


ARTICLE

DOI: 10.1038/s42003-018-0086-7

OPEN

Genomic evidence for the degradation of terrestrial organic matter by pelagic Arctic Ocean Chloroflexi bacteria

David Colatrisano¹, Patricia Q. Tran¹, Celine Guéguen², William J. Williams³, Connie Lovejoy ^{4,5} & David A. Walsh¹

The Arctic Ocean currently receives a large supply of global river discharge and terrestrial dissolved organic matter. Moreover, an increase in freshwater runoff and riverine transport of organic matter to the Arctic Ocean is a predicted consequence of thawing permafrost and increased precipitation. The fate of the terrestrial humic-rich organic material and its impact on the marine carbon cycle are largely unknown. Here, a metagenomic survey of the Canada Basin in the Western Arctic Ocean showed that pelagic Chloroflexi from the Arctic Ocean are replete with aromatic compound degradation genes, acquired in part by lateral transfer from terrestrial bacteria. Our results imply marine Chloroflexi have the capacity to use terrestrial organic matter and that their role in the carbon cycle may increase with the changing hydrological cycle.

¹Department of Biology, Concordia University, 7141 Sherbrooke St. West, Montreal, QC H4B 1R6, Canada. ²Department of Chemistry and School of the Environment, Trent University, 1600 West bank Drive, Peterborough, ON K9J 7B8, Canada. ³Fisheries and Oceans Canada, Institute of Ocean Sciences, 9860 West Saanich Road, Sidney, BC V8V 4L1, Canada. ⁴Département de biologie, Institut de Biologie Intégrative et des Systèmes (IBIS) and Québec-Océan, Université Laval, Laval, QC G1K 7P4, Canada. ⁵Takuvik Joint International Laboratory, Centre National de la Recherche Scientifique (France, CNRS UMI 3376), Université Laval, Laval, QC G1V 0A6, Canada. Correspondence and requests for materials should be addressed to D.A.W. (email: david.walsh@concordia.ca)

The Arctic Ocean accounts for 1.4% of global ocean volume but receives 11% of global river discharge¹. Up to 33% of the dissolved organic matter in the Arctic Ocean is of terrestrial origin and a major fraction of this terrestrial dissolved organic matter (tDOM) originates from carbon-rich soils and peatlands^{2,3}. With thawing permafrost and increased precipitation occurring across the Arctic⁴, increases in freshwater runoff and riverine transport of organic matter to the Arctic Ocean are predicted, which will increase tDOM fluxes and loadings^{5,6}. The additional tDOM may represent new carbon and energy sources for the Arctic Ocean microbial community and contribute to increased respiration, which would result in the Arctic being a source of dissolved inorganic carbon to the ocean. Alternatively, as it moves from its source of origin to the Arctic Ocean tDOM could become more recalcitrant to bacterial metabolism and represent a long-term sequestration of the newly released carbon making the Arctic more carbon neutral^{7,8}. However, an estimated 50% of Arctic Ocean tDOM is removed before being released to the Atlantic, at least in part by microbial processes⁹. As input of tDOM increases, knowledge on its microbial transformation will be critical for understanding changes in Arctic carbon cycling.

The marine SAR202 is a diverse and uncultivated clade of Chloroflexi bacteria that comprise roughly 10% of planktonic cells in the dark ocean^{10–14}. SAR202 is also common in marine sediments and deep lakes^{15–17}. It has long been speculated that SAR202 may have a role in the degradation of recalcitrant organic matter^{11,14}, and the recent analysis of SAR202 single-cell-amplified genomes (SAGs) lends support to this notion¹⁸. More generally, Chloroflexi, including those in the SAR202 clade, are also present in the upper layers of the Arctic Ocean¹⁹, leading to the hypothesis that recalcitrant organic compounds present in high Arctic tDOM could be utilized by this group.

Results

In this study, we analyzed Chloroflexi metagenome-assembled genomes (MAGs) generated from samples collected from the vertically stratified waters of the Canada Basin in the Western Arctic Ocean (Fig. 1a). A metagenomic co-assembly was generated from samples originating from the surface layer (5–7 m), the subsurface chlorophyll maximum (25–79 m) and a layer corresponding to the terrestrially-derived DOM fluorescence (FDOM) maximum previously described within the cold Canada Basin halocline comprised of Pacific-origin waters (177–213 m)²⁰. The Pacific-origin FDOM maximum is due to sea ice formation and interactions with bottom sediments on the Beaufort and Chukchi shelves, which themselves are influenced by coastal erosion and river runoff²⁰. Binning based on tetranucleotide frequency and coverage resulted in 360 MAGs from a diversity of marine microbes (Fig. 1b). Six near-complete Chloroflexi MAGs were identified. Based on 16S rRNA gene phylogeny, these MAGs represented three distinct SAR202 subclades (SAR202-II, -VI, -VII), the AncK29 clade and the TK10 clade (Fig. 2a). Estimated MAG completeness ranged from 77 to 99%, while contamination ranged from 0 to 2.3% (Table 1). All MAGs exhibited highest coverage just below the subsurface chlorophyll maximum (Fig. 2b) which is consistent with earlier findings on SAR202 distribution in the North Pacific Ocean¹². However, the concentration and composition of the FDOM maximum in the Canada Basin is significantly different compared to the North Pacific Ocean²¹ and the North Atlantic Subtropical Gyre²². A concatenated protein phylogeny demonstrated that the SAR202 MAGs were distinct from previously published MAGs from the deep ocean¹⁸ and oxygen minimum zones²³ (Supplementary Figure 1). Fragment recruitment of 21 TARA Ocean metagenomic datasets spanning epipelagic to mesopelagic waters at 7

locations and 4 separate bathypelagic metagenomes indicated that the Canada Basin Chloroflexi MAGs were not widely distributed in the oceans (Fig. 2c, Supplementary Data 1). These findings are evidence that the Chloroflexi MAGs represent genotypes that are rare outside Arctic marine waters.

The Chloroflexi MAGs contained many genes implicated in the degradation of aromatic compounds typically associated with humic-rich tDOM (Supplementary Data 2). A single MAG (SAR202-VII-2) from a previously undescribed clade (SAR202-VII) exhibited a striking enrichment in these genes (Fig. 3a). Partial pathways for the catabolism of aromatic compounds were recently reported from deep ocean SAR202 SAGs¹⁸. To assess whether the abundance and diversity of SAR202-VII-2 genes involved in aromatic compound catabolism is unique to Arctic Ocean MAGs or is a more broad characteristic of marine Chloroflexi, we compared gene content between SAR202-VII-2 and two SAGs (SAR202-V-AB-629-P13 and SAR202-III-AAA240-O15) reported in Landry et al.¹⁸ Of the 117 SAR202-VII-2 orthologs implicated in aromatic compound degradation, 12 were identified in SAR202-III-AAA240-O15 and only one was identified in SAR202-V-AB-629-P13, implying distinct and less diverse pathways in deep ocean SAR202 compared to the Arctic Ocean populations (Supplementary Data 2).

Proteins for the modification and degradation of monoaryl and biaryl compounds were predicted, including a diversity of aromatic ring-cleaving dioxygenases^{24–26}. A total of 42 ring-cleaving dioxygenases targeting compounds related to catechol, protocatechuate and gentisate were present in the six MAGs, with 25 dioxygenases predicted in SAR202-VII-2 alone (Fig. 3a, b). Ring demethylation, hydroxylation, and decarboxylation are important prerequisite steps to prime diverse aromatic compounds for downstream oxidative cleavage^{27,28}. Thirty ring-demethylating monooxygenases, ten ring-hydroxylating dioxygenases, and eleven ring-decarboxylases were annotated in the SAR202-VII-2 MAG (Fig. 3c, Supplementary Data 2). Proteins involved in the conversion of ring-cleavage products to central intermediates of the citric acid cycle were also present in the SAR202-VII-2 MAG, including dehydrogenases (i.e., 2,3-dihydroxy-2,3-dihydrophenylpropionate dehydrogenase), decarboxylases (i.e., oxaloacetate B-decarboxylase), aldolases (i.e., HMG aldolase and 4-carboxymuconolactone decarboxylase), hydratases (i.e., 4-oxalmaleonate hydratase and 2-oxopent-4-enoate hydratase), isomerases (i.e., mycothiol maleulpyruvate isomerase and muconolactone isomerase) and hydrolases (i.e., 3-oxoadipate enol-lactonase and β -keto adipate enol-lactone hydrolase) (Supplementary Data 2, Supplementary Figure 2). We note that we were unable to identify a single complete reference pathway for humic-like aromatic compound degradation. Since estimated genome completeness for SAR202-VII-2 was 99%, it is unlikely the genes were missed due to an incomplete genome. Another explanation is that marine Chloroflexi genomes encode novel pathway variants. Indeed, numerous metal-dependent hydrolases, hydrolases of the HAD family and NAD(P)-dependent dehydrogenase were clustered in genomic regions with the ring-modifying oxygenases, decarboxylases, and demethylases described above. In addition to the array of aromatic compound degradation genes, the SAR202-VII-2 MAG also contained 34 copies of the flavin mononucleotide (FM)/F420-dependent monooxygenase catalytic subunit (FMNO) proteins previously implicated in activation of recalcitrant organic compounds in the deep ocean¹⁸. These results are consistent with Chloroflexi in the Arctic Ocean having the metabolic potential to access carbon and energy available in aromatic compounds typically associated with tDOM.

The diversity of SAR202-VII-2 genes implicated in aromatic compound degradation lead us to hypothesize that they may have

originated by lateral gene transfer from terrestrial bacteria. To test this, we targeted aromatic compound degradation genes (the ring-cleaving dioxygenases, specifically) in the Chloroflexi MAGs for in-depth phylogenetic analyses. The genomic diversity of

marine Chloroflexi was expanded in our analysis by including 130 Chloroflexi MAGs recently assembled and binned from the TARA Oceans project²⁹. A number of the SAR202-VII-2 ring-cleaving dioxygenase homologs were most closely related to

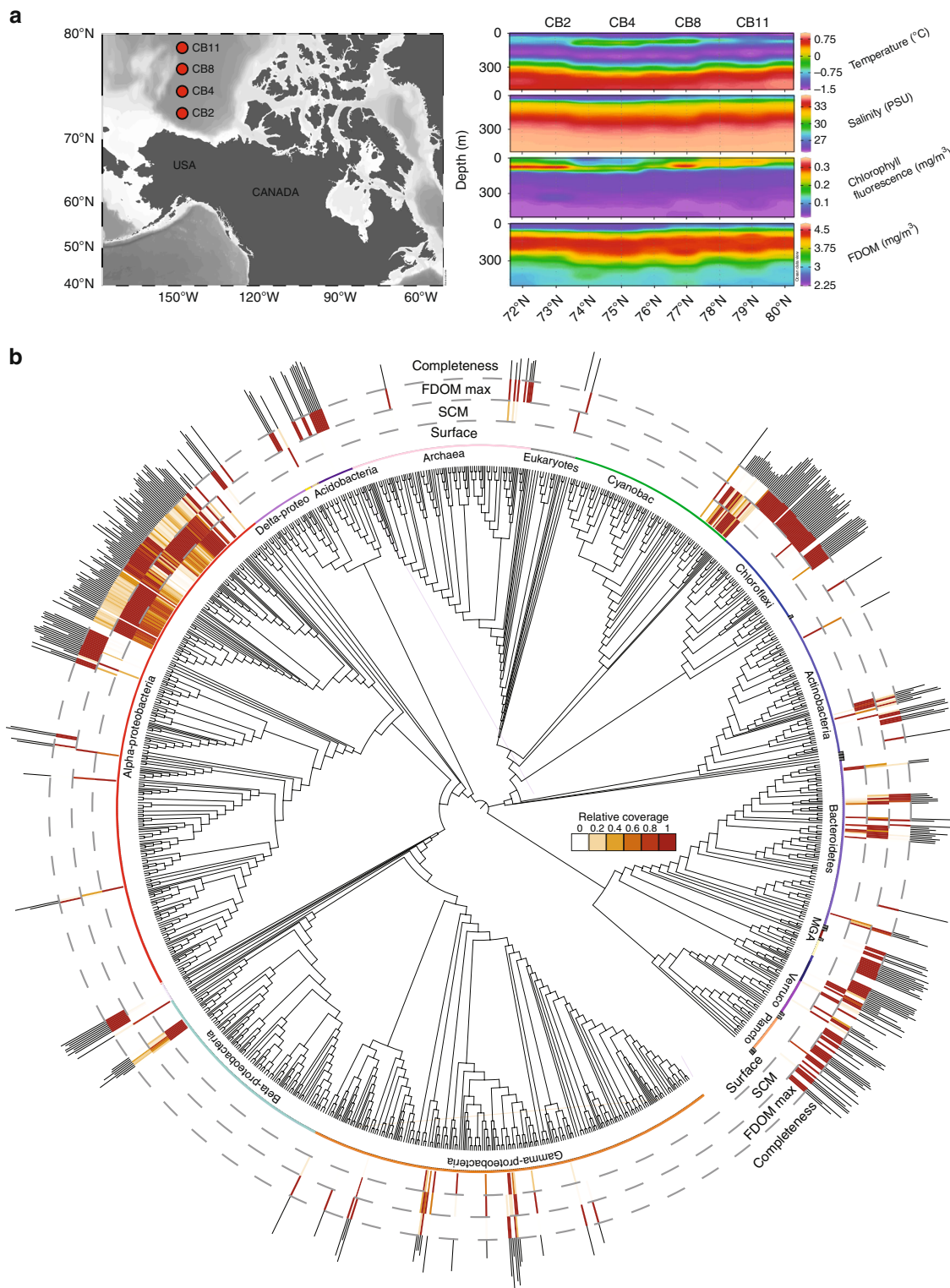


Fig. 1 Metagenomic survey of microbial diversity in the Canada Basin. **a** Sampling locations and environmental profiles of the Canada Basin. Sample locations and the associated environmental datasets were plotted using Ocean Data View version 4.7.7⁴². **b** Concatenated protein phylogeny of 360 Arctic Ocean MAGs inferred by MetaWatt and visualized in iTOL. The three inner tracks present relative coverage of MAGs averaged across samples collected from surface waters, subsurface chlorophyll maximum (SCM) and the fluorescent dissolved organic matter maximum (FDOM max). The outer track presents estimated MAG completeness as inferred by MetaWatt. MAG completeness ranged from 25 to 94%

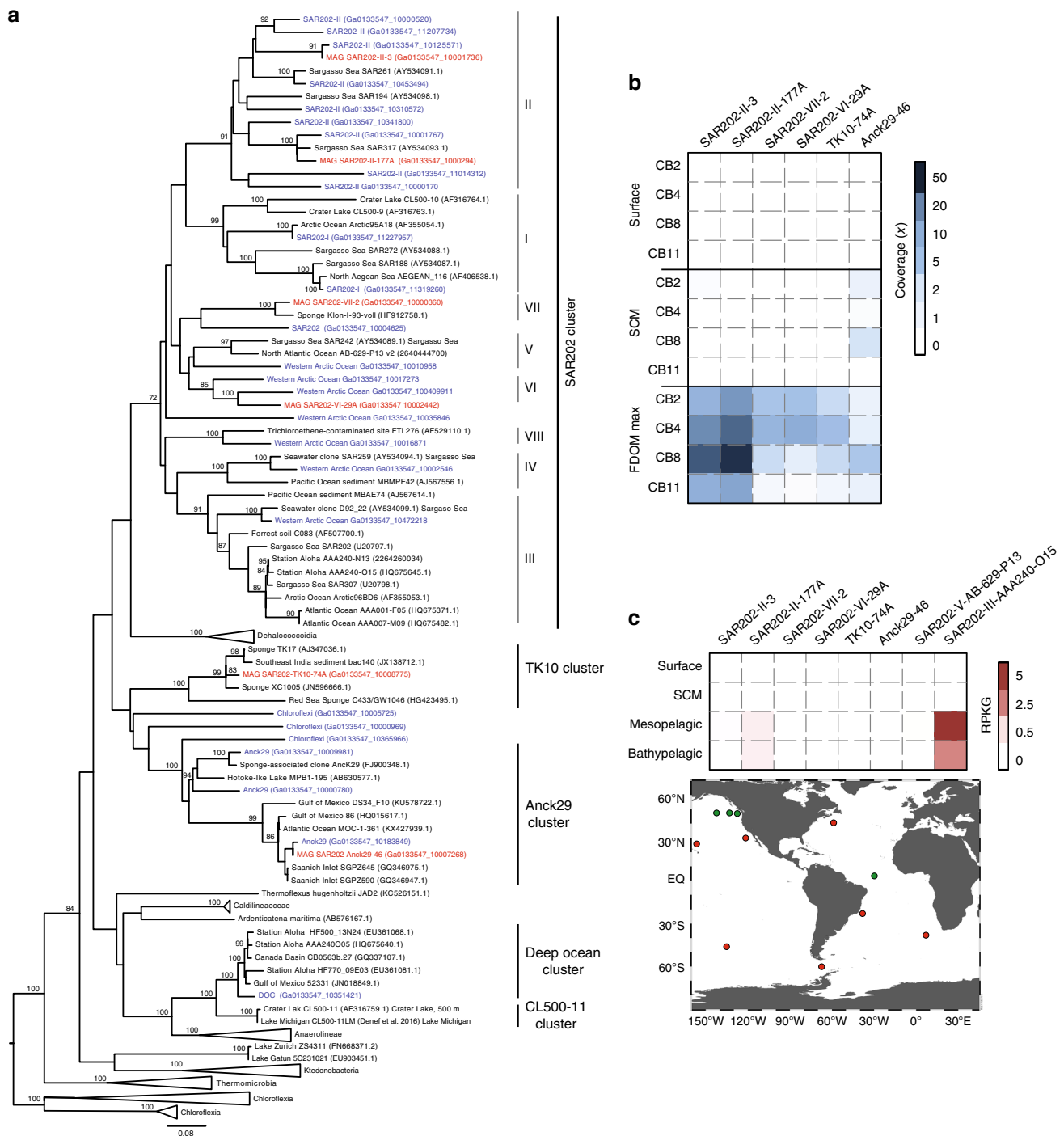


Fig. 2 Diversity and distribution of Arctic Ocean Chloroflexi MAGs. **a** Maximum likelihood tree of Chloroflexi based on partial 16S rRNA gene sequences. Blue taxa are from Canada Basin metagenomes and red taxa are from the six Chloroflexi MAGs. Bootstrap values of >70% are shown (100 replicates). **b** Distribution of MAGs in the Canada Basin based on metagenome coverage at the surface, subsurface chlorophyll maximum (SCM) and fluorescent dissolved organic matter maximum (FDOM max). **c** Distribution of MAGs in global ocean metagenomes based on fragment recruitment. Two deep ocean SAR202 SAGs from Landry et al.¹⁸ were included for comparison. Location of TARA ocean metagenomes (red) and bathypelagic metagenomes (green) are shown on the map generated using Ocean Data View version 4.7.7⁴²

proteins from the TARA Ocean Chloroflexi MAGs and other marine originating genomes, particularly the catechol dioxygenases (Supplementary Figure 3), gentisate 1,2-dioxygenases (Fig. 4) and methylgallate dioxygenases (Supplementary Figure 4) indicating that aromatic compound degradation in Chloroflexi is not restricted to the Arctic Ocean. However, lateral gene transfer from terrestrial bacteria was also evident. For example, an

annotated gentisate dioxygenase gene was positioned within a clade of terrestrial Actinomycetes (Fig. 4). Additional genes involved in the degradation of structures related to catechol, protocatechuate and gentisate were phylogenetically associated with homologs from terrestrial Acidobacteria (Supplementary Figure 5), Actinobacteria (Supplementary Figure 6), Armatimonadetes (Fig. 4), Delta-proteobacteria (Supplementary Figures 3

Table 1 Genomic characteristics of MAGs

	Size (Mb)	Cov (x)	GC (%)	Completeness (%)	Contamination (%)	N50 (kb)	# of Contigs
SAR202-II-3	1.36	68	39	80	0	35	46
SAR202-II-177A	1.62	101	42	82	1	36	52
SAR202-VII-2	2.78	16	59	99	0	246	8
SAR202-VI-29A	1.52	16	46	97	2	59	2
TK10-74A	2.45	12	69	81	2.3	38	76
Anck29-46	1.11	11	32	77	0	75	22

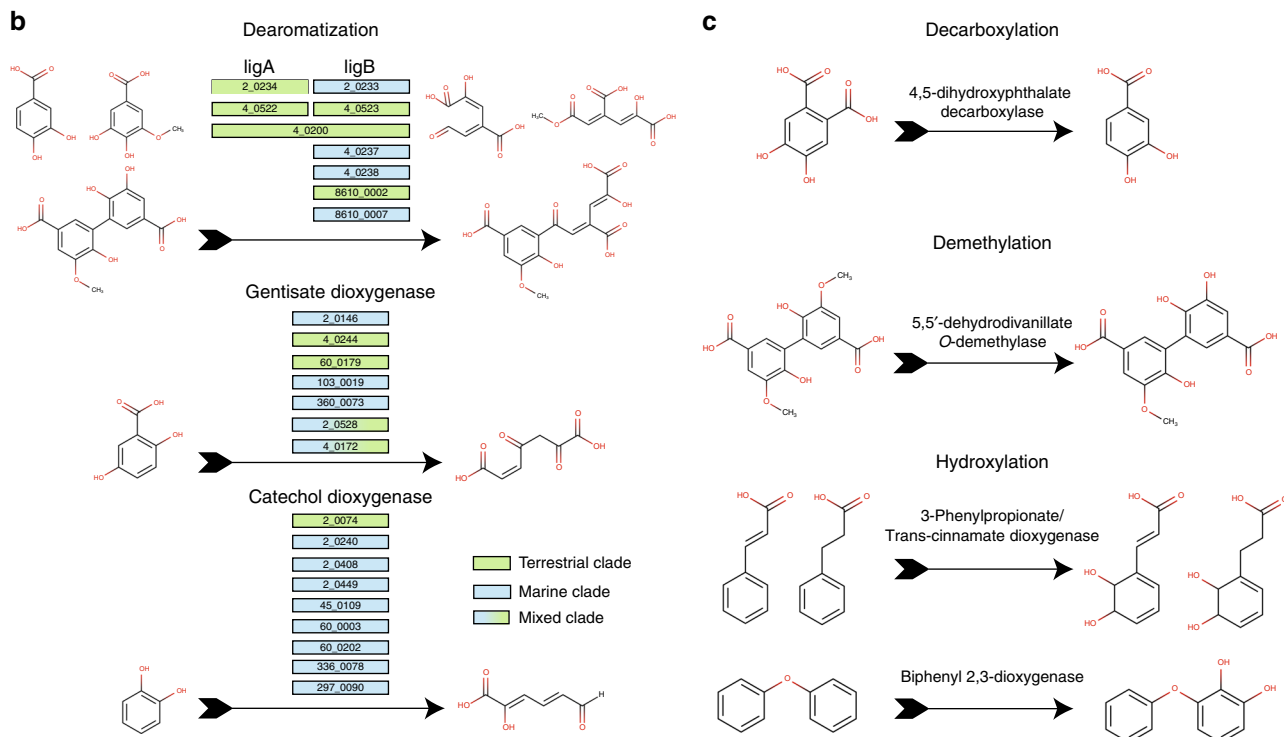
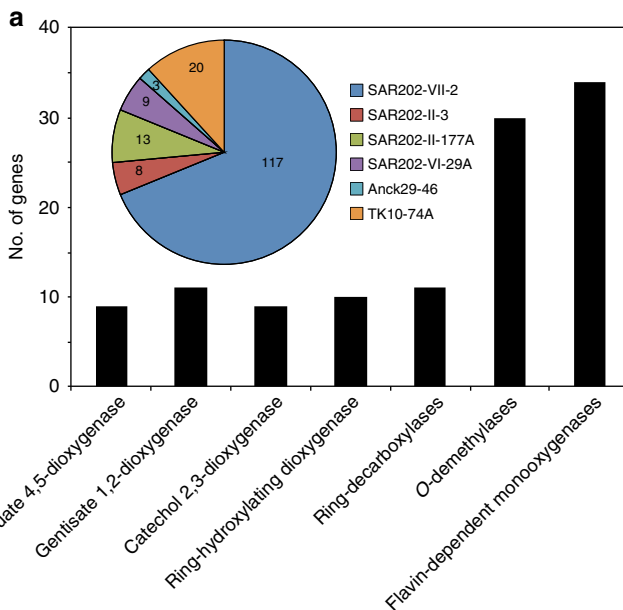


Fig. 3 Aromatic compound degradation genes and pathways in Arctic Ocean Chloroflexi MAGs. **a** Abundance of aromatic compound degradation genes in Chloroflexi MAGs from the Canada Basin (pie chart) and a breakdown of those found specifically in SAR202-VII-2 (column plot). **b** Predicted aromatic ring-opening enzymatic reactions identified in SAR202-VII-2 with gene loci displayed in the colored boxes. Genes in green and blue boxes were most closely related to homologs from terrestrial or marine bacteria, respectively. Genes in the blue/green boxes were in clades containing diverse environmental bacteria. **c** Examples of predicted aromatic ring-modifying enzymatic reactions identified in SAR202-VII-2

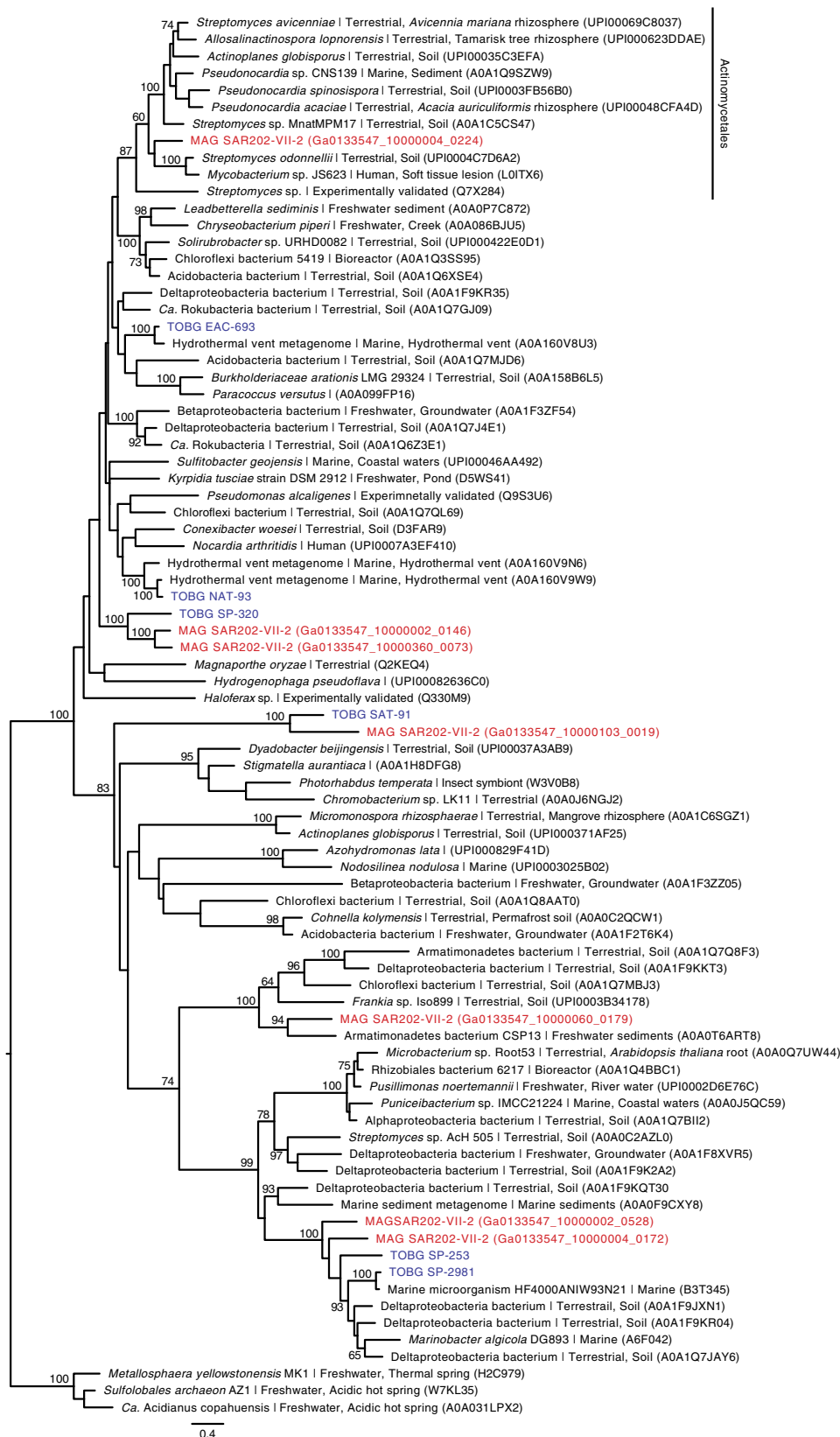


Fig. 4 Maximum likelihood tree of predicted gentisate 1,2-dioxygenases. Bootstrap values of >60% are shown (100 replicates). Homologs from SAR202-VII-2 are highlighted in red and homologs from TARA Ocean MAGs are highlighted in blue

and 6), Beta-proteobacteria (Supplementary Figure 7) and a clade of diverse terrestrial phyla (Supplementary Figure 8). Additionally, 2 gentisate 1,2-dioxygenase genes and 1 protocatechuate dioxygenase *ligB* gene were phylogenetically associated to a clade of genes from both terrestrial Delta-proteobacteria and marine microbes (Fig. 4 and Supplementary Figure 8). These putative gene acquisitions were unlikely due to contaminating scaffolds because the genes were located on long scaffolds that were assigned to Chloroflexi with high confidence based on tetranucleotide frequencies and the phylogenetic identity of housekeeping genes. Such a phylogenetic pattern supports the hypothesis that marine Chloroflexi acquired the capacity for aromatic compound degradation, at least in part, by lateral gene transfer from terrestrial bacteria.

Conclusions

In total, these results are consistent with Chloroflexi having a role in tDOM transformation in waters of the Arctic Ocean. This is the first study to our knowledge to associate a specific microbial group with tDOM metabolism in the Arctic Ocean and it expands on recent studies contributing to our understanding of the metabolic diversity of the abundant yet uncultivated marine Chloroflexi^{18,23}. Moreover, lateral gene transfer from terrestrial bacteria appears to have contributed to the evolution of aromatic compound degradation capabilities within marine Chloroflexi, particularly in regions of the Arctic Ocean impacted by terrestrial input.

The majority of MAGs were restricted to the humic-rich Pacific-origin halocline of the Canada Basin, however it is the surface waters that will be most immediately affected by increased freshwater input¹. Hence, our initial observations suggest a need for further research on the distribution of tDOM-utilizing microbes in other Arctic water masses with an aim to establish how common and phylogenetically widespread tDOM metabolism is in the Arctic Ocean. These water masses could include coastal surface waters at the mouth of the Mackenzie River, as well as regions of differing DOM composition such as the East Siberian Sea²⁰. Moreover, metagenomic studies such as this are, in essence, hypothesis-generating and future work that includes targeted cultivation, in situ gene expression analysis, and rate measurement-based approaches are required to validate and quantify microbial metabolic contributions to nutrient cycling. Overall, it is likely that marine Chloroflexi have the capacity to degrade tDOM, and their role in the Arctic carbon cycle may increase as Arctic warming leads to greater inputs of terrestrial organic matter.

Methods

Sampling and DNA extraction. Twelve samples for metagenomics were collected in September 2015 during the Joint Ocean Ice Study cruise to the Canada Basin. For each sample, 4–8 L of seawater was sequentially filtered through a 50 µm pore mesh, followed by a 3 µm pore size polycarbonate filter and a 0.22 µm pore size Sterivex filter (Durapore; Millipore, Billerica, MA, USA). Filters were preserved in RNAlater and stored at –80 °C until processed in the laboratory. DNA was extracted from the Sterivex filter using the following method: filters were thawed on ice and RNAlater was removed. The Sterivex was then rinsed twice with a sucrose-based lysis buffer, and filled with 1.8 mL of the lysis buffer. Filters were treated with 100 µL of 125 mg mL⁻¹ lysozyme and 20 µL of 10 µg mL⁻¹ RNase A and left to rotate at 37 °C for 1 h. After incubation, 100 µL of 10 mg mL⁻¹ proteinase K and 100 µL of 20% SDS was added. Filters were left to rotate for 2 h at 55 °C. Lysate was removed from the filters. Protein was precipitated and removed with 0.583 volumes of MPC Protein Precipitation Reagent (Epicentre, Madison, WI, USA) and centrifugation at 10,000×g at 4 °C for 10 min. The supernatant was transferred to a clean tube. DNA was precipitated with cold isopropanol, and resuspended in low TE buffer.

Metagenomic sequencing, assembly, annotation, and binning. DNA sequencing of 12 samples was performed at the Department of Energy Joint Genome Institute (Walnut Creek, CA, USA) on the HiSeq 2500-1TB (Illumina) platform.

Paired-end sequences of 2 × 150 bp were generated for all libraries. A metagenome co-assembly of all raw reads was generated using MEGAHIT³⁰ with kmer sizes of 23,43,63,83,103,123. Gene prediction and annotation was performed using the DOE Joint Genome Institute's Integrated Microbial Genomes (IMG) database tool (version 4.11.0)³¹. Metagenomic binning was performed on scaffolds ≥10 kb in length using MetaWatt³². Relative weight of coverage binning was set to 0.75 and the optimize bins and polish bins options were set to on. The taxonomic identity of MAGs was assessed using a concatenated phylogenetic tree based on 138 single copy conserved genes as implemented in MetaWatt³². Visualization of the tree and mapping of data on to taxa was performed with iTOL³³. Estimation of MAG completeness and contamination was performed using CheckM³⁴. Six Chloroflexi MAGs were selected for further analysis based on the presence of a 16S rRNA gene, high completeness and low contamination. Manual curation of the six Chloroflexi MAGs was performed and suspected contaminating scaffolds (single copy genes most similar to non-Chloroflexi taxa) were removed prior to further analysis of MAGs.

16S rRNA phylogenetic analysis. Chloroflexi diversity in the metagenomic assembly was assessed by 16S rRNA gene analysis. All 16S rRNA genes in the co-assembly were assigned to taxonomic groups using mothur³⁵ and the Wang method with a bootstrap value cutoff of 60%³⁶. Chloroflexi 16S rRNA genes greater than 360 bp were included in a phylogenetic analysis with Chloroflexi reference sequences. A multiple sequence alignment was generated using MUSCLE (implemented in MEGA6)³⁷. Phylogenetic reconstructions were conducted by maximum likelihood using MEGA6-v.0.6 and the following settings: general time reversible model, gamma distribution model for the rate variation with four discrete gamma categories, and the nearest-neighbor interchange (NNI) heuristic search method³⁸ with a bootstrap analysis using 100 replicates.

Single protein and concatenated protein phylogenies. A concatenated protein phylogeny was constructed using 30 Chloroflexi reference genomes and the 6 Canada Basin Chloroflexi MAGs. Orthologous genes in the 36 Chloroflexi genomes were identified using ProteinOrtho³⁸. Fifty orthologs present in at least 34 of the 36 genomes were selected for concatenated phylogenetic analysis (Supplementary Data 3). Each orthologous protein family was aligned using MUSCLE (implemented in MEGA6) and alignment positions were masked using the probabilistic masker ZORRO³⁹, masking columns with weights <0.5. The concatenated alignment consisted of 14,815 amino acid positions. Phylogenetic reconstructions were conducted by maximum likelihood using MEGA6-v.0.6 and the following settings: JTT substitution model, gamma distribution with invariant sites model for the rate variation with four discrete gamma categories, and the nearest-neighbor interchange (NNI) heuristic search method⁴⁰ with a bootstrap analysis using 100 replicates.

For phylogenetic analysis of ring-cleaving dioxygenase sequences identified in SAR202-VII-2, query sequences were searched against UniRef90 and 130 Chloroflexi MAGs constructed from the TARA Oceans dataset²⁹. The TARA Ocean Chloroflexi MAGs were used as is with no manual curation. UniRef90 sequences and the top TARA Ocean MAG hits for each dioxygenase were aligned with their respective SAR202-VII-2 homologs with MUSCLE (implemented in MEGA6) and alignment positions were masked using the probabilistic masker ZORRO³⁹, masking columns with weights less than 0.5. Phylogenetic reconstruction was conducted using the same settings as the concatenated phylogeny.

Comparative genomics and metabolic reconstruction. The distribution of orthologs across Arctic Ocean genomes, as well as the identification of orthologs shared with the deep ocean SAGs, was determined using proteinortho³⁸. Inference of protein function and metabolic reconstruction was based on the IMG annotations provided by the JGI, including KEGG, Pfam, EC numbers, and Metacyc annotations. Metabolic reconstruction was also facilitated by generated pathway genome databases for each MAG using the pathologic software available through Pathway Tools⁴¹.

Metagenomic fragment recruitment. The distribution of the Canada Basin MAGs in the global ocean was determined using best-hit reciprocal blast analysis similar to Landry et al.¹⁸ Unassembled metagenomic data from 25 samples (Supplementary Data 1) was first recruited to the six Canada Basin Chloroflexi MAGs as well as two SAR202 SAGs originating from the deep North Pacific Ocean from the Hawaiian ocean time series (HOTS) and the deep North Atlantic Ocean¹⁸. Metagenomes from the TARA Ocean project used here were representative of the surface, chlorophyll maximum and mesopelagic waters from the North Atlantic, South Atlantic, North Pacific, Coastal North Pacific, South Pacific, Coast of Brazil and the Antarctic peninsula. To reduce computational demand, only part 1 (1 Gbp of a random subset of reads) of each metagenomic dataset available at EBI was used (Supplementary Data 1). Additional bathypelagic metagenomes from the North Pacific and South Atlantic Oceans (LineP P04, P12 and P26, and Knorr S15 2500 m) were also included. All hits from the initial blast were then reciprocally queried against the Canada Basin Chloroflexi MAGs, bathypelagic SAR202 SAGs, and 130 Chloroflexi MAGs constructed from the

TARA oceans data. The best-hit was reported. Only hits with an alignment length ≥ 100 bp and a percent identity of 95% or more were counted (lower % identity cutoffs did not alter the number of reads recruited in any significant manner). To compare the results among the different datasets, the number of recruited reads was normalized to total number of reads in each sample. The final coverage results were expressed as the number of reads per kilobase of the MAG per gigabase of metagenome (rpk).

Data availability. The metagenomic data generated in this study are available in the Integrated Microbial Genomes database at the Joint Genome Institute at <https://img.jgi.doe.gov>, GOLD Project ID: Ga0133547. Metagenome-assembled genome projects have been deposited at DDBJ/ENA/GenBank under the Bioproject PRJNA471535 and accession numbers QGNM000000000 (for SAR202-II-3), QGNN000000000 (for SAR202-II-177A), QGNO000000000 (for SAR202-VI-29A), QEVV000000000 (for SAR202-VII-2), QGNP000000000 (for Anck29-46), and QGNQ000000000 (for TK10-74A). The versions described in this paper are versions QGNM010000000 (for SAR202-II-3), QGNN010000000 (for SAR202-II-177A), QGNO010000000 (for SAR202-VI-29A), QEVV010000000 (for SAR202-VII-2), QGNP010000000 (for Anck29-46), and QGNQ010000000 (for TK10-74A).

Received: 6 November 2017 Accepted: 24 May 2018

Published online: 05 July 2018

References

- Carmack, E. C. et al. Freshwater and its role in the Arctic marine system: sources, disposition, storage, export, and physical and biogeochemical consequences in the Arctic and global oceans. *J. Geophys. Res. Biogeosci.* **121**, 675–717 (2016).
- Benner, R., Benitez-Nelson, B., Kaiser, K. & Amon, R. M. W. Export of young terrigenous dissolved organic carbon from rivers to the Arctic Ocean. *Geophys. Res. Lett.* **31**, 1–4 (2004).
- Opsahl, S., Benner, R. & Amon, R. M. W. Major flux of terrigenous dissolved organic matter through the Arctic Ocean. *Limnol. Oceanogr.* **44**, 2017–2022 (1999).
- Bintanja, R. & Andry, O. Towards a rain-dominated Arctic. *Nat. Clim. Change* **7**, 263–267 (2017).
- Frey, K. E. & McClelland, J. W. Impacts of permafrost degradation on arctic river biogeochemistry. *Hydrol. Process* **23**, 169–182 (2009).
- Vonk, J. E. et al. Activation of old carbon by erosion of coastal and subsea permafrost in Arctic Siberia. *Nature* **489**, 137–140 (2012).
- Jiao, N. et al. Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nat. Rev. Microbiol.* **8**, 593–599 (2010).
- Hansell, D. A. & Carlson, C. A. *Biochemistry of Marine Dissolved Organic Matter* 2nd ed (Academic Press, Amsterdam, 2015).
- Kaiser, K., Benner, R. & Amon, R. M. W. The fate of terrigenous dissolved organic carbon on the Eurasian shelves and export to the North Atlantic. *J. Geophys. Res.* **122**, 4–20 (2017).
- Giovannoni, S. J., Rappé, M. S., Vergin, K. L. & Adair, N. L. 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the green non-sulfur bacteria. *Proc. Natl Acad. Sci. USA* **93**, 7979–7984 (1996).
- DeLong, E. F. et al. Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**, 496–503 (2006).
- Morris, R. M., Rappé, M. S., Urbach, E., Conon, S. A. & Giovannoni, S. J. Prevalence of the chloroflexi-related SAR202 bacterioplankton cluster throughout the mesopelagic zone and deep ocean. *Appl. Environ. Microbiol.* **70**, 2836–2842 (2004).
- Schattenhofer, M. et al. Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic ocean. *Environ. Microbiol.* **11**, 2078–2093 (2009).
- Varela, M. M., van Aken, H. M. & Herndl, G. J. Abundance and activity of Chloroflexi-type SAR202 bacterioplankton in the meso- and bathypelagic waters of the (sub)tropical Atlantic. *Environ. Microbiol.* **10**, 1903–1911 (2008).
- Urbach, E. et al. Unusual bacterioplankton community structure in ultra-oligotrophic Crater Lake. *Limnol. Oceanogr.* **46**, 557–572 (2001).
- Urbach, E., Vergin, K. L., Larson, G. L. & Giovannoni, S. J. Bacterioplankton communities of Crater Lake, OR: dynamic changes with euphotic zone food web structure and stable deep water populations. *Hydrobiologia* **574**, 161–177 (2007).
- Yamada, T. & Sekiguchi, Y. Cultivation of uncultured Chloroflexi subphyla: significance and ecophysiology of formerly uncultured Chloroflexi “subphylum I” with natural and biotechnological relevance. *Microbes Environ.* **24**, 205–216 (2009).
- Landry, Z., Swan, B. K., Herndl, G. J., Stepanauskas, R. & Giovannoni, S. J. SAR202 genomes from the dark ocean predict pathways for the oxidation of recalcitrant dissolved organic matter. *mBio* **8**, e00413–e00417 (2017).
- Bano, N. & Hollibaugh, J. T. Phylogenetic composition of bacterioplankton assemblages from the Arctic ocean. *Appl. Environ. Microbiol.* **68**, 505–528 (2002).
- Guéguen, C. et al. The nature of colored dissolved organic matter in the southern Canada Basin and East Siberian Sea. *Deep Sea Res. Part 2* **81–84**, 102–113 (2012).
- Dainard, P. G. & Guéguen, C. Distribution of PARAFAC modeled CDOM components in the North Pacific Ocean, Bering, Chukchi and Beaufort Seas. *Mar. Chem.* **157**, 216–223 (2013).
- Dainard, P. G., Guéguen, C., McDonald, N. & Williams, W. J. Photobleaching of fluorescent dissolved organic matter in Beaufort Sea and North Atlantic Subtropical Gyre. *Mar. Chem.* **177**, 630–637 (2015).
- Thrash, J. C. et al. Metabolic roles of uncultivated bacterioplankton lineages in the Northern Gulf of Mexico “dead zone”. *mBio* **8**, e01017–e01017 (2017).
- Barry, K. P. & Taylor, E. A. Characterizing the promiscuity of LigAB, a lignin catabolite degrading extradiol dioxygenase from *Sphingomonas paucimobilis* SYK-6. *Biochemistry* **52**, 6724–6736 (2013).
- Kasai, D., Masai, E., Miyauchi, K., Katayama, Y. & Fukuda, M. Characterization of the gallate dioxygenases are involved in syringate degradation by *Sphingomonas paucimobilis* SYK-6. *J. Bacteriol.* **187**, 5067–5074 (2005).
- Werwath, J., Arfmann, H.-A., Pieper, D. H., Timmis, K. N. & Wittich, R.-M. Biochemical and genetic characterization of a gentisate 1,2-dioxygenase from *Sphingomonas* sp. Strain RW5. *J. Bacteriol.* **180**, 4171–4176 (1998).
- Fetzner, S. Ring-cleaving dioxygenases with a cupin fold. *Appl. Environ. Microbiol.* **78**, 2505–2514 (2012).
- Fuchs, G., Boll, M. & Heider, J. Microbial degradation of aromatic compounds—from one strategy to four. *Nat. Rev. Microbiol.* **9**, 803–816 (2011).
- Tully, B., Graham, E. D. & Heidelberg, J. F. The reconstruction of 2,631 draft metagenome-assembled-genomes from the global oceans. *Sci. Data* **5**, 170203 (2018).
- Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
- Huntemann, M. et al. The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP v.4). *Stand. Genom. Sci.* **11**, 1–17 (2016).
- Strous, M., Kraft, B., Bisdorf, R. & Tegetmeyer, H. E. The binning of metagenomic contigs for microbial physiology of mixed cultures. *Front. Microbiol.* **3**, 1–11 (2012).
- Letunic, I. & Bork, P. Interactive tree of life (iTOL)v3: an online tool for display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, 242–245 (2016).
- 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).
- Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
- Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
- Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
- Lechner, M. et al. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* **12**, 124 (2011).
- Wu, M., Chatterji, S. & Eisen, J. A. Accounting for alignment uncertainty in phylogenomics. *PLoS ONE* **7**, e30288 (2012).
- Tamura, K., Stecher, G., Peterson, D., Filipiński, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
- Karp, P. D., Latendresse, M. & Caspi, R. The pathway tools pathway prediction algorithm. *Stand. Genom. Sci.* **5**, 424–429 (2011).
- Schlitzer, R. Ocean data view. <http://odv.awi.de> (2016).

Acknowledgements

The data were collected aboard the CCGS Louis S. St-Laurent in collaboration with researchers from Fisheries and Oceans Canada at the Institute of Ocean Sciences and Woods Hole Oceanographic Institution's Beaufort Gyre Exploration Program and are available at <http://www.who.edu/beaufortgyre>. We would like to thank both the Captain and crew of the CCGS Louis S. St-Laurent and the scientific teams aboard. The work was conducted in collaboration with the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, and was supported under Contract No.

DE-AC02-05CH11231. Funding from the Canadian Natural Science and Engineering Research Council (NSERC) Discovery grants (D.W., C.G. and C.L.) and the Canada Research Chair Program (D.W., C.G.) are acknowledged. D.C. was supported by FRQNT and Concordia's Institute for Water, Energy and Sustainable Systems.

Authors contributions

D.C. and D.W. designed and carried out the metagenomic experiments. W.J.W., C.L. and C.G. designed the sampling strategy. W.J.W. provided oceanographic data. D.W. collected samples. D.C. extracted the environmental genomic DNA and analyzed sequencing data. P.T. contributed to the bioinformatic analyses. D.C. and D.W. with contributions from C.G. and C.L. wrote the manuscript. All authors commented on the manuscript.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

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

Additional information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s42003-018-0086-7>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018