



OPEN

# Genome characteristics of *Kordia antarctica* IMCC3317<sup>T</sup> and comparative genome analysis of the genus *Kordia*

Yeonjung Lim<sup>1</sup>, Ilnam Kang<sup>2</sup>✉ & Jang-Cheon Cho<sup>1</sup>✉

The genus *Kordia* is one of many genera affiliated with the family *Flavobacteriaceae* of the phylum *Bacteroidetes*, well known for its degradation of high molecular weight organic matters. The genus *Kordia* currently comprises eight species, type strains of which have been isolated from a diverse range of marine environments. As of this report, four genome sequences have been submitted for cultured strains of *Kordia*, but none are complete nor have they been analyzed comprehensively. In this study, we report the complete genome of *Kordia antarctica* IMCC3317<sup>T</sup>, isolated from coastal seawater off the Antarctic Peninsula. The complete genome of IMCC3317<sup>T</sup> consists of a single circular chromosome with 5.5 Mbp and a 33.2 mol% of G+C DNA content. The IMCC3317<sup>T</sup> genome showed features typical of chemoheterotrophic marine bacteria and similar to other *Kordia* genomes, such as complete gene sets for the Embden–Meyerhof–Parnas glycolysis pathway, tricarboxylic acid cycle and oxidative phosphorylation. The genome also encoded many carbohydrate-active enzymes, some of which were clustered into approximately seven polysaccharide utilization loci, thereby demonstrating the potential for polysaccharide utilization. Finally, a *nosZ* gene encoding nitrous oxide reductase, an enzyme that catalyzes the reduction of N<sub>2</sub>O to N<sub>2</sub> gas, was also unique to the IMCC3317<sup>T</sup> genome.

The phylum *Bacteroidetes* is among the most abundant phyla in the ocean, accounting for an average of approximately 10% of marine surface bacterioplankton cells<sup>1,2</sup>. Marine *Bacteroidetes* specializes in the utilization of particulate and high molecular weight organic matter predominately from micro- or macroalgae<sup>3</sup>, and many studies have shown that *Bacteroidetes* increases in abundance during phytoplankton blooms in coastal waters, occupying more than 50% of the bacterial community at certain times<sup>4,5</sup>. Of the many groups within the phylum *Bacteroidetes*, the family *Flavobacteriaceae* is the most abundant in marine environments<sup>4</sup>, and has been subjected to several cultivation and genome-based studies<sup>6–9</sup>. Currently, more than 100 genera with validly published names have been described within the family *Flavobacteriaceae*.

The genus *Kordia* belonging to the family *Flavobacteriaceae* was proposed by Sohn et al.<sup>10</sup> and currently comprises eight species: *K. ulvae*<sup>11</sup>, *K. algicida*<sup>10</sup>, *K. aquimaris*<sup>12</sup>, *K. zhangzhouensis*<sup>13</sup>, *K. jejudonensis*<sup>14</sup>, *K. periserrulae*<sup>15</sup>, *K. zosteriae*<sup>16</sup>, and *K. antarctica*<sup>17</sup>. The type strains of these *Kordia* species have been isolated from a range of marine habitats, including surface seawater, a connection between the ocean and a freshwater spring, seaweed surfaces<sup>16</sup>, and the digestive tract of a marine polychaete. Despite enriched taxonomic diversity and wide distribution, physiological, ecological, and genomic studies on the genus *Kordia* have yet to be performed in-depth beyond those on the algicidal activity of *K. algicida* strains<sup>10,18–20</sup> and a recent report detailing microdiversity among uncultured *Kordia* species identified by single-amplified genomes from a seawater sample of the Indian Ocean<sup>21</sup>.

Currently, whole-genome sequences of four strains of *Kordia* species are publicly available: *K. algicida*, *K. zhangzhouensis*, *K. jejudonensis*, and *K. periserrulae*. However, all of these genome sequences remain in the draft stage, and the only published genome sequence is that of *K. algicida* via a brief single-page genome announcement<sup>22</sup>. Recently, a high-quality, metagenome-assembled genome belonging to the genus *Kordia* was obtained by sequencing a non-axenic culture of the marine diatom *Skeletonema marinoi*<sup>23</sup>. This genome sequence

<sup>1</sup>Department of Biological Sciences, Inha University, Incheon 22212, Republic of Korea. <sup>2</sup>Center for Molecular and Cell Biology, Inha University, Incheon 22212, Republic of Korea. ✉email: ikang@inha.ac.kr; chojc@inha.ac.kr

has been described as being acquired from the *Kordia* sp. strain SMS9, but according to reports, attempts to obtain a pure culture of strain SMS9 have been unsuccessful.

This report provides the first complete genome sequence of the genus *Kordia* obtained from *K. antarctica* strain IMCC3317<sup>T</sup>, as well as comparative genome analyses with the four existing genomes of other *Kordia* species. Strain IMCC3317<sup>T</sup> is a Gram-negative, chemoheterotrophic, yellow-pigmented, non-motile, flexirubin-negative, facultative anaerobic bacterium that was isolated from a coastal seawater sample at the Antarctic Peninsula<sup>17</sup>. To our knowledge, strain IMCC3317<sup>T</sup> is the only *Kordia* strain isolated from the polar environment thus far. The complete genome sequence of strain IMCC3317<sup>T</sup> confirmed a chemoheterotrophic lifestyle and showed the potential to utilize polysaccharides and synthesize secondary metabolites. Comparative analyses also indicated that the IMCC3317<sup>T</sup> genome contains a *nosZ* gene encoding nitrous oxide (N<sub>2</sub>O), unique to this strain among the *Kordia* species.

## Results and discussion

**General genome features.** Strain IMCC3317<sup>T</sup> was isolated from a coastal seawater sample at King George Island in western Antarctica (62°14' S 58°47' W) using a standard dilution-plating method and was established as *K. antarctica* based on phylogenetic, biochemical, and physiological characterization<sup>17</sup>. General features of strain IMCC3317<sup>T</sup> are summarized in Baek et al.<sup>17</sup> and are also presented in Table 1. Comparison of 16S rRNA gene sequences with other type strains of the genus *Kordia* showed that strain IMCC3317<sup>T</sup> is most closely related to *K. zosteriae* ZO2-23<sup>T</sup> (97.7%, sequence similarity), followed by *K. ulvae* SC2<sup>T</sup> (97.0%), *K. jejudonensis* SSK3-3<sup>T</sup> (96.5%), *K. algicida* OT-1<sup>T</sup> (96.6%), and *K. periserrulae* IMCC1412<sup>T</sup> (96.1%). Since a complete genome of an axenic strain has yet to be reported for the genus *Kordia*, we derived the complete genome of IMCC3317<sup>T</sup> using PacBio sequencing.

The complete genome of strain IMCC3317<sup>T</sup> was obtained by de novo assembly followed by five rounds of polishing and comprised a single circular contig. The total length of the genome was 5,500,985 bp and the DNA G+C content was 33.23 mol% (Fig. 1). A total of 4,761 genes were predicted for the genome via the IMG-ER pipeline, including 4,697 protein-coding genes, nine rRNA genes, and 49 tRNA genes. Among the 4,697 protein-coding genes, 62.7% (2,944) could be assigned a putative function. The major COG categories were translation, ribosomal structure and biogenesis (7.74%), coenzyme transport and metabolism (5.97%), and inorganic ion transport and metabolism (5.92%). Overall information on the sequencing and annotation of the IMCC3317<sup>T</sup> genome is presented in Table 1.

ANI values among the five *Kordia* genomes (including the IMCC3317<sup>T</sup> genome) ranged from 76 to 80%. These ANI values are lower than 95–96%, a widely accepted threshold for prokaryotic species demarcation<sup>24,25</sup>, and therefore are consistent with the fact that these five genomes are from five different species of the genus *Kordia*. When compared using BLASTn, the IMCC3317<sup>T</sup> genome showed similarities to the other four *Kordia* genomes over the entire length of the genome (Fig. 2). However, many genomic regions not similar to the four other *Kordia* genomes were also found for IMCC3317<sup>T</sup>. Metabolic genes that were predicted to be unique to the IMCC3317<sup>T</sup> genome, such as *cysC* and *nosZ* (see below), were found within these regions (Fig. 2). Comparison with other *Kordia* genomes based on the inference of orthologous protein clusters showed that the IMCC3317<sup>T</sup> genome shared around 55% of its protein clusters (2,324 among 4,221) with all other four *Kordia* genomes, while 900 protein clusters were not found in any other *Kordia* genome (Fig. 3).

A long, 50,835 bp gene was predicted in the IMCC3317<sup>T</sup> genome. This 'giant gene' (IMG gene ID: 2713618241; hereafter, ten-digit numbers in the parentheses following gene or protein names are their IMG gene IDs) had a G+C content of ~42 mol% that deviated considerably from the average<sup>26</sup> (Fig. 1). This gene encoded a protein of 16,944 amino acids in length with more than 100 tandemly-repeated, single Pfam domains (PF03160; Calx-beta). Though a specific function could not be assigned to this protein, the prediction of a signal peptide and a TIGR04131 (gliding motility-associated C-terminal) motif indicated the protein might be involved in cell surface function. BLASTp analysis against the NCBI nr database (accessed at Jul, 2020) showed that many similar proteins are found mostly (but not exclusively) in the family *Flavobacteriaceae*, including other *Kordia* species. The top 10 best hit proteins were also composed of a very large number of amino acids (3,864–12,288), suggesting a wide distribution of long-length proteins similar to the one predicted in the IMCC3317<sup>T</sup> genome.

**Carbon metabolism.** The metabolic capacity of strain IMCC3317<sup>T</sup> was analyzed using KEGG pathway maps obtained from BlastKOALA and the IMG-ER annotation. The other four *Kordia* genomes available in IMG were also analyzed by BlastKOALA and used for comparative analyses. Since the completeness of other *Kordia* genomes (99.48–99.59%; calculated by CheckM) was only slightly lower than that of the IMCC3317<sup>T</sup> genome (99.92%), these genome sequences were confident to be used for comparative genome analyses.

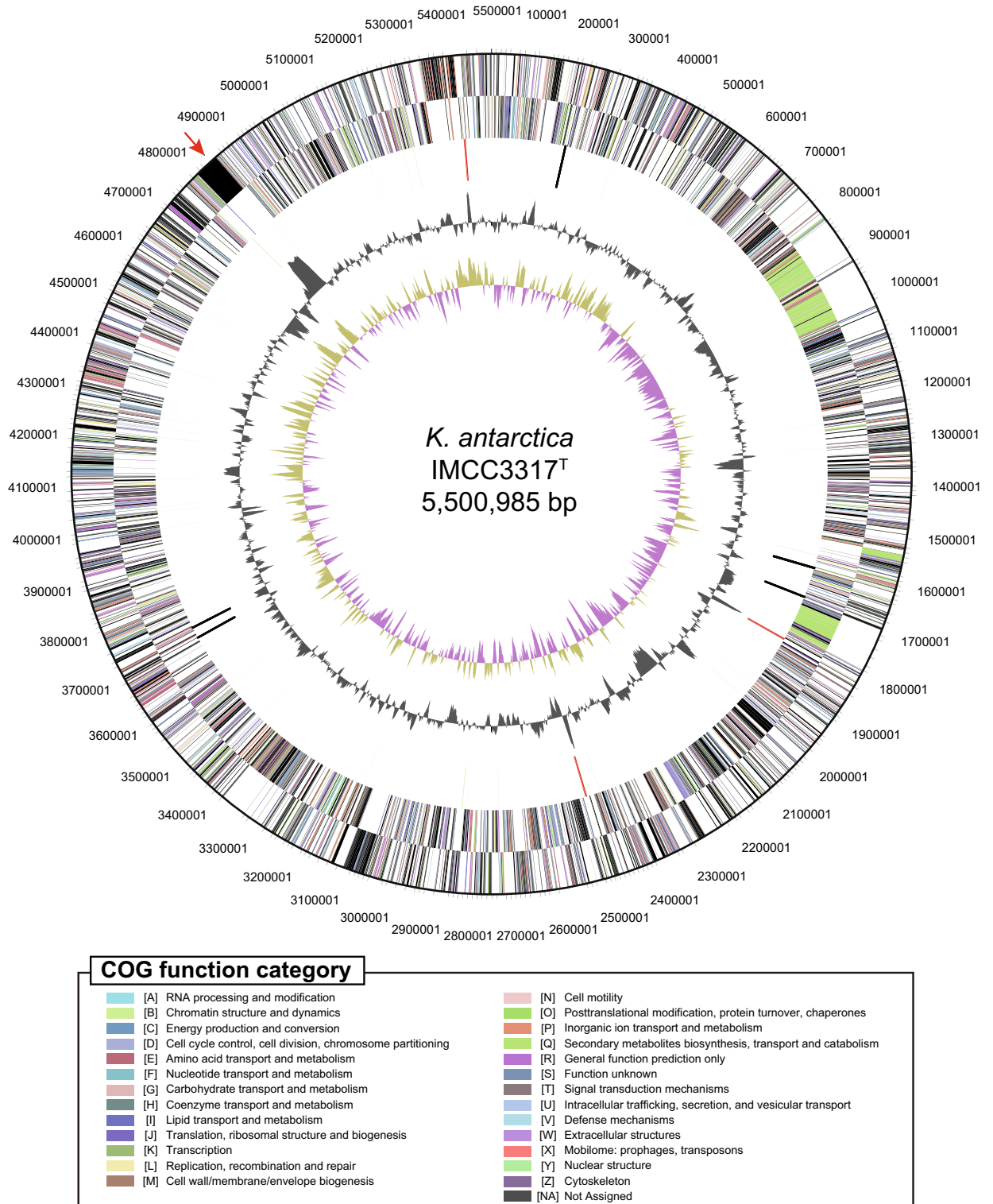
Strain IMCC3317<sup>T</sup> and other four *Kordia* type strains were all predicted to possess complete pathways for central carbon metabolism, such as the Embden–Meyerhof–Parnas (EMP) glycolysis pathway, the tricarboxylic acid (TCA) cycle, and the non-oxidative branch of the pentose phosphate pathway (Table 2). However, strain IMCC3317<sup>T</sup> was distinct from other strains in the number of enzymes it possessed for several steps in these pathways. While the other four strains encoded but a single copy of 6-phosphofructokinase, one of the most important regulatory enzymes of the EMP pathway, IMCC3317<sup>T</sup> encoded two copies thereof, with one copy (2713615308) showing more than 90% similarity to homologs from other strains and the other (2713616029) having with less than 50% similarity to these homologs. In addition, while the other four strains each had one class II fructose-bisphosphate aldolase, IMCC3317<sup>T</sup> encoded the class I fructose-bisphosphate aldolase (2713616027) as well as the class II enzyme. All five strains contained NAD-dependent malate dehydrogenase (EC 1.1.1.37), an enzyme that converts malate to oxaloacetate in the TCA cycle, but strain IMCC3317<sup>T</sup> additionally harbored

Property	Term
<b>General features</b>	
Classification	Domain <i>Bacteria</i>
	Phylum <i>Bacteroidetes</i>
	Class <i>Flavobacteriia</i>
	Order <i>Flavobacteriales</i>
	Family <i>Flavobacteriaceae</i>
	Genus <i>Kordia</i>
	Species <i>Kordia antarctica</i>
	Strain: IMCC3317 <sup>T</sup> (= KCTC 32292 <sup>T</sup> = NBRC 109401 <sup>T</sup> )
Gram stain	Negative
Cell shape	Rod
Motility	Non-motile
Temperature range; optimum	4–42 °C; 20 °C
pH range; optimum	6.0–9.0; 8.0
Salinity	0.5–5.0% NaCl (w/v)
Oxygen requirement	Facultatively anaerobic
<b>Investigation</b>	
Submitted to INSDC	CP019288 (GenBank)
Investigation type	Bacteria_Archaea
Project name	Genome sequencing of <i>Kordia antarctica</i> IMCC3317 <sup>T</sup>
<b>Environment</b>	
Geographic location	King George Island, West Antarctica
Sample collection	2005–2012
Latitude and longitude	62°14' S, 58°47' W
Depth	1 m
Environment (biome)	Marine biome (ENVO: 00000447)
Environment (feature)	Coastal water body (ENVO: 02000049)
Environment (material)	Coastal sea water (ENVO: 00002150)
Isolation and growth conditions	PMID: 23606478
<b>Sequencing</b>	
Sequencing platform	PacBio RS II (P6-C4 chemistry)
Fold coverage	93.32 ×
Assembly method	SMRT Analysis v.2.3.0
Annotation source	IMG/ER
<b>Genome features</b>	
Size	5,500,985
DNA G+C content (%)	33.23
CDS	4,697
rRNA genes	9
tRNA genes	49
CDS assigned to COG	2,057
CDS assigned to Pfam	3,097
CDS assigned to KO	1,408

**Table 1.** General features and genome sequencing information of *Kordia antarctica* strain IMCC3317<sup>T</sup> according to the MIGS recommendations.

a malate dehydrogenase (quinone) (EC 1.1.5.4; 2713614246), a membrane-associated enzyme that can donate electrons to quinone (Table 2).

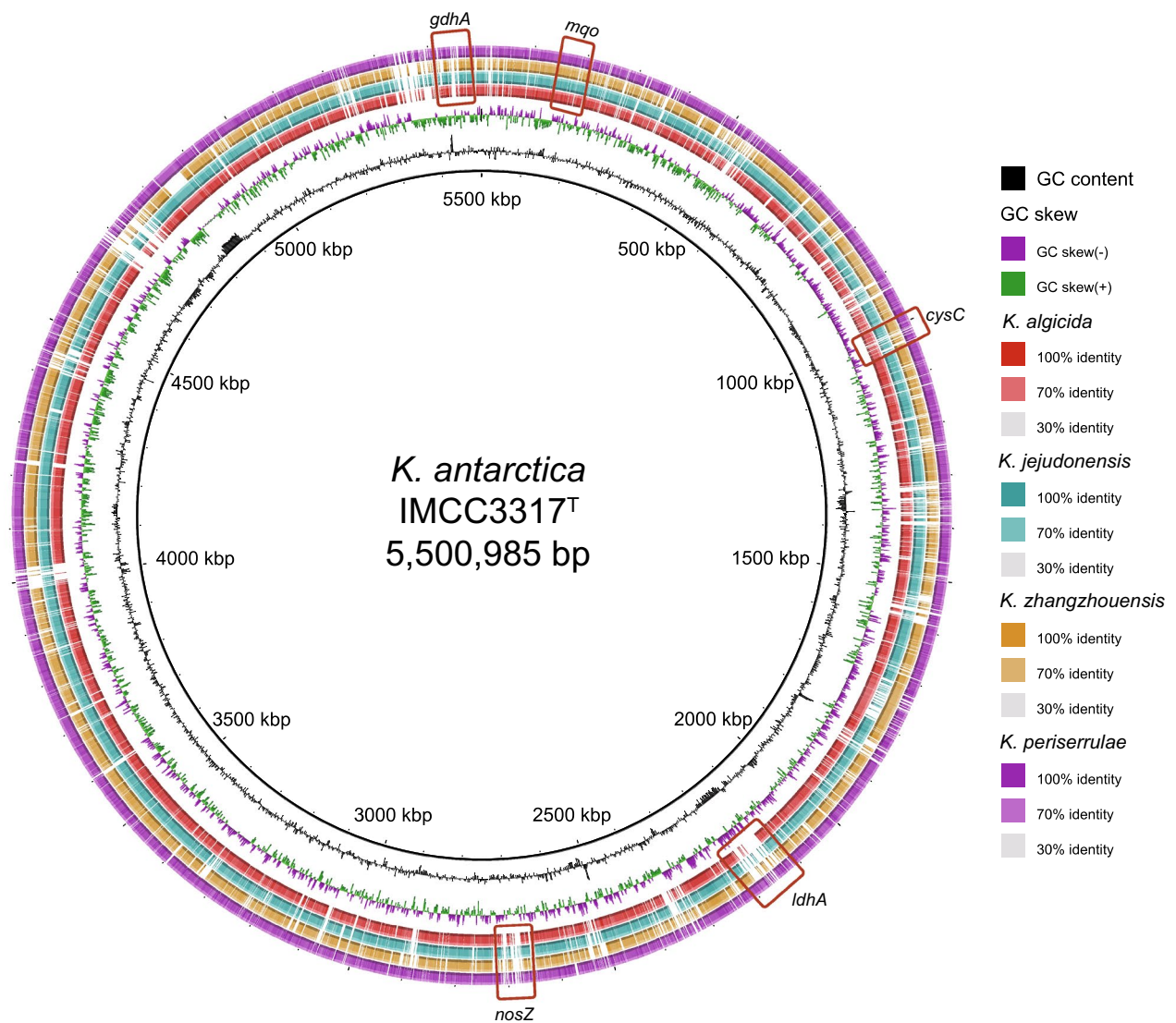
Several differences were found amongst the *Kordia* genomes regarding accessory carbon metabolism. Strain IMCC3317<sup>T</sup> was predicted to lack a glyoxylate cycle, while the other four strains were equipped with isocitrate lyase (EC 4.1.3.1) and malate synthase (EC 2.3.3.9), two key enzymes for this cycle (Table 2). Strains IMCC3317<sup>T</sup> and *K. periserrulae* IMCC1412<sup>T</sup> encoded genes for the Leloir galactose utilization pathway, including galactokinase (EC 2.7.1.6) and galactose-1-phosphate uridylyltransferase (EC 2.7.7.12). Only strain IMCC3317<sup>T</sup> possessed a lactate dehydrogenase (EC 1.1.1.28; 2713616031), which may facilitate fermentation under anaerobic conditions. All five *Kordia* genomes encoded phosphoenolpyruvate carboxylase (EC 4.1.1.31), an anaplerotic enzyme that converts pyruvate to oxaloacetate via the incorporation of CO<sub>2</sub>. However, pyruvate carboxylase (EC 6.4.1.1), another representative anaplerotic enzyme, was not found in any *Kordia* genomes.



**Figure 1.** Circular map of the *Kordia antarctica* IMCC3317<sup>T</sup> genome. From outside to center: Genes on forward strand (colored by COG categories), genes on reverse strand (colored by COG categories). RNA genes: tRNAs, green; rRNAs, red; other RNAs, black. GC content, black; GC skew, purple/olive. Color codes for COG function categories shown below genome map. Position of the ‘giant gene’ indicated by red arrow.

**Nitrogen related metabolism.** All five *Kordia* genomes encoded genes for glutamine synthetase and glutamate synthase for ammonia assimilation (Table 2). All genomes possessed two glutamine synthetase genes of different lengths located next to each other in opposite directions. One group of glutamine synthetase genes

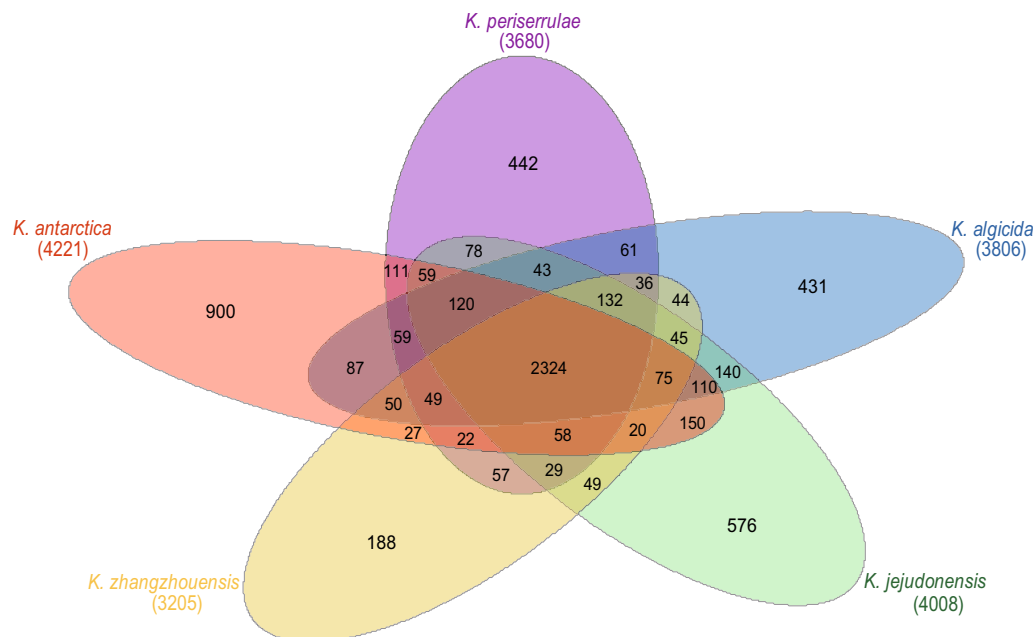




**Figure 2.** Genomic comparisons among *K. antarctica* IMCC3317<sup>T</sup> and other *Kordia* strains *K. algicida*, *K. jejudonensis*, *K. zhangzhouensis* and *K. periserrulae*. Innermost rings show G+C content (black) and GC skew (purple/green) of IMCC3317<sup>T</sup> genome. Other rings show BLASTn-based similarities of IMCC3317<sup>T</sup> genome to other *Kordia* genomes. Relative shading density within each circle represents level of nucleotide identity. Red boxes indicate the approximate locations of genes unique to the IMCC3317<sup>T</sup> genome. Abbreviations: *gdhA*, glutamate dehydrogenase (NAD(P)); *mgo*, malate dehydrogenase (quinone); *cysC*, adenylylsulfate kinase; *ldhA*, D-lactate dehydrogenase; *nosZ*, nitrous oxide reductase.

encoded a type III enzyme of ~ 730 amino acids and the other encoded enzymes of ~ 340 amino acids. In the IMCC3317<sup>T</sup> genome, genes for an ammonia transporter and nitrogen regulatory protein P-II were located just upstream of glutamate synthase. The five *Kordia* genomes also contained genes for glutamate dehydrogenase (GdhA) responsible for another ammonia assimilation pathway that synthesizes glutamate directly from ammonia. There are three types of GdhA, each showing different cofactor specificities: NAD-dependent (EC 1.4.1.2), NADP-dependent (EC 1.4.1.4), and NAD(P)-dependent (EC 1.4.1.3). All *Kordia* genomes possessed the NADP-dependent type (EC 1.4.1.4), but only the IMCC3317<sup>T</sup> genome encoded an additional GdhA annotated as a dual cofactor-specific type<sup>27</sup> (EC 1.4.1.3; 2713618789). Interestingly, a *nosZ* gene was found in the IMCC3317<sup>T</sup> genome, but not in other *Kordia* genomes. *nosZ* encodes a nitrous oxide reductase that reduces nitrous oxide to nitrogen, the final step in the denitrification pathway. No other genes involved in denitrification were found in the IMCC3317<sup>T</sup> genome. A more detailed analysis of the *nosZ* gene found in strain IMCC3317<sup>T</sup> is presented below.

**Sulfur metabolism.** Only strain IMCC3317<sup>T</sup> had a complete assimilatory sulfate reduction pathway wherein sulfate is reduced to sulfide and sequentially incorporated into the biomass, usually via sulfur-containing amino acids such as methionine and cysteine (Table 2). In many bacteria, assimilatory sulfate reduction is accomplished through four steps sequentially mediated by sulfate adenylyltransferase (CysND; EC 2.7.7.4),



**Figure 3.** Venn diagram showing number of shared and unique orthologous protein clusters among five *Kordia* strains: *K. antarctica*, *K. periserrulae*, *K. algicida*, *K. jejudonensis*, and *K. zhangzhouensis*. Total number of protein clusters of each genome indicated in parentheses below strain names.

Metabolic pathway and enzyme	<i>K. antarctica</i>	<i>K. algicida</i>	<i>K. zhangzhouensis</i>	<i>K. jejudonensis</i>	<i>K. periserrulae</i>
<b>Carbon</b>					
Glycolysis/gluconeogenesis	O	O	O	O	O
Citrate cycle (TCA cycle)	O	O	O	O	O
Malate dehydrogenase (quinone) [EC:1.1.5.4]	o	x	x	x	x
Pentose phosphate pathway (non-oxidative)	O	O	O	O	O
Glyoxylate cycle	X	O	O	O	O
<b>Nitrogen</b>					
Ammonia assimilation	O	O	O	O	O
Glutamine synthetase/glutamate synthase	o	o	o	o	o
Glutamate dehydrogenase (NADP) [EC:1.4.1.4]	o	o	o	o	o
Glutamate dehydrogenase (NAD(P)) [EC:1.4.1.3]	o	x	x	x	x
Nitrous oxide reductase [EC:1.7.2.4]	o	x	x	x	x
<b>Sulphur</b>					
Assimilatory sulfate reduction	O	X	X	X	X
Sulfate adenylyltransferase [EC:2.7.7.4]	o	o	o	x	x
Adenylylsulfate kinase [EC:2.7.1.25]	o	x	x	x	x
Phosphoadenylylsulfate reductase (thioredoxin) [EC:1.8.4.8]	o	o	o	x	x
Sulfite reductase (ferredoxin) [EC:1.8.7.1]	o	o	o	x	x
Sulfide:quinone oxidoreductase [EC:1.8.5.4]	x	x	o	x	o
<b>Phosphorus</b>					
High-affinity phosphate ABC transporter [EC:7.3.2.1]	x	x	x	x	x
Alkaline phosphatase [EC:3.1.3.1]	o	o	o	o	o

**Table 2.** Distribution of major metabolic pathways and genes among *K. antarctica* strain IMCC3317<sup>T</sup> and four other *Kordia* strains. Large and small symbols indicate pathways and genes, respectively (presence, O or o; absence, X or x).

adenylylsulfate kinase (CysC; EC 2.7.1.25), phosphoadenylylsulfate reductase (thioredoxin) (CysH; EC 1.8.4.8), and sulfite reductase (ferredoxin) (Sir; EC 1.8.7.1). In the genomes of the strains *K. algicida*, *K. zhangzhouensis*, and *K. antarctica*, a homologous gene cluster containing *cysND*, *cysH*, and *sir* genes, but not the *cysC* gene, was found in the vicinity of methionine biosynthesis-related genes. The genomes of *K. algicida* and *K. zhangzhouensis* did not encode a *cysC* gene, suggesting an incomplete assimilatory sulfate reduction pathway in these strains. In contrast, the IMCC3317<sup>T</sup> genome encoded a *cysC* gene (2713614888) in another gene cluster that contained an additional copy of the *cysND* gene (2713617813, 2713617814), located at the 5' region of a non-ribosomal peptide synthetase (NRPS) biosynthetic gene cluster. This finding indicates that only IMCC3317<sup>T</sup> possesses a complete pathway for assimilatory sulfate reduction, which may confer a competitive advantage when reduced sulfur compounds are scarce. Genes for the Sox pathway of sulfur oxidation that enables sulfur compound utilization as an energy source were not predicted for any of the *Kordia* genomes. However, a sulfide:quinone oxidoreductase that mediates the electron transfer from sulfide to the quinone pool was found in the *K. zhangzhouensis* (2628586580) and *K. periserrulae* genomes (2735934522) (Table 2).

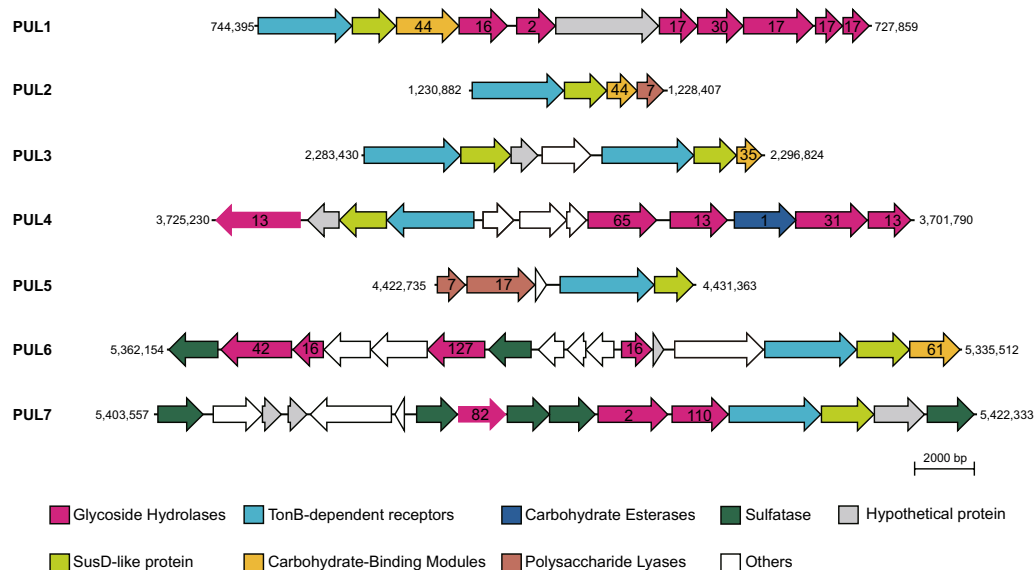
**Phosphorus metabolism.** Regarding phosphorus metabolism, strain IMCC3317<sup>T</sup> and other *Kordia* strains appear to specialize in the utilization of organophosphate but not inorganic phosphate. Many representative marine bacterial groups abundant in pelagic ocean, such as the SAR11 clade and *Prochlorococcus*, possess a high-affinity phosphate ABC transporter (EC 7.3.2.1; PstSCAB), which may be an adaptation to phosphate-depleted oligotrophic pelagic waters. However, all *Kordia* genomes lacked this ABC transporter, implying that the genus *Kordia* has not adapted to phosphate-depleted oligotrophic condition. Instead, all *Kordia* genomes encoded genes for alkaline phosphatase D (PhoD) and PhoPR, a two-component signal transduction system. Alkaline phosphatases (EC 3.1.3.1) can liberate phosphate from organic phosphorus compounds via hydrolysis of phosphoester bonds. Among the three prokaryotic alkaline phosphatase groups known to date, PhoA, PhoX, and PhoD, PhoD was more abundant than the other two groups among global ocean sampling (GOS) metagenomes<sup>28</sup>. Given that all *Kordia* PhoD proteins contain signal peptides, it is likely that *Kordia* strains secrete PhoD into the periplasm or extracellular milieu and subsequently uptake the phosphate released from dissolved organic phosphorus compounds. The possession of secretable alkaline phosphatase suggests that the genus *Kordia* may be adapted to coastal waters replete with organophosphate compounds. Whether and how the expression of PhoD is regulated by the PhoPR system in the *Kordia* strains, however, remains unclear.

**Secondary metabolites.** Analyses by antiSMASH showed that the IMCC3317<sup>T</sup> genome contained nine putative biosynthetic gene clusters (BGCs) with lengths ranging from 21 to 225 kb. These BGCs included a range of clusters, such as NRPS, T1PKS [Type I polyketide synthase (PKS)], T3PKS (Type III PKS), transAT-PKS (Trans-AT PKS), CDPS (tRNA-dependent cyclodipeptide synthase), arylpolyene, lanthipeptide, and terpene. Considering that BGCs are found rarely in marine bacterial groups that have streamlined genomes and are abundant in pelagic waters (e.g. the SAR11 clade), the possession of multiple BGCs that can produce secondary metabolites may be an adaptation to coastal waters where the interactions with other organisms are expected to be more active than pelagic ocean.

The longest BGC, spanning ~ 225 kb (802,427–1,027,059 bp), was identified as a polyketide-NRPS hybrid gene cluster and predicted to synthesize a polymer consisting of at least 30 amino acids (Suppl. Fig. S1). The antiSMASH results showed that gene clusters similar to this BGC were found in certain flavobacterial strains, such as *Flavobacterium spartansii* MSU, *Flavobacterium chilense* DSM 24724, *Flavobacterium johnsoniae* UW101, *Flavobacterium* sp. WG21, and *Kordia zhangzhouensis* MCCC 1A00726, suggesting wide distribution among the family *Flavobacteriaceae*. Of note, however, none of these gene clusters have been fully characterized, leading to a low similarity between this *K. antarctica* BGC and those curated in the MIBiG, a database of BGCs and their products. This *K. antarctica* BGC shared only 3% to 4% of its genes with the most similar BGCs found in the MIBiG database, and there was no homology with the core biosynthetic genes of this BGC, indicating the *K. antarctica* BGC and similar BGCs found in other flavobacterial genomes might synthesize novel secondary metabolites.

**Carbohydrate-active enzymes (CAZymes).** Many CAZymes and SusCD proteins were found in the IMCC3317<sup>T</sup> genome, suggesting a metabolic potential for polysaccharide utilization. Members of the marine *Flavobacteriaceae* specialize in degradation and utilization of polysaccharides produced primarily by algae and phytoplankton<sup>3,29</sup>, making this heterotrophic bacterial group a major contributor to the turnover of dissolved or particulate high-molecular weight organic matter<sup>30</sup>. Recent studies on genomics, transcriptomics, and proteomics of representative marine flavobacterial strains showed that a diverse set of proteins, including transporters, CAZymes, and sulfatases, are critical to polysaccharide metabolism; moreover, the genes encoding these proteins are usually gathered into clusters termed polysaccharide utilization loci (PULs)<sup>8,31</sup>.

As such, we screened and analyzed PULs from the IMCC3317<sup>T</sup> genome, focusing on CAZymes and SusCD, the two major constituents of PULs, to reveal the genomic potential of marine flavobacterial strain IMCC3317<sup>T</sup> for polysaccharide metabolism. The *susC* and *susD* genes, located adjacently in many PULs, encode a TonB-dependent transporter and carbohydrate-binding lipoprotein, respectively. SusC and SusD form a complex in the outer membrane to bind and uptake carbohydrates<sup>32</sup>. CAZymes found in PULs include glycoside hydrolases (GH), glycosyltransferases (GT), carbohydrate esterases (CE), and polysaccharide lyase (PL), and are involved in the further breakdown and metabolism of various carbohydrates. Analysis of the IMCC3317<sup>T</sup> genome using dbCAN showed a total of 194 putative CAZyme proteins, including 53 GH (20 families), 57 GT (12 families), 36 CE (8 families), nine PL (5 families), and 45 carbohydrate-binding modules (14 families). PFAM and TIGRFAM-based searches for SusC and SusD revealed that the IMCC3317<sup>T</sup> genome possessed 10 SusCD pairs. An



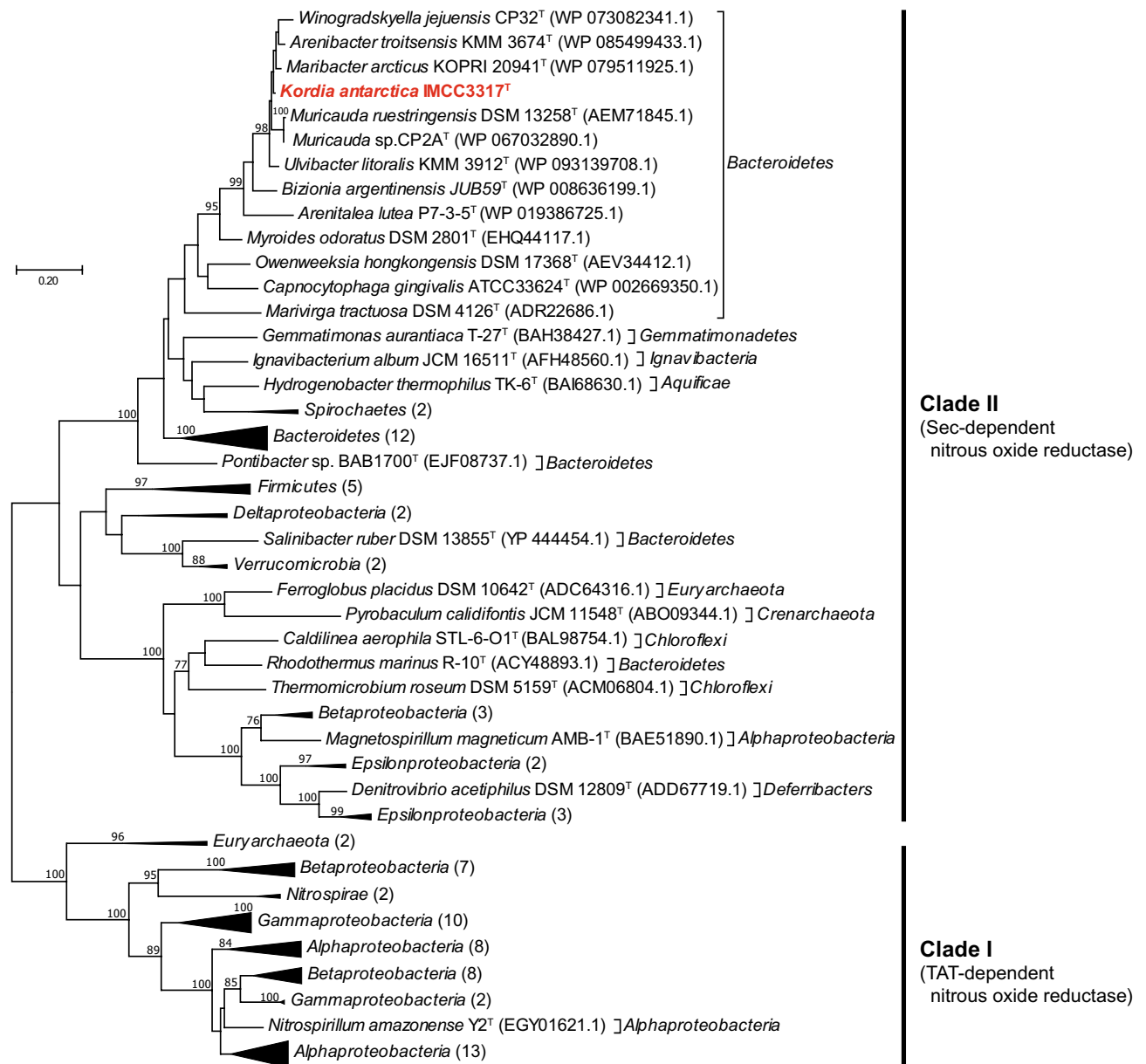
**Figure 4.** Seven polysaccharide utilization loci (PUL) of *K. antarctica* IMCC3317<sup>T</sup> predicted based on eight SusCD pairs and adjacent CAZymes. Genes are colored according to function (color key at bottom). Families of CAZyme genes indicated by numbers within arrows. Numbers on either side of each PUL map indicate start and end positions in the genome. 'Others' refers to annotated genes with functions other than CAZyme, SusCD, and sulfatase.

inspection of genomic regions around these pairs showed that eight SusCD pairs constituted seven PULs together with neighboring CAZymes (Fig. 4). GntR family transcriptional regulators or sulfatases were also found within many PULs. One of the PULs (PUL1, Fig. 4) contained GH families 16, 17, and 30, suggesting the utilization of laminarin<sup>8</sup>. Collectively, the IMCC3317<sup>T</sup> genome has many genes necessary for polysaccharide utilization, which may be regarded as an adaptation to coastal seas where phytoplankton and algae may produce a large amount of polysaccharides.

**Nitrous oxide reductase gene (*nosZ*).** A *nosZ* gene (2713616369) encoding nitrous oxide reductase was found in the IMCC3317<sup>T</sup> genome, but not in the other *Kordia* genomes. The NosZ enzyme localizes to the periplasmic space and is involved in the final step of denitrification; that is, the reduction of nitrous oxide (N<sub>2</sub>O) to nitrogen gas (N<sub>2</sub>)<sup>33</sup>. According to a recent phylogenetic analysis<sup>37</sup>, there are two distinct clades of NosZ, each with different signal peptides related to different translocation pathways. While the clade I NosZ localizes to the periplasm via the twin-arginine translocation pathway, the clade II NosZ is transported via the Sec-dependent translocation pathway. An analysis of signal peptides showed that the NosZ of strain IMCC3317<sup>T</sup> belongs to clade II. Since no other enzymes involved in the denitrification pathway (such as nitrate reductase and nitrite reductase) were found in the IMCC3317<sup>T</sup> genome, like many other prokaryotes with clade II NosZ, IMCC3317<sup>T</sup> was designated as a non-denitrifying N<sub>2</sub>O reducer. It has been suggested that phylogenetically diverse non-denitrifying N<sub>2</sub>O reducers with a clade II NosZ may be important N<sub>2</sub>O sinks since these organisms can consume, but not produce, N<sub>2</sub>O in the environment<sup>34</sup>. From a climate change perspective, given that N<sub>2</sub>O is a major greenhouse gas, at around 300 times more warming potential than CO<sub>2</sub> and implicated in ozone depletion, the existence of non-denitrifying N<sub>2</sub>O reducers, such as IMCC3317<sup>T</sup>, is critical for the conversion of N<sub>2</sub>O to harmless N<sub>2</sub><sup>35</sup>.

The *nosZ* gene was found only in IMCC3317<sup>T</sup>, and analyses were thereby performed to infer the phylogenetic position of the IMCC3317<sup>T</sup> NosZ. In a BLASTp search against the NCBI nr database, top hits for IMCC3317<sup>T</sup> NosZ were predominantly from marine flavobacterial genera phylogenetically related to *Kordia*. The IMCC3317<sup>T</sup> NosZ sequence and several best BLASTp hits were used for the construction of maximum-likelihood trees together with representative NosZ sequences<sup>36,37</sup> and IMCC3317<sup>T</sup> NosZ, and its best BLASTp hits formed a robust branch within the *Bacteroidetes* group of clade II NosZ (Fig. 5). These results suggest that IMCC3317<sup>T</sup> NosZ is similar to that of other phylogenetically-related flavobacterial genera, notwithstanding the absence of NosZ in other species of the same genus. This finding is consistent with previous research indicating that NosZ phylogeny was predominantly shaped by vertical transfer, and gene gain or loss has led to a variable distribution of NosZ even among closely related organisms<sup>37,38</sup>. Further analyses of the family *Flavobacteriaceae* genomes (available in the IMG database) revealed that the variable distribution of NosZ among the same genera was not limited to *Kordia*, but widespread throughout the family *Flavobacteriaceae*. Many genera, including *Tenacibaculum*, *Maribacter*, *Muricauda*, *Aquimarina*, and *Winogradskyella*, had both NosZ-possessing genomes and NosZ-lacking genomes (Suppl. Table S1). More elaborate analyses of the phylogenomics of flavobacterial strains and phylogeny of their NosZ proteins will be necessary to reconstruct the evolutionary history of NosZ in the family *Flavobacteriaceae* and provide a plausible explanation for the distribution of NosZ in *Kordia* among the





**Figure 5.** Maximum-likelihood phylogenetic tree of NosZ proteins from different bacterial groups. NosZ of IMCC3317<sup>T</sup> is bolded in red. Designation of NosZ clades I and II based on Hallin et al.<sup>37</sup> Phylogenetic affiliation at phylum/class level indicated after brackets to the right of taxa names. For visualization purposes, certain NosZ genes from the same phylum/class were grouped together and described as wedges, with the number of grouped proteins indicated in parentheses following phylum/class name. Bootstrap values ( $\geq 70$ ; 100 replicates) indicated above the branches.

following two scenarios: all *Kordia* strains, except for *K. antarctica* IMCC3317<sup>T</sup>, have lost the *nosZ* gene inherited from the common ancestor of the genus *Kordia*; or, the common *Kordia* ancestor lacked the *nosZ* gene and strain IMCC3317<sup>T</sup> acquired it via horizontal gene transfer from closely related organisms.

In many *nosZ*-harboring genomes, *nosZ* is clustered with other genes involved in protein assembly and transport of copper, a cofactor of NosZ. The *nosZ* gene of IMCC3317<sup>T</sup> is located within a gene cluster arranged as *nosZ*, *nosL* (copper chaperone; 2713616367), *nosD* (accessory protein; 2713616366), *nosF* (Cu-processing system ATP-binding protein; 2713616365), and *nosY* (Cu-processing system permease protein; 2713616364). This gene order is highly conserved among *nosZ*-possessing flavobacterial strains closely related to the genus *Kordia*.

## Conclusion

The genome sequence of *K. antarctica* IMCC3317<sup>T</sup> reported in this study represents the first complete, published genome of the genus *Kordia*. Analyses of this complete genome showed that strain IMCC3317<sup>T</sup> shared certain chemoheterotroph metabolic features with other *Kordia* strains. Also, like the two *Kordia* genomes (*K. algicida*

and *K. sp.* SMS9) listed in PULDB<sup>39</sup>, the IMCC3317<sup>T</sup> genome possessed many PULs, suggesting the utilization of polysaccharides originated by marine phytoplankton or algae. Based on the prediction of a *nosZ* gene in the IMCC3317<sup>T</sup> genome in the absence of other denitrification genes, strain IMCC3317<sup>T</sup> is considered a non-denitrifying N<sub>2</sub>O-reducing bacteria. Further studies will be required to explore whether strain IMCC3317<sup>T</sup> actually reduce N<sub>2</sub>O, a significant greenhouse gas in climate change.

## Methods

**Genome sequencing.** Strain IMCC3317<sup>T</sup> was resuscitated by incubating a loopful of 10% glycerol stock on marine agar 2216 plates at 20°C for 7 days. Genomic DNA was extracted from colonies grown on these plates with a DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Genome sequencing of IMCC3317<sup>T</sup> was conducted on the PacBio RS II platform (Pacific Bioscience, Menlo Park, CA, USA) with a 20 kb SMRTbell library generation. De novo assembly was conducted with the SMRT Analysis RS\_HGAP\_Assembly.2 protocol (v2.3.0) using ~38,000 filtered reads (~583 Mbp total; N50 read length, 20,089 bp), resulting in a single contig with approximately ~93× coverage. This single contig was circularized using Circlator (v1.5.5; 'all' command with default options)<sup>40</sup> and polished five times with the SMRT Analysis RS\_Resequencing.1 protocol until no variants were called, generating a final, error-corrected genome sequence. The GenBank accession number of the complete IMCC3317<sup>T</sup> genome is CP019288.

**Genome annotation and comparison.** The complete genome sequence of IMCC3317<sup>T</sup> was submitted to the IMG-ER system<sup>41</sup> for detailed annotation and comparative analyses. Prokka (v1.12)<sup>42</sup> was also used for local annotation (with the options '--rfam' and '--rnammer'). Protein coding sequences, ribosomal RNA genes, transfer RNA genes, and non-coding RNAs were predicted using the tools (implemented in Prokka) Prodigal (v2.6.3), RNAmmer (v1.2), ARAGORN (v1.2), and Infernal (v1.1), respectively. The GenBank file produced by Prokka was submitted to the RAST server for improved functional annotation<sup>43</sup> with classic RAST annotation scheme and 'preserve original gene calls' option. Where necessary, annotation information from Prokka and the RAST server were used for comparison with IMG-ER annotation. HMMs from dbCAN<sup>44</sup> were used to predict and classify CAZymes with default parameters. SusCD proteins were predicted based on the presence of PFAM (for SusD; PF07980, PF12741, PF12771, and PF14322) and TIGRFAM (for SusC; TIGR04056 and TIGR04057) domains. Biosynthetic gene clusters for secondary metabolites were predicted using antiSMASH 5.0<sup>45</sup> with default settings.

Comparative analyses with four other *Kordia* genomes were generally performed using the functionalities provided by the IMG-ER system. All four genomes included in comparative analyses were from the type strains of the genus *Kordia* available in the IMG database: *K. algicida* OT-1 (IMG genome ID; 641380434), *K. zhangzhouensis* MCCC 1A00726 (2627853697), *K. jejudonensis* SSK3-3 (26364156330), and *K. perserrulae* DSM 25731 (2734482288). The completeness of the genome sequences was calculated using CheckM (v1.1.2) with default setting<sup>46</sup>. Average nucleotide identity (ANI) values between genomes were calculated using JSpeciesWS<sup>47</sup>. Strain genome sequences were downloaded and used to draw a plot for genomic comparison between the *K. antarctica* IMCC3317<sup>T</sup> genome and the other four *Kordia* genomes using BRIG<sup>48</sup>. Protein sequences were also downloaded and used for the clustering of homologous proteins among the five *Kordia* genomes by OrthoFinder (v2.2.7)<sup>49</sup> at default settings, and a Venn diagram showing the number of shared or unique protein clusters was generated using the R packages 'limma' and 'gplot'. Protein sequences were also analyzed by BlastKOALA (ran on August 13, 2018)<sup>50</sup> to assign them to KEGG orthology (KO) groups and reconstruct metabolic pathways. A comparative analysis of metabolic pathways was conducted using the KO ID, EC number, COG ID, and Pfam ID of major metabolic enzymes of five *Kordia* genomes in searches against multiple databases, including MetaCyc. When necessary, protein sequences were searched against the NCBI nr database.

**Phylogenetic analysis of NosZ protein.** Protein sequences highly similar to that of the NosZ protein found in the IMCC3317<sup>T</sup> genome were searched for and retrieved from the NCBI RefSeq database using BLASTP 2.8.1+ (ran on September 11, 2018). These sequences were used for phylogenetic analyses together with 109 representative NosZ proteins selected based on previous studies on the distribution and phylogeny of NosZ<sup>37,51</sup>. The collected protein sequences were aligned using Muscle and the aligned sequences were then used for a tree building by the Maximum Likelihood method based on the JTT model using MEGA<sup>752</sup>.

We analyzed the distribution of the *nosZ* gene among members of *Flavobacteriaceae* using genomes publicly available in the IMG database (accessed at Oct, 2018). Genomes classified as *Flavobacteriaceae* in IMG were analyzed using the Function Profile utility. Any genome harboring gene(s) assigned to K00376, a KEGG ortholog corresponding to nitrous oxide reductase (NosZ), was considered to contain a *nosZ* gene. Single-cell genomes, metagenome-assembled genomes, and genomes from bacterial isolates not classified at the genus level were excluded from the analysis.

Received: 3 February 2020; Accepted: 10 August 2020

Published online: 07 September 2020

## References

1. Yilmaz, P. *et al.* Ecological structuring of bacterial and archaeal taxa in surface ocean waters. *FEMS Microbiol.* **81**, 373–385 (2012).
2. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
3. Fernández-Gómez, B. *et al.* Ecology of marine Bacteroidetes: A comparative genomics approach. *ISME J.* **7**, 1026–1037 (2013).

4. Buchan, A., LeCleir, G. R., Gulvik, C. A. & González, J. M. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat. Rev. Microbiol.* **12**, 686–698 (2014).
5. Teeling, H. *et al.* Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. *Elife* **5**, e11888. <https://doi.org/10.7554/eLife.11888> (2016).
6. Hahnke, R. L. & Harder, J. Phylogenetic diversity of Flavobacteria isolated from the North Sea on solid media. *Syst. Appl. Microbiol.* **36**, 497–504 (2013).
7. Hahnke, R. L. *et al.* Dilution cultivation of marine heterotrophic bacteria abundant after a spring phytoplankton bloom in the North Sea. *Environ. Microbiol.* **17**, 3515–3526 (2015).
8. Kappelmann, L. *et al.* Polysaccharide utilization loci of North Sea Flavobacteria as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *ISME J.* **13**, 76–91 (2018).
9. Zhang, H. *et al.* Repeated evolutionary transitions of flavobacteria from marine to non-marine habitats. *Environ. Microbiol.* **21**, 648–666 (2019).
10. Sohn, J. H. *et al.* Kordia algicida gen. nov., sp. nov., an algicidal bacterium isolated from red tide. *Int. J. Syst. Evol. Microbiol.* **54**, 675–680 (2004).
11. Qi, F., Huang, Z., Lai, Q., Li, D. & Shao, Z. Kordiaulvae sp. nov., a bacterium isolated from the surface of green marine algae Ulva sp.. *Int. J. Syst. Evol. Microbiol.* **66**, 2623–2628 (2016).
12. Hameed, A. *et al.* Kordiaaquimaris sp. nov., a zeaxanthin-producing member of the family Flavobacteriaceae isolated from surface seawater, and emended description of the genus Kordia. *Int. J. Syst. Evol. Microbiol.* **63**, 4790–4796 (2013).
13. Du, J. *et al.* Kordia zhangzhouensis sp. nov., isolated from surface freshwater. *Int. J. Syst. Evol. Microbiol.* **65**, 3379–3383 (2015).
14. Park, S., Jung, Y.-T. & Yoon, J.-H. Kordia jejudonensis sp. nov., isolated from the junction between the ocean and a freshwater spring, and emended description of the genus Kordia. *Int. J. Syst. Evol. Microbiol.* **64**, 657–662 (2014).
15. Choi, A., Oh, H.-M., Yang, S.-J. & Cho, J.-C. Kordia periserrulae sp. nov., isolated from a marine polychaete Periserrula leucophryna, and emended description of the genus Kordia. *Int. J. Syst. Evol. Microbiol.* **61**, 864–869 (2011).
16. Kim, D. I., Lee, J. H., Kim, M. S. & Seong, C. N. Kordia zosteriae sp. nov., isolated from the seaweed, Zostera marina. *Int. J. Syst. Evol. Microbiol.* **67**, 4790–4795 (2017).
17. Baek, K., Choi, A., Kang, I., Lee, K. & Cho, J.-C. Kordia antarctica sp. nov., isolated from Antarctic seawater. *Int. J. Syst. Evol. Microbiol.* **63**, 3617–3622 (2013).
18. Paul, C. & Pohnert, G. Interactions of the algicidal bacterium Kordia algicida with diatoms: Regulated protease excretion for specific algal lysis. *PLoS ONE* **6**, e21032. <https://doi.org/10.1371/journal.pone.0021032> (2011).
19. Bigalke, A., Meyer, N., Papanikolopoulou, L. A., Wiltshire, K. H. & Pohnert, G. The algicidal bacterium Kordia algicida shapes a natural plankton community. *Appl. Environ. Microbiol.* **85**, e02779–e12718. <https://doi.org/10.1128/AEM.02779-18> (2019).
20. Saha, M. & Weinberger, F. Microbial, “gardening” by a seaweed holobiont: Surface metabolites attract protective and deter pathogenic epibacterial settlement. *J. Ecol.* **107**, 2255–2265 (2019).
21. Royo-Llonch, M. *et al.* Exploring microdiversity in Novel Kordia sp. (Bacteroidetes) with Proteorhodopsin from the tropical indian ocean via single amplified genomes. *Front. Microbiol.* **8**, 1317. <https://doi.org/10.3389/fmicb.2017.01317> (2017).
22. Lee, H. S., Kang, S. G., Kwon, K. K., Lee, J.-H. & Kim, S.-J. Genome sequence of the algicidal bacterium Kordia algicida OT-1. *J. Bacteriol.* **193**, 4031–4032 (2011).
23. Pinder, M. I. *et al.* Genome sequence of Kordia sp. strain SMS9 identified in a non-axenic culture of the diatom Skeletonema marinoi. *J. Genomics.* **7**, 46–49 (2019).
24. Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. USA* **106**, 19126–19131 (2009).
25. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantindis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 1–8 (2018).
26. Reva, O. & Tümmler, B. Think big—giant genes in bacteria. *Environ. Microbiol.* **10**, 768–777 (2008).
27. Hudson, R. & Daniel, R. L-glutamate dehydrogenases: distribution, properties and mechanism. *Comp. Biochem. Physiol. B.* **106**, 767–792 (1993).
28. Luo, H., Benner, R., Long, R. A. & Hu, J. Subcellular localization of marine bacterial alkaline phosphatases. *Proc. Natl. Acad. Sci. USA* **106**, 21219–21223 (2009).
29. Gómez-Pereira, P. R. *et al.* Genomic content of uncultured Bacteroidetes from contrasting oceanic provinces in the North Atlantic Ocean. *Environ. Microbiol.* **14**, 52–66 (2012).
30. Kirchman, D. L. The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS Microbiol. Ecol.* **39**, 91–100 (2002).
31. Bennke, C. M. *et al.* Polysaccharide utilisation loci of Bacteroidetes from two contrasting open ocean sites in the North Atlantic. *Environ. Microbiol.* **18**, 4456–4470 (2016).
32. Glenwright, A. J. *et al.* Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* **541**, 407–411 (2017).
33. Zumft, W. G. Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* **61**, 533–616 (1997).
34. Sanford, R. A. *et al.* Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. *Proc. Natl. Acad. Sci. USA* **109**, 19709–19714 (2012).
35. Ravishankara, A., Daniel, J. S. & Portmann, R. W. Nitrous oxide (N<sub>2</sub>O): the dominant ozone-depleting substance emitted in the 21st century. *Science* **326**, 123–125 (2009).
36. Graf, D. R., Jones, C. M. & Hallin, S. Intergenomic comparisons highlight modularity of the denitrification pathway and underpin the importance of community structure for N<sub>2</sub>O emissions. *PLoS ONE* **9**, e114118. <https://doi.org/10.1371/journal.pone.0114118> (2014).
37. Hallin, S., Philippot, L., Löffler, F. E., Sanford, R. A. & Jones, C. M. Genomics and ecology of novel N<sub>2</sub>O-reducing microorganisms. *Trends Microbiol.* **36**, 497–504 (2017).
38. Jones, C. M., Stres, B., Rosenquist, M. & Hallin, S. Phylogenetic analysis of nitrite, nitric oxide, and nitrous oxide respiratory enzymes reveal a complex evolutionary history for denitrification. *Mol. Biol. Evol.* **25**, 1955–1966 (2008).
39. Terrapon, N. *et al.* PULDB: The expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res.* **46**, D677–D683 (2017).
40. Hunt, M. *et al.* Circlator: Automated circularization of genome assemblies using long sequencing reads. *Genome Biol.* **16**, 294 (2015).
41. Chen, I.-M.A. *et al.* IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* **45**, D507–D516 (2016).
42. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
43. Aziz, R. K. *et al.* The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics.* **9**, 75 (2008).
44. Yin, Y. *et al.* dbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, W445–W451 (2012).
45. Blin, K. *et al.* antiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47**, 81–87 (2019).
46. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
47. Richter, M., Rosselló-Móra, R., Oliver Glöckner, F. & Peplies, J. JSpeciesWS: Web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* **32**, 929–931 (2015).

48. Alikhan, N.-F., Petty, N. K., Zakour, N. L. B. & Beatson, S. A. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genomics* **12**, 402 (2011).
49. Emms, D. M. & Kelly, S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
50. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).
51. Zumft, W. G. & Kroneck, P. M. Respiratory transformation of nitrous oxide (N<sub>2</sub>O) to dinitrogen by Bacteria and Archaea. *Adv. Microb. Physiol.* **52**, 107–227 (2006).
52. Kumar, S., Stecher, G. & Tamura, K. Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

## Acknowledgements

This research was supported by the Collaborative Genome Program of the Korea Institute of Marine Science and Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries (No. 20180430), and also by grants from the National Research Foundation of Korea (NRF). (NRF-2016R1A6A3A11934789 and NRF-2019R1I1A1A01063401 to IK and NRF-2018R1A5A1025077 and NRF-2019-R1A2B5B02070538 to JCC).

## Author contributions

Y.L., I.K., and J.-C.C. conceived the experiments and designed the work. Y.L. performed experiments, analyzed data, prepared tables and figures, and wrote the draft manuscript. I.K. and J.-C.C. supervised data analysis and critically curated the manuscript. I.K. and J.-C.C. are responsible for the publication. All authors contributed to manuscript revision and read and approved the submitted version.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-71328-9>.

**Correspondence** and requests for materials should be addressed to I.K. or J.-C.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020