frontiers in Microbiology

ORIGINAL RESEARCH
published: 31 January 2017
doi: 10.3389/fmicb.2017.00099

Metabolomic and Metagenomic Analysis of Two Crude Oil Production Pipelines Experiencing Differential Rates of Corrosion

Vincent Bonifay[1], Boris Wawrik[1], Jan Sunner[1,2], Emily C. Snodgrass[1], Egemen Aydin[1], Kathleen E. Duncan[1], Amy V. Callaghan[1], Athenia Oldham[3], Turid Liengen[4] and Iwona Beech[1,2]**

[1] Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, USA, [2] Institute for Energy and the Environment, University of Oklahoma, Norman, OK, USA, [3] Department of Biology, University of Texas of the Permian Basin, Odessa, TX, USA, [4] Research Centre Porsgrunn, Statoil ASA, Herøya Industripark, Porsgrunn, Norway

OPEN ACCESS

Edited by:
Pankaj Kumar Arora, M. J. P. Rohilkhand University, India

Reviewed by:
Christopher L. Hemme, University of Rhode Island, USA
Santosh Kr Karn, National Institute of Technology Raipur, India

***Correspondence:**
Boris Wawrik
bwawrik@ou.edu
Iwona Beech
ibeech@ou.edu

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *28 October 2016*
Accepted: *13 January 2017*
Published: *31 January 2017*

Citation:
Bonifay V, Wawrik B, Sunner J, Snodgrass EC, Aydin E, Duncan KE, Callaghan AV, Oldham A, Liengen T and Beech I (2017) Metabolomic and Metagenomic Analysis of Two Crude Oil Production Pipelines Experiencing Differential Rates of Corrosion. Front. Microbiol. 8:99. doi: 10.3389/fmicb.2017.00099

Corrosion processes in two North Sea oil production pipelines were studied by analyzing pig envelope samples via metagenomic and metabolomic techniques. Both production systems have similar physico-chemical properties and injection waters are treated with nitrate, but one pipeline experiences severe corrosion and the other does not. Early and late pigging material was collected to gain insight into the potential causes for differential corrosion rates. Metabolites were extracted and analyzed via ultra-high performance liquid chromatography/high-resolution mass spectrometry with electrospray ionization (ESI) in both positive and negative ion modes. Metabolites were analyzed by comparison with standards indicative of aerobic and anaerobic hydrocarbon metabolism and by comparison to predicted masses for KEGG metabolites. Microbial community structure was analyzed via 16S rRNA gene qPCR, sequencing of 16S PCR products, and MySeq Illumina shotgun sequencing of community DNA. Metagenomic data were used to reconstruct the full length 16S rRNA genes and genomes of dominant microorganisms. Sequence data were also interrogated via KEGG annotation and for the presence of genes related to terminal electron accepting (TEA) processes as well as aerobic and anaerobic hydrocarbon degradation. Significant and distinct differences were observed when comparing the 'high corrosion' (HC) and the 'low corrosion' (LC) pipeline systems, especially with respect to the TEA utilization potential. The HC samples were dominated by sulfate-reducing bacteria (SRB) and archaea known for their ability to utilize simple carbon substrates, whereas LC samples were dominated by pseudomonads with the genetic potential for denitrification and aerobic hydrocarbon degradation. The frequency of aerobic hydrocarbon degradation genes was low in the HC system, and anaerobic hydrocarbon degradation genes were not detected in either pipeline. This is in contrast with metabolite analysis, which demonstrated the presence of several succinic acids in HC samples that are diagnostic of anaerobic hydrocarbon metabolism. Identifiable aerobic metabolites were confined to the LC samples, consistent with the metagenomic

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/48166

data. Overall, these data suggest that corrosion management might benefit from a more refined understanding of microbial community resilience in the face of disturbances such as nitrate treatment or pigging, which frequently prove insufficient to alter community structure toward a stable, less-corrosive assemblage.

Keywords: microbially influenced corrosion (MIC), reservoir microbiology, metabolomics, mass spectrometry, hydrocarbon degradation, anaerobic microbiology

INTRODUCTION

Corrosion of ferrous alloys in oilfield environments is of considerable concern to the oil and gas industry worldwide. In industrialized economies, the cost of corrosion-related damage to industrial infrastructure has been estimated to be as high as 4% of GNP (Heitz, 1992). In the U.S. alone, the National Association of Corrosion Engineers (NACE) estimated that the direct cost of deterioration of metallic materials was $276 billion in 1998, or 3.1% of GNP, with incurred indirect monetary loss being equally high (Koch et al., 2002). Whether in natural or man-made environments, including oilfields, corrosion of metallic materials results from physico-chemical interactions between the metal and its surroundings. These interactions are often strongly influenced by the presence of complex microbial communities thriving as biofilms on internal or external surfaces of metals (Geesey et al., 2000; Beech et al., 2014). It is often assumed that the biofilm-driven corrosion process, referred to as microbiologically influenced corrosion (MIC) or biocorrosion, is initiated, enhanced, or facilitated by the metabolic activity of sessile microbial populations (Beech et al., 2005). In the absence of biofilms, abiotically formed corrosion products adhere to surfaces of deteriorating metals, thus providing a diffusion barrier to reactants and slowing the corrosion reaction (Beech and Sunner, 2004). In the presence of biofilms, however, the metabolic activity of organisms within the biofilm community can considerably alter interfacial electrochemistry between the biofilm matrix and the colonized metallic substrata, either enhancing or inhibiting corrosion reactions (Little and Ray, 2002; Ornek et al., 2002).

The activities of biofilm-associated prokaryotes, such as sulfide-producing prokaryotes (SPP), which in particular include sulfate-reducing bacteria (SRB; Hamilton, 1985; Lee et al., 1995), iron oxidizing/reducing bacteria, manganese-oxidizing bacteria, and bacteria that secrete extracellular polymeric substances (EPS) and organic acids (Beech and Gaylarde, 1999), are associated with metal deterioration and corrosion failures. In oilfield installations, SPPs are frequently invoked as a main cause of MIC (Gu, 2012). Despite extensive efforts to elucidate the roles that biofilms play in MIC and the mechanisms by which a single species or a complex microbial community can mediate corrosion processes, the mechanisms are still disputed. Notably, the relationship between biofilm community dynamics and MIC is poorly understood. Cultivation-based approaches are typically of limited utility, given that the vast majority of microbes cannot be brought into culture using standard cultivation techniques (Amann et al., 1995). As a consequence, nucleic acid based techniques have emerged as the standard approach for profiling of microbial community composition *in situ*. To date, several studies have characterized planktonic microbial communities implicated in MIC via molecular techniques (Dahle et al., 2008; Duncan et al., 2009; Gittel et al., 2009; Stevenson et al., 2011; Suflita et al., 2012). For example, 16S rRNA gene surveys of produced water from the North Sea Troll formation indicated the presence of bacteria belonging to *Firmicutes*, *Bacteroidetes*, and δ-proteobacteria, as well as Archaea related to the genera *Methanococcus* and *Methanolobus* (Dahle et al., 2008). Similar taxonomic groupings have been detected at other production facilities and oil reservoir produced waters, although considerable variability among sites have also been observed, reflecting variations in *in situ* environmental parameters (e.g., reservoir temperature, facility production practices, and availability of oxygen, nitrate, and sulfate; Stevenson et al., 2011). Importantly, when comparing bacterial community structure in production waters from two adjacent North Sea oil reservoirs, which differed solely in the method of controlling sulfide production, it was revealed that nitrate treatment was associated with lower microbial diversity yet higher cell densities, as well as bacterial taxa known to have the capability of nitrate reduction (Gittel et al., 2009).

Although insightful, requisite studies thus far have lacked a broader perspective of *in situ* microbial metabolic capabilities and, importantly, activities. In addition, reports on biofilm community structure and their potential function and potential contribution to corrosion are scarce (Stevenson et al., 2011; Lenhart et al., 2014). Recent advances in DNA sequencing methods (Buermans and den Dunnen, 2014) and rapid progress in techniques of mass spectrometry, in particular metabolomics (Patti et al., 2012; Monteiro et al., 2013), have greatly enhanced our capacity to interrogate both the metabolic potential and the actual metabolic activity of biofilm communities implicated in corrosion (Bonifay et al., 2013; Sunner and Beech, 2015).

Here, we report a combination of metabolomic and metagenomic measurements conducted on material recovered from pigging of two North Sea production pipelines. The aim was to gain insight into the contribution of prokaryotic biofilms to corrosion of these structures, given that they experienced very different levels of deterioration.

MATERIALS AND METHODS

Field Sample Collection

Samples comprising pigging debris were collected from the production pipelines of two adjacent oil-fields, both located in the Norwegian sector of the North Sea. At the time of sampling,

the pipeline of one of the fields had been experiencing ongoing and severe corrosion, whereas the corrosion status in the other pipeline was moderate and of no concern. The pipelines will be designated herein as "HC" (high corrosion) and "LC" (low corrosion), respectively. Pigging debris from sequential pigging runs was collected during routine cleaning operations. For each pipeline, early and late pigging material was recovered by sampling from the third and eleventh pig run, respectively. The samples are here identified as LC3, HC3, LC11, and HC11. Pigging material was collected into sterile glass bottles that had been filled with nitrogen gas. Bottles were filled to exclude head space, placed on ice, and delivered to a laboratory, where approximately 100 g of each sample was removed, acidified, and subsequently frozen and stored at −70°C.

Metabolite Analysis

For each of the four samples, metabolic profiling was performed in triplicate, using ultra-high performance liquid chromatography/high-resolution mass spectrometry (UPLC/HRMS) with electrospray ionization (ESI) in both positive and negative ion modes. For each sample, 1 g was removed and further acidified by the addition of 5 mL of 4 M HCl (ACS grade, EMD Millipore). Five milliliters of a 50/50 (v/v) mixture of ethyl acetate and dichloromethane (ACS grade) were added; the mixture was vortexed for 2 min; and sonicated in a water bath for 15 min. The phases were allowed to separate, and the aqueous layer was removed. The remaining organic solution was evaporated to dryness under a stream of N_2 gas, and the residue was reconstituted in 100 μL of isopropyl alcohol (HPLC grade, EMD Millipore). Five microliters of each solution were injected on an Agilent 1290 Infinity UPLC system, coupled to an Agilent 6538 quadrupole time-of-flight mass spectrometer (QTOF-MS; Agilent, Santa Clara, CA, USA) equipped with a dual-spray ESI source. For negative ion mode, an Acquity UPLC® HSS C18 SB column (1.8 μm, 2.1 mm × 100 mm; Waters Inc., USA) was used for separation with a flow rate of 0.4 mL min^{-1}. The initial concentration of 23.5% of HPLC grade acetonitrile (EMD Millipore) in HPLC grade water (EMD Millipore) was maintained for 5 min, and this was followed by a linear gradient from 23.5 to 95.5% over 35 min, and by 95.5% maintained for an additional 5 min. For positive ion mode, a SeQuant® ZIC®-HILIC column (5 μm, 150 mm × 4.6 mm, The Nest Group, Inc., Southborough, MA, USA) was used at a flow rate of 0.30 mL min^{-1}. An initial concentration of 80% HPLC grade acetonitrile in HPLC grade water was maintained for 5 min; followed by a linear gradient from 80 to 20% over 30 min; and isocratic operation at 5% acetonitrile for an additional 8 min. In positive ion mode only, HPLC solvent mixtures contained 0.1% HPLC grade formic acid (Sigma Aldrich) to improve ionization efficiency. In both positive and negative ion modes, MS parameters were as follows: gas temperature, 325°C; capillary voltage, 3,500 V; fragmentor voltage, 160 V; m/z range, 50–1100; detector signal acquisition rate, 4 GHz; and spectrum storage rate, 1 s^{-1}.

Raw MS data were processed under IDEOM (version 19; Creek et al., 2012) workflow control. IDEOM uses XCMS Centwave (Tautenhahn et al., 2008) for peak detection and mzMatch.R (Scheltema et al., 2011) for alignment of peaks, filtering, and storage of data in peakML files. XCMS (Smith et al., 2006) and mzMatch (Scheltema et al., 2011) are scripts in the R environment. Abundance values are here reported in instrument-specific units (i.u.), and an abundance of 1,000 i.u. corresponds to the detection of about 80 ions. An abundance threshold of three times the noise level, as determined by XCMS, was used for peak-picking. For alignment of features between triplicates, samples and ionization modes at a maximum mass difference of 6 ppm and a retention time difference of 0.25 min were allowed. The parameter settings used for peak-picking and alignment are listed in Supplementary Table S1.

Putative Identification of Metabolites

Detected features were matched against metabolites in IDEOM's version of the KEGG [Kyoto Encyclopedia of Genes and Genomes database (Kanehisa et al., 2014)], as well as the Metacyc (Caspi et al., 2014), Lipidmaps (Fahy et al., 2009), and HMDB (Wishart et al., 2013) databases using the criterion that the experimentally determined exact mass should be within 6 ppm of the theoretical mass of the metabolite. For features detected in positive ion mode (about 72% of the total), identification was also based on comparing experimental and calculated retention times. Creek et al. (2011) described improved metabolite identification using retention time prediction for HILIC chromatography, and the method is implemented in IDEOM. The retention time calculator is based on a quantitative structure-retention relationship model that incorporates six physico-chemical variables in a multiple-linear regression based on 120 authentic standard metabolites. For the present work, 33 metabolite standards were used to calibrate the retention time predictor (Supplementary Table S2). Putative metabolites, whose experimentally observed retention time was more than 45% off from that calculated by the model, were rejected. Metabolites matching eight of the standards were observed in the pigging samples.

With the use of an RT calculator, an identification level of 2 can be claimed (Sumner et al., 2007; Salek et al., 2013), but given the level of uncertainty associated with identification, detected metabolites herein are referred to as "putative." The eight metabolites matching standards, however, were identified with a higher degree of certainty. The IDEOM software assigns a level of confidence to each identification using a numerical value that ranges from 1 to 10, where 10 reflects the highest level of confidence. The confidence value assigned is based on the magnitude of the mass error, on the difference between experimental and expected retention times (for features detected in positive ion mode), and on a database priority order internal to IDEOM. For any feature, the metabolite with the highest confidence score is selected.

After the putative identification of extracted features with metabolites and other compounds in the KEGG database, the (putatively) identified metabolites were matched with the Pathos (Leader et al., 2011) metabolic pathway database to identify 'active' pathways. The grouping of metabolites into named pathways in Pathos is identical to that used in the KEGG database. Finally, targeted analysis was performed for alkylsuccinates and

alkylbenzylsuccinates, known anaerobic degradation products of alkanes and alkylated aromatics, respectively (Agrawal and Gieg, 2013; Rabus et al., 2016). Positive identification required that compounds were observed with the correct mass (± 2 ppm) and retention time (± 10%). Two standard alkylsuccinates, hexylsuccinate, and decylsuccinate, were used. The retention times of other alkylsuccinates were determined using samples from anaerobic alkane degradation cultures known to produce prominent series of alkylsuccinate peaks. For identification of benzylsuccinates, the standards ethylbenzylsuccinate and benzylsuccinate were used.

DNA Extraction and Metagenome Sequencing

DNA was extracted from 3 g of pigging sample using the Maxwell 16 Tissue LEV Total RNA Purification Kit with the automated Maxwell 16 Research System (Promega, Madison, WI, USA). Briefly, 1.5 mL of 10 mM Tris:50 mM EDTA, 500 μL of 10% SDS, and 50 μL of proteinase K (>600 U mL^{-1} solution; QIAGEN, Valencia, CA, USA) were added to the 3 g of sample in a 50-mL Falcon tube. To facilitate cell lysis, the sample mixture was vortexed for 30 s and incubated at room temperature (∼21°C) for 25 min, followed by 5 min at 65°C. To each sample, one tube of lysing matrix E beads (MP Biomedicals, Solon, OH, USA) was added, and samples were bead-beaten on the high setting for 2 min using a vortex adapter (MOBIO Laboratories, Carlsbad, DA, USA). Samples were centrifuged at 6,000 × *g* for 2 min, and the supernatants were transferred to four pre-dispensed reagent cartridges with the addition of 250 μL of RNA lysis buffer (RLA) and 250 μL of RNA dilution buffer (RDB). Samples were processed using the pre-programmed RNA isolation mode with the DNA removal steps omitted. Post-extraction, DNA extracts were centrifuged for 1 min on a low setting to pellet particulate carryover from the extraction procedure. The extracts were pooled and quantified using the Qubit dsDNA BR Assay kit with the Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA, USA) following the manufacturer's instructions. Metagenome sequencing was conducted by generating libraries with 400 bp fragment sizes by first fragmenting DNA with a Covaris Focused-ultrasonicator (Covaris). Separate Illumina libraries with unique index tags were generated for each sample. DNA was then pooled and sequenced using the Illumina MiSeq PE250 platform at the Oklahoma Medical Research Foundation sequencing core.

qPCR for 16S Bacteria and Archaea

16S rRNA gene copy numbers were determined by quantitative PCR (qPCR), as previously described (Wawrik et al., 2012). In brief, amplification was conducted using bacterial 16S rRNA gene qPCR forward primer 27F (5′-AGA GTT TGA TCM TGG CTC AG-3′) and reverse primer 519R (5′-GWA TTA CCG CKG CTG-3′), as well as archaeal-specific qPCR forward primer 8AF (5′-TCC GGT TGA TCC TGC C-3′) and reverse primer A344R (5′-TCG CGC CTG CTC CIC CCC GT-3′). The qPCR reactions were performed in 30 μL of sample using *Power* SYBR® Green PCR Master Mix (Applied Biosystems) with 250 nM of each primer and 2 μL of template DNA in an Applied Biosystems ABI 7300 Real Time PCR System. PCR conditions were as follows: 50°C for 2 min and 95°C for 8 min, followed by 40 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s. Genomic DNA of *Roseobacter denitrificans* Och114 was used as a standard for Bacteria, while a plasmid (pCRII-TOPO®, Invitrogen) containing the complete 16S rRNA gene of *Methanospirillum hungatei* JF-1 served as a standard for Archaea.

16S rRNA Gene-Based Community Analysis

DNA was amplified by targeting partial 16S rRNA genes with the universal (Bacteria and Archaea) primers S-D-Arch-0519-a-S-15 (5′-CAG CMG CCG CGG TAA-3′) and S-D-Bact-785-a-A-21 (5′-GAC TAC HVG GGT ATC TAA TCC-3′) as previously described (Klindworth et al., 2013). Based on *in silico* analysis, these primers exclude eukaryotes, but cover 86.5 and 87.1% of bacterial and archaeal phyla, respectively. If one mismatch is allowed, only candidate divisions WS6, TM7, and OP11, as well as phylum Nanoarchaeota are undetectable with these primers (Klindworth et al., 2013). Forward primers were 5′-M13 tagged to allow the addition of Illumina primers and barcodes (Wawrik et al., 2012). PCR reactions were carried with 2X PCR Master Mix (Fermentas/Thermo Scientific), and products were checked by gel electrophoresis to confirm that a single band was produced. PCR products were cleaned using a QIAquick® PCR Purification Kit (Qiagen), and amplicons were tagged for MiSeq Illumina sequencing by including a unique 8 bp barcode into each amplicon and mixing products in equimolar ratios (Wawrik et al., 2012). MiSeq Illumina sequencing was performed as previously described (Caporaso et al., 2010b) with the modification of an added CC between the adapter and the barcode.

16S rRNA Gene Sequence Classification

For Illumina sequence libraries of 16S rRNA gene PCR products, adapter sequences were removed from raw Illumina reads using Cutadapt (Martin, 2011). Sequences were trimmed using a quality score of 30, and paired-end reads were joined. All unpaired reads were removed from further analysis. Sequences were clustered and assigned taxonomy using the QIIME pipeline (Caporaso et al., 2010b). After de-multiplexing, sequences were clustered into Operational Taxonomic Units (OTUs) using UCLUST at the 95% identity level and screened for the presence of chimeric sequences. A randomly selected set of representative set of sequences were then aligned to the SILVA small subunit rRNA reference alignment[1] using the PyNAST algorithm (Caporaso et al., 2010a). For the analysis of 16S rRNA genes contained in metagenomic datasets, full length 16S sequences were extracted using EMIRGE (Miller, 2013). Adapter and quality trimmed metagenomic reads were utilized (see below). Forty iterations were conducted using the SILVA small subunit rRNA reference alignment with otherwise default parameters. Full length sequences were exported to the Ribosomal RNA Database Project for classification. The most closely related type strains were aligned with metagenomic 16S sequences, and Neighbor-Joining, Parsimony, Maximum

[1]www.arb-silva.de

Likelihood phylogenetic trees were generated using MEGA 6 (Tamura et al., 2013).

Metagenome Analysis

Metagenomic sequence reads were trimmed to remove Illumina index tags and bar codes using Cutadapt (Martin, 2011). Reads were subsequently quality trimmed with a minimum quality score cutoff of 30. Sequences shorter than 50 bp were discarded. Metagenomes were analyzed as both unassembled and assembled datasets. Assembly of paired reads was conducted using Ray (Boisvert et al., 2010) with a K-mer of 31. Open reading frames (ORF) were predicted using the Prokaryotic Dynamic Programming Gene Finding Algorithm (MetaProdigal; Hyatt et al., 2010, 2012). ORF frequency read counts were generated via mapping trimmed reads onto ORFs using Bowtie 2 (Langmead and Salzberg, 2012). All predicted ORFs were compared to the NCBI NR, COG (clusters of orthologous groups), and KEGG databases using the blastN and blastX functions of CAMERA (Sun et al., 2011). ORFs that were assigned to COGs related to S and N cycling, as well as methanogenesis, were compared to the NR database via blastX. Taxonomic identifiers were retrieved from each best hit and then analyzed via MEGAN (Huson et al., 2007) to evaluate the phylogenetic breakdown of each COG in the dataset (Supplementary Table S3). Genome scaffolds of dominant microorganisms were reconstructed via contig and tetranucleotide frequencies as described by Albertsen (Albertsen et al., 2013). Scaffolds were uploaded to Rapid Annotation using Subsystem Technology (RAST[2]) for annotation and genome recruitment analysis. Unassembled reads were exported to MG-RAST for annotation, PCoA and rarefaction analysis, heatmap generation, and KEGG map analysis.

RESULTS

Site Characteristics

The two production pipelines investigated in this study were selected because one of them was experiencing on-going and severe corrosion ("HC"), whereas the other was experiencing low levels of corrosion ("LC"). This difference in corrosion was interesting because the pipelines were associated with adjacent fields in the North Sea, and their physico-chemical environments were very similar to each other (see **Table 1**). Specifically, the temperature (≈80°C), oil:water ratios, acetate, and propionate concentrations were similar in both pipelines; however, the sulfate concentration was approximately eightfold higher in water obtained from the LC compared to the HC pipe. At the time of sampling, injection water at both fields had been treated with nitrate over a period of several years. Sea water injection at HC started in 1999, while nitrate treatment did not commence until 2001 after significant souring and corrosion problems were first noticed. Production in the LC system started in 1997. Biocide treatment with glutaraldehyde of the injection water (sea water) occurred between 1997 and 2007, after which nitrate treatment was started and biocide treatment ended.

[2]rast.nmpdr.org

16S rRNA Gene Surveys

The microbial community composition was assessed via sequencing of 16S rRNA gene PCR products (Supplementary Figure S1). A total of 1,126 OTU's were detected in bacterial PCR-based 16S rRNA gene sequence data across all samples. Of these, 15 OTUs accounted for >80% of all read data, indicating a high degree of enrichment for a small number of lineages. The LC pipeline was overwhelmingly dominated by Gammaproteobacteria in both the early (74% in LC3) and the late pigging material (72% in LC11; Supplementary Figure S1B). In contrast, Gammaproteobacteria only accounted for 3.2 and 2.7% of reads in the HC3 and HC11 pigging material, respectively. HC sequence libraries contained comparatively greater proportions of sequences classified within the Thermotogae, Synergistes, and Firmicutes. Sequences classified as methanogenic archaeal lineages fell within several methanomicrobial OTUs as well as a single OTU within the Methanococci. Methanomicrobial lineages collectively accounted for less than 0.04% of methanococcal reads in each of the four libraries. The methanococcal OTU accounted for less than 0.2% of the reads in LC samples but for 3 and 21% of reads in the HC3 and HC11 samples, respectively. A significant portion of reads were classified within marine epipelagic taxa, accounting for 12% of LC3 and LC11 reads, and for 4.3 and 4.8% of HC3 and HC11 16S RNA sequences observed (Supplementary Figure S2). Among the detected epipelagic marine lineages were *Synechococcus*, *Prochlorococcus*, Marine Group I Crenarchaeota, *Pelagibacter ubique* (a.k.a. Sar11), *Oceanospirillum*, and a diverse group of 16S gene sequences originating from marine chlorophytes and chromalveolates including heterokonts, haptophytes, and dinoflagellates. This suggests that the LC pipeline contained a greater proportion of recently introduced surface seawater than the HC pipeline.

Metagenomic Surveys

To generate a deeper classification of dominant microbes, full length 16S rRNA gene sequences were reconstructed from metagenomic data using EMIRGE (Miller, 2013) (**Figure 1**). This analysis produced a total of 30 full length bacterial and archaeal 16S rRNA sequences. With respect to the dominant microbial populations, this analysis mirrored results from PCR-based 16S rRNA gene sequencing. Gammaproteobacterial lineages that dominated in the LC samples were most closely related to marine pseudomonads and to halo- and alkali-tolerant halomonads and *Methylophaga* spp. Early pigging material at the HC site (HC3) was dominated by bacteria most closely related to thiosulfate-reducing Firmicutes within the genus *Clostridium* and deltaproteobacteria closely related to sulfate/thiosulfate-reducing *Desulfonauticus* spp. Bacteria related to *Desulfonauticus* spp. were also most abundant in the 16S sequences from late HC pigging debris (HC11), which also contained a dominant population of bacteria most closely related to thermophilic fermentative *Thermoanaerobacter* spp. Emirge analysis also reconstructed 16S rRNA genes for lineages closely related to *Clostridium, Methanothermococcus, Thermovirga, Desulfothermus, Fusibacter, Kosmotoga*, and *Thermococcus*.

TABLE 1 | Site characteristics.

Site	LC pipeline	HC pipeline
In situ corrosion rate*	<0.1 mm y^{-1}	0.5–5.3 mm y^{-1}
Nitrate treatment	From inception	Started after aggressive corrosion was observed
Pressure support	Nitrate treated SW	Nitrate treated SW
In situ biomass	High	Low
Pipeline	12 km, 12″ carbon steel (X65)	10 km, 10″ carbon steel (0.5% Cr)
Temperature	> 80°C	80°C
Oil:water ratio	1:3	2:5
Sulfate (mg L^{-1})	980	121
Phosphate (mg L^{-1})	<0.5	<0.5
Acetate (mg L^{-1})	330	311
Propionate (mg L^{-1})	30	31
Pigs analyzed	3 and 11	3 and 11
# Bases sequenced	2.20×10^9 and 2.11×10^9	2.20×10^9 and 7.54×10^8

**As determined by coupon weight loss (n = 19).*

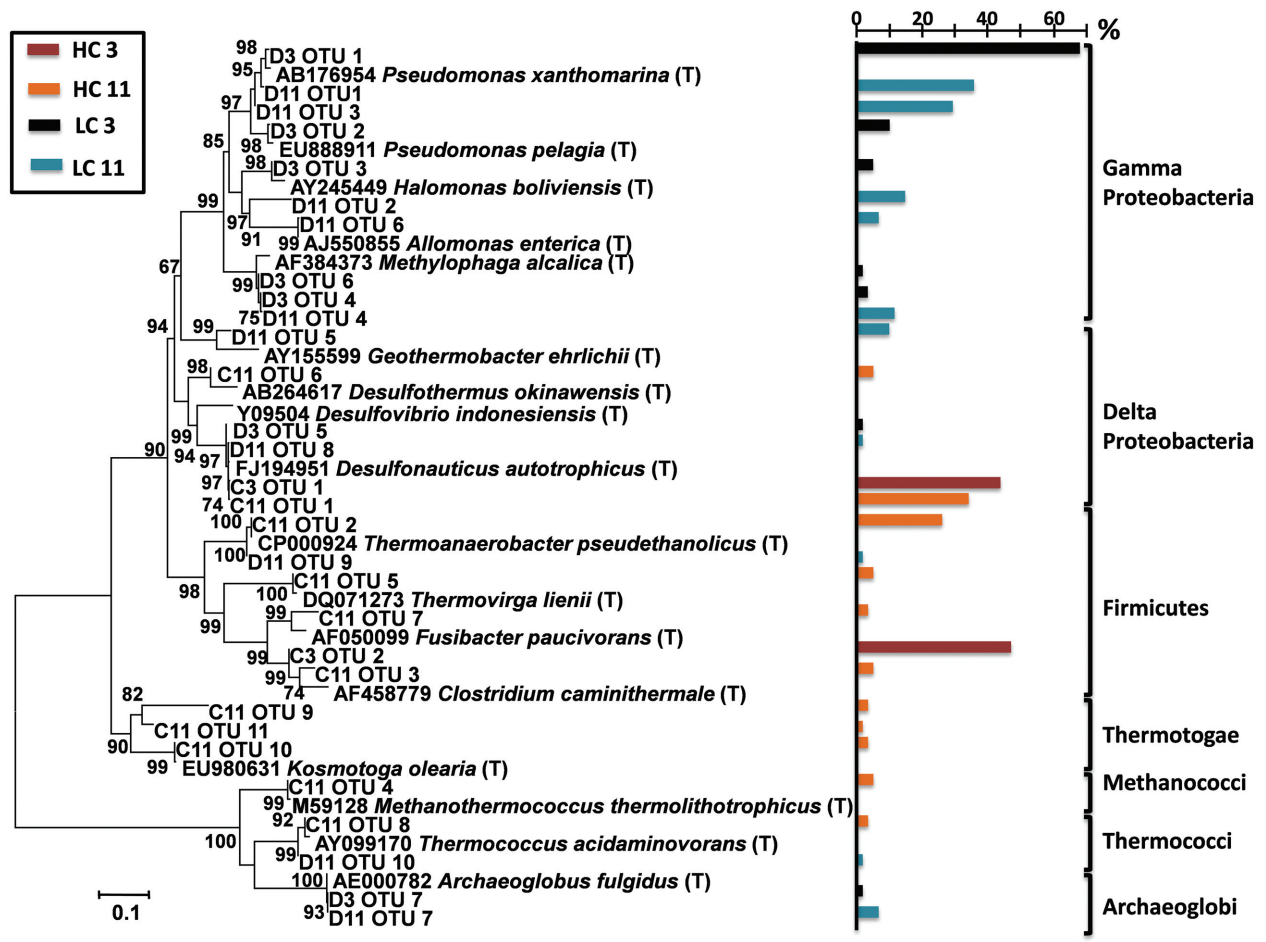


FIGURE 1 | Phylogenetic analysis of full length 16S rRNA genes extracted from metagenomic data using EMIRGE (Miller, 2013). Operational Taxonomic Units (OTUs) were compared to the RDP database, and the most closely related 16S rRNA genes from type strains (designated with a T) were used for phylogenetic analysis of metagenomic OTUs. A bootstrapped (5,000 replicates) Neighbor-Joining tree is shown. The colored bars indicated the proportion of 16S reads that were assembled into respective full length 16S sequences for each of the detected phylotypes (percentage scale shown in scale bar above). Bacterial and archaeal divisions/families are indicated and color coded. D3 = LC3; D11 = LC11; C3 = HC3; and C11 = HC11.

To gain insight into the metabolic capabilities of these microbial populations, several analyses were performed. First, all metagenomic reads were classified into COGs via CAMERA. Reads associated with individual COGs for denitrification, methanogenesis, and sulfate reduction (Supplementary Table S3) were blasted against NCBI's non-redundant nucleotide database, and the identities of their best matches were used to extract phylogenetic information. The phylogenetic breakdown of a subset of these COGs is shown in **Figure 2**. COG analysis indicated that methanogens were limited to the HC11 sample and belonged to the class of Methanococci (**Figure 2A**). Nitrogen cycling genes were primarily associated with Gammaproteobacteria found mostly in the LC samples. In the HC11 sample, Archaeoglobi appear to contain genes for nitrogen fixation (COGs 1348 and 2710; **Figure 2B**). Sulfate reduction COGs were more diverse (**Figure 2C**) and suggested the presence of sulfate-reducing Clostridia, Deltaproteobacteria, and Archaeoglobi in the HC samples. In LC samples, these COGs yielded a number of hits to Gammaproteobacteria. However, the latter are not known to reduce sulfate, and these hits are more likely the result of homology among domain structures of enzymes with unrelated functions.

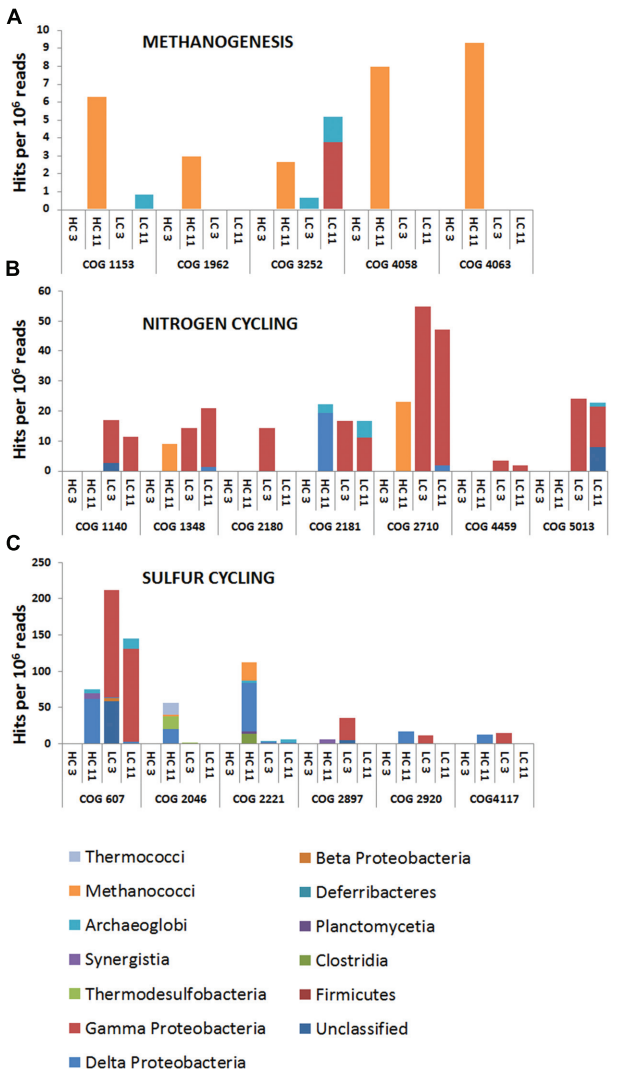


FIGURE 2 | Cluster of Orthologous Groups (COG) analysis of metagenomic data from high corrosion (HC) and low corrosion (LC) pigging samples. Open reading frames (ORF) were predicted from assembled, paired-end read data. Predicted ORFs were compared to the COG database in CAMERA (Sun et al., 2011) and are shown as hit frequencies per million reads (*y*-axis) where the frequencies were calculated by mapping read data onto predicted ORFs and normalizing to the respective metagenome size. Frequencies for COGs involved in **(A)** methanogenesis, **(B)** nitrogen cycling, and **(C)** sulfur cycling.

For more targeted analysis, genome scaffolds for dominant microorganisms were reconstructed using multi-metagenomic analysis (Albertsen et al., 2013) by clustering assembled contigs based on coverage, %GC, phylogenetic affiliation of housekeeping genes, and tetranucleotide frequencies (Supplementary Figure S3). This analysis allowed the reconstruction of 15 binned genomes (referred to as genome 'scaffolds'; **Table 2**), which were annotated and assigned to their closest relative based on recruitment analysis of predicted genes via RAST (Overbeek et al., 2014). Annotations were interrogated for the presence of key N and S cycling genes, revealing that denitrification genes were exclusively present in two LC scaffolds most closely related to *Pseudomonas stutzeri* (Supplementary Table S4). Sulfate reduction genes, in contrast, were detected in both LC and HC scaffolds and were associated with the reconstructed genomes of the Deltaproteobacteria *Desulfonatronospira thiodismutans* and *Desulfohalobium retbaense*, as well as the archaeon *Archaeoglobus fulgidis* (Supplementary Table S5). These data suggest that the LC and HC communities differ in their potential to utilize sulfate and nitrate as terminal electron acceptors. It can be estimated from the size of the metagenomes and the coverage of the reconstructed genomes that 29 and 34% of cells in the HC3 and HC11 but only 1.5 and 1.6% in the LC3 and LC11 pigging communities, respectively, are capable of sulfate reduction (**Figure 3**). Conversely, it is estimated that 78 and 39% of cells in LC3 and LC11 samples contain genes for denitrification, while such genes were absent from any of the HC genome scaffolds. Similar results with respect to denitrification potential are obtained when read data are considered directly via MG-RAST analysis (**Table 3A**). The frequency of reads annotated as nitrogen cycling genes in LC samples was consistent with the interpretation that the majority of cells contained the capability to reduce nitrate. In contrast, N-cycling genes were almost absent from HC metagenomes. MG-RAST based blast analysis of S cycling genes produced hits in all four metagenomes (**Table 3B**), a result that was consistent with the analysis of full length gene sequences generated by assembly and genome binning (Supplementary Table S5).

Given the presence of crude oil in the HC and LC pipelines, pigging envelope metagenomes were further analyzed for the presence of genes associated with aerobic and anaerobic hydrocarbon degradation by evaluating the MG-RAST annotations. A search for reads similar to aerobic hydrocarbon degradation genes revealed large differences in

TABLE 2 | Genome recruitment of genome scaffolds extracted from metagenomic data.

Site	Contigs	Total length	N50	% G+C	Coverage pig #3	Coverage pig #11	# of Open reading frames (ORFS)	Best recruitment	Recruitment hits
Low Corrosion (LC)	123	4749114	58352	61.42	161.5	19.1	4720	*Pseudomonas stutzeri*	2279
	21	3244921	436384	44.66	72.3	21.1	3087	*Methylophaga thiooxidans*	1301
	118	4521009	56215	52.31	52.3	28.6	4075	*Shewanella amazonensis*	436
	43	4853824	189984	62.79	37.2	175.5	2942	*Pseudomonas stutzeri*	2467
	67	2261788	43799	48.08	28.6	6.5	2571	*Archaeoglobus fulgidis*	2196
	126	1895321	28962	51.91	13.7	7.5	2070	*Thermococcus kodakarensis*	646
	116	2136066	33623	37.97	8.5	13.4	2185	*Desulfonatronospira thiodismutans*	319
High Corrosion (HC)	51	2132562	63261	37.94	220	111	2144	*Desulfonatronospira thiodismutans*	342
	40	2131658	85501	51.87	191.3	1.9	2350	*Thermococcus kodakarensis*	1121
	84	2335091	38988	34.24	89.1	7.9	2421	*Thermoanaerobacter pseudethanolicus*	703
	21	1256481	81700	32.1	38.9	0.9	1255	*Methanothermococcus okinawensis*	284
	36	2640115	123629	34.43	20.5	9.7	2638	*Desulfohalobium retbaense*	164
	85	1907033	30617	41.69	19.2	3.8	1838	*Kosmotoga olearia*	1645
	142	3318928	33498	33.18	12.9	33.1	3198	*Alkaliphilus metalliredigens*	370
	138	1633891	13445	30.53	10.6	9.5	1611	*Thermosipho africanus*	1482

the frequencies of oxygen-dependent, hydrocarbon activation enzymes (i.e., mono- and di-oxygenases) among samples. The LC metagenomes contained, on average, >250-fold more monooxygenase >4- to 8-fold more dioxygenase genes per million reads than the corresponding HC metagenomes. An analysis of the dominant types of mono- and di-oxygenases indicated high potential for the degradation of cyclic ketones, mono-cyclic aromatic compounds, and alkanes by LC-, but not HC-, associated communities (**Table 4A**). Analysis of reconstructed genome scaffolds painted a similar picture, with few genes annotated as mono- and di-oxygenases in the HC scaffolds. LC scaffolds contained 6.8-fold more hits to mono- and di-oxygenases than HC scaffolds. (Supplementary Table S6). In particular, one LC scaffold which is most similar to *Pseudomonas stutzeri*, as per recruitment analysis, stands out for containing a range of genes annotated as dioxygenases involved in aromatic hydrocarbon degradation (Supplementary Table S6B).

Genes associated with known anaerobic hydrocarbon degradation pathways were essentially absent from both the HC and the LC metagenomes as determined by MG-RAST analysis. To confirm this observation, blastP searches of predicted proteins from each genome scaffold were conducted against a custom database of known proteins involved in anaerobic hydrocarbon degradation (Callaghan and Wawrik, 2016) revealing no significant hits. In total, only four reads (three in the HC and one in the LC samples; **Table 4B**) with significant blast hits against the catalytic subunit of naphthylmethylsuccinate synthase (NmsA) in the Deltaproteobacterium NaphS6 (Annweiler et al., 2000) were observed, but further analysis (data not shown) did not support a phylogeny that would classify these reads as proteins that catalyze hydrocarbon addition to fumarate, but rather placed them in association with known pyruvate formate lyases. Similarly, a number of reads produced blast hits to acetophenone carboxylase, but closer analysis suggested that these sequences contained a motif common to many types of carboxylases that are not related to hydrocarbon degradation.

Metabolomic Surveys

Pigging debris from oil pipelines contain a large diversity of chemicals, including microbial metabolites, crude oil constituents, material introduced with the seawater, industrial chemicals such as corrosion inhibitors, as well as metal ions from pipeline corrosion and their complexes. The main objective of the metabolomic analysis was to identify groups of features that potentially distinguish two systems experiencing differential corrosion rates. Total ion chromatograms for the four samples are shown in Supplementary Figure S4. In all,

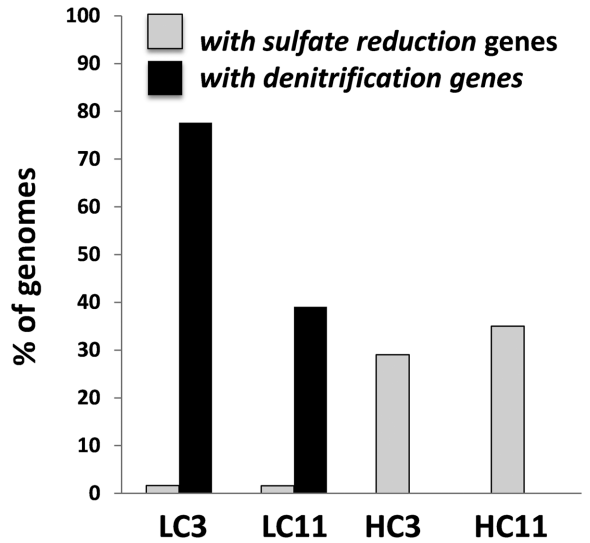


FIGURE 3 | Estimated relative abundance of cells in each of the pigging debris samples carrying sulfate reduction and denitrification pathway genes. The estimates were generated by considering the proportion of all metagenomic reads that assembled into individual genome scaffolds reconstructed from metagenomic data using multi-metagenome (Albertsen et al., 2013) (**Table 2**; Supplementary Figure S3). Each genome scaffold was then annotated via Rapid Annotation using Subsystem Technology (RAST) to identify the presence of denitrification and sulfate reduction pathways (Supplementary Tables S3–S5).

9,824 aligned features were detected, with 5,917 features in positive ion mode and 3,964 features in negative ion mode. Included in both numbers are 57 features that were found in both ionization modes. Each feature is associated with 12 abundance values (i.e., three for each set of triplicate analyses of the four pipeline samples). Features with an abundance below <1,000 i.u. in all four samples were removed from further analysis. Below this abundance value, the probability of detection decreases substantially, as determined from a log/log plot of the number of detected features as a function of the peak abundance (Suflita et al., 2016). The remaining 5,465 features, of which 3,041 were detected in all four samples, ranged in abundance by more than four orders of magnitude among the samples. The reproducibility of the abundance measurements in this set of features was excellent as demonstrated by the observation that for ∼90% of the triplicate measurements the distribution of relative standard deviations was in agreement with what is expected based on ion counting shot noise alone (Suflita et al., 2016; data not shown).

When comparing the HC with the LC system, the sum of the abundances of Pigs 3 and 11 were used. The ratio of the abundances in the two systems ranged over more than six orders of magnitude; however, for more than 50% of the features, the abundances in the two systems were similar, differing by a ratio smaller than a factor of 3 (Supplementary Figure S5). However, more than 2,000 features were over- or under-represented in one or other of the systems by a factor of three or more. It is thus apparent that the metabolic activities in the two systems are very different.

Of particular interpretive value are those features that both have high abundance and exhibit large abundance differences among samples. Using the criterion of a minimum abundance (>10,000 i.u.) in at least one of the pipelines and a minimum abundance ratio of 5, about 400 features were found to be more abundant in the HC system, versus about 220 features that were more abundant in the LC system.

The relative similarities and differences between the metabolomes of the four samples are illustrated by the result of a principal component analysis (PCA; **Figure 4B**). The figure shows that all four metabolomes were distinctly different, though LC3 and LC11 are the pair of samples most similar to each other. The PCA pattern in **Figure 4B** was very robust. For example, while only abundant features (>10,000 i.u. in at least one sample) were used for this analysis, a very similar graph resulted when including all 5,465 features or whether a linear or logarithmic abundance scale was used. The notable exception was that the relative similarity of the HC3 and LC3 metabolomes was shown to be largely due to the presence of high-abundance compounds. Most of these were not identified in the KEGG database and, presumably, they were not of microbial origin, but likely derived from injected seawater, extracted petroleum, and/or industrial chemicals added to the systems.

Of the 5,465 features retained for analysis, 2,352 were putatively matched with metabolites in the KEGG database (Supplementary Table S7). The percentage of features that was putatively identified with known metabolites was 43% overall, but it was substantially different in the four samples. Thus, for the 200 most abundant features in each respective sample, the ratios of identified to non-identified features were 1.47, 0.87, 0.80, and 0.35 for HC11, HC3, LC11, and LC3, respectively.

TABLE 3A | Frequency of N-cycling genes in metagenomic data as determined by MG-RAST annotation.

	NITROGEN CYCLING										
Sample	***amoA***	***narB***	***nasA***	***nifH***	***ureC***	***nirS***	***nirK***	***nosZ***	***norB***	***narG***	***napA***
LC 3	200.5	0.0	17.2	116.3	9.3	4.0	26.5	331.5	506.8	798.2	393.6
LC 11	175.3	0.0	11.7	133.6	35.8	22.3	12.3	173.9	503.7	709.3	104.0
HC 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.6	0.0	1.6	0.8
HC 11	0.0	10.5	0.0	0.8	0.3	0.0	0.0	0.0	1.0	0.5	0.0

 # hits/10^6 reads. The numbers of reads annotated with respective gene identifiers (columns) per million metagenomic reads are shown.

TABLE 3B | Frequency of S cycling genes in metagenomic data as determined by MG-RAST annotation.

	SULFATE REDUCTION								
Sample	***dsrA***	***dsrB***	***dsrk***	***dsrJ***	***dsrP***	***dsrM***	***dsrE***	***dsrF***	***dsrH***
LC 3	9.6	8.2	8.2	0.0	0.0	3.1	67.5	54.8	27.4
LC 11	47.6	25.6	4.1	0.0	0.9	0.9	56.7	72.8	36.7
HC 3	0.0	8.8	132.8	0.0	4.8	58.4	49.6	1.6	1.6
HC 11	20.1	22.2	64.7	0.5	7.4	54.5	58.4	12.0	0.8

	SULFUR OXIDATION					THIOSULFATE REDUCTION		
Sample	***soxA***	***soxB***	***soxX***	***soxY***	***soxZ***	**Rhodanese domain**	**Thiosulfate sulfotransferase**	**Thiosulfate reductase**
LC 3	15.2	0.3	2.8	3.4	42.6	987.4	314.2	87.0
LC 11	42.6	0.0	0.0	0.0	138.9	828.8	282.0	19.7
HC 3	3.2	0.0	0.0	0.0	0.0	547.2	80.0	32.8
HC 11	18.4	0.0	0.0	0.0	2.5	71.6	23.5	7.9

 # hits/10^6 reads. The numbers of reads annotated with respective gene identifiers (columns) per million metagenomic reads are shown.

Among the 2,353 putative metabolites, peptides and lipids were the numerically dominating compound groups, comprising 671 di-, tri-, and tetra-peptides and 590 lipids (sterol lipids, sphingolipids, prenols, polyketides, glycerophospholipids, and fatty acyls). While peptides were more numerous, lipids contributed ∼4X as much as the peptides to the total metabolite abundance. A large group of putatively identified metabolites (247) were mapped to "housekeeping" processes (Supplementary Table S8), including metabolism of amino acids (168) and carbohydrates (37). Secondary metabolism was represented by 110 putative compounds, and a total of 735 putatively identified metabolites were not associated with any specific metabolic pathway in the KEGG database. The "richness" of the HC metabolome was largely due to a wide variety of high-abundance lipids. Other putative metabolites preferentially present in the HC system were involved in biosynthesis and/or metabolism of steroid lipids, phenylpropanoids, porphyrins, arachidonic acid, unsaturated fatty acids, taurine, and isoprenoids. Further, iron ions were detected at a high abundance in both HC samples, and they were four times more abundant in HC11 than in HC3.

Given the dominating presence of petroleum in the pipelines, targeted searches were also performed for compounds that are indicative of known aerobic and anaerobic hydrocarbon degradation pathways (Supplementary Table S8). A number of both aerobic and anaerobic degradation products were detected. The metabolites with abundances above 5×10^3 are listed in **Table 5**. Alkylsuccinates and benzylalkylsuccinates are of particular importance as they are strongly indicative of anaerobic degradation of alkanes and aromatics via addition to fumarate (Agrawal and Gieg, 2013; Rabus et al., 2016) respectively, and these compounds were identified using standards. A wide range of both alkylsuccinates and alkylbenzylsuccinates were detected in all samples. The corresponding extracted ion chromatograms can be found in Supplementary Figure S6. Signatures consistent with carboxylated phenanthrene were also observed suggesting potential for anaerobic aromatic ring carboxylation, primarily in the HC samples.

DISCUSSION

The problem addressed in this study is the observation that different rates of corrosion occur in two very similar oil pipeline systems in adjacent fields in the North Sea, and the ultimate objective is to approach an explanation for this observation. Both metagenomic and metabolomic analyses point to strong microbial activity in both systems. This is consistent with aggressive MIC observed in the HC system, but does not explain the relatively low rates of corrosion in the LC pipeline. The physico-chemical characteristics of produced waters in both systems are very similar (**Table 1**) suggesting that abiotic processes may not be the differentiating factor, but rather that differences in corrosion rates are rooted in the biological activity occurring in the two systems.

Community Structure

Data described here reveal fundamental differences in the microbial community structure and metabolism associated with LC and HC pig envelopes. With respect to the metabolome, chemicals detected in oil-field systems have both microbial and non-microbial origins. The latter group (referred to as "exochemicals" here), originate from seawater, petroleum, and industrial equipment or are added industrial chemicals (e.g., as corrosion inhibitors). Using the ratio of metabolomics features identified via comparison with catalogs of biological molecules (i.e., KEGG) to non-identified features might thus be a useful measure of the relative importance of microbial metabolites versus exochemicals. This analysis indicates that a significantly larger fraction of the features detected in the HC samples (55%) were putatively identified as known metabolites than in the LC samples (32%). For the 200 most abundant features in each respective sample, the ratios of identified to non-identified features were 1.47, 0.87, 0.80, and 0.35 for HC11, HC3, LC11, and LC3, respectively. In addition to indicating a more dominant microbial signal in the HC system, biological contributions also appear more prominent in later pigging samples (i.e., pig run 11 vs. 3), perhaps reflecting the physical structure and

TABLE 4 | Frequency of hydrocarbon degradation genes in metagenomic data as determined by MG-RAST.

(A) Aerobic hydrocarbon degradation

Sample	(Total) monooxygenases	Alkane monooxygenase	Cyclopentanone monooxygenase	Cyclohexanone monooxygenase	Cyclodecanone monooxygenase	Cyclododecanone monooxygenase	Phenol monooxygenase
LC 3	1331.8	136.7	0.3	79.6	0.0	6.2	3.1
LC 11	1453.0	0.6	0.0	227.0	0.0	0.0	0.0
HC 3	3.2	0.0	0.0	0.0	0.0	0.0	0.0
HC 11	7.6	0.0	0.0	0.8	0.0	0.0	0.0

Sample	Phenylacetone monooxygenase	Xylene monooxygenase	Cytochrome P450 monooxygenase	Alkane hydroxylase	(Other) monooxygenase
LC 3	29.6	16.7	0.3	19.5	1039.9
LC 11	37.0	0.0	0.0	0.0	1188.3
HC 3	0.0	0.0	0.0	0.0	3.2
HC 11	0.3	0.0	0.0	0.0	6.6

Sample	(Total) dioxygenases	Ethylbenzene dioxygenase	Nitropropane dioxygenase	Isopropylcatechol dioxygenase	Hydroxyphenylpyruvate dioxygenase	Phthalate dioxygenase	Benzoate dioxygenase
LC 3	4262.0	250.7	440.2	140.0	557.3	11.6	147.1
LC 11	2401.7	59.0	382.4	28.8	347.5	0.0	19.4
HC 3	572.0	0.0	215.2	0.0	0.0	0.0	0.0
HC 11	523.6	0.0	202.4	0.3	0.0	0.0	0.0

Sample	Ring opening dioxygenases	Catechol dioxygenase	Homogentisate dioxygenase	Isopropylbenzene dioxygenase	Protocatechuate dioxygenase	Toluate dioxygenase	(Other) dioxygenase
LC 3	181.3	399.8	246.2	325.3	32.2	331.2	1199.1
LC 11	90.8	91.0	286.1	96.3	7.9	71.1	921.4
HC 3	0.0	0.0	0.0	0.0	0.0	0.0	356.8
HC 11	0.8	1.0	0.3	0.0	0.0	0.5	318.4

(B) Anaerobic hydrocarbon degradation

Sample	Alkylsuccinate synthase	Benzylsuccinate synthase	Ethylbenzene dehydrogenase	Phenylphosphate carboxylase	Benzene carboxylase	Acetophenone carboxylase
LC 3	0.0	0.0	0.0	0.0	0.0	3.4
LC 11	0.3	0.0	0.0	0.0	0.0	31.1
HC 3	1.6	0.0	0.0	0.0	0.0	0.8
HC 11	0.3	0.0	0.0	0.0	0.0	0.0

 # hits/10^6 reads. The numbers of reads annotated with respective gene identifiers (columns) per million metagenomic reads are shown.

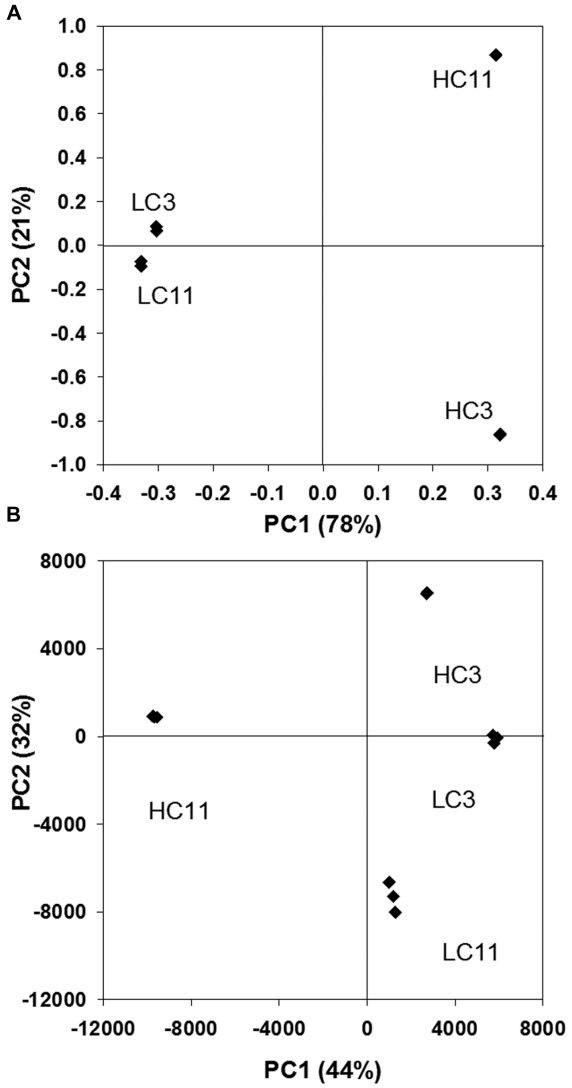


FIGURE 4 | Principal component analysis results (PCA) showing principal components 1 and 2 for (A) metagenomic DNA sequence reads as determined via MG-RAST; and for **(B)** metabolomic data from each triplicate analysis of each of the four samples. Only features with an abundance of at least 10,000 in one of the four samples were included.

densities of microbial biofilms associated with the pipeline walls.

With respect to the metagenomic data, a sharp differentiation between LC and HC samples is reflected by apparent differences in the potential for terminal electron acceptor utilization by dominant community members. This is despite that fact that nitrate addition to injection water had occurred in both systems for several years. Many different bacterial groups have been associated with MIC, including SRB, iron oxidizers, and manganese-oxidizing bacteria (Little and Lee, 2007), though sulfate reducers in particular are often the focus of corrosion management due to the corrosivity of sulfides. In the HC samples, >40% of genome equivalents included the capacity to respire sulfate (**Figure 3**), and the dominant binned genome contained the genes for dissimilatory sulfate reduction. Overall, the features encoded by this genome bin were most similar to the genome of *Desulfonatronospira thiodismutans* (**Table 2**), which is known for high sulfidogenesis under haloalkaline conditions (Sorokin et al., 2008). The type strain of *D. thiodismutans* can perform sulfite/thiosulfate dismutation to sulfide. It can also grow both autotrophically with formate or hydrogen, and utilize short chain organic molecules such as lactate, pyruvate, ethanol, and acetate, which are present in high concentrations in HC production waters (**Table 1**). The second most abundant taxon in the HC metagenomic data was a genome bin most similar to the archaeon *Thermococcus kodakarensis*, which is known for sulfidogenic, heterotrophic growth on organic substrates in the presence of elemental sulfur (Atomi et al., 2004). We note that neither aerobic nor anaerobic hydrocarbon degradation genes were detected in the HC genome scaffolds. HC metagenomic and physiochemical data are therefore consistent with sulfur cycling that is dependent on short chain organic acids, but not hydrocarbons. Conversely, only a small population of sulfate reducers (<5% of cells) was present the LC samples, and by far the most abundant bacterial population was most closely related to *Pseudomonas stutzeri*. *P. stutzeri* is well known for its ability to reduce nitrate, and requisite genome bins contained the necessary genes for the dissimilatory nitrate reduction machinery (Supplementary Table S4). Also present in the LC *P. stutzeri* genome bins are a large diversity of mono- and di-oxygenases related to hydrocarbon utilization. Similarly, this is a feature common to *P. stutzeri* strains, which have been shown to be involved in the utilization of a range of hydrocarbons (Lalucat et al., 2006). LC microbial communities therefore appear to be supported by dissimilatory nitrate reduction, potentially coupled to the utilization of oil hydrocarbons.

Principal component analysis analysis (**Figure 4**) indicates that the patterns of overall differences in the metagenomes and the metabolomes of the four samples analyzed here are quite similar, with the exception that the LC3 and LC11 metagenomes are relatively more similar than the two corresponding metabolomes. This does not necessarily mean that the metabolism of the LC microbial community changes with increasing depth into the deposits, but may simply reflect that the concentrations of oilfield- and other exo-chemicals decrease with increasing depth into the deposits.

Metabolomic Analysis of Hydrocarbon Degradation Pathways

A targeted search of more than 200 hydrocarbon degradation products reported in the literature was conducted in an effort to identify degradation pathways and to determine whether degradation occurred under aerobic and/or anaerobic conditions. Although, fewer positive identifications were made than expected, meaningful results were obtained where standards were available (**Table 5**). Particularly interesting in this context was the presence of several alkylsuccinates and benzylsuccinates, which are diagnostic for anaerobic hydrocarbon degradation

TABLE 5 | Selected hydrocarbon degradation products detected in the LC and HC systems (EIC of the corresponding ions are reproduced in Supplementary Figure S6).

Degradation product	Pathway/substrate[1]		Formula	Experimental mass (Da)	Mass error (mDa)	Retention time (min)	High abundance in samples	Abundance/i.u.
Ethylsuccinic acid	AN	Ethane	$C_6H_{10}O_4$	146.0579	0.1	1.20	HC11	162000
Propylsuccinic acid	AN	Propane	$C_7H_{12}O_4$	160.0735	0.2	1.67	HC11	39000
Butylsuccinic acid	AN	Butane	$C_8H_{14}O_4$	174.0892	0.3	2.51	HC11	11500
Pentylsuccinic acid	AN	Pentane	$C_9H_{16}O_4$	188.1045	0.3	4.12	HC11	42000
Hexylsuccinic acid	AN	Hexane	$C_{10}H_{18}O_4$	202.1205	0.4	5.98	HC11	15400
1-Phenylethyl-succinates or methyl benzylsuccinate	AN	Ethylbenzene or xylene	$C_{12}H_{14}O_4$	222.0892	0.0	4.6	HC11	1080000
Methylbenzylalcohol	A	Xylene	$C_8H_{10}O$	122.0732	−0.5	15.17	LC11	175000
Dimethylcatechol or isomers	A	Ethylbenzene or xylene	$C_8H_{10}O_2$	138.0681	1.1	5.99	LC11 LC3 HC3	9000 6600 7100

[1] *AN, anaerobic and A, aerobic degradation pathway.*

(Agrawal and Gieg, 2013; Rabus et al., 2016). Requisite succinates were more than an order of magnitude more abundant in HC11 than in any other sample. Although this points to anaerobic microbial hydrocarbon degradation in HC, the necessary genetic signatures (i.e., alkyl- and benzyl-succinate synthase genes), which are diagnostic of hydrocarbon degradation via addition to fumarate (Callaghan et al., 2010), were not observed in the metagenomes (**Table 4B**). This discrepancy might potentially be explained by several properties of the system. Anaerobic hydrocarbon degraders may be present in the HC sample, but their abundance may have been too low to be detectable in metagenomic data given the sequencing depth applied here. Alternatively, anaerobic degradation metabolites may reflect processes in the source formation but not the pipeline. It is also possible that HC bacterial communities contain succinate-forming metabolic pathways for which the genetic signatures are unknown. More likely, however, the data are consistent with a relatively minor contribution of anaerobic hydrocarbon degradation to the overall microbial metabolism in the HC samples.

Detectable aerobic degradation products included hydroxylated derivatives of xylene, ethylbenzene, naphthalene, and phenanthrene. These were limited to the LC system and their relative abundance was greater in the later (LC11) vs. the earlier (LC3) pig sample (**Table 5**), which is consistent with the observed distribution of aerobic hydrocarbon degradation genes (see above). Thus, hydrocarbon degradation may have preferentially occurred deeper into the wall deposits. In the outer layer of the deposits, substrate requirements may well be met by diffusion from the water phase of partially oxidized compounds such as volatile fatty acids, which are preferentially used for oxidation. The role that biofilms play in the degradation of hydrocarbons in pipeline systems is, however, still poorly understood. It is, for example, unclear how requisite deposits were structured. Outer layers of the biofilm may have been dominated by abiotic products and the patterns observed here may simply reflect the distribution of biomass.

Organonitrogen Metabolism

Denitrification, as a strictly inorganic process, is not directly targeted by conventional metabolomic analysis. Nitrate (NO_3^-; m/z 62) is detectable in the negative ion mode, but wasn't found in any of the samples. Quite apart from the inorganic compounds involved, the biochemical denitrification process is likely associated with unique metabolic signatures (Kampschreur et al., 2012). The related signals, however, must be characterized using laboratory cultures before they can be reliable identified in metabolic data of complex communities. Our ability to extract requisite signatures related to denitrification or sulfate reduction from the LC and HC metabolomes, respectively, are therefore limited. With high-resolution mass spectrometry, it is, however, possible to study both patterns and abundance of nitrogen-containing organic compounds without the need for identifications. Organic compounds that contain an odd number of nitrogen atoms are unusual in that they have an odd integral mass (McLafferty and Turecek, 1993). The relative abundance of odd molecular masses in the metabolome, RN (Equation 1: RN=

$$\frac{\sum Abundances\ of\ features\ derived\ from\ odd{-}mass\ molecules}{\sum Abundances\ of\ all\ features}),$$

might therefore be used as a measure of the importance of organonitrogen compounds, and presumably their metabolism, in the metabolome. While this assumption is hypothetical, a striking pattern is indeed observed: While RN values are similar in LC3 and HC3 at ≈40%, the value increases substantially with depth of deposits in the LC system to 55% in LC11, but decreases in the HC system to 30% in HC11. Even more strikingly, the RN value for those features that were given putative identifications was only 24% in HC11, while it was 62% in LC11. These observations are consistent with a more dominant role for the metabolism of nitrogen-containing organic compounds and the accumulation of such compounds in the LC deposit. What role, if any, denitrifying organisms such as *P. stutzeri* play in this remains unclear from our data. The observation may nonetheless present

future opportunities for the development of predictive capacity with respect to corrosion, by correlating metabolomic profiles to microbial community structure and their corrosivity.

Microbial Community Resilience

A notable feature of the HC and LC systems is the observation that dramatically different microbial community organizations are maintained over many years, even though both the physico-chemical drivers and chemical field treatments are very similar. While it is possible that there are significant, but overlooked, environmental differences between the two systems, it is also possible that a biological mechanism dominates. We note that at the time of sampling nitrate treatment of injection water was occurring in both systems, but while seawater injection in HC was initiated in 1999, nitrate treatment did not start until 2001. This could suggests that the HC system retains a corrosive community structure due to its inherent resilience to disturbance via requisite nitrate treatments. Similarly, the LC system might contain persistent denitrifying community assemblages that resist invasion by more corrosive SRBs. Ecological resilience theory indicates that communities can often assemble in many different stable states (Botton et al., 2006; Shade et al., 2012). Alternate community states often persist for no other reason than their historical context and resilience to disturbance. Whether nitrate treatment should be considered a severe, potentially community-transforming, "disturbance" for the established sulfate-reducing HC community is not at all obvious with the data at hand.

Survival of sulfate reduction metabolism in the HC system may seem surprising in view of the long-term nitrate treatment. For example, it has previously been reported that the addition of nitrate not only inhibits sulfate reduction but also leads to the reduction of sulfide already present in a biofilm (Halim et al., 2012). Similarly, nitrate reducing bacteria (NRB) appear to outcompete SRB for access to degradable oil organics (Hubert and Voordouw, 2007). Using a packed-bed bioreactor, the same study found that souring control in a reactor that received 12.5 mM lactate required the addition of 10 mM nitrate, irrespective of the sulfate concentration. Thus, if the (usable) oxidative capacity of organics, such as lactate or VFAs, is larger than the reductive capacity of nitrate added to the injection water (da Silva et al., 2014), most of the organics will remain in the water phase by the time the denitrification process is completed, creating ideal conditions for sulfate reduction. This is not inconsistent with observations from the LC and HC systems. Acetic acid concentrations were as high as 5.2 mM (**Table 1**), while nitrate additions were ca. 2.5 mM, potentially stoichiometrically insufficient to allow complete VFA depletion.

One may also consider that NRB are known to suppress SRB activity, not only by competition for electron donors, but also by producing nitrite, which inhibits sulfite reduction by binding to the dissimilatory sulfate reductase (DsrAB) enzyme (Haveman et al., 2004). In thin biofilms, nitrate and nitrite may penetrate by diffusion, offering a possible explanation for observations in experiments that demonstrate nitrate inhibition, even when nitrate levels would not allow the stoichiometric depletion of available carbon sources. In thicker, more stable biofilms, however, such as the ones frequently found in oil pipeline systems, requisite inhibition might not be observed, and the exact mechanism of SRB inhibition by NRB is likely to be more complicated. Other factors that might influence this interplay are the history of the system, the rates of transport, and the size and geometry of a system. Inoculation with nitrate reducing-sulfur oxidizing bacteria are effective in suppressing souring in experiments performed in serum bottles (Nemati et al., 2001) but might not be effective in packed-reactors or in thick pipeline deposits.

Corrosion Mechanisms

It is in the profound differences in metabolic activity and organization between the two microbial systems that the explanations for the differences in corrosion rates are likely rooted. Aspects of the metabolic conditions in the HC systems that would seem to be particularly relevant to aggressive corrosion include the accumulation of sulfides and/or H_2S, greater biomass, the presence of fermentation products, and the generation of a unique chemical environment.

Sulfides and H_2S are traditionally associated with MIC corrosion, yet 80 years after the cathodic depolarization theory was first articulated (von Wolzogen Kuhr and van der Vlugt, 1934) the corrosion mechanisms occurring in the field are still disputed (Beech and Sunner, 2007; Enning and Garrelfs, 2014). It is assumed that the presence of sulfides affects the stability of protective oxide layers. However, the sulfide layers that replace them are sometimes protective (Ning et al., 2015). Sulfide layers may also catalyze reduction reactions, detected as cathodic depolarization (King and Miller, 1977). The ratio of sulfide generation by reduction and iron ion generation by corrosion will determine the extent to which free H_2S or HS^-_{aq} is formed. The dominant role of sulfate reduction in the HC samples, and the fact that both samples had a very strong odor of H_2S, support that free H_2S is indeed formed in HC. Hydrogen sulfide is, independently of other sulfides, recognized as a corrosion agent in the oilfields (Zheng et al., 2015), and significant enhancements of both cathodic and anodic currents on (X65) pipeline steel can be observed in its presence (Zheng et al., 2014).

Lastly, anaerobic fermentation may have been an important contributor. Fermentation produces a range of low-molecular weight organic compounds, such as acetate, which was observed in high concentration in produced waters. Furthermore, the potential for fermentative processes is somewhat supported by the observation that two of the dominant genome bins were most similar to Firmicutes lineages, which are well known for their fermentative metabolism. The genomic content of the more abundant of these bins was most closely related to the genome of *Thermoanaerobacter pseudethanolicus*, which is known to ferment simple sugars, starch, and cellobiose (Onyenwoke et al., 2007). However, *T. pseudethanolicus* is also known to respire via the reduction of thiosulfate to H_2S, and it may have therefore been a contributor to sulfidogenesis. This distinction is likely dependent on the availability of suitable electron acceptors, but cannot be directly assessed via metagenomic analysis. The other dominant Firmicute genome bin is most similar to the

genome of *Alkaliphilus metalliredigens*, which is best known for its ability to reduce iron in the presence of organic substrates (Ye et al., 2004), rather than fermentative metabolism. Metagenomic data are therefore not necessarily supportive of the notion that fermentative metabolism was a driving force in the corrosion of HC.

CONCLUSION

A combined metagenomic and metabolomic approach was used to investigate significant differences in the severity of corrosion observed in two, otherwise very similar, production pipelines. Distinct microbial communities with dramatically different metabolic and genomic signatures were observed, with notable differences in the complexity and abundance of organonitrogen metabolites, the potential to utilize nitrate or sulfate as terminal electron acceptors, and the ability to degrade hydrocarbons. These results are not easily reconciled when it is assumed that nitrate addition is the primary driver (or disturbance) of community assembly. Put another way, it is typically expected that the addition of nitrate will create a push disturbance that, due to changes in redox potential, allows less-corrosive NRBs to outcompete SRBs. Our data indicate, however, that nitrate addition might sometimes result in an insufficient push, and that a lasting alternate stable community structure is not always achieved in the face of nitrate treatments. The problem of corrosion management in pipelines like HC may therefore be as much about assessing stability and creating the right kind of disturbance as it is about changing the redox potential. It is likely that localized, if not systemically occurring, biofilms will grow to sufficient thickness between pigging operations to allow the creation of nitrite-free locales in which SRB activity will have opportunities to persist. This would explain why nitrate addition has been reported to be effective in cultures as well as reservoirs, but does not always produce the desired results in pipeline systems. At this point, it therefore remains unclear how it is possible to effectively suppress SRB activity in production systems in which significant biofilms and abiotic deposits form. We suggest that corrosion management approaches and attempts to understand corrosion in complex engineered systems might benefit from a more refined understanding of the forces related to community resilience and the impact of disturbances such as pigging and nitrate addition.

AUTHOR CONTRIBUTIONS

VB conducted MS analyses and assisted with data analysis and writing. BW was responsible for molecular data QC, bioinformatics analysis, and python code. BW took the lead in writing the manuscript and data analysis. JS took the lead in metabolome data analysis and contributed to the writing of the manuscript. ES was involved in the bioinformatics analysis presented in the paper. EA assisted in MS analyses. KD was involved in sample processing, DNA extraction, qPCR analysis, and 16S rRNA gene sequencing as well as editing the manuscript. AC contributed to metagenome analysis for aerobic and anaerobic hydrocarbon degradation genes by providing reference databases for gene searches. AC also guided the analysis of metabolome data with respect to microbial hydrocarbon degradation by providing search targets for metabolomic surveys. AO (Duncan Lab) was involved in sample processing, DNA extraction, qPCR analysis, and 16S rRNA gene sequencing as well as editing the manuscript. TL was responsible for sampling in the field and metadata collection. IB initiated the project, assembled the team, and contributed to the ideas and writing.

FUNDING

This work was supported in part by grants from the University of Oklahoma Biocorrosion Center, the National Science Foundation (OCE 1634630 and MCB 1329890) and BP (The Gulf of Mexico Research Initiative, Project No. 130206). The instrumentation for the metabolomic analysis was funded by ONR through a DURIP grant (Award no. N000140910797), and the method development by ONR through a MURI grant (Award no. N000141010946).

ACCESSION NUMBERS

Metagenomic reads can be found in NCBI's Short Read Archive. BioProject: PRJNA356928; BioSamples: SAMN06130235 and SAMN06130234.

ACKNOWLEDGMENT

We wish to thank STATOIL for generously providing field samples.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.00099/full#supplementary-material

REFERENCES

Agrawal, A., and Gieg, L. M. (2013). In situ detection of anaerobic alkane metabolites in subsurface environments. *Front. Microbiol.* 4:140. doi: 10.3389/fmicb.2013.00140

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., and Nielsen, P. H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31, 533–541. doi: 10.1038/nbt.2579

Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995). Phylogenetic identification and *in situ* detection of individual microbial-cells without cultivation. *Microbiol. Rev.* 59, 143–169.

Annweiler, E., Materna, A., Safinowski, M., Kappler, A., Richnow, H. H., Michaelis, W., et al. (2000). Anaerobic degradation of 2-methylnaphthalene by

a sulfate-reducing enrichment culture. *Appl. Environ. Microbiol.* 66, 5329–5333. doi: 10.1128/AEM.66.12.5329-5333.2000

Atomi, H., Fukui, T., Kanai, T., Morikawa, M., and Imanaka, T. (2004). Description of *Thermococcus kodakaraensis* sp. nov., a well studied hyperthermophilic archaeon previously reported as *Pyrococcus* sp. KOD1. *Archaea* 1, 263–267. doi: 10.1155/2004/204953

Beech, I., and Sunner, J. (2007). "Sulphate-reducing bacteria and their role in corrosion of ferrous materials," in *Sulphate-Reducing Bacteria – Environmental and Engineered Systems*, eds W. A. Hamilton and L. L. Barton (Cambridge: Cambridge University Press).

Beech, I. B., and Gaylarde, C. C. (1999). Recent advances in the study of biocorrosion - an overview. *Rev. Microbiol.* 30, 177–190. doi: 10.1590/S0001-37141999000300001

Beech, I. B., Sunner, J. A., and Hiraoka, K. (2005). Microbe-surface interactions in biofouling and biocorrosion processes. *Int. Microbiol.* 8, 157–168.

Beech, I. B., Sztyler, M., Gaylarde, C. C., Smith, W. L., and Sunner, J. (2014). "Biofilms and Biocorrosion," in *Understanding Biocorrosion: Fundamentals and Applications*, eds T. Liengen, D. Féron, R. Basséguy, and I. B. Beech (Sawston: Woodhead Publishing). doi: 10.1533/9781782421252.1.33

Beech, W. B., and Sunner, J. (2004). Biocorrosion: towards understanding interactions between biofilms and metals. *Curr. Opin. Biotechnol.* 15, 181–186. doi: 10.1016/j.copbio.2004.05.001

Boisvert, S., Laviolette, F., and Corbeil, J. (2010). Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J. Comput. Biol.* 17, 1519–1533. doi: 10.1089/cmb.2009.0238

Bonifay, V., Smith, W., Beech, I., and Sunner, J. (2013). "Metabolomics study of corroding microbial systems," in *Proceedings of the 61st ASMS Conference on Mass Spectrometry and Allied Topics: American Society for Mass Spectrometry* (San Diego, CA: ASMS).

Botton, S., Van Heusden, M., Parsons, J. R., Smidt, H., and Van Straalen, N. (2006). Resilience of microbial systems towards disturbances. *Crit. Rev. Microbiol.* 32, 101–112. doi: 10.1080/10408410600709933

Buermans, H. P. J., and den Dunnen, J. T. (2014). Next generation sequencing technology: advances and applications. *Biochim. Biophys. Acta Mol. Basis Dis.* 1842, 1932–1941. doi: 10.1016/j.bbadis.2014.06.015

Callaghan, A. V., Davidova, I. A., Savage-Ashlock, K., Parisi, V. A., Gieg, L. M., Suflita, J. M., et al. (2010). Diversity of benzyl- and alkylsuccinate synthase genes in hydrocarbon-impacted environments and enrichment cultures. *Environ. Sci. Technol.* 44, 7287–7294. doi: 10.1021/es1002023

Callaghan, A. V., and Wawrik, B. (2016). *AnHyDeg: A Curated Database of Anaerobic Hydrocarbon Degradation Genes*. Available at: https://github.com/AnaerobesRock/AnHyDeg

Caporaso, J. G., Bittinger, K., Bushman, F. D., Desantis, T. Z., Andersen, G. L., and Knight, R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26, 266–267. doi: 10.1093/bioinformatics/btp636

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., et al. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 42, D459–D471. doi: 10.1093/nar/gkt1103

Creek, D. J., Jankevics, A., Breitling, R., Watson, D. G., Barrett, M. P., and Burgess, K. E. V. (2011). Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: improved metabolite identification by retention time prediction. *Anal. Chem.* 83, 8703–8710. doi: 10.1021/ac2021823

Creek, D. J., Jankevics, A., Burgess, K. E., Breitling, R., and Barrett, M. P. (2012). IDEOM: an Excel interface for analysis of LC-MS-based metabolomics data. *Bioinformatics* 28, 1048–1049. doi: 10.1093/bioinformatics/bts069

da Silva, M. L. B., Soares, H. M., Furigo, A., Schmidell, W., and Corseuil, H. X. (2014). Effects of nitrate injection on microbial enhanced oil recovery and oilfield reservoir souring. *Appl. Biochem. Biotechnol.* 174, 1810–1821. doi: 10.1007/s12010-014-1161-2

Dahle, H., Garshol, F., Madsen, M., and Birkeland, N. K. (2008). Microbial community structure analysis of produced water from a high-temperature North Sea oil-field. *Antonie Van Leeuwenhoek* 93, 37–49. doi: 10.1007/s10482-007-9177-z

Duncan, K. E., Gieg, L. M., Parisi, V. A., Tanner, R. S., Tringe, S. G., Bristow, J., et al. (2009). Biocorrosive thermophilic microbial communities in Alaskan North Slope oil facilities. *Environ. Sci. Technol.* 43, 7977–7984. doi: 10.1021/es9013932

Enning, D., and Garrelfs, J. (2014). Corrosion of iron by sulfate-reducing bacteria: new views of an old problem. *Appl. Environ. Microbiol.* 80, 1226–1236. doi: 10.1128/AEM.02848-13

Fahy, E., Subramaniam, S., Murphy, R. C., Nishijima, M., Raetz, C. R. H., Shimizu, T., et al. (2009). Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.* 50, S9–S14. doi: 10.1194/jlr.R800095-JLR200

Geesey, G. G., Beech, I., Bremmer, P. J., Webster, B. J., and Wells, D. (2000). "Biocorrosion," in *Biofilms II: Process Analysis and Applications*, ed. J. Bryers (Hoboken, NJ: Wiley-Liss), 281–326.

Gittel, A., Sorensen, K. B., Skovhus, T. L., Ingvorsen, K., and Schramm, A. (2009). Prokaryotic community structure and sulfate reducer activity in water from high-temperature oil reservoirs with and without nitrate treatment. *Appl. Environ. Microbiol.* 75, 7086–7096. doi: 10.1128/AEM.01123-09

Gu, T. (2012). New understandings of biocorrosion mechanisms and their classification. *Microb. Biochem. Technol.* 4, iii–vi. doi: 10.4172/1948-5948.1000e107

Halim, A., Watkin, E., and Gubner, R. (2012). Short term corrosion monitoring of carbon steel by bio-competitive exclusion of thermophilic sulphate reducing bacteria and nitrate reducing bacteria. *Electrochim. Acta* 77, 348–362. doi: 10.1016/j.electacta.2012.06.040

Hamilton, W. A. (1985). Sulfate-reducing bacteria and anaerobic corrosion. *Annu. Rev. Microbiol.* 39, 195–217. doi: 10.1146/annurev.mi.39.100185.001211

Haveman, S. A., Greene, E. A., Stilwell, C. P., Voordouw, J. K., and Voordouw, G. (2004). Physiological and gene expression analysis of inhibition of *Desulfovibrio vulgaris* Hildenborough by nitrite. *J. Bacteriol.* 186, 7944–7950. doi: 10.1128/JB.186.23.7944-7950.2004

Heitz, E. (1992). *A Working Party Report on a Microbiological Degradation of Materials and Methods of Protection*, eds E. Heitz, A. D. Mercer, W. Sand and A. K. Tiller. (London: The institute of Materials).

Hubert, C., and Voordouw, G. (2007). Oil field souring control by nitrate-reducing *Sulfurospirillum* spp. that outcompete sulfate-reducing bacteria for organic electron donors. *Appl. Environ. Microbiol.* 73, 2644–2652. doi: 10.1128/AEM.02332-06

Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386. doi: 10.1101/gr.5969107

Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Hyatt, D., Locascio, P. F., Hauser, L. J., and Uberbacher, E. C. (2012). Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* 28, 2223–2230. doi: 10.1093/bioinformatics/bts429

Kampschreur, M. J., Kleerebezem, R., Picioreanu, C., Bakken, L., Bergaust, L., De Vries, S., et al. (2012). Metabolic modeling of denitrification in *Agrobacterium tumefaciens*: a tool to study inhibiting and activating compounds for the denitrification pathway. *Front. Microbiol.* 3:370. doi: 10.3389/fmicb.2012.00370

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205. doi: 10.1093/nar/gkt1076

King, R. A., and Miller, J. D. A. (1977). Corrosion of ferrous-metal by bacterially produced iron sulphides and its control by cathodic protection. *Anti Corrosion Methods Mater.* 24, 9–11. doi: 10.1108/eb007036

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41:e1. doi: 10.1093/nar/gks808

Koch, G. H., Brongers, M. P. H., Thompson, N. G., Virmani, Y. P., and Payer, J. H. (2002). *Corrosion Cost and Preventive Strategies in the United States*. Alexandria, VA: National Technical Information Service.

Lalucat, J., Bennasar, A., Bosch, R., Garcia-Valdes, E., and Palleroni, N. J. (2006). Biology of *Pseudomonas stutzeri*. *Microbiol. Mol. Biol. Rev.* 70, 510–547. doi: 10.1128/MMBR.00047-05

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–U354. doi: 10.1038/nmeth.1923

Leader, D. P., Burgess, K., Creek, D., and Barrett, M. P. (2011). Pathos: a web facility that uses metabolic maps to display experimental changes in metabolites identified by mass spectrometry. *Rapid Commun. Mass Spectrom.* 25, 3422–3426. doi: 10.1002/rcm.5245

Lee, W., Lewandowski, Z., Nielsen, P. H., and Hamilton, W. A. (1995). Role of sulfate-reducing bacteria in corrosion of mild-steel - a review. *Biofouling* 8, 165–194. doi: 10.1080/08927019509378271

Lenhart, T. R., Duncan, K. E., Beech, I. B., Sunner, J. A., Smith, W., Bonifay, V., et al. (2014). Identification and characterization of microbial biofilm communities associated with corroded oil pipeline surfaces. *Biofouling* 30, 823–835. doi: 10.1080/08927014.2014.931379

Little, B. J., and Lee, J. S. (2007). "Causative organisms and possible mechanisms," in *Microbiologically Influenced Corrosion*, eds B. J. Little and J. S. Lee (Hoboken, NJ: Wiley & Sons, Inc). doi: 10.1002/047011245X

Little, B. J., and Ray, R. (2002). A perspective on corrosion inhibition by biofilms. *Corrosion* 58, 424–428. doi: 10.5006/1.3277632

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J.* 17, 10–12. doi: 10.14806/ej.17.1.200

McLafferty, F. W., and Turecek, F. (1993). *Interpretation of Mass Spectra*. Mill Valley, CA: University Science Books.

Miller, C. S. (2013). Assembling full-length rRNA genes from short-read metagenomic sequence datasets using EMIRGE. *Methods Enzymol.* 531, 333–352. doi: 10.1016/B978-0-12-407863-5.00017-4

Monteiro, M. S., Carvalho, M., Bastos, M. L., and Guedes De Pinho, P. (2013). Metabolomics analysis for biomarker discovery: advances and challenges. *Curr. Med. Chem* 20, 257–271. doi: 10.2174/092986713804806621

Nemati, M., Jenneman, G. E., and Voordouw, G. (2001). Mechanistic study of microbial control of hydrogen sulfide production in oil reservoirs. *Biotechnol. Bioeng.* 74, 424–434. doi: 10.1002/bit.1133

Ning, J., Zheng, Y. G., Brown, B., Young, D., and Nesic, S. (2015). A thermodynamic model for the prediction of mild steel corrosion products in an aqueous hydrogen sulfide environment. *Corrosion* 71, 945–960. doi: 10.5006/1566

Onyenwoke, R. U., Kevbrin, V. V., Lysenko, A. M., and Wiegel, J. (2007). *Thermoanaerobacter pseudethanolicus* sp. nov., a thermophilic heterotrophic anaerobe from Yellowstone National Park. *Int. J. Syst. Evol. Microbiol.* 57, 2191–2193. doi: 10.1099/ijs.0.65051-0

Ornek, D., Wood, T. K., Hsu, C. H., Sun, Z., and Mansfeld, F. (2002). Pitting corrosion control of aluminum 2024 using protective biofilms that secrete corrosion inhibitors. *Corrosion* 58, 761–767. doi: 10.5006/1.3277659

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi: 10.1093/nar/gkt1226

Patti, G. J., Yanes, O., and Siuzdak, G. (2012). Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* 13, 263–269. doi: 10.1038/nrm3314

Rabus, R., Boll, M., Heider, J., Meckenstock, R. U., Buckel, W., Einsle, O., et al. (2016). Anaerobic microbial degradation of hydrocarbons: from enzymatic reactions to the environment. *J. Mol. Microbiol. Biotechnol.* 26, 5–28. doi: 10.1159/000443997

Salek, R. M., Steinbeck, C., Viant, M. R., Goodacre, R., and Dunn, W. B. (2013). The role of reporting standards for metabolite annotation and identification in metabolomic studies. *Gigascience* 2:13. doi: 10.1186/2047-217X-2-13

Scheltema, R. A., Jankevics, A., Jansen, R. C., Swertz, M. A., and Breitling, R. (2011). PeakML/mzMatch: a file format, Java library, R library, and tool-chain for mass spectrometry data analysis. *Anal. Chem.* 83, 2786–2793. doi: 10.1021/ac2000994

Shade, A., Peter, H., Allison, S. D., Baho, D. L., Berga, M., Burgmann, H., et al. (2012). Fundamentals of microbial community resistance and resilience. *Front. Microbiol.* 3:417. doi: 10.3389/fmicb.2012.00417

Smith, C. A., Want, E. J., O'maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* 78, 779–787. doi: 10.1021/ac051437y

Sorokin, D. Y., Tourova, T. P., Henstra, A. M., Stams, A. J. M., Galinski, E. A., and Muyzer, G. (2008). Sulfidogenesis under extremely haloalkaline conditions by *Desulfonatronospira thiodismutans* gen. nov., sp nov., and *Desulfonatronospira delicata* sp nov - a novel lineage of Deltaproteobacteria from hypersaline soda lakes. *Microbiology* 154, 1444–1453. doi: 10.1099/mic.0.2007/015628-0

Stevenson, B. S., Drilling, H. S., Lawson, P. A., Duncan, K. E., Parisi, V. A., and Suflita, J. M. (2011). Microbial communities in bulk fluids and biofilms of an oil facility have similar composition but different structure. *Environ. Microbiol.* 13, 1078–1090. doi: 10.1111/j.1462-2920.2010.02413.x

Suflita, J. M., Aktas, D. F., Oldham, A. L., Perez-Ibarra, B. M., and Duncan, K. (2012). Molecular tools to track bacteria responsible for fuel deterioration and microbiologically influenced corrosion. *Biofouling* 28, 1003–1010. doi: 10.1080/08927014.2012.723695

Suflita, J. M., Sunner, J., Bonifay, V., Aydin, E., and Aktas, D. F. (2016). "Metabolic profiling and metabolomic procedures for investigating the biodegradation of hydrocarbons," in *Hydrocarbon and Lipid Microbiology Protocols: Genetic, Genomic and System Analyses of Communities*, eds T. J. McGenity, K. N. Timmis, and B. Nogales (Heidelberg: Springer-Verlag).

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., et al. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics* 3, 211–221. doi: 10.1007/s11306-007-0082-2

Sun, S. L., Chen, J., Li, W. Z., Altintas, I., Lin, A., Peltier, S., et al. (2011). Community cyberinfrastructure for advanced microbial ecology research and analysis: the CAMERA resource. *Nucleic Acids Res.* 39, D546–D551. doi: 10.1093/nar/gkq1102

Sunner, J., and Beech, I. B. (2015). "Mass Spectrometry-based Metabolomics – a tool to guide MIC Management in the Field," in *ISMOS-5 Abstract Book*, eds C. Boccadoro, D. Durand, C. Whitby, and T. L. Skovhus (Stavanger: ISMOS).

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Tautenhahn, R., Bottcher, C., and Neumann, S. (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* 9:504. doi: 10.1186/1471-2105-9-504

von Wolzogen Kuhr, C. A. H., and van der Vlugt, S. (1934). The graphitization of cast iron as an electrobiochemical process in anaerobic soil. *Water Hague* 18, 147–165.

Wawrik, B., Mendivelso, M., Parisi, V. A., Suflita, J. M., Davidova, I. A., Marks, C. R., et al. (2012). Field and laboratory studies on the bioconversion of coal to methane in the San Juan Basin. *FEMS Microbiol. Ecol.* 81, 26–42. doi: 10.1111/j.1574-6941.2011.01272.x

Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y. F., et al. (2013). HMDB 3.0-the human metabolome database in 2013. *Nucleic Acids Res.* 41, D801–D807. doi: 10.1093/nar/gks1065

Ye, Q., Roh, Y., Carroll, S. L., Blair, B., Zhou, J., Zhang, C. L., et al. (2004). Alkaline anaerobic respiration: isolation and characterization of a novel alkaliphilic and metal-reducing bacterium. *Appl. Environ. Microbiol.* 70, 5595–5602. doi: 10.1128/AEM.70.9.5595-5602.2004

Zheng, Y. G., Brown, B., and Nesic, S. (2014). Electrochemical study and modeling of H_2S corrosion of mild steel. *Corrosion* 70, 351–365. doi: 10.5006/0937

Zheng, Y. G., Ning, J., Brown, B., and Nesic, S. (2015). Electrochemical model of mild steel corrosion in a mixed H_2S/CO_2 aqueous environment in the absence of protective corrosion product layers. *Corrosion* 71, 316–325. doi: 10.5006/1287

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Bonifay, Wawrik, Sunner, Snodgrass, Aydin, Duncan, Callaghan, Oldham, Liengen and Beech. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.