

RESEARCH ARTICLE

Identification and widespread environmental distribution of a gene cassette implicated in anaerobic dichloromethane degradation

Robert W. Murdoch¹ | Gao Chen^{1,2} | Fadime Kara Murdoch^{1,3} | E. Erin Mack⁴ |
Manuel I. Villalobos Solis³ | Robert L. Hettich³ | Frank E. Löffler^{1,2,3,5,6} 

¹Center for Environmental Biotechnology, University of Tennessee, Knoxville, Tennessee, USA

²Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, Tennessee, USA

³Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

⁴Corteva Environmental Remediation, Corteva Agriscience, Wilmington, Delaware, USA

⁵Department of Microbiology, University of Tennessee, Knoxville, Tennessee, USA

⁶Department of Biosystems Engineering and Soil Science, University of Tennessee, Knoxville, Tennessee, USA

Correspondence

Frank E. Löffler. Center for Environmental Biotechnology, University of Tennessee, 676 Dabney Hall, 1416 Circle Drive, Knoxville, TN 37996-1605, USA.
Email: frank.loeffler@utk.edu

Present address

Robert W. Murdoch and Fadime Kara Murdoch, Battelle Memorial Institute, Columbus, Ohio, USA

Funding information

The Chemours Company; University Consortium for Field-Focused Groundwater Research

Abstract

Anthropogenic activities and natural processes release dichloromethane (DCM, methylene chloride), a toxic chemical with substantial ozone-depleting capacity. Specialized anaerobic bacteria metabolize DCM; however, the genetic basis for this process has remained elusive. Comparative genomics of the three known anaerobic DCM-degrading bacterial species revealed a homologous gene cluster, designated the methylene chloride catabolism (*mec*) gene cassette, comprising 8–10 genes encoding proteins with 79.6%–99.7% amino acid identities. Functional annotation identified genes encoding a corrinoid-dependent methyltransferase system, and shotgun proteomics applied to two DCM-catabolizing cultures revealed high expression of proteins encoded on the *mec* gene cluster during anaerobic growth with DCM. In a DCM-contaminated groundwater plume, the abundance of *mec* genes strongly correlated with DCM concentrations ($R^2 = 0.71$ – 0.85) indicating their potential value as process-specific bioremediation biomarkers. *mec* gene clusters were identified in metagenomes representing peat bogs, the deep subsurface, and marine ecosystems including oxygen minimum zones (OMZs), suggesting a capacity for DCM degradation in diverse habitats. The broad distribution of anaerobic DCM catabolic potential infers a role for DCM as an energy source in various environmental systems, and implies that the global DCM flux (i.e., the rate of formation minus the rate of consumption) might be greater than emission measurements suggest.

KEYWORDS

anaerobic degradation, bioremediation, climate change, dichloromethane flux, ozone destruction

1 | INTRODUCTION

Dichloromethane (DCM, methylene chloride) is a widely distributed halomethane, produced both naturally and industrially. While anthropogenic DCM has received attention due to widespread groundwater contamination and, more recently ozone destruction potential (Hossaini et al., 2017), analysis of Antarctic ice cores has demonstrated that DCM was present in the atmosphere prior to the industrial era at

approximately 10% of modern levels (Trudinger et al., 2004). The natural sources of DCM are diverse, encompassing both abiotic (Isidorov et al., 1990; Kanters & Louw, 1996) and biotic (Eustáquio et al., 2008; Hoekstra et al., 1998; Wuosmaa & Hager, 1990) processes and are estimated to contribute up to one third of total emissions (Gribble, 2010). Since the 1960s, atmospheric DCM concentrations rose steadily (An et al., 2021; Hossaini et al., 2017) although the reported worldwide production and use has been steady or declining

since 2010 (McCulloch, 2017). Possible explanations include undocumented production, rogue emissions, or increased natural emissions reflecting environmental (e.g., climate) change responses.

Atmospheric measurements and corresponding efforts to extrapolate to global-scale emissions have led to the perception that marine systems and biomass combustion (e.g., wildfires) are the primary non-industrial sources of DCM (Gribble, 2010), releasing estimated amounts of 190 and 60 Gg of DCM each year, respectively. Natural and deliberate forest fires have increased in frequency and size (Haines et al., 2020), a global trend that can be expected to lead to further formation and release of DCM. Wetlands emit up to 2 Gg DCM/year (Cox et al., 2004; Hu et al., 2017; Kolusu et al., 2018) (Supplementary Information), and volcanic activity contributes an estimated amount of 0.021 Gg DCM/year (Gribble, 2010). Halomethanes occur in crustal minerals, and DCM release from rocks from the near-surface and the deep subsurface have been reported (Mulder et al., 2013; Svensen et al., 2009). Knowledge gaps remain and not-yet identified environmental sources of DCM are likely. The current understanding of global DCM fluxes (i.e., the sum of DCM production and DCM removal rates), as opposed to emissions (i.e., the amount of DCM released to the atmosphere), is very limited (Figure 1) (McCulloch, 2017).

The rates of DCM production and consumption are unclear, and biological attenuation in oxic and anoxic systems prior to release to the atmosphere has not been considered in existing atmospheric emission models (Hossaini et al., 2017).

Anaerobic metabolism of DCM was described in a bacterial isolate, *Dehalobacterium formicoaceticum* strain DMC (Defo) (Mägli et al., 1996). Two additional bacterial populations responsible for DCM metabolism in anaerobic enrichment cultures have been identified, '*Ca. Dichloromethanomonas elyunquensis*' (Diel) (Justicia-Leon et al., 2012; Kleindienst et al., 2016, 2017) and '*Ca. Formimonas warabiya*' (Fowa) (Holland et al., 2021). All three anaerobic DCM degraders are members of the family *Peptococcaceae* within the phylum Firmicutes. Bacterial dechlorination of DCM under oxidizing conditions is catalyzed by the glutathione-S-transferase (GST) DcmA (Hayoun et al., 2020; Muller et al., 2011); however, GST enzymes rarely occur in obligate anaerobes (Allocati et al., 2012). Accordingly, anaerobic DCM degraders utilize a distinct metabolic strategy, in which the C_1 group

is transferred to methylene-tetrahydrofolate (THF), a process that likely uses a corrinoid-dependent methyltransferase. The resulting methylene-THF is then channeled into the Wood-Ljungdahl pathway (WLP). Interestingly, Diel generates hydrogen during DCM mineralization to CO_2 and inorganic chloride, which necessitates a syntrophic partnership with a hydrogen-consuming population (Chen, Kleindienst et al., 2017). In contrast, Fowa ferments DCM to acetate and inorganic chloride (Holland et al., 2021) and axenic Defo cultures ferment DCM to acetate, formate, and chloride (Chen, Murdoch et al., 2017). Diel expresses reductive dehalogenases during growth with DCM (Kleindienst et al., 2019); however, the genomes of Defo and Fowa do not encode such enzyme systems and the modes of energy conservation during DCM catabolism remain unresolved.

Due to the genetic intractability of anaerobic DCM-degrading bacteria, the genes implicated in anaerobic DCM metabolism have remained elusive. This study used comparative genomic approaches to unravel underlying conserved genes involved in DCM metabolism, which led to the discovery of a novel gene cassette. A multiple lines of evidence approach including proteomics applied to DCM-degrading Diel and Defo cultures, the analysis of metatranscriptome data collected from a DCM groundwater plume, the study of novel DCM-degrading enrichment cultures, and targeted qPCR assays supported a role of this gene cassette in anaerobic DCM metabolism. The identification of homologs of this gene cluster among public metagenome datasets motivated collection of samples from a pristine marine system for more sensitive analyses using targeted qPCR assays. The new findings suggest DCM presence and consumption in diverse natural ecosystems and provide new opportunities for assessing how global environmental changes impact DCM fluxes and emissions to the atmosphere.

2 | METHODS

2.1 | Comparative genomics

Initial identification of homologous genes shared between the genomes of Diel and Defo was performed by blastp-based reciprocal best hit (RBH) analysis within the Integrated Microbial Genomes

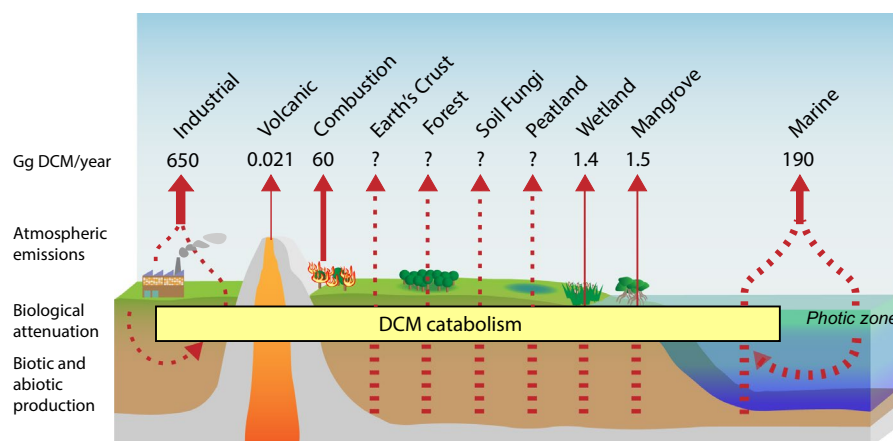


FIGURE 1 Major reported and potential dichloromethane (DCM) atmospheric emission sources. Width of the solid arrows is log-proportional to the magnitude of DCM emission estimates or potential as described in the text. Dashed lines indicate putative DCM sources and emissions that have not been directly investigated

(IMG) system (Chen, Chu et al., 2019) and by using GView (Petkau et al., 2010). A homologous gene cluster was identified in the genome of '*Ca. Formimonas warabiya*' by application of local blastp searches (Altschul et al., 1990). Functional annotations (COG, pfam, KEGG, TIGRFAM) for Defo and Diel genes were obtained from the IMG system, while those of '*Ca. Formimonas warabiya*' were assigned using the WebMGA server (Wu et al., 2011) for COG (Tatusov et al., 2000), pfam (El-Gebali et al., 2019), and TIGRFAM (Haft et al., 2012) annotations and GhostKOALA (Kanehisa et al., 2016) for KEGG annotations (Kanehisa et al., 2017).

2.2 | Metagenome searches

All 18,314 metagenomes in the IMG database publicly available as of January 7, 2020 were subjected to blastp query with the Defo putative DCM dehalogenating methyltransferase MecE protein sequence with a minimum bit score cutoff of 150 (approximately 40% identity). The resulting protein set was further filtered by applying an RBH criterion, retaining only proteins whose closest blastp hit to the IMG genomes database was found among the MecE sequences encoded on the putative *mec* gene cassettes. The candidate metagenome MecE homologs were then further filtered by retaining only proteins encoded by genes co-localized with at least one other *mec* cassette protein, applying the same RBH criterion. All proteins encoded by genes located on scaffolds where the Mec protein homologs were identified were downloaded, subjected to local blastp query using the ten *mec* gene cassette-encoded proteins from Defo, and plotted using GenoPlotR and custom R scripts (Guy et al., 2010). Gene copy per genome for each metagenomic *mec* cassette, provided in Dataset S1, was calculated by dividing the read depth of the corresponding scaffold by the average read depth of ten single copy conserved protein-encoding genes identified in the same metagenome (ribosomal proteins L11 (COG0080), L1 (COG0081), L3 (COG0087), L4 (COG0088), L2 (COG0090), L22 (COG0091), L5 (COG0094), L15 (COG0200), L10 (COG0244), and L29 (COG0255)).

2.3 | Phylogenetic reconstruction

The top 20 most similar proteins to each of the proteins encoded by each of the Defo genes located on the *mec* cassette were obtained by searching the IMG genome database using blastp with a confidence threshold of $1e^{-5}$. All *mec* cassette genes located in both genomes and metagenomes were aligned and subjected to phylogenetic reconstruction alongside the top 20 most similar genes located in microbial genomes in the IMG database. Proteins encoded by genes from metagenomes were clustered at 80% similarity using CD-hit (Li & Godzik, 2006). Sequences were aligned using MAFFT G-INS-I with 1000 maximum iterations (Kato & Standley, 2013), trimmed using trimAl-gappout (Capella-Gutiérrez et al., 2009) and subjected to phylogenetic reconstruction using FastTree2 maximum-likelihood estimation (Gamma-LG model) (Price et al., 2010). The resulting

Newick tree files were visualized using the Interactive Tree of Life (Letunic & Bork, 2016).

2.4 | Preparation of Diel and Defo cultures for global proteomics

The acquisition of metaproteomics data for the DCM-degrading consortium RM harboring '*Ca. Dichloromethanomonas elyunquensis*' (Diel) was described previously (Kleindienst et al., 2019). The axenic culture *Dehalobacterium formicoaceticum* (Defo) was grown in triplicate in 100 ml of anoxic mineral basal salt medium with 0.2 mM sodium sulfide, 0.2 mM L-cysteine (Löffler et al., 2005) and 30 mM bicarbonate (pH 7.3) under a headspace of N_2/CO_2 (80:20, vol/vol) with 156 μ mol (10 μ l) of DCM as the sole energy source. Cultures were initiated with a 5% (vol/vol) inoculum, incubated at 30°C in the dark without agitation, and provided one additional feeding of DCM once the initial amendment was consumed. Biomass for metaproteomic analyses was collected after 2 weeks of incubation when approximately 95% of the second DCM feedings were consumed. Culture suspensions were passed through Sterivex™ 0.22 μ m membrane filter units (EMD Millipore Corporation, Billerica, MA, US) to capture cells. The outlets of the filter units were capped, and 1.5 ml of boiling SDS lysis buffer (4% SDS in 100 mM Tris/HCl buffer, pH 8.0) were added to each of them. Filter unit inlets were then capped and the cartridges placed in a laboratory rocker for 1-hour at room temperature. The SDS lysis buffer was removed by connecting 3 ml plastic syringes to the inlets of the cartridges, and then holding the syringes and filter units vertically and pushing air into each cartridge in order to withdraw as much lysate as possible by back pressure. In addition, filters were rinsed once more with 0.5 ml of fresh SDS lysis buffer. Lysate mixtures were centrifuged at 21,000 *g* for 15 min and the clean protein supernatant transferred to fresh Eppendorf plastic tubes. Proteins were precipitated with trichloroacetic acid (TCA), denatured in 8 M urea, reduced with dithiothreitol (DTT), alkylated with iodoacetamide (IAM), and digested with sequencing grade trypsin (Promega, 1:50 trypsin-to-protein [wt/wt]) (Yang et al., 2012). Protein and peptide concentrations were estimated with the BCA assay (Pierce Biotechnology) and peptide extracts were stored at -80°C for subsequent LC-MS/MS analysis.

2.5 | Global proteomics analyses of Diel and Defo cultures

Global proteomics analyses were performed with an Orbitrap Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) equipped with a nano-electrospray source (ESI) interfaced with a Proxeon EASY-nLC™ 1200 system. Peptides (2 μ g) from each sample were suspended in solvent A (2% acetonitrile/0.1% formic acid in water) and injected onto a C18 resin 75 μ m microcapillary column (1.7 μ m, 100 Å, Phenomenex). Separation was accomplished at a constant flow rate of 250 nl/min with a 90-min gradient from

2% to 30% solvent B (0.1% formic acid/80% acetonitrile in water) followed by an increase to 40% solvent B within 10 min. Tandem mass spectrometry (MS/MS) data were collected using the Thermo Xcalibur software version 4.2.47 with similar parameters as reported before (Ganusova et al., 2021). Raw proteomics spectral data previously generated from Diel (Kleindienst et al., 2019) were re-analyzed alongside newly generated Defo data using identical programs and parameters to avoid methodological biases. Raw spectral files were searched against protein databases from the IMG annotated genomes of enrichment culture RM (which contains Diel) and Defo (IMG genome IDs 3300005804 and 2811995020, respectively), to which common laboratory contaminant proteins were appended. For standard database searching, the peptide MS/MS data were searched using Proteome Discoverer v2.4. The MS/MS data were searched using the SEQUEST HT algorithm (Eng et al., 1994) which was configured to derive fully tryptic peptides with the following settings: Maximum of 2 missed cleavage sites per peptide, minimum peptide length of 2, MS1 mass tolerance of 10 ppm and a MS2 tolerance of 0.02 Da. In addition, carbamidomethylations on cysteines (+57.0214 Da) and methionine oxidations (+5.9949 Da) were searched on peptides as static and dynamic modifications, respectively. Peptide spectrum match (PSM) confidence was evaluated with Percolator (Käll et al., 2007). PSMs and peptides were considered identified at a *q* value of <.01. Abundance values were converted to log₂ values for ease of visualization. The IMG gene IDs of detected culture RM proteins were mapped to proteins contained in the IMG annotated genome of Diel (IMG genome ID 2627853586, Dataset S2) by blastp and manual curation of results to require perfect protein-level identities and identical contig-level gene order.

2.6 | Dichloromethane measurements

DCM was quantified by manual headspace injections (0.1 ml) into an Agilent 7890 gas chromatograph (GC) (Santa Clara) equipped with a DB-624 column (60 m length, 0.32 mm i.d., 1.8 mm film thickness) and a flame ionization detector (FID). To analyze DCM concentrations in groundwater, 1 ml samples were collected, immediately transferred to sealed 20 ml glass vials, and the DCM concentration determined in the headspace. Aqueous phase DCM concentrations were determined using a dimensionless Henry's law constant of 0.0895 (Gossett, 1987).

2.7 | Environmental samples

Anaerobic digester sludge was collected from two wastewater treatment plants, the Knoxville Utilities Board (KUB) Kuwahee Wastewater Treatment Plant and the Lenoir City (LC) Wastewater Treatment Plant. Groundwater samples from six monitoring wells representing within plume, fringe and outside locations at a DCM-contaminated site were obtained from CDM Smith (Wright et al., 2017). The groundwater samples were shipped with an overnight

carrier in a cooler with ice and analyzed immediately upon receipt. Frank Stewart (Montana State University) provided archived DNA samples from two vertical transects from the Eastern Tropical North Pacific Oxygen Minimum Zone (ETNP OMZ).

2.8 | Dichloromethane enrichments

Microcosms were established in 160 ml glass serum bottles containing 98 ml of anoxic mineral basal salt medium amended with 156 µmol (10 µl) of DCM. The microcosms were seeded with 2 ml of digester sludge, and additional DCM feedings occurred upon the depletion of DCM. Microcosms showing DCM degradation were sequentially transferred to fresh anoxic medium with DCM as the sole electron donor with an inoculation volume of 3 ml. After eight consecutive transfers, solid-free enrichment cultures were obtained that degraded DCM under anoxic conditions. DNA samples were extracted from the new DCM enrichment cultures and used to examine the presence of *mecE* and *mecF* genes by qPCR.

2.9 | DNA extraction

DNA extraction from 1 ml anaerobic digester sludge was performed using the DNeasy PowerSoil DNA extraction kit (Qiagen, Valencia, CA). DNA from Defo and Diel cultures was extracted from 5 ml culture suspensions collected onto 0.22 µm Durapore membrane filters (Millipore) and DNA using the DNeasy PowerLyzer PowerSoil DNA extraction kit (Qiagen) following the manufacturer's instructions. Biomass from groundwater samples (950 ml) was collected on Supor® 0.2 µm membrane filters (Pall Lab.). Each filter was cut in half using a sterile scalpel and each piece was placed into a separate bead-beating tube for extraction with the DNeasy PowerLyzer PowerSoil DNA kit. The extracted DNA was concentrated with the Zymo DNA Clean and Concentrator-25 Kit (Zymo Research). DNA concentrations were determined with fluorometry and DNA was stored at -80°C until qPCR analysis.

2.10 | Primer design, PCR, and qPCR analyses

Primer sets were developed for *mecE* and *mecF*. The design was based on the target gene alleles identified in the genomes of Diel, Defo, Fowa, and *Dehalobacter* sp. strain UNSWDHB and the most similar homologs from peat bog metagenomes. Additional primers were designed based on the most common *mecE* and *mecF* alleles identified in Eastern Pacific OMZ metagenomes (IMG genome Ga0066828, gene IDs 100177932 [*mecE*] and 100027434 [*mecF*]). The respective target gene sequences were aligned using ClustalW and primer sets were designed using the Primer 3 plug-in in Geneious R11.0.2 (Kearse et al., 2012) for PCR and SYBR qPCR assays. The primer sequences were blasted against the NCBI nr database using the Primer BLAST program with default parameters to verify specificity of

the assays. The primers were obtained from a commercial supplier (Integrated DNA Technologies). For quantification of total bacterial 16S rRNA genes, Bac1055YF/Bac1392R primers (Ritalahti et al., 2006) were used for groundwater and enrichment cultures and EUB338F/EUB518R primers (Lane, 1991; Muyzer et al., 1993) were used for OMZ samples. Primer sequences are listed in Table 1.

qPCR was performed in 10 µl volumes consisting of 5 µl 2X Power SYBR Green PCR Master Mix (Applied Biosystems), 0.5 µl of each primer (final concentration of 300 nM) and 2 µl of template DNA (undiluted, 1:10 and 1:100 dilutions). Genomic DNA from three strains that lack *mecE* and *mecF* homologs, *Dehalobacter* sp. strain CF, *Dehalococcoides mccartyi* strain GT, and *Dehalogenimonas lykanthroporepellens* strain BL-DC-9 were used as negative controls for *mecE* and *mecF* qPCR assays. qPCR analysis was conducted using an Applied Biosystems QuantStudio 12K Flex Real Time qPCR System (Life Technologies), and the PCR temperature cycling program was as follows: Initial denaturation step for 10 min at 95°C, followed by 40 cycles of 15 s at 95°C and 1 min at 60°C. Specific amplification was confirmed by melt curve analysis and agarose gel electrophoresis as previously described (Hatt & Löffler, 2012). Standard curves were established with linear GeneArt DNA fragments of target genes. Standards were run as triplicate on each plate using tenfold dilution series in the range of 10^1 to 10^8 gene copies/µl. The amplification efficiencies (AEs), linear dynamic range, slope, Y-intercept and R^2 values are listed in Dataset S3. AEs were calculated using the equation $10^{(-1/\text{slope})} - 1$.

2.11 | Metatranscriptomics

Unprocessed Illumina metatranscriptome sequencing data generated from groundwater from a DCM contamination plume was provided by CDM Smith. This same site was previously the subject of 16S rRNA gene amplicon library analysis (Wright et al., 2017). Raw

data were trimmed using Trimmomatic v0.35 with a 6:25 sliding window quality trim, Illumina adapter read-through contamination removal, and final minimum length of 25 bp (Bolger et al., 2014). Trimmed reads were assembled *de novo* using Trinity v2.8.5 under default parameters (Grabherr et al., 2011). Prokaryotic ribosomal RNA genes were detected in the transcript contigs using barrnap v0.9-2 (Seeman, 2018) and removed using a custom R script employing Biostrings v2.5.4 (Pagès et al., 2019). Remaining transcripts were queried against a database consisting of all Defo *mec* genes using tblastx. Transcript coverage was calculated using kallisto (Bray et al., 2016) and TPM values were calculated using the Trinity utility script align_and_estimate_abundance.pl (Grabherr et al., 2011).

3 | RESULTS

3.1 | Anaerobic dichloromethane degraders share a common gene cassette

Comparative analysis of the Defo and Diel genomes revealed that the eight most similar genes (and 10 of the top 25 most similar genes, Dataset S4) in terms of percent predicted amino acid identity were located in genetic clusters. The genome of Defo harbors a single 10-gene cluster, and Diel has two highly similar clusters A and B (Figure 2).

The gene arrangement in these clusters is identical (i.e., syntenic), and amino acid identities between proteins encoded by homologous genes range from 79.6% to 99.7% (Figure 2). This novel, conserved 10-gene *mec* cassette harbors *mecA* through *mecJ* implicated in methylene chloride catabolism. A homologous gene cassette was also found in the newly described DCM degrader 'Ca. Formimonas warabiya' (Holland et al., 2019), although lacking *mecG* and *mecH*. Aside from *mecJ*, the only close homologs in any prokaryotic genome to any of the *mec*-encoded proteins are found in the genomes of three chloroform (CF) degrading bacteria; a homologous 10-gene

TABLE 1 qPCR primer sets targeting *mecE*, *mecF* and bacterial 16S rRNA genes in environmental samples

Primer	Sequence (5'3')	Target ^a	Applied to	Reference
mecE 828F mecE 1007R	ACCATATTGTCTTTTGCCYCAG TACCGCCCAATTTTCTGCTC	<i>mecE</i> ^b	DCM enrichment cultures, DCM plume samples	This study
mecF 554F mecF 641R	TGCTTGACATGGCCGTAMTGGAC GCAGGATADCCATATTTGTCTTT	<i>mecF</i> ^b	DCM enrichment cultures, DCM plume samples	This study
mecE 98F mecE 191R	ACGGCCTGACCTACAATGTC GCCGTGATGTCATAGCCGTA	<i>mecE</i> ^c	ETNP OMZ samples	This study
mecF 612F mecF 698R	GCTCAAGGACAAGTACGGCT CCGTATTGCTTCTTGCCGTG	<i>mecF</i> ^c	ETNP OMZ samples	This study
EUB338F EUB518R	ACTCCTACGGGAGGCAGCAG ATTACCGCGGCTGCTGG	Bacterial 16S rRNA gene	ETNP OMZ samples	Lane (1991); Muyzer et al. (1993)
Bac1055YF Bac1392R	ATGGYTGTGCTCAGCT ACGGGCGGTGTGTAC	Bacterial 16S rRNA gene	DCM enrichment cultures, DCM plume samples	Ritalahti et al. (2006)

^aDifferent *mecE* and *mecF* primers were designed for specificity to *mecE* and *mecF* homologs from different *mec* cassettes.

^bSpecific to *mecE* and *mecF* alleles found in *mec* cassettes from genomes of DCM degraders, *Dehalobacter* sp. strain UNSWDHB, and peat metagenomes.

^cSpecific to *mecE* and *mecF* alleles identified in ETNP OMZ metagenomes.

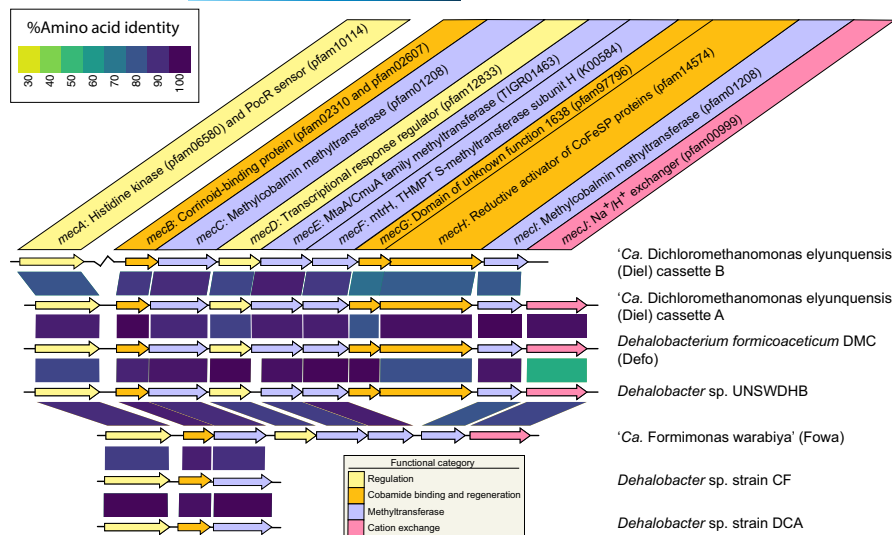


FIGURE 2 *mec* metabolic gene cassettes and close homologs identified in genomes. Colored boxes separating genes from different cassettes represent blastp amino acid identity scores. The arrow colors represent the general functional category of the encoded protein. Function was inferred from functional annotation systems, with priority given to the TIGRFAM and KEGG systems

cassette in *Dehalobacter* sp. strain UNSWDHB and partial gene cassettes in *Dehalobacter* sp. strain CF and *Dehalobacter* sp. strain DCA (Figure 2). Functional annotation of the *mec* cassette gene products revealed a histidine kinase sensory protein and an associated regulatory protein (MecAD), an MtaA/CmuA methyltransferase (MecE), an MtrH methyltransferase (MecE), two methyltransferases of indeterminate function (MecCI), a corrinoid-binding protein (MecB), a cation transporter/antiporter (MecJ), a reductive activator of corrinoid proteins (MecH), and a protein with a conserved domain of unknown function (MecG) (Table 2 and Supplementary Text).

3.2 | Mec proteins are expressed during growth on dichloromethane

A total of 1790 proteins were detected in the axenic Defo culture (Dataset S5), and 1743 proteins were detected in the metaproteome of mixed culture RM, 793 of which were assigned to Diel (Dataset S2). The majority of proteins encoded by the *mec* gene cassettes were detected in the proteomes of both DCM degraders (Figure 3, Table S1).

All proteins of the Defo *mec* cassette and Diel *mec* cassette B were detected, with the exception of MecJ in both cases. The corrinoid-binding protein MecB was the fourth and third most abundant protein in Defo and Diel proteomes, respectively. The three methyltransferases MecC, MecE, and MecF were in the top 1% most abundant proteins in both proteomes. The methyltransferase MecI and the corrinoid protein reductive activator MecH were all in the upper quartile of detected proteins. In neither case was MecJ detected, although its predicted eleven transmembrane alpha helices suggest strong association with the cytoplasmic membrane, likely hindering detection in the proteomics measurements (Vit & Petrak, 2017). Consistent with a previous study (Kleindienst et al., 2019), two reductive dehalogenases were detected in the Diel proteome (8th and 256th most abundant proteins), but no such enzymes are encoded on the Defo genome.

3.3 | Enrichment with dichloromethane selects for bacteria harboring *mec* genes

Alleles of *mecE* and *mecF*, which encode the putative dehalogenating methyltransferase and MtrH-like methyltransferase, respectively, were chosen for qPCR assay design. *mecB* and *mecC* encode a corrinoid protein and methyltransferase, respectively, were both highly abundant in Diel and Defo proteomes, but were excluded from consideration due to the presence of close homologs in the genomes of CF-respiring bacteria. When applied to DNA extracted from anaerobic digester sludge from KUB and LC municipal wastewater treatment plants, qPCR assays for the *mecE* and *mecF* genes did not yield quantifiable signals. Following enrichment with DCM using the same anaerobic sludge materials as inocula, 1×10^6 to 5×10^7 gene copies/ml of both *mecE* and *mecF* were measured in transfer cultures (Figure S1).

Groundwater samples from a DCM plume provided a unique opportunity to explore *mec* gene abundance and expression in response to varying DCM concentrations. The *mecE*- and *mecF*-targeted qPCR assays yielded signals in a DCM dose-dependent manner, with ratios of *mec* gene copy number versus total bacterial 16S rRNA gene copy numbers ranging between 4% and 10% in groundwater from wells with 2.0 g/L DCM or higher (Figure 4A). At lower DCM concentrations of 1.5 and 2.7 mg/L, the relative abundance of *mec* genes dropped to 0.0024%–0.036% (2.4×10^{-5} to 3.6×10^{-4}) and in wells at the fringe of the plume with no detectable DCM, *mec* gene-to-16S rRNA gene ratios dropped to 0.00036%–0.00041% (3.6×10^{-6} to 4.1×10^{-6}) (Figure 4A). Absolute *mec* gene copy numbers followed this trend, except for samples collected from the well with the highest DCM concentration of 7.2 g/L (approaching the 13 g/L solubility limit of DCM at 25°C (Yalkowsky et al., 2010)), where lower abundances of bacterial 16S rRNA genes were observed (i.e., 1.44×10^6 vs. 3.57×10^9 to 1.60×10^{10} 16S rRNA genes per L at locations with lower DCM concentrations) (Figure 4B). Absolute target gene copy numbers and ratios of target genes to total bacterial 16S rRNA copy numbers both covaried with measured DCM concentrations, with

TABLE 2 Consensus functional annotations of the putative *mec* cassette gene products

Protein	Annotation	Description
MecA	pfam06580	Histidine kinase
	pfam10114	Sensory domain found in PocR
MecB	COG5012	Methanogenic corrinoid protein MtbC1
	pfam02310	B ₁₂ -binding
	pfam02607	B ₁₂ -binding 2
	TIGR02370	Methyltransferase cognate corrinoid proteins
	K00548	MetH, 5-methyl-THF-homocysteine methyltransferase
MecC	COG0407	Uroporphyrinogen III decarboxylase
	pfam01208	Uroporphyrinogen decarboxylase
	K01599	Uroporphyrinogen decarboxylase
MecD	pfam00072	Translational response regulator receiver domain
	pfam12833	Helix-turn-helix domain
MecE	COG0407	Uroporphyrinogen III decarboxylase
	pfam01208	Uroporphyrinogen decarboxylase
	TIGR01463	Methyltransferase, MtaA/CmuA family
MecF	COG1962	Tetrahydromethanopterin S-methyltransferase, subunit H
	pfam02007	MtrH, tetrahydromethanopterin S-methyltransferase subunit H
	TIGR01114	N5-methyltetrahydromethanopterin:coenzyme M methyltransferase subunit H
	K00584	MtrH, tetrahydromethanopterin S-methyltransferase subunit H
MecG	pfam07796	Domain of unknown function 1638
MecH	COG3894	Uncharacterized 2Fe-2 and 4Fe-4S clusters-containing protein
	pfam14574	C-terminal of reductive activator of CoFeSP (RACo)
	pfam00111	2Fe-2S iron-sulfur cluster binding domain
MecI	pfam01208	Uroporphyrinogen decarboxylase
MecJ	COG0475	Kef-type K ⁺ transport system, membrane component KefB
	pfam00999	Sodium/hydrogen exchanger family
	K03455	Monovalent cation:H ⁺ antiporter-2, CPA2 family
	K03499	KtrA, trk system potassium uptake protein

Abbreviations: COG, clusters of orthologous genes; KEGG, kyoto encyclopedia of genes and genomes; PFAM, protein families; TIGRFAM, the institute for genomic research's database of protein families.

no detections outside the plume and *mec* gene-to-16S rRNA gene ratios of up to 10% within the plume, indicating that roughly 1 in 10 bacterial genomes in the plume harbored a *mec* cassette.

Analysis of metatranscriptomes, previously obtained for groundwater microbiomes collected from the same DCM-contaminated site, revealed expression of *mec* cassette genes. *mecE* and *mecF* were identified on 12 and 11 distinct transcript contigs, respectively, and co-localized on the same transcript on eight occasions. The most detected *mec* gene transcript encoded *mecBCEFGH* with amino acid identities to the corresponding Defo Mec proteins of 76.6%–96.5%. Expression of *mec* cassette genes reached their highest values of

up to 49.6 transcripts per million transcripts (TPM) in wells located in the plume fringes. *mec* cassette transcripts were also detected in groundwater samples in the core plume at relative expression levels up to 22.2 TPM (Table S2).

3.4 | Environmental distribution of the *mec* cassette

The search of over 18,000 metagenomes, anchored by identification of homologs of the *mecE* gene encoding a putative dehalogenating

methyltransferase, identified similar gene cassettes in 41 metagenomes from peatland, the deep subsurface, and marine systems (Figure 5, Figure S2). The SI provides gene and genome IDs and detailed blastp results (Table S3, Dataset S6, Dataset S7).

mec cassettes were identified in 13 of 117 metagenomes from ombrotrophic peat bogs in the Marcell experimental forest in Minnesota, USA. The *mec* cassettes were predominantly (i.e., 10 out of 13 gene cassettes) identified in metagenomes from samples collected between 1 and 1.5 m depth, where the pH was approximately

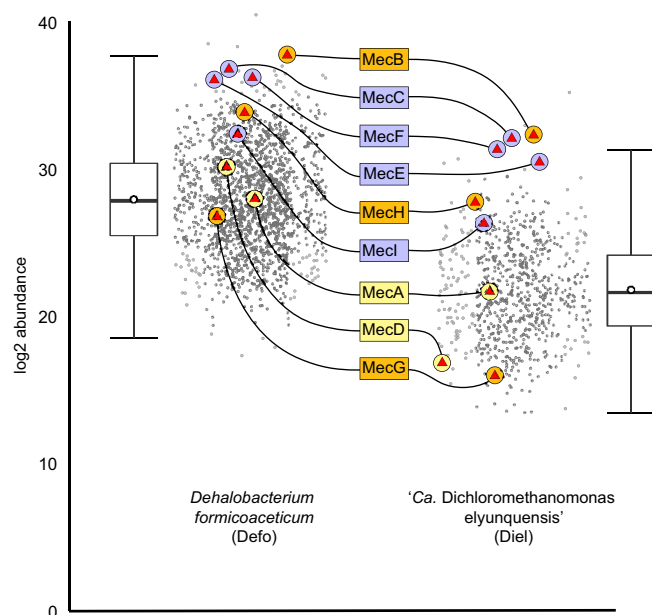


FIGURE 3 Mec protein abundance in the proteomes of Defo and Diel when grown anaerobically with dichloromethane (DCM) as the sole energy source. Protein products of the *mec* cassette genes are labelled with red triangles, with general functional category indicated by the encircling color (orange, corrinoid-related; purple, methyltransferase; yellow, regulatory). Box and whisker plots show median (central horizontal line), upper and lower quartile range (box), highest and lowest values excluding outliers (upper and lower whiskers) and mean value (open circle). The Diel proteome is a subset of the DCM enrichment culture RM metaproteome, and proteins corresponding to non-Diel community members are not shown

4.5 and anoxic conditions prevailed (Chris Schadt, personal communication). Two cassettes organized as *mecABCDEFI* displayed predicted amino acid identity scores to the Defo *mec* cassette genes above 80% (Figure 2) and close phylogenetic affiliation (Figure S3). This *mecABCDEFI* cassette was most prevalent, present in nearly 1% of all genome copies (i.e., approximately 1 out of 100 prokaryotic cells in the community harbor a *mec* cassette) (Dataset S7).

A total of 23 *mec* gene cassettes were identified in metagenomes from samples collected beneath the photic zone of the oceanic water column. The majority of cassettes was identified in metagenomes from the Eastern Pacific oxygen minimum zone (OMZ) at depths of 150–400 m, wherein oxygen was below the limit of detection (Thamdrup et al., 2019). Additional cassettes detected in metagenomes derived from three coastal sites, Monterey Bay, CA, Jervis Inlet in British Columbia, Canada, and Amundsen Gulf, in Arctic northern Canada, and from an open ocean sample from the Eastern Tropical North Pacific (ETNP). The marine *mec* cassettes were syntenic, with the order *mecBCFGHE* (Figure 5). *mecE* and *mecF* were most similar to the corresponding Defo homologs, with average amino acid identities of 49.7%–51.6%. All of the marine *mec* genes were more closely related to one another and to the *mec* cassette genes of the characterized DCM degraders than to any other gene in assembled metagenomes or genomes available in NCBI or the IMG database. The marine *mec* gene clusters formed distinct, deeply branching clades (Figure S3). The highest marine *mec* cassette occurrence was observed in a sample from Coal Oil Point, CA, a natural marine petroleum seep area (present in 1.02% of total genome copies) (Dataset S7). qPCR applied to water column samples from two ETNP OMZ locations detected *mecE* and *mecF* at a higher frequency (9 out of 11) than they were found in the ETNP OMZ metagenome assemblies (16 out of 90), with target gene-to-bacterial 16S rRNA gene ratios of 0.01–0.06% and gene copy numbers ranging from 1.2×10^4 to 1.3×10^5 per L (Figure 6, Dataset S8).

Evidence was obtained for the presence of *mecEFG* in anoxic porewater from a hydrogen-amended borehole in Opalinus Clay rock situated 300 m beneath Mt. Terri, Switzerland (Bagnoud et al., 2016). The amino acid identities for the putative methyltransferases MecE and MecF were 64.0% and 66.7%, respectively. Accordingly, phylogenetic analysis revealed close relationships with *mecE* and *mecF*

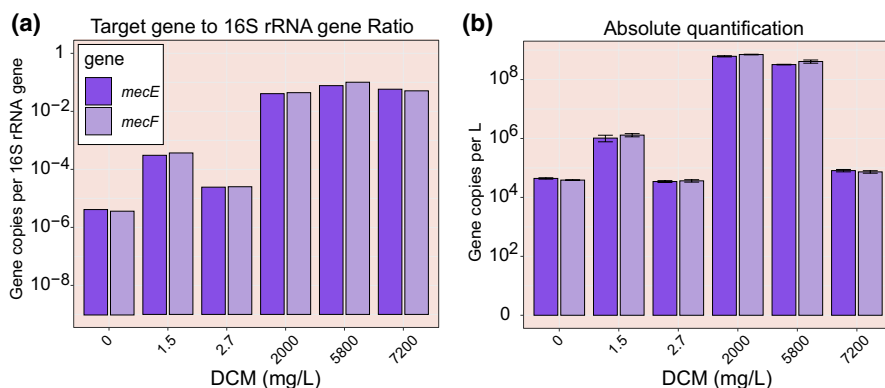


FIGURE 4 qPCR targeting *mecE* and *mecF* genes in groundwater samples from a dichloromethane plume. Panel (a) presents ratios of *mecE* or *mecF* gene to 16S rRNA gene copies. Panel (b) shows the *mecE* and *mecF* gene copy numbers per liter of groundwater with standard deviations represented by error bars ($n = 3$)

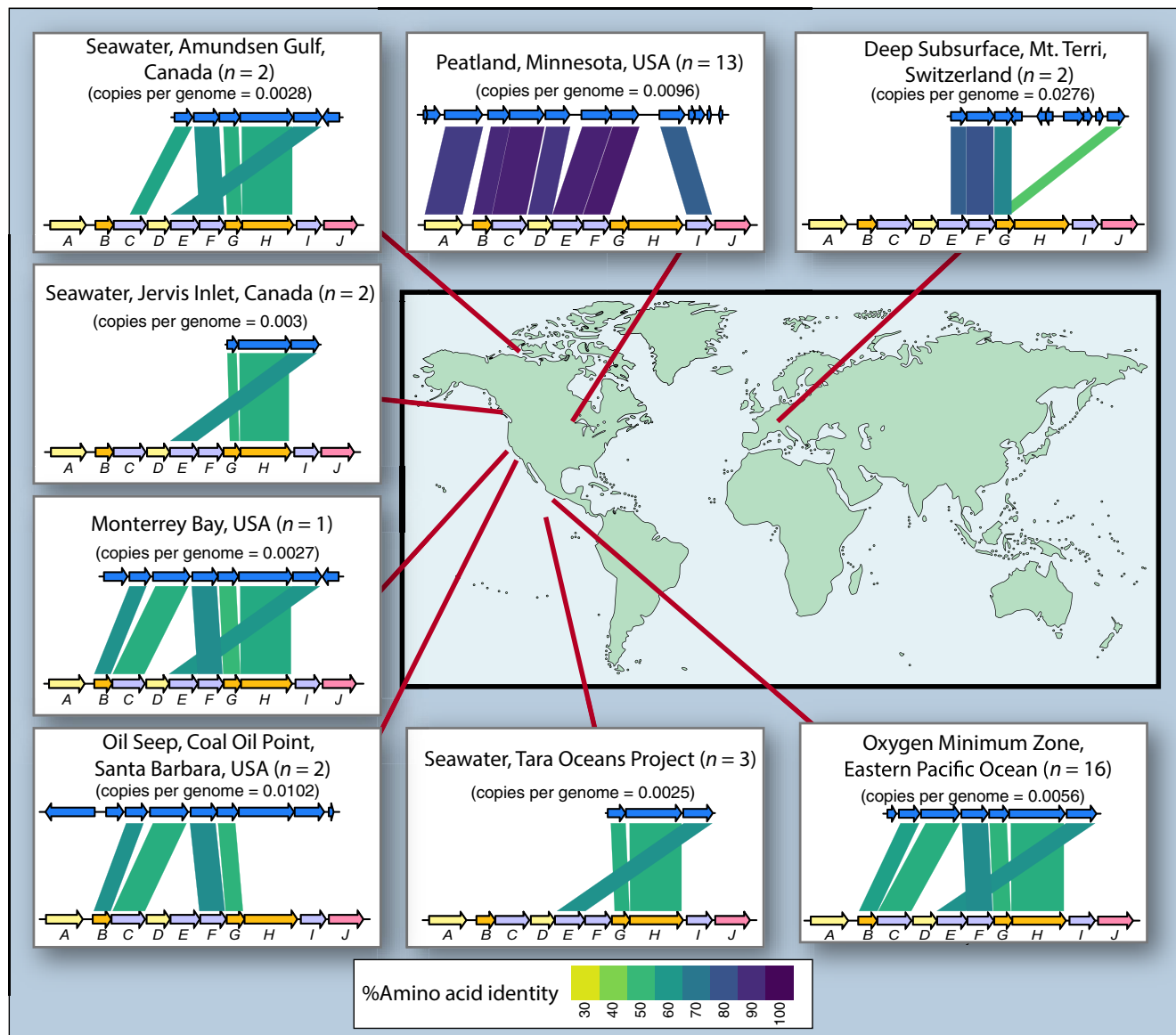


FIGURE 5 Environmental distribution of 41 gene cassettes with homology and synteny to the Defo *mec* cassette. Cassettes were identified among all 18,314 public metagenomes available in the IMG metagenome database as of January 7, 2020. Details for all cassettes are provided in Dataset S7. Shown are representative *mec* cassettes identified in eight distinct environmental ecosystems. The red lines point to the approximate locations where the metagenomes were derived. In each gene cassette illustration, the lower, multicolor gene cassette represents the *Dehalobacterium formicoaceticum mec* cassette, while the blue, upper illustration represents the gene cassette found in the indicated environmental metagenome. "n" indicates the number of individual metagenomes in which a syntenic cassette was identified, while "copies per genome" indicates the estimate of cassette copies per total genomes present, as determined by comparing cassette read depth to the average read depth of ten single-copy, protein-encoding genes throughout the metagenome. Colored boxes between genes indicate the blastp-derived amino acid identity allowing comparisons of gene-product identities, where color corresponds to degree of identity as indicated in the legend

of the known DCM degraders (Figure S3). The deep subsurface *mec* cassettes were present in 2.8% of total estimated genome copies.

3.5 | Genomes of chloroform-respiring organisms harbor *mec* gene orthologs

The *mec* gene orthologs comprise cohesive, deeply branching clades (Figure S3). Aside from *mecJ*, no close homologs (i.e., proteins with

>35% amino acid identity) to the *mec* genes are found in any publicly available bacterial or archaeal genomes, with three notable exceptions. Homologs of *mecA*, *mecB*, and *mecC* are found on genomes of the CF respirers *Dehalobacter* sp. strain CF and *Dehalobacter* sp. strain DCA, both of which were reported to generate DCM as an end product during growth with CF as electron acceptor (Grostern et al., 2010). In both of these genomes, the *mecABC* homologs are found immediately adjacent to the CF reductive dehalogenase and anchor protein encoding genes *cfrAB*, suggesting functional association

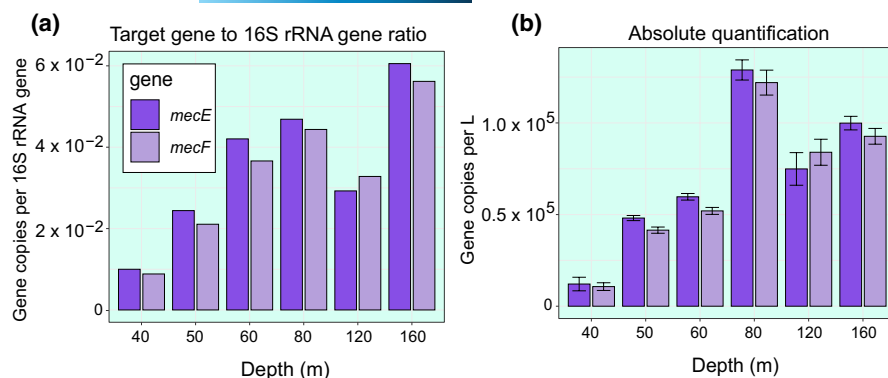


FIGURE 6 qPCR enumeration of *mecE* and *mecF* in samples from the Eastern Tropical North Pacific OMZ water column. Panel a presents ratios of *mecE* or *mecF* gene to 16S rRNA gene copies. Panel b presents absolute quantification of gene copies per liter of seawater with standard deviations represented by error bars ($n = 3$)

between the two gene clusters. In the third case, a complete 10-gene *mec* cassette is located on the genome of *Dehalobacter* sp. strain UNSWDHB (Figure 2), also a CF-respiring organism that lacks the ability to use DCM (Wong et al., 2016). Close inspection of the *mec* cassette of strain UNSWDHB reveals that *mecE*, implicated in the initial chloromethyltransfer reaction, is truncated at the 5' end, which is projected to lead to a ~70 amino acid shorter protein (Figure S4), consistent with a loss of function (Wong et al., 2016).

4 | DISCUSSION

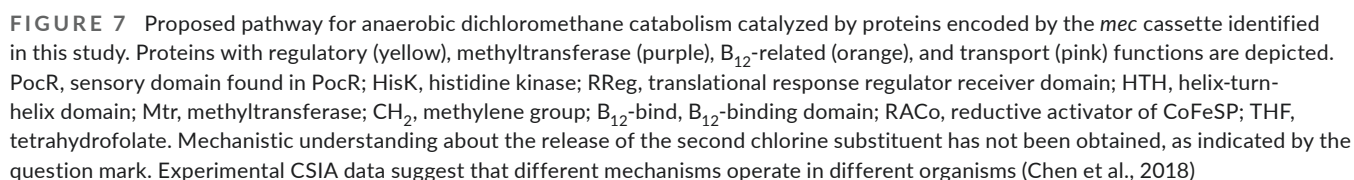
4.1 | Identification of dichloromethane biomarker genes

The 10-gene *mec* cassette was initially identified by comparative genomic analyses between the anaerobic DCM-degrading bacteria *Dehalobacterium formicoaceticum* (Defo) and "*Ca. Dichloromethanomonas elyunquensis*" (Diel). A third highly similar and syntenic gene cassette was also identified in the more recently described DCM degrader '*Ca. Formimonas warabiya*' (Fowa). This high degree of gene identity and gene synteny among DCM degraders and the near total absence of any closely related gene cassette in any other bacterial genome strongly suggest that this gene cassette is involved in anaerobic DCM catabolism. In support of the hypothesis, Mec proteins were among the most abundant proteins in the proteomes of DCM-grown Defo and Diel. A recent report details growth of Diel with dichloroacetate (Chen et al., 2021), and a re-examination of the published proteomic data revealed that the MecB, MecC, MecE, MecF, MecH, and MecI proteins Diel encodes on *mec* cassette B were 7.45–8.38 log₂fold more abundant when growth occurred on DCM versus dichloroacetate (Table S4). Enrichment of anaerobic digester sludge with DCM led to an increase in the abundance of *mec* genes from undetectable to levels similar to those observed in Defo and Diel cultures growing with DCM (i.e., 10⁶ to 10⁸ gene copies per ml). An enrichment of *mec* genes was also apparent in the microbiomes associated with a DCM contaminant plume where *mec* gene expression correlated with DCM concentrations. Thus, multiple lines of evidence support a strong and exclusive coordination between anaerobic DCM metabolism and *mec* cassette gene, transcript and protein abundances.

4.2 | Functional annotations of the *mec* cassette are consistent with a dichloromethane dehalogenating methyltransferase system

The DCM-degrading bacteria are strict anaerobes and no genes consistent with aerobic DCM catabolism by a GST (e.g., *dcmA*) were identified in the genomes of Defo, Diel, or Fowa. Previous biochemical studies on Defo demonstrated that DCM is channeled into the WLP via methylene-THF catalyzed by a corrinoid methyltransferase (Chen et al., 2018, 2020; Mägli et al., 1996, 1998). Guided by functional annotation (Table 2) and comparison to analogous metabolic systems, the proteins encoded by the *mec* cassette genes assemble to a DCM catabolic pathway that is compatible with available biochemical evidence (Figure 7).

The core system involved in C₁ group transfer is possibly composed of two methyltransferases, MecE and MecF, and a corrinoid-binding protein, MecB. MecE shares 23%–26% identity and 39%–45% similarity with proteins of the MtaA/CmuA family (TIGR01463), an ortholog family defined by methyltransferases employed during metabolism of methylamine (MtbA), methanol (MtaA), methanethiol (MtsA), and chloromethane (CmuA). MtbA, MtaA, and MtsA are corrinoid-CoM methyltransferases used during methylotrophic methanogenesis by *Methanosarcina barkeri* and related organisms (Paul et al., 2000). CmuA is a dual-domain methyltransferase/corrinoid protein that catalyzes dehalogenation and an initial methyl transfer reaction during chloromethane utilization by *Hyphomicrobium chloromethanicum* strain CM2 (McAnulla et al., 2001). The affiliation between MecE and the methyltransferase domain of CmuA renders MecE a candidate for catalyzing an initial chloromethyl transfer reaction concomitant with the release of one chlorine substituent from DCM during C₁ group transfer to the corrinoid-binding protein MecB. Such a reaction would yield a hypothetical chloromethyl corrinoid protein (Figure 7). Methyltransferase MecF is a member of the MtrH enzyme family responsible for transfer of a corrinoid-bound methyl group to the organic methyl carrier tetrahydromethanopterin (THMPT), a cofactor for C₁ transfer analogous to tetrahydrofolate (THF). Based on this orthology, MecF is the best candidate for the final methyltransfer reaction to THF generating methylene-THF, for which biochemical evidence exists (Mägli et al., 1998). This methyltransfer reaction could lead to C-Cl bond cleavage and release of the second chlorine substituent. Alternatively, another enzyme system



bacteria (Price et al., 2018). Thus, MecGH may function as a reductive activator for MecB. While close *mecG* and *mecH* homologs are absent in the Fowa *mec* cassette, the Fowa genome is rich in genes encoding proteins with the same functional domains (six with the pfam14574 RACo domain and nine with the pfam07796 DUF1638 domain); one RACo-encoding gene (IMG_2718344850, GenBank, ATW27267.1) shares 43% predicted amino acid identity with the Defo MecG and lies in close proximity to a DUF1638-encoding gene. MecJ is predicted to be a monovalent cation:proton antiporter, possibly supporting DCM catabolism by maintaining acid-base homeostasis (Pinner et al., 1994; Roosild et al., 2010), counteracting acidification of the cytosol by hydrochloric acid (i.e., proton) generation during dechlorination of DCM (Ferguson et al., 2000). Co-localization of dehalogenase genes (e.g., reductive dehalogenases, haloacid dehalogenases) and *mecJ* homologs is common among related organisms (Table S5). The gene

cluster *cmuABC* for aerobic chloromethane utilization is 6 kb upstream and in the same orientation of a gene (IMG ID 650984269) encoding a putative chloride carrier/channel (CIC) protein, a protein family also implicated in acid-base homeostasis (Iyer et al., 2002), and which was required for efficient growth on DCM via heterologous expression of *dcmA* (Michener et al., 2014). In organisms employing DcmA for DCM degradation, regulation is accomplished by DcmR (La Roche & Leisinger, 1991); however, genes encoding proteins with the DcmR sensor domain (METHanogen/methylotroph, DcmR Sensory domain (MEDS), pfam14417) were absent from Diel and Defo genomes, while one such gene was identified in Fowa (GenBank WP_148133634), although at a distant location and with a different domain organization than found in DcmR. Among the *mec* cassette proteins, functional annotations suggest that MecA represents a DCM sensor kinase that regulates the DNA-binding protein MecD following detection of DCM via the PocR (pfam10114) domain (Anantharaman & Aravind, 2005). In the marine *mec* cassettes, the regulatory genes *mecA* and *mecD* were not detected; however, in two cases, homologs of *dcmR*, were identified directly adjacent to the *mec* cassette (Figure 8) and could fulfill regulatory functions. Furthermore, the close proximity of marine *mec* cassettes to homologs of a gene encoding a proven DCM sensor and regulator add an additional line of support to the proposed role of the *mec* cassette in DCM catabolism.

MecC and MecI both have weak identity to MecE (23% full-length identity and maximum 27% identity across a 40% alignment, respectively), weak similarity with the MtaA/CmuA enzyme

family (up to 40% amino acid similarity across partial alignments), and no further similarities with any other protein(s) of known function. Both were abundant in proteomic experiments, MecC in the top 1% along with MecE and MecF, and MecI in the top quartile. *mecC* was also frequently found in environmental *mec* cassettes, although *mecI* was detected only in some of the peat systems (Figure 5). The apparent involvement of additional methyltransferases is not unprecedented but has yet to be explained mechanistically. For example, the chloromethane utilization gene cluster in *Methylobacterium* sp. strain CM4 also contains a third putative methyltransferase encoding gene, *cmuC*, which is required for growth with chloromethane but whose function has yet to be revealed (Vannelli et al., 1999). Methanol-grown *Desulfotomaculum kuznetsovii*, which uses a related methanol corrinoid methyltransferase system, also expresses three methyltransferases all located in the same gene cluster (Sousa et al., 2018). Additional details of the functional annotations of the *mec* cassette genes are provided in Supplementary Information.

4.3 | Highly similar *mec* cassettes are broadly distributed in the environment

A search of public metagenomes led to identification of *mec* cassettes in disparate ecosystems, including peatland, marine OMZs, and the deep subsurface. There is good reason to suspect each of

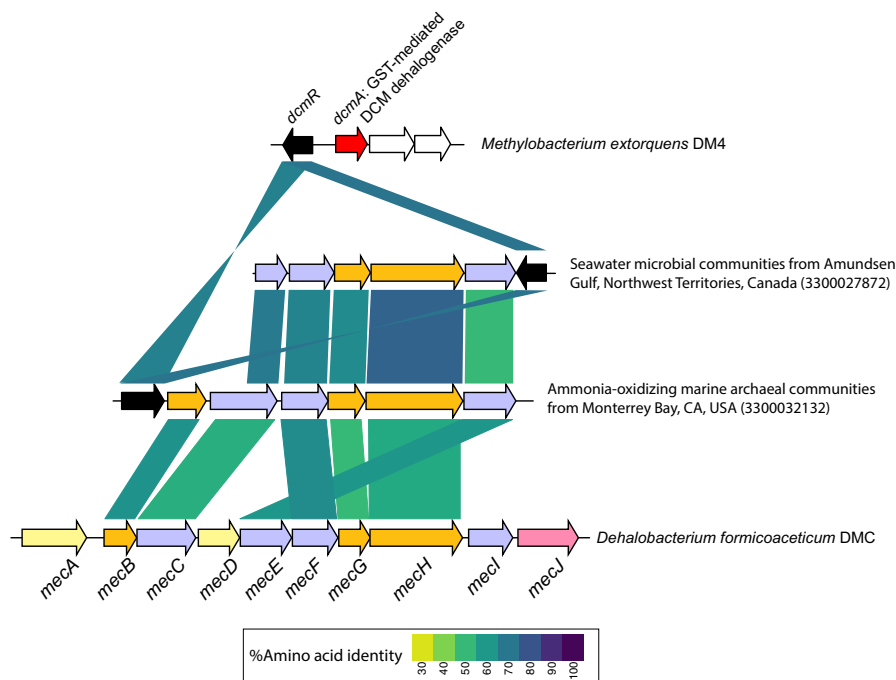


FIGURE 8 Dichloromethane (DCM) regulatory gene *dcmR* homologs are adjacent to marine *mec* cassettes. Two distinct *mec* cassettes identified in metagenomes from Monterey Bay, CA and Amundsen Gulf, CA are flanked by genes, colored in solid black, encoding proteins annotated as DcmR sensory domain and regulator of DCM dehalogenase DcmA (KEGG K17071). IMG ID numbers for these two metagenomes are provided in parentheses. These proteins share 70.6% amino acid identity and are 44.6%–46.4% identical to DcmR of *Methylobacterium extorquens* strain DM4. The *Methylobacterium extorquens* gene cassette (*dcmRA*) encodes glutathione-S-transferase DCM dehalogenase (DcmA), colored red, and associated sensor and regulatory protein DcmR, colored in solid black

these environments as being hotspots for DCM flux. Peat bogs have been demonstrated to be rich in chlorinated organic material (Biester et al., 2004) and have been identified as a source of halomethanes (Dimmer et al., 2001). The occurrence of halomethanes in subsurface rock formations has been demonstrated (Mulder et al., 2013), although information about quantities and fluxes is lacking. Marine systems are net producers of DCM and considered major emitters to the atmosphere (Gribble, 2010). In the water column, DCM concentrations peak along with chlorophyll concentrations (Ooki & Yokouchi, 2011), consistent with production by phytoplankton and the abiotic chlorination of planktonic iodo- and bromomethanes (Ooki & Yokouchi, 2011). DCM is detected beneath the photic zone (Ooki & Yokouchi, 2011), suggesting that mixing events carry DCM into deeper waters, or other sources of DCM exist in the deep ocean, possibly hydrothermal vents (Eklund et al., 1988; Isidorov et al., 1990; Jordan et al., 2000) or settling dead biomass (Wever & Barnett, 2017). Of note, the analysis of ETNP OMZ samples using targeted qPCR led to a much higher detection rate (9 of 11) than was obtained by metagenome analyses (16 of 90). This increase in frequency of *mec* gene detection is likely due to the higher sensitivity of qPCR compared to shotgun metagenomics (Suttner et al., 2020), and implies that the *mec* cassette is more broadly distributed in marine systems than the metagenomics survey results suggest.

4.4 | Implications of widespread anaerobic dichloromethane degradation potential

DCM predates the anthropocene (Trudinger et al., 2004), and the mechanisms underlying natural releases are far from fully characterized. Based on the available information, DCM is an energy source readily available to microorganisms (Cox et al., 2004; Gribble, 2010; Kolusu et al., 2018), and the anaerobic microbial consumption of DCM is likely a major attenuation factor, eliminating DCM in anoxic environments prior to atmospheric release, and thus a relevant process for reducing DCM emissions. Environmental change, including global warming, has high potential to alter the flux of DCM with unpredictable consequences for emissions and the integrity of the ozone layer. Whether low-oxygen marine systems and peat bogs, for example, are net producers or net sinks of DCM is currently unclear. OMZs are expanding at accelerating rates worldwide (Stramma et al., 2008), and uncertainty exists over the impact of environmental change on net DCM emissions. Likewise, climate change induced melting of permafrost in the northern hemisphere will create more active peat bogs, and warming of peat bogs is expected to mobilize recalcitrant carbon and stimulate the breakdown of accumulated organic materials (Gill et al., 2017), which will likely increase halomethane, including DCM formation, in these critical zone environments. As is the case with low oxygen marine systems, the degree to which DCM production is counterbalanced by consumption in peat bogs is unknown. A widely distributed phenotype for

anaerobic DCM catabolism is likely to affect DCM fluxes and pool sizes, and thus the global DCM budget, which until present considers atmospheric releases and abiotic stratospheric breakdown, but not microbial attenuation (i.e., sinks prior to release to the atmosphere). The increased knowledge of microbial DCM catabolism offers opportunities to include relevant sink and attenuation terms and generate refined flux models with more predictive power.

5 | CONCLUSIONS

DCM is an increasing threat to stratospheric ozone with both anthropogenic and natural emission sources. Anaerobic bacterial metabolism of DCM has not yet been taken into consideration as a factor in the global DCM cycle. The discovery of the *mec* gene cassette associated with anaerobic bacterial DCM metabolism and its widespread distribution in environmental systems highlight an attenuation potential for DCM. Knowledge of the *mec* cassette offers new opportunities to delineate DCM sources, enables more robust estimates of DCM fluxes, supports refined DCM emission modeling and simulation of the stratospheric ozone layer, reveals a novel, broadly distributed C_1 carbon metabolic system, and provides prognostic and diagnostic tools supporting bioremediation of groundwater aquifers impacted by DCM.

ACKNOWLEDGMENTS

We thank F. Stewart and A. Bertagnolli for valuable discussions and providing Eastern Tropical North Pacific Oxygen Minimum Zone DNA samples for qPCR analysis. T. Macbeth, K. Sorenson, and R. Chenenko from CDM Smith graciously provided groundwater samples and metatranscriptomic data from a DCM-contaminated site. The authors acknowledge N. Ivanova and R. Seshadri from the Joint Genome Institute for providing invaluable insight into use of the IMG database.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All genomes, genes, and protein sequences studied are available in the IMG database (Chen et al., 2019) under the specified ID numbers. Corresponding NCBI ID numbers for *mec* cassettes located on genomes are provided in Dataset S9. All IMG-derived genes and proteins from genomic *mec* cassettes, proteins from metagenomic *mec* cassettes, all *mec* homolog gene and protein sequences detailed in Figure S3, and all supplemental materials can be downloaded directly from Figshare (<https://doi.org/10.6084/m9.figshare.17698604.v1>). Code used for data processing and figure creation is available via RPubs links located in the methods section. All spectral data collected from the Defo axenic cultures used in this study have been deposited in the MassIVE and ProteomeXchange repositories with identifiers MSV000087235 and PXD025479, respectively. Spectral data collected from the Diel axenic cultures

used in this study have been deposited in the MassIVE and ProteomeXchange repositories with identifiers MSV000086520 and PXD022742, respectively. Scripts used to perform gene copy per genome copy calculations, pairwise gene cassette alignments, and to construct gene cassette synteny plots can be found at https://rpubs.com/rmurdoch/mec_cassette_abundance_and_synteny. Detailed description of gene phylogeny pipelines can be found at https://rpubs.com/rmurdoch/mec_cassette_trees. Full metatranscriptome analysis pipeline description and scripts used to operate programs can be found at https://rpubs.com/rmurdoch/mec_transcriptomes. Processing of proteome data to produce plots shown in Figure 3 is documented at https://rpubs.com/rmurdoch/mec_proteome_jitterplots.

ORCID

Frank E. Löffler  <https://orcid.org/0000-0002-9797-4279>

REFERENCES

- Allocati, N., Federici, L., Masulli, M., & Di Ilio, C. (2012). Distribution of glutathione transferases in Gram-positive bacteria and Archaea. *Biochimie*, 94(3), 588–596. <https://doi.org/10.1016/J.BIOCHI.2011.09.008>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- An, M., Western, L. M., Say, D., Chen, L., Claxton, T., Ganesan, A. L., Hossaini, R., Krummel, P. B., Manning, A. J., Mühle, J., O'Doherty, S., Prinn, R. G., Weiss, R. F., Young, D., Hu, J., Yao, B., & Rigby, M. (2021). Rapid increase in dichloromethane emissions from China inferred through atmospheric observations. *Nature Communications*, 12(1), 1–9. <https://doi.org/10.1038/s41467-021-27592-y>
- Anantharaman, V., & Aravind, L. (2005). MEDS and PocR are novel domains with a predicted role in sensing simple hydrocarbon derivatives in prokaryotic signal transduction systems. *Bioinformatics*, 21(12), 2805–2811. <https://doi.org/10.1093/bioinformatics/bti418>
- Bagnoud, A., Chourey, K., Hettich, R. L., de Bruijn, I., Andersson, A. F., Leupin, O. X., Schwyn, B., & Bernier-Latmani, R. (2016). Reconstructing a hydrogen-driven microbial metabolic network in Opalinus Clay rock. *Nature Communications*, 7(1), 12770. <https://doi.org/10.1038/ncomms12770>
- Biester, H., Keppler, F., Putschew, A., Martinez-Cortizas, A., & Petri, M. (2004). Halogen retention, organohalogen, and the role of organic matter decomposition on halogen enrichment in two Chilean peat bogs. *Environmental Science and Technology*, 38(7), 1984–1991. <https://doi.org/10.1021/es0348492>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34(5), 525–527. <https://doi.org/10.1038/nbt.3519>
- Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)*, 25(15), 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Chen, G., Fisch, A. R., Gibson, C. M., Mack, E. E., Seger, E. S., Campagna, S. R., & Löffler, F. E. (2020). Mineralization versus fermentation: Evidence for two distinct anaerobic bacterial degradation pathways for dichloromethane. *The ISME Journal*, 14, 959–970. <https://doi.org/10.1038/s41396-019-0579-5>
- Chen, G., Jiang, N., Solis, M. I. V., Murdoch, F. K., Murdoch, R. W., Xie, Y., Swift, C. M., Hettich, R. L., & Löffler, F. E. (2021). Anaerobic microbial metabolism of dichloroacetate. *mBio*, 12(2), e00537-21. https://doi.org/10.1128/MBIO.00537-21/SUPPL_FILE/MBIO.00537-21-ST003.XLSX
- Chen, G., Kleindienst, S., Griffiths, D. R., Mack, E. E., Seger, E. S., & Löffler, F. E. (2017). Mutualistic interaction between dichloromethane- and chloromethane-degrading bacteria in an anaerobic mixed culture. *Environmental Microbiology*, 19(11), 4784–4796. <https://doi.org/10.1111/1462-2920.13945>
- Chen, G., Murdoch, R. W., Mack, E. E., Seger, E. S., & Löffler, F. E. (2017). Complete genome sequence of *Dehalobacterium formicoaceticum* strain DMC, a strictly anaerobic dichloromethane-degrading bacterium. *Genome Announcements*, 5(37), e00897-17. <https://doi.org/10.1128/genomeA.00897-17>
- Chen, G., Shouakar-Stash, O., Phillips, E., Justicia-Leon, S. D., Gilevska, T., Sherwood Lollar, B., Mack, E. E., Seger, E. S., & Löffler, F. E. (2018). Dual carbon–chlorine isotope analysis indicates distinct anaerobic dichloromethane degradation pathways in two members of *Peptococcaceae*. *Environmental Science & Technology*, 52(15), 8607–8616. <https://doi.org/10.1021/acs.est.8b01583>
- Chen, I.-M.-A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M., Varghese, N., White, J. R., Seshadri, R., Smirnova, T., Kirton, E., Jungbluth, S. P., Woyke, T., Elie-Fadrosh, E. A., Ivanova, N. N., & Kyrpides, N. C. (2019). IMG/M v.5.0: An integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Research*, 47(D1), D666–D677. <https://doi.org/10.1093/nar/gky901>
- Cox, M. L., Fraser, P. J., Sturrock, G. A., Siems, S. T., & Porter, L. W. (2004). Terrestrial sources and sinks of halomethanes near Cape Grim, Tasmania. *Atmospheric Environment*, 38(23), 3839–3852. <https://doi.org/10.1016/J.ATMOSENV.2004.03.050>
- Dimmer, C. H., Simmonds, P. G., Nickless, G., & Bassford, M. R. (2001). Biogenic fluxes of halomethanes from Irish peatland ecosystems. *Atmospheric Environment*, 35(2), 321–330. [https://doi.org/10.1016/S1352-2310\(00\)00151-5](https://doi.org/10.1016/S1352-2310(00)00151-5)
- Eklund, G., Pedersen, J. R., & Stromberg, B. (1988). Methane, hydrogen chloride and oxygen form a wide range of chlorinated organic species in the temperature range 400°C–950°C. *Chemosphere*, 17(3), 575–586. [https://doi.org/10.1016/0045-6535\(88\)90032-X](https://doi.org/10.1016/0045-6535(88)90032-X)
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, 47, D427–D432. <https://doi.org/10.1093/nar/gky995>
- Eng, J. K., McCormack, A. L., & Yates, J. R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, 5(11), 976–989. [https://doi.org/10.1016/1044-0305\(94\)80016-2](https://doi.org/10.1016/1044-0305(94)80016-2)
- Eustáquio, A. S., Pojer, F., Noel, J. P., & Moore, B. S. (2008). Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nature Chemical Biology*, 4(1), 69–74. <https://doi.org/10.1038/nchembio.2007.56>
- Ferguson, G. P., Booth, I. R., Evans, G. J., & Vuilleumier, S. (2000). Growth inhibition of *Escherichia coli* by dichloromethane in cells expressing dichloromethane dehalogenase/glutathione S-transferase. *Microbiology*, 146(11), 2967–2975. <https://doi.org/10.1099/00221287-146-11-2967>
- Ganusova, E. E., Vo, L. T., Abraham, P. E., O'Neal Yoder, L., Hettich, R. L., & Alexandre, G. (2021). The *Azospirillum brasilense* core chemotaxis proteins CheA1 and CheA4 link chemotaxis signaling with nitrogen metabolism. *mSystems*, 6(1), e01354-20. <https://doi.org/10.1128/msystems.01354-20>
- Gill, A. L., Giasson, M.-A., Yu, R., & Finzi, A. C. (2017). Deep peat warming increases surface methane and carbon dioxide emissions in a black

- spruce-dominated ombrotrophic bog. *Global Change Biology*, 23(12), 5398–5411. <https://doi.org/10.1111/gcb.13806>
- Gossett, J. M. (1987). Measurement of Henry's Law constants for C1 and C2 chlorinated hydrocarbons. *Environmental Science and Technology*, 21(2), 202–208. <https://doi.org/10.1021/es00156a012>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644–652. <https://doi.org/10.1038/nbt.1883>
- Gribble, G. W. (2010). *Naturally occurring organohalogen compounds—A comprehensive update* (Vol. 91). Springer Vienna. <https://doi.org/10.1007/978-3-211-99323-1>
- Groster, A., Duhamel, M., Dworatzek, S., & Edwards, E. A. (2010). Chloroform respiration to dichloromethane by a *Dehalobacter* population. *Environmental Microbiology*, 12(4), 1053–1060. <https://doi.org/10.1111/j.1462-2920.2009.02150.x>
- Guy, L., Roat Kultima, J., & Andersson, S. G. E. (2010). genoPlotR: Comparative gene and genome visualization in R. *Bioinformatics*, 26(18), 2334–2335. <https://doi.org/10.1093/bioinformatics/btq413>
- Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., & Beck, E. (2012). TIGRFAMs and genome properties in 2013. *Nucleic Acids Research*, 41(D1), D387–D395. <https://doi.org/10.1093/nar/gks1234>
- Haines, A., Ebi, K. L., & Li, S. (2020). Wildfires, global climate change, and human health. *The New England Journal of Medicine*, 383(22), 2173–2181. <https://doi.org/10.1056/NEJMSr2028985>
- Hatt, J. K., & Löffler, F. E. (2012). Quantitative real-time PCR (qPCR) detection chemistries affect enumeration of the *Dehalococcoides* 16S rRNA gene in groundwater. *Journal of Microbiological Methods*, 88(2), 263–270. <https://doi.org/10.1016/j.mimet.2011.12.005>
- Hayoun, K., Geersens, E., Laczy, C. C., Halder, R., Sánchez, C. L., Manna, A., Bringel, F., Ryckelynck, M., Wilmes, P., Muller, E. L., Alpha-Bazin, B., Armengaud, J., & Vuilleumier, S. (2020). Dichloromethane degradation pathway from unsequenced *Hyphomicrobium* sp. MC8b rapidly explored by pan-proteomics. *Microorganisms*, 8(12), 1876. <https://doi.org/10.3390/MICROORGANISMS8121876>
- Hoekstra, E. J., Verhagen, F. J. M., Field, J. A., Leer, E. W. B. D., & Brinkman, U. A. T. (1998). Natural production of chloroform by fungi. *Phytochemistry*, 49(1), 91–97. [https://doi.org/10.1016/S0031-9422\(97\)00984-9](https://doi.org/10.1016/S0031-9422(97)00984-9)
- Holland, S. I., Edwards, R. J., Ertan, H., Wong, Y. K., Russell, T. L., Deshpande, N. P., Manefield, M., & Lee, M. J. (2019). Whole genome sequencing of a novel, dichloromethane-fermenting *Peptococcaceae* from an enrichment culture. *PeerJ*, 7, e7775. <https://doi.org/10.7287/peerj.preprints.27718v1>
- Holland, S. I., Ertan, H., Montgomery, K., Manefield, M. J., & Lee, M. (2021). Novel dichloromethane-fermenting bacteria in the *Peptococcaceae* family. *The ISME Journal*, 15, 1709–1721. <https://doi.org/10.1038/s41396-020-00881-y>
- Hossaini, R., Chipperfield, M. P., Montzka, S. A., Leeson, A. A., Dhomse, S. S., & Pyle, J. A. (2017). The increasing threat to stratospheric ozone from dichloromethane. *Nature Communications*, 8(1), 15962. <https://doi.org/10.1038/ncomms15962>
- Hu, S., Niu, Z., Chen, Y., Li, L., & Zhang, H. (2017). Global wetlands: Potential distribution, wetland loss, and status. *Science of the Total Environment*, 586, 319–327. <https://doi.org/10.1016/J.SCITOTENV.2017.02.001>
- Isidorov, V. A., Zenkevich, I. G., & Ioffe, B. V. (1990). Volatile organic compounds in sulfataric gases. *Journal of Atmospheric Chemistry*, 10, 329. <https://doi.org/10.1007/BF00053867>. <https://link.springer.com/content/pdf/10.1007%2FBF00053867.pdf>
- Iyer, R., Iverson, T. M., Accardi, A., & Miller, C. (2002). A biological role for prokaryotic CIC chloride channels. *Nature*, 419(6908), 715–718. <https://doi.org/10.1038/nature01000>
- Jordan, A., Harnisch, J., Borchers, R., Le Guern, F., & Shinohar, H. (2000). Volcanogenic halocarbons. *Environmental Science & Technology*, 34(6), 1122–1124. <https://doi.org/10.1021/ES990838Q>
- Justicia-Leon, S. D., Ritalahti, K. M., Mack, E. E., & Löffler, F. E. (2012). Dichloromethane fermentation by a *Dehalobacter* sp. in an enrichment culture derived from pristine river sediment. *Applied and Environmental Microbiology*, 78(4), 1288–1291. <https://doi.org/10.1128/AEM.07325-11>
- Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., & MacCoss, M. J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods*, 4(11), 923–925. <https://doi.org/10.1038/nmeth1113>
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(D1), D353–D361. <https://doi.org/10.1093/nar/gkw1092>
- Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology*, 428(4), 726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>
- Kanters, J., & Louw, R. (1996). Thermal and catalysed halogenation in combustion reactions. *Chemosphere*, 32(1), 89–97. [https://doi.org/10.1016/0045-6535\(95\)00230-8](https://doi.org/10.1016/0045-6535(95)00230-8)
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., & Valencia, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)*, 28(12), 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kleindienst, S., Chourey, K., Chen, G., Murdoch, R. W., Higgins, S. A., Iyer, R., Campagna, S. R., Mack, E. E., Seger, E. S., Hettich, R. L., & Löffler, F. E. (2019). Proteogenomics reveals novel reductive dehalogenases and methyltransferases expressed during anaerobic dichloromethane metabolism. *Applied and Environmental Microbiology*, 85(6), e02768–e2818. <https://doi.org/10.1128/AEM.02768-18>
- Kleindienst, S., Higgins, S. A., Tsementzi, D., Chen, G., Konstantinidis, K. T., Mack, E. E., & Löffler, F. E. (2017). '*Candidatus* Dichloromethanomonas elyunquensis' gen. nov., sp. nov., a dichloromethane-degrading anaerobe of the *Peptococcaceae* family. *Systematic and Applied Microbiology*, 40(3), 150–159. <https://doi.org/10.1016/j.syapm.2016.12.001>
- Kleindienst, S., Higgins, S. A., Tsementzi, D., Konstantinidis, K. T., Mack, E. E., & Löffler, F. E. (2016). Draft genome sequence of a strictly anaerobic dichloromethane-degrading bacterium. *Genome Announcements*, 4(2), e00037–e116. <https://doi.org/10.1128/GENOMEA.00037-16>
- Kolusu, S. R., Schlünzen, K. H., Grawe, D., & Seifert, R. (2018). Determination of chloromethane and dichloromethane in a tropical terrestrial mangrove forest in Brazil by measurements and modelling. *Atmospheric Environment*, 173, 185–197. <https://doi.org/10.1016/J.ATMOENV.2017.10.057>
- La Roche, S. D., & Leisinger, T. (1991). Identification of *dcmR*, the regulatory gene governing expression of dichloromethane dehalogenase in *Methylobacterium* sp. strain DM4. *Journal of Bacteriology*, 173(21), 6714–6721.
- Lane, D. J. (1991). 16S/23S rRNA sequencing. In E. Stackebrandt, & M. Goodfellow (Eds.), *Nucleic acid techniques in bacterial systematics* (pp. 115–175). John Wiley & Sons, Ltd.

- Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, 44(W1), W242–W245. <https://doi.org/10.1093/nar/gkw290>
- Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>
- Löffler, F. E., Sanford, R. A., & Ritalahti, K. M. (2005). Enrichment, cultivation, and detection of reductively dechlorinating bacteria. *Methods in Enzymology*, 397, 77–111. [https://doi.org/10.1016/S0076-6879\(05\)97005-5](https://doi.org/10.1016/S0076-6879(05)97005-5)
- Mägli, A., Messmer, M., & Leisinger, T. (1998). Metabolism of dichloromethane by the strict anaerobe *Dehalobacterium formicoaceticum*. *Applied and Environmental Microbiology*, 64(2), 646–650. <http://www.ncbi.nlm.nih.gov/pubmed/16349505>
- Mägli, A., Wendt, M., & Leisinger, T. (1996). Isolation and characterization of *Dehalobacterium formicoaceticum* gen. nov. sp. nov., a strictly anaerobic bacterium utilizing dichloromethane as source of carbon and energy. *Archives of Microbiology*, 166(2), 101–108. <https://doi.org/10.1007/s002030050362>
- McAnulla, C., Woodall, C. A., McDonald, I. R., Studer, A., Vuilleumier, S., Leisinger, T., & Murrell, J. C. (2001). Chloromethane utilization gene cluster from *Hyphomicrobium chloromethanicum* strain CM2T and development of functional gene probes to detect halomethane-degrading bacteria. *Applied and Environmental Microbiology*, 67(1), 307–316. <https://doi.org/10.1128/AEM.67.1.307-316.2001>
- McCulloch, A. (2017). Dichloromethane in the environment: A note prepared for the European Chlorinated Solvents Association (ECSA) and the Halogenated Solvents Industry Alliance (HSIA). <https://www.chlorinated-solvents.eu/wp-content/uploads/2021/05/Dichloromethane-paper.pdf>
- Michener, J. K., Neves, A. A. C., Vuilleumier, S., Bringel, F., & Marx, C. J. (2014). Effective use of a horizontally-transferred pathway for dichloromethane catabolism requires post-transfer refinement. *eLife*, 3(November), 1–16. <https://doi.org/10.7554/ELIFE.04279>
- Mulder, I., Huber, S. G., Krause, T., Zetzsch, C., Kotte, K., Dultz, S., & Schöler, H. F. (2013). A new purge and trap headspace technique to analyze low volatile compounds from fluid inclusions of rocks and minerals. *Chemical Geology*, 358, 148–155. <https://doi.org/10.1016/J.CHEMGEO.2013.09.003>
- Muller, E. E. L., Bringel, F., & Vuilleumier, S. (2011). Dichloromethane-degrading bacteria in the genomic age. *Research in Microbiology*, 162(9), 869–876. <https://doi.org/10.1016/J.RESMIC.2011.01.008>
- Muyzer, G., De Waal, E. C., & Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology*, 59(3), 695–700. <https://doi.org/10.1128/aem.59.3.695-700.1993>
- Ooki, A., & Yokouchi, Y. (2011). Dichloromethane in the Indian Ocean: Evidence for *in-situ* production in seawater. *Marine Chemistry*, 124(1–4), 119–124. <https://doi.org/10.1016/J.MARCHEM.2011.01.001>
- Pagès, H., Aboyou, P., Gentleman, R., & DebRoy, S. (2019). Biostrings: Efficient manipulation of biological strings (R package version 2.54.0).
- Paul, L., Ferguson, D. J., & Krzycki, J. A. (2000). The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of *Methanosarcina barkeri* contain in-frame and read-through amber codons. *Journal of Bacteriology*, 182(9), 2520. <https://doi.org/10.1128/JB.182.9.2520-2529.2000>
- Petkau, A., Stuart-Edwards, M., Stothard, P., & Van Domselaar, G. (2010). Interactive microbial genome visualization with GView. *Bioinformatics*, 26(24), 3125–3126. <https://doi.org/10.1093/bioinformatics/btq588>
- Pinner, E., Padan, E., & Schuldiner, S. (1994). Kinetic properties of NhaB, a Na⁺/H⁺ antiporter from *Escherichia coli*. *The Journal of Biological Chemistry*, 269, 26274–26279. [https://doi.org/10.1016/S0021-9258\(18\)53333-0](https://doi.org/10.1016/S0021-9258(18)53333-0)
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS One*, 5(3), e9490. <https://doi.org/10.1371/journal.pone.0009490>
- Price, M. N., Zane, G. M., Kuehl, J. V., Melnyk, R. A., Wall, J. D., Deutschbauer, A. M., & Arkin, A. P. (2018). Filling gaps in bacterial amino acid biosynthesis pathways with high-throughput genetics. *PLOS Genetics*, 14(1), e1007147. <https://doi.org/10.1371/journal.pgen.1007147>
- Ritalahti, K. M., Amos, B. K., Sung, Y., Wu, Q., Koenigsberg, S. S., & Löffler, F. E. (2006). Quantitative PCR targeting 16S rRNA and reductive dehalogenase genes simultaneously monitors multiple *Dehalococcoides* strains. *Applied and Environmental Microbiology*, 72(4), 2765–2774. <https://doi.org/10.1128/AEM.72.4.2765-2774.2006>
- Roosild, T. P., Castronovo, S., Healy, J., Miller, S., Pliotas, C., Rasmussen, T., Bartlett, W., Conway, S. J., & Booth, I. R. (2010). Mechanism of ligand-gated potassium efflux in bacterial pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, 107(46), 19784–19789. <https://doi.org/10.1073/pnas.1012716107>
- Seeman, T. (2018). *tseemann/barnnap: Bacterial ribosomal RNA predictor (0.9-2)*. <https://github.com/tseemann/barnnap>
- Sousa, D. Z., Visser, M., van Gelder, A. H., Boeren, S., Pieterse, M. M., Pinkse, M. W. H., Verhaert, P. D. E. M., Vogt, C., Franke, S., Kümmel, S., & Stams, A. J. M. (2018). The deep-subsurface sulfate reducer *Desulfotomaculum kuznetsovii* employs two methanol-degrading pathways. *Nature Communications*, 9(1), 239. <https://doi.org/10.1038/s41467-017-02518-9>
- Stramma, L., Johnson, G. C., Sprintall, J., & Mohrholz, V. (2008). Expanding oxygen-minimum zones in the tropical oceans. *Science*, 320(5876), 655–658. <https://doi.org/10.1126/science.1153847>
- Suttner, B. J., Johnston, E. R., Orellana, L. H., Rodriguez-R, L. M., Hatt, J. K., Carychao, D., Carter, M. Q., & Cooley, M. B. (2020). Potential and limitations of metagenomics as a public health risk assessment tool in a study of natural creek sediments influenced by agricultural and livestock runoff. *Applied and Environment Microbiology*, 86(6), e02525–e2619. <https://doi.org/10.1128/AEM.02525-19>
- Svensen, H., Planke, S., Polozov, A. G., Schmidbauer, N., Corfu, F., Podladchikov, Y. Y., & Jamtveit, B. (2009). Siberian gas venting and the end-Permian environmental crisis. *Earth and Planetary Science Letters*, 277(3–4), 490–500. <https://doi.org/10.1016/J.EPSL.2008.11.015>
- Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000). The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 28(1), 33–36. <https://doi.org/10.1093/nar/28.1.33>. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC102395/>
- Thamdrup, B., Steinsdóttir, H. G. R., Bertagnolli, A. D., Padilla, C. C., Patin, N. V., Garcia-Robledo, E., Bristow, L. A., & Stewart, F. J. (2019). Anaerobic methane oxidation is an important sink for methane in the ocean's largest oxygen minimum zone. *Limnology and Oceanography*, 64(6), 2569–2585. <https://doi.org/10.1002/lno.11235>
- Trudinger, C. M., Etheridge, D. M., Sturrock, G. A., Fraser, P. J., Krummel, P. B., McCulloch, A., & Etheridge, M. (2004). Atmospheric histories of halocarbons from analysis of Antarctic firn air: Methyl bromide, methyl chloride, chloroform, and dichloromethane. *Journal of Geophysical Research*, 109, 22310. <https://doi.org/10.1029/2004JG004932>
- Vannelli, T., Messmer, M., Studer, A., Vuilleumier, S., & Leisinger, T. (1999). A corrinoid-dependent catabolic pathway for growth of a *Methylobacterium* strain with chloromethane. *Proceedings of the National Academy of Sciences of the United States of America*, 96(8), 4615–4620. <https://doi.org/10.1073/pnas.96.8.4615>
- Vit, O., & Petrak, J. (2017). Integral membrane proteins in proteomics. How to break open the black box? *Journal of Proteomics*, 153, 8–20. <https://doi.org/10.1016/J.JPROT.2016.08.006>
- Wever, R., & Barnett, P. (2017). Vanadium chloroperoxidases: The missing link in the formation of chlorinated compounds and chloroform

- in the terrestrial environment? *Chemistry—An Asian Journal*, 12(16), 1997–2007. <https://doi.org/10.1002/asia.201700420>
- Wong, Y. K., Holland, S. I., Ertan, H., Manefield, M., & Lee, M. (2016). Isolation and characterization of *Dehalobacter* sp. strain UNSWDHB capable of chloroform and chlorinated ethane respiration. *Environmental Microbiology*, 18(9), 3092–3105. <https://doi.org/10.1111/1462-2920.13287>
- Wright, J., Kirchner, V., Bernard, W., Ulrich, N., McLimans, C., Campa, M. F., Hazen, T., Macbeth, T., Marabello, D., McDermott, J., Mackelprang, R., Roth, K., & Lamendella, R. (2017). Bacterial community dynamics in dichloromethane-contaminated groundwater undergoing natural attenuation. *Frontiers in Microbiology*, 8, 2300. <https://doi.org/10.3389/fmicb.2017.02300>
- Wu, S., Zhu, Z., Fu, L., Niu, B., & Li, W. (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics*, 12(1), 444. <https://doi.org/10.1186/1471-2164-12-444>
- Wuosmaa, A. M., & Hager, L. P. (1990). Methyl chloride transferase: A carbocation route for biosynthesis of halometabolites. *Science*, 249(4965), 160–162. <https://doi.org/10.1126/science.2371563>
- Yalkowsky, S. H., He, Y., & Jain, P. (2010). *Handbook of aqueous solubility data* (2nd ed.). CRC Press.
- Yang, S., Giannone, R. J., Dice, L., Yang, Z. K., Engle, N. L., Tschaplinski, T. J., Hettich, R. L., & Brown, S. D. (2012). *Clostridium thermocellum* ATCC27405 transcriptomic, metabolomic and proteomic profiles after ethanol stress. *BMC Genomics*, 13(1), 336. <https://doi.org/10.1186/1471-2164-13-336>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Murdoch, R. W., Chen, G., Kara Murdoch, F., Mack, E. E., Villalobos Solis, M. I., Hettich, R. L., & Löffler, F. E. (2022). Identification and widespread environmental distribution of a gene cassette implicated in anaerobic dichloromethane degradation. *Global Change Biology*, 00, 1–17. <https://doi.org/10.1111/gcb.16068>