

2022-05-11

On the Origin of Thermophilic Endospores

Gittins, Daniel Anthony

Gittins, D. A. (2022). On the Origin of Thermophilic Endospores (Doctoral thesis, University of Calgary, Calgary, Canada). Retrieved from <https://prism.ucalgary.ca>.

<http://hdl.handle.net/1880/114666>

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

On the Origin of Thermophilic Endospores

by

Daniel Gittins

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOLOGICAL SCIENCES

CALGARY, ALBERTA

MAY, 2022

© Daniel Gittins 2022

Abstract

Marine sediment covers ~70% of Earth's surface. Diverse microbial populations inhabit the subseafloor biosphere substantially contributing to global biomass. Patterns demonstrating microbial biogeography are well established in many environments, but conspicuous examples of thermophilic endospores in permanently cold seabed sediments highlight how the ecological processes controlling biogeography in the deep biosphere remain poorly understood.

This thesis examines environmental selection in subsurface petroleum reservoirs and assesses large scale microbial dispersal associated with these habitats to better understand the factors shaping microbial communities. The diversity in the global petroleum reservoir microbiome was assessed by examining 16S rRNA gene amplicon and shotgun metagenomic libraries from oil reservoirs around the world. Taxonomic composition varies among reservoirs with different physicochemical characteristics, and by geographic location, yet gene composition analysis highlights a common functional core. Shared functions include diverse capabilities for carbon acquisition and energy conservation consistent with metabolisms characteristic of the deep biosphere. Taxonomic variation with functional redundancy demonstrates environmental selection acting in these subsurface biogeochemical hotspots. *Firmicutes* are among the most prevalent taxa, while genetic potential for sporulation is widespread. Correlation of geophysical and geochemical evidence of hydrocarbon seepage with biogeographic patterns in the seabed distributions of endospores of thermophilic bacteria reveals geofluid-facilitated cell migration pathways connecting petroleum reservoirs with the surface. Genomic sequencing in high temperature incubations highlight adaptations of these microorganisms to life in anoxic petroleum systems, while phylogenetic comparisons reveal close resemblance to oil reservoir microbiomes globally. Microbial activity and selection of diverse populations of *Firmicutes* in crude oil-amended incubations further validates the origin of these thermophilies. Upon transport out of the subsurface, viable endospores re-enter the geosphere by sediment burial, enabling germination and environmental selection at depth where new petroleum systems establish. This newly termed 'microbial dispersal loop' circulates living biomass in and out of the deep biosphere.

Microbial ecology and Earth system processes are tightly linked. Deep biosphere populations possess extensive physiological and metabolic diversity and influence biogeochemical cycling on a global scale. By connecting geology and geological frameworks with factors that influence survival and evolution, the geosphere can be a model system to better understand microbial ecology.

Preface

Chapter 1 and Chapter 5 are original, unpublished, independent work by the author, Daniel Gittins, with feedback from Dr. Casey Hubert.

Chapter 2 is currently under review at *Environmental Microbiology* as Gittins DA, Bhatnagar S, Hubert CRJH. “Environmental selection in subsurface petroleum reservoirs from around the world”. DAG and SB compiled and processed the sequence data. DAG interpreted the data and wrote the manuscript with feedback from CRJH and SB during refinement and finalization.

Chapter 3 is currently under review at *Science Advances* as Gittins DA, Desiage P-A, Morrison N, Rattray JE, Bhatnagar S, Chakraborty A, Zorz J, Li C, Horanszky O, Cramm MA, Bisiach F, Webb J, MacDonald A, Fowler M, Campbell DC, Hubert CRJ. “Geological processes mediate a microbial dispersal loop in the deep biosphere”. P-AD, NM, DCC and AM acquired, processed, and analyzed the geophysical data. DAG, AC, CL, MAC, JW and CRJH collected marine sediment samples for testing. MF, JW and DAG generated and analyzed the hydrocarbon geochemical data. DAG, AC, CL, OH, MAC, FB and CRJH generated and analyzed the microbial community data. DAG and SB curated and analyzed the reservoir microbiome data. DAG and JZ processed and analyzed the metagenomic data. JR measured and interpreted dipicolinic acid signals. DAG and CRJH wrote the manuscript with feedback from all authors during refinement and finalization. The crew of CCGS *Hudson* and Natural Resources Canada collected sediment cores and assisted with onboard core processing. Dr. Steve Larter, Dr. Lisa Gieg, Dr. Bo Barker Jørgensen, Dr. Will Richardson and Dr. Rhonda Clark offered valuable discussions and provided research support.

Chapter 4, “Cold seabed thermophilic endospores germinate and grow in the presence of high concentrations of crude oil”, is original, unpublished work based on a collaboration with Po Sze (Cynthia) Kwan. DG and PSK designed and conducted the experiments. DAG synthesized and interpreted the data. DAG wrote the chapter with feedback from Dr. Casey Hubert. Jody Sandel and the crew of *R/V GeoExplorer* collected and processed piston cores from the Eastern Gulf of Mexico. Dr. James Brooks and Dr. Bernie Bernard provided geochemical data for these cores. The crew of CCGS *Hudson* and Natural Resources Canada collected and helped process sediment cores

from offshore Nova Scotia. Dr. Martin Fowler and Jamie Webb provided hydrocarbon geochemical data for these cores. Dr. Emmanuel Alexis Rodrigues Duran provided the Mizzen crude oil sample and associated metadata.

Additional research contributed during my PhD, but not included in this thesis:

Khot V*, Zorz J*, Gittins D, Chakraborty A, Bell E, Bautista MA, Paquette AJ, Hawley AK, Novotnik B, Hubert CRJ, Strous M, Bhatnagar S. (2022) CANT-HYD: A curated database of phylogeny-derived Hidden Markov Models for annotation of marker genes involved in hydrocarbon degradation. *Front. Microbiol.* **12**, 764058. *authors contributed equally to work.

Chakraborty A, Ellefson E, Li C, Gittins D, Brooks JM, Bernard BB, Hubert CRJ. (2018) Thermophilic endospores associated with migrated thermogenic hydrocarbons in deep Gulf of Mexico marine sediments. *ISME J.* **12**, 1895–1906.

Acknowledgements

Many people have supported me during my PhD, and I consider myself fortunate to know each of you. First and foremost, I would like to thank my supervisor, Dr. Casey Hubert, for your constant support, guidance, and vision that has provided me with the opportunity to develop both personally and professionally. Thank you also to my supervisory committee members Dr. Lisa Gieg and Dr. Steve Larter, and my external examiners Dr. Steven Bryant and Dr. Jennifer Biddle.

This research would not have been possible without funding from Genome Canada, Mitacs, and the Nova Scotia Department of Energy and Mines. I would like to wholeheartedly thank each of these sources for providing me with financial security during my PhD. I would also like to thank the Nova Scotia Department of Energy and Mines, the Nova Scotia Offshore Energy Research Association, and Natural Resources Canada for providing ship time funding and support that allowed me to take part in numerous incredible offshore expeditions.

I consider myself fortunate to be a member of the Microbial Markets and Geomicrobiology (MMG) working group, co-led by Dr. Casey Hubert and Dr. Marc Strous. I would like to thank Casey, Marc, and every member of our group past and present for creating such an enjoyable and inspiring environment to work in. Special thanks to Carmen Li for your guidance with molecular work, to Dr. Srijak Bhatnagar for teaching me what bioinformatics is and how to use it, and Dr. Rhonda Clark for supporting us in the lab and the field. A huge thank you also to Dr. Anirban Chakraborty whose patience, supervision, and support has undoubtedly made me a better scientist.

Lastly, I want to appreciate my friends and family who encouraged and supported me throughout my PhD journey. To my parents, thank you for your endless support of my ambitions and endeavors, even though these have led me to live on the other side of the world from you. To my partner, Jackie Zorz, your love and support means the world to me.

Table of Contents

Abstract	i
Preface	iii
Acknowledgements	v
List of Tables	ix
List of Figures	xii
<i>Chapter 1 - Introduction</i>	1
1.1 Biogeography of microorganisms	1
1.2 The microbial deep biosphere	2
1.3 Petroleum reservoir ecosystems	6
1.4 Microbial survival strategies	8
1.5 Endospore-forming bacteria	10
1.6 Endospores as model organisms for microbial dispersal	11
1.7 Thesis overview.....	13
<i>Chapter 2 - Environmental selection in subsurface petroleum reservoirs from around the world</i>	15
2.1 Abstract	16
2.2 Introduction	17
2.3 Results	19
2.3.1 Overview of the petroleum reservoir microbiome	19
2.3.2 Factors associated with the community structure within oil reservoirs.....	22
2.3.3 Functional potential of the oil reservoir microbiome	24
2.3.4. Hydrocarbon biodegradation potential	28
2.4 Discussion	29
2.4.1 Petroleum reservoirs lack a core microbiome	29
2.4.2 Secondary oil recovery imposes a selective pressure on the oil reservoir microbiome	31
2.4.3 Is hydrocarbon degradation universal in petroleum reservoirs?.....	32
2.4.4 Provenance of the oil reservoir microbiome.....	34

2.5 Methods.....	34
2.5.1 Data acquisition	34
2.5.2 16S rRNA gene sequence processing.....	35
2.5.3 Metagenomic sequence processing	37
2.5.4 Statistical analysis and data visualization.....	38
<i>Chapter 3 - Geological processes mediate a microbial dispersal loop in the deep biosphere ...</i>	<i>40</i>
3.1 Abstract	41
3.2 Introduction.....	42
3.3 Results.....	43
3.4 Discussion	50
3.5 Methods.....	53
3.5.1 Subsurface imaging via seismic data acquisition and processing	53
3.5.2 Marine sediment sampling.....	54
3.5.3 Hydrocarbon geochemical analysis	55
3.5.4 Sediment incubation at elevated temperatures	56
3.5.5 16S rRNA gene amplicon sequencing.....	56
3.5.6 16S rRNA gene amplicon sequence processing	57
3.5.7 Metagenome sequencing	58
3.5.8 Metagenome sequence processing.....	58
3.5.9 Analysis of global oil reservoir microbiome sequences	59
3.5.10 Statistical analysis and data visualization.....	60
3.5.11 Phylogenetic analysis	61
3.5.12 Dipicolinic acid (DPA) measurement.....	61
<i>Chapter 4 - Thermophilic endospores in the cold seabed germinate and become active in the presence of high concentrations of crude oil.....</i>	<i>74</i>
4.1 Abstract	75
4.2 Introduction.....	76
4.2 Results.....	78
4.2.1 Migrated thermogenic hydrocarbons in surface sediments indicate seepage.....	78
4.2.2 Microbial diversity and community structure	81
4.2.3 Prevalence of seep associated taxa	86

4.2.4 Identification of seep associated species in different locations	87
4.3 Discussion	90
4.4 Methods.....	93
4.4.1 Sediment and crude oil sample descriptions.....	93
4.4.2 Hydrocarbon geochemical analysis	94
4.4.3 Crude oil-amended sediment incubations.....	96
4.4.4 Sulfate and organic acid measurements.....	97
4.4.5 16S rRNA gene amplicon sequencing.....	97
4.4.6 Sequence processing.....	98
4.4.7 Comparative sequence analysis	99
4.4.8 Statistical analysis.....	99
 <i>Chapter 5 - Conclusion</i>	 102
5.1 Summary	102
5.2 Outlook.....	105
 References	 107
 Appendix	 128

List of Tables

Table S2.1. Sample metadata of low throughput clone libraries of 16S rRNA genes, high throughput 16S rRNA gene and metagenome libraries.

Table S2.2. Sequence totals, number of observed phylotypes, and alpha diversity estimates of high throughput 16S rRNA gene libraries.

Table S2.3. Taxonomic classifications of low throughput clone libraries of 16S rRNA genes, high throughput 16S rRNA gene and metagenome libraries.

Table S2.4. Correlation of environmental and technical factors with alpha diversity indices for high throughput 16S rRNA gene libraries, high throughput 16S rRNA gene-based microbial community composition, and KEGG Ortholog assignments.

Table S2.5. *IndicSpecies* Stat values of significant associations ($P < 0.05$) of phylotypes to different reservoir temperature intervals.

Table S2.6. Metabolic genes and pathways identified in metagenome libraries based on KEGG Ortholog assignments. Extended list of samples KEGG Ortholog (KO) assignments annotated using GhostKOALA. Select sporulation genes detected in metagenome libraries based on KEGG Ortholog assignments. Hydrocarbon activation genes detected in metagenome libraries.

Table S2.7. Sequence counts of low throughput clone libraries of 16S rRNA genes, high throughput 16S rRNA gene and metagenome libraries.

Table S3.1. Marine sediment core site locations and sample descriptions.

Table S3.2. Hydrocarbon geochemical measurements on marine sediment cores.

Table S3.3. ASV table with taxonomic classification of 16S rRNA gene amplicon libraries of 14 core sediments, in triplicate, before incubation (0 d; post-pasteurization) and after 28 (28 d) or 56 (56 d) days of incubation at 40, 50, or 60°C, subsampled to 4,635 sequences.

Table S3.4. Analysis of similarity (ANOSIM) test results indicating microbial community composition differences between subsets of the 16S rRNA gene amplicon libraries for groups of core sediments.

Table S3.5. *IndicSpecies* Stat values of significant associations ($P < 0.05$) of *Firmicutes* ASVs to sediments with geochemical evidence of thermogenic hydrocarbons ($n=2$) when tested against

sediments without geochemical evidence of thermogenic hydrocarbons ($n=8$) after combined 28 and 56 days of incubation.

Table S3.6. Taxonomic assignment (Silva version 138) of 10,860,283 curated 16S rRNA sequences from 59 oil reservoir sequencing surveys.

Table S3.7. Metagenome-assembled genomes (MAGs). MAG to amplicon sequence variant (ASV) correlation.

Table S3.8. Summary of genes detected in metagenome-assembled genomes (MAGs).

Table S3.9. Sporulation genes annotated against Pfam, TIGRFAM and Swiss-Prot databases.

Table 4.1. Composition of the experimental treatments.....78

Table S4.1. Marine sediment core site and crude oil well locations and sample descriptions.

Table S4.2. Hydrocarbon geochemical measurements on EGM and NS marine sediment core.

Table S4.3. Sequence totals, number of observed ASVs and alpha diversity estimates of 16S rRNA gene libraries randomly subsampled to 1000 reads.

Table S4.4. Concentrations of sulfate and six organic acids before and after 14 days of incubation at 50°C under the respective experimental conditions. ASVs statistically associated with experimental conditions resulting in high levels of sulfate reduction when compared to conditions with low or no sulfate reduction.

Table S4.5. ASV table of sequence counts with taxonomic classification of 16S rRNA gene amplicon libraries.

Table S4.6. Correlation of microbial community composition before incubation (i.e., post-pasteurization) and after 14 days of incubation. Sequence count data converted to relative abundance prior to statistical tests to account for variable library sizes.

Table S4.7. ASV table with taxonomic classification of 16S rRNA gene amplicon libraries of 111 Eastern Gulf of Mexico core sediments after 14 days of incubation at 50°C, subsampled to 9,434 sequences (*experimental conditions detailed in Chakraborty *et al.* 2018). ASV table with taxonomic classification of 16S rRNA gene amplicon libraries of 14 offshore Nova Scotia core sediments, in triplicate, after 28 (28 d) or 56 (56 d) days of incubation at 40, 50, or 60°C, subsampled to 4,635 sequences (*table reproduced from *Chapter 3*).

Table S4.8. Seep ‘indicator species’ identified following high temperature incubations of sediments from offshore Nova Scotia and the Eastern Gulf of Mexico.

Table S4.9. Abundances of ASVs from high temperature incubations of Nova Scotia sediments that share $\geq 95\%$ sequence identity (BLAST) to seep indicator species identified from sediments in the Eastern Gulf of Mexico. Abundances of ASVs from high temperature incubations of Gulf of Mexico sediments that share $\geq 95\%$ sequence identity (BLAST) to seep indicator species identified from sediments offshore Nova Scotia.

List of Figures

Figure 2.1. Petroleum reservoir microbiome. Average relative sequence abundance of the ten most abundant phyla and all classes >1% relative sequence abundance in 295 high throughput 16S rRNA gene amplicon sequencing libraries, in addition to their corresponding abundance in 25 metagenome libraries. These 10 phyla represent 93% and 78% of the 16S and metagenomes libraries, respectively, and classes 90% and 70%, respectively. Assessing whether or not petroleum reservoirs around the world harbour a core microbiome revealed that the bacterial phyla *Proteobacteria* and *Firmicutes* and the archaeal phylum *Halobacterota* were observed across >75% of the 295 samples analysed by high throughput 16S rRNA gene sequencing. Accordingly, the classes *Gammaproteobacteria*, *Alphaproteobacteria*, *Methanosarcinia*, *Methanomicrobia*, *Clostridia* and *Bacilli* were identified in 82–60% of the samples. In metagenomic datasets, *Firmicutes* and *Proteobacteria*, along with *Euryarchaeota*, were observed in all samples, and at the genus level *Thermovirga*, *Halanaerobium* and *Petrotoga* were observed in more than 75% of the samples. Considering oil reservoir samples associated with primary recovery ($n = 98$), all high throughput 16S rRNA libraries showed Archaea, while *Firmicutes* and *Proteobacteria* were detected in 97 out of 98. Comparative analysis of reservoirs produced by secondary recovery methods showed *Proteobacteria* and *Halobacterota* were the most prevalent phyla, detected in 75 and 69% of the 198 libraries, respectively.....21

Figure 2.2. Relationship between microbial community composition and environmental factors. High throughput 16S rRNA gene library comparisons were performed based on the use of archaea-specific primers (**a**, **b**) or universal primers (**c**, **d**). **a**, **c**, Bray-Curtis dissimilarity in microbial community composition between samples from reservoirs at different depths and *in situ* temperatures. Libraries prepared using archaeal primers (**a**) exhibit greater dissimilarity between samples under the tested environmental conditions (Mantel, temperature $r = 0.55$, depth $r = 0.55$) than libraries prepared with universal primers (**c**; Mantel, temperature $r = 0.33$, depth $r = 0.43$), but both represent significant dissimilarity ($P < 0.05$). **b**, **d** Correlation between the geographic location of a petroleum reservoir and Bray-Curtis dissimilarity in the observed microbiome assessed in amplicon libraries using archaeal (**b**) and universal (**d**) primers. Visual trends as well as statistical testing (Mantel, $P < 0.05$) indicate that geographically distant reservoirs have more dissimilar microbial communities.....23

Figure 2.3. Gene composition of samples from different reservoirs. Nonmetric multidimensional scaling of the Bray-Curtis dissimilarity in the relative abundance of unique gene compositions represented as Kegg Ortholog (KO) assignments in samples from different reservoirs (**Table S2.6**). Unlike taxonomy-based associations, no significant correlation was observed between environmental factors (e.g., temperature and depth) and gene composition of the samples (**Table S2.4**). Ellipses indicate one standard deviation to highlight co-occurrence trends.....25

Figure 2.4. Metabolic potential of the petroleum reservoir microbiome. Occurrence of metabolic genes and pathways in metagenomes from microbiomes derived from primary and secondary produced reservoirs. Complete gene lists can be found in *Table S2.6*.....27

Figure 3.1. Deep subsurface to surface geofluid migration. **a**, Seafloor surface map derived from autonomous underwater vehicle (AUV) multibeam bathymetric sonar data. **b**, 3D seismic cross section showing a buried salt diapir, the location and direction of crestal faults (white dashed lines), including an interval with direct hydrocarbon indicators (DHI). **c**, Combined mosaic of side-scan sonar data and shaded relief bathymetry of the area surrounding a seep structure, indicating a pockmark feature as well as a small mound morphology. High backscatter intensity, related to distinctive properties of near-surface sediment, is shown in light-yellowish tones. **d**, AUV-based sub-bottom profiling showing localized acoustic blanking under the seep structure, indicative of upward migration of fluid originating deeper in the sediment.....45

Figure 3.2. Hydrocarbon geochemistry and microbial community variance between seabed sampling sites. **a**, Gas ($\sum C_2-nC_4$, gas wetness, and $C_1 \delta^{13}C$), liquid hydrocarbon extract (nC_{17}/nC_{27} and odd-over-even predominance) and biomarker (% 27 d β S and $C_{30} \alpha\beta/C_{31} \alpha\beta$ 22R) measurements to assess the presence of thermogenic hydrocarbons. Each parameter is scaled between 0 and 1 as shown in the heatmap. Cores are represented by maximum values in instances where multiple depths from a core were tested (see *Table S3.2*). Hierarchical clustering of the Euclidean distance between scaled values highlights groups of sites where the evidence for the presence of thermogenic hydrocarbons is strong (red), inconclusive (green) or not detected (blue). **b**, The same three groupings are reflected in Bray-Curtis dissimilarity in microbial community composition after sediment incubation at 50°C (ellipses indicate standard deviations of weighted averaged means of within-group distances for each of the three groups). Results from 40 and 60°C incubations showed similar clustering (see *Figure S3.2*). This demonstrates that sites with strong thermogenic hydrocarbon signals have distinct microbial populations after high temperature incubation (see also *Table S3.4*), relative to the sites without thermogenic hydrocarbon signals.....46

Figure 3.3. Subsurface oil reservoir origins of seep-associated thermophiles. **a**, Correlation of thermophilic spore-forming bacterial amplicon sequence variants (ASVs) with thermogenic hydrocarbons (*IndicSpecies*, $P < 0.05$). Highest ranking ASVs from each of 15 different genera are shown (representing 42 hydrocarbon-correlated ASVs in total). **b**, Prevalence of these genera in 59 oil reservoir microbiome assessments (11 million 16S rRNA gene sequences in total). **c**, Maximum likelihood phylogeny showing the 15 representative hydrocarbon-correlated ASVs and close relatives (see *Figure S3.4* for all 42 indicator ASVs). Black circles at the branch nodes indicate >80% bootstrap support (1,000 re-samplings), and the scale bar indicates 10% sequence divergence as inferred from PhyML. **d**, Metagenome-assembled genomes (MAGs) matching ASVs of interest (corresponding brown bands in **a–c**) were assessed for anaerobic alkane degradation, sporulation, and other metabolic features characteristic of deep biosphere populations living in oil reservoirs. Metagenome bin numbers

are indicated below the corresponding taxonomy (see **Table S3.8** for pathway definitions and additional gene annotations). Asterisk (*) denote bins with exact 16S rRNA gene sequence matching to ASVs.....48

Figure 3.4. Subsurface microbial dispersal loop mediated by seepage, sedimentation, dormancy and environmental selection. **a**, Endospore-forming microbial populations that are active and abundant in deep petroleum systems get dispersed upwards as dormant spores by hydrocarbon seepage along geological conduits. Endospores entering the deep-sea are dispersed laterally by bottom water currents as part of the rare biosphere. **b**, Endospores get deposited on the seabed and undergo burial. **c**, Dipicolinic acid was quantified in the upper metres of the seabed in two Scotian Slope sediment cores where hydrocarbons were not detected. Based on the average sedimentation rate in this region, this corresponds to a burial rate on the order of 10^8 spores $m^{-2} y^{-1}$. The dashed regression line reflects the global average estimated for endospores in the marine subseafloor biosphere (Wörmer *et al.* 2019). Survival of individual endospores over long burial time-scales (Fang *et al.* 2017; de Rezende *et al.* 2013; Morono *et al.* 2020) enables environmental selection (i.e., germination, activity and population replenishment) by suitable substrates and favorable conditions that exist in petroleum-bearing sediments such as oil reservoirs (Magoon and Dow, 1994). The sequence shown in **a** and **b** completes a subsurface ‘microbial dispersal loop’ that incorporates cell transport and biogeochemical cycling in Earth’s deep biosphere.....51

Figure S3.1. Deep sea study sites in the NW Atlantic Ocean. Sediment coring locations on the Scotian Slope. The inset shows the extent of the 20,000 km² study area, off the east coast of Nova Scotia, Canada. Bathymetric map from the General Bathymetric Chart of the Oceans (GEBCO, www.gebco.net) and National Oceanic and Atmospheric Administration (NOAA, www.ngdc.noaa.gov/)63

Figure S3.2. Microbial community variance between core sites. Non-metric multidimensional scaling of the Bray-Curtis dissimilarity of the microbial community composition after sediment incubation at **(a)** 40°C and **(b)** 60°C (for 50°C incubations, see **Figure 3.2b**). Red symbols indicate sites with strong geochemical evidence of hydrocarbons ($n=2$), green symbols indicate sites with inconclusive hydrocarbon signals ($n=4$), and blue symbols indicate sites where hydrocarbons were not detected ($n=8$). Triplicate amplicon libraries are plotted for each condition. Sites with strong thermogenic hydrocarbon signals have distinct microbial populations after high temperature incubation (see also **Table S3.4**), relative to the sites without thermogenic hydrocarbon signals. This is indicated by standard deviation ellipses of the hydrocarbon groups.....64

Figure S3.3. Endospore germination and enrichment after high temperature incubation. Relative sequence abundance of *Firmicutes* in 372 libraries of 16S rRNA gene amplicons before (d0; post-pasteurisation) and after 28 and 56 days of anoxic incubation at 40, 50, or 60°C (see **Table S3.3** for detailed microbial community compositions). Each blue bar represents a single library of 4,635 subsampled reads, and demonstrates an increased proportion of Firmicutes following 28 and 56 days

of incubation. The proportion of *Firmicutes* in d0 libraries was on average 0.9%, and was even lower in libraries from unpasteurized surface sediments (data not shown).....65

Figure S3.4. Phylogenetic association of seep-associated sequences with sequences from other environments. Maximum likelihood tree showing phylogenetic relationships between 42 seep-associated ASVs (bold) and close relatives in the GenBank database. Black circles at the branch nodes indicate >80% bootstrap support (1,000 re-samplings). Scale bar indicates 10% sequence divergence as inferred from PhyML. *Pseudomonas aeruginosa* (accession number Z76672) was used as an outgroup to root the tree.....69

Figure S3.5. Phylogenetic relationships of putative glycyl-radical enzymes with alkylsuccinate synthases. Putative anaerobic alkane-degrading pyruvate-formate lyase enzyme variants from *Desulfohalotomaculum*, *Caminicella* and *Parameldivibacter* thermophilic spores in this study (shown in blue) cluster together with homologous pflD gene sequences found in the oil reservoir bacteria *Petromonas tenebris* (Christman *et al.* 2020) and oil reservoir archaea *Archaeoglobus fulgidus* and *Thermococcus sibiricus*, shown in red. Phenotypic evidence of alkane hydrocarbon biodegradation has been observed at high temperature under anaerobic conditions for these *Archaeoglobus* and *Thermococcus* strains (Khelifi *et al.* 2014; Birkeland *et al.* 2017; Mardanov *et al.* 2009). Reference sequences of alkane succinate synthase (AssA and MasD) genes with corresponding experimental verification of anaerobic alkane degradation are shown in bold. Benzyl succinate synthase (BssA) and naphthyl-2-methyl-succinate synthase (NmsA) sequences are represented by collapsed clades. Black circles at the branch nodes indicate >80% bootstrap support (1,000 re-samplings). Scale bar indicates 10% sequence divergence as inferred from PhyML. A sequence of pyruvate formate lyase (Pfl) from *E. coli* was used to root the tree.....70

Figure S3.6. Seabed hydrocarbon seepage. ROV imagery of hydrocarbon seepage in the vicinity of sites with strong thermogenic hydrocarbon signals on the Scotian Slope. Vertical hydrocarbon migration rates from a subsurface reservoir to the seabed in the study area can be estimated using buoyancy models incorporating Darcy’s law (Arp, 1992). Transport facilitated by vertical migration of hydrocarbons through subsurface sediments at rates of 0.1 m d⁻¹ (Arp, 1992; Rice, 2022) would require on the order of 50 years for dispersal from hydrocarbon reservoirs in 2,000 mbsf Cretaceous sediments (Deptuck and Kendell, 2020) up to the surface and into the marine bottom water (**Figure 3.4a**). However, modelled migration rates predict that gas will migrate 40,000 times faster than oil for any given sediment pore size (Rice, 2022). If migration rates of 0.1 m d⁻¹ apply for gas (density 0.087 g / cm³), oil will correspondingly migrate at 2.5 × 10⁻⁵ m d⁻¹ (0.8 g / cm³, 45 API gravity) leading to timescales on the order of eight million years to reach the surface in a similar system. Larger pore sizes especially those associated with faults and fractures (see **Figure 3.1b**) would facilitate faster dispersal from deep sediments up to the surface seabed72

Figure S3.7. Endospores remain viable during burial. Detection of endospore-forming bacteria with the same taxonomic classification as seep-associated ASVs (cf. **Table S3.5**) in deeper Scotian Slope

sediment layers following incubation at 50°C. Bubble size indicates the maximum relative sequence abundance of a taxon in unrarefied sequence libraries at the indicated depth.....73

Figure 4.1. Deep sea seabed sampling sites and hydrocarbon geochemistry. **a**, Location of 111 sediment cores in the Eastern Gulf of Mexico and 14 sediment cores offshore Nova Scotia tested for geochemical evidence of migrated thermogenic hydrocarbons. The red points labelled ‘*EGM147*’ and ‘*NS 18-7*’ indicate the location of the EGM and NS cores. **b**, Unresolved complex mixture (UCM) concentration and total scanning fluorescence (TSF) intensity averaged across three depth intervals of individual gravity cores at the 111 seabed locations in the Eastern Gulf of Mexico. The blue box highlights thresholds of the UCM and TSF for the unambiguous presence of migrated thermogenic hydrocarbons in sediment extracts. TSF thresholds were raised five-fold from the industry standard of 10,000 intensity units to account for additional terrigenous input of organic matter in the Gulf of Mexico. The red point (labelled ‘*EGM147*’) indicates geochemical parameters of the EGM core sediment used for the experimental treatment conditions presented in this study. Data and plot reproduced from Chakraborty *et al.* (2018). **c**, Average nC_{17}/nC_{27} ratio and % 27 d β S (diasterane) in liquid hydrocarbon extracts at intervals from 14 sediment core locations on the Scotian Slope (cf. *Figure S3.1*). The red point (and ‘*NS 18-7*’ label) indicates geochemical parameters of the NS sediment used in the experiments performed here (cf. *Table 4.1*). Black, green, and white points represent other sites with strong, inconclusive, and absent geochemical signals of thermogenic hydrocarbons, respectively (based on these and other geochemical parameters). Additional geochemical data supporting these classifications is provided in *Table S4.2*. An overview of the NS sediments is also provided in *Figure 3.2a*.....80

Figure 4.2. Chemical signals of metabolism in the endospore-forming population. **a**, Changes in sulfate and organic acid concentrations after 14 days of incubation of the EGM and NS sediments corresponding to the different experimental treatment conditions (cf. *Table 4.1* and *Table S4.4*) Negative control incubations (no sediment; no crude oil) did not reveal any change in sulfate or organic acid concentrations (not shown). **b**, Average number of ASVs affiliated with the phylum *Firmicutes* and average relative abundance of *Firmicutes* in sequence libraries before incubation (“0 days”; i.e., post pasteurisation) and after 14 days of incubation (“14 days”) underneath the corresponding experimental conditions.....83

Figure 4.3. *Firmicutes* compositional changes under different experimental conditions. Relative sequence abundance of 16S rRNA gene amplicons affiliated with family-level groups belonging to the phylum *Firmicutes*. Data reflects 0 days of incubation (i.e., post pasteurisation at 80°C but prior to incubation at 50°C) and 14 days of incubation under the respective treatment conditions (see *Table S4.5* for additional data). Treatment 3 (oil and medium without sediment) is omitted from the graph due to insufficient sequence reads to reliably profile the microbial community, likely owing to the absence of germination and activity under this condition (*Figure 4.2* shows corresponding absence of chemical signals of metabolism in Treatment 3). Taxa shown combine to represent greater than 25%

of the total sequence reads across all libraries. Bubble colours indicate class-level taxonomy indicated in the colour key at the bottom.....85

Figure 4.4. Prevalence of organisms phylogenetically related to seep indicator species under different experimental conditions. Count of indicator species hits (y-axis) is the number of unique ASVs in the libraries of each experimental treatment that share $\geq 95\%$ 16S rRNA gene sequence identity with 48 indicator species. Abundance of indicator species hits (x-axis) is the total relative abundance of the same ASVs averaged across the replicates of a treatment condition (see **Table S4.8** for non-averaged data). Squares and circles represent sediment from EGM and NS, respectively, with the exception of Treatment 6 (no sediment) denoted by the white diamond. Arrows highlight results without and with crude oil amendment, indicating that the addition of crude oil resulted in detection of a higher number and abundance of organisms phylogenetically related to 48 seep indicator species, provided that the sediment had not been autoclaved. This suggests crude oil-amendment can improve the detection of seep bioindicators in deep sea sediments.....87

Fig. 4.5. Applicability of regionally-defined seep indicator species for assessment of hydrocarbon seepage in other prospective areas. Sequence data derived from high temperature incubations of Eastern Gulf of Mexico and offshore Nova Scotia sediments reported in Chakraborty *et al.* (2018) and *Chapter 3*, respectively. **a**, Points represent the total relative abundance of sequence reads from one region that share $\geq 95\%$ sequence identity to indicator species defined in the other region. For the Gulf of Mexico (left) this total is plotted for sequence libraries derived from 111 high temperature sediment incubations (111 locations), based on comparisons with the 42 NS indicator species. For Nova Scotia (right) this total is plotted for sequence libraries derived from 178 high temperature sediment incubations (14 locations), based on comparisons with the 6 EGM indicator species. Comparative indicator assessments and relative abundance data is further summarized in **Table S4.9**. Sequencing libraries are grouped according to sites where hydrocarbon geochemical signals of seepage were present (“HC positive”) vs sites where hydrocarbon geochemical signals of seepage were absent (“HC negative”). Boxplots illustrate distributions of the data shown in the individual points. In both locations, quantifying organisms related to indicator species identified from the other region produced a higher average abundance for HC positive sediments than HC negative sediments. **b**, Relative abundance of two taxonomic groups that were found to better differentiate between HC positive and negative site groupings (**Table S4.9**).....89

Figure S4.1. Microbial community composition under different experimental conditions. Nonmetric multidimensional scaling of the Bray-Curtis dissimilarity in microbial community composition after 14 days of incubation under the respective treatment conditions. Squares and circles represent sediment from EGM and NS, respectively, with the exception of Treatment 6 (no sediment) denoted by the white diamond. Statistically significant differences in microbial communities were observed between the different treatment conditions (PERMANOVA, r^2 0.35, $P < 0.05$).....101

Chapter 1

Introduction

1.1 Biogeography of microorganisms

Vast functional diversity enables microorganisms to influence whole ecosystems and global biogeochemical cycles (Falkowski *et al.* 2008). Studies describing microbial diversity across different temporal and spatial scales are numerous (e.g., Turnbaugh *et al.* 2007, Thompson *et al.* 2017; Sunagawa *et al.* 2020), but challenges remain in explaining the complexities of community assembly in many ecosystems. Microbial biogeography identifies patterns of microbial diversity in an attempt to uncover the underlying factors which influence the distribution of microorganisms (Hanson *et al.* 2012).

The restriction of microbes to a defined location or a specific environment type, along with genetic similarities among the inhabiting populations, are evidence of organisms displaying biogeographic patterns (Martiny *et al.* 2006). Four fundamental ecological and evolutionary processes — selection, drift, mutation, and dispersal — are proposed to create and maintain these patterns (Velland, 2010; Hanson *et al.* 2012). Selection results from the growth of microbes that are relatively better adapted to the conditions of their surrounding environment. Drift is the stochastic and inherent characteristic of a microbial population that leads to chance changes in the frequency of a species, independent of their environment or phenotypic variations among populations. Mutation is a change in the genetic sequence of a genotype. Dispersal is the physical movement of microbes to a new geographic location. Understanding the effect of these four processes on the relationship between microbial community compositional similarity and spatial distance, referred to as the distance–decay relationship, is important to uncover the ecological roles and environmental niches of microorganisms (Hanson *et al.* 2012).

In some environments, for example Earth's subsurface, it is unclear how certain ecological and evolutionary processes could occur and therefore the extent to which these 'drivers' actually drive community composition (Becraft *et al.* 2021). Additional complexity is introduced by the fact that certain microorganisms may be found everywhere (Baas Becking, 1934) and acting as a 'seed bank' for selection under the right environmental conditions (Finlay, 2002; Whittaker and Rynearson, 2017), whereas some may be dispersal limited and confined to certain geographic locations (Whittaker *et al.* 2003; Müller *et al.* 2014). As such, efforts are required to disentangle the processes that influence biogeography to better understand the ecology of microorganisms.

1.2 The microbial deep biosphere

Marine sediment covers ~70% of Earth's surface. Since the first discovery of subseafloor microbial populations (ZoBell and Anderson, 1936), subsequent scientific ocean drilling has revealed a 'deep biosphere' that exists below the first few meters of the surface to a depth of ~2.5 km below the seafloor (Inagaki *et al.* 2015). The size of the subseafloor biosphere has been estimated from direct cell counts (Parkes *et al.* 1994; Whitman *et al.* 1998) and lipid biomarkers (Lipp *et al.* 2008). Studies investigating cell abundance in a geographically diverse range of marine sediments using advanced analytical and statistical approaches estimate between 2.9×10^{29} (Kallmeyer *et al.* 2012) and 5.4×10^{29} (Parkes *et al.* 2014) cells in the marine deep biosphere. This equates to a habitat comprising 12–45% of Earth's living microbial biomass (Morono *et al.* 2020) and Earth's largest pool of organic carbon (Orsi *et al.* 2013). On average, cell and biomarker abundances decrease with sediment depth, stoking interest in the survival of microbial life at the extremes of energy and physical limitations. In addition, deviations from the depth-cell number decay trend raise more questions about the interplay between each of the fundamental ecological and evolutionary processes — selection, drift, mutation, and dispersal — in the deep biosphere.

Evidence of microbial activity in deep marine sediments lies at the intersection of geochemical and microbiological research. Biogeochemical profiles (D'Hondt *et al.* 2002; Røy *et al.* 2012; Heuer *et al.* 2020), the presence of intact polar lipids (Lipp *et al.* 2008; Xie *et al.* 2013), gene expression studies (Orsi *et al.* 2013), active uptake of isotope-labelled substrates (Morono *et al.* 2011; 2020), and enrichments of viable cells (D'Hondt *et al.* 2004; Fang *et al.* 2017, Heuer *et al.*

2020; Beulig *et al.* 2020) highlight an active deep biosphere. These microorganisms require energy that is ultimately conserved from the catalysis of redox reactions, with the amount of energy derived dependent on free energy under *in situ* conditions with both biotic, e.g., syntrophy (Morris *et al.* 2013), and abiotic, e.g., temperature (Amend *et al.* 2001), influences. In regions of the ocean with high sedimentation rates, oxygen is typically removed by aerobic respiration in surface sediments, necessitating anaerobic respiration in underlying sediments using (reducing) a sequential series of electron acceptors with decreasing energy yields (e.g., NO_3^- , Mn^{4+} , Fe^{3+} , SO_4^{2-} , and CO_2). A corresponding zonation of dissolved metabolic products (NH_4^+ , Mn^{2+} , Fe^{2+} , S^{2-} , and CH_4) can be observed as a result (D'Hondt *et al.* 2004). Succession between the different terminal electron acceptors varies in depth, largely constrained by organic carbon flux to the seabed and subsequently to the underlying sediments (Parkes *et al.* 2005; Jørgensen and Kasten, 2006). In low productivity ocean environments, such as subtropical gyres and the abyssal plains, oxygen has been shown to penetrate deeper into the seabed and even to the basement, supporting the activity of aerobic communities in sediments deposited over a hundred million years ago (Røy *et al.* 2012; D'Hondt *et al.* 2009, 2015; Morono *et al.* 2020).

Due to the relatively low concentrations of higher energy electron acceptors and their preferential consumption in surficial sediments, sulfate reduction, methanogenesis, and in particular fermentation are the predominant forms of energy conservation in the deep biosphere (D'Hondt *et al.* 2004; Orsi *et al.* 2013; Bowles *et al.* 2014). These metabolisms have important effects on global biogeochemical cycling of carbon, sulfur, and other nutrients. Intervals of active sulfate reduction, methanogenesis, and manganese reduction highlight the co-occurrence of certain metabolic processes in deep sediments (D'Hondt *et al.* 2004). This has been attributed to factors including severe energy limitation, presumably linked to the availability of buried organic and inorganic compounds, preventing the dominance of a single microbial metabolism (Parkes *et al.* 2005), the presence of non-competitive substrates, e.g., methylated compounds that are only consumed by methanogens (Oremland and Polcin, 1982; Valentine, 2011), and *in situ* free energies of the metabolisms differing from standard free energies (D'Hondt *et al.* 2002; LaRowe and Amend, 2016; Bradley *et al.* 2020). An outcome of this is that favourable electron acceptors can remain in sediments during burial. Novel metabolisms such as ethanogenesis and propanogenesis may also be occurring in the deep biosphere (Hinrichs *et al.* 2006). Divergence from standard redox zonation

typically observed with increasing depth and the apparent metabolic versatilities of the microbial populations highlights the complexity of energy conservation in deep marine sediments.

Organic matter input to the sedimentary seafloor biosphere originates from photosynthetic activity in the overlying ocean or on land. Under anoxic conditions, complex compounds such as proteins, carbohydrates, and lipid molecules are broken down by heterotrophic microbes using hydrolytic extracellular and membrane-bound enzymes (Biddle *et al.* 2006; Lloyd *et al.* 2013; Takano *et al.* 2010). The products are consumed by fermentative (Soffientino *et al.* 2009) and acetogenic (Lever *et al.* 2010; 2011) microorganisms, producing metabolites, e.g., hydrogen and acetate, which are then utilised in terminal electron-accepting processes. During this sequential breakdown, some organic compounds are re-assimilated into cells creating new *in situ* biomass. The remains of dead microbial cells likely contribute biomolecules to the deep subsurface over geological timescales (Burdige, 2007; Bradley *et al.* 2018), with necromass recycling times, estimated from slow chemical racemization between D- and L- amino acids, to be on the order of hundreds to thousands of years (Lomstein *et al.* 2012). Burial by sedimentation generally constrains the supply of organic matter to the deep biosphere. The reactivity of organic matter decreases by more than tenfold with each tenfold increase in age (Parkes *et al.* 1994; Jørgensen, 2006), and microbes at deeper depths generally have to rely on increasingly refractory organic matter to survive.

Despite the limited supply of organic matter from above, the deep biosphere still supports a huge microbial population. As such, alternative organic and inorganic compound sources could be important. Hydrogen can be generated by abiotic processes including the thermal activation and aromatisation of organic matter during burial (Parkes *et al.* 2007), mineral weathering and oxidation reactions (Stevens and McKinley, 1995), tectonic fracturing and faulting events (Freund *et al.* 2002), free radical reactions resulting in water hydrolysis (Parkes *et al.* 2011), and aqueous alteration of mantle rocks (Klein *et al.* 2013; Brovarone *et al.* 2020) benefitting diverse hydrogen-utilising (hydrogenotrophic) microorganisms. Thermochemical alteration of organic matter that persists during burial can produce acetate and petroleum that supports significant microbial populations (Wellsbury *et al.* 1997; Horsfield *et al.* 2006). Other geological features such as cold seeps (Ruff *et al.* 2015; Chakraborty *et al.* 2020; Joye, 2020), hydrothermal vents (Dick, 2019), mid-ocean ridges (Summit and Baross, 2001; Schrenk *et al.* 2010), and mud volcanoes (Niemann

et al. 2006; Ijiri *et al.* 2018) also transport electron donors to subsurface and surface microbes offering more abundant growth substrates. Sustaining life in the deep biosphere in effect involves a complex interplay of biological, geological, and geochemical processes.

Compared to surface environments, rates of metabolic activity and growth in the deep biosphere are extremely slow. Cell generation times increase with depth and community turnover rates are estimated to be on the order of 100–2,000 years or longer (Biddle *et al.* 2006; Parkes *et al.* 2000; Jørgensen and Marshall, 2016). Metabolic rates generally decrease by an order of magnitude over a similar depth interval (Lomstein *et al.* 2012; Hoehler and Jørgensen, 2013; Jørgensen and Marshall, 2016), however, recent research has shown that elevated rates of energy metabolism may be sustained by thermally driven reactions of sedimentary organic matter (Beulig *et al.* 2022). Exact cell specific rate measurements are challenging though due to the size and variability of subsurface environments, as well as their inaccessibility for sampling.

The lower boundaries of the deep biosphere are likely defined by the bioavailable energy fluxes described above, coupled with physical environmental constraints (Takai *et al.* 2008; Inagaki *et al.* 2015). The temperature of marine sediments increases with depth due to an average global geothermal gradient of 20–40°C km⁻¹ presenting increasingly challenging conditions for microbial populations to meet their cellular maintenance energy requirements (Hoehler, 2004). A recent investigation of microbial life in the Nankai Trough subduction zone revealed biological methane production and oxidation at temperatures of 85°C, and isotopic and cell concentration-based evidence of acetate-degrading hyperthermophiles in 120°C sediments, with the latter defining the current boundary for temperature limits to life in the deep subsurface (Heuer *et al.* 2020). Lower temperature limits have been reported under different environmental conditions (Wilhelms *et al.* 2001; Roussel *et al.* 2008; LaRowe *et al.* 2017). As with temperature, hydrostatic pressure increases by ~15 MPa per kilometer of depth below the sea surface (Cario *et al.* 2019) with microbes employing metabolic and physiological specialisations to cope (Allen and Bartlett, 2002; Cario *et al.* 2015). The direct effect pressure has on metabolism and growth of deep biosphere microbes is still unclear though (Cario *et al.* 2019), as is the indirect effect pressure exerts on subsurface populations by reducing permeability and porosity, and by association limiting substrate transport and microbial motility. Ultimately both temperature and pressure likely control

the limits of deep biosphere cell viability and highlight the important influence of Earth system processes on microbiology.

1.3 Petroleum reservoir ecosystems

Petroleum systems, in a geological context, consist of a petroleum source rock, migration pathway, reservoir rock, trap, and seal (Magoon and Dow, 1994). Heat-driven alteration (maturation) of organic matter in deep subsurface source rocks generates petroleum compounds, e.g., saturated and aromatic hydrocarbons, that migrate to a reservoir rock. Sufficient porosity and permeability of the reservoir, coupled with an impermeable trap or seal restricting further migration, leads to an accumulation of petroleum. The relative timing of the formation of these geological elements and generation-migration-accumulation processes determines hydrocarbon preservation (Magoon and Dow, 1994). The formation of petroleum systems creates conditions for environmental selection in the deep biosphere by influencing the form and distribution of organic carbon, oxidants, and nutrients in the subsurface (Head *et al.* 2003; 2014). Petroleum reservoirs have been reported to support microbial cell densities as high as 10^7 cells cm^{-3} (Orphan *et al.* 2000; Bennett *et al.* 2013; Vigneron *et al.* 2017) at depths where cell counts in surrounding sediments are typically one or more orders of magnitude lower.

Reservoir microbiomes are metabolically diverse (Magot *et al.* 2000). Consistent with the majority of the deep biosphere, microbes indigenous to petroleum reservoirs respire anaerobically *in situ*. Sulfate-, nitrate- and iron-reducing, methanogenic, and fermentative metabolisms have been reported in oil fields (Ollivier and Magot, 2005). Methanogenic pathways in reservoirs contribute to hydrocarbon biodegradation over geological timescales of hundred of thousands to millions of years (Zengler *et al.* 1999; Jones *et al.* 2008). Metabolism of petroleum compounds typically proceeds through syntrophic partnerships of hydrocarbon-degrading bacteria and methanogenic archaea (Dolfing *et al.* 2008; Gieg *et al.* 2008), however recent studies have shown that archaeal species can combine the degradation of alkanes with methanogenesis without a syntrophic partner (Laso-Pérez *et al.* 2019; Borrel *et al.* 2019). This metabolism was most recently shown in *Candidatus* Methanoliparum cultured from an oil reservoir (Zhou *et al.* 2021). Subsurface hydrocarbon biodegradation is linked to sulfate-reduction and iron-reduction metabolisms if

sufficient concentrations of sulfate and iron are present (Reuter *et al.* 1994; Gieg *et al.* 2010; Lovley *et al.* 1989). Similarly, fermentation of hydrocarbons is thermodynamically feasible and proceeds when coupled with methanogenic conversion of hydrogen and acetate (Zengler *et al.* 1999; Liu *et al.* 2018).

Microbial degradation of hydrocarbons involves a variety of biochemical reactions. Under oxic conditions, initial activation of C–H bonds, representing the first step in the degradation process, occurs via oxygenase-catalyzed reactions that introduce oxygen atom(s) into the substrate (Xu *et al.* 2018). Under anoxic conditions, addition to fumarate, carboxylation, hydroxylation, and methylation are mechanisms of initial activation. Addition of *n*-alkanes and aromatic compounds to fumarate, catalyzed by the glyceryl radical family of enzymes — alkylsuccinate synthase, benzylsuccinate synthase or naphthyl-2-methylsuccinate synthase, encoded by *ass*, *bss*, and *nms* genes respectively — represents the best described pathway (Rabus *et al.* 2016). The genes encoding the enzymes can be used to assess hydrocarbon utilization in genomes and metagenomic datasets from diverse environments (Khot *et al.* 2022).

Members of the phyla *Firmicutes*, *Thermotogae*, *Thermodesulfobacteria*, *Euryarchaeota*, *Crenarchaeota*, and *Proteobacteria* have been reported in microbial assessments of oilfield samples. Physical factors that exert controls on the composition of the deep biosphere in general, such as temperature and pressure, are believed to influence community assembly in petroleum reservoirs (Hu *et al.* 2016). Microbiological characterisation of subsurface oil fields at temperatures >50°C showed that *Firmicutes* was the dominant phylum (Hubert *et al.* 2012), while *Thermococcales* and *Archaeoglobales*, belonging to the phylum *Euryarchaeota*, have been isolated from petroleum reservoirs with *in situ* temperatures >70°C (Gittel *et al.* 2009; Gao *et al.* 2016). Related to this, temperature is considered an important control on the occurrence of biodegraded oil reservoirs. Geochemical observations of *in situ* crude oil biodegradation in deeply buried and tectonically uplifted systems supports the palaeosterilization model of oil reservoir habitability, which posits 80–90°C as an upper limit of activity by hydrocarbon degrading organisms in these environments (Connan, 1984; Wilhelms *et al.* 2001; Head *et al.* 2003). Heat-sterilized oil reservoirs that are subsequently uplifted to depths where temperatures are below 80°C, still do not experience biodegradation despite these habitats containing plentiful substrates. An implication of these findings is that dispersal of microbial cells with hydrocarbon-degrading

metabolisms into these oil-bearing sediments is not a process influencing the microbial biogeography of the deep biosphere (Parkes and Maxwell, 1993; Røling *et al.* 2003).

Chemical factors influence microbial metabolisms, and by association community composition in petroleum reservoirs. Biodegradation gradients are observed in reservoirs with the most biodegraded oil at the oil-water transition zone (Head *et al.* 2003), potentially owing to electron donor supply from the above oil leg and electron acceptor and nutrient supply from the below water leg fueling microbial communities at this interface. Nutrient supply and salinity are also controls on the net *in situ* degradation rate of hydrocarbons (Larter *et al.* 2006). A less researched topic is the selective pressure on microorganisms owing to petroleum toxicity. Hydrocarbons have been shown to interact with membrane structures of microbes, changing fluidity and protein configurations and affecting membrane-bound and embedded enzyme activity leading to disruptions of functions important for energy conservation (Sikkema *et al.* 1995; Van Hamme *et al.* 2003). The effect of petroleum toxicity when coupled to other environmental stressors in the subsurface, e.g., high temperature, is unclear, but presumably reservoir microorganisms must be adapted to cope.

As a function of the variable physicochemical characteristics, often related to the depositional setting and tectonic history of the basin where a petroleum system forms, defining the microbial community and functional potential that is actually indigenous to the reservoir is challenging. Additionally, the potential for exogenous introduction of microorganisms during exploration and production activities that precede sampling, e.g., drilling and seawater injection to maintain reservoir pressure, likely influences the microbial community composition of the petroleum reservoir ecosystem.

1.4 Microbial survival strategies

Specialization to survive extreme conditions is a requisite for most life in deep subsurface sediments. Microbial communities employ a diverse range of physiological adaptations. Changes in cell morphology, such as becoming smaller and more spherical, reduces surface area and the energy requirements of a cell (Lever *et al.* 2015). Increased numbers of transporter proteins or

surface proteins with a high substrate affinity enhance nutrient uptake (Lever *et al.* 2015). Efficient ATP synthesis pathways (Schink, 1997), decreasing membrane permeability to ions (Lipp *et al.* 2008, Lloyd *et al.* 2013), increasing transmembrane permeability to protons (Lolkema *et al.* 1994), and limiting biomolecule decay (Orsi *et al.* 2013; Onstott *et al.* 2014) are also physiologies that would favour survival under extreme conditions. Research suggests that these subsurface survival traits are selected for and are not acquired during burial (Starnawski *et al.* 2017), and that microbes persisting during sedimentation in the deep biosphere are capable of living there before they get there. An example of this may be the apparent requirement for minimal evolution shown by extreme genome conservation — possibly reflecting high-fidelity DNA replication and repair mechanisms — in *Candidatus Desulforudis audaxviator* which inhabits geographically widespread deep subsurface locations that were last connected before the physical breakup of Pangea 165 and 55 Ma ago (Becraft *et al.* 2021).

An additional explanation of survival in the deep biosphere is that cells can enter a physiologically inactive state, known as microbial dormancy. Endospore formation is the most resistant form of dormancy, and represents a strategy for microbial populations to survive unfavourable environmental conditions (Lennon and Jones, 2011). Accordingly, endospore-forming bacteria are often reported in environments where conditions fluctuate and favour intermittent survival stages, although, endospores are also abundant in the relatively stable deep biosphere (Lomstein *et al.* 2012). Recent research estimates a significant population of 2.5×10^{28} to 1.9×10^{29} bacterial endospores in the uppermost kilometer of marine sediment, which would indicate that the total biomass of endospores exceeds that of vegetative cells (Wörmer *et al.* 2019). The viability of endospores in sediments ~2.5 km beneath the seafloor (Fang *et al.* 2017; Trembath-Reichert *et al.* 2017) highlights the effectiveness of this survival mechanism. Interestingly, the rate of depth decay of endospore-forming populations varies between mesophiles and thermophiles, the latter presenting more resistant forms with shallower decay trends (Wörmer *et al.* 2019). As with vegetative cells, this may infer the role of selection of existing physiological traits even within dormant populations, rather than *in situ* adaptive evolution, to define deep biosphere microbial community composition.

Viable dormant microbial populations that persist in the subsurface form a genetically and functionally diverse microbial ‘seed bank’ that can be revived under the right environmental

conditions (Pedrós-Alió, 2006; Lennon and Jones, 2011; Mestre and Höfer, 2020). The abundance of electron donors and acceptors in petroleum reservoirs could be one such example of the ‘right conditions’, fueling deep biosphere microbial growth. This conveys the importance of ecological processes such as environmental selection in influencing microbial community assembly and biogeographic patterns of microbes in the deep biosphere.

1.5 Endospore-forming bacteria

Formation of endospores, termed sporulation, offers a mechanism to protect the genome of the cell enabling persistence in an unfavourable environment (Nicholson *et al.* 2000). Sporulation is a characteristic of many members of the phylum *Firmicutes*, one of the earliest branching bacterial phyla (Ciccarelli *et al.* 2006). Endospore formation is distributed widely across the phylum, with spore-forming species in most classes, including the primarily aerobic *Bacilli* and primarily anaerobic *Clostridia*. Deprivation of key nutrients is thought to be the primary stimulus for sporulation (Lopez *et al.* 1979; Grossman and Losick, 1988; Stragier and Losick, 1996; Sorg and Sonenshein, 2008), while exposure of anaerobes to oxygen and DNA damage also induces this physiological change (Sauer *et al.* 1995; Ireton *et al.* 1995). These stressors are consistent with the environmental challenges of energy limited deep subsurface sediments and may contribute to the high abundances of endospores in this biosphere. Interestingly, triggers such as increasing cell density and pH fluctuations associated with the accumulation of organic acids during the exponential phase of cell growth promote sporulation in the presence of excess nutrients (Burbulys *et al.* 1991; Fujita and Losick, 2005; Long *et al.* 1984), indicating dormancy even during favourable conditions. This potentially highlights a bet-hedging strategy of thriving populations to ensure survival in the event of sudden adverse environmental changes (Lennon and Jones, 2011; Ellegaard and Ribeiro, 2018). Variability in the signal which induces sporulation in a population may reflect the characteristics of the environments they inhabit (Dürre and Hollergschwandner, 2004; Schultz *et al.* 2009). However, studies which examine sporulation triggers with respect to physiological adaptations to environmental conditions are limited.

Sporulation is a form of cell differentiation with a series of distinct, energy intensive morphological changes driven by and coupled to gene expression. The sporulation process

involves hundreds of genes, although these genes are not conserved between taxa, and defining the minimum genomic requirement for sporulation has proved challenging (Galperin *et al.* 2012). Gene regulation typically relies on the stationary-phase sigma factor σ^H and the master transcriptional regulator Spo0A, encoded by the *spo0H/sigH* and *spo0A* genes, respectively (Alsaker and Papoutsakis *et al.* 2005; Hoch, 1993). The sporulation process involves asymmetric cell division preceding differentiation into two cell types, the mother cell and the endospore. The genome of the endospore is contained within a dehydrated core where water has been replaced by 2,6-pyridine dicarboxylic acid (dipicolinic acid, DPA) — encoded by dipicolinic acid synthetase, *dpa* (Daniel and Errington, 1993) — in a complex with divalent calcium ions (Ca^{2+}). DPA is considered a diagnostic endospore biomarker, and accurate quantification of this compound has been used for estimates of spore abundance in diverse samples and environments (Fichtel *et al.* 2007; Rattray *et al.* 2021), including marine sediments resulting in the global abundance estimates reported above (Wörmer *et al.* 2019). DNA is protected by binding to α/β type small, acid-soluble proteins (SASP) less than 100 amino acids in length encoded by the *ssp* genes (Setlow, 1988). The outermost layer is a complex proteinaceous coat that confers resistivity to environmental stress but may also contribute to environmental selection during germination under favorable conditions (Al-Hinai *et al.* 2015). Diversity in spore physiology, like the factors which initiate sporulation, may be driven by adaptation to a variety of niches permitting spores to survive in diverse environments (McKenney *et al.* 2012).

1.6 Endospores as model organisms for microbial dispersal

Physiological resilience of bacterial endospores to extreme physical and chemical conditions, e.g., nutrient limitation, desiccation, and extremes of temperature and pressure (Nicholson *et al.* 2000; Margosch *et al.* 2006; Setlow, 2014), means that spores can persist and remain viable for very long periods, and potentially over geological timescales of millions of years (Cano and Borucki, 1995; Vreeland *et al.* 2000). Presumably this confers an evolutionary advantage by increasing their ability to be dispersed and by association increasing the possible extent of dispersion. The inactivity and genetic stability of these dormant populations means that the effect of dispersal can be investigated without the confounding influences of selection, drift, and mutation which

typically shape the biogeography of active microbial populations. Non-random patterns in the spatial distribution of endospores should therefore be attributable to dispersal limitation decreasing the ubiquity of microorganisms (Martiny *et al.* 2006; Müller *et al.* 2014).

Conspicuous occurrence of thermophilic endospores in cold environments to which their metabolisms are not suited has been proposed as evidence for cell dispersal from warm environments (Isaksen *et al.* 1994; de Rezende *et al.* 2013). Based on high temperature endospore germination experiments, Hubert *et al.* (2009) estimated a constant flux of thermophilic bacteria into the cold Arctic seabed at a rate exceeding 10^8 spores $m^{-2} y^{-1}$. In identifying the source habitat(s) of these ‘misplaced’ microbes, the dispersal vectors and dispersal histories can be studied in a biogeographic context. Phylogenetic analysis has revealed populations of seabed thermophiles have close relatives in mid-ocean ridge and petroleum reservoir ecosystems (Hubert *et al.* 2009; Müller *et al.* 2014; Chakraborty *et al.* 2018). Petroleum reservoirs are connected to the overlying ocean by hydrocarbon seepage driven by overpressure and the buoyancy of hydrocarbons accumulated in the subsurface. Hydrocarbons migrate upwards through networks of faults and fractures, and along vertical geological structures such as salt diapirs, to the seafloor (Abrams, 2005). This geological process potentially represents a dispersal vector for cell transport from deep to shallow sediment layers (Ruff *et al.* 2019; Chakraborty *et al.* 2020). The connection between petroleum reservoir habitats and the surface means that hydrocarbon seeps represent portals into the subsurface geosphere. Despite the obvious potential for contributing to and influencing the seabed microbiome, investigations of dispersal from petroleum systems that combine surface and subsurface observations are lacking.

Long-distance and long-term dispersal of microorganisms from favorable to unfavorable environmental conditions can result in physiological cycles of activity and dormancy. If this spatial dispersal eventually leads to transport of certain individual cells back to their original starting environment, selection of the source population occurs. This process implies that the dispersal of microorganisms could be spatially cyclical. Based on this logic, Mestre and Höfer (2020) proposed an oceanic ‘microbial conveyor belt’ linked to global ocean circulation. In this model, microorganisms that, for example, prefer high hydrostatic pressure, thrive during sinking but become dormant during upwelling, while the converse would be the case for photosynthetic microorganisms that become dormant while sinking. Mestre and Höfer (2020) indicate how this

could control the oceanic distribution of microorganisms, influencing cycling of organic and inorganic matter. However, they also note that complexities in this model are introduced when it is connected with or influenced by other biospheres, e.g., the submarine sedimentary biosphere.

1.7 Thesis overview

This thesis aims to understand environmental selection in subsurface petroleum reservoirs, as well as evaluate thermophilic endospores as model organisms to assess global scale microbial dispersal originating from these habitats. These new ecological insights will support descriptions of future biotechnological applications of this knowledge.

This thesis includes three results chapters (2–4) preceded by the *Introduction* above.

Chapter 2, “*Environmental selection in subsurface petroleum reservoirs from around the world*”, is a systematic study of the petroleum reservoir microbiome. Analysis of 368 samples comprising over 500 million sequences from oil reservoirs globally are assessed for evidence of a taxonomic and functional core microbiome. Community composition is evaluated with respect to differences in *in situ* temperature regime and between reservoirs in different geographic locations. Genomic capabilities for carbon acquisition and energy conservation are investigated across diverse reservoir conditions and compared to other deep biosphere metabolisms. With these assessments, environmental selection in petroleum reservoir habitats is uncovered, highlighting potential microbial responses associated with oil recovery.

Chapter 3, “*Geological processes mediate a microbial dispersal loop in the deep biosphere*”, investigates thermophilic endospores conspicuously occurring in the cold seabed to uncover the interplay between dispersal and selection in the marine subsurface. Combining geophysics, geochemistry, microbiology, and genomics, cell transport via geofluids from deep petroleum bearing sediment layers up into the ocean is examined. Measurement of an endospore specific biomarker assesses dispersal back into the deep biosphere via sedimentation, where high energy compounds provided in petroleum-bearing sediments allow re-establishment of the originally

dispersed populations. This study examines a potential for recurrent and spatially cyclical dispersion of living microorganisms through the deep biosphere.

Chapter 4, “*Cold seabed thermophilic endospores germinate and become active in the presence of high concentrations of crude oil*”, aims to provide evidence supporting the oil reservoir provenance of cold seabed thermophiles. Survival in the presence of crude oil is necessary for bacteria that share genomic relationships to microbes inhabiting petroleum reservoirs to originate in these environments. A series of high temperature incubations with crude oil and sediment from areas with proven cold seeps are established. The results of this analysis lead to an assessment of subsurface reservoir-derived microbes as bioindicators of marine cold seeps. Comparing the prevalence of seep indicator taxa in two prospective areas (Nova Scotia and Gulf of Mexico) is further used to examine if seep detection bioassays based on this approach have applicability in new areas of the ocean.

A summary of findings is provided in Chapter 5, along with considerations for future biotechnological applications of this research in relation to the interplay between ecology and usefulness.

Chapter 2

Environmental selection in subsurface petroleum reservoirs from around the world

Daniel A. Gittins¹, Sriyak Bhatnagar¹, Casey R. J. Hubert¹

¹Department of Biological Sciences, University of Calgary, Canada.

2.1 Abstract

Petroleum reservoirs within the deep biosphere are extreme environments characterised by high temperature. Highly diverse microbial communities inhabit these environments creating biogeochemical hotspots in the subsurface. Despite their ecological and industrial importance, systematic studies of the oil reservoir microbiome are lacking, such that knowledge of core microbial taxa and associated genomic attributes are limited. This study fills knowledge gaps in the field of petroleum microbiology by compiling and comparing 343 16S rRNA gene amplicon libraries, as well as 25 shotgun metagenomic libraries, from oil reservoirs around the world. Taxonomic composition varies among reservoirs with different physicochemical characteristics, and by geographic location. Oil reservoirs lack a taxonomic core microbiome, yet gene composition analysis highlights a common functional core. Shared functions include diverse capabilities for carbon acquisition and energy conservation consistent with metabolisms characteristic of the deep biosphere. Genes for anaerobic hydrocarbon degradation were only observed in a subset of the samples and are therefore not considered to represent core biogeochemical functions across all oil reservoirs. Metabolic redundancy within the petroleum reservoir microbiome reveals a system poised to respond to changes in redox biogeochemistry highlighting the potential of genomics for predicting microbial responses associated with oil recovery.

2.2 Introduction

The subsurface biosphere is the largest microbial habitat on Earth (Bar-On *et al.* 2018). Microbes inhabiting this environment play an important role in mediating global scale biogeochemical processes, such as element and nutrient cycling (Orcutt *et al.* 2013). Understanding the factors that shape microbial community structure and functional potential is important to uncovering these cycles and understanding how geology and biology interact. The deep biosphere includes diverse terrestrial and marine habitats such as aquifer systems, basaltic ocean crust, deeply buried sediments and petroleum reservoirs. The latter represents an interesting microbial setting in the subsurface context in that energy rich petroleum compounds offer plentiful substrates and electron donors, while at the same time hydrocarbons are generally considered toxic to microorganisms. It has been observed that petroleum reservoirs contain about an order of magnitude more microorganisms than surrounding sediments at similar depths (Bennett *et al.* 2013), indicating that oil reservoirs may be deep biosphere ‘hot spots’ of microbial activity.

The physiology and metabolism of the petroleum reservoir microbiome influences the physicochemical properties of these environments. Crude oil biodegradation changes the composition and physical properties of both liquid and gas components of petroleum via the sequential metabolism of hydrocarbons and other compounds (Head *et al.* 2003). Methanogenic oxidation of petroleum hydrocarbons over geological timescales is mainly understood as being catalyzed *in situ* by consortia of bacteria and archaea, altering both the chemistry of the remaining oil as well as the levels of CO₂ and CH₄ (Jones *et al.* 2008; Milkov, 2011) producing gas caps in some reservoirs (Larter and Di Primio, 2005; Gray *et al.* 2009). Interestingly, this oil-altering biogeochemistry seems to only take place in reservoirs that have a burial history that has not included depths hotter than 80–90°C, which apparently inactivate these populations by heat sterilization (Wilhelms *et al.* 2001). In situations where sulfate is present, anaerobic respiration of sulfate to sulfide, also known as souring, can be coupled to the oxidation of hydrocarbons directly, or of organic acids or hydrogen indirectly derived from crude oil (Aitken *et al.* 2013; Gieg *et al.* 2010; Sherry *et al.* 2013). Accurate characterisation of these metabolic processes catalysed by the petroleum reservoir microbiome enables useful predictions of reservoir conditions.

Microbial communities in nature are typically highly complex, comprising thousands of species. A ‘core microbiome’ is considered to be the consistent component of complex microbial assemblages present in a given habitat type, as confirmed by observations of common taxa across several different sampling sites (Turnbaugh *et al.* 2007; Hamady and Knight, 2009; Shade and Handelsman, 2011). Identifying the occurrence of a core microbiome is important for unravelling the ecology of an ecosystem. Commonly occurring microorganisms are likely also important for the normal biogeochemical functioning, such that the microbiome helps to define dominant metabolic processes and microbial interactions (Neu *et al.* 2021). It is equally important to ascertain if a given habitat type does not contain core microbial taxa, and the reasons why. Temporal and spatial heterogeneity in microbiota can also occur, resulting in dynamic shifts in functioning across complex microbiomes (Astudillo-García *et al.* 2017). In petroleum reservoirs, identifying common taxa may guide engineering interventions with the potential to manipulate these subsurface communities to achieve desired outcomes.

The most common method for microbial community profiling in oil reservoirs has been to sequence the 16S rRNA gene following its PCR amplification in reservoir samples. This approach has dominated the field of petroleum microbiology for the past 25 years (Voordouw *et al.* 1996), with lower read count clone libraries giving way to much larger high throughput amplicon sequencing libraries. Shotgun metagenome sequencing is now also applied to oil reservoir samples with increasing regularity (Hu *et al.* 2016; Vigneron *et al.* 2017; Christman *et al.* 2020). Metagenomic sequencing advances understanding of the microbial ecology of a given ecosystem relative to single-gene amplicon sequencing by moving past taxonomic assignments to reveal more information about the functional potential within the sampled community. These techniques are routinely applied in other sectors for biotechnological applications ranging from medicine to bioenergy, and are now being applied to oil reservoirs from around the world. Here we provide a meta-analysis of 16S rRNA gene and metagenome sequencing data from published microbial assessments of hydrocarbon reservoir fluids, unveiling diverse taxonomic assemblages and ecosystem functioning, and identifying global patterns that fill knowledge gaps in oil reservoir microbiology.

2.3 Results

2.3.1 Overview of the petroleum reservoir microbiome

Depth and temperature for 368 petroleum reservoir samples spanning six continents ranged from 270–3,550 meters below surface (mbsf; defined as the seabed in offshore settings) and 8–110°C (**Table S2.1**). Analysing nine million quality-controlled sequences from 295 high throughput 16S RNA gene amplicon libraries shows a negative correlation between depth and alpha diversity (Spearman's rank correlation $r = -0.47$ and -0.49 , respectively, $P < 0.05$). The deepest and hottest samples ($> 2,000$ m depth and 74°C average reservoir temperature; $n = 81$; **Table S2.1**) have an average Shannon diversity index H' of 1.6, whereas the shallowest and coolest samples (< 500 m depth and 12°C average reservoir temperature; $n = 76$) have an average Shannon diversity index H' of 2.8 (**Table S2.2**). This correlation of temperature with microbial diversity is apparent for oil reservoirs situated both onshore ($n = 231$) and offshore ($n = 64$).

Oil reservoirs undergoing secondary recovery operations ($n = 197$) are less diverse (average $H' = 1.7$) than reservoirs sampled during primary recovery without any exogenous fluid injection ($n = 98$; $H' = 2.8$). Whereas primary oil recovery is mainly driven by *in situ* reservoir pressure, secondary oil recovery involves re-establishing pressure in the reservoir by injecting fluids. Accordingly, there is a greater opportunity for exogenous organisms to be introduced into reservoir microbiomes during secondary recovery, which could be expected to result in higher diversity as measured by the Shannon index and other metrics. The data here reveal the opposite, suggesting that secondary recovery stimulates enrichment of opportunistic populations within the microbiome.

A total of 2,473 phylotypes (i.e., the highest taxonomic classification applied here as described in the Methods section) were inferred from high throughput 16S rRNA gene amplicon sequencing (**Table S2.3**). Phylotypes encompassed 96 phyla comprising 2,283 genera in total. At the phylum level, *Proteobacteria* (30%), *Euryarchaeota* (19%), *Halobacterota* (15%), *Firmicutes* (9%) and *Campilobacterota* (5%) were most abundant. Classes and genera present at $>1\%$ average relative sequence abundance all belong to the ten most abundant phyla in the dataset (**Figure 2.1**). At the genus level, *Thermococcus*, *Methanosaeta*, and *Methanothermobacter* cumulatively account for

8–4% of the sequences. Oil reservoir clone libraries (48 samples comprising 2,850 reads) support findings from the more extensive high throughput 16S rRNA gene data, with *Halobacterota*, *Proteobacteria* and *Firmicutes* accounting for 26, 20 and 13%, respectively, at the phylum level. Oil reservoir metagenomes corroborate 16S rRNA gene sequencing approaches in terms of the most prevalent taxonomic groups, but highlight differences in relative abundance estimates (**Figure 2.1**). The most abundant phyla in metagenome libraries were *Firmicutes* (27%), *Proteobacteria* (20%), *Euryarchaeota* (9%), *Desulfobacterota* (8%), and *Thermotogota* (6%), corresponding with the most abundant genera in the metagenomes being *Caminicella*, *Pseudomonas*, *Desulfonauticus*, *Petrotoga*, and *Thermoanaerobacter*, each accounting for 15–4% of the genus-level assignments. Inconsistencies in community composition using these different DNA sequencing strategies may reflect true variations between the reservoirs sampled (since few reservoirs had both amplicon and metagenome sequencing applied), or could alternatively reflect well-known methodological differences, e.g., preferential amplification of certain phylotypes during 16S rRNA gene sequence library preparation skewing the reported community composition.

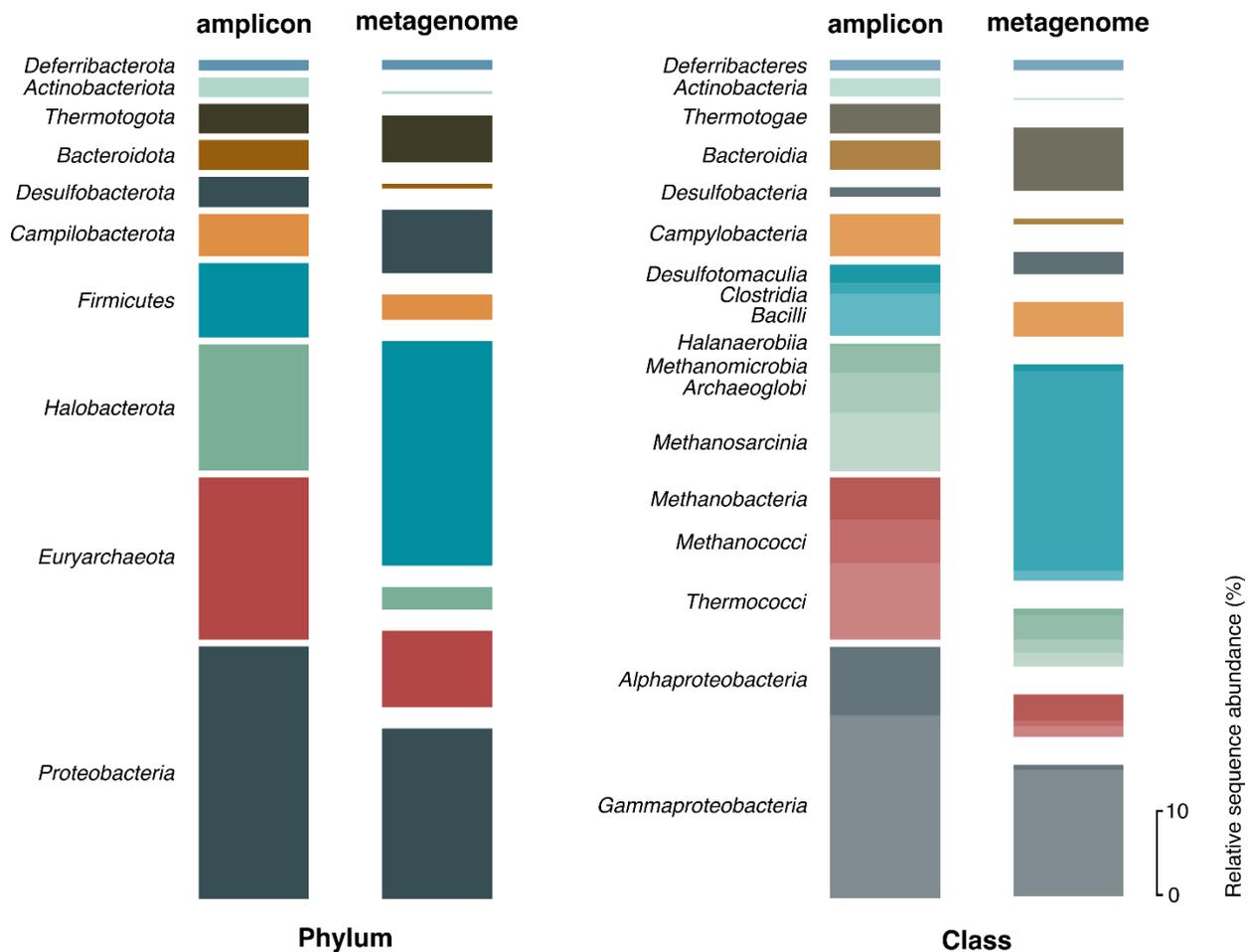


Figure 2.1. Petroleum reservoir microbiome. Average relative sequence abundance of the ten most abundant phyla and all classes >1% relative sequence abundance in 295 high throughput 16S rRNA gene amplicon sequencing libraries, in addition to their corresponding abundance in 25 metagenome libraries. These 10 phyla represent 93% and 78% of the 16S and metagenomes libraries, respectively, and classes 90% and 70%, respectively. Assessing whether or not petroleum reservoirs around the world harbour a core microbiome revealed that the bacterial phyla *Proteobacteria* and *Firmicutes* and the archaeal phylum *Halobacterota* were observed across >75% of the 295 samples analysed by high throughput 16S rRNA gene sequencing. Accordingly, the classes *Gammaproteobacteria*, *Alphaproteobacteria*, *Methanosarcinia*, *Methanomicrobia*, *Clostridia* and *Bacilli* were identified in 82–60% of the samples. In metagenomic datasets, *Firmicutes* and *Proteobacteria*, along with *Euryarchaeota*, were observed in all samples, and at the genus level *Thermovirga*, *Halanaerobium* and *Petrotoga* were observed in more than 75% of the samples. Considering oil reservoir samples associated with primary recovery ($n = 98$), all high throughput 16S rRNA libraries showed Archaea, while *Firmicutes* and *Proteobacteria* were detected in 97 out of 98. Comparative analysis of reservoirs produced by secondary recovery methods showed *Proteobacteria* and *Halobacterota* were the most prevalent phyla, detected in 75 and 69% of the 198 libraries, respectively.

2.3.2 Factors associated with the community structure within oil reservoirs

Reservoir temperature and depth, as with alpha diversity, are correlated with greater variation in high throughput 16S rRNA gene-based community composition, such that as the difference in temperature and depth of two compared samples increases, community composition variance increases (**Figure 2.2a, c; Table S2.4**). Assessments of phylotype presence and abundance for distinct reservoir temperature intervals using *indicspecies* tests (De Cáceres *et al.* 2010; **Table S2.5**) show that most members within the commonly observed phyla (e.g., *Proteobacteria*, *Halobacterota* and *Firmicutes*) were not significantly associated with an individual temperature interval. By contrast, at the genus level, *Thermococcus* and *Petrotoga* demonstrated strong correlations with temperature, being prevalent community members in reservoirs between 71–80 and 81–90°C, respectively. This was consistent with results from metagenomes, particularly for *Petrotoga* (**Table S2.3**).

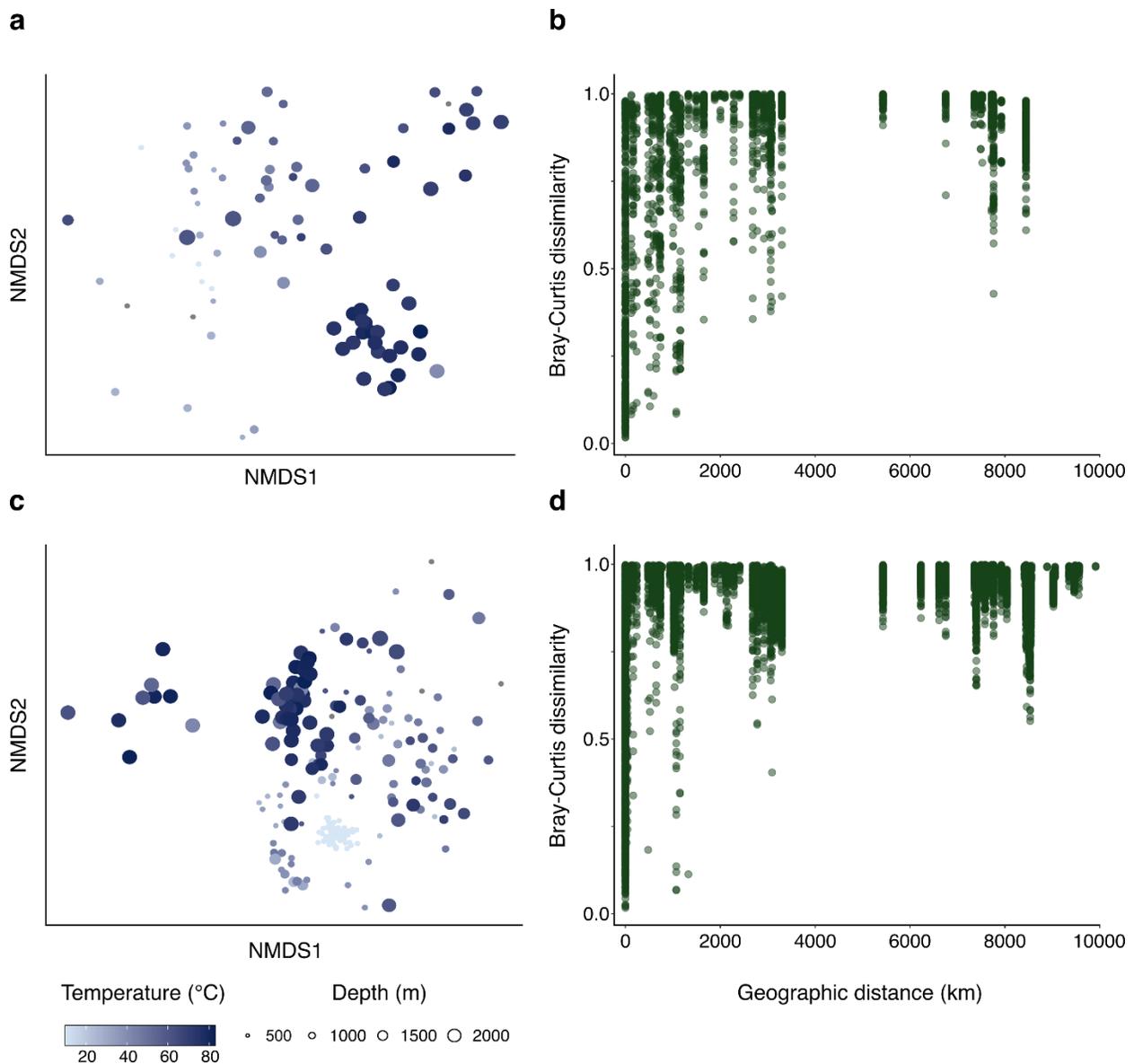


Figure 2.2. Relationship between microbial community composition and environmental factors.

High throughput 16S rRNA gene library comparisons were performed based on the use of archaea-specific primers (**a**, **b**) or universal primers (**c**, **d**). **a**, **c**, Bray-Curtis dissimilarity in microbial community composition between samples from reservoirs at different depths and *in situ* temperatures. Libraries prepared using archaeal primers (**a**) exhibit greater dissimilarity between samples under the tested environmental conditions (Mantel, temperature $r = 0.55$, depth $r = 0.55$) than libraries prepared with universal primers (**c**; Mantel, temperature $r = 0.33$, depth $r = 0.43$), but both represent significant dissimilarity ($P < 0.05$). **b**, **d** Correlation between the geographic location of a petroleum reservoir and Bray-Curtis dissimilarity in the observed microbiome assessed in amplicon libraries using archaeal (**b**) and universal (**d**) primers. Visual trends as well as statistical testing (Mantel, $P < 0.05$) indicate that geographically distant reservoirs have more dissimilar microbial communities.

Instances of individual studies incorporating several oil reservoirs revealed variability in community composition was greater between studies than between the sampled reservoirs. This confirms that sampling techniques may have important impacts on community structure observations (Rachel and Gieg, 2020). Primary or secondary oil recovery methods accounted for only a small fraction of the variation in the microbial community (**Table S2.4**). Slightly weaker correlations were apparent between geographic location and the reservoir microbiome similarity (**Figure 2.2b, d**). As such, reservoirs in similar locations tend to have more similar microbial community compositions than reservoirs separated by large distances. This distance decay relationship is more prominent if only reservoirs produced by primary recovery are considered (Mantel, archaeal $r = 0.87$, universal $r = 0.80$, $P < 0.05$), suggesting a biogeography in pristine oil reservoir deep biosphere environments. Much weaker correlations with geographic distance and oil reservoir microbiomes are observed when reservoirs produced by secondary recovery are incorporated into this analysis (Mantel, archaeal $r = 0.44$, universal $r = 0.30$, $P < 0.05$). Re-pressurization via fluid injection into subsurface oil reservoirs during secondary recovery may exert a normalizing effect regardless of geographic location. This is not necessarily because exogenous fluids contain similar organisms but rather due to the fluid chemistry provoking a re-organization of microbial populations in the reservoir, e.g., introducing oxidants into a highly reduced biogeochemical system.

2.3.3 Functional potential of the oil reservoir microbiome

Metagenomes from 25 samples from reservoirs produced via primary and secondary recovery and ranging from 457–3350 mbsf and 26–102°C were assembled for gene prediction to assess the functional potential of the oil reservoir microbiome. NMDS of the relative abundance of 7,046 unique KEGG Orthology (KO) assignments show both distinct and overlapping gene compositions in samples from different studies (**Figure 2.3; Table S2.4**). Unlike taxonomy-based associations of microbiomes with environmental factors (**Figure 2.1, 2.2**), no correlation was observed between reservoir type (i.e., primary vs secondary recovery), temperature, depth or geographic location and the gene composition of a reservoir (**Table S2.4**).

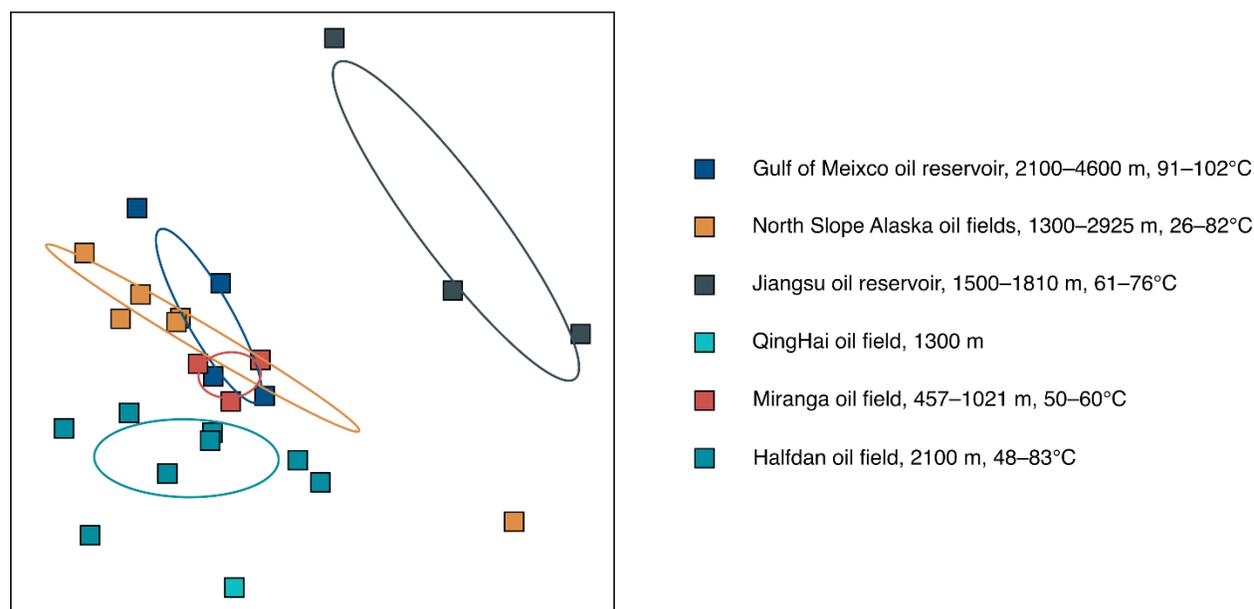


Figure 2.3. Gene composition of samples from different reservoirs. Nonmetric multidimensional scaling of the Bray-Curtis dissimilarity in the relative abundance of unique gene compositions represented as Kegg Ortholog (KO) assignments in samples from different reservoirs (**Table S2.6**). Unlike taxonomy-based associations, no significant correlation was observed between environmental factors (e.g., temperature and depth) and gene composition of the samples (**Table S2.4**). Ellipses indicate one standard deviation to highlight co-occurrence trends.

Reservoir metagenomes were compared to better understand functional potential for carbon acquisition and energy conservation (**Figure 2.4; Table S2.6**). Diverse capabilities for carbohydrate, peptide, and lipid catabolism, as well as carbon fixation, are widespread among the reservoirs. Mixed-acid fermentation appears to be a universal strategy in oil reservoirs. Acetate production from pyruvate was observed in all metagenomes, consistent with acetate being measured in oil field waters (Carothers, Kharaka, 1978; Barth, 1991) and acetogenesis being important in the deep biosphere in general (Lever, 2012). Energy conservation through anaerobic respiration is also prevalent. Genes required for the reduction of sulfate to sulfide (*sat*, *aprAB* and *dsrAB*) were observed in 24 of the 25 metagenomes. Reductases required for nitrate reduction to nitrite, ammonium and nitrogen were detected in 18, 15, and 8 of the metagenomes, respectively. Observations of sulfate and nitrate metabolism do not correlate with primary or secondary recovery

practices, indicating that these respiratory pathways are inherent features of indigenous microbial communities. Sulfide:quinone oxidoreductase (*sqr*), which catalyzes the oxidation of sulfide to elemental sulfur, was detected in all samples. Sulfide oxidation can be coupled to denitrification or dissimilatory reduction of nitrate to ammonia (Hubert and Voordouw, 2007), which are couplings that these metagenomes reveal the oil reservoir microbiome has capacity for. This metabolism also offers a strategy to detoxify sulfide accumulation (Marcia *et al.* 2009), produced by both biogenic and thermochemical sulfate reduction in petroleum reservoirs.

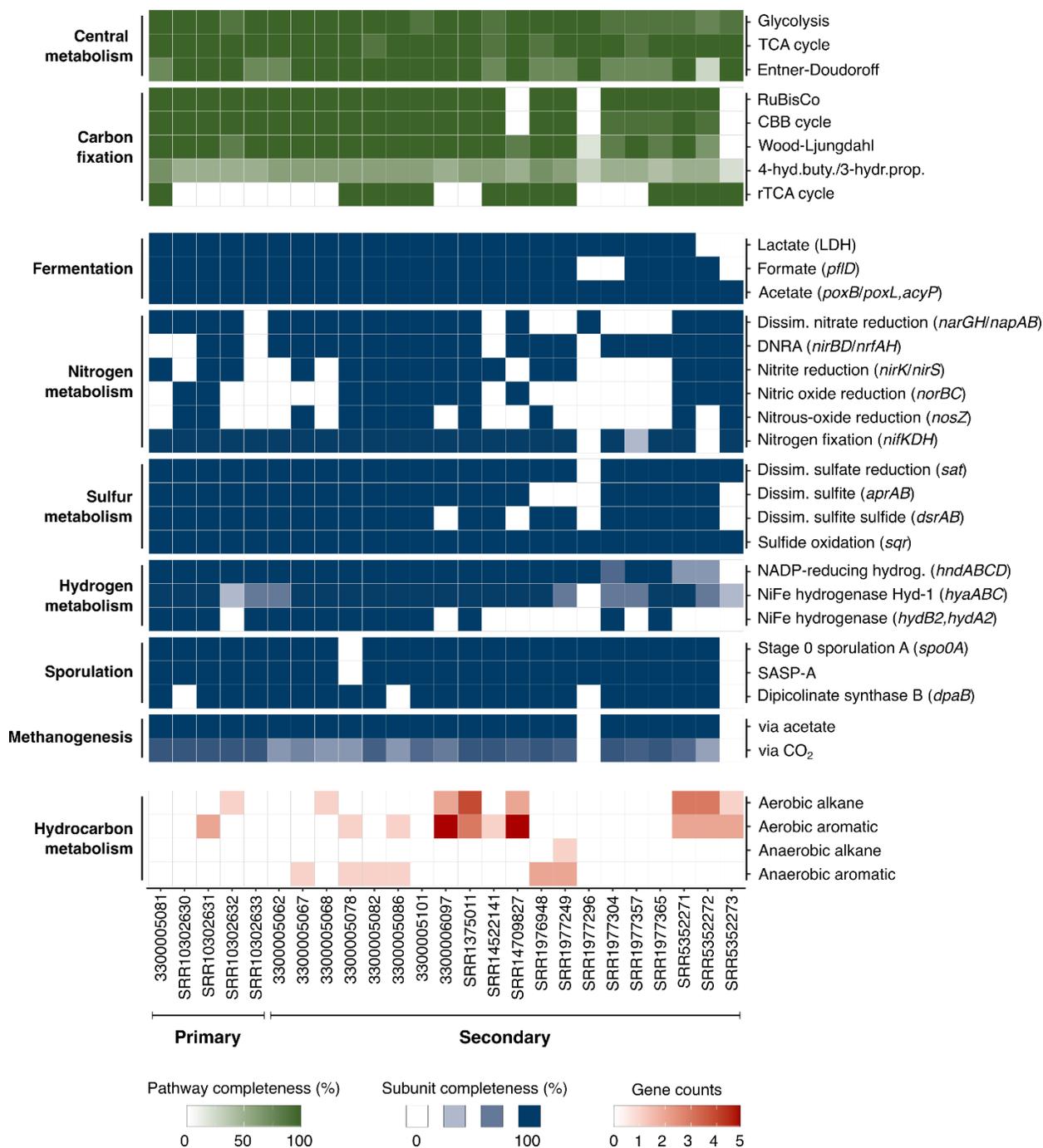


Figure 2.4. Metabolic potential of the petroleum reservoir microbiome. Occurrence of metabolic genes and pathways in metagenomes from microbiomes derived from primary and secondary produced reservoirs. Complete gene lists can be found in *Table S2.6*.

Hydrogen metabolism in oil reservoirs can be linked to both fermentative and respiratory pathways. Membrane-bound Group 1b respiratory H₂-uptake hydrogenases (*hydAB*) and Group 1d oxygen-tolerant NiFe hydrogenase (*hyaABC*) are both observed (**Figure 2.4; Table S2.6**). These enzymes pair hydrogen oxidation with the liberation of electrons for anaerobic respiration (Greening *et al.* 2016). The consistency in the co-occurrence of *hyaABC* with the dissimilatory sulfate reduction gene sulfate adenylyltransferase (*sat*) suggests potential coupling of this form of hydrogen oxidation with sulfate respiration. Genes encoding the Group 3b cytosolic NiFe hydrogenase (*hynABCD*) were present in nearly all metagenomes. This enzyme can work bidirectionally, but generally couples reoxidation of NAD(P)H to the evolution of H₂ produced by fermentation (Greening *et al.* 2016). Together with oxidative hydrogenases, this suggests that both hydrogen production and consumption may be widespread features of oil reservoir biogeochemistry.

Microbial dormancy in the form of endospore formation is considered to be an important process in the deep biosphere (Wörmer *et al.* 2019). Accordingly, key marker genes for sporulation including the *spo0A* master transcriptional response regulator and small, acid-soluble protein (SASP-A) and dipicolinic acid (*dpaB*) synthesis genes are present in these subsurface oil reservoir metagenomes (**Figure 2.4; Table S2.6**). This is consistent with the *Firmicutes* phylum, which known endospore-forming microbes belong to, demonstrating the highest average relative abundance in 16S rRNA gene and metagenome libraries (**Figure 2.1**).

2.3.4. Hydrocarbon biodegradation potential

To identify a potential for microbial degradation of hydrocarbons, functional marker genes encoding enzymes that initiate aerobic or anaerobic hydrocarbon biodegradation by activating either alkane or aromatic compounds were examined using a set of 37 Hidden Markov Models (Khot *et al.* 2022; **Figure 2.4; Table S2.6**). Of the 28 aerobic and 9 anaerobic marker genes tested, 17 were identified in 16 of the 25 metagenomes. Among the genes identified, 14 are associated with the aerobic biodegradation of hydrocarbons. Associated aerobic enzymes initiate metabolism of monoaromatic hydrocarbons (MAH; e.g., benzene, toluene, ethylbenzene, and xylene) via the

introduction of oxygen into the aromatic ring by a multi-component dioxygenase. Also prevalent is the genetic potential for the aerobic oxidation of *n*-alkanes with chain lengths C₁₀–C₁₆ by *alkB* alkane hydroxylase and short- and medium-chain-length *n*-alkanes by the cytochrome P450 CYP153 family alkane hydroxylase. Genes involved in the aerobic degradation of long chain alkanes (e.g., *almA* and *ladA/B*) and polyaromatic hydrocarbons (e.g., *ndoB/C* and *dszC*) were less frequently observed.

A well understood mechanism for the anaerobic degradation of *n*-alkanes and aromatic hydrocarbons is initiated by their addition to fumarate. Sequences encoding the catalytic subunit of benzylsuccinate synthase (*bssA*) were detected in six samples from reservoirs in the North Sea and North Slope Alaska, with the latter also hosting the catalytic subunits of alkylsuccinate synthase (*assA*) and naphthylmethylsuccinate synthase (*nmsA*). Metabolisms encoded by these genes can couple hydrocarbon degradation to the reduction of an electron acceptor such as sulfate, or to fermentative or syntrophic metabolisms that proceed in conjunction with methanogenesis — a process that is understood to be a primary mechanism for hydrocarbon metabolism in reservoirs worldwide over geological timescales (Jones *et al.* 2008; Gray *et al.* 2011). Genes involved in acetoclastic and hydrogenotrophic methanogenesis needed to facilitate this were detected in 23 of the 25 metagenomes. Similarly, the gene encoding the Group 3a: F₄₂₀ hydrogenase, *frhB* (Kegg ortholog KO00441), which directly couples oxidation of H₂ to reduction of F₄₂₀ during methanogenesis (Mills *et al.* 2013), was widespread in 23 of the 25 metagenomes (**Table S2.6**).

2.4 Discussion

2.4.1 Petroleum reservoirs lack a core microbiome

This study reveals that oil reservoirs lack a core microbiome. This assessment is based on defining a core microbiome as requiring the presence of common species in all samples (Shade and Handelsman, 2011). Identifying the occurrence of a core microbiome is important for understanding the ecology of an environment, with prevalent taxa likely having an important role in that ecosystem's functional potential (e.g., Lundberg *et al.* 2012; Ainsworth *et al.* 2015). In a petroleum reservoir context, understanding this potential can inform more rational predictions

about the interplay between physicochemical conditions and biological consequences, thus informing operational decisions. The data presented here suggest that, instead of a core microbiome in oil reservoirs, environmental factors result in niche habitat partitioning. Temperature and depth influence bacterial and archaeal population structure and diversity in petroleum reservoirs (**Figure 2.2**; **Table S2.4**). This is consistent with other deep biosphere settings, and the environmental and energetic challenges of sustaining life in subsurface environments (Jørgensen and D'Hondt, 2006; Heuer *et al.* 2020). Cell numbers in the sedimentary deep biosphere show a log-log decrease with depth (Kallmeyer *et al.* 2012). Petroleum reservoirs are known to harbour cell densities higher than those in surrounding sediments (Bennett *et al.* 2013), likely owing to a rich supply of organic carbon and nutrients (Head *et al.* 2003; Orphan *et al.* 2001; Vigneron *et al.* 2017); however, temperature and depth are shown here to still constrain microbial diversity and corresponding metabolic potential in oil reservoirs.

The lack of a core species-level taxonomy characterising the global oil reservoir microbiome implies a functional redundancy that results in similar biogeochemical processes, such as *in situ* hydrocarbon biodegradation or reservoir souring — phenomena widely observed in petroleum reservoirs around the world (Head *et al.* 2003). The growing number of oil reservoirs being assessed by shotgun metagenomics (not just 16S rRNA gene amplicon sequencing) allows this to be assessed in detail. Despite taxonomic differences between metagenomes from different petroleum reservoirs, the sampled environments share a remarkable number of orthologous genes. Redundant functional potential in different oil reservoirs suggests trait selection in these settings despite the absence of core taxonomic groups. A diversity of genes encoding these functions is indicative of metabolic versatility being an important feature in the oil reservoir microbiome. Genetic capacity for carbohydrate, peptide, and lipid metabolism indicates that degradation of microbial necromass and sedimentary detrital material could be an important process in petroleum reservoirs, or their precursor sediments. This is consistent with the role of residual organic matter recycling in the deep biosphere in general (Lomstein *et al.* 2012; Lloyd *et al.* 2013; Orsi *et al.* 2013). Similarly, the widespread occurrence of genes involved in acetogenesis in these metagenomes is consistent with this being an important process in anaerobic subsurface environments (Lever, 2012). The diversity and prevalence of both respiratory and fermentative pathways indicates these may be co-occurring processes or metabolisms used sequentially by

individual organisms in response to changing conditions. Whereas respiration has favourable energetics relative to fermentation, the limited availability of electron acceptors in the subsurface is likely an important determinant for the predominance of fermentative metabolism, which may additionally reflect a strategy to access energy from more challenging substrates such as hydrocarbons. It is likely that fermentation products contribute substrates for further sulfate reduction or methanogenesis, depending on the prevailing redox conditions and the presence of sulfate. Not surprisingly, genes for sulfate reduction and methanogenesis were widespread in the samples studied here. The potential for respiration may also explain the rapid transition to souring and souring control scenarios following the introduction of sulfate-rich seawater and injected nitrate, respectively.

2.4.2 Secondary oil recovery imposes a selective pressure on the oil reservoir microbiome

The number of reservoir samples assessed in this study enables an empirical assessment of the consequences of secondary oil recovery on the microbiome present in oil reservoirs. It is well known that seawater injection is associated with oil reservoir souring. While the introduction of new organisms (Bell *et al.* 2020) or cooling of the near injection wellbore region to more permissive temperatures (Eden *et al.* 1993) are sometimes cited as contributing factors, the large amount of sulfate present in seawater is understood to be a main driver for this outcome. This has led some oil companies to consider costly strategies for nanofiltration to remove sulfate from injected fluids (Bilstad, 1992). What is less clear is whether or not the sulfate-reducing microorganisms (SRM) responsible are introduced exogenously or are persisting in the reservoir until conditions that favour this metabolic response arise. The former interpretation is challenged by seawater being an oxic environment where SRM may not be plentiful or active (Bell *et al.* 2020), whereas the latter would similarly require SRM to persist in the reservoir in the absence of sulfate — a situation that may be explained by SRM that exhibit fermentative activity in the absence of sulfate (Muyzer and Stams, 2008). Similar questions can be asked regarding nitrate reducing microorganisms (NRM) that are activated by nitrate injection for souring control, which they can achieve by coupling organotrophic metabolism to nitrate reduction to out-compete SRM, or by oxidizing sulfide directly with nitrate (Carlson and Hubert, 2019). Metagenomes from

several different oil reservoirs summarized here reveal that the potential for sulfate reduction, nitrate reduction and sulfide oxidation regardless of whether the microbiome sampled is from a reservoir undergoing primary or secondary recovery. This suggests that the metabolic potential for biogeochemical shifts that accompany operational changes is a pre-existing feature of deep biosphere communities in oil reservoirs. Detecting these genes in metagenomes suggests that they are not necessarily harboured by minor or dormant constituents of pristine oil reservoir microbial communities, and may thus be auxiliary metabolisms in bacteria capable of fermentation in the absence of electron acceptors. The 295 high throughput 16S rRNA gene libraries from oil reservoirs produced by primary and secondary recovery assessed here reveal that overall microbial diversity is lower during secondary recovery. Taken together the meta-analysis here provides strong evidence from oil reservoirs around the world to support an interpretation that secondary oil recovery imposes selective pressure on the microbiome, resulting in an enrichment of community constituents that are well-adapted to the changing conditions.

2.4.3 Is hydrocarbon degradation universal in petroleum reservoirs?

In pristine oil reservoirs, prior to secondary recovery or even reservoir discovery, the absence of oxygen or other electron acceptors dictates that methanogenic conditions should prevail. In this context, fermentation reactions that initiate anaerobic hydrocarbon biodegradation are coupled to methanogenic archaea rapidly consuming acetate, CO₂ and hydrogen to ensure thermodynamic feasibility of syntrophic partnerships (Schink, 1997; Dolfing *et al.* 2008). Recent work demonstrates that similar metabolism (conversion of oil to methane) can also be facilitated by oil reservoir methanogens in the absence of a syntrophic partner (Zhou *et al.* 2021). Genetic potential for acetoclastic and hydrogenotrophic methanogenesis was identified in all of the oil reservoirs examined here, consistent with the understanding that this may be a default biogeochemical regime in pristine petroleum-bearing sediments — a view supported by thermodynamic modelling (Dolfing *et al.* 2008), carbon dioxide and methane stable carbon isotopic measurements (Jones *et al.* 2008), and radiotracer experiments (Mayumi *et al.* 2011) that predict hydrogenotrophic CO₂ reduction to be the primary route for crude oil hydrocarbon biodegradation in oil reservoirs.

Assuming syntrophic degradation of hydrocarbons, anaerobic hydrocarbon activation genes are expected to be correspondingly widespread; however, *assA*, *bssA* and *nmsA* were only detected in two of the nine reservoirs examined here. The discrepancy between widespread genomic potential for methanogenesis with geochemical evidence of biogenic methane production in reservoirs (Horstad and Larter, 1997; Larter and di Primio, 2005), and observations here of a limited occurrence of anaerobic hydrocarbon degradation genes, potentially indicates that other mechanisms for anaerobic hydrocarbon biodegradation (e.g., Lack and Fuchs, 1994; Heider *et al.* 2016; Rabus *et al.* 2016), or as yet unknown forms of hydrocarbon activation by microbiomes inhabiting anoxic petroleum reservoirs. For example, analysis of variants of glyceryl-radical enzymes proposed to mediate anaerobic alkane biodegradation via addition to fumarate has revealed a clade of genes encoding alkylsuccinate synthase (*assA*) that is divergent from canonical *assA* found in *Proteobacteria*. These *assA*-like pyruvate formate-lyase (*pflD*) genes are found in taxa from different petroleum reservoirs such as *Archaeoglobus fulgidus* (Khelifi *et al.* 2014) and ^U*Petromonas tenebris* (Christman *et al.* 2020).

The prevalence of aerobic genes in the metagenomes analysed here could be considered surprising. Despite possible influx of oxygen-bearing meteoric waters (Palmer, 1993; Parkes and Maxwell, 1993), anoxic conditions are understood to prevail in subsurface petroleum reservoirs (Head *et al.* 2003). Interestingly, metagenomes containing aerobic genes were not restricted to samples associated with secondary oil recovery; long chain alkane and monoaromatic dioxygenase enzymes were detected in samples from an offshore reservoir in the Gulf of Mexico (Christman *et al.* 2020) and an onshore reservoir in northeastern Brazil (Sierra-Garcia *et al.* 2020) that had not undergone secondary recovery prior to sampling for DNA analysis. These results are consistent with observations of high proportions of genes for enzymes involved in aerobic hydrocarbon metabolism in various anaerobic hydrocarbon resource environments (An *et al.* 2013), and may arise from enrichment of initially minor groups of organisms capable of aerobic respiration coming into contact with air during sampling and sample transport. This is substantiated by genes encoding alkane hydroxylase (*alkB*) and monoaromatic dioxygenase (MAH) showing high sequence identity (based on BLAST against the non-redundant nucleotide database) with *Marinobacter* spp. present in low relative abundance *in situ*. Enrichment of aerobic bacteria *ex situ* during sampling is ecologically similar to enrichment of sulfate- or nitrate-reducing bacteria *in situ* in response to

reservoir conditions changing, and again points more generally to an inactive or latent metabolic potential inherent to oil reservoir microbial communities.

2.4.4 Provenance of the oil reservoir microbiome

A biogeographic pattern indicated by increasing community variance with increasing distance between reservoirs raises questions about the provenance of microbes and the establishment of the global reservoir microbiome. Heuer *et al.* (2020) postulated that vegetative cells and endospores that are deposited in surface sediments and undergo burial over geological timescales can then be revived under the right selective conditions. Coupled with this, limited dispersal vectors into the deep biosphere have been proposed (Stetter *et al.* 1993). This evidence combines to suggest that microbial communities in oil reservoirs are inherited from populations that are present during the proximal deposition of sediments that eventually form reservoirs. Environmental selection during burial likely contributes to the differences observed here as a function of geography between oil reservoirs from different parts of the world.

2.5 Methods

2.5.1 Data acquisition

To determine the community composition and functional potential of the petroleum reservoir microbiome, published studies with microbial assessments of reservoir derived samples were identified. To be included in this meta-analysis, studies had to (i) have samples originating from a petroleum reservoir, (ii) have sample collection points at the well head or directly associated infrastructure, and (iii) have publicly available sequences. Metadata including sample collection point and type, reservoir name (an unconnected reservoir assumed if different location or oil field reported), reservoir (and/or collection point) location, temperature and depth, and the use of primary or secondary (e.g., water or steam injection) production was compiled directly from the respective publications (***Table S2.1***). In instances where geographic coordinates of the reservoir

were not provided, latitude and longitude have been approximated based on the location of the reported oil field.

2.5.2 16S rRNA gene sequence processing

Raw high throughput 16S rRNA gene amplicon sequence data was obtained from the National Center for Biotechnology Information's (NCBI) Sequence Read Archive (SRA; Leinonen *et al.* 2011) by compiling sequence accession lists and implementing the *prefetch* and *fastq-dump* commands from the SRA Toolkit. The dataset comprised a total of 53 million raw sequence reads. Initial sequence processing was performed using the open source VSEARCH version 2.11.1 (Rognes *et al.* 2016). In instances where paired-end sequence data was available from the SRA, read pairs were merged based on a minimum read overlap length of 10 base pairs (bp) and a maximum permitted mismatch of 20% of the length of the overlap. Merged reads were filtered with a maximum expected error of 0.5 for all bases in the read, and minimum and maximum read lengths of 150 and 500 bp (informed by the reported amplicon size), respectively. Identical reads were dereplicated and annotated with their associated total abundance for each sample, prior to *de novo* chimera detection and removal using default parameters. Accounts of the reads retained at each processing stage are provided in **Table S2.7**. The quality-controlled sequences are available in the figshare online open access repository at <https://doi.org/10.6084/m9.figshare.c.5801735.v1>.

The studies compiled in this meta-analysis were published over a period during which sequencing and accepted standards have been evolving. Despite applying quality control steps designed to include a larger range of studies and enable a broader global meta-analysis, in some instances as few as 1.7% of the raw reads in a given library and 8.4% of raw reads in an entire study could be retained. This highlights how data quality can be an important and sometimes unresolved issue in microbial ecology investigations.

Quality-filtered 16S rRNA gene amplicon sequences from each high throughput library were concatenated into a single file for further processing. To account for the different hypervariable regions amplified within studies and between studies, required for microbial community comparisons, identical sequences were considered and compared to each other as members of

phylotypes representing nested taxonomic classifications for a sequence or group of sequences. For example, sequences in different libraries that sequenced different hypervariable regions were grouped into a common phylotype if their taxonomy was consistent, allowing all libraries to be compared regardless of PCR methodology. Taxonomy was assigned using Mothur version 1.41.3 (Schloss *et al.* 2009) and Ribosomal Database Project's k-mer-based naïve Bayesian classifier (Wang *et al.* 2007) with the SILVA SSU Ref NR version 138 database (Quast *et al.* 2013). A bootstrap cut-off value was set to 80% to return only taxonomies above this confidence threshold. A custom R script written in base R version 3.6.1 (R Core Team, 2014) parsed the output taxonomic assignments into a sample-by-phylotype table that accounted for earlier dereplication of the sequences. Sequence processing included the removal of singleton phylotypes, phylotypes with no domain level taxonomic classification (likely artifacts of sequence preprocessing), and phylotypes classified as *Vertebrata*, *Mitochondria* and *Chloroplast*. High throughput sequencing libraries with <1,000 reads after quality filtering were removed from the dataset. To reduce the loss of data, subsampling was only used for alpha diversity calculations and for comparisons of the effect of variable library sizes (results are reported alongside non-subsampled libraries in specific instances as described herein). Subsampling was performed without replacement to 1,000 reads using the *phyloseq* R package (McMurdie and Holmes, 2013). For all other analyses, subsampling was not performed.

The majority of oil reservoir amplicon sequencing studies to date have employed clone libraries (e.g., Voordouw *et al.* 1996; Gittel *et al.* 2009; Hubert *et al.* 2012). Amplicon sequencing reads derived from cloning were downloaded from NCBI's GenBank database (Sayers *et al.* 2019) and analysed separately from the high throughput 16S rRNA gene amplicon sequence data described above, in order to account for the varying sampling depth and data outputs. No sequence processing steps were implemented prior to the analysis of these 'low throughput' clone libraries except for using the reverse complement sequence for taxonomic classification in instances where this produced a better classification. Analysed clone sequences are compiled in the figshare online open access repository at <https://doi.org/10.6084/m9.figshare.c.5801735.v1>. Taxonomy was again assigned using Mothur version 1.41.3 (Schloss *et al.* 2009) and Ribosomal Database Project's k-mer-based naïve Bayesian classifier (Wang *et al.* 2007) with the SILVA SSU Ref NR version 138 database (Quast *et al.* 2013). A bootstrap cut-off value was set to 80%. In this way, phylotype-

level community structures between high throughput and low throughput amplicon libraries were comparable. Counts of the respective classifications were used to assess prevalence across the dataset at various taxonomic resolutions.

2.5.3 Metagenomic sequence processing

The metagenomic dataset comprised sequences from 25 metagenome libraries – 16 raw libraries accounting for 457 million raw sequence reads were downloaded from NCBI's SRA (Leinonen *et al.* 2011), in addition to 9.6 million assembled reads from nine libraries downloaded from The Integrated Microbial Genomes & Microbiomes system (IMG/M; Chen *et al.* 2021) when raw sequence data was not publicly available (**Table S2.7**). Raw, unassembled reads were quality-controlled by trimming technical sequences (primers and adapters) and low-quality additional bases, and filtering artifacts (phiX), low-quality reads and contaminated reads using BBDuk (BBTools suite, <http://jgi.doe.gov/data-and-tools/bbtools>). Ribosomal rRNA genes in quality-controlled reads were reconstructed and classified using phyloFlash (Gruber-Vodicka *et al.* 2017) with mapping against the SILVA SSU Ref NR version 138 database (Quast *et al.* 2013). Ribosomal rRNA genes in assembled reads were identified using rRNAFinder (<https://github.com/xiaoli-dong/rRNAFinder>) after contigs <500 bp were removed, and taxonomy was assigned using Mothur version 1.41.3 (Schloss *et al.* 2009) with the SILVA SSU Ref NR version 138 database (Quast *et al.* 2013). Detailed accounts of the reads retained during processing, and assembly statistics are provided in **Table S2.7**. Known challenges associated with 16S rRNA gene assembly from metagenomes (Yuan *et al.* 2015) likely account for the lower gene counts in prior assembled data as compared to unassembled data. For the quality-controlled reads not already assembled, the metagenome libraries were assembled using MEGAHIT version 1.2.2 (Li *et al.* 2015) with default parameters, and contigs <500 bp removed. Protein coding genes in each assembly were predicted using Prodigal version 2.6.3 (Hyatt *et al.* 2010). Predicted gene sequences were compiled and made available in the figshare online open access repository at <https://doi.org/10.6084/m9.figshare.c.5801735.v1>. Predicted genes were annotated with KEGG Orthology (KO) using GhostKOALA (Kanehisa *et al.* 2016). Metabolic pathways from KO assignments were reconstructed using KEGG Decoder (Graham *et al.* 2018). Sporulation genes

were identified by manual searching for KO numbers of specific marker genes. Genes involved in the aerobic and anaerobic activation of hydrocarbon compounds, and therefore indicators of hydrocarbon biodegradation functions, were annotated using CANT-HYD (Khot *et al.* 2022).

2.5.4 Statistical analysis and data visualization

Statistical analyses and visualization were performed using base R version 3.6.1 (R Core Team, 2017) or the specific R packages as indicated below. Richness and alpha diversity metrics were calculated for subsampled high throughput 16S rRNA gene sequence libraries using *phyloseq* (McMurdie and Holmes, 2013). Spearman's rank correlation coefficient assessed correlations between alpha diversity indices and two-group variables. Kruskal-Wallis tests assessed alpha diversity variance with respect to multi-group and non-numeric variables.

Prior to beta diversity calculations the non-subsampled and subsampled high throughput 16S rRNA gene amplicon datasets were split according to the implied and probable use of universal or archaeal primers; libraries comprising $\geq 75\%$ archaea (based on relative sequence abundance) were defined here as archaeal libraries and the remainder defined as universal libraries indicating non-taxon specific 16S rRNA gene amplification. Non-metric multidimensional scaling (NMDS) of Bray-Curtis dissimilarity between high throughput 16S rRNA gene-based sample communities, as well as between metagenome KO assignments, was calculated from relative abundance data using *phyloseq* (McMurdie and Holmes, 2013) and visualized using *ggplot2* (Wickham and Chang, 2015). The statistical difference between sample community dissimilarities (Bray-Curtis) in relation to non-numeric variables was assessed using Permutational Multivariate Analysis of Variance (PERMANOVA) tests in *vegan* (Oksanen *et al.* 2007). Biogeographical patterns were assessed using Mantel tests of sample community dissimilarities (Bray-Curtis) and the Euclidean distance between environmental parameters or haversine (geographic) distance between locations. Haversine distance was calculated using the *geosphere* package (Hijmans, 2016) and Mantel tests performed using *vegan* (Oksanen *et al.* 2007).

Microbial indicator sequence analysis, designed to test the association of a single taxon with an environment through multilevel pattern analysis, was used to identify phylotypes that best

represent a specific environmental condition based on both phylotype presence/absence and relative abundance patterns. Indicator phylotypes were calculated using the *multipatt* function of the *indicspecies* package in R, employing a point-biserial correlation index (De Cáceres *et al.* 2010). Tests were performed on non-subsampled libraries (see **Table S2.1** for temperature subsets).

Chapter 3

Geological processes mediate a microbial dispersal loop in the deep biosphere

Daniel A. Gittins¹, Pierre-Arnaud Desiagne², Natasha Morrison³, Jayne E. Rattray¹, Sriyak Bhatnagar¹, Anirban Chakraborty¹, Jackie Zorz¹, Carmen Li¹, Oliver Horanszky¹, Margaret A. Cramm¹, Francesco Bisiach¹, Robbie Bennett², Jamie Webb⁴, Adam MacDonald³, Martin Fowler⁴, D. Calvin Campbell², and Casey R. J. Hubert¹

¹Geomicrobiology Group, Department of Biological Sciences, University of Calgary; Calgary, Canada.

²Natural Resources Canada, Geological Survey of Canada-Atlantic; Dartmouth, Canada.

³Department of Natural Resources and Renewables, Government of Nova Scotia; Halifax, Canada.

⁴Applied Petroleum Technology; Calgary, Canada.

Submitted to *Science Advances* (abn348)

3.1 Abstract

The deep biosphere is the largest microbial habitat on Earth and features abundant bacterial endospores. Whereas dormancy and survival at theoretical energy minima are hallmarks of microbial physiology in the subsurface, ecological processes like dispersal and selection in the deep biosphere remain poorly understood. We investigated the biogeography of dispersing bacteria in the deep sea where upward hydrocarbon seepage was confirmed by acoustic imagery and geochemistry. Thermophilic endospores in the permanently cold seabed correlated with underlying seep conduits reveal geofluid-facilitated cell migration pathways originating in deep petroleum-bearing sediments. Endospore genomes highlight adaptations to life in anoxic petroleum systems and bear close resemblance to oil reservoir microbiomes globally. Upon transport out of the subsurface, viable thermophilic endospores re-enter the geosphere by sediment burial, enabling germination and environmental selection at depth where new petroleum systems establish. This geological microbial dispersal loop circulates living biomass in and out of the deep biosphere.

3.2 Introduction

Identifying dispersal vectors that distribute microorganisms throughout the biosphere is critical to understanding biogeography and Earth system functioning. Whereas distributions of animals and plants have been studied since the time of Darwin (Darwin, 1859), related ecological processes are harder to elucidate in the microbial realm where the effects of dispersal and environmental selection must be disentangled (Baas Becking, 1934; Hanson *et al.* 2012; Ward *et al.* 2021). Dormant populations of microbes retain viability while enduring inhospitable conditions in relation to growth requirements, allowing dispersal to be studied directly without the influence of conflating factors like selection, drift, or mutation (Hanson *et al.* 2012). Bacterial endospores are equipped to survive dispersal over long distances and timescales (Setlow *et al.* 2006), with reports of viable spores ~2.5 km beneath the seafloor (Fang *et al.* 2017) suggesting dispersal journeys lasting millions of years. This points to a genetically and functionally diverse seed bank of microbes that can be revived if subsurface environmental conditions select for their traits (Lennon and Jones, 2011; Mestre and Höfer, 2021).

The marine subsurface biosphere contains an estimated 10^{29} microbial cells contributing up to 2% of the total living biomass on Earth (Bar-On *et al.* 2018). Whereas deep biosphere populations exhibit exponentially decreasing numbers with depth (Heuer *et al.* 2020), endospores experience less pronounced declines and are estimated to outnumber vegetative cells in deeper marine sediments (Wörmer *et al.* 2019; Lomstein *et al.* 2012). Measurements of the endospore-specific biomarker dipicolinic acid indicate remarkably high numbers of endospores in warm buried sediments, with depth profiles revealing that temperature influences sporulation and germination (Heuer *et al.* 2020). This is consistent with the prevalence of endospore forming *Firmicutes* in microbiome surveys of hot oil reservoirs from around the world (Hubert *et al.* 2012) where they actively contribute to biogeochemical cycling. In the deep biosphere, these petroleum systems represent energy rich oases (Orphan *et al.* 2001; Vigneron *et al.* 2017) that select for thermophilic organotrophs, including anaerobic hydrocarbon degraders, sulfate reducers and other groups. Accordingly, cell densities in oil reservoirs can be an order of magnitude higher than those in surrounding sediments (Bennett *et al.* 2013).

Hydrocarbon seepage up and out of deep petroleum systems is widespread in the ocean (Judd, 2003). Studies of thermophilic spores in cold surface sediments globally (Müller *et al.* 2014; Hanson *et al.* 2012) have invoked warm-to-cold dispersal routes like hydrocarbon seeps to explain these observations (Hubert *et al.* 2009). In the Gulf of Mexico where cold seeps are common (MacDonald *et al.* 2015), spore-forming thermophiles in surface sediments are correlated with the presence of migrated liquid hydrocarbons (Chakraborty *et al.* 2018). While buoyant gas migration mediates upward dispersal of cells in the top few centimeters (Chakraborty *et al.* 2020), whether or not viable cells from deeper and hotter subsurface layers can be similarly circulated over greater depths and timescales remains hypothetical. Here we compare deep-sea sediments from the NW Atlantic Ocean (**Figure S3.1**) using geophysics, hydrocarbon geochemistry, spore germination dynamics and genomics to demonstrate dispersal of viable cells throughout the marine subsurface. This geologically mediated microbial dispersal loop transports living biomass via upward seepage and downward burial, and represents a previously overlooked mechanism for ecological maintenance and preservation of life in the subsurface biosphere.

3.3 Results

Structural geology indicative of deep subsurface to surface geofluid conduits with the potential to transport microbial cells was determined by multichannel 2D and 3D seismic reflection surveys along the NW Atlantic Scotian Slope. Geophysical surveys covered $\sim 70,000$ km² and obtained $\sim 10,000$ m of subsurface stratigraphic imagery in up to 3,400 m water depth (**Figure 3.1a**). Global estimates of the extent of the deep biosphere predict that marine sediments in this study area harbour at least 10^{25} microbial cells (Kallmeyer *et al.* 2012), and even higher numbers of bacterial endospores (Wörmer *et al.* 2019). Detecting seabed hydrocarbon seepage in deep-sea settings like this is very challenging (Abrams, 2020), thus a multi-disciplinary strategy was employed. Co-location at the seabed of the up-dip limit of deep-seated faults and seismic reflection anomalies considered to be direct hydrocarbon indicators were used to identify potential subsurface seep networks (Nanda, 2016) (**Figure 3.1b**). These large-scale geophysical survey results were refined through high-resolution seismic reflection, side-scan sonar and multibeam bathymetry. Morphological features included a mounded structure with high backscatter intensity intersected

by an elongated fracture-like depression (*Figure 3.1c*) and a circular pockmark, suggesting the presence of a seep-like structure. Immediately beneath these features, high-resolution subsurface seismic profiling revealed a localized acoustic blanking zone (*Figure 3.1d*) suggesting the presence of gas (Judd and Hovland, 2007) and a hydrocarbon migration pathway through subsurface sediments. On this basis, sites were prioritized for sediment coring to enable further geochemistry, microbiology, and genomics.

Locations showing seismic evidence of migrated hydrocarbons originating from a deep subsurface source (*Table S3.1*) were examined in greater detail by comparing chemical differences in hydrocarbon signals from 14 different sites (*Figure 3.2a; Table S3.2*). Higher concentrations of thermogenic C_2 - nC_4 compounds, elevated wet gas ratios and heavy $\delta^{13}C$ values for methane (-42 to -52‰) in interstitial gas, coupled with liquid hydrocarbon extracts featuring elevated nC_{17}/nC_{27} ratios and a lack of odd-over-even alkane distributions in the nC_{23-33} range, provided clear evidence of thermogenic hydrocarbons at two sites. Thermogenic hydrocarbons are formed at depth, making these sites candidates for geofluid dispersal of deep biosphere microbes. A deep origin of these hydrocarbons is further supported by higher proportions of thermally derived diasteranes relative to regular steranes (% 27 $d\beta$ S) and more thermally mature terpane distributions (C_{30} $\alpha\beta$ relative to C_{31} $\alpha\beta$ 22R hopane) in these two cores. At the 12 other sites, thermogenic hydrocarbon signals were either inconclusive ($n=4$) or not detected ($n=8$).

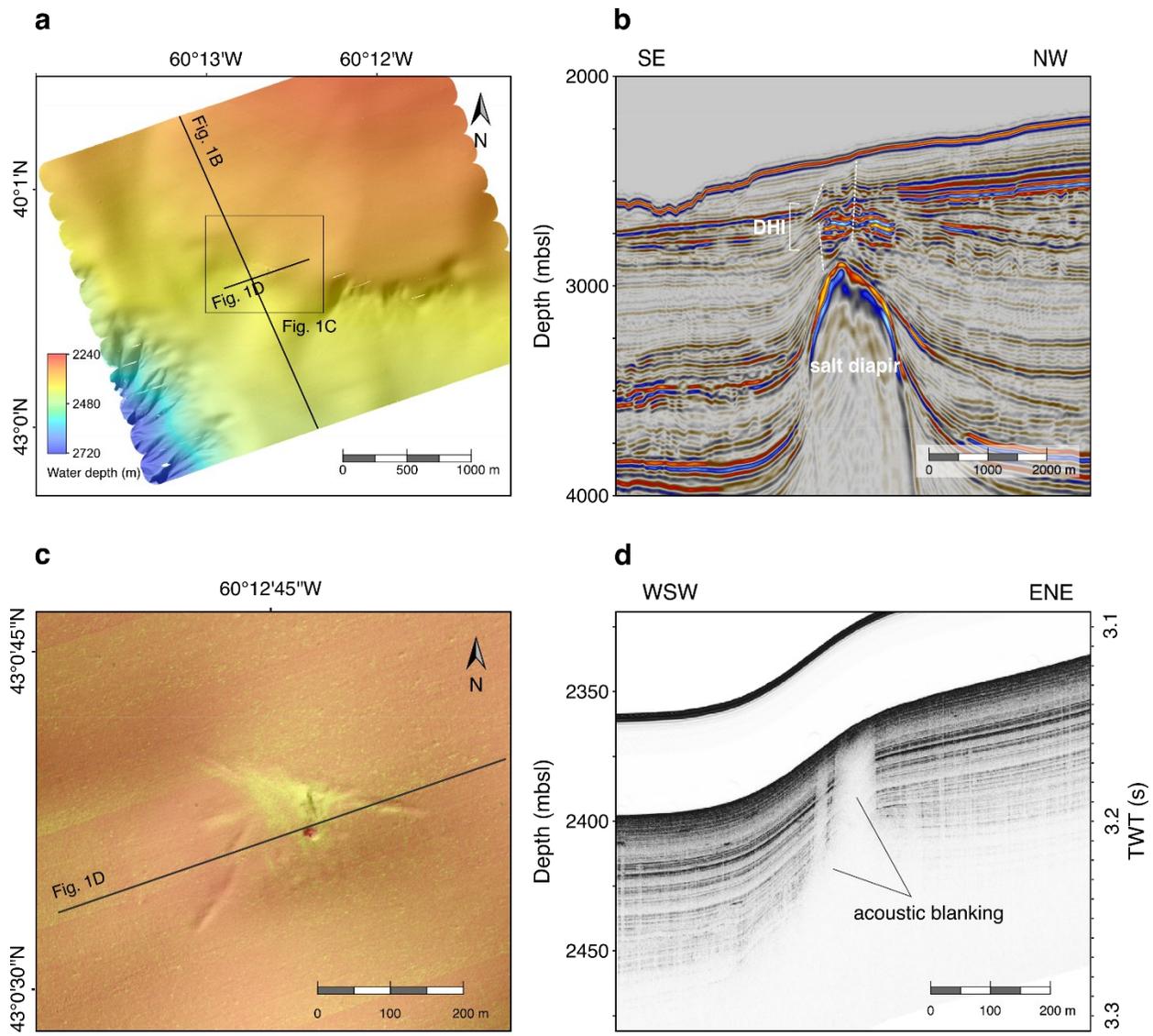


Figure 3.1. Deep subsurface to surface geofluid migration. **a**, Seafloor surface map derived from autonomous underwater vehicle (AUV) multibeam bathymetric sonar data. **b**, 3D seismic cross section showing a buried salt diapir, the location and direction of crestal faults (white dashed lines), including an interval with direct hydrocarbon indicators (DHI). **c**, Combined mosaic of side-scan sonar data and shaded relief bathymetry of the area surrounding a seep structure, indicating a pockmark feature as well as a small mound morphology. High backscatter intensity, related to distinctive properties of near-surface sediment, is shown in light-yellowish tones. **d**, AUV-based sub-bottom profiling showing localized acoustic blanking under the seep structure, indicative of upward migration of fluid originating deeper in the sediment.

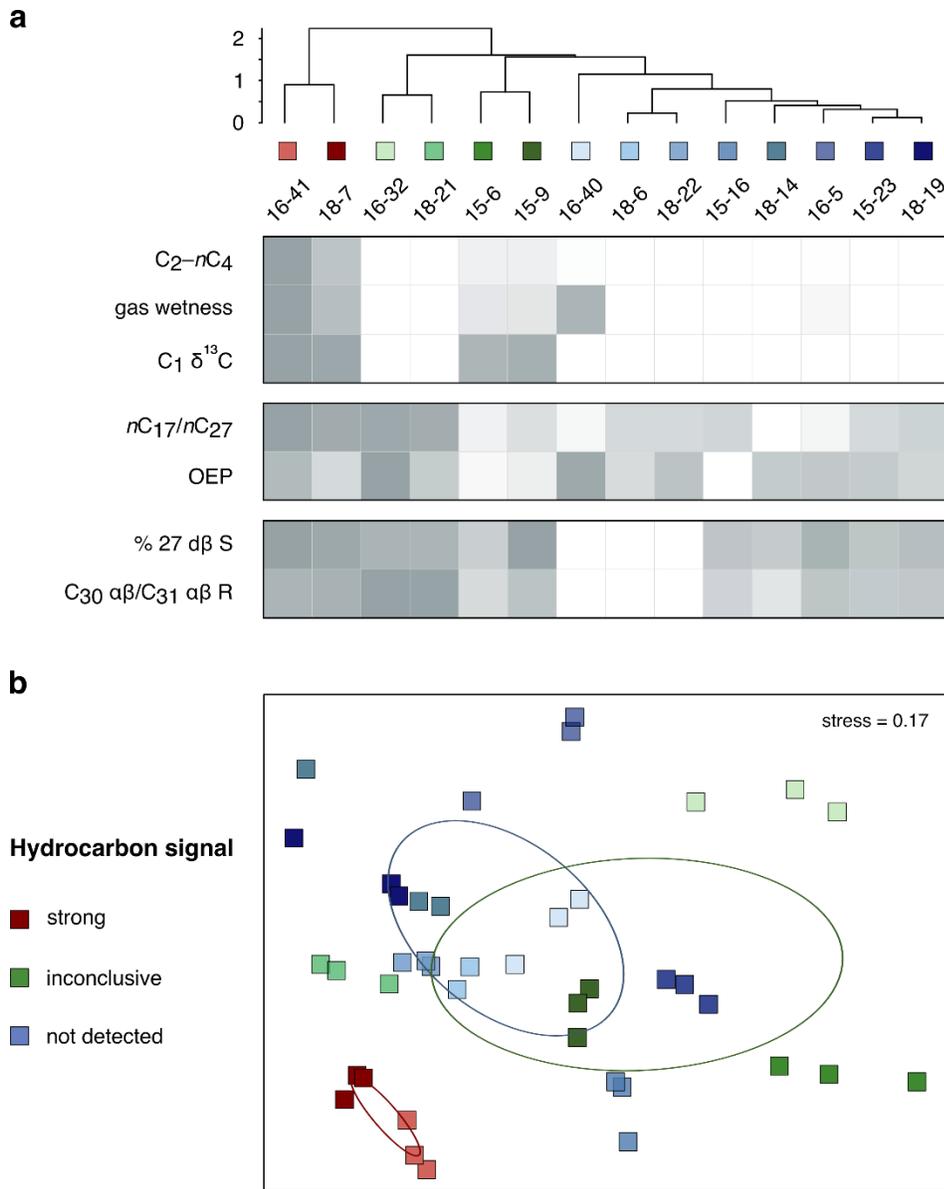


Figure 3.2. Hydrocarbon geochemistry and microbial community variance between seabed sampling sites. a, Gas ($\sum C_2-nC_4$, gas wetness, and $C_1 \delta^{13}C$), liquid hydrocarbon extract (nC_{17}/nC_{27} and odd-over-even predominance) and biomarker ($\% 27 d\beta S$ and $C_{30} \alpha\beta/C_{31} \alpha\beta 22R$) measurements to assess the presence of thermogenic hydrocarbons. Each parameter is scaled between 0 and 1 as shown in the heatmap. Cores are represented by maximum values in instances where multiple depths from a core were tested (see **Table S3.2**). Hierarchical clustering of the Euclidean distance between scaled values highlights groups of sites where the evidence for the presence of thermogenic hydrocarbons is strong (red), inconclusive (green) or not detected (blue). **b,** The same three groupings are reflected in Bray-Curtis dissimilarity in microbial community composition after sediment incubation at $50^\circ C$ (ellipses indicate standard deviations of weighted averaged means of within-group distances for each of the three groups). Results from $40^\circ C$ and $60^\circ C$ incubations showed similar clustering

(see **Figure S3.2**). This demonstrates that sites with strong thermogenic hydrocarbon signals have distinct microbial populations after high temperature incubation (see also **Table S3.4**), relative to the sites without thermogenic hydrocarbon signals.

To compare thermophilic spore-forming bacterial populations in cores with and without evidence of thermogenic hydrocarbons, endospore germination and thermophile enrichment was stimulated in high temperature anoxic incubations (40–60°C following pasteurization at 80°C; **Figure S3.3**). Assessing microbial community composition by 16S rRNA gene profiling of incubated surface sediments from all 14 locations showed divergent profiles of enriched communities of thermophiles in hydrocarbon-positive locations (**Figure 3.2b**; **Table S3.3**). These differences were further assessed by statistical comparisons that revealed 42 unique amplicon sequence variants (ASVs), all belonging to spore-forming *Firmicutes* (Galperin, 2013), correlated with upward seepage of thermogenic hydrocarbons (*indicspecies*, $P < 0.05$; **Table S3.5**). Putative fermentative organotrophs such as *Paramaledivibacter* and *Caminicella*, as well as sulfate-reducing *Desulfotomaculales* and *Candidatus Desulforudis*, showed strong hydrocarbon associations (**Figure 3.3a**). None of these groups were detected by applying the same DNA sequencing method to unincubated sediment (**Table S3.3**), likely owing both to the corresponding endospores *in situ* having low relative abundance (de Rezende *et al.* 2017) and the multi-layered endospore coat not yielding to the standard cell lysis protocol used here for DNA extraction from vegetative cells (Wunderlin *et al.* 2014).

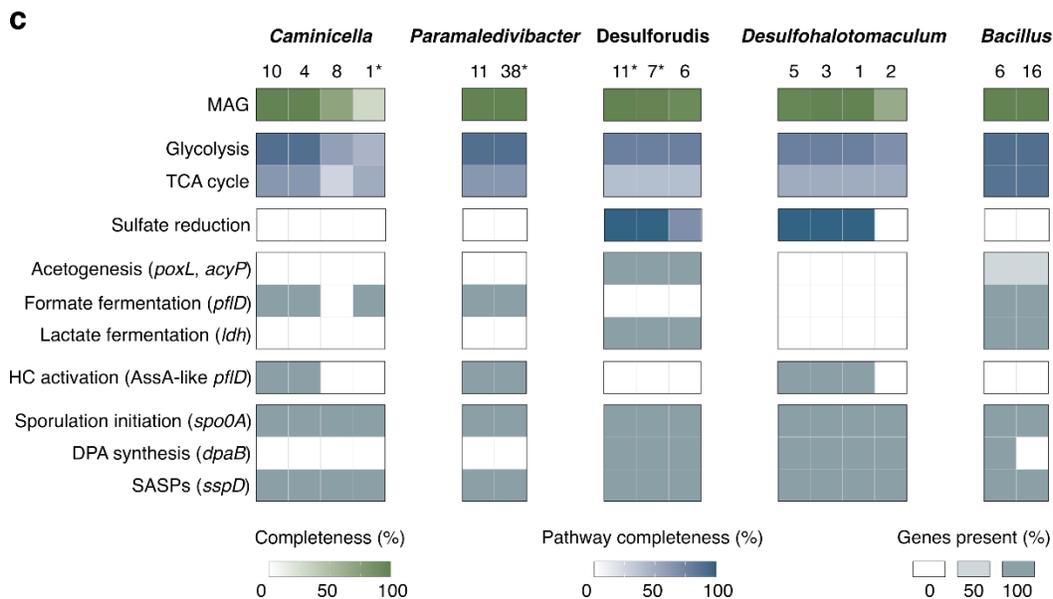
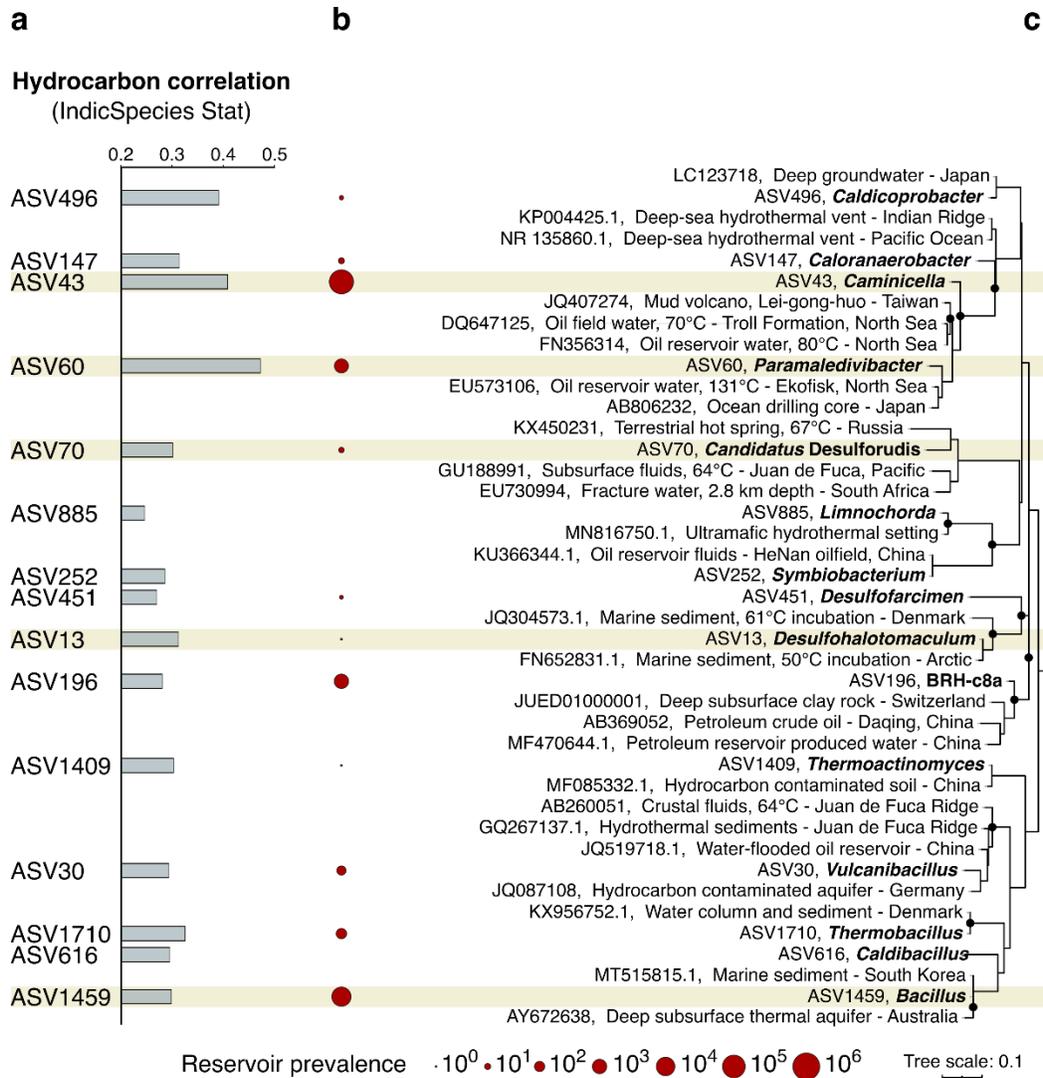


Figure 3.3. Subsurface oil reservoir origins of seep-associated thermophiles. **a**, Correlation of thermophilic spore-forming bacterial amplicon sequence variants (ASVs) with thermogenic hydrocarbons (*IndicSpecies*, $P < 0.05$). Highest ranking ASVs from each of 15 different genera are shown (representing 42 hydrocarbon-correlated ASVs in total). **b**, Prevalence of these genera in 59 oil reservoir microbiome assessments (11 million 16S rRNA gene sequences in total). **c**, Maximum likelihood phylogeny showing the 15 representative hydrocarbon-correlated ASVs and close relatives (see **Figure S3.4** for all 42 indicator ASVs). Black circles at the branch nodes indicate >80% bootstrap support (1,000 re-samplings), and the scale bar indicates 10% sequence divergence as inferred from PhyML. **d**, Metagenome-assembled genomes (MAGs) matching ASVs of interest (corresponding to brown bands in **a–c**) were assessed for anaerobic alkane degradation, sporulation, and other metabolic features characteristic of deep biosphere populations living in oil reservoirs. Metagenome bin numbers are indicated below the corresponding taxonomy (see **Table S3.8** for pathway definitions and additional gene annotations). Asterisk (*) denote bins with exact 16S rRNA gene sequence matches to ASVs.

To assess the prevalence of these bacteria in deep petroleum systems, we curated a dataset of bacterial and archaeal 16S rRNA gene sequences from 59 different oil reservoir microbiomes from around the world (**Table S3.6**; *Chapter 2*). Seep-associated thermophilic endospore lineages identified in the cold deep-sea sediments analysed here are found in high proportions in subsurface petroleum systems, especially *Caminiaceae* and *Desulfotomaculales* which each make up 2–3% of the global oil reservoir microbiome (**Figure 3.3b**). ASV assessment at finer taxonomic resolution confirms close genetic relatedness between thermophilic endospores in Scotian Slope sediments and bacteria found in different subsurface oil reservoirs (**Figure 3.3c**; **Figure S3.4**). Metagenomic analysis of heated sediments revealed that these dormant spores encode the potential for anaerobic hydrocarbon biodegradation, favouring their selection and growth in deep petroleum-bearing sediments (**Figure 3.3d**). Metagenome-assembled genomes (MAGs) of *Caminiella*, *Paramaledivibacter*, *Desulfohalotomaculum* and *Bacillus* included contigs with rRNA sequences matching the indicator ASVs (**Figure 3.3**; **Table S3.7**), and encode glycyl-radical enzymes related to alkylsuccinate synthases proposed to mediate anaerobic alkane biodegradation via addition to fumarate (Jones *et al.* 2008; Gray *et al.* 2009). Based on newly developed Hidden Markov Models for annotating alkylsuccinate synthases (Khot *et al.* 2022), putative *assA* gene sequences in thermophilic spores are shown here to diverge from canonical *assA* found in mesophilic *Proteobacteria* (**Figure S3.5**). This divergent clade includes thermophiles from hot oil

reservoirs such as ^U*Petromonas tenebris* (Christman *et al.* 2020) and *Archaeoglobus fulgidus* (Khelifi *et al.* 2014). Sediment incubation MAGs also contain sporulation genes (**Figure 3.3d**) including the *spo0A* master transcriptional response regulator (Hoch, 1993) as well as genes for synthesizing α/β -type small, acid-soluble proteins (e.g., *sspD*) and dipicolinic acid (e.g., *dpaB*) involved in DNA protection (Setlow *et al.* 2006) (for a full list of sporulation genes see **Table S3.9**). Other genomic features relevant to life in deep hot oil reservoirs are also summarized in **Figure 3.4d**.

3.4 Discussion

Maintenance of dormancy has been proposed as a necessary pre-requisite for microbial taxa to exhibit biogeographic patterns over large distances and timescales (Mestre and Höfer, 2021). The thermophilic endospores revealed here as originating from deep petroleum-bearing sediments exemplify large-scale biogeography by connecting anaerobic hydrocarbon biodegradation and other microbial activities in these subsurface habitats with intervening periods of large-scale migration in a dormant, sporulated state. Recurrent cyclical dispersal facilitates this scenario, consistent with a ‘microbial conveyor belt’ framework for understanding biogeography where the same environment features as both the origin and eventual destination for migrating populations (Mestre and Höfer, 2021). Transport facilitated by vertical migration of hydrocarbons through subsurface sediments at rates of 0.1 m d^{-1} (Arp, 1992; Rice, 2022) would require on the order of 50 years for dispersal from hydrocarbon reservoirs in 2,000 mbsf Cretaceous sediments (Deptuck and Kendall, 2020) up to the surface and into the marine bottom water (**Figure 3.4a**; **Figure S3.6**). Further transport of bacterial spores via benthic currents on the order of 0.9 m s^{-1} (Hannah *et al.* 2001) would precede their eventual re-entry into the seabed (**Figure 3.4b**). In the cold surface sediment of the Scotian Slope, thermophilic spores were detected in all sampling locations, including those lacking geochemical evidence of hydrocarbon seepage (**Table S3.3**). Dipicolinic acid concentrations within the top few metres in non-seep sites demonstrate constant deposition and burial of endospores (**Figure 3.4c**), with numbers in this study area similar to or exceeding the seabed global average (Wörmer *et al.* 2019). High temperature anoxic incubation of sediments from the deepest cores (down to 9 mbsf) causes thermophilic spores to germinate, demonstrating

that viability is maintained during burial (*Figure S3.7*). Survival during burial (Morono *et al.* 2020) enables sedimentation to disperse dormant populations down into deep, hot environments that can support high rates of metabolism by thermophiles (Beulig *et al.* 2022). Burial to the equivalent reservoir depth of present day Cretaceous reservoirs at rates consistent with deposition in the study area (0.15 mm a^{-1}) would require 13.3 Ma. This process completes a cyclical subsurface migration via passive dispersal of viable cells looping out of and back into the deep biosphere (*Figure 3.4*). In buried sediments that get filled with petroleum from deeper organic-rich source rocks to create oil reservoirs (Magoon and Dow, 1994), dispersing populations are further replenished.

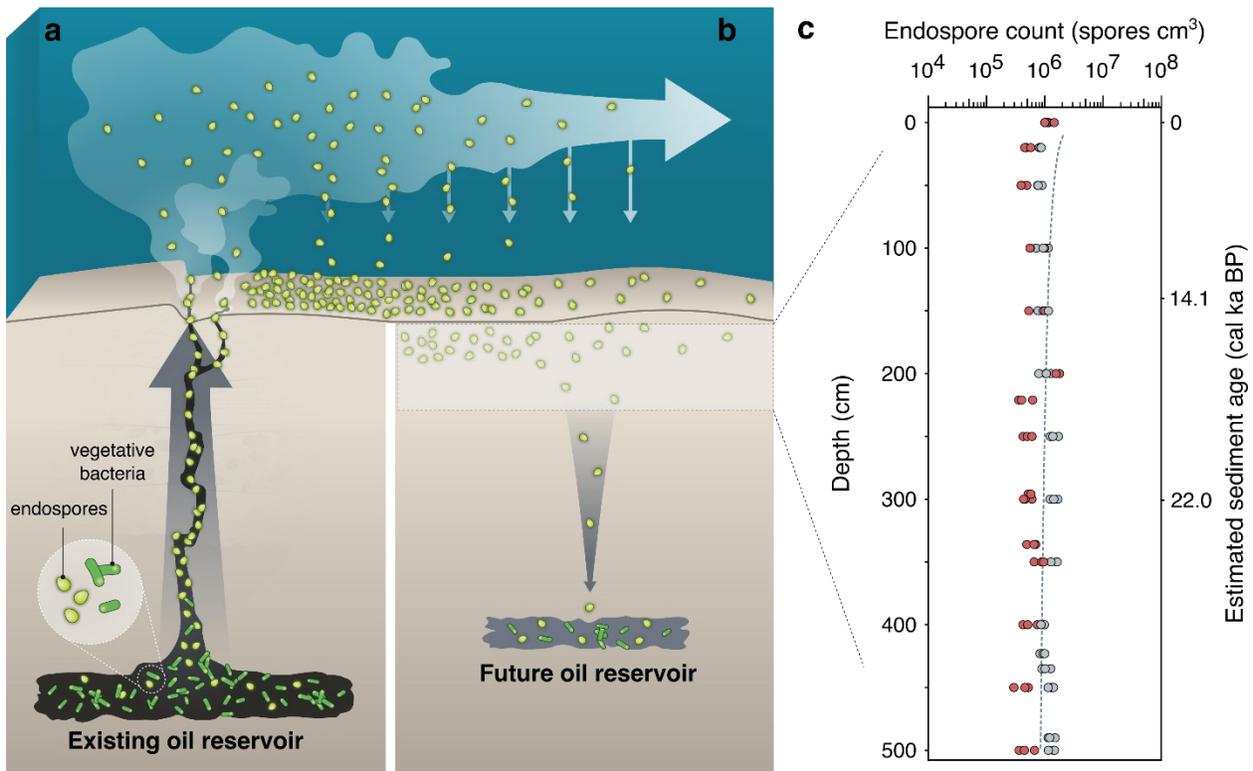


Figure 3.4. Subsurface microbial dispersal loop mediated by seepage, sedimentation, dormancy and environmental selection. **a**, Endospore-forming microbial populations that are active and abundant in deep petroleum systems get dispersed upwards as dormant spores by hydrocarbon seepage along geological conduits. Endospores entering the deep-sea are dispersed laterally by bottom water currents as part of the marine rare biosphere. **b**, Endospores get deposited on the seabed and undergo burial. **c**, Dipicolinic acid was quantified in the upper metres of the seabed in two Scotian Slope

sediment cores where hydrocarbons were not detected. Based on the average sedimentation rate in this region, this corresponds to a burial rate on the order of 10^8 spores $\text{m}^{-2} \text{y}^{-1}$. The dashed regression line reflects the global average estimated for endospores in the marine subseafloor biosphere (Wörmer *et al.* 2019). Survival of individual endospores over long burial time-scales (Fang *et al.* 2017; de Rezende *et al.* 2013; Morono *et al.* 2020) enables environmental selection (i.e., germination, activity and population replenishment) by suitable substrates and favorable conditions that exist in petroleum-bearing sediments such as oil reservoirs (Magoon and Dow, 1994). The sequence shown in **a** and **b** completes a subsurface ‘microbial dispersal loop’ that incorporates cell transport and biogeochemical cycling in Earth’s deep biosphere.

Marine sediments contain 12–45% of Earth’s microbial biomass and are central to the planet’s biogeochemical cycling (Morono *et al.* 2020). This is especially true in subsurface petroleum systems that control and are controlled by deep biosphere communities (Jones *et al.* 2008; Wilhelms *et al.* 2001). Despite the importance of these processes, research on the subsurface microbiome rarely focuses on ecological factors like dispersal and selection, preventing a more complete understanding of this vast ecosystem (Biddle *et al.* 2012). The results presented here demonstrate that geological processes of geofluid flow and sedimentation connect deep petroleum systems with the ocean, where thermophiles contribute to the marine rare biosphere seed bank (Lynch and Neufeld, 2015) and can travel great distances (Müller *et al.* 2014). While some dispersing thermophiles may encounter and colonize suitable new environments, e.g., at warm mid-ocean ridges, many of the endospores ejected from the deep biosphere will re-enter the seabed as sedimentary particles, thus mediating extensive microbial dispersal throughout the subsurface. This circulation of living biomass is uniquely characterized by intermittent episodes of microbial activity in petroleum-bearing sediments interspersed by long intervals of passive dispersal — an ecological sequence that is difficult to clearly delineate in other environmental settings (Biddle *et al.* 2012). By connecting the physical and physiological factors that govern survival and evolution in the deep biosphere, this subsurface microbial dispersal loop between major Earth compartments showcases the geosphere as a model system for understanding the interplay between dispersal and selection in the biosphere at large spatial and temporal scales.

3.5 Methods

3.5.1 Subsurface imaging via seismic data acquisition and processing

Multiple two- and three-dimensional multi-channel seismic surveys performed here for the identification of seafloor hydrocarbon seeps relied on an earlier regional 28,000 km² 2D seismic survey. The earlier survey was shot in a 6 km grid, acquiring 14 seconds of data with 80–106-fold and a 2 millisecond sampling interval. The 1998 vintage used was processed to pre-stack time migrated data. 2D seismic survey interpretations were refined using the Shelburne 3D Wide Azimuth Seismic survey. This survey was acquired over 12,000 km² in the deep-water Shelburne sub-basin at 6.25 x 50 m bin spacing with a fold of 100. This vintage of data utilized both full 3D Anisotropic Kirchhoff pre-stack time migration (PSTM) and full volume anisotropic Kirchhoff pre-stack depth migration (PSDM) with vertical transverse isotropy. PSTM had a processed bin size of 12.5 x 25 m, while PSDM had an output bin size of 25 x 25 m. Data were interpreted using the Petrel E&P Software Platform (Schlumberger Limited).

High-resolution seismic reflection profiles (data not shown) were used to investigate the subsurface stratigraphy in the vicinity of seabed fluid flow prospects to inform autonomous underwater vehicle (AUV) survey and piston coring locations. Profiles were collected during three expeditions between 2015 and 2018 onboard the CCGS *Hudson* (Campbell and MacDonald, 2016; Campbell, 2019; Campbell and Normandeau, 2019) using a Hunttec single-channel Deep Tow Seismic (DTS) sparker system. Tow depth was ~100 m beneath the sea surface with the source fired at a moving time interval between 1 and 3 seconds. The peak frequency for the Hunttec DTS sparker is approximately 1,500 Hz and spans from 500–2,500 Hz. Raw sparker data was processed using the VISTA Desktop Seismic Data Processing Software (Schlumberger Limited) and included Ormsby band-pass filtering, scaling correction, automatic gain control and trace mixing.

An AUV was deployed from the vessel *Pacific Constructor* to collect high-resolution geophysical data over a 2.5 x 2.5 km area at the location of cores 16-41 and 18-7 in August 2020. The HUGIN 6000 AUV (Kongsberg Maritime) was steered approximately 40 m above the seafloor for multibeam bathymetric, side-scan sonar and sub-bottom profiling data collection using a Kongsberg EM 2040 multibeam echosounder and a EdgeTech 2205 sonar system, respectively.

EM2040 Multibeam bathymetric data was acquired at a frequency of 400 kHz, with a continuous waveform (CW) pulse and synchronized with Doppler velocity log. Multibeam bathymetric data was processed using Caris and Eiva suite. Side Scan Sonar data was acquired at a frequency of 230 kHz and post-processing of the data was completed in Sonarwiz. The sub-bottom profiler was operated over the frequency range of 1–9 kHz with a 20-millisecond pulse. The high-resolution seismic data was integrated and analyzed using IHS Kingdom Suite (IHS Markit Ltd). Acoustic travel times for high resolution sub-bottom profiler lines were converted into depths by using an average seismic velocity of 1,500 m.s⁻¹.

3.5.2 Marine sediment sampling

Seabed surface sediments in 2,000 to 3,400 m water depth were collected by piston and gravity coring from different locations on the Scotian Slope, offshore Nova Scotia, Canada (**Figure S3.1**; **Table S3.1**) during May-June expeditions aboard the CCGS *Hudson* in 2015, 2016 and 2018 (Campbell and MacDonald, 2016; Campbell, 2019; Campbell and Normandeau, 2019). Piston cores, trigger weight cores (i.e., smaller cores that release the head weight of the piston core as they enter the seabed) and gravity cores ranged from 0.18 to 8.34 m in length. Upon recovery, cores were split longitudinally onboard the ship. Sediment intervals from the base of the core (5–10 cm) were transferred to gas-tight IsoJars[®] (Isotech Laboratories Inc., USA), immediately flushed with nitrogen, and stored at -20°C prior to interstitial gas analysis. Similar intervals from variable depths along the cores, selected based on indications of visible hydrocarbon staining or odour, fluorescence, or sandy lithology, were stored in aluminum foil at -20°C for eventual hydrocarbon analysis, and in sterile Whirl-Pak[®] bags or glass jars at 4°C for eventual high temperature endospore germination experiments. Sediment intervals at the top of the core (either 0–10 or 0–20 cm below seabed) were similarly transferred to sterile Whirl-Pak[®] bags or sterile glass jars. Sedimentation rates of 0.1 and 0.2 mm a⁻¹ were determined in core 18-19 for sections between 0–141 cm and 141–302 cm, respectively, using a chronostratigraphic framework based on the correlation with sediment cores from the Scotian slope (Jenner *et al.* in prep) and the identification at 141 cm of the brick red mud regional marker bed “d” dated at 14 ka BP (Piper and Skene, 1998).

3.5.3 Hydrocarbon geochemical analysis

Interstitial gas analysis was performed on aliquots of IsoJar[®] (Isotech Laboratories Inc., USA) headspace transferred into Exetainers[®] (Labco Limited, UK). Sample volumes of 1 mL were injected into an Agilent 7890 RGA Gas Chromatograph (Agilent Technologies, USA). A flame ionisation detector determined C₁–C₅ hydrocarbon gas concentrations that were used to calculate gas wetness ($(\sum C_{2-n}C_4) / (\sum C_{1-n}C_4) * 100$). Carbon isotopic composition ($\delta^{13}C$) of hydrocarbon gas components was determined by gas chromatography combustion isotope ratio mass spectrometry; headspace aliquots were analyzed on a Trace 1310 Gas Chromatograph (Thermo Fisher Scientific, USA) interfaced to a Delta V Isotope Ratio Mass Spectrometer (Thermo Fisher Scientific, USA).

Sediments were analyzed for hydrocarbon biomarkers in subsamples where sufficient extract yields were recovered. Accordingly, no extract yield, or insufficient yields to determine biomarker concentrations, were considered indicative of the absence of hydrocarbon seepage. Organic matter was extracted from sediment by adding dichloromethane with 7% (v/v) methanol, mixing the solution in an ultrasonic bath for 15 min and then leaving at room temperature for 24 h. Extractable organic matter (EOM) was evaporated to dryness and weighed. Asphaltenes were removed by pentane addition in excess (40 times the volume of EOM), storage for 12 h, and centrifugation. Gas chromatography analysis of the EOM was performed on an Agilent 7890A Gas Chromatograph (Agilent Technologies, USA). Saturate and aromatic hydrocarbon fractions showing possible evidence of deeply sourced thermogenic hydrocarbons were analyzed further using a Micromass ProSpec Gas Chromatography-Mass Spectrometer (Waters Corporation, USA). Geochemical analyses were performed by Applied Petroleum Technology, Norway, to the standards used in industrial hydrocarbon assessments. Geochemistry data was collectively interpreted for evidence of thermogenic hydrocarbons likely derived from subsurface hydrocarbon seeps. Hierarchical clustering (complete linkage clustering based on Euclidean distance) of geochemical measurements scaled between 0 and 1 over the range of values (0 representing weakest thermogenic signal and 1 representing the strongest thermogenic signal) was used to further assess and visualise groups of sites with similar geochemical signatures (*Figure 3.2a*).

3.5.4 Sediment incubation at elevated temperatures

Sediments were investigated for the germination and growth of dormant bacterial endospores. Following homogenizing by stirring within the sample container, up to 100 g of sediment was transferred into separate 250 mL serum bottles that were sealed with butyl rubber stoppers (Chemglass Life Sciences, Canada) and the headspace exchanged with N₂:CO₂ (90:10%). Sediment slurries were prepared in a 1:2 (w/w) ratio with sterile, anoxic, synthetic seawater medium (Isaksen *et al.* 1994) containing 20 mM sulfate and amended with acetate, butyrate, formate, lactate, propionate, and succinate (5 mM each for surface sediments and either 5 mM or 1 mM each for deeper sediments). The same medium was used with all sediment samples. Master slurries for a given sediment sample were subdivided into replicate, sterile, anoxic 50 mL serum bottles sealed with butyl rubber stoppers. Slurries were pasteurized at 80°C for 1.5 h to kill vegetative cells and select for heat-resistant endospores. Triplicate pasteurized slurries were immediately incubated at 40, 50 or 60°C for up to 56 days to promote germination and growth of thermophilic endospore-forming bacteria. Subsamples (2 mL) were periodically removed using sterile N₂:CO₂-flushed syringes and stored at -20°C for molecular analysis.

3.5.5 16S rRNA gene amplicon sequencing

Genomic DNA was extracted from triplicate slurries subsampled immediately before incubation (i.e., post-pasteurization), and periodically during the incubation, using the DNeasy PowerLyzer PowerSoil Kit (Qiagen, USA). Extractions were performed on 300 µL of slurry according to the manufacturer's protocol, except for inclusion of a 10 min incubation at 70°C immediately after the addition of Solution C1 to enhance cell lysis. Extraction blanks (Milli-Q water) were processed in parallel. DNA was quantified using the Qubit dsDNA High Sensitivity assay kit on a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Canada). The V3 and V4 hypervariable regions of the 16S rRNA gene were amplified in triplicate PCR reactions per extraction using the primer pair SD-Bact-341-bS17/SD-Bact-785-aA21 (Klindworth *et al.* 2013) modified with Illumina MiSeq overhang adapters. This primer pair has been shown to preferentially amplify bacterial 16S rRNA genes, and provide excellent overall coverage in detecting bacterial diversity (Klindworth *et al.*

2013). PCR reactions included 30 cycles and were performed in triplicate. All DNA extraction blanks and PCR reagent blanks were confirmed for negative amplification using agarose gel electrophoresis. Triplicate PCR products were pooled, purified using a NucleoMag NGS Clean-up and Size Select kit (Macherey-Nagel Inc., USA) and indexed. Sizes of indexed amplicons were verified using the High Sensitivity DNA kit on an Agilent 2100 Bioanalyzer system (Agilent Technologies, Canada). Indexed amplicons were pooled in equimolar amounts and sequenced on an in-house Illumina MiSeq benchtop sequencer (Illumina Inc., USA) using Illumina's v3 600-cycle reagent kit to obtain 300 bp paired-end reads.

3.5.6 16S rRNA gene amplicon sequence processing

A total of 20,589,990 raw paired-end reads were generated across six separate runs on an Illumina MiSeq sequencer. Primers were trimmed using Cutadapt version 2.7 (Martin, 2011) prior to amplicon sequence variant (ASV) inference using DADA2 version 1.16 (Callahan *et al.* 2016) in base R version 3.6.1 (R Core Team, 2013). Forward and reverse read pairs were trimmed to a run-specific length defined by a minimum quality score of 25. Read pairs were filtered allowing no ambiguous bases and requiring each read to have less than two expected errors, and PhiX sequences removed. Reads were dereplicated providing unique sequences with their corresponding abundance. Error rates were estimated from sequence composition and quality by applying a core denoising algorithm for each sequencing run to account for run-to-run variability. Unique ASVs were inferred independently from the forward and reverse reads of each sample, using the run-specific error rates, and then pairs were merged if they overlapped with no mismatches. Chimeras were identified and removed, then an additional length trimming step removed sequence variants shorter than 400 nucleotides and larger than 435 nucleotides. A total of 32,018 ASVs were resolved from 11,355,683 quality-controlled reads. Taxonomy was assigned using the Ribosomal Database Project's k-mer-based naïve Bayesian classifier with the DADA2-formatted Silva database version 138 (Quast *et al.* 2013). Reads were randomly subsampled without replacement to the smallest library size ($n=4,635$) using the *phyloseq* R package (McMurdie and Holmes, 2013) prior to comparative analysis.

3.5.7 Metagenome sequencing

Genomic DNA extracted from four separate sediment slurries after 56 days of incubation was used for metagenomic sequencing. Library preparation and sequencing was conducted at the Center for Health Genomics and Informatics in the Cumming School of Medicine, University of Calgary. DNA was sheared using a Covaris S2 ultrasonicator (Covaris, USA), and fragment libraries prepared using a NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs, USA). Metagenomic libraries were sequenced on the Illumina NovaSeq platform (Illumina Inc., USA) using an S4 flow cell with Illumina 300 cycle (2×150 bp) V1.5 sequencing kit.

3.5.8 Metagenome sequence processing

A total of 65,786,766 raw reads from four metagenomic libraries were quality-controlled by trimming technical sequences (primers and adapters) and low-quality additional bases, and filtering artifacts (phiX), low-quality reads and contaminated reads using BBDuk (BBTools suite, <http://jgi.doe.gov/data-and-tools/bbtools>). Trimmed and filtered reads from each metagenome were assembled separately, as well as co-assembled, using MEGAHIT version 1.2.2 (Li *et al.* 2015) using default parameters and with <500 bp contigs removed. Binning of the four assemblies and one co-assembly was performed using MetaBAT 2 version 2.12.1 (Kang *et al.* 2019). Contamination and completeness of the resulting MAGs were estimated using CheckM version 1.0.11 (Parks *et al.* 2015) with the lineage-specific workflow. Ribosomal rRNA genes were identified in unbinned reads using phyloFlash (Gruber-Vodicka *et al.* 2017) and in binned reads using rRNAFinder implemented in MetaErg version 1.2.0 (Dong and Strous, 2019). Protein coding genes were predicted and annotated against curated protein sequence databases (Pfam, TIGRFAM, and Swiss-Prot) using MetaErg version 1.2.0 (Dong and Strous, 2019). Metabolic pathways were identified using KEGG Decoder (Graham *et al.* 2018) to parse genes annotated with KEGG Orthology using BlastKOALA (Kanehisa *et al.* 2016). Hydrocarbon degradation genes were additionally annotated using CANT-HYD (Khot *et al.* 2022) following gene predictions made using Prodigal version 2.6.3 (Hyatt *et al.* 2010). MAGs were classified with GTDB-Tk version 1.3.0 (Chaumeil *et al.* 2019) and by alignment with Silva database version 138 (Quast *et al.* 2013)

using mothur version 1.39.5 (Schloss *et al.* 2009) in instances where 16S rRNA gene was recovered by rRNAFinder (Dong and Strous, 2019).

MAGs for the seep-associated taxa were identified by alignment of predicted 16S rRNA gene sequences recovered from bins with the seep indicator ASV sequences highlighted by *IndicSpecies* (see below). An alignment identity of 100% across the full length of the amplicon was required to confirm association. In instances where a V3-V4 overlapping 16S rRNA gene sequence was not recovered in the MAG, taxonomic classification of the partial 16S rRNA gene, the MAG (GTDB-Tk version 1.3.0; Chaumeil *et al.* 2019), or the sample by phyloFlash (Gruber-Vodicka *et al.* 2017) was used to identify possible associations to seep-associated taxa. If the most abundant ASV in an unrarefied 16S rRNA gene amplicon library, with the same taxonomic classification as the recovered 16S rRNA gene or MAG, corresponded to the most abundant ASV in that sample, a probabilistic association was assumed and the MAG was retained for further analysis. Replicate MAGs were identified from cluster groups based on metagenome distance estimation using a rapid primary algorithm (Mash) and average nucleotide identity (ANI) using dRep version 2.3.2 (Olm *et al.* 2017) and included in the analysis.

3.5.9 Analysis of global oil reservoir microbiome sequences

Raw high-throughput sequence data, totalling 53,019,792 reads from ten separate studies, was obtained from the National Center for Biotechnology Information's (NCBI) Sequence Read Archive (SRA; Leinonen *et al.* 2011) by compiling sequence accession lists and using the SRA Toolkit. Initial sequence data processing was performed using VSEARCH version 2.11.1 (Rognes *et al.* 2016). If necessary, paired-end sequence files were merged based on a minimum overlap length of 10 base pairs (bp) and a maximum permitted mismatch of 20% of the length of the overlap. Merged reads were filtered with a maximum expected error of 0.5 for all bases in the read, and minimum and maximum read lengths of 150 and 500 bp, respectively. Identical reads were dereplicated and annotated with their associated total abundance for each sample, prior to *de novo* chimera detection. Re-replication resulted in 10,857,433 quality-controlled reads. In addition to these amplicons generated by high-throughput sequencing platforms, 2,850 near full length

amplicon sequences from 49 separate clone library and/or cultivation-based studies were downloaded from the NCBI's GenBank database using published accession numbers. Taxonomy was assigned to the combined 10,860,283 sequences using the Ribosomal Database Project's k-mer-based naïve Bayesian classifier with the Silva database version 138 (McMurdie and Holmes, 2013). Genus-level taxonomic comparisons were made to ASVs derived from marine sediment incubation experiments, since not all of the compiled oil reservoir studies sequenced the same V3–V4 hypervariable region of the 16S rRNA gene that heated sediment amplicon libraries targeted.

3.5.10 Statistical analysis and data visualization

Statistical analyses and visualization were performed using base R version 3.6.1 (R Core Team, 2013), or the specific R packages described below. Non-metric multidimensional scaling (NMDS) of Bray-Curtis dissimilarity was calculated using the *metaMDS* function of the *vegan* package (Oksanen *et al.* 2007) in R and visualized using the *ggplot2* package (Wickham, 2009). Analysis of similarity (ANOSIM) tests measured significant differences between sediment communities and were performed using the *anosim* function of the *vegan* package (Oksanen *et al.* 2007).

Microbial indicator sequence analysis, designed to test the association of a single ASV with an environment through multilevel pattern analysis, was used to identify sequences that best represent specific sediments or groups of sediments under variable test conditions based on both ASV presence/absence and relative abundance patterns. Indicator ASVs were calculated using the *multipatt* function of the *indicspecies* package in R, employing a point-biserial correlation index (De Cáceres *et al.* 2010). Tests were performed on amplicon libraries constructed after 28 and 56 days of high temperature incubation, omitting pre-incubation (day 0) libraries as representing samples prior to endospore enrichment. Among the 32,018 ASVs, only those present in >1% relative abundance in at least one sample across the entire dataset were included in the analysis. The strength of the association is represented by the *IndicSpecies* Stat value (plotted in **Figure 3a**). Only observations with $P < 0.05$ were considered statistically significant and reported.

3.5.11 Phylogenetic analysis

ASVs associated with thermogenic hydrocarbons, together with their five most closely related sequences from Genbank (determined by BLAST searches), were aligned using the web-based multiple sequence aligner SINA (Pruesse *et al.* 2012). Aligned sequences were imported into the ARB-SILVA 138 SSU Ref NR 99 database (Quast *et al.* 2013) and visualized using the open-source ARB software package (Ludwig *et al.* 2004). A maximum likelihood (PhyML) tree was calculated with near full length (>1,300 bases) bacterial 16S rRNA gene reference sequences as well as those from closest cultured isolates. In total, 172 sequences were used to calculate phylogeny (bootstrapped with 1,000 re-samplings), accounting for 1,006 alignment positions specified based on positional variability and termini filters for bacteria. Using the ARB Parsimony tool, ASV and Genbank sequences were added to the newly calculated tree using positional variability filters covering the length of the representative sequences for each sequence without changing the overall tree topology (**Figure S3.4**). Trees were annotated using iTOL version 5.5 (Letunic and Bork, 2019).

3.5.12 Dipicolinic acid (DPA) measurement

Sediment samples from cores 18-14 and 18-19 that showed no geochemical evidence of thermogenic hydrocarbons were prepared in triplicate using the methods described in Lomstein and Jørgensen (2012) and Rattray *et al.* (2021). To extract DPA, 0.1 g of freeze-dried sediment was hydrolysed by addition of 6M HCl and heating at 95°C for 4 hours, before quenching on ice to stop hydrolysis. The hydrolysate was freeze dried, reconstituted in Milli-Q water, frozen and freeze dried again. Samples were then dissolved in 1M sodium acetate and aluminium chloride was added. Sediment extracts were filtered (0.2 µm) and mixed with terbium (Tb³⁺) prepared in 1M sodium acetate. DPA was separated and eluted using gradient chromatography over a Kinetex 2.6 µm EVO C18 100Å LC column (150 x 4.5 mm; Phenomenex, USA) fitted with a guard column. Solvent A was 1M sodium acetate amended with 1M acetic acid to pH 5.6 and solvent B was 80% methanol: 20% water pumped with a Thermo RS3000 pump (Thermo Scientific Dionex, USA). The sample injection volume was 50 µl and the total run time was 10 min (including

flushing). Detection was performed using a Thermo FLD-3000RS fluorescence detector (Thermo Scientific Dionex, USA) set at excitation wavelength 270 nm and emission 545 nm. To determine DPA concentrations under the limit of detection, samples were analysed using standard addition (Ratray *et al.* 2021). For this, a known concentration of DPA standard /Tb³⁺ sodium acetate was sequentially added to the sediment exact and analysed. Concentrations were calculated using methods described in Lomstein and Jørgensen (2012). DPA concentrations were converted into endospore abundances by assuming 2.24 femtomole DPA per endospore as determined previously (Fichtel *et al.* 2009) and used in other studies (Heuer *et al.* 2020; Wörmer *et al.* 2019; Ratray *et al.* 2021).

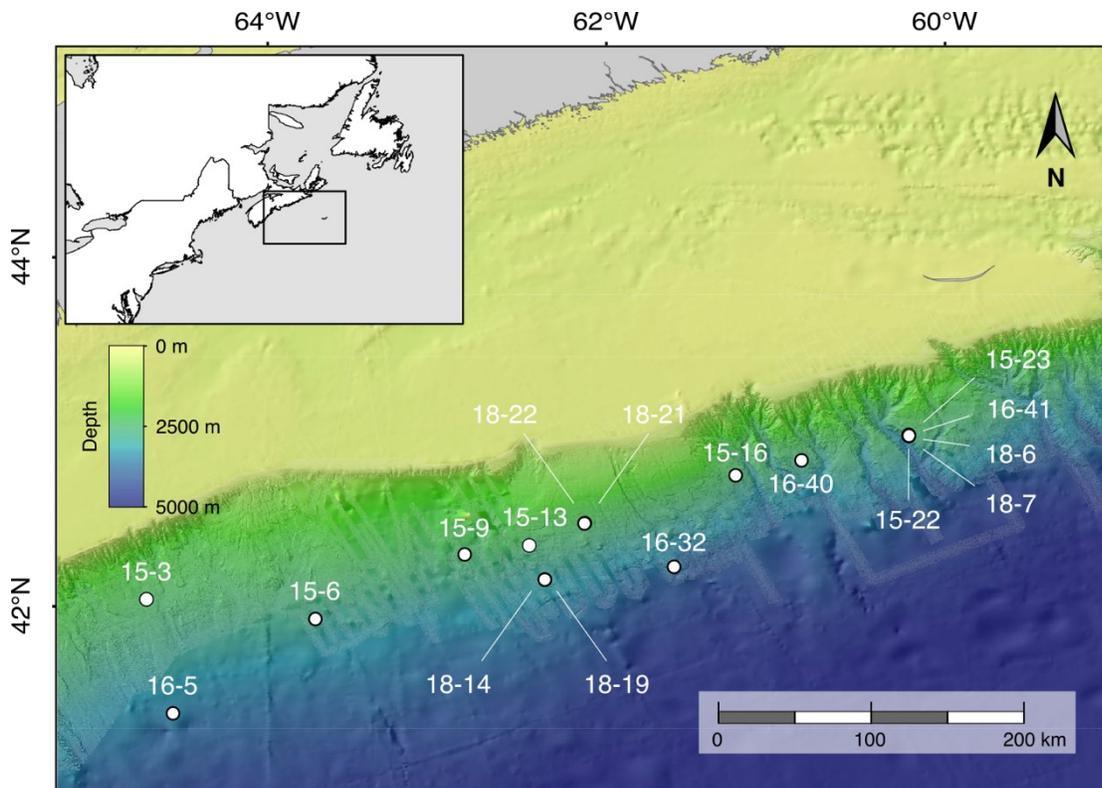


Figure S3.1. Deep sea study sites in the NW Atlantic Ocean. Sediment coring locations on the Scotian Slope. The inset shows the extent of the 20,000 km² study area, off the east coast of Nova Scotia, Canada. Bathymetric map from the General Bathymetric Chart of the Oceans (GEBCO, www.gebco.net) and National Oceanic and Atmospheric Administration (NOAA, www.ngdc.noaa.gov/).

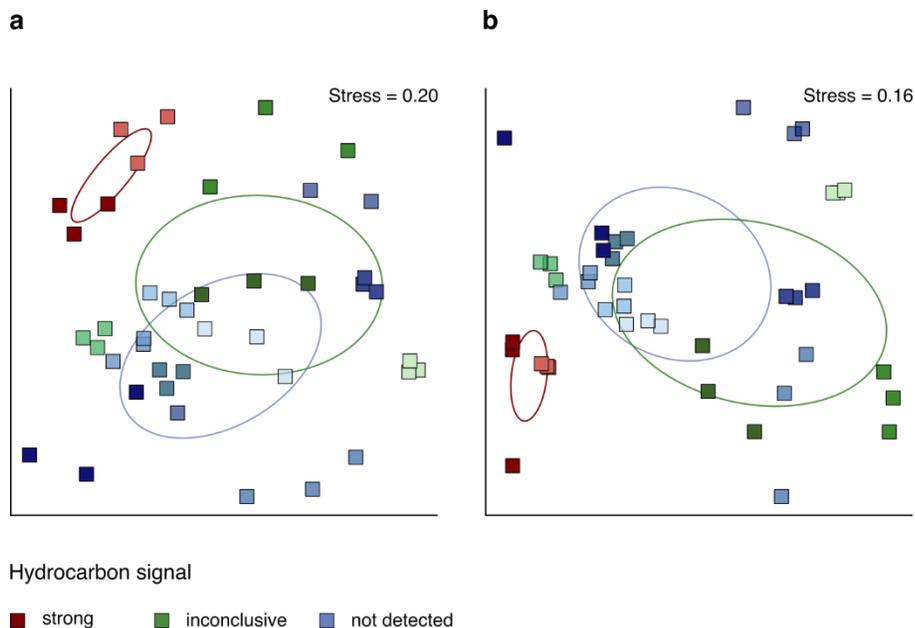


Figure S3.2. Microbial community variance between core sites. Non-metric multidimensional scaling of the Bray-Curtis dissimilarity of the microbial community composition after sediment incubation at (a) 40°C and (b) 60°C (for 50°C incubations, see **Figure 3.2b**). Red symbols indicate sites with strong geochemical evidence of hydrocarbons ($n=2$), green symbols indicate sites with inconclusive hydrocarbon signals ($n=4$), and blue symbols indicate sites where hydrocarbons were not detected ($n=8$). Triplicate amplicon libraries are plotted for each condition. Sites with strong thermogenic hydrocarbon signals have distinct microbial populations after high temperature incubation (see also **Table S3.4**), relative to the sites without thermogenic hydrocarbon signals. This is indicated by standard deviation ellipses of the hydrocarbon groups.

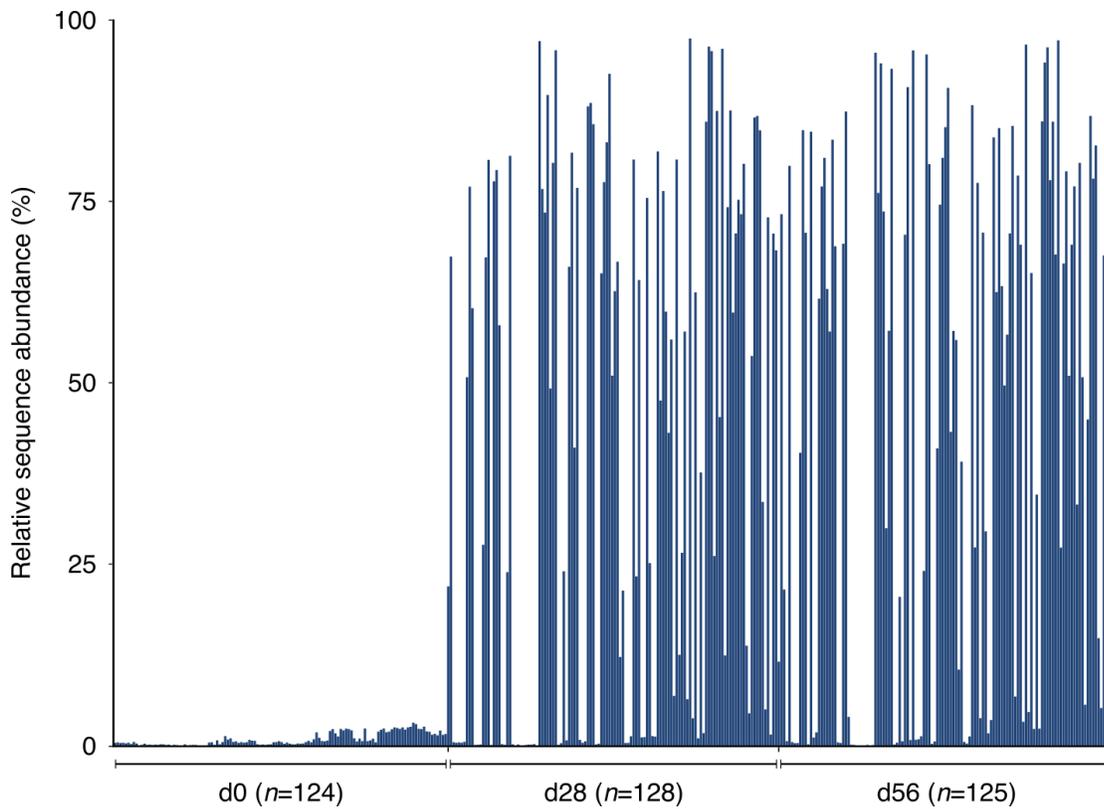
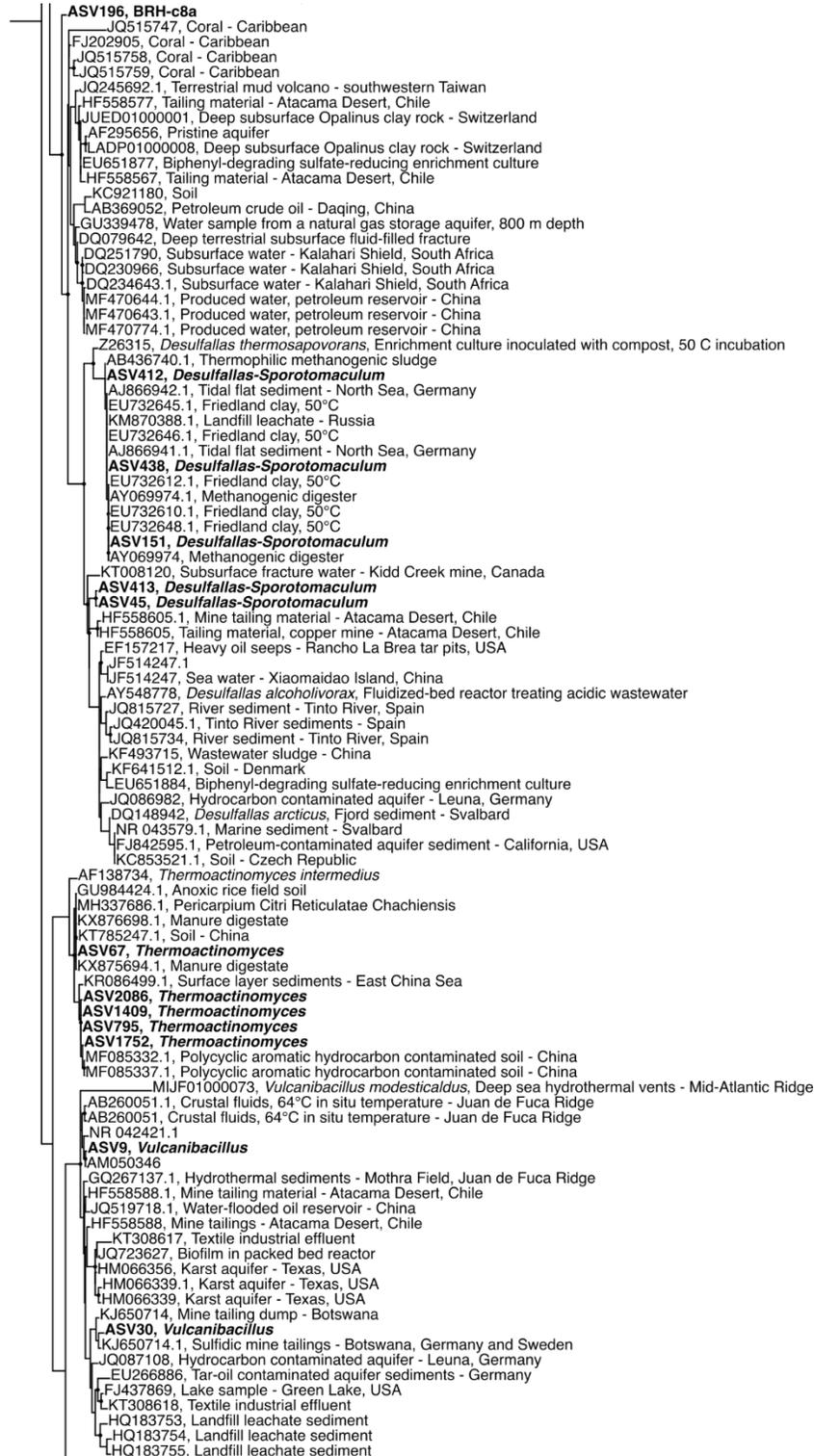


Figure S3.3. Endospore germination and enrichment after high temperature incubation. Relative sequence abundance of *Firmicutes* in 372 libraries of 16S rRNA gene amplicons before (d0; post-pasteurisation) and after 28 and 56 days of anoxic incubation at 40, 50, or 60°C (see **Table S3.3** for detailed microbial community compositions). Each blue bar represents a single library of 4,635 subsampled reads, and demonstrates an increased proportion of *Firmicutes* following 28 and 56 days of incubation. The proportion of *Firmicutes* in d0 libraries was on average 0.9%, and was even lower in libraries from unpasteurized surface sediments (data not shown).

MK766147.1, H2-producing continuous bioreactor
 LC123718.1, Deep subterrestrial environments - Japan
 LC123718, Deep groundwater - southeastern Kyushu, Japan
ASV279, *Caldicoprobacter*
ASV496, *Caldicoprobacter*
 GU118676, Coral - Caribbean
 GU118676.1, Coral - Caribbean
 GU118758.1, Coral - Caribbean
 JQ515772, Coral - Caribbean
 JQ515767.1, Coral - Caribbean
 JQ515767, Coral - Caribbean
 AF113543, *Thermohalobacter berrensii*, Solar saltern - Berre Lagoon, France
 KP004425.1, Deep-sea hydrothermal vent - Southwest Indian Ridge, East Pacific and South Atlantic
 KC533829.1, Hydrothermal vent - East Pacific Ocean
 KC901624.1, Hydrothermal vent sediment, 57.6°C fluids - Guaymas Basin, Mexico
 KP004433.1, Deep-sea hydrothermal vent - Southwest Indian Ridge, East Pacific and South Atlantic
ASV366, *Caloranaerobacter*
 NR 135860.1, Deep-sea hydrothermal vent, 30-75°C incubation - Pacific Ocean
 FN396790.1, Marine surface sediment, 50°C incubation - Smeerenburgfjorden, Svalbard
 KX955408.1, Water column and sediment - Aarhus Bay, Denmark
 FN396787.1, Marine surface sediment, 50°C incubation - Smeerenburgfjorden, Svalbard
 FN396776.1, Marine surface sediment, 50°C incubation - Smeerenburgfjorden, Svalbard
 FN396778.1, Marine surface sediment, 50°C incubation - Smeerenburgfjorden, Svalbard
ASV147, *Caloranaerobacter*
ASV516, *Caloranaerobacter*
ASV615, *Caloranaerobacter*
 AJ431243, *Alvinella pompejana* white tubes - East Pacific Rise
 AJ320233, *Caminicella sporogenes*, Deep sea hydrothermal vent - East Pacific Rise
 AJ431244, *Alvinella pompejana* white tubes - East Pacific Rise
 AJ431245, *Alvinella pompejana* white tubes - East Pacific Rise
 AJ874301, Hydrothermal black chimney, 60°C enrichment culture - Rainbow field, Mid-Atlantic Ridge
 AJ874305, Hydrothermal black chimney, 60°C enrichment culture - Rainbow field, Mid-Atlantic Ridge
 AJ874310, Hydrothermal black chimney, 60°C enrichment culture - Rainbow field, Mid-Atlantic Ridge
 AJ874312, Hydrothermal black chimney, 60°C enrichment culture - Rainbow field, Mid-Atlantic Ridge
ASV43, *Caminicella*
 KX956036.1, Water column and sediment - Aarhus Bay, Denmark
 MN463066.1, Marine sediment
 JN539926.1, Hypersaline microbial mat, Guerrero Negro - Baja California Sur, Mexico
 JN538308.1, Hypersaline microbial mat, Guerrero Negro - Baja California Sur, Mexico
 JN539026.1, Hypersaline microbial mat, Guerrero Negro - Baja California Sur, Mexico
 FN667354, Compost - Lahti, Finland
 FN396772, Marine surface sediment, 50°C incubation - Smeerenburgfjorden, Svalbard
 KM823684, River sediment - China
 JQ407283, Mud volcano, Lei-gong-huo - eastern Taiwan
 HQ916619, Mud volcano, Lei-gong-huo - eastern Taiwan
 HQ916607, Mud volcano, Lei-gong-huo - eastern Taiwan
 KF758687, Deep sea water
 KF964589, Wetland soil - Ebinur lake
 JN539026, Hypersaline mat, Guerrero Negro - Baja California Sur, Mexico
 JN537682, Hypersaline mat, Guerrero Negro - Baja California Sur, Mexico
 JQ407274, Mud volcano, Lei-gong-huo - eastern Taiwan
 FN356285, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 FN356288, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 FN356295, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 DQ647144, Produced water, 70°C in situ oil field temperature - Troll Formation, North Sea
 FN356335, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 FN356327, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 FN356332, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 FN356342, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 FN356328, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 FN356302, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 FN356340, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 FN356237, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 FN356323, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 DQ647125, Produced water, 70°C in situ oil field temperature - Troll Formation, North Sea
 FN356239, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 FN356341, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 FN356304, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 FN356314, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 FN356303, Produced water, 80°C in situ oil reservoir temperature - Dan and Halfdan oil fields, North Sea
 X77837
 GU118213, Coral - Caribbean
 KC668910, Coral - Red Sea
 KC668846.1, Coral - Red Sea
 KC668846, Coral - Red Sea
 KC668837, Coral - Red Sea
 KP305708.1, Reef coral - Luhuitou fringing reef, China
 HQ606285.1, Marine sediments - South China Sea
 KT973497.1, Intertidal outcrops - Isla de Mona, Puerto Rico
 GQ267134.1, Hydrothermal sediments - Mothra Field, Juan de Fuca Ridge
 HQ696463.1, Deep sea sediment - Indian Ocean
ASV41, *Paramaledivibacter*
 FR695371.1, Marine sediment, 25°C incubation - Aarhus Bay, Denmark
 DQ831102.1
ASV2045, *Paramaledivibacter*
ASV60, *Paramaledivibacter*
 KX062021, Hot spring, Polichnitos - Lesvos, Greece
 AF458779, *Paramaledivibacter caminithermalis*, Deep sea hydrothermal chimney - Atlantic Ocean Ridge
 FJ203551.1, Coral - Caribbean
ASV165, *Paramaledivibacter*
 KC668883.1, Coral - Red Sea
 KC668869.1, Coral - Red Sea
 KC668910.1, Coral - Red Sea
 EU573106, Produced water, 131°C in situ oil reservoir temperature - Ekofisk oil field, North Sea
 HQ696463, Deep sea sediment - Indian Ocean
 AB806232, Ocean drilling core - Shimokita Peninsula, Japan
 FJ203551, Coral - Caribbean
 FJ202390, Coral - Caribbean
 JQ515752, Coral - Caribbean
 KC631808, Hypersaline microbial mat - Kiribati
 EF123532, Coral - Caribbean
 DQ446118, Coral - Caribbean
 JX391232, Surface marine sediment - Hong Kong, China
 KC668891, Coral - Red Sea
 KT783480, *Wukongibacter baidiensis*, Deep sea hydrothermal field - Southwest Indian Ridge
 JMSU01000863, Marine intertidal flat - Wadden Sea, Germany
 IAB806231, Ocean drilling core - Shimokita Peninsula, Japan

AM777975, Subterrestrial alkaline groundwater - Cabeco de Vide Aquifer, Portugal
 AM777972, Subterrestrial alkaline groundwater - Cabeco de Vide Aquifer, Portugal
 AY741712, Deep borehole water, gold mine - South Africa
 AM778013, Subterrestrial alkaline groundwater - Cabeco de Vide Aquifer, Portugal
 KX450231, *Thermodesulfitimonas autotrophica*, Terrestrial hot spring, 67 °C - Kuril Islands, Russia
ASV70, *Candidatus Desulfuridis*
 KP151232, Thermophilic chicken dung
 AY604055, Dolomite aquifer, 896 m depth - Chuniespoort group, South Africa
 FJ712408, Submarine mud volcano, Kazan - East Mediterranean Sea
 FJ712600, Submarine mud volcano, Kazan - East Mediterranean Sea
 KP151255.1, Thermophilic chicken dung
 KR013605.1, Anaerobic reactor sludge
 LR640649.1, Wastewater treatment system
 KP151232.1, Thermophilic chicken dung
 LR640519.1, Wastewater treatment system
 AM777962, Subterrestrial alkaline groundwater - Cabeco de Vide Aquifer, Portugal
 GU188991, Subsurface observatory, 64°C fluids - Juan de Fuca, Pacific Ocean
 EU731007, Fracture water from a borehole, 2.8 km depth - Witwatersrand Basin, South Africa
 EU730994, Fracture water from a borehole, 2.8 km depth - Witwatersrand Basin, South Africa
 EU730986, Fracture water from a borehole, 2.8 km depth - Witwatersrand Basin, South Africa
 EU730992, Fracture water from a borehole, 2.8 km depth - Witwatersrand Basin, South Africa
 EU730977, Fracture water from a borehole, 2.8 km depth - Witwatersrand Basin, South Africa
 EU730976, Fracture water from a borehole, 2.8 km depth - Witwatersrand Basin, South Africa
 EU730999, Fracture water from a borehole, 2.8 km depth - Witwatersrand Basin, South Africa
 EU730997, Fracture water from a borehole, 2.8 km depth - Witwatersrand Basin, South Africa
 EU731003, Fracture water from a borehole, 2.8 km depth - Witwatersrand Basin, South Africa
 CP000860, Fracture water from a borehole, 2.8 km depth - Witwatersrand Basin, South Africa
 EU730996, Fracture water from a borehole, 2.8 km depth - Witwatersrand Basin, South Africa
 AY741715, Deep borehole water, gold mine - South Africa
 AY741695, Deep borehole water, gold mine - South Africa
 FR749980, *Caldicoprobacter faecalis*, Sewage sludge
ASV885, *Limnochorda*
 AP014924
 MN816750.1, Ultramafic hydrothermal setting
 MG855670.1, Ultramafic hydrothermal setting
 MG855671.1, Ultramafic hydrothermal setting
 NR 136767.1, Sediment of meromictic lake - Japan
 JF218025, Skin
 AB374129, Hyperthermophilic, 65-80°C, anaerobic reactor
 KF026007, thermophilic methane fermentation reactor
 FN667455, Compost - Lahti, Finland
 FN667408.1, Municipal drum compost
 AB451833.1
 MF080064.1, Sewage sludge
 GU325833.1, Thermophilic, 58-65°C, sludge from wastewater - Co. Kerry, Ireland
 KX955546.1, Water column and sediment - Aarhus Bay, Denmark
 KU366344.1, Oil reservoir production fluids - HeNan oilfield, China
 NR 075044.2
 FR716211.1, Rice paddy soils
 FN868427.1, Biogas reactors, mesophilic and thermophilic
 KY313610.1, Thermophilic microbial fuel cell
ASV252, *Symbiobacterium*
ASV237, *Symbiobacterium*
 AB361629, *Symbiobacterium ostreiconchae*, Oyster shell - Shizuoka, Japan
 GU325835, Thermophilic, 58-65°C, sludge from wastewater - Co. Kerry, Ireland
 FN667307, Compost - Lahti, Finland
 AF190460, Compost
 AB551168, Rice field soil - Yamagata, Japan
 AP006840, *Symbiobacterium thermophilum* IAM 14863, Compost
 JX133588.1, Soil
 JX133637.1, Soil
ASV1878, *Desulfofarcimen*
ASV451, *Desulfofarcimen*
 JX133637, Soil
 JX133588, Soil
 AB778029, Freshwater lake sediment, 42-45°C incubation - Lake Mizugaki, Japan
 KT223435.1, Ship hull
 NR 114368.1, Freshwater lake sediment - Lake Mizugaki, Japan
 NR 114369.1, Freshwater lake sediment - Lake Mizugaki, Japan
 AB778017, *Desulfofarcimen intricatum*, Freshwater lake sediment - Lake Mizugaki, Japan
 AB778020, Freshwater lake sediment, 42-45°C incubation - Lake Mizugaki, Japan
 AB778025, Freshwater lake sediment, 42-45°C incubation - Lake Mizugaki, Japan
 JX133608, Soil
 JX861507, *Desulfohalotomaculum peckii*, Anaerobic digester treating abattoir wastewater - Tunisia
 JQ741980, Marine sediment - Yellow Sea, China
 JQ741985, Marine sediment - Yellow Sea, China
ASV575, *Desulfohalotomaculum*
ASV775, *Desulfohalotomaculum*
ASV910, *Desulfohalotomaculum*
 JQ304573.1, Marine sediment, 61°C incubation - Aarhus Bay, Denmark
 JQ741980.1, Marine sediment - Yellow Sea, China
ASV299, *Desulfohalotomaculum*
 JQ741985.1, Marine sediment - Yellow Sea, China
 JQ304695, Marine sediment, 28-85°C incubation - Smeerenburgfjorden, Svalbard
 JQ304694, Marine sediment, 28-85°C incubation - Smeerenburgfjorden, Svalbard
 JQ304680.1, Marine sediment, 46°C incubation - Aarhus Bay, Denmark
ASV13, *Desulfohalotomaculum*
ASV655, *Desulfohalotomaculum*
 FN652832.1, Marine sediment, 50°C incubation - Arctic
 FN652831.1, Marine sediment, 50°C incubation - Arctic
 FN652813.1, Marine sediment, 50°C incubation - Arctic
 FN652828.1, Marine sediment, 50°C incubation - Arctic
 FN652839.1, Marine sediment, 50°C incubation - Arctic



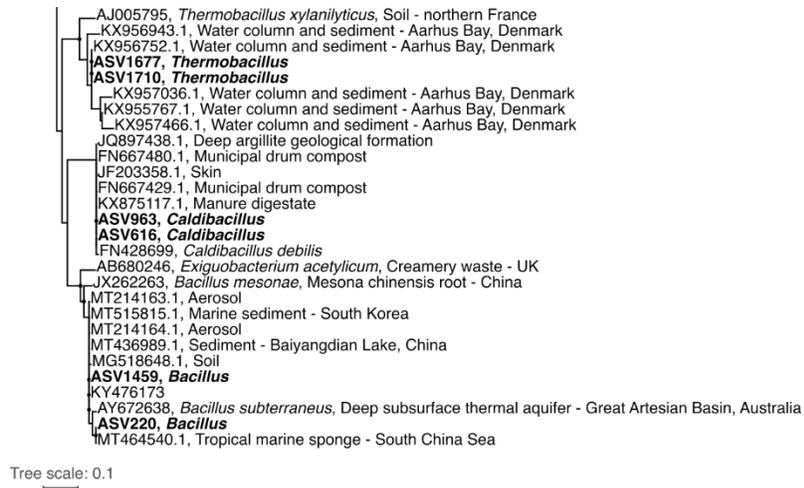


Figure S3.4. Phylogenetic association of seep-associated sequences with sequences from other environments. Maximum likelihood tree showing phylogenetic relationships between 42 seep-associated ASVs (bold) and close relatives in the GenBank database. Black circles at the branch nodes indicate >80% bootstrap support (1,000 re-samplings). Scale bar indicates 10% sequence divergence as inferred from PhyML. *Pseudomonas aeruginosa* (accession number Z76672) was used as an outgroup to root the tree.

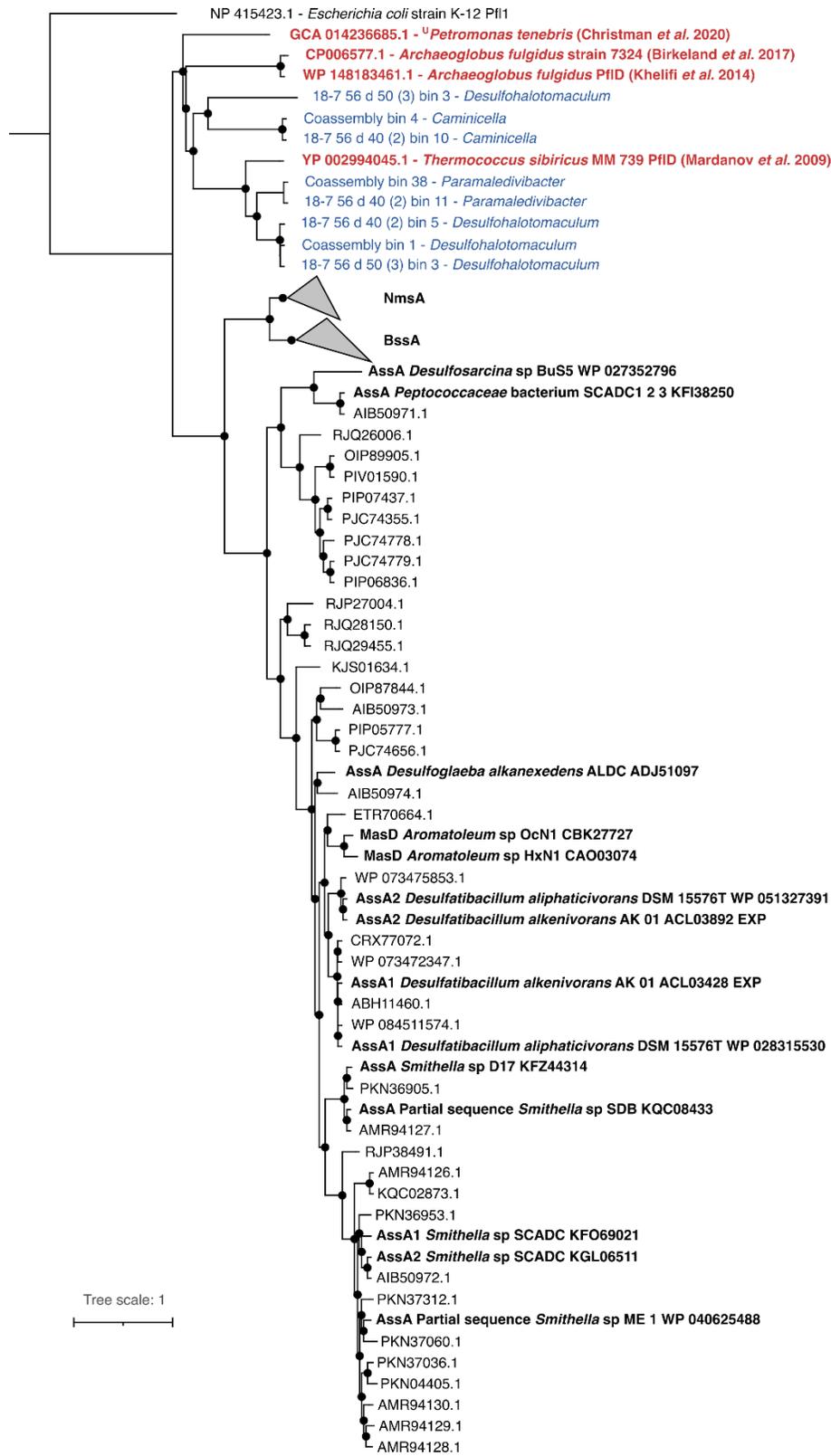


Figure S3.5. Phylogenetic relationships of putative glycyl-radical enzymes with alkylsuccinate synthases. Putative anaerobic alkane-degrading pyruvate-formate lyase enzyme variants from *Desulfohalotomaculum*, *Caminicella* and *Parameldivibacter* thermophilic spores in this study (shown in blue) cluster together with homologous pflD gene sequences found in the oil reservoir bacteria ^U*Petromonas tenebris* (Christman *et al.* 2020) and oil reservoir archaea *Archaeoglobus fulgidus* and *Thermococcus sibiricus*, shown in red. Phenotypic evidence of alkane hydrocarbon biodegradation has been observed at high temperature under anaerobic conditions for these *Archaeoglobus* and *Thermococcus* strains (Khelifi *et al.* 2014; Birkeland *et al.* 2017; Mardanov *et al.* 2009). Reference sequences of alkane succinate synthase (AssA and MasD) genes with corresponding experimental verification of anaerobic alkane degradation are shown in bold. Benzyl succinate synthase (BssA) and naphthyl-2-methyl-succinate synthase (NmsA) sequences are represented by collapsed clades. Black circles at the branch nodes indicate >80% bootstrap support (1,000 re-samplings). Scale bar indicates 10% sequence divergence as inferred from PhyML. A sequence of pyruvate formate lyase (Pfl) from *E. coli* was used to root the tree.



Figure S3.6. Seabed hydrocarbon seepage. ROV imagery of hydrocarbon seepage in the vicinity of sites with strong thermogenic hydrocarbon signals on the Scotian Slope. Vertical hydrocarbon migration rates from a subsurface reservoir to the seabed in the study area can be estimated using buoyancy models incorporating Darcy's law (Arp, 1992). Transport facilitated by vertical migration of hydrocarbons through subsurface sediments at rates of 0.1 m d^{-1} (Arp, 1992; Rice, 2022) would require on the order of 50 years for dispersal from hydrocarbon reservoirs in 2,000 mbsf Cretaceous sediments (Deptuck and Kendell, 2020) up to the surface and into the marine bottom water (**Figure 3.4a**). However, modelled migration rates predict that gas will migrate 40,000 times faster than oil for any given sediment pore size (Rice, 2022). If migration rates of 0.1 m d^{-1} apply for gas (density 0.087 g / cm^3), oil will correspondingly migrate at $2.5 \times 10^{-5} \text{ m d}^{-1}$ (0.8 g / cm^3 , 45 API gravity) leading to timescales on the order of eight million years to reach the surface in a similar system. Larger pore sizes especially those associated with faults and fractures (see **Figure 3.1b**) would facilitate faster dispersal from deep sediments up to the surface seabed.

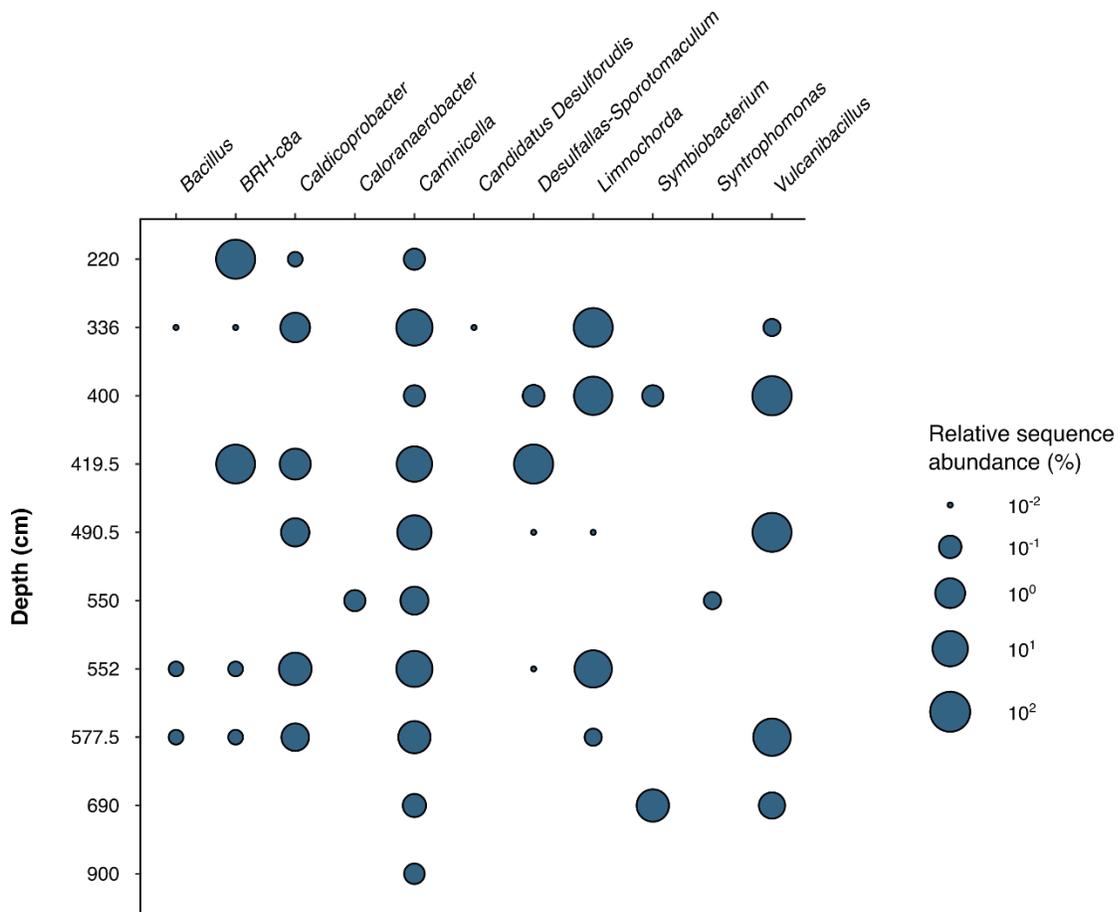


Figure S3.7. Endospores remain viable during burial. Detection of endospore-forming bacteria with the same taxonomic classification as seep-associated ASVs (cf. *Table S3.5*) in deeper Scotian Slope sediment layers following incubation at 50°C. Bubble size indicates the maximum relative sequence abundance of a taxon in unrefined sequence libraries at the indicated depth.

Chapter 4

Thermophilic endospores in the cold seabed germinate and become active in the presence of high concentrations of crude oil

4.1 Abstract

Endospores of thermophilic bacteria can be detected in cold seabed sediment samples following high temperature incubation. This apparent misplacement raises questions about their origins. Marine cold seeps are characterized by the migration of gas and oil from deep subsurface reservoirs up to the seafloor. Geofluid flow in seep systems has been proposed as a dispersal vector for deep-to-shallow cell transport connecting the surface with hot oil reservoirs where thermophilic populations are known to reside; however, investigations of endospore germination and activity in petroleum reservoir-like conditions are lacking. Here the enrichment of thermophilic endospore-forming bacteria in a series of incubations with high concentrations of crude oil (1% v/v) is assessed using seep-associated sediments from Eastern Gulf of Mexico and offshore Nova Scotia. Distinct microbial populations emerging under different test conditions reveal *Firmicutes* with a high degree of relatedness to hydrocarbon seep indicator species are selected following crude oil amendment. Providing growth conditions consistent with what microbes inhabiting petroleum reservoirs encounter substantiates a deep biosphere origin. Substrate concentration measurements indicate prominent sulfate-reducing and fermentative metabolisms among seabed thermophilics from the Eastern Gulf of Mexico and offshore Nova Scotia, respectively, potentially revealing important biogeochemical reactions in corresponding subsurface oil reservoir ecosystems. The applicability of seep indicator species identified in one location to predict seep occurrence in the seabed in a new location was evaluated in two hydrocarbon prospective basins with a combined 125 seabed sampling sites and 48 indicator species. This revealed that subsets of indicator species may be more likely to successfully rank different seabed samples. Overall, the approach presented here highlights the potential application of endospores as proximity markers that extend the detectable signal of hydrocarbon seeps beyond what can be measured by geochemical testing alone.

4.2 Introduction

Endospores of thermophilic bacteria are routinely detected in cold environments that do not match their metabolic requirement for higher temperatures. A constant flux of these microbes into seabed sediments indicates prominent microbial dispersal mechanisms (Hubert *et al.* 2009; de Rezende *et al.* 2013; Müller *et al.* 2014, Chakraborty *et al.* 2018). Integration of deep-sea geophysics, geochemistry, microbiology, and genomics has shown how cells can be transported via geofluids from warm oil reservoirs up into the cold seabed (*Chapter 3*). This evidence supports that thermogenic hydrocarbon seepage originating in deep hot oil-bearing sediment layers is an important vector for microbial dispersal, connecting the deep biosphere to surface environments.

Bacterial cells being supplied from petroleum reservoir habitats into the overlying ocean by migrating hydrocarbons contribute to a low abundance biosphere and dormant ‘seed bank’ (Jones and Lennon, 2010; Locey *et al.* 2017). These seed banks, consisting of microbes that can become active under favourable environmental conditions, increase marine phylogenetic and functional diversity, and can prolong the persistence of genotypes (Sogin *et al.* 2006; Lennon *et al.* 2021). The simultaneous influx of energy rich hydrocarbons at seepage sites supports accelerated biogeochemical reactions in the seabed, resulting in biological oases of chemosynthetic communities (Joye, 2020). Efforts to identify and characterise the occurrence of seabed cold seeps are important within an ecological context (Vigneron *et al.* 2017a; Dong *et al.* 2019), but are also needed for more accurate estimates of global hydrocarbon and greenhouse gas releases into the hydrosphere and atmosphere (Solomon *et al.* 2009; Boetius and Wenzhöfer, 2013; Ciotoli *et al.* 2020). Whereas shallow water seep systems are well studied (Suess, 2010), estimates of the extent of seepage in the deep ocean, e.g., >1,000 m water depth, are less well constrained, as is the composition of the microbial rare biosphere in these environments.

The ability of certain *Firmicutes* to withstand extreme temperatures has been used to highlight their potential for being effective colonisers of hot environments, including geothermal systems related to oceanic spreading centres and subsurface petroleum reservoirs (O’Sullivan *et al.* 2015). Abundant crude oil compounds in petroleum reservoirs, including diverse aliphatic and aromatic hydrocarbons, present both ecological challenges and opportunities for microbial populations. Higher solubility petroleum compounds, as well as certain metabolic intermediates formed during

their biodegradation, exert toxic pressures on microbes, inhibiting cell growth and activity (Sikkema *et al.* 1995; Chibwe *et al.* 2015). Despite this, certain microorganisms have been shown to be tolerant to and grow in the presence of hydrocarbons (Inoue and Horikoshi, 1989; Ramos *et al.* 2002), while numerous microorganisms have evolved a capability for utilizing hydrocarbons as a carbon and energy source for growth (Das and Chandran, 2011; Rabus *et al.* 2016). For endospores of thermophilic bacteria in seabed sediments that share genomic relationships to microbes inhabiting petroleum reservoirs, activity in the presence of high concentrations of crude oil would be necessary for survival. Similarly, given the prevalence of *Firmicutes* and widespread occurrence of sporulation genes reported in petroleum reservoirs from around the world (Hubert *et al.* 2012; *Chapter 2*), crude oil may be expected to harbour viable dormant endospores.

The Gulf of Mexico petroleum basin is well known for widespread natural seepage of hydrocarbons sourced from numerous deeply buried petroleum reservoirs. The basin covers an area of 1.6 million km², with oil seep discharge rates of over 160,000 m³ yr⁻¹ (Coleman *et al.* 2003; Kennicutt, 2007). The Scotian Basin, offshore Nova Scotia, extends over an area of 260,000 km², where seafloor seepage has been linked to the presence of an active petroleum system(s) (Kidston *et al.* 2002). To further validate a subsurface oil reservoir provenance of specific seabed thermophiles, a series of incubations with high concentrations of crude oil and sediment from locations of hydrocarbon seepage in the Eastern Gulf of Mexico (EGM; Chakraborty *et al.* 2018) and offshore Nova Scotia (NS; *Chapter 3*) were established (**Figure 4.1a**; **Table 4.1**). Distinct microbial populations under the different test conditions used here reveal thermophilic endospore-forming taxa that are correlated with occurrences of seabed hydrocarbon seepage are also selected for by crude oil amendment in high temperature incubations. This is consistent with these microbes inhabiting petroleum reservoirs and validates their use as bioindicators of marine cold seeps and for characterising underlying petroleum systems.

Table 4.1. Composition of the experimental treatments.

Treatment	Sediment (g)	Crude oil (mL) *	Medium (mL)	Ratio †
1	15 (triple-autoclaved)	0.3	20	1 : 0.01 : 2
2	15	0.3 (triple autoclaved)	20	1 : 0.01 : 2
3	0	0.3	30	0.01 : 1
4	15	0.3	20	1 : 0.01 : 2
5	15 (triple-autoclaved)	0	20	1 : 2
6	0	0.3 (triple autoclaved)	30	0.01 : 1
7 ⁺⁺	15	0	20	1 : 2
negative control	0	0	30	-

* approximately 1% v/v crude oil in the oil-amended microcosms

† ratio of sediment : crude oil : medium

⁺⁺ positive control; experiments as described in Chapter 3

4.2 Results

4.2.1 Migrated thermogenic hydrocarbons in surface sediments indicate seepage

The presence of migrated petroleum (i.e., thermogenic hydrocarbons) was confirmed by separate geochemical analysis of sediments from the Eastern Gulf of Mexico (EGM) and offshore Nova Scotia (NS) obtained at different seabed depth intervals following gravity coring (**Figure 4.1**). Hydrocarbon measurements and parameters were compared to those reported at other sample sites within the Gulf of Mexico (Chakraborty *et al.* 2018) or Nova Scotia offshore (*Chapter 3*), respectively, to further validate the occurrence of hydrocarbon seepage.

In the EGM core, the presence of hydrocarbons was assessed in two ways. Unresolved complex mixture (UCM) concentrations measuring extractable hydrocarbons that are not separated by gas

chromatography averaged 20 µg/g dry sediment. Total scanning fluorescence (TSF) intensities, measuring aromatic hydrocarbon contributions averaged 178,017 maximum intensity units (**Figure 4.1**). Values for the three tested intervals of the EGM core are above thresholds (TSF $\geq 50,000$ max. intensity units and UCM ≥ 10 µg/g) used for defining the unambiguous presence of thermogenic hydrocarbons in the Gulf of Mexico seabed (Chakraborty *et al.* 2018). An average thermogenic/diagenetic (T/D) ratio of 4.1 further supports a preferential contribution of thermogenic hydrocarbons over recent biological organic matter (**Table S4.2**). These findings confirm the EGM core site to be a cold seep location.

NS sediment was analysed using a different geochemical approach, as outlined in *Chapter 3*. Elevated nC_{17}/nC_{27} ratios in liquid hydrocarbon extracts demonstrate higher concentrations of thermally derived mid-weight *n*-alkanes compared to higher carbon numbered *n*-alkanes preferentially contributed from recent terrestrial sources. Proportions of thermally rearranged steranes (e.g., % 27 d β S) further support a thermogenic hydrocarbon contribution at this site (**Figure 4.1**). Additional geochemical indicators of thermogenic hydrocarbons included elevated concentrations of C_2 – nC_4 compounds and heavy $\delta^{13}C$ values for methane (-52‰), a lack of odd-over-even alkane distributions in the nC_{23-33} range, and thermally mature terpane distributions (C30 $\alpha\beta$ relative to C31 $\alpha\beta$ 22R hopane) (**Table S4.2**). These measurements, coupled to geophysical indicators of hydrocarbon seepage reported previously at the same site (*Chapter 3*), validate the occurrence of a cold seep at this location.

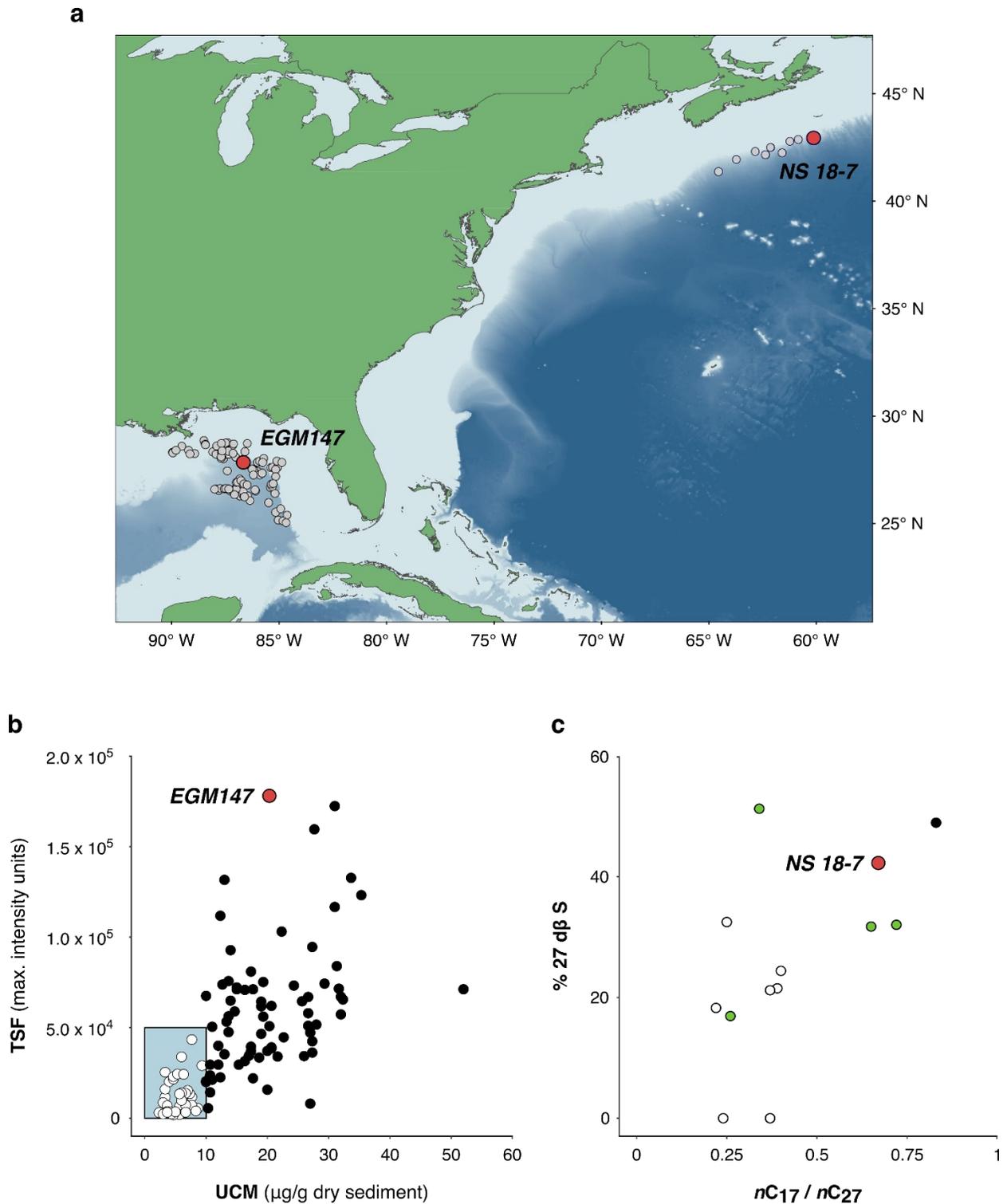


Figure 4.1. Deep sea seabed sampling sites and hydrocarbon geochemistry. **a**, Location of 111 sediment cores in the Eastern Gulf of Mexico and 14 sediment cores offshore Nova Scotia tested for geochemical evidence of migrated thermogenic hydrocarbons. The red points labelled ‘*EGM147*’ and

‘NS 18-7’ indicate the location of the EGM and NS cores. **b**, Unresolved complex mixture (UCM) concentration and total scanning fluorescence (TSF) intensity averaged across three depth intervals of individual gravity cores at the 111 seabed locations in the Eastern Gulf of Mexico. The blue box highlights thresholds of the UCM and TSF for the unambiguous presence of migrated thermogenic hydrocarbons in sediment extracts. TSF thresholds were raised five-fold from the industry standard of 10,000 intensity units to account for additional terrigenous input of organic matter in the Gulf of Mexico. The red point (labelled ‘EGM147’) indicates geochemical parameters of the EGM core sediment used for the experimental treatment conditions presented in this study. Data and plot reproduced from Chakraborty *et al.* (2018). **c**, Average nC_{17}/nC_{27} ratio and % 27 d β S (diasterane) in liquid hydrocarbon extracts at intervals from 14 sediment core locations on the Scotian Slope (cf. **Figure S3.1**). The red point (and ‘NS 18-7’ label) indicates geochemical parameters of the NS sediment used in the experiments performed here (cf. **Table 4.1**). Black, green, and white points represent other sites with strong, inconclusive, and absent geochemical signals of thermogenic hydrocarbons, respectively (based on these and other geochemical parameters). Additional geochemical data supporting these classifications is provided in **Table S4.2**. An overview of the NS sediments is also provided in **Figure 3.2a**.

4.2.2 Microbial diversity and community structure

High temperature anoxic incubation of either NS or EGM sediment according to the experimental conditions outlined in **Table 4.1** resulted in a decrease in average species richness and alpha diversity over 14 days (**Table S4.3**). Incubations with sediment that had not been autoclaved in advance (Treatments 2, 4, and 7) were more diverse than the incubations with the same sediments that had undergone triple autoclaving (Treatments 1 and 5). Introducing a high concentration of crude oil (~1% vol/vol) in non-autoclaved sediment incubations (Treatments 2 and 4) resulted in only a slightly lower alpha diversity (Shannon index average $H' = 2.8$) than equivalent sediment incubations that were not amended with crude oil and thus had a low concentration of hydrocarbons (~26 ppm vol/vol; i.e., ~400 times lower than the crude oil amendments) owing to naturally-present compounds in the seep sediment samples (Treatment 7; Shannon index average $H' = 3.1$). Low species diversity and low numbers of unique ASVs were observed in the triple-autoclaved crude oil incubation (Treatment 6). Insufficient numbers of sequence reads passing quality thresholds were recovered to reliably assess the microbial community composition of the crude oil incubation without sediment (Treatment 3) or the medium-only negative control.

Substantial changes in sulfate and organic acid concentration were only measured in incubations containing sediment that was not autoclaved prior to 50°C incubation (*Figure 4.2a; Table S4.4*). Complete removal of sulfate — likely the result of anaerobic respiration utilizing sulfate as a terminal electron acceptor — was observed in incubations containing non-autoclaved EGM sediment. Corresponding removal of succinate, lactate, formate, propionate, and butyrate indicates that sulfate respiration is primarily coupled to the oxidation of the low molecular weight organic acids supplied in the medium. Interestingly, complete removal of sulfate (15.3 mM) was also observed in Treatment 7 when only 1.3 mM of formate and negligible amounts of other organic acids appeared to be removed or produced. According to the redox stoichiometry of sulfate reduction coupled to formate oxidation (*Eq. 4.1*), an insufficient amount of formate was removed to account for the complete removal of sulfate, suggesting the presence of additional electron donors derived from the EGM sediment matrix.



Sulfate removal was less evident in NS sediment incubations, yet changes in organic acid concentrations were still observed. The largest increases in the number and average relative abundance of *Firmicutes* — the phylum to which known endospore-forming bacteria belong — were observed in incubations exhibiting changes in sulfate and organic acid concentrations (*Figure 4.2b; Table S4.5*).

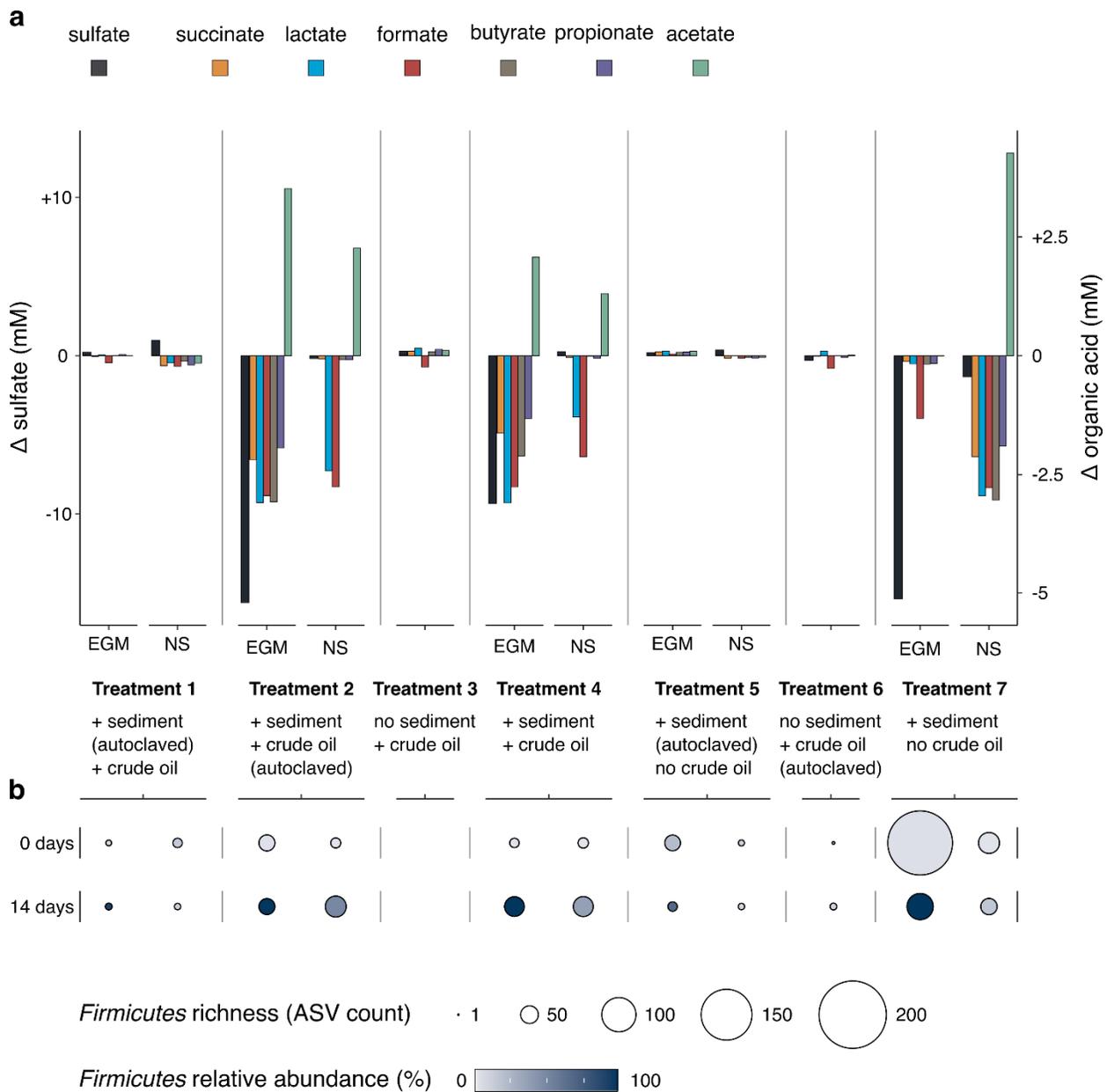


Figure 4.2. Chemical signals of metabolism in the endospore-forming population. **a**, Changes in sulfate and organic acid concentrations after 14 days of incubation of the EGM and NS sediments corresponding to the different experimental treatment conditions (cf. **Table 4.1** and **Table S4.4**). Negative control incubations (no sediment; no crude oil) did not reveal any change in sulfate or organic acid concentrations (not shown). **b**, Average number of ASVs affiliated with the phylum *Firmicutes* and average relative abundance of *Firmicutes* in sequence libraries before incubation (“0 days”; i.e., post pasteurisation) and after 14 days of incubation (“14 days”) underneath the corresponding experimental conditions.

Significant shifts in microbial community composition (PERMANOVA, $P = 0.1$; **Table S4.6**) were observed following 14 days of high temperature incubation under all test conditions containing EGM sediment, whereas significant shifts in community composition were observed sporadically across 14-day incubations containing NS sediment. A similar trend was also observed when a subset of ASVs affiliated with the phylum *Firmicutes* were investigated (**Table S4.6**). Incubations without sediment, i.e., those containing only crude oil and medium, led to no shift in microbial community being observed following incubation. Prevalent taxa varied between different experimental conditions (**Figure 4.3**; **Figure S4.1**; **Table S4.5**), consistent with the diversity trends observed between different experimental conditions as well as measured changes in concentrations of sulfate and organic acids. Diverse families known to encode sulfate-reducing metabolism (e.g., *Desulfotomaculales*, *Desulfitibacteraceae* and *Desulfallas-Sporotomaculum*) were abundant in incubations containing EGM sediment, where complete sulfate removal was measured (**Table S4.4** and **S4.5**). Multilevel pattern *IndicSpecies* analysis (De Cáceres *et al.* 2010) showed ASVs affiliated with the genera *Desulfallas-Sporotomaculum*, *Thermincola* and *Desulfitibacter*, as well as the families *Sulfobacillaceae* and *Heliobacteriaceae*, were correlated ($P < 0.05$) with instances of significant sulfate reduction (**Table S4.4**). Triple autoclaving of EGM sediment was highly selective for *Limnochordaceae*, primarily represented by a single ASV, in both the presence and absence of additional crude oil. This taxon shares high (>99%) 16S rRNA gene sequence identity to the moderate thermophile *Limnochorda pilosa* which does not use sulfate as an electron acceptor for growth with organic substrates (Watanabe *et al.* 2015), in agreement with the absence of sulfate removal in the same triple-autoclaved incubations (**Figure 4.2**).

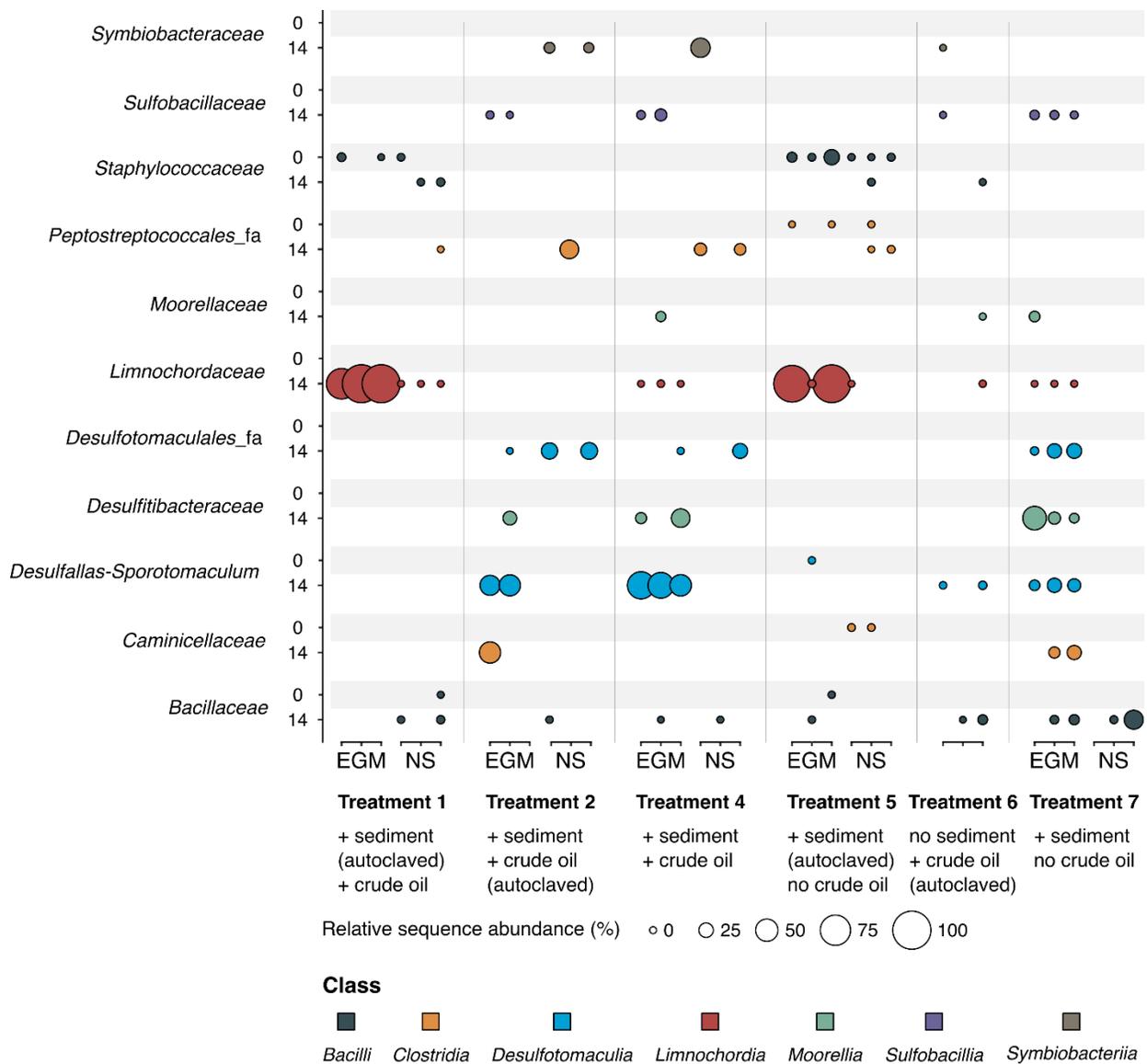


Figure 4.3. *Firmicutes* compositional changes under varying experimental conditions. Relative sequence abundance of 16S rRNA gene amplicons affiliated with family-level groups belonging to the phylum *Firmicutes*. Data reflects 0 days of incubation (i.e., post pasteurisation at 80°C but prior to incubation at 50°C) and 14 days of incubation under the respective treatment conditions (see **Table S4.5** for additional data). Treatment 3 (oil and medium without sediment) is omitted from the graph due to insufficient sequence reads to reliably profile the microbial community, likely owing to the absence of germination and activity under this condition (**Figure 4.2** shows corresponding absence of chemical signals of metabolism in Treatment 3). Taxa shown combine to represent greater than 25% of the total sequence reads across all libraries. Bubble colours indicate class-level taxonomy indicated in the colour key at the bottom.

4.2.3 Prevalence of seep associated taxa

Microbes associated with cold seeps offshore Nova Scotia were determined (*Chapter 3*) by correlating the occurrence and abundance of *Firmicutes* ASVs (**Table S4.7**; *IndicSpecies*, De Cáceres *et al.* 2010) with geophysical and geochemical evidence of migrated hydrocarbons (**Figure 4.1**; **Table S4.2**). A total of 42 unique ASVs belonging to endospore-forming bacterial taxa correlate with hydrocarbon seepage (see list in **Table S4.8**). By applying the same statistical approach to *Firmicutes* ASVs identified in heated sediments from 111 different sites in the Eastern Gulf of Mexico (**Table S4.7**; Chakraborty *et al.* 2018), a further six ASVs were found here to correlate with geochemical evidence of thermogenic hydrocarbons (**Figure 4.1**; **Table S4.2**; see list in **Table S4.8**). As a result of these correlations to geochemical parameters, the 48 ASVs are inferred to be (and correspondingly referred to as) seep ‘indicator species’.

Organisms phylogenetically related to these 48 ‘indicator species’ were identified in the different experimental conditions employed in the present study using ASV-based BLAST comparisons with a sequence identity threshold of $\geq 95\%$. Sequence hits above this cut-off were quantified to assess the experimental conditions that best select for indicator taxa. *Firmicutes* sequence hits ($\geq 95\%$) were detected in all treatments (**Table S4.8**; **Figure 4.4**). Incubations containing non-autoclaved sediment (Treatments 2, 4, and 7) had the highest number of hits. Triple autoclaving sediments resulted in the highly selective conditions described above, meaning the number of sequence hits was low, but the corresponding total abundance of those sequences was high. Notably, the addition of crude oil to incubations of NS and EGM sediments (Treatment 2 and 4) led to an increase in the number and abundance of sequences with a high degree of relatedness ($\geq 95\%$) to the 48 indicator species, as compared to sediment incubations without crude oil (Treatment 7).

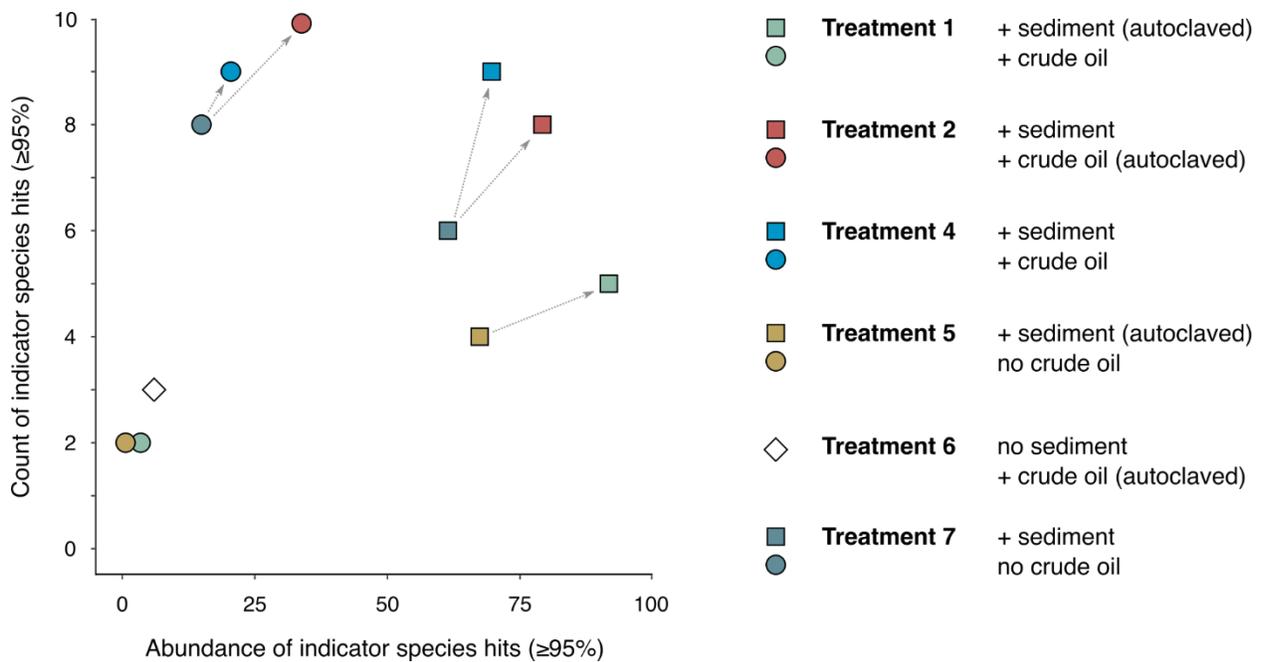


Figure 4.4. Prevalence of organisms phylogenetically related to seep indicator species under different experimental conditions. Count of indicator species hits (y-axis) is the number of unique ASVs in the libraries of each experimental treatment that share $\geq 95\%$ 16S rRNA gene sequence identity with 48 indicator species. Abundance of indicator species hits (x-axis) is the total relative abundance of the same ASVs averaged across the replicates of a treatment condition (see *Table S4.8* for non-averaged data). Squares and circles represent sediment from EGM and NS, respectively, with the exception of Treatment 6 (no sediment) denoted by the white diamond. Arrows highlight results without and with crude oil amendment, indicating that the addition of crude oil resulted in detection of a higher number and abundance of organisms phylogenetically related to 48 seep indicator species, provided that the sediment had not been autoclaved. This suggests crude oil-amendment can improve the detection of seep bioindicators in deep sea sediments.

4.2.4 Identification of seep associated species in different locations

The generalizability of the defined ‘indicator species’ for predicting seeps in different parts of the ocean was examined by estimating the prevalence of related organisms in different areas. Sequences that share $\geq 95\%$ identity to indicator species from the Eastern Gulf of Mexico ($n = 6$) were quantified in high temperature incubations of sediment from 14 Nova Scotia sample sites. Sequences that share $\geq 95\%$ identity to indicator species from offshore Nova Scotia ($n = 42$) were quantified in high temperature incubations of sediment from 111 Eastern Gulf of Mexico sample

sites. The abundance of sequences closely related to the respective sets of indicator species was subsequently compared in locations where hydrocarbon geochemistry signalled the presence of a seep (i.e., HC positive) vs sites where such geochemical signals were absent (i.e., HC negative) (**Figure 4.1; Table S4.2**). Sequences with a high degree of relatedness ($\geq 95\%$) to indicator species were more abundant in HC positive sites than HC negative sites, with total abundances 18 and 26% greater in HC positive sites in the Gulf of Mexico and Nova Scotia, respectively (**Figure 4.5a; Table S4.9**). This comparative difference did not meet the $P < 0.05$ significance threshold, however, further analysis of individual taxa driving this variance revealed ASVs sharing a high degree of relatedness ($\geq 95\%$) to seep indicators affiliated with the genus-level BRH-c8a group (family *Desulfallas-Sporotomaculum*) had large abundance differences between HC positive and negative sites (**Figure 4.5b**).

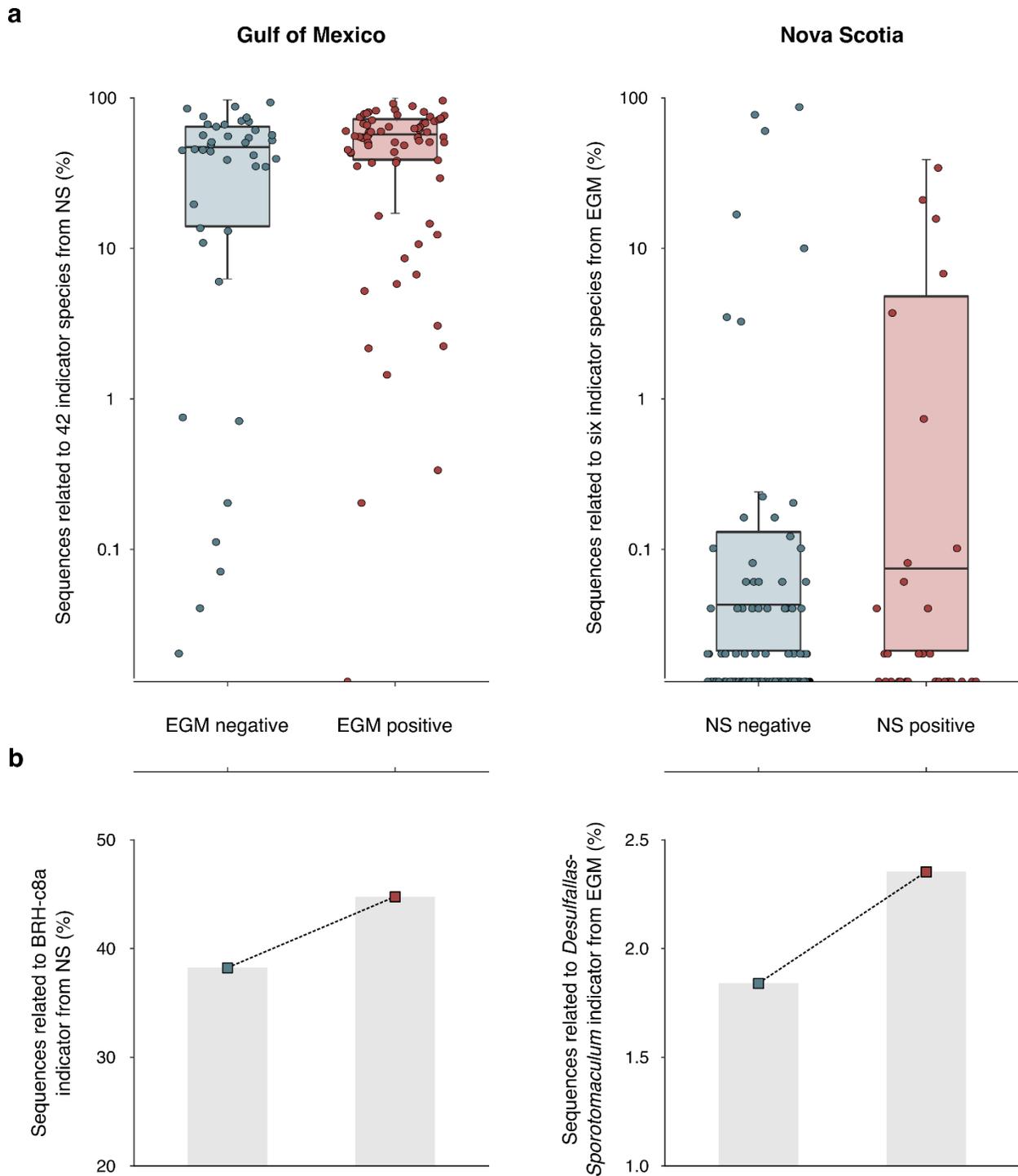


Fig. 4.5. Applicability of regionally-defined seep indicator species for assessment of hydrocarbon seepage in other prospective areas. Sequence data derived from high temperature incubations of Eastern Gulf of Mexico and offshore Nova Scotia sediments reported in Chakraborty *et al.* (2018) and *Chapter 3*, respectively. **a**, Points represent the total relative abundance of sequence reads from one region that share $\geq 95\%$ sequence identity to indicator species defined in the other region. For the Gulf

of Mexico (left) this total is plotted for sequence libraries derived from 111 high temperature sediment incubations (111 locations), based on comparisons with the 42 NS indicator species. For Nova Scotia (right) this total is plotted for sequence libraries derived from 178 high temperature sediment incubations (14 locations), based on comparisons with the 6 EGM indicator species. Comparative indicator assessments and relative abundance data is further summarized in **Table S4.9**. Sequencing libraries are grouped according to sites where hydrocarbon geochemical signals of seepage were present (“HC positive”) vs sites where hydrocarbon geochemical signals of seepage were absent (“HC negative”). Boxplots illustrate distributions of the data shown in the individual points. In both locations, quantifying organisms related to indicator species identified from the other region produced a higher average abundance for HC positive sediments than HC negative sediments. **b**, Relative abundance of two taxonomic groups that were found to better differentiate between HC positive and negative site groupings (**Table S4.9**).

4.3 Discussion

The inactivity of dormant thermophilic bacterial endospores conspicuously found in cold seabed sediments allows investigation of their biogeography without conflating influences of ecological processes such as environmental selection, drift and mutation that typically shape microbial distributions (Hanson *et al.* 2012; 2019). Germination of thermophilic anaerobes and their subsequent enrichment in selective media has enabled focussed assessments of the marine microbial seed bank (Isaksen *et al.* 1994; Hubert *et al.* 2009; de Rezende *et al.* 2013) leading to proposals that these thermophiles arrive here as a result of dispersal from warm anoxic environments. Possible source environments that are consistent with this general description include parts of oceanic spreading centres and petroleum reservoirs (Müller *et al.* 2014, Chakraborty *et al.* 2018; *Chapter 3*). Bacteria originating in petroleum reservoirs should be able to survive in the presence of crude oil and under high temperature conditions. Additionally, crude oil sampled from a reservoir could also be expected to harbour viable endospores. In this study, enrichment from crude oil, as well as the selective pressure crude oil exerts on thermophilic endospores and their ability to germinate was investigated to further validate a petroleum reservoir origin of certain cold seabed thermophiles.

Thermophilic endospore-forming microbes are shown here to be viable members of the rare biosphere in sediments from sites of hydrocarbon seepage in the Eastern Gulf of Mexico and

offshore Nova Scotia, as evidenced by dramatic increases in the abundance of *Firmicutes* concurrent with substrate concentration changes during high temperature incubation (**Figure 4.2**). Microbial activity was detected in the presence and absence of additional crude oil, but only in incubations containing sediment. In incubations with crude oil but without sediment, low species diversity (**Table S4.3**), a lack of detectable sulfate or organic acid concentration changes (**Figure 4.2**), and no significant shifts in microbial community composition during incubation (**Table S4.6**) indicate that the crude oil from the Mizzen reservoir in the Flemish Pass Basin does not harbour thermophilic endospores culturable under the tested conditions.

Selection of diverse populations of microbes that are closely related to seep indicator species by the addition of high concentrations (1% vol/vol) of crude oil to non-autoclaved sediment shows that certain members of the dormant microbiome found in the seabed can germinate and become active under these conditions. A similar observation was made with sulfate-reducing microorganisms in temperate estuary sediment that actively reduce sulfate at high temperature (60°C) and in the presence of crude oil (Bell *et al.* 2020). Interestingly, the microbial community of the estuary sediment included members of lineages that have been isolated from oil reservoirs. The results presented here, and in the literature, therefore support the proposal that these thermophiles originate from hot oily environments, such as subsurface petroleum reservoirs. Assuming this source, seabed detection of reservoir derived bacteria presents a potential strategy for identifying hydrocarbon seeps. Results from this study show that detecting these taxa can be aided by the addition of high concentrations of crude oil to sediment incubations.

Predominant metabolisms of seabed thermophiles may represent the functional potential that they contribute within oil reservoir microbiomes. Complete removal of sulfate in incubations of sediment from the Gulf of Mexico, coupled with high average abundance of *Firmicutes*, shows that sulfate reduction is an important energy conservation process for EGM spore-forming populations, whereas thermophiles enriched/detected in seep sediments from Nova Scotia appear to be less dominated by sulfate reducers. The latter suggests other metabolisms such as fermentation are likely important for energy conservation. In line with this, highly abundant species affiliated with clades that include known sulfate-reducing bacteria such as *Desulfallas-Sporotomaculum* and *Desulfitibacter*, and putative fermenters like *Caloranaerobacter* and *Tepidimicrobium* were generally predominant in EGM and NS heated sediment libraries,

respectively (**Table S4.5**). Notwithstanding variable germination dynamics of endospore populations with different metabolisms (Volpi *et al.* 2017), sulfate reduction and fermentation may reflect important metabolisms in the petroleum reservoirs that these thermophilic bacteria are dispersed from. These experiments therefore potentially provide a window into deep biosphere metabolism, overcoming challenges associated with sampling these deeply-buried, inaccessible environments.

The stoichiometric discrepancy between formate removal and sulfate reduction in the Gulf of Mexico sediment incubation without crude oil addition (**Table S4.4; Eq. 4.1**) suggests the utilisation of electron donors other than the organic acids supplied in the medium. Hydrocarbons are an obvious candidate substrate in sediments sampled from hydrocarbon seeps, as well as in petroleum reservoirs. Although generally more refractory than low molecular weight organic acids, an oil reservoir derived microbial community may be primed to degrade these compounds. Vigneron *et al.* (2017b) noted that members of the order *Clostridiales* — which were abundant in the incubations performed here (**Table S4.5**) — are adapted to living on hydrocarbons and complex refractory organic matter. Alternative to direct oxidation of hydrocarbons by sulfate-reducing bacteria, fermentative biodegradation of hydrocarbon substrates may result in the production of hydrogen that can serve as an electron donor for sulfate reduction. Organic acids other than the ones measured here may also be fermentation products of crude oil degradation in this way. Frequent detection of *Firmicutes* in high temperature subsurface oil fields (Hubert *et al.* 2012) that are genetically capable of fermentative, sulfate-reducing, and syntrophic metabolisms (*Chapter 2*) support these findings.

The applicability of seep indicator species identified in one location to predict hydrocarbon seepage in a new location was tested by estimating the prevalence of organisms with a high degree of relatedness ($\geq 95\%$). Total abundance differences of related sequences between seep sites and non-seep sites were observed (18 and 26% higher abundance in seep sites), although this comparison did not meet a significance threshold of $P < 0.05$. Explanations for this could be technical or ecological. Different geochemistry approaches to defining the occurrence of thermogenic hydrocarbons were used to categorize EGM and NS samples. This correspondingly influenced the definition of indicator species in one region and subsequent assessments of close relatives in a different region. One way to improve a bioassay for the detection of seepage in new

prospective areas could be to select a subset of indicator species showing the greatest abundance difference between seep and non-seep sites, such as those affiliated BRH-c8a and *Desulfallas-Sporotomaculum* (e.g., **Figure 4.5b**). Another factor potentially affecting indicator species comparisons between the two locations are different experimental conditions, e.g., varying incubation times and temperatures, leading to enrichment of different thermophile populations, which would again influence these ASV-based comparisons (**Figure 4.3; Table S4.7**).

An alternative explanation for comparisons not meeting a significance threshold between quantitative detection in positive and negative sites could be that endospores are ideally equipped for passive dispersal in bottom water currents, as depicted in **Figure 3.4** of the previous chapter. The Gulf of Mexico is a prolific basin containing large volumes of petroleum. The region has been described as an archetype for hydrocarbon seepage based on the geologic history and setting being suited for the upward migration of petroleum (Kennicutt, 2007). Based on the density and prominence of these seeps, the weaker statistical significance between positive and negative sites may be attributable spores of indicator species being dispersed to sediments that lack detectable thermogenic hydrocarbons. Refinement of the bioassay to quantify this dispersal and biogeographical patterns in the distribution of thermophilic endospores highlights a potential utility of these seep associated species as non-regional bioindicators of cold seeps, but also as proximity markers extending the detectable signal of hydrocarbon seepage beyond that measured by geochemical testing.

4.4 Methods

4.4.1 Sediment and crude oil sample descriptions

Seabed surface sediments were collected from the Mississippi Canyon area of the Eastern Gulf of Mexico in February 2011 aboard the *R/V GeoExplorer* (labelled here as EGM) and from the Scotian Slope, offshore Nova Scotia, Canada in May 2018 aboard the CCGS *Hudson* (labelled here as NS) (**Table S4.1**). EGM and NS sediments were collected by gravity coring in 2,929 and 2,405 m water depth, respectively. For geochemical assessment to confirm hydrocarbon seepage at these locations, sub-samples from deeper sediment layers, i.e., 2.0–5.5 m and 0.32–0.43 m below

seafloor, were collected from three core intervals at the EGM and NS sites, respectively. Sediment samples for microbiological analysis were taken from the top of the cores (0–20 cm mbsf) and stored in either a sterile Whirl-Pak bag at -20°C (EGM) or in a sterile glass jar at 4°C (NS). Crude oil was collected from the Mizzen O-16 well in the Flemish Pass Basin, offshore Newfoundland and Labrador, Canada (**Table S4.1**). The crude oil sample was taken by a drill stem test from a formation at 3,213–3,224 meters measured depth below seabed (mMD), corresponding to an approximate *in situ* temperature of 85°C in the reservoir.

4.4.2 Hydrocarbon geochemical analysis

EGM sediments and NS sediments were analysed using different approaches in different labs. Geochemical analyses of EGM samples were performed by TDI-Brooks International Inc. Additional details of these methods are provided in Bernard *et al.* (2008). Geochemical analyses of NS samples were performed by Applied Petroleum Technology, Norway.

EGM sediments were analysed by measuring total scanning fluorescence (indicative of aromatic hydrocarbon compounds), unresolved complex mixture (UCM) concentrations, and distributions of C₁₅₊ hydrocarbons in sediment solvent extracts from each of the tested core intervals. Organic matter was extracted from 15 g of pre-dried sediment using a Dionex ASE200 Accelerate Solvent Extractor (Thermo Scientific, USA), and the extracts concentrated to a final volume of 8 mL using a TurboVap II Concentration Workstation (Zymark, USA). A three-dimensional fluorescence spectrum was acquired for each extract using a LS-50B Fluorescence Spectrometer (PerkinElmer, USA). The extracts were further concentrated by solvent reduction, an internal standard added, and injected into a capillary gas chromatography column. The UCM concentration was measured as the total area under the chromatogram between the retention time of *n*C₁₅ and *n*C₃₄, minus the total area of all resolved analytes. C₁₅₊ hydrocarbon concentrations were calculated based on the concentration and response of internal standard compounds. A thermogenic/diagenetic ratio assessing the abundance of recent plant wax derived odd carbon number *n*-alkanes ($(\sum n\text{-alkanes} - (nC_{27} + nC_{29} + nC_{31} + nC_{33})) / (nC_{27} + nC_{29} + nC_{31} + nC_{33})$) was used to further assess hydrocarbon origin. Standard thresholds for the presence of migrated thermogenic hydrocarbons in sediment

extracts around the world are a UCM ≥ 10 $\mu\text{g/g}$ and TSF $\geq 10,000$ max. intensity units (Bernard *et al.* 2008). To account for the large input of recent terrigenous organic matter supplying aromatic compounds to the Gulf of Mexico, a five-fold increase in the TSF threshold (50,000 max. intensity units) was used to confirm the unambiguous presence of thermogenic hydrocarbons (Chakraborty *et al.* 2018). TSF ($\geq 50,000$ max. intensity units) and UCM (≥ 10 $\mu\text{g/g}$) thresholds were applied to confirm seepage at EGM sites.

NS sediments were analysed for gas and liquid hydrocarbon compounds. Interstitial gas analysis was performed on aliquots of IsoJar[®] (Isotech Laboratories Inc., USA) headspace transferred into Exetainers[®] (Labco Limited, UK). Sample volumes of 1 mL were injected into an Agilent 7890 RGA Gas Chromatograph (Agilent Technologies, USA). A flame ionisation detector determined C₁–C₅ hydrocarbon gas concentrations allowing calculation of gas wetness ($(\sum C_{2-n}C_4) / (\sum C_{1-n}C_4) * 100$). Carbon isotopic composition ($\delta^{13}\text{C}$) of gas components was determined by gas chromatography combustion isotope ratio mass spectrometry; headspace aliquots were analyzed on a Trace 1310 Gas Chromatograph (Thermo Fisher Scientific, USA) interfaced to a Delta V Isotope Ratio Mass Spectrometer (Thermo Fisher Scientific, USA). Organic matter was extracted from sediment by adding dichloromethane with 7% (v/v) methanol. A subsample of the extractable organic matter (EOM) was evaporated to dryness and weighed. Asphaltenes were removed from a separate subsample of EOM by pentane addition, the solution filtered, and the asphaltenes weighed. The de-asphalted EOM was separated into saturate, aromatic, and polar fractions using medium pressure liquid chromatography with a Kieselgel 100 pre-column and LiChroprep Si60 main column, and each fraction weighed. Sediment extracts were analyzed for hydrocarbon biomarkers — compounds derived from a biological precursor that can be used to assess origin and thermal maturity — in subsamples where sufficient extract yields were recovered. Gas chromatography analysis of the EOM was performed on an Agilent 7890A Gas Chromatograph (Agilent Technologies, USA). Saturate and aromatic hydrocarbon fractions were analyzed further using a Micromass ProSpec Gas Chromatography-Mass Spectrometer (Waters Corporation, USA).

4.4.3 Crude oil-amended sediment incubations

Germination and activity of dormant bacterial endospores from sediments and crude oil was investigated, as well as the extent to which additional crude oil and triple-autoclaving selects for hydrocarbon seep associated populations present in either the sediment or the oil. Seven experimental treatments were established and a negative control containing medium only. The experimental set up for the different incubations is summarized in **Table 4.1**. Treatment 7 has been previously shown to enrich reservoir associated thermophilic endospore-forming bacteria (Chakraborty *et al.* 2018; *Chapter 3*).

Incubations were prepared in triplicate using sterile 50 mL serum bottles with anoxic, synthetic seawater medium (Widdel and Bak, 1992) amended with 20 mM of sulfate and 5 mM each of six organic acids, and an N₂:CO₂ (90 : 10 %) headspace as described in *Chapter 3*. All microcosms had an approximate final volume of 30 mL. Microcosms containing sediments were prepared by transferring 15 g of sediment to three separate serum bottles (resulting in two sets of triplicates) inside a 4°C walk-in fridge, before sealing and exchanging the headspace to make the bottles anoxic. Sediments for triple-autoclave treatments were autoclaved three times, with incubation at 50°C for 12 h between each autoclave run. In addition to amendment with medium, 0.3 mL of either crude oil or triple-autoclaved crude oil (prepared using the same method as for triple autoclaving of sediment) was introduced to certain microcosms in the ratios specified in **Table 4.1**. The negative control was prepared with 30 mL of medium only. All microcosms were pasteurized at 80°C for 1.5 h to select for heat-resistant endospores and immediately incubated at 50°C for 14 days. Subsamples (2 mL) were removed using sterile N₂:CO₂-flushed syringes and stored at -20°C for chemical and molecular analysis.

Final concentrations of crude oil were on the order of 1% volume/volume (v/v) in all oil-amended microcosms. In non-oil amended sediment microcosms, the average concentration of hydrocarbons (i.e., saturate and aromatic fractions) was 0.058 mg / g of extractable organic matter (**Table S4.2**; measurements for Nova Scotia sediment). The 15 g of sediment in these microcosms was accordingly estimated to comprise ~0.87 mg of hydrocarbons. Based on the estimated relative density of hydrocarbon compounds in crude oil (0.88 g / mL) and the approximate final volume of each microcosm (30 mL), the concentration of hydrocarbons in non-oil amended sediment

microcosms was estimated to be approximately 26 ppm v/v, or ~400 times lower than oil amended microcosms. Although concentration measurements were not available, sediment from the Eastern Gulf of Mexico is anticipated to have similar hydrocarbon levels.

4.4.4 Sulfate and organic acid measurements

Indications of microbial activity were assessed by measuring changes in sulfate and organic acid concentrations before incubation (i.e., immediately after 80°C pasteurization) and after 14 days of incubation at 50°C. Subsamples were centrifuged (14,000 g, 5 min) and the supernatant filtered through 13 mm, 0.22 µm pore size syringe filters. Sulfate concentrations were measured using a Dionex ICS-5000 reagent-free ion chromatography system (Thermo Scientific) equipped with an anion-exchange column (Dionex IonPac AS22; 4 x 250 mm; Thermo Scientific), an EGC-500 K₂CO₃ eluent generator cartridge, and a conductivity detector. Concentrations of six organic acids (acetate, butyrate, formate, lactate, propionate, and succinate) were measured using reversed-phase separation followed by UV detection (210 nm) on a Dionex Ultimate-3000 HPLC system (Thermo Scientific) equipped with an Aminex HPX-87H column (9 µm; 7.8 x 300 mm; Bio-Rad). The eluent was 5 mM H₂SO₄ with a pH of 2.3–2.4.

4.4.5 16S rRNA gene amplicon sequencing

Genomic DNA was extracted from triplicate incubation subsamples taken immediately before incubation (i.e., post-pasteurization) and after 14 days of incubation using the DNeasy PowerLyzer PowerSoil Kit (Qiagen, USA). Extractions were performed on 300 µL of subsample according to the manufacturer's protocol, except for an additional 10 min incubation at 70°C immediately after the addition of Solution C1 to improve the efficiency of cell lysis. Extraction blanks (Milli-Q water) were processed in parallel. DNA was quantified using the Qubit dsDNA High Sensitivity assay kit on a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Canada). The V3 and V4 hypervariable regions of the 16S rRNA gene were amplified in triplicate PCR reactions per DNA extract using the bacteria-specific primer pair SD-Bact-341-bS17/SD-Bact-785-aA21 (Klindworth

et al. 2013) modified with Illumina MiSeq overhang adapters. All PCR reactions were performed in triplicate and confirmed for positive amplification using agarose gel electrophoresis. All DNA extraction and PCR reagent blanks were confirmed for negative amplification. As a result of an absence of quantifiable DNA and no positive amplification, subsamples from the medium-only negative control (**Table 4.1**) were not analysed further. Triplicate PCR products were pooled, purified using a NucleoMag NGS Clean-up and Size Select kit (Macherey-Nagel Inc., USA) and indexed. Sizes of indexed amplicons were verified using the High Sensitivity DNA kit on an Agilent 2100 Bioanalyzer system (Agilent Technologies, Canada). Indexed amplicons were pooled in equimolar amounts and sequenced on an in-house Illumina MiSeq benchtop sequencer (Illumina Inc., USA) using Illumina's v3 600-cycle reagent kit to obtain 300 bp paired-end reads.

4.4.6 Sequence processing

A total of 3.4 million raw paired-end reads were generated from 72 sequence read libraries across three separate MiSeq runs performed for this study. Primers were trimmed using Cutadapt version 2.747 (Martin, 2011) prior to amplicon sequence variant (ASV) inference using DADA2 version 1.1648 (Callahan *et al.* 2016) in base R version 3.6.1 (R Core Team, 2014). Forward and reverse read pairs were trimmed to a run-specific length based on a minimum quality score of 25. Read pairs were filtered allowing no ambiguous bases and requiring each read to have less than two expected errors, and PhiX sequences removed. Reads were dereplicated providing unique sequences and their corresponding abundance. Error rates were estimated from sequence composition and quality by applying a core denoising algorithm for each of the three sequencing runs to account for run-to-run variability. Unique ASVs were inferred independently from the forward and reverse reads of each sample, using the run-specific error rates, and then pairs were merged if they overlapped with no mismatches. Chimeras were identified and removed, then an additional length trimming step removed sequence variants shorter than 400 nucleotides and larger than 435 nucleotides. Taxonomic classification to the genus level was performed using Mothur version 1.41.3 (Schloss *et al.* 2009) and Ribosomal Database Project's k-mer-based naïve Bayesian classifier (Wang *et al.* 2007) with the Silva version 138 database (Quast *et al.* 2013). Singleton ASVs, ASVs with no kingdom level taxonomic classification (likely artifacts of sequence

processing), and ASVs classified as *Vertebrata*, *Mitochondria* and *Chloroplast* were removed. In addition, samples with <1,000 reads were removed prior to community profiling, beta diversity calculations and statistical analyses. To reduce the loss of data, subsampling was only implemented for alpha diversity calculations. Random subsampling was performed without replacement to 1,000 reads using the *phyloseq* R package (McMurdie and Holmes, 2013). As a function of quality control sequence processing, all crude oil and medium-only (Treatment 3) libraries, plus one library from each of Treatments 1, 2, 4, and 6 were removed due to low read counts (<1,000 qc reads), leaving 62 sequence read libraries in the dataset. A total of 4,858 ASVs were resolved from 1.1 million quality-controlled reads in non-subsampled libraries.

4.4.7 Comparative sequence analysis

For identification of seep associated species (cf. *Statistical analysis* below) and comparative analysis of their presence in Eastern Gulf of Mexico and offshore Nova Scotia seabed sediments, sequence data from high temperature sediment incubations reported in Chakraborty *et al.* (2018) and *Chapter 3*, respectively, was downloaded and analysed. Raw sequence data was accessed from the National Center for Biotechnology Information's (NCBI) Sequence Read Archive (SRA; Leinonen *et al.* 2011) by compiling sequence accession lists from BioProject accession numbers PRJNA415828 and PRJNA604781 and implementing the *prefetch* and *fastq-dump* commands from the SRA Toolkit. ASV inference and taxonomic classification were performed using the same workflow described above for the 16S rRNA gene amplicon sequences generated in the present study. Random subsampling was performed to the minimum number of reads in the libraries of the respective studies; accordingly, Gulf of Mexico libraries were subsampled to 9,434 sequences and Nova Scotia libraries subsampled to 4,635 sequences.

4.4.8 Statistical analysis

Statistical analyses and visualization were performed using base R version 3.6.1 (R Core Team, 2017) or the specific R packages as indicated below. Richness and alpha diversity metrics on

randomly subsampled libraries were calculated using *phyloseq* (McMurdie and Holmes, 2013). Non-metric multidimensional scaling (NMDS) of Bray-Curtis distance dissimilarity was calculated using the *metaMDS* function of the *vegan* package (Oksanen *et al.* 2007) and visualized using the *ggplot2* package (Wickham, 2009). The statistical difference between sample community dissimilarities was assessed using Permutational Multivariate Analysis of Variance (PERMANOVA) tests in *vegan* (Oksanen *et al.* 2007). The association of an ASV with experimental conditions or geoscience indicators of hydrocarbon seepage at a site or a group of sites (thereby contributing to the definition of ‘seep associated species’) was assessed through multilevel pattern analysis using the *multipatt* function of the *indicspecies* package, employing a point-biserial correlation index (De Cáceres *et al.* 2010).

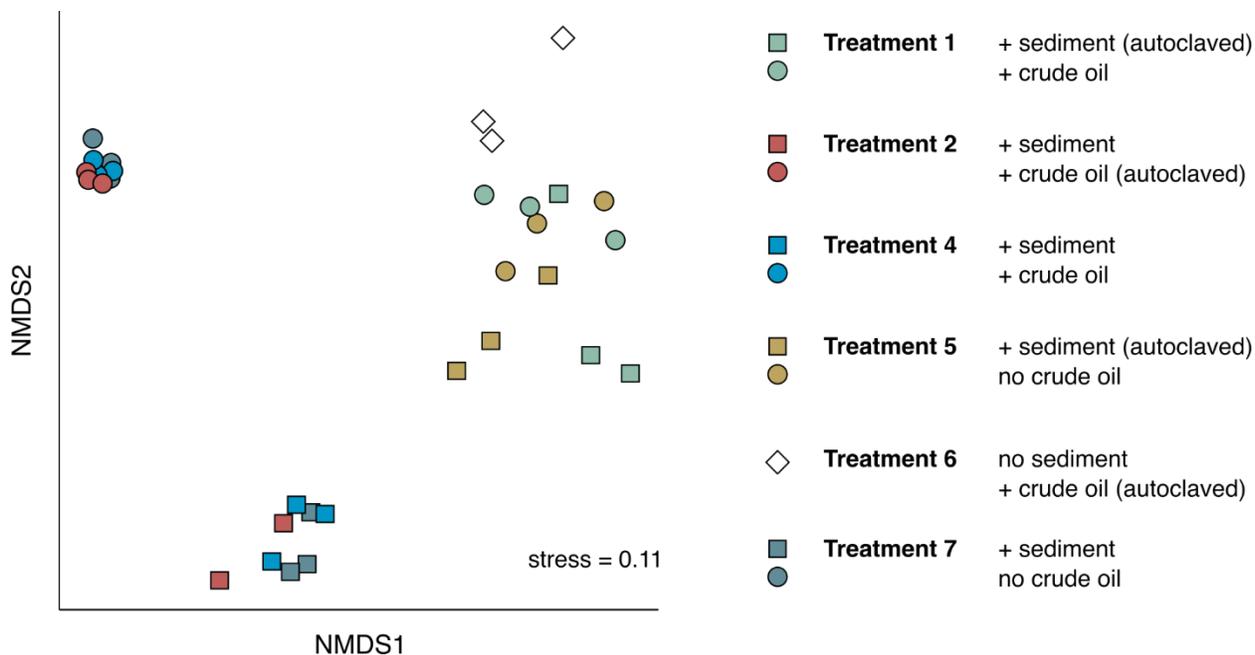


Figure S4.1. Microbial community composition under different experimental conditions. Nonmetric multidimensional scaling of the Bray-Curtis dissimilarity in microbial community composition after 14 days of incubation under the respective treatment conditions. Squares and circles represent sediment from EGM and NS, respectively, with the exception of Treatment 6 (no sediment) denoted by the white diamond. Statistically significant differences in microbial communities were observed between the different treatment conditions (PERMANOVA, r^2 0.35, $P < 0.05$).

Conclusion

5.1 Summary

The work presented in this thesis investigated deep biosphere environmental selection and the recent dispersal patterns of thermophilic endospores in cold seabed sediments.

Chapter 2 provided a systematic assessment of the composition and functional potential of the petroleum reservoir microbiome. The study compared 343 16S rRNA gene amplicon libraries and 25 shotgun metagenomic libraries from oil reservoirs around the world. Taxonomic analysis revealed the absence of a core microbiome, with diversity and compositional variance observed among reservoirs with different *in situ* temperatures and between reservoirs in different geographic locations. Despite these taxonomic variances, gene-centric analysis of metagenomes showed that a functional core microbiome exists. Common metabolic potentials for energy conservation are consistent with other deep biosphere environments, e.g., prevalent sulfate reduction and methanogenesis pathways, and are conserved across diverse reservoir conditions. Metagenomic analysis also highlighted how microbial populations in petroleum reservoirs are genetically capable of responding to changes in redox biogeochemistry associated with oil production activities, illustrated by the lower taxonomic diversity in reservoirs undergoing secondary production.

Microbiomes in hydrocarbon-rich petroleum reservoirs present an alternative to deep biosphere settings in most subseafloor sediments where life becomes increasingly constrained by energy limitation. Petroleum reservoir ecosystems appear to be a biogeochemical ‘hotspot’ for microbial activity in the subsurface, with populations benefitting from deeper thermal transformations of organic matter into petroleum. Abundant *Firmicutes* and genetic potential for sporulation — shown by the widespread occurrence of genes encoding Spo0A (master transcriptional regulator for sporulation), DPA (dipicolinic acid), and SASP (small, acid-soluble spore proteins) —

indicates spore-forming microbes are prevalent in these ecosystems. This provided a basis for investigation of cells capable of long-term and long-distance dispersal being transported from petroleum reservoirs to the surface biosphere by geological processes.

Chapter 3 explored the biogeography of microbial populations in the marine sedimentary biosphere. Using thermophilic endospore-forming bacteria in the cold seabed, microbial dispersal and environmental selection were investigated revealing new insights into subsurface ecology. Geophysical imaging and geochemistry revealed deep-to-shallow geofluid conduits and chemical signals indicating migration of thermogenic hydrocarbons in the Scotian Slope, off Canada's eastern coast in the NW Atlantic. Microbiological investigation, using high temperature incubation of sediment samples and 16S rRNA gene amplicon sequencing allowed sites classified according to the strength of the geoscience indicators to be compared. This revealed correlations of certain thermophilic endospore-forming taxa with hydrocarbon seepage. The prevalence of these bacteria in deep petroleum systems was assessed by taxonomic comparisons with the dataset compiled in Chapter 2, confirming that closely related lineages are abundant members of the global oil reservoir microbiome. Accordingly the seabed taxa represent candidate 'indicator species' for hydrocarbon seepage out of deeply buried petroleum systems.

Genome centric analysis of seep-correlated species unveiled in this study indicate the presence of glycy radical enzymes related to alkylsuccinate synthases — proposed to mediate alkane biodegradation via addition to fumarate — highlighting a potential for anaerobic biodegradation of hydrocarbons. This metabolism would favour selection in petroleum reservoirs by facilitating growth on abundant petroleum compounds as an energy source, further supporting the hypothesis that these bacteria originate in oil reservoir environments. Sporulation capability of these same indicator species is consistent with observations of endospores of thermophilic bacteria in the overlying permanently cold seabed arriving there after large scale dispersal from warm petroleum reservoir habitats. Spores dispersed in this way are deposited in surface sediments and buried, as shown by DPA concentrations exceeding the global seabed average. Burial by sedimentation represents a dispersal vector that returns viable cells back into the deep biosphere, where some of these populations eventually encounter opportunities to 're-charge' upon encountering high energy organic compounds delivered to the cells via migration of petroleum into sedimentary geologic traps. Germination and growth here represent selection of the starting population of microbes

capable of another round of long-term dispersal, via seepage out of these oil reservoirs. Synthesis of this multidisciplinary geophysical, geochemical, microbiological, and genomic dataset highlights a ‘microbial dispersal loop’ circulating living biomass through the deep biosphere.

This advancement of deep biosphere science beyond comparisons of numbers of cells and spores presents an ecological framework for understanding microbial ecology within Earth’s largest habitat. By employing dormant spores in cold sediments this study system disentangles microbial dispersal and environmental selection, and observes them occurring in discrete phases of a cycle. Accordingly, this work enhances our understanding of the relationship between key ecological processes not just in the subsurface but in the rest of the biosphere where microbial biogeography questions remain challenging.

Chapter 4 adds another line of evidence to support the petroleum reservoir origin of cold seabed thermophiles reported in Chapter 3. A requisite physiology for bacteria originating in petroleum reservoirs is survival under the selective pressures exhibited by the *in situ* environmental conditions, namely high concentrations of crude oil and high temperature. A series of incubations where crude oil was amended to sediment sampled from sites of proven hydrocarbon seepage in the Eastern Gulf of Mexico and offshore Nova Scotia were established. Substrate concentration changes indicate microbial activity in the presence of crude oil amendment to the sediment. Selection of diverse populations of *Firmicutes* that are closely related to seep indicator species under these conditions shows that certain members of the dormant microbiome populating the seabed as endospores can germinate and become active in the presence of high concentrations of crude oil. This is consistent with the growth conditions for microbes inhabiting petroleum reservoirs, providing validation of a system presented in Chapter 3 in which cold seabed thermophiles detected at seabed hydrocarbon seeps are dispersed from deep biosphere petroleum reservoirs.

Evidence supporting a subsurface petroleum reservoir origin of these thermophilic endospores highlights their utility as bioindicators of marine cold seeps. The results presented here suggest the discovery of seep indicator species and/or their detection can be enhanced by the addition of crude oil to experimental microcosms typically employed to investigate these populations. Comparing the prevalence of seep indicator species across 125 sampling sites from two hydrocarbon

prospective areas (Nova Scotia and Gulf of Mexico), revealed that seep detection bioassays based on this approach are valid in new areas of the deep sea, while thermophilic endospores may also serve as proximity markers extending the detectable signal of hydrocarbon seepage beyond that measured by geochemical testing.

5.2 Outlook

A primary aim of microbial ecology research is to determine patterns and drivers of community distribution, interaction, and assembly across diverse ecosystems. Advances in DNA sequencing technologies have transformed our capability to explore complex microbial communities. These developments have led to rapid advances in microbiome studies in many fields of science, which in turn has rapidly advanced our awareness of the microbial world's importance. Coinciding with this ever-increasing knowledge, microbiome research is increasingly contributing to novel biotechnology applications underpinned by discoveries from basic science.

The research presented in this thesis represents an in-depth investigation of microbial biogeography. Environmental selection and large-scale dispersal were explored. Multidisciplinary techniques were applied, and the results synthesized to better understand the interactions of microorganisms with their environment as well as patterns in their distributions. The studies describe a methodological approach to arrive at a better understanding of ecological processes that can be employed to investigate new microbiomes in new environments. Opportunities to apply this approach exist in both natural and engineered systems. As an example, any effective response for avoiding the effects of climate change will require multiple large-scale solutions. Hydrogen has emerged as a key future, low-carbon energy carrier. To enable hydrogen as a renewable energy resource, inter-seasonal and longer-term storage solutions that can balance the intermittency of supply and demand are required. A potential solution is hydrogen storage in suitable geological formations. Depleted petroleum reservoirs that have previously produced oil and gas are one option for this, however, assessments of subsurface microbiome interactions with stored hydrogen will have important implications for the viability of this approach. Genomic potential for hydrogen consumption has been shown in this thesis to be a widespread feature of the oil reservoir microbiome. Examining how this translates to rates of hydrogen oxidation and predicting the

products of this metabolism are important considerations. Also important for the success of subsurface hydrogen storage will be evaluating and continuously monitoring the sealing capacity of reservoir traps and caprocks to leakage. Hydrocarbon seeps have been demonstrated in this thesis as a vector for microbial dispersal out of the subsurface. Identification of microbes dispersed from formations containing stored hydrogen could be used as an indication of the efficiency of storage. The same principles founded on scientific understanding of an ecosystem's microbiome are applicable to examining risks to long-term subsurface storage of carbon dioxide from carbon capture and storage technologies.

In exploring fundamental questions relating to microbial biogeography, the results presented in this thesis also demonstrate how bioinformatic tools can be used to mine microbial datasets to examine untested or otherwise unapparent features of complex microbiomes. One example of this is the conserved functional potential across gradients of petroleum reservoir depth and temperature. Insights such as these are difficult to identify without sufficiently large-scale individual studies or, as was fundamental here, assimilation of datasets from separate studies. The success of extracting useful information from diverse datasets is dependent on the availability of the data and the computational power to store and process it. To deal with rapid increases in the amount of sequencing data, inexpensive cloud computing resources may become increasingly important. Additionally, artificial intelligence, such as machine learning, will continue to offer a means to efficiently and accurately mine, process, and analyze available information.

The deep biosphere is one of the largest, yet least explored, microbial habitats on the planet. Microbial populations here possess extensive physiological and metabolic diversity. Continued fundamental and applied research of the deep biosphere and other environments has the potential to deliver novel discoveries as well as innovative solutions to societal challenges in energy, the environment and beyond.

References

- Abrams MA. (2005) Significance of hydrocarbon seepage relative to petroleum generation and entrapment. *Mar. Pet. Geol.* **22**, 457–477.
- Abrams MA. (2020) Marine seepage variability and its impact on evaluating the surface migrated hydrocarbon seep signal. *Mar. Pet. Geol.* **121**.
- Ainsworth T, *et al.* (2015) The coral core microbiome identifies rare bacterial taxa as ubiquitous endosymbionts. *ISME J.* **9**, 2261–2274.
- Aitken CM, *et al.* (2013) Evidence that crude oil alkane activation proceeds by different mechanisms under sulfate-reducing and methanogenic conditions. *Geochim. Cosmochim. Acta* **109**, 162–174.
- Al-Hinai MA, Jones SW, Papoutsakis ET. (2015) The *Clostridium* sporulation programs: diversity and preservation of endospore differentiation. *Microbiol. Mol. Biol. Rev.* **79**, 19–37.
- Allen EE, Bartlett DH. (2002) Structure and regulation of the omega-3 polyunsaturated fatty acid synthase genes from the deep-sea bacterium *Photobacterium profundum* strain SS9. *Microbiol.* **148**, 1903–13.
- Alsaker KV, Papoutsakis ET. (2005) Transcriptional program of early sporulation and stationary-phase events in *Clostridium acetobutylicum*. *J Bacteriol.* **187**, 7103–7118.
- Amend JP, Shock EL. (2001) Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS Microbiol. Rev.* **25**, 175–243.
- An D, *et al.* (2013) Metagenomics of hydrocarbon resource environments indicates aerobic taxa and genes to be unexpectedly common. *Environ. Sci. Technol.* **47**, 10708–10717.
- Arp GK. (1992) Effusive microseepage: A first approximation model for light hydrocarbon movement in the subsurface. *AAPG Bulletin* **8**, 1–17.
- Astudillo-García C, *et al.* (2017) Evaluating the core microbiota in complex communities: a systematic investigation. *Environ. Microbiol.* **19**, 1450–62.
- Baas Beeking LGM. (1934) *Geobiologie of inleiding tot de milieukunde*. The Hague, the Netherlands: W.P. Van Stockum & Zoon.

- Bar-On YM, Phillips R, Milo R. (2018) The biomass distribution on Earth. *Proc. Natl. Acad. Sci.* **115**, 6506–6511.
- Barth T. (1991) Organic-acids and inorganic-ions in waters from petroleum reservoirs, Norwegian continental-shelf: a multivariate statistical-analysis and comparison with American reservoir formation waters. *Appl. Geochem.* **6**, 1–15.
- Becraft ED, *et al.* (2021) Evolutionary stasis of a deep subsurface microbial lineage. *ISME J.* **15**, 2830–2842.
- Bell E, *et al.* (2020) Sediment cooling triggers germination and sulfate reduction by heat-resistant thermophilic spore-forming bacteria. *Environ. Microbiol.* **22**, 456–465.
- Bennett B, *et al.* (2013) The controls on the composition of biodegraded oils in the deep subsurface - Part 3. The impact of microorganism distribution on petroleum geochemical gradients in biodegraded petroleum reservoirs. *Org. Geochem.* **56**, 94–105.
- Bernard BB, *et al.* (2008) Surface geochemical exploration and heat flow surveys in fifteen (15) frontier Indonesian basins. *Proceedings Indonesian Petroleum Association.*
- Beulig F, *et al.* (2022) Rapid metabolism fosters microbial survival in the deep, hot subseafloor biosphere. *Nat. Comm.* **13**, 1–9.
- Biddle JF, *et al.* (2006) Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru. *Proc. Natl. Acad. Sci.* **103**, 3846–3851.
- Biddle JF, *et al.* (2012) Prospects for the study of evolution in the deep biosphere. *Front. Microbiol.* **2**, 285.
- Bilstad T. (1992) Sulphate separation from seawater by nanofiltration. In *Produced Water* (pp. 503–509). Springer, Boston, MA.
- Birkeland NK, Schönheit P, Poghosyan L, Fiebig A, Klenk HP. (2017) Complete genome sequence analysis of *Archaeoglobus fulgidus* strain 7324 (DSM 8774), a hyperthermophilic archaeal sulfate reducer from a North Sea oil field. *Stand. Genom. Sci.* **12**, 1–9.
- Boetius A, Wenzhöfer F. (2013) Seafloor oxygen consumption fuelled by methane from cold seeps. *Nat. Geosci.* **6**, 725–734.
- Borrel G, *et al.* (2019) Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. *Nat. Microbiol.* **4**, 603–613.

- Bowles MW, Mogollón JM, Kasten S, Zabel M, Hinrichs KU. (2014) Global rates of marine sulfate reduction and implications for sub-sea-floor metabolic activities. *Science* **344**, 889–891.
- Bradley JA, Amend JP, LaRowe DE. (2018) Necromass as a limited source of energy for microorganisms in marine sediments. *Journal of Geophysical Research: J. Geophys. Res. Biogeosci.* **123**, 577–590.
- Bradley JA, *et al.* (2020) Widespread energy limitation to life in global subseafloor sediments. *Sci. Adv.* **6**.
- Brovarone AV, *et al.* (2020) Subduction hides high-pressure sources of energy that may feed the deep subsurface biosphere. *Nat. Comm.* **11**, 1–1.
- Burbulys D, Trach KA, Hoch JA. (1991) Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay. *Cell* **64**, 545–552.
- Burdige DJ. (2007) Preservation of organic matter in marine sediments: controls, mechanisms, and an imbalance in sediment organic carbon budgets? *Chem. Rev.* **107**, 467–485.
- Callahan BJ, *et al.* (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583.
- Campbell DC, MacDonald AWA. (2016) CCGS Hudson Expedition 2015-018 Geological investigation of potential seabed seeps along the Scotian Slope, June 25–July 9, 2015 (Geological Survey of Canada, Open File 8116).
- Campbell DC, Normandeau A. (2019) CCGS Hudson Expedition 2018-041: high-resolution investigation of deep-water seabed seeps and landslides along the Scotian Slope, offshore Nova Scotia, May 26–June 15, 2018 (Geological Survey of Canada, Open File 8567).
- Campbell DC. (2019) CCGS Hudson Expedition 2016-011, phase 2. Cold seep investigations on the Scotian Slope, offshore Nova Scotia, June 15–July 6, 2016 (Geological Survey of Canada, Open File 8525).
- Cano RJ, Borucki MK. (1995) Revival and identification of bacterial spores in 25-to 40-million-year-old Dominican amber. *Science* **268**, 1060–1064.
- Cario A, Gina O, Rogers K. (2019) Exploring the deep marine biosphere: challenges, innovations, and opportunities earth science. *Front. Earth Sci.* **7**, 225.
- Cario A, Grossi V, Schaeffer P, Oger PM. (2015) Membrane homeoviscous adaptation in the piezo-hyperthermophilic archaeon *Thermococcus barophilus*. *Front. Microbiol.* **6**, 1152.

- Carlson HK, Hubert CRJ. (2019) Mechanisms and monitoring of oil reservoir souring control by nitrate or perchlorate injection. *Microbial Communities Utilizing Hydrocarbons and Lipids: Members, Metagenomics and Ecophysiology* 225–249.
- Carothers WW, Kharaka YK. (1978) Aliphatic acid anions in oil-feld waters - implications for the origin of natural gas. *Amer. Assoc. Pet. Geol. Bull.* **62**, 2441–2453.
- Chakraborty A, *et al.* (2018) Thermophilic endospores associated with migrated thermogenic hydrocarbons in deep Gulf of Mexico marine sediments. *ISME J.* **12**, 1895–1906.
- Chakraborty A, *et al.* (2020) Hydrocarbon seepage in the deep seabed links subsurface and seafloor biospheres. *Proc. Natl. Acad. Sci.* **117**, 11029–11037.
- Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. (2019) GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927.
- Chen IM, *et al.* (2021) The IMG/M data management and analysis system v. 6.0: new tools and advanced capabilities. *Nucleic Acids Res.* **49**, 751–763.
- Chibwe L, *et al.* (2015) Aerobic bioremediation of PAH contaminated soil results in increased genotoxicity and developmental toxicity. *Environ. Sci. Tech.* **49**, 13889–13898.
- Christman GD, León-Zayas RI, Zhao R, Summers ZM, Biddle JF. (2020) Novel clostridial lineages recovered from metagenomes of a hot oil reservoir. *Sci. Rep.* **10**, 8048.
- Ciccarelli FD, *et al.* (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283–1287.
- Ciotoli G, Procesi M, Etiope G, Fracassi U, Ventura G. (2020) Influence of tectonics on global scale distribution of geological methane emissions. *Nat. Comm.* **11**, 1–8.
- Coleman J, *et al.* (2003) Oil in the sea III: Inputs, fates, and effects. *Committee on Oil in the Sea: Inputs and Effects, Ocean Studies Board and Marine Board, Divisions of Earth and Life Studies and Transportation Research Board, National Research Council.* National Academies Press, Washington, DC, USA.
- Connan J. (1984) Biodegradation of crude oils in reservoirs. In *Advances in Petroleum Geochemistry.* J Brooks, DH Welte (eds.). London: Academic Press, 299–330.
- Deptuck ME, Kendell KL. (2020) Atlas of 3D seismic surfaces and thickness maps, central and southwestern Scotian Slope.

D'Hondt S, Rutherford S, Spivack AJ. (2002) Metabolic activity of subsurface life in deep-sea sediments. *Science* **295**, 2067–2070.

D'hondt S, *et al.* (2015) Presence of oxygen and aerobic communities from seafloor to basement in deep-sea sediment. *Nat. Geosci.* **8**, 299–304.

Daniel RA, Errington J. (1993) Cloning, DNA sequence, functional analysis and transcriptional regulation of the genes encoding dipicolinic acid synthetase required for sporulation in *Bacillus subtilis*. *J Molec. Biol.* **232**, 468–483.

Darwin C. (1859) On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. London: J. Murray.

Das N, Chandran P. (2011) Microbial degradation of petroleum hydrocarbon contaminants: an overview. *Biotechnol. Res. Int.*

De Cáceres M, Legendre P, Moretti M. (2010) Improving indicator species analysis by combining groups of sites. *Oikos* **119**, 1674–1684.

de Rezende JR, Hubert CR, Røy H, Kjeldsen KU, Jørgensen BB. (2017) Estimating the abundance of endospores of sulfate-reducing bacteria in environmental samples by inducing germination and exponential growth. *Geomicrobiol. J.* **4**, 338–345.

de Rezende JR, *et al.* (2013) Dispersal of thermophilic *Desulfotomaculum* endospores into Baltic Sea sediments over thousands of years. *ISME J.* **7**, 72–84.

D'Hondt S, *et al.* (2004) Distributions of microbial activities in deep subseafloor sediments. *Science* **306**, 2216–2221.

D'Hondt S, *et al.* (2009) Subseafloor sedimentary life in the South Pacific Gyre. *Proc. Natl Acad. Sci.* **106**, 11651–11656.

Dick GJ. (2019) The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. *Nat. Rev. Microbiol.* **17**, 271–283.

Dolfing J, Larter SR, Head IM. (2008) Thermodynamic constraints on methanogenic crude oil biodegradation. *ISME J.* **4**, 442–452.

Dong X, Strous M. (2019) An integrated pipeline for annotation and visualization of metagenomic contigs. *Front. Genet.* **10**.

- Dong X, *et al.* (2019) Metabolic potential of uncultured bacteria and archaea associated with petroleum seepage in deep-sea sediments. *Nat. Comm.* **10**, 1–12.
- Dürre P, Hollergschwandner C. (2004) Initiation of endospore formation in *Clostridium acetobutylicum*. *Anaerobe* **10**, 69–74.
- Eden B, Laycock PJ, Fielder M. (1993) *Oilfield Reservoir Souring*. Sudbury: HSE Books.
- Ellegaard M, Ribeiro S. (2018) The long-term persistence of phytoplankton resting stages in aquatic ‘seed banks’. *Biol. Rev.* **93**, 166–183.
- Falkowski PG, Fenchel T, Delong EF. (2008) The microbial engines that drive earth's biogeochemical cycles. *Science* **320**, 1034–1039.
- Fang J, *et al.* (2017) Predominance of viable spore-forming piezophilic bacteria in high-pressure enrichment cultures from ~1.5 to 2.4 km deep coal-bearing sediments below the ocean floor. *Front. Microbiol.* **8**, 137.
- Fichtel J, Köster J, Rullkötter J, Sass H. (2007) Spore dipicolinic acid contents used for estimating the number of endospores in sediments. *FEMS Microbiol. Ecol.* **61**, 522–532.
- Finlay BJ. (2002) Global dispersal of free-living microbial eukaryote species. *Science* **296**, 1061–1063.
- Freund F, Dickinson JT, Cash M. (2002) Hydrogen in rocks: an energy source for deep microbial communities. *Astrobiology* **2**, 83–92.
- Fujita M, Losick R. (2005) Evidence that entry into sporulation in *Bacillus subtilis* is governed by a gradual increase in the level and activity of the master regulator Spo0A. *Genes Dev.* **19**, 2236–2244.
- Galperin MY. (2013) Genome diversity of spore-forming Firmicutes. *Microbiol. Spectr.* **1**.
- Galperin MY, *et al.* (2012) Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. *Environ. Microbiol.* **14**, 2870–2890.
- Gao P, *et al.* (2016). Spatial isolation and environmental factors drive distinct bacterial and archaeal communities in different types of petroleum reservoirs in China. *Sci. Rep.* **6**, 20174.
- Gieg LM, Davidova IA, Duncan KE, Suflita JM. (2010) Methanogenesis, sulfate reduction and crude oil biodegradation in hot Alaskan oilfields. *Environ. Microbiol.* **12**, 3074–3086.
- Gieg LM, Duncan KE, Suflita JM. (2008) Bioenergy production via microbial conversion of residual oil to natural gas. *Appl. Environ. Microbiol.* **74**, 3022–3029.

- Gittel A, Sørensen KB, Skovhus TL, Ingvorsen K, 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.
- Graham ED, Heidelberg JF, Tully BJ. (2018) Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME J.* **350**, 1–6.
- Gray ND, *et al.* (2009) Biogenic methane production in formation waters from a large gas field in the North Sea. *Extremophiles* **13**, 511–519.
- Gray ND, *et al.* (2011) The quantitative significance of Syntrophaceae and syntrophic partnerships in methanogenic degradation of crude oil alkanes. *Environ. Microbiol.* **13**, 2957–2975.
- Greening C, *et al.* (2016) Genomic and metagenomic surveys of hydrogenase distribution indicate H₂ is a widely utilised energy source for microbial growth and survival. *ISME J.* **10**, 761–777.
- Grossman AD, Losick R. (1988) Extracellular control of spore formation in *Bacillus subtilis*. *Proc. Natl. Acad. Sci.* **85**, 4369–4373.
- Gruber-Vodicka HR, Seah BK, Pruesse E. (2017) Seah, phyloFlash – Rapid SSU rRNA profiling and targeted assembly from metagenomes. *mSystems* **5**, 1–16.
- Hamady M, Knight R. (2009). Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Res.* **19**, 1141–1152.
- Hannah GC, Shore JA, Loder JW, Naimie CE. (2001) Seasonal circulation on the western and central Scotian Shelf. *J. Phys. Oceanogr.* **31**, 591–615.
- Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.* **10**, 497–506.
- Hanson CA, *et al.* (2019) Historical factors associated with past environments influence the biogeography of thermophilic endospores in Arctic marine sediments. *Front. Microbiol.* **28**, 245.
- Head IM, Gray ND, Larter SR. (2014) Life in the slow lane; biogeochemistry of biodegraded petroleum containing reservoirs and implications for energy recovery and carbon management. *Front. Microbiol.* **5**.
- Head IM, Jones DM, Larter SR. (2003) Biological activity in the deep subsurface and the origin of heavy oil. *Nature* **426**, 344–352.

- Heider J, Szalaniec M, Suenwoldt K, Boll M. (2016) Ethylbenzene dehydrogenase and related molybdenum enzymes involved in oxygen-independent alkyl chain hydroxylation. *J. Mol. Microbiol. Biotechnol.* **26**, 45–62.
- Heuer VB, *et al.* (2020) Temperature limits to deep seafloor life in the Nankai Trough subduction zone. *Science* **370**, 1230–1234.
- Hijmans RJ, Williams E, Vennes C. (2015) Geosphere: spherical trigonometry. *R package*. <https://CRAN.R-project.org/package=geosphere>.
- Hinrichs KU, *et al.* (2006) Biological formation of ethane and propane in the deep marine subsurface. *Proc. Natl. Acad. Sci.* **103**, 14684–14689.
- Hoch JA. (1993) Regulation of the phosphorelay and the initiation of sporulation in *Bacillus subtilis*. *Ann. Rev. Microbiol.* **47**, 441–465.
- Hoehler TM, Jørgensen BB. (2013) Microbial life under extreme energy limitation. *Nat. Rev. Microbiol.* **11**, 83–94.
- Hoehler TM. (2004) Biological energy requirements as quantitative boundary conditions for life in the subsurface. *Geobiol.* **2**, 205–215.
- Horsfield B, *et al.* (2006) Living microbial ecosystems within the active zone of catagenesis: implications for feeding the deep biosphere. *Earth Planet Sci. Lett.* **246**, 55–69.
- Horstad I, Larter SR. (1997) Petroleum migration, alteration and remigration within Troll Field. *Norwegian North Sea Bull Am Assoc Petrol Geol* **81**, 222–248.
- Hu P, *et al.* (2016) Genome-resolved metagenomic analysis reveals roles for candidate phyla and other microbial community members in biogeochemical transformations in oil reservoirs. *MBio* **7**.
- Hubert C, Judd A. (2010) Using microorganisms as prospecting agents in oil and gas exploration. In: Timmis KN, (ed.) *Handbook of Hydrocarbon and Lipid Microbiology*. Springer-Verlag: Berlin, Heidelberg, 2711–2725.
- Hubert C, *et al.* (2009) A constant flux of diverse thermophilic bacteria into the cold Arctic seabed. *Science* **325**, 1541–1544.
- Hubert C, 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.

- Hubert CR, *et al.* (2012) Massive dominance of Epsilonproteobacteria in formation waters from a Canadian oil sands reservoir containing severely biodegraded oil. *Environ. Microbiol.* **14**, 387–404.
- Hyatt D, *et al.* (2010) Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119.
- Ijiri A, *et al.* (2018) Deep-biosphere methane production stimulated by geofluids in the Nankai accretionary complex. *Sci. Adv.* **4**.
- Inagaki F, *et al.* (2015) Deep biosphere. Exploring deep microbial life in coal-bearing sediment down to ~2.5 km below the ocean floor. *Science* **349**, 420–424.
- Inoue A, Horikoshi K. (1989) A *Pseudomonas* thrives in high concentrations of toluene. *Nature* **338**, 264–266.
- Ireton K, Jin S, Grossman AD, Sonenshein AL. (1995) Krebs cycle function is required for activation of the SpoOA transcription factor in *Bacillus subtilis*. *Proc. Natl. Acad. Sci.* **92**, 2845–2849.
- Isaksen MF, Bak F, Jørgensen BB. (1994) Thermophilic sulfate-reducing bacteria in cold marine sediment. *FEMS Microbiol. Ecol.* **14**, 1–8.
- Jenner KA, Campbell DC, Barnett JM, Higgins J, Normandeau A. Piston cores and supporting high-resolution seismic data, CCGS Hudson Expedition 2015018, Scotian Slope, Canada. (Geological Survey of Canada, Open File 8637, In prep).
- Jones DM, *et al.* (2008) Crude-oil biodegradation via methanogenesis in subsurface petroleum reservoirs. *Nature* **451**, 176–180.
- Jones SE, Lennon JT. (2010) Dormancy contributes to the maintenance of microbial diversity. *Proc. Natl. Acad. Sci.* **107**, 5881–5886.
- Jørgensen BB, D'Hondt S. (2006) A starving majority deep beneath the seafloor. *Science* **314**, 932–934.
- Jørgensen BB, Kasten S. (2006) Sulfur cycling and methane oxidation. H.D. Schulz, M. Zabel (Eds.), *Marine Geochemistry*, Springer, Berlin (2006), pp. 271–309.
- Jørgensen BB, Marshall IPG. (2016) Slow microbial life in the seabed. *Ann. Rev. Mar. Sci.* **8**, 311–332.
- Jørgensen BB. (2006) Bacteria and marine biogeochemistry. H.D. Schulz, M. Zabel (Eds.), *Marine Geochemistry*, Springer, Berlin (2006), 169–206.

- Joye SB. (2020) The geology and biogeochemistry of hydrocarbon seeps. *Annu. Rev. Earth Planet Sci.* **48**, 205–231.
- Judd AG, Hovland M. (2007) *Seabed Fluid Flow: The Impact on Geology, Biology and the Marine Environment*. Cambridge University Press, Cambridge, 163–178.
- Judd AG. (2003) The global importance and context of methane escape from the seabed. *Geo-Mar. Lett.* **23**, 147–154.
- Kallmeyer J, Pockalny R, Adhikari R, Smith DC, D'Hondt S. (2012) Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc. Natl Acad. Sci.* **109**, 16213–16216.
- Kanehisa M, Sato Y, Morishima K. (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731.
- Kang DD, *et al.* (2019) MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ.* **7**.
- Kennicutt MC. (2017) Oil and gas seeps in the Gulf of Mexico. In *Habitats and biota of the Gulf of Mexico: Before the deepwater horizon oil spill 2017*. Springer, New York, NY, 275–358.
- Khelifi N, *et al.* (2014) Anaerobic oxidation of long-chain n-alkanes by the hyperthermophilic sulfate-reducing archaeon, *Archaeoglobus fulgidus*. *ISME J.* **8**, 2153–2166.
- Khot V, *et al.* (2022) CANT-HYD: A curated database of phylogeny-derived hidden markov models for annotation of marker genes involved in hydrocarbon degradation. *Front. Microbiol.* **12**, 764058.
- Kidston AG, Brown DE, Altheim B. Smith BM. (2002) Hydrocarbon potential of the deep-water Scotian Slope: Overview.
- Klein F, Bach W, McCollom TM. (2013) Compositional controls on hydrogen generation during serpentinization of ultramafic rocks. *Lithos* **178**, 55–69.
- Klindworth A, *et al.* (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucl. Acids Res.* **41**, 1–11.
- Lack A, Fuchs G. (1994) Evidence that phenol phosphorylation to phenylphosphate is the first step in anaerobic phenol metabolism in a denitrifying *Pseudomonas* sp. *Arch. Microbiol.* **161**, 132–139.
- LaRowe DE, Amend JP. (2016) The energetics of anabolism in natural settings. *ISME J.* **10**, 1285–1295.

- LaRowe DE, Burwicz E, Arndt S, Dale AW, Amend JP. (2017) Temperature and volume of global marine sediments. *Geology* **45**, 275–278.
- Larter S, *et al.* (2006) The controls on the composition of biodegraded oils in the deep subsurface: Part II – Geological controls on subsurface biodegradation fluxes and constraints on reservoir-fluid property prediction. *Am. Assoc. Pet. Geol. Bull.* **90**, 921–938.
- Larter SR, di Primio R. (2005) Effects of biodegradation on oil and gas field PVT properties and the origin of oil rimmed gas accumulations. *Org. Geochem.* **36**, 299–310.
- Laso-Pérez R, *et al.* (2019) Anaerobic degradation of non-methane alkanes by “Candidatus Methanoliparia” in hydrocarbon seeps of the Gulf of Mexico. *MBio* **10**.
- Leinonen R, Sugawara H, Shumway M. (2011) The Sequence Read Archive. *Nucleic Acids Res.* **39**, 19–21.
- Lennon JT, den Hollander F, Wilke-Berenguer M, Blath J. (2021) Principles of seed banks and the emergence of complexity from dormancy. *Nat. Comm.* **12**, 1–6.
- Lennon JT, Jones SE. (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat. Rev. Microbiol.* **9**, 119–130.
- Letunic I, Bork P. (2019) Interactive Tree of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, 256–259.
- Lever MA. (2011) Acetogenesis in the energy-starved deep biosphere—a paradox. *Front. Microbiol.* **2**, 284.
- Lever MA, *et al.* (2010) Acetogenesis in deep seafloor sediments of the Juan de Fuca Ridge Flank: a synthesis of geochemical, thermodynamic, and gene-based evidence. *Geomicrobiol. J.* **27**, 183–211.
- Lever MA, *et al.* (2015) Life under extreme energy limitation: a synthesis of laboratory- and field-based investigations. *FEMS Microbiol. Rev.* **39**, 688–728.
- Li D, Liu CM, Luo R, Sadakane K, Lam TW. (2015) MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676.
- Lipp JS, Morono Y, Inagaki F, Hinrichs K-U. (2008) Significant contribution of Archaea to extant biomass in marine subsurface sediments. *Nature* **454**, 991–94.

- Liu YF, *et al.* (2018) Metabolic capability and in situ activity of microorganisms in an oil reservoir. *Microbiome* **6**, 5.
- Lloyd KG, May MK, Kevorkian RT, Steen AD. (2013) Meta-analysis of quantification methods shows that archaea and bacteria have similar abundances in the subseafloor. *Appl. Environ. Microbiol.* **79**, 7790–7799.
- Lloyd KG, *et al.* (2013) Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496**, 215–218.
- Locey KJ, Fisk MC, Lennon JT. (2017) Microscale insight into microbial seed banks. *Front. Microbiol.* **7**, 2040.
- Lolkema JS, Speelmans G, Konings WN. (1994) Na⁺-coupled versus H⁺-coupled energy transduction in bacteria. *Biochim. Biophys. Acta* **1187**, 211–15.
- Lomstein BA, Jørgensen BB. (2012) Pre-column liquid chromatographic determination of dipicolinic acid from bacterial endospores. *Limnol. Oceanogr. Methods* **10**, 227–233.
- Lomstein BA, Langerhuus AT, D’Hondt S, Jørgensen BB, Spivack A. (2012) Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. *Nature* **484**, 101–104.
- Long S, Jones DT, Woods DR. (1984) The relationship between sporulation and solvent production in *Clostridium acetobutylicum* P262. *Biotechnol. Lett.* **6**, 529–534.
- Lopez JM, Marks CL, Freese E. (1979) The decrease of guanine nucleotides initiates sporulation of *Bacillus subtilis*. *Biochim. Biophys. Acta* **587**, 238–252.
- Lovley DR, *et al.* (1989) Oxidation of aromatic contaminants coupled to microbial iron reduction. *Nature* **339**, 297–300.
- Ludwig W, *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371.
- Lundberg DS, *et al.* (2012) Defining the core *Arabidopsis thaliana* root microbiome. *Nature* **488**, 86–90.
- Lynch MD, Neufeld JD. (2015) Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* **13**, 217–229.
- MacDonald IR. (2015) Natural and unnatural oil slicks in the Gulf of Mexico. *J. Geophys. Res. Oceans* **120**, 8364–8380.

- Magoon LB, Dow WG. (1994) The Petroleum System — From Source to Trap (Am. Assoc. 537 Petroleum Geologists, Tulsa, Mem. 60).
- Magot M, Ollivier B, Patel BKC. (2000) Microbiology of petroleum reservoirs. *Antonie Van Leeuwenhoek* **77**, 103–116.
- Marcia M, Ermler U, Peng G, Michel H. (2009) The structure of Aquifex aeolicus sulfide:quinone oxidoreductase, a basis to understand sulfide detoxification and respiration. *Proc. Natl. Acad. Sci.* **106**, 9625–9630.
- Mardanov AV, *et al.* (2009) Metabolic versatility and indigenous origin of the archaeon *Thermococcus sibiricus*, isolated from a siberian oil reservoir, as revealed by genome analysis. *Appl. Environ. Microbiol.* **75**, 4580–4588.
- Margosch D, *et al.* (2006) High-pressure-mediated survival of *Clostridium botulinum* and *Bacillus amyloliquefaciens* endospores at high temperature. *Appl. Environ. Microbiol.* **72**, 3476–3481.
- Martin M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12.
- Martiny JB, *et al.* (2006) Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102–112.
- Mayumi D, *et al.* (2011) Evidence for syntrophic acetate oxidation coupled to hydrogenotrophic methanogenesis in the high-temperature petroleum reservoir of Yabase oil field (Japan). *Environ. Microbiol.* **13**, 1995–2006.
- McKenney PT, Eichenberger P. (2012) Dynamics of spore coat morphogenesis in *Bacillus subtilis*. *Microbiol. Mol. Biol. Rev.* **83**, 245–260.
- McMurdie PJ, Holmes S. (2013) Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, 61217.
- Mestre M, Höfer J. (2021) The microbial conveyor belt: connecting the globe through dispersion and dormancy. *Trends Microbiol.* **29**, 482–492.
- Milkov AV. (2011) Worldwide distribution and significance of secondary microbial methane formed during petroleum biodegradation in conventional reservoirs. *Org. Geochem.* **42**, 184–207.
- Mills DJ, Vitt S, Strauss M, Shima S, Vonck J. (2013) De novo modeling of the F420-reducing [NiFe]-hydrogenase from a methanogenic archaeon by cryo-electron microscopy. *Elife*, **2**.

- Morono Y, *et al.* (2011) Carbon and nitrogen assimilation in deep seafloor microbial cells. *Proc. Natl. Acad. Sci.* **108**, 18295–18300.
- Morono Y, *et al.* (2020) Aerobic microbial life persists in oxic marine sediment as old as 101.5 million years. *Nat. Comm.* **11**, 3626.
- Morris BE, Henneberger R, Huber H, Moissl-Eichinger C. (2013) Microbial syntrophy: interaction for the common good. *FEMS Microbiol. Rev.* **37**, 384-406.
- Müller AL, *et al.* (2014) Endospores of thermophilic bacteria as tracers of microbial dispersal by ocean currents. *ISME J.* **8**, 1153–1165.
- Muyzer G, Stams AJ. (2008) The ecology and biotechnology of sulphate-reducing bacteria. *Nat. Rev. Microbiol.* **6**, 441–454.
- Nanda NC. (2016) Direct Hydrocarbon Indicators (DHI) in *Seismic data interpretation and evaluation for hydrocarbon exploration and production*. N. C. Nanda (Springer, Switzerland), pp. 103–113.
- Neu AT, Allen EE, Roy K. (2021) Defining and quantifying the core microbiome: Challenges and prospects. *Proc. Natl. Acad. Sci.* **118**.
- Nicholson WL, Munakata N, Horneck G, Melosh HJ, Setlow P. (2000) Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. *Microbiol. Mol. Biol. Rev.* **64**, 548–572.
- Niemann H, *et al.* (2006) Novel microbial communities of the Haakon Mosby mud volcano and their role as a methane sink. *Nature* **443**, 854–858.
- Oksanen J, Kindt R, Legendre P, O'Hara B. (2007) *The Vegan Package—Community Ecology Package*. R package version 2.0-9.
- Ollivier B, Magot M, Cunningham-Rundles S. (2005) *Petroleum microbiology*. Washington, DC: ASM Press.
- Olm MR, Brown CT, Brooks B, Banfield JF. (2017) dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868.
- Onstott TC, *et al.* (2014) Does aspartic acid racemization constrain the depth limit of the subsurface biosphere? *Geobiology* **12**, 1–19.

- Orcutt BN, *et al.* (2013) Microbial activity in the deep marine biosphere: progress and prospects. *Front. Extr. Microbiol.* **4**, 189.
- Oremland RS, Polcin S. (1982) Methanogenesis and sulfate reduction: competitive and noncompetitive substrates in estuarine sediments. *Appl. Environ. Microbiol.* **44**, 1270–1276.
- Orphan VJ, Taylor LT, Hafenbradl D, Delong EF. (2000) Culture-dependent and culture-independent characterization of microbial assemblages associated with high-temperature petroleum reservoirs. *Appl. Environ. Microbiol.* **66**, 700–711.
- Orsi WD, Edgcomb VP, Christman GD, Biddle JF. (2013) Gene expression in the deep biosphere. *Nature* **499**, 205–208.
- O'Sullivan LA, *et al.* (2015) Survival of *Desulfotomaculum* spores from estuarine sediments after serial autoclaving and high-temperature exposure. *ISME J.* **9**, 922–933.
- Palmer ES. (1993) “Effect of biodegradation and water washing on crude oil composition,” in *Organic Geochemistry Principles and Applications*, eds M. Engel and S. A. Macko (New York, NY: Plenum Press), 511–534.
- Parkes J, Maxwell J. (1993) Some like it hot (and oily). *Nature* **365**, 694–695.
- Parkes RJ, Cragg BA, Wellsbury P. (2000) Recent studies on bacterial populations and processes in subseafloor sediments: a review. *Hydrogeol. J.* **8**, 11–28.
- Parkes RJ, Maxwell JR. (1993) Some like it hot (and oily). *Nature* **365**: 694.
- Parkes RJ, *et al.* (1994) Deep bacterial biosphere in Pacific Ocean sediments. *Nature* **371**, 410–413.
- Parkes RJ, *et al.* (2005) Deep sub-seafloor prokaryotes stimulated at interfaces over geological time. *Nature* **436**, 390–394.
- Parkes RJ, *et al.* (2007) Temperature activation of organic matter and minerals during burial has the potential to sustain the deep biosphere over geological timescales. *Org. Geochem.* **38**, 845–852.
- Parkes RJ, *et al.* (2011) Prokaryotes stimulate mineral H₂ formation for the deep biosphere and subsequent thermogenic activity. *Geology* **39**, 219–222.
- Parkes RJ, *et al.* (2014) A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere:geosphere interactions. *Mar. Geol.* **352**, 409–425.

- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. (2015) CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055.
- Pedrós-Alío C. (2006) Marine microbial diversity: can it be determined? *Trends Microbiol.* **14**, 257–263.
- Piper DJW, Skene KI (1998). Latest Pleistocene ice-rafting events on the Scotian Margin (eastern Canada) and their relationship to Heinrich events. *Paleoceanography* **13**, 205–214.
- Pruesse E, Peplies J, Glöckner FO (2012) SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823–1829.
- Quast C, *et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl. Acids Res.* **41**, 590–596.
- R Core Team, R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria, 2014); www.R-project.org/.
- Rabus R, *et al.* (2016) Anaerobic microbial degradation of hydrocarbons: from enzymatic reactions to the environment. *J. Mol. Microbiol. Biotechnol.* **26**, 5–28.
- Rachel NM, Gieg LM. (2020) Preserving microbial community integrity in oilfield produced water. *Front. Microbiol.* **11**, 2536.
- Ramos JL, *et al.* (2002) Mechanisms of solvent tolerance in gram-negative bacteria. *Annual Reviews in Microbiology.* 2002 Oct;56(1):743-68.
- Rattray JE, *et al.* (2021) Sensitive quantification of dipicolinic acid from bacterial endospores in soils and sediments. *Environ. Microbiol.* **23**, 1397–1406.
- Rice GK. (2022) Vertical migration in theory and in practice. *Interpretation* **10**, 1.
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, 2584.
- Röling WF, Head IM, Larter SR. (2003) The microbiology of hydrocarbon degradation in subsurface petroleum reservoirs: perspectives and prospects. *Res. Microbiol.* **154**, 321–328.
- Roussel EG, *et al.* (2008) Extending the sub-sea-floor biosphere. *Science* **320** 1046.

- Røy H, *et al.* (2012) Aerobic microbial respiration in 86-million-year-old deep-sea red clay. *Science* **336**, 922–925.
- Rueter P, *et al.* (1994) Anaerobic oxidation of hydrocarbons in crude oil by new types of sulphate-reducing bacteria. *Nature* **372**, 455–458.
- Ruff SE, *et al.* (2015) Global dispersion and local diversification of the methane seep microbiome. *Proc. Natl. Acad. Sci.* **112**, 4015–4020.
- Ruff SE, *et al.* (2019) In situ development of a methanotrophic microbiome in deep-sea sediments. *ISME J.* **13**, 197–213.
- Sauer U, Santangelo JD, Treuner A, Buchholz M, Dürre P. (1995) Sigma factor and sporulation genes in *Clostridium*. *FEMS Microbiol. Rev.* **17**, 331–340.
- Sayers EW, *et al.* (2019) GenBank. *Nucleic Acids Res.* **47**, 94–99.
- Schink B. (1997) Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol. Mol. Biol. Rev.* **61**, 262–280.
- Schloss PD, *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541.
- Schrenk MO, Huber JA, Edwards KJ. (2010) Microbial provinces in the seafloor. *Annu. Rev. Mar. Sci.* **2**, 279–306.
- Schultz D, Wolynes PG, Jacob EB, Onuchic JN. (2009) Deciding fate in adverse times: Sporulation and competence in *Bacillus subtilis*. *Proc. Natl. Acad. Sci.* **106**, 21027–21034.
- Setlow B, Atluri S, Kitchel R, Koziol-Dube K, Setlow P. (2006) Role of dipicolinic acid in resistance and stability of spores of *Bacillus subtilis* with or without DNA-protective α/β -type small acid-soluble proteins. *J Bacteriol.* **188**, 3740–3747.
- Setlow P. (1988) Small, acid-soluble spore proteins of *Bacillus* species: structure, synthesis, genetics, function, and degradation. *Ann. Rev. Microbiol.* **42**, 319–38.
- Setlow P. (2014) Spore resistance properties. *Microbiol. Spec.* **2**, 2–5.
- Shade A, Handelsman J. (2011) Beyond the Venn diagram: the hunt for a core microbiome. *Environ. Microbiol.* **14**, 4–12.

- Sherry A, *et al.* (2013) Anaerobic biodegradation of crude oil under sulphate-reducing conditions leads to only modest enrichment of recognized sulphate-reducing taxa. *Int. Biodeter. Biodegr.* **81**, 105–113.
- Sierra-Garcia IN, *et al.* (2020) In depth metagenomic analysis in contrasting oil wells reveals syntrophic bacterial and archaeal associations for oil biodegradation in petroleum reservoirs. *Sci. Total. Environ.* **715**, 136646.
- Sikkema J, de Bont JA, Poolman B. (1995) Mechanisms of membrane toxicity of hydrocarbons. *Microbiol. Rev.* **59**, 201–222.
- Soffientino B, Spivack A, Smith DC, D'Hondt S. (2009) Hydrogenase activity in deeply buried sediments of the Arctic and North Atlantic oceans. *Geomicrobiol. J.* **26**, 537–545.
- Sogin ML, *et al.* (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere.” *Proc. Natl. Acad. Sci.* **103**, 12115–12120.
- Solomon EA, Kastner M, MacDonald IR, Leifer I. (2009) Considerable methane fluxes to the atmosphere from hydrocarbon seeps in the Gulf of Mexico. *Nat. Geosci.* **2**, 561–565.
- Sorg JA, Sonenshein AL. (2008) Bile salts and glycine as cogerminants for *Clostridium difficile* spores. *J. Bacteriol.* **190**, 2505–2512.
- Starnawski P, *et al.* (2017) Microbial community assembly and evolution in subseafloor sediment. *Proc. Natl. Acad. Sci.* **114**, 2940–2945.
- Stetter KO, *et al.* (1993) Hyperthermophilic archaea are thriving in deep North Sea and Alaskan oil reservoirs. *Nature* **365**, 743–745.
- Stevens TO, McKinley JP. (1995) Lithoautotrophic microbial ecosystems in deep basalt aquifers. *Science* **270**, 450–455.
- Stragier P, Losick R. (1996) Molecular genetics of sporulation in *Bacillus subtilis*. *Annu. Rev. Genet.* **30**, 297–341.
- Suess E. (2010) Marine cold seeps. In: Timmis, K. N. *Handbook of Hydrocarbon and Lipid Microbiology* (Springer-Verlag, Berlin).
- Summit M, Baross JA. (2001). A novel microbial habitat in the mid-ocean ridge subseafloor. *Proc. Natl. Acad. Sci.* **98**, 2158–2163.
- Sunagawa S, *et al.* (2020) Tara Oceans: towards global ocean ecosystems biology. *Nat. Rev. Microbiol.* **18**, 428–445.

- Takai K, *et al.* (2008) Cell proliferation at 122C and isotopically heavy CH₄ production by a hyperthermophilic methanogen under high-pressure cultivation. *Proc. Natl. Acad. Sci.* **105**, 10949–10954.
- Takano Y, *et al.* (2010) Sedimentary membrane lipids recycled by deep-sea benthic archaea. *Nat. Geosci.* **3**, 858–861.
- Thompson LR, *et al.* (2017) A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature* **551**, 457–463.
- Trembath-Reichert E, *et al.* (2017). Methyl-compound use and slow growth characterize microbial life in 2-km-deep seafloor coal and shale beds. *Proc. Natl. Acad. Sci.* **44**, 9206–9215.
- Turnbaugh PJ, *et al.* (2007) The Human Microbiome Project. *Nature* **449**, 804–810.
- Valentine DL. (2011) Emerging topics in marine methane biogeochemistry. *Ann. Rev. Mar. Sci.* **3**, 147–171.
- Van Hamme JD, Singh A, Ward OP. (2003) Recent advances in petroleum microbiology. *Microbiol. Molec. Biol. Rev.* **67**, 503–549.
- Vellend M. (2010) Conceptual synthesis in community ecology. *Q. Rev. Biol.* **85**, 183–206.
- Vigneron A, *et al.* (2017a) Comparative metagenomics of hydrocarbon and methane seeps of the Gulf of Mexico. *Sci. Rep.* **7**, 1–12.
- Vigneron A, *et al.* (2017b) Succession in the petroleum reservoir microbiome through an oil field production lifecycle. *ISME J.* **11**, 2141–2154.
- Volpi M, *et al.* (2017) Identity, abundance and reactivation kinetics of thermophilic fermentative endospores in cold marine sediment and seawater. *Front. Microbiol.* **8**, 131.
- Voordouw G, *et al.* (1996) Characterization of 16S rRNA genes from oil field microbial communities indicates the presence of a variety of sulfate-reducing, fermentative, and sulfide-oxidizing bacteria. *Appl. Environ. Microbiol.* **62**, 1623–1629.
- Vreeland RH, Rosenzweig WD, Powers DW. (2000) Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. *Nature* **407**, 897–900.
- Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007) Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267.

- Ward BA, Cael BB, Collins S, Young CR (2021) Selective constraints on global plankton dispersal. *Proc. Natl. Acad. Sci.* **118**.
- Watanabe M, Kojima H, Fukui M. (2015) *Limnochorda pilosa* gen. nov., sp. nov., a moderately thermophilic, facultatively anaerobic, pleomorphic bacterium and proposal of *Limnochordaceae* fam. nov. *Limnochordales* ord. nov. and *Limnochordia classis* nov. in the phylum Firmicutes. *Int. J. Syst. Evol. Microbiol.* **65**, 2378–2384.
- Wellsbury P, *et al.* (1997) Deep marine biosphere fuelled by increasing organic matter availability during burial and heating. *Nature* **388**, 573–576.
- Whitaker RJ, Grogan DW, Taylor JW. (2003) Geographic barriers isolate endemic populations of hyperthermophilic Archaea. *Science* **301**, 976–978.
- Whitman WB, Coleman DC, Wiebe WJ. (1998) Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci.* **95**, 6578–6583.
- Whittaker KA, Ryneerson TA. (2017) Evidence for environmental and ecological selection in a microbe with no geographic limits to gene flow. *Proc. Natl. Acad. Sci.* **114**, 2651–2656.
- Wickham H, Chang W. (2015) ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*. <https://ggplot2.tidyverse.org>.
- Wickham H. (2009) *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York); <http://ggplot2.org>.
- Widdel F, Bak F. (1992) Gram-negative mesophilic sulfate-reducing bacteria. In *The prokaryotes* (3352–3378). Springer, New York, NY.
- Wilhelms A, *et al.* (2001) Biodegradation of oil in uplifted basins prevented by deep-burial sterilisation. *Nature* **411**, 1034–1037.
- Wörmer L, *et al.* (2019) Microbial dormancy in the marine subsurface: global endospore abundance and response to burial. *Sci. Adv.* **5**, 1024.
- Wunderlin T, Junier T, Roussel-Delif L, Jeanneret N, Junier P. (2014) Endospore-enriched sequencing approach reveals unprecedented diversity of Firmicutes in sediments. *Environ. Microbiol. Rep.* **6**, 631–639.
- Xie S, Lipp JS, Wegener G, Ferdelman TG, Hinrichs K-U. (2013) Turnover of microbial lipids in the deep biosphere and growth of benthic archaeal populations. *Proc. Natl. Acad. Sci.* **10**, 6010–6014.

Xu X, *et al.* (2018) Petroleum hydrocarbon-degrading bacteria for the remediation of oil pollution under aerobic conditions: a perspective analysis. *Front. Microbiol.* **9**.

Yuan C, Lei J, Cole J, Sun Y. (2015) Reconstructing 16S rRNA genes in metagenomic data. *Bioinformatics* **31**, 35–43.

Zengler K, Richnow HH, Rossello-Mora R, Michaelis W, Widdel F. (1999) Methane formation from long-chain alkanes by anaerobic microorganisms. *Nature* **401**, 266–269.

Zhou Z, *et al.* (2021) Non-syntrophic methanogenic hydrocarbon degradation by an archaeal species. *Nature* **601**, 257–262.

ZoBell CE, Anderson DQ. (1936) Vertical distribution of bacteria in marine sediments. *Bull. Am. Assoc. Petroleum Geol.* **20**, 258–269.

Appendix

Co-author permissions

Chapter 3 – Geological processes mediate a microbial dispersal loop in the deep biosphere

Daniel Gittins

From: Desiage, Pierre-Arnaud
Sent: May 11, 2022 9:57 AM
To: Daniel Gittins
Subject: RE: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

[△EXTERNAL]

Dear Daniel,

You have permission to include the research manuscript titled “Geological processes mediate a microbial dispersal loop in the deep biosphere” by Daniel A. Gittins, Pierre-Arnaud Desiage, Natasha Morrison, Jayne E. Rattray, Srijak Bhatnagar, Anirban Chakraborty, Jackie Zorz, Carmen Li, Oliver Horanszky, Margaret A. Cramm, Francesco Bisiach, Robbie Bennett, Jamie Webb, Adam MacDonald, Martin Fowler, D. Calvin Campbell, and Casey R. J. Hubert in your thesis.

Pierre-Arnaud Desiage, Ph. D

Daniel Gittins

From: Morrison, Natasha M
Sent: May 11, 2022 11:00 AM
To: Daniel Gittins
Subject: RE: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

[△EXTERNAL]

Dear Daniel,

You have permission to include the research manuscript titled “Geological processes mediate a microbial dispersal loop in the deep biosphere” by Daniel A. Gittins, Pierre-Arnaud Desiage, Natasha Morrison, Jayne E. Rattray, Srijak Bhatnagar, Anirban Chakraborty, Jackie Zorz, Carmen Li, Oliver Horanszky, Margaret A. Cramm, Francesco Bisiach, Robbie Bennett, Jamie Webb, Adam MacDonald, Martin Fowler, D. Calvin Campbell, and Casey R. J. Hubert in your thesis.

Natasha Morrison

Daniel Gittins

From: Jayne Rattray
Sent: May 11, 2022 9:56 AM
To: Daniel Gittins
Subject: Re: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

Dear Daniel,

You have permission to include the research manuscript titled "Geological processes mediate a microbial dispersal loop in the deep biosphere" by Daniel A. Gittins, Pierre-Arnaud Desiage, Natasha Morrison, Jayne E. Rattray, Srijak Bhatnagar, Anirban Chakraborty, Jackie Zorz, Carmen Li, Oliver Horanszky, Margaret A. Cramm, Francesco Bisiach, Robbie Bennett, Jamie Webb, Adam MacDonald, Martin Fowler, D. Calvin Campbell, and Casey R. J. Hubert in your thesis.

Jayne E Rattray

Daniel Gittins

From: Srijak Bhatnagar
Sent: May 11, 2022 9:57 AM
To: Daniel Gittins
Subject: Re: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

[△EXTERNAL]

Dear Daniel,

You have permission to include the research manuscript titled "Geological processes mediate a microbial dispersal loop in the deep biosphere" by Daniel A. Gittins, Pierre-Arnaud Desiage, Natasha Morrison, Jayne E. Rattray, Srijak Bhatnagar, Anirban Chakraborty, Jackie Zorz, Carmen Li, Oliver Horanszky, Margaret A. Cramm, Francesco Bisiach, Robbie Bennett, Jamie Webb, Adam MacDonald, Martin Fowler, D. Calvin Campbell, and Casey R. J. Hubert in your thesis.

Srijak Bhatnagar

Daniel Gittins

From: Anirban Chakraborty
Sent: May 12, 2022 10:09 AM
To: Daniel Gittins
Subject: Re: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

[△EXTERNAL]

Dear Daniel,

You have permission to include the research manuscript titled "Geological processes mediate a microbial dispersal loop in the deep biosphere" by Daniel A. Gittins, Pierre-Arnaud Desiage, Natasha Morrison, Jayne E. Rattray, Srijak Bhatnagar, Anirban Chakraborty, Jackie Zorz, Carmen Li, Oliver Horanszky, Margaret A. Cramm, Francesco Bisiach, Robbie Bennett, Jamie Webb, Adam MacDonald, Martin Fowler, D. Calvin Campbell, and Casey R. J. Hubert in your thesis.

Anirban Chakraborty

Daniel Gittins

From: Jackie Zorz
Sent: May 11, 2022 9:56 AM
To: Daniel Gittins
Subject: Re: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

Dear Daniel,

You have permission to include the research manuscript titled "Geological processes mediate a microbial dispersal loop in the deep biosphere" by Daniel A. Gittins, Pierre-Arnaud Desiage, Natasha Morrison, Jayne E. Rattray, Srijak Bhatnagar, Anirban Chakraborty, Jackie Zorz, Carmen Li, Oliver Horanszky, Margaret A. Cramm, Francesco Bisiach, Robbie Bennett, Jamie Webb, Adam MacDonald, Martin Fowler, D. Calvin Campbell, and Casey R. J. Hubert in your thesis.

Best,
Jackie Zorz

Daniel Gittins

From: Carmen Li
Sent: May 11, 2022 10:12 AM
To: Daniel Gittins
Subject: RE: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

Dear Daniel,

You have permission to include the research manuscript titled "Geological processes mediate a microbial dispersal loop in the deep biosphere" by Daniel A. Gittins, Pierre-Arnaud Desiage, Natasha Morrison, Jayne E. Rattray, Srijak Bhatnagar, Anirban Chakraborty, Jackie Zorz, Carmen Li, Oliver Horanszky, Margaret A. Cramm, Francesco Bisiach, Robbie Bennett, Jamie Webb, Adam MacDonald, Martin Fowler, D. Calvin Campbell, and Casey R. J. Hubert in your thesis.

Carmen Li

Daniel Gittins

From: Oliver Horanszky
Sent: May 12, 2022 9:33 AM
To: Daniel Gittins
Subject: Re: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

[△EXTERNAL]

Dear Daniel,

You have permission to include the research manuscript titled "Geological processes mediate a microbial dispersal loop in the deep biosphere" by Daniel A. Gittins, Pierre-Arnaud Desiage, Natasha Morrison, Jayne E. Rattray, Srijak Bhatnagar, Anirban Chakraborty, Jackie Zorz, Carmen Li, Oliver Horanszky, Margaret A. Cramm, Francesco Bisiach, Robbie Bennett, Jamie Webb, Adam MacDonald, Martin Fowler, D. Calvin Campbell, and Casey R. J. Hubert in your thesis.

Cheers,
Oliver Horanszky

Daniel Gittins

From: Margaret Anne Cramm
Sent: May 11, 2022 10:00 AM
To: Daniel Gittins
Subject: RE: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

[△EXTERNAL]

Dear Daniel,

You have permission to include the research manuscript titled “Geological processes mediate a microbial dispersal loop in the deep biosphere” by Daniel A. Gittins, Pierre-Arnaud Desiage, Natasha Morrison, Jayne E. Rattray, Sriyak Bhatnagar, Anirban Chakraborty, Jackie Zorz, Carmen Li, Oliver Horanszky, Margaret A. Cramm, Francesco Bisiach, Robbie Bennett, Jamie Webb, Adam MacDonald, Martin Fowler, D. Calvin Campbell, and Casey R. J. Hubert in your thesis.

I can't wait to celebrate your PhD finish!! Good luck!!

Sincerely,

Margaret Cramm

Daniel Gittins

From: Francesco Bisiach
Sent: May 11, 2022 3:55 PM
To: Daniel Gittins
Subject: RE: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

Dear Daniel,

You have permission to include the research manuscript titled “Geological processes mediate a microbial dispersal loop in the deep biosphere” by Daniel A. Gittins, Pierre-Arnaud Desiage, Natasha Morrison, Jayne E. Rattray, Sriyak Bhatnagar, Anirban Chakraborty, Jackie Zorz, Carmen Li, Oliver Horanszky, Margaret A. Cramm, Francesco Bisiach, Robbie Bennett, Jamie Webb, Adam MacDonald, Martin Fowler, D. Calvin Campbell, and Casey R. J. Hubert in your thesis.

Francesco Bisiach

Daniel Gittins

From: Bennett, Robbie
Sent: May 11, 2022 11:23 AM
To: Daniel Gittins
Subject: RE: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

[△EXTERNAL]

Dear Daniel,

You have permission to include the research manuscript titled “Geological processes mediate a microbial dispersal loop in the deep biosphere” by Daniel A. Gittins, Pierre-Arnaud Desiagne, Natasha Morrison, Jayne E. Rattray, Srijak Bhatnagar, Anirban Chakraborty, Jackie Zorz, Carmen Li, Oliver Horanszky, Margaret A. Cramm, Francesco Bisiach, Robbie Bennett, Jamie Webb, Adam MacDonald, Martin Fowler, D. Calvin Campbell, and Casey R. J. Hubert in your thesis.

All the best,

Robbie Bennett

Daniel Gittins

From: Jamie Webb
Sent: May 11, 2022 10:16 AM
To: Daniel Gittins
Subject: Re: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

[△EXTERNAL]

Dear Daniel,

You have permission to include the research manuscript titled “Geological processes mediate a microbial dispersal loop in the deep biosphere” by Daniel A. Gittins, Pierre-Arnaud Desiagne, Natasha Morrison, Jayne E. Rattray, Srijak Bhatnagar, Anirban Chakraborty, Jackie Zorz, Carmen Li, Oliver Horanszky, Margaret A. Cramm, Francesco Bisiach, Robbie Bennett, Jamie Webb, Adam MacDonald, Martin Fowler, D. Calvin Campbell, and Casey R. J. Hubert in your thesis.

Jamie Webb

Daniel Gittins

From: MacDonald, Adam W
Sent: May 11, 2022 11:47 AM
To: Daniel Gittins
Subject: RE: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

[△EXTERNAL]

Hi Daniel,

I approve. One comment is that we are now the Department of Natural Resources and Renewables. Dept of Energy and Mines is no more.

Adam

Daniel Gittins

From: Martin Fowler
Sent: May 11, 2022 8:00 PM
To: Daniel Gittins
Subject: Re: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

[△EXTERNAL]

Dear Daniel,

You have permission to include the research manuscript titled “Geological processes mediate a microbial dispersal loop in the deep biosphere” by Daniel A. Gittins, Pierre-Arnaud Desiage, Natasha Morrison, Jayne E. Rattray, Srijak Bhatnagar, Anirban Chakraborty, Jackie Zorz, Carmen Li, Oliver Horanszky, Margaret A. Cramm, Francesco Bisiach, Robbie Bennett, Jamie Webb, Adam MacDonald, Martin Fowler, D. Calvin Campbell, and Casey R. J. Hubert in your thesis.

Martin Fowler

Daniel Gittins

From: Campbell, Calvin
Sent: May 12, 2022 10:35 AM
To: Daniel Gittins
Subject: RE: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

[△EXTERNAL]

Dear Daniel,

You have permission to include the research manuscript titled “Geological processes mediate a microbial dispersal loop in the deep biosphere” by Daniel A. Gittins, Pierre-Arnaud Desiagne, Natasha Morrison, Jayne E. Rattray, Srijak Bhatnagar, Anirban Chakraborty, Jackie Zorz, Carmen Li, Oliver Horanszky, Margaret A. Cramm, Francesco Bisiach, Robbie Bennett, Jamie Webb, Adam MacDonald, Martin Fowler, D. Calvin Campbell, and Casey R. J. Hubert in your thesis.

- D. Calvin Campbell, PhD

Daniel Gittins

From: Casey Hubert
Sent: May 11, 2022 11:14 AM
To: Daniel Gittins
Subject: Re: Co-author copyright approval - Chapter 3 - Microbial dispersal loop

Dear Daniel,

You have permission to include the research manuscript titled “Geological processes mediate a microbial dispersal loop in the deep biosphere” by Daniel A. Gittins, Pierre-Arnaud Desiagne, Natasha Morrison, Jayne E. Rattray, Srijak Bhatnagar, Anirban Chakraborty, Jackie Zorz, Carmen Li, Oliver Horanszky, Margaret A. Cramm, Francesco Bisiach, Robbie Bennett, Jamie Webb, Adam MacDonald, Martin Fowler, D. Calvin Campbell, and Casey R. J. Hubert in your thesis.

Casey