carrageenan, and fucoidan as carbon and energy sources, while other
513 clades in the *Kiritimatiellales* order may specialize on other substrates available in their living
514 environments. Some genomes in the *Pontiellaceae* family do not have high number of GHs or
515 sulfatases either. This may indicate that these carbohydrate-related genes could be laterally
516 transferred into the *Pontiella* genus but some were lost during evolution living in the environments
517 where other available substrates were preferred. For example, such phenomenon was reported that
518 the lateral gene transfer of porphyranases was from the marine *Bacteroidota*, *Zobellia*
519 *galactanivorans* to the human gut bacterium *Bacteroides plebeius* (80). Another explanation would
520 be that these MAGs were incomplete, and the GHs or sulfatases investigated were not easy to be
521 captured.

522 **3.6 Nitrogen fixation**

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

532 Mo-dependent nitrogenase is the most common and widely studied enzyme that performs nitrogen
533 fixation. It contains two components: an Fe protein as the reductase (*nifH*) collecting and transferring
534 electrons, and a MoFe protein (*nifDK*) which binds dinitrogen (N_2) and converts it to ammonia (NH_3)
535 (81). Genes encoding both nitrogenase iron protein (*nifH*, PF00142) and nitrogenase molybdenum-
536 iron protein alpha and beta subunits (*nifDK*, PF00148) are present in the genome, which together
537 form a complete pathway of nitrogen fixation. No alternative vanadium-iron nitrogenase or iron-only
538 nitrogenase was found. In addition to *nifHDK*, both *nifB* and *nifE* involved in the biosynthesis of
539 nitrogenase MoFe cofactor are present in the genome. Two genes coding for nitrogen regulatory
540 protein PII were present, which are important for the regulation of nitrogen fixation in response to
541 nitrogen source availability (82). The rop-like protein is uncharacterized but often found in nitrogen
542 fixation operons and may play a role in regulation (83). There are various other *nif* genes present in
543 other parts of the genome including *nifA, M, S, U, V* which together may help regulate the function of
544 nitrogenase (**Table S6**).

545 Nitrogenases are highly oxygen-sensitive, but even though there are diverse anaerobes in the PVC
546 superphylum, only a few studies demonstrated nitrogen fixation in this superphylum (84–87) and no
547 reports in the *Kiritimatiellota* phylum. Moreover, we have little knowledge as to where *nif* genes
548 were acquired from by the nitrogen-fixing members in the PVC superphylum. We found 5 genomes
549 in this phylum housing a *nifH* gene. Three were from *P. desulfatans*, *P. sulfatireligans*, and isolate
550 S94, and two were from the marine sediments at the hydrothermal vent of South Mid-Atlantic Ridge
551 (SZUA-380 and SZUA-494). All *nifH* genes in this clade were classified as cluster III, which is
552 dominated by distantly related obligate anaerobes (88). All 6 *nifH* genes from the *Kiritimatiellota*
553 phylum form a monophyletic clade with a bootstrap value of 89 (**Figure S6**). They also cluster
554 together with sequences from *Chlorobi*, *Bacteroidota*, and *Delataproteobacteria* (mainly the
555 *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).

4 Conclusion

In this study, we reported a novel anaerobic bacterial strain NLcol2^{T} 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 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, non-motile 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 10-60 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, iota-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 non-gaseous 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^{T} (= DSM 113125^{T} = MCCC 1K08672^{T}), was isolated from microbial mats on the surface of marine sediments offshore Santa Barbara, California. 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^{T} is OQ749723, and the genome of strain NLcol2^{T} was deposited at NCBI under the accession number JARVCO000000000.

Funding

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).

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

596 on GC-MS, and Professor Craig Carlson at UC Santa Barbara for assistance with polysaccharide
597 measurements. The sequencing was carried out at the DNA Technologies and Expression Analysis
598 Cores at the UC Davis Genome Center. The genomic analysis work used Bridges and Bridges-2 at
599 Pittsburg Supercomputing Center through allocation DEB190007 from the Extreme Science and
600 Engineering Discovery Environment (XSEDE) (89) and the Advanced Cyberinfrastructure
601 Coordination Ecosystem: Services & Support (ACCESS) program (90), which was supported by
602 National Science Foundation grant number #1830033.

603 **References**

604 1. Krause-Jensen D, Duarte CM. 2016. Substantial role of macroalgae in marine carbon
605 sequestration. 10. Nat Geosci 9:737–742.

606 2. National Academies of Sciences E and Medicine. 2022. A Research Strategy for Ocean-based
607 Carbon Dioxide Removal and Sequestration. The National Academies Press, Washington, DC.
608 https://nap.nationalacademies.org/catalog/26278/a-research-strategy-for-ocean-based-carbon-
609 dioxide-removal-and-sequestration.

610 3. Ortega A, Geraldi NR, Alam I, Kamau AA, Acinas SG, Logares R, Gasol JM, Massana R,
611 Krause-Jensen D, Duarte CM. 2019. Important contribution of macroalgae to oceanic carbon
612 sequestration. 9. Nat Geosci 12:748–754.

613 4. Popper ZA, Michel G, Hervé C, Domozych DS, Willats WGT, Tuohy MG, Kloareg B, Stengel
614 DB. 2011. Evolution and Diversity of Plant Cell Walls: From Algae to Flowering Plants. 1.
615 Annu Rev Plant Biol 62:567–590.

616 5. Wei N, Quarterman J, Jin Y-S. 2013. Marine macroalgae: an untapped resource for producing
617 fuels and chemicals. 2. Trends Biotechnol 31:70–77.

618 6. Helbert W. 2017. Marine Polysaccharide Sulfatases. Front Mar Sci 4.

619 7. Glöckner FO, Kube M, Bauer M, Teeling H, Lombardot T, Ludwig W, Gade D, Beck A,
620 Borzym K, Heitmann K, Rabus R, Schlesner H, Amann R, Reinhardt R. 2003. Complete

621 genome sequence of the marine planctomycete Pirellula sp. strain 1. 14. Proc Natl Acad Sci
622 100:8298–8303.

623 8. Martinez-Garcia M, Brazel DM, Swan BK, Arnosti C, Chain PSG, Reitenga KG, Xie G,
624 Poulton NJ, Gomez ML, Masland DED, Thompson B, Bellows WK, Ziervogel K, Lo C-C,
625 Ahmed S, Gleasner CD, Detter CJ, Stepanauskas R. 2012. Capturing Single Cell Genomes of
626 Active Polysaccharide Degraders: An Unexpected Contribution of *Verrucomicrobia*. 4. PLOS
627 One 7:e35314.

628 9. Kim JW, Brawley SH, Prochnik S, Chovatia M, Grimwood J, Jenkins J, LaButti K, Mavromatis
629 K, Nolan M, Zane M, Schmutz J, Stiller JW, Grossman AR. 2016. Genome Analysis of
630 Planctomycetes Inhabiting Blades of the Red Alga *Porphyra umbilicalis*. 3. PLOS One
631 11:e0151883.

632 10. Cardman Z, Arnosti C, Durbin A, Ziervogel K, Cox C, Steen AD, Teske A. 2014.
633 *Verrucomicrobia* Are Candidates for Polysaccharide-Degrading Bacterioplankton in an Arctic
634 Fjord of Svalbard. 12. Appl Environ Microbiol 80:3749–3756.

635 11. Rivas-Marín E, Devos DP. 2018. The Paradigms They Are a-Changin': past, present and future
636 of PVC bacteria research. 6. Antonie van Leeuwenhoek 111:785–799.

637 12. Spring S, Bunk B, Spröer C, Schumann P, Rohde M, Tindall BJ, Klenk H-P. 2016.
638 Characterization of the first cultured representative of *Verrucomicrobia* subdivision 5 indicates
639 the proposal of a novel phylum. 12. ISME J 10:2801–2816.

640 13. Oren A, Garrity GM. 2021. Valid publication of the names of forty-two phyla of prokaryotes.
641 10. Int J Syst Evol Microbiol 71:005056.

642 14. Spring S, Brinkmann N, Murrja M, Spröer C, Reitner J, Klenk H-P. 2015. High Diversity of
643 Culturable Prokaryotes in a Lithifying Hypersaline Microbial Mat. 3–4. Geomicrobiol J 32:332–
644 346.

645 15. van Vliet DM, Palakawong Na Ayudthaya S, Diop S, Villanueva L, Stams AJM, Sánchez-
646 Andrea I. 2019. Anaerobic Degradation of Sulfated Polysaccharides by Two Novel
647 *Kiritimatiellales* Strains Isolated From Black Sea Sediment. Front Microbiol 10:253.

648 16. van Vliet DM, Lin Y, Bale NJ, Koenen M, Villanueva L, Stams AJM, Sánchez-Andrea I. 2020.
649 *Pontiella desulfatans* gen. nov., sp. nov., and *Pontiella sulfatireligans* sp. nov., Two Marine
650 Anaerobes of the *Pontiellaceae* fam. nov. Producing Sulfated Glycosaminoglycan-like
651 Exopolymers. 6. Microorganisms 8:920.

652 17. Mu D-S, Zhou L-Y, Liang Q-Y, Chen G-J, Du Z-J. 2020. *Tichowtungia aerotolerans* gen. nov.,
653 sp. nov., a novel representative of the phylum *Kiritimatiellaeota* and proposal of
654 *Tichowtungiaceae* fam. nov., *Tichowtungiales* ord. nov. and *Tichowtungiia* class. nov. 9. Int J
655 Syst Evol Microbiol 70:5001–5011.

656 18. Ding H, Valentine DL. 2008. Methanotrophic bacteria occupy benthic microbial mats in
657 shallow marine hydrocarbon seeps, Coal Oil Point, California. G1. J Geophys Res:
658 Biogeosciences 113:2007JG000537.

659 19. Eichhubl P, Greene HG, Naehr T, Maher N. 2000. Structural control of fluid flow: offshore
660 fluid seepage in the Santa Barbara Basin, California. J Geochem Explor 69–70:545–549.

661 20. Washburn L, Clark JF, Kyriakidis P. 2005. The spatial scales, distribution, and intensity of
662 natural marine hydrocarbon seeps near Coal Oil Point, California. 4. Mar Pet Geol 22:569–578.

663 21. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
664 BLAST+: architecture and applications. 1. BMC Bioinf 10:421.

665 22. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012.
666 The SILVA ribosomal RNA gene database project: improved data processing and web-based
667 tools. D1. Nucleic Acids Res 41:D590–D596.

668 23. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: Accurate high-throughput multiple sequence
669 alignment of ribosomal RNA genes. 14. Bioinformatics 28:1823–1829.

670 24. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated
671 alignment trimming in large-scale phylogenetic analyses. 15. Bioinformatics 25:1972–1973.

672 25. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
673 large phylogenies. 9. Bioinformatics 30:1312–1313.

674 26. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new
675 developments. W1. Nucleic Acids Res 47:W256–W259.

676 27. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
677 data. 15. Bioinformatics 30:2114–2120.

678 28. Joshi N, Fass J. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for
679 FastQ files (Version 1.33) [Software]. (1.33). C.

680 29. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node
681 solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. 10.
682 Bioinformatics 31:1674–1676.

683 30. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. 4. Nat Methods
684 9:357–359.

685 31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
686 1000 Genome Project Data Processing Subgroup. 2009. 1000 Genome Project Data Processing
687 Subgroup. 2009. The sequence alignment/map format and samtools. Bioinformatics 25:2078–
688 2079.

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

691 33. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the
692 quality of microbial genomes recovered from isolates, single cells, and metagenomes. 7.
693 Genome Res 25:1043–1055.

694 34. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal:
695 prokaryotic gene recognition and translation initiation site identification. 1. BMC Bioinf 11:119.

696 35. Eddy SR. 2011. Accelerated Profile HMM Searches. 10. PLOS Comput Biol 7:e1002195.

697 36. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ,
698 Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn
699 RD. 2019. The Pfam protein families database in 2019. D1. Nucleic Acids Res 47:D427–D432.

700 37. Haft DH. 2001. TIGRFAMs: a protein family resource for the functional identification of
701 proteins. 1. Nucleic Acids Res 29:41–43.

702 38. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang H-Y, El-
703 Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F,

704 Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G, Orengo C, Pandurangan AP, Paysan-
705 Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND, Redaschi N, Richardson LJ,
706 Rivoire C, Salazar GA, Sangrador-Vegas A, Sigrist CJA, Sillitoe I, Sutton GG, Thanki N,
707 Thomas PD, Tosatto SCE, Yong S-Y, Finn RD. 2019. InterPro in 2019: improving coverage,
708 classification and access to protein sequence annotations. D1. Nucleic Acids Res 47:D351–
709 D360.

710 39. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer
711 RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z,
712 Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017. CDD/SPARCLE: functional
713 classification of proteins via subfamily domain architectures. D1. Nucleic Acids Res 45:D200–
714 D203.

715 40. Leimbach A. 2016. bac-genomics-scripts: Bovine E. coli mastitis comparative genomics edition
716 (bovine_ecoli_mastitis). Zenodo.

717 41. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM,
718 Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S,
719 Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes.
720 1. BMC Bioinf 4:41.

721 42. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for
722 Functional Characterization of Genome and Metagenome Sequences. 4. J Mol Biol 428:726–
723 731.

724 43. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer:
725 consistent and rapid annotation of ribosomal RNA genes. 9. Nucleic Acids Res 35:3100–3108.

726 44. Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of
727 transfer RNA genes. W1. Nucleic Acids Res 44:W54–W57.

728 45. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and
729 interpretation of large-scale molecular data sets. D1. Nucleic Acids Res 40:D109–D114.

730 46. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK, Paley
731 S, Subhraveti P, Karp PD. 2020. The MetaCyc database of metabolic pathways and enzymes - a
732 2019 update. D1. Nucleic Acids Res 48:D445–D453.

733 47. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify
734 genomes with the Genome Taxonomy Database. 6. Bioinformatics 36:1925–1927.

735 48. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized
736 analyses of microbial genomes and metagenomes
737 https://doi.org/10.7287/PEERJ.PREPRINTS.1900V1.

738 49. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018.
739 dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. W1. Nucleic
740 Acids Res 46:W95–W101.

741 50. Barbeyron T, Brillet-Guéguen L, Carré W, Carrière C, Caron C, Czjzek M, Hoebeke M, Michel
742 G. 2016. Matching the diversity of sulfated biomolecules: Creation of a classification database
743 for sulfatases reflecting their substrate specificity. 10. PLOS One 11:e0164846.

744 51. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, Von
745 Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural
746 networks. 4. Nat Biotechnol 37:420–423.

747 52. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
748 throughput. 5. Nucleic Acids Res 32:1792–1797.

749 53. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2—a
750 multiple sequence alignment editor and analysis workbench. 9. Bioinformatics 25:1189–1191.

751 54. Folch J, Lees M, Stanley GHS. 1957. A simple method for the isolation and purification of total
752 lipides from animal tissues. 1. J Biol Chem 226:497–509.

753 55. Gilmore SP, Lankiewicz TS, Wilken StE, Brown JL, Sexton JA, Henske JK, Theodorou MK,
754 Valentine DL, O'Malley MA. 2019. Top-Down Enrichment Guides in Formation of Synthetic
755 Microbial Consortia for Biomass Degradation. 9. ACS Synth Biol 8:2174–2185.

756 56. Peng X, Wilken StE, Lankiewicz TS, Gilmore SP, Brown JL, Henske JK, Swift CL, Salamov A,
757 Barry K, Grigoriev IV, Theodorou MK, Valentine DL, O'Malley MA. 2021. Genomic and
758 functional analyses of fungal and bacterial consortia that enable lignocellulose breakdown in
759 goat gut microbiomes. 4. Nat Microbiol 6:499–511.

760 57. Engel A, Händel N. 2011. A novel protocol for determining the concentration and composition
761 of sugars in particulate and in high molecular weight dissolved organic matter (HMW-DOM) in
762 seawater. 1. Mar Chem 127:180–191.

763 58. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby
764 J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured
765 bacteria and archaea using 16S rRNA gene sequences. 9. Nat Rev Microbiol 12:635–645.

766 59. Konstantinidis KT, Rosselló-Móra R, Amann R. 2017. Uncultivated microbes in need of their
767 own taxonomy. 11. ISME J 11:2399–2406.

768 60. Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for
769 prokaryotes. 7. Proc Natl Acad Sci 102:2567–2572.

770 61. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput
771 ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. 1. Nat Commun
772 9:5114.

773 62. Santarella-Mellwig R, Franke J, Jaedicke A, Gorjanacz M, Bauer U, Budd A, Mattaj IW, Devos
774 DP. 2010. The Compartmentalized Bacteria of the Planctomycetes-Verrucomicrobia-
775 Chlamydiae Superphylum Have Membrane Coat-Like Proteins. 1. PLOS Biol 8:e1000281.

776 63. Hehemann J-H, Correc G, Thomas F, Bernard T, Barbeyron T, Jam M, Helbert W, Michel G,
777 Czjzek M. 2012. Biochemical and Structural Characterization of the Complex Agarolytic
778 Enzyme System from the Marine Bacterium *Zobellia galactanivorans*. 36. J Biol Chem
779 287:30571–30584.

780 64. Ficko-Blean E, Préchoux A, Thomas F, Rochat T, Larocque R, Zhu Y, Stam M, Génicot S, Jam
781 M, Calteau A, Viart B, Ropartz D, Pérez-Pascual D, Correc G, Matard-Mann M, Stubbs KA,
782 Rogniaux H, Jeudy A, Barbeyron T, Médigue C, Czjzek M, Vallenet D, McBride MJ, Duchaud
783 E, Michel G. 2017. Carrageenan catabolism is encoded by a complex regulon in marine
784 heterotrophic bacteria. 1. Nat Commun 8:1685.

785 65. Sichert A, Corzett CH, Schechter MS, Unfried F, Markert S, Becher D, Fernandez-Guerra A,
786 Liebeke M, Schweder T, Polz MF, Hehemann J-H. 2020. Verrucomicrobia use hundreds of
787 enzymes to digest the algal polysaccharide fucoidan. 8. Nat Microbiol 5:1026–1039.

788 66. Reisky L, Préchoux A, Zühlke M-K, Bäumgen M, Robb CS, Gerlach N, Roret T, Stanetty C,
789 Larocque R, Michel G, Song T, Markert S, Unfried F, Mihovilovic MD, Trautwein-Schult A,

790 Becher D, Schweder T, Bornscheuer UT, Hehemann J-H. 2019. A marine bacterial enzymatic
791 cascade degrades the algal polysaccharide ulvan. 8. Nat Chem Biol 15:803–812.

792 67. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. 2022. The carbohydrate-
793 active enzyme database: functions and literature. D1. Nucleic Acids Res 50:D571–D577.

794 68. Appel MJ, Bertozzi CR. 2015. Formylglycine, a post-translationally generated residue with
795 unique catalytic capabilities and biotechnology applications. 1. ACS Chem Biol 10:72–84.

796 69. Berteau O, Guillot A, Benjdia A, Rabot S. 2006. A New Type of Bacterial Sulfatase Reveals a
797 Novel Maturation Pathway in Prokaryotes *. 32. J Biol Chem 281:22464–22470.

798 70. Wagner M, Horn M. 2006. The *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* and sister phyla
799 comprise a superphylum with biotechnological and medical relevance. 3. Curr Opin Biotechnol
800 17:241–249.

801 71. Zhang Q, Li N, Liu X, Zhao Z, Li Z, Xu Z. 2004. The structure of a sulfated galactan from
802 *Porphyra haitanensis* and its in vivo antioxidant activity. 1. Carbohydr Res 339:105–111.

803 72. Araki C. 1956. Structure of the agarose constituent of agar-agar. 4. Bull Chem Soc Jpn 29:543–
804 544.

805 73. Anderson NS, Dolan TCS, Rees DA. 1965. Evidence for a common structural pattern in the
806 polysaccharide sulphates of the rhodophyceae. 4976. Nature 205:1060–1062.

807 74. Duckworth M, Yaphe W. 1971. The structure of agar: Part I. Fractionation of a complex
808 mixture of polysaccharides. 1. Carbohydr Res 16:189–197.

809 75. Cong Q, Chen H, Liao W, Xiao F, Wang P, Qin Y, Dong Q, Ding K. 2016. Structural
810 characterization and effect on anti-angiogenic activity of a fucoidan from *Sargassum fusiforme*.
811 Carbohydr Polym 136:899–907.

812 76. Hehemann J-H, Kelly AG, Pudlo NA, Martens EC, Boraston AB. 2012. Bacteria of the human
813 gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from
814 extrinsic microbes. 48. Proc Natl Acad Sci 109:19786–19791.

815 77. Glover JS, Ticer TD, Engevik MA. 2022. Characterizing the mucin-degrading capacity of the
816 human gut microbiota. 1. Sci Rep 12:8456.

817 78. Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H. 2017. Polysaccharide Utilization
818 Loci: Fueling Microbial Communities. J Bacteriol 199:10.1128/jb.00860-16.

819 79. Skerker JM, Prasol MS, Perchuk BS, Biondi EG, Laub MT. 2005. Two-Component Signal
820 Transduction Pathways Regulating Growth and Cell Cycle Progression in a Bacterium: A
821 System-Level Analysis. 10. PLOS Biol 3:e334.

822 80. Hehemann J-H, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. 2010. Transfer of
823 carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. 7290. Nature
824 464:908–912.

825 81. Seefeldt LC, Hoffman BM, Dean DR. 2009. Mechanism of Mo-Dependent Nitrogenase. 1.
826 Annu Rev Biochem 78:701–722.

827 82. Arcondéguy T, Jack R, Merrick M. 2001. P_{II} signal transduction proteins, pivotal players in
828 microbial nitrogen control. 1. Microbiol Mol Biol Rev 65:80–105.

829 83. Buchko GW, Robinson H, Addlagatta A. 2009. Structural characterization of the protein
830 cce_0567 from *Cyanothece 51142*, a metalloprotein associated with nitrogen fixation in the
831 DUF683 family. 4. Biochim Biophys Acta (BBA) - Proteins Proteom 1794:627–633.

832 84. Khadem AF, Pol A, Jetten MSM, Op Den Camp HJM. 2010. Nitrogen fixation by the
833 verrucomicrobial methanotroph *'Methylacidiphilum fumariolicum'* SolV. 4. Microbiology
834 156:1052–1059.

835 85. Wertz JT, Kim E, Breznak JA, Schmidt TM, Rodrigues JLM. 2012. Genomic and Physiological
836 Characterization of the *Verrucomicrobia* Isolate *Diplosphaera colitermitum* gen. nov., sp. nov.,
837 Reveals Microaerophily and Nitrogen Fixation Genes. 5. Appl Environ Microbiol 78:1544–
838 1555.

839 86. Cabello-Yeves PJ, Ghai R, Mehrshad M, Picazo A, Camacho A, Rodriguez-Valera F. 2017.
840 Reconstruction of Diverse Verrucomicrobial Genomes from Metagenome Datasets of
841 Freshwater Reservoirs. Front Microbiol 8:2131.

842 87. Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, McLellan SL, Lücker S, Eren
843 AM. 2018. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in
844 surface ocean metagenomes. 7. Nat Microbiol 3:804–813.

845 88. Zehr JP, Jenkins BD, Short SM, Steward GF. 2003. Nitrogenase gene diversity and microbial
846 community structure: a cross-system comparison. 7. Environ Microbiol 5:539–554.

847 89. Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, Hazlewood V, Lathrop S,
848 Lifka D, Peterson GD, Roskies R, Scott JR, Wilkins-Diehr N. 2014. XSEDE: Accelerating
849 Scientific Discovery. 5. Comput Sci Eng 16:62–74.

850 90. Boerner TJ, Deems S, Furlani TR, Knuth SL, Towns J. 2023. ACCESS: Advancing innovation:
851 NSF's advanced cyberinfrastructure coordination ecosystem: Services & support, p. 173–176.
852 *In* Practice and Experience in Advanced Research Computing. Association for Computing
853 Machinery, New York, NY, USA.

854

855 **Tables**

856 **Table 1.** Comparison of phenotypic characteristics between strain NLcol2^T and four other isolated
857 strains in the *Kiritimatiellota* phylum. Notations: NA, data not available. +, positive; -, negative; +/-,
858 unstable, ceasing growth upon the second transfer. Data for strains other than NLcol2^T were
859 referenced from literatures: a) van Vliet et al., 2020 (16); b) Spring et al., 2016 (12); c) Mu et al.,
860 2020 (17). * Substrates were D-galactose for strain NLcol2^T and D-glucose for other strains. ** For
861 strain S-5007^T, acetate production was predicted from genomic data.

Strains	*P. agarivorans* NLcol2^T	*P. desulfatans* F1$^{T\ a)}$	*P. sulfatireligans* F21$^{T\ a)}$	*K. glycovorans* L21-Fru-AB$^{T\ b)}$	*T. aerotolerans* S-5007$^{T\ c)}$
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*-C12:0, *i*-C18:0	C18:0, *i*-C12:0, *i*-C14:0	C18:0, *i*-C12:0, *i*-C18:0	*i*-C14:0, C18:0	C18:0, *i*-C12:0, *i*-C18:0, *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
iota-Carrageenan	+	-	+	+/-	NA
Xylan	+	-	-	NA	NA
Agar	+	-	-	-	-
Major non-gaseous fermentation products *	Succinate, acetate, malate	Acetate, ethanol, lactate	Acetate, ethanol, lactate	Ethanol, acetate	Acid (maybe acetate **)
Nitrogen sources	N_2, NH_4^+	NH_4^+	NH_4^+	NH_4^+	NH_4^+

862