# Turning Down the Gas: What is the potential for microbial detection of methane leakage from soils?

**Thomas Bott**

**Supervisors**

Dr Simon Gregory, Prof. George Shaw,

Dr Barbara Palumbo-Roe, Prof. Neil Crout

Thesis Submitted to the University of Nottingham for Completion of a PhD in Soil Science

August 2022

## 0.1 Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. It has not been submitted, in part or whole, to any university or institution for any degree, diploma, or other qualification.

In accordance with university guidelines, this thesis does not exceed 100,000 words.

*Signed:*

12/01/2022

## 0.2 Covid-19 Impact Statement

The pandemic did unfortunately impact the work in this thesis. This was mostly due to access problems and the subsequent loss of time in the laboratory. Whilst all work needed to complete the thesis and test the hypothesis was undertaken there are still pieces of laboratory work I would have liked to complete. These would have offered a little more depth or insight into communities being studied, for example RNA extraction for use in PCR would have shed light some on community activity. Another field site or more field work at the sites presented below would also have been ideal. Overall, despite the difficulties posed by Covid-19 the work presented here still tackles the questions asked, offers novel insights into methanotrophic communities and identifies areas for future work.

## 0.3 Abstract

Methane is a potent greenhouse gas. Emissions from soils, wetlands and subsurface sources are therefore a concern when trying to understand greenhouse gas emission and fluxes in the environment. Soil microbial communities can act as a sink for methane, mitigating its release. The purpose of the thesis was to test if terrestrial natural gas flux could be detected using changes in the soil microbiology, focussing upon changes in the relative abundance of methanotrophs. Two methane mono-oxygenase genes were focussed upon, *pmoA*, *mmoX* and *Methylocella spp. mmoX*. To detect these genes PCR and qPCR assays for use with soil samples were developed using existing primer sets. Next-generation amplicon sequencing, of the 16S rRNA gene, was also used to better explore community changes.

Assays were developed and tested in laboratory incubated soils and field samples. Field work used these molecular tools to develop a method of surveying sites to attempt to identify hotspots of natural gas emission using microbial indicators. Four field sites were surveyed: the two sites with clear methane releases, Hardstoft One and La Fontaine Ardent, appeared to have related increases in the relative abundance of methanotrophs. Across all four field sites spatial separation of *pmoA*, *mmoX* and *Methylocella spp. mmoX* genes appeared to be occurring suggesting either habitat preferences, differing growth strategies, differing substrate preferences or a combination of these.

Within the laboratory, soils were incubated with a methane headspace to develop methanotrophic communities. A methane pulse was simulated in a long-term experiment to develop an understanding of how soil microbial communities are impacted by methane fluxes. Overall, increases in the relative abundance of pmoA were observed alongside an increase in methane oxidation rates; the increased relative abundance of *pmoA* remained the duration of the experiment but methane oxidation rates declined through time.

The role of methanotrophs in natural attenuation was also explored using contaminated land samples. This work aimed to explore the use of molecular techniques to investigate the relative abundance of methane mono-oxygenase genes in contaminated land sites. This hoped to assess the use of these in future studies exploring the natural attenuation of organic contaminants, with the aim of identifying future research directions. Problems with DNA recovery and amplification limited the scope of this work but highlighted potential avenues for future work.

Cumulatively the thesis demonstrates the ability of qPCR to survey methane oxidising microbial communities over scales of tens of metres. These methods have the potential to help detect and monitor methane emissions from former oil and gas fields or sites where methane emission is suspected, for example landfill sites. The thesis also develops the understanding of methanotrophic community dynamics in soils.

# Contents

# List of Figures

# List of Tables

## 0.4 Acknowledgments

Firstly, I must thank my supervisors, Simon Gregory, George Shaw, Barbara Palumbo-Roe and Neil Crout, for their help and guidance. Help from Megan Barnett, Kay Green and Jess Mackie, also at the British Geological Survey, has also been invaluable. The microbiology group at the BGS has shared ideas, space, methods, equipment, and my general clutter: without their patience this thesis would not have been possible. I would like to thank Simon for finding time for my continuous questioning whilst also giving me the opportunity and trusting me to try some unusual approaches in the laboratory. I must also thank John Corrie, and the rest of the technical staff at Gateway, for dealing with my often-odd orders and helping me navigate the world of university finances. I also need to thank my family for tolerating my choice to do a PhD. Kirsty has been incredibly supportive and understanding. As for the dog, I do not think she cared what I was doing so long as she got fed and walked, but I apologise as she probably missed out on a few walks. I should thank her for always being there (did she have anywhere else to go?) and providing an uncomplaining sounding board and sometimes sharing her toys.

### 0.4.1 Chapter Acknowledgments

**Section 2 - Methods**

Megan Barnett wrote and tested the SOP for the wet chemistry in the Nanopore Sequencing section. Megan's assistance was also invaluable in design and implementation of the data pipe also outlined in this section.

**Section 4 – Field Surveys at Sub-Alpine Macro-Seeps**

Wolfram Kloppmann, Eric Proust and Fredrick Gal, of Bureau de Recherches Géologiques et Minières (BGRM), for all their assistance setting up the fieldwork and hosting me. BGRM also completed and provided all the gas data for use in this chapter. I am also indebted to Megan Barnett and Simon Gregory for their assistance in field (and helping me prepare for fieldwork abroad).

**Section 5 – Field Survey at Dukes Wood Eakring**

I am hugely indebted to Simon Gregory for accompanying me to the site, in rain, and thick clay mud, to assist me in collecting the soil samples. Also for his help with the risk assessments needed for the work to be undertaken. I am also grateful for the assistance of Helen Taylor-Curran and Andi Smith (BGS) for pore gas sampling

equipment and sampling containers. Finally also to the Nottingham Wildlife Trust for granting me access.

**Section 6 – Long Term Impacts of a Methane Pulse Event**

I am grateful for the patience of other laboratory staff in K106. Large pipes and trays of soils have cluttered the floor for several years. I am also indebted to Andrew Smith (BGS) for access, and assistance with his laboratory's gas chromatography equipment. Finally, thanks to the staff in BGS Sample Preparation for allowing me to use their workspace for soil sieving.

**Section 7 – Relative Abundance of Mono-Oxygenases at Former Gasworks**

This work would not have been possible without the guidance and assistance of Darren Beriro. Furthermore, my thanks are extended to the BGS sample preparation team, Mark Kalra and Charles Gowing, for allowing me access to the facilities to handle samples with a potentially high toxicity. Access to sites and consultants was made possible thanks to National Grid Property Holdings (NGPH). The proposal development and data analysis was assisted by Russell Thomas (WSP). Further funding to help cover the cost of chemical analysis was provided by NGPH and STARs.

## 0.5 Abbreviations and Conventions

Typographical Conventions

*Italic*:

All genes (*pmoA*) and binomials (*Methylocella stellata*) are in italics.

`Monospaced` Font:

Computer packages (`phyloseq`) and code are presented monospaced.

### 0.5.1 List of Abbreviations:

| Abbreviation | Definition |
|---|---|
| bp | Base Pair - DNA |
| CuMMO | Copper Methane Mono-Oxygenase |
| DI - Water | Deionised Water |
| dNTPs | Deoxynucleotide triphosphates |
| E-Value | Efficiency Value for qPCR |
| GC-Content | Guanine Cytosine Content (in DNA) |
| IDW | Inverse Distance Weighted |
| kb | Kilo-base (1000 base pairs of DNA) |
| LB | Lysogeny Broth |
| LoD | Limit of Detection |
| MIQE | Minimum Information for publication of qPCR Experiments |
| MMO | Methane Mono-Oxygenase |
| PCR | Polymerase Chain Reaction |
| qPCR | Quantitative Polymerase Chain Reaction |
| RMSE | Root Mean Square Error |
| SCA | Short Chain Alkanes |
| SD | Standard Deviation |
| SDIMO | Soluble Di-Iron Mono-Oxygenase |
| SDS | Sodium dodecyl sulphate (surfactant for DNA extraction) |
| SOP | Standard Operating Procedure |
| TAE | Tris-Acetate-EDTA (buffer) |

# Chapter 1

# Project Introduction

Methane ($CH_4$) is the simplest alkane; composed of a single carbon atom and four hydrogen atoms. At high concentrations, methane is considered to be a pollutant, but it is ubiquitous in the environment at low concentrations. Soils, sediments and the free atmosphere can all contain detectable concentrations of methane. Within the free atmosphere methane is a potent trace greenhouse gas, with a radiative effect 25 times that of carbon dioxide (mole for mole) (Rodhe, 1990). Atmospheric concentrations of methane have risen since 1750 and despite a decade of stability between 1996 and 2006 have been rising since 2007 with atmospheric methane at 1803ppb in 2011 (Ciais *et al.*, 2013) and reaching 1909 ppb in March 2022 (Dlugokencky and NOAA/GML, 2022). An understanding of methane's biogeochemical cycle is, therefore, important in monitoring and mitigating the potential risks of rising atmospheric methane concentrations. Furthermore, an understanding of how society's activities and industry can perturb that cycle, to either increase or decrease emissions, is needed.

One perturbation is the use of methane as a common fuel source in natural gas, usually in the production of electricity and in domestic heating systems. Compared to other hydrocarbon fuels, methane is considered relatively clean; its combustion leads to the production of only water and carbon dioxide instead of the complex mix of compounds produced during the combustion of oil and coal. However, the processes of extracting, transporting and storing methane can result in fugitive emissions, offsetting some of the climate benefits of natural gas compared to coal (Alvarez *et al.*, 2012). After extraction has ceased there is also potential emission from abandoned extraction wells.

## 1.1 A Legacy of Oil and Gas Extraction

The United Kingdom (UK) has a legacy of onshore petroleum exploration and extraction. The first recorded onshore oil well was sunk in Hardstoft, Derbyshire, in 1915. There are now 2249 onshore wells in the UK (UK Oil and Gas Authority, 2019), the majority of which are decommissioned and abandoned. The potential for the UK to produce natural gas from shales through onshore, unconventional extraction (fracking) has been the source of controversy. While currently not allowed by the planning regime, fracking highlights that the UK's onshore legacy may increase in coming years.

### 1.1.1 Fracking

Unconventional extraction releases natural gas from the reservoir/source rock by fracturing it. Simply, a well is sunk into the target strata and high-pressure fluids pumped from the surface into the well and rock. The pressure causes the rock to fracture and release gases from its matrix. The UK government (as of November 2019), no longer supports unconventional gas extraction and there is currently a moratorium on the activity due to the unpredictable seismic activity generated by the practice. Up to that point, fracking within the UK had been the subject of intense public, political and scientific scrutiny (for example Bomberg, 2017; Jaspal and Nerlich, 2014; Howarth *et al.*, 2011), with fears of water and ground pollution caused during gas extraction alongside potential impacts upon human health. Currently, unconventional wells seem to have a greater risk of experiencing well integrity issues (for example Ingraffea *et al.*, 2014)). However, no direct comparison has been made and the seemingly increased prevalence of integrity failure may be due to tougher regulations leading to increased violations being reported (Jackson, 2014). That said, unconventional wells are more technically complex (they deviate from vertical to horizontal) and operate under higher pressures than conventional gas wells, both factors that could increase the risks to well integrity (Jackson, 2014).

### 1.1.2 Abandonment, Decommissioning and Seeps

Once a well is no longer economically viable it is decommissioned which involves restoration to a condition of equal, or better quality, than it was before drilling began (UK Government, 2015). The well, in accordance with health and safety guidance, should have been constructed and decommissioned (plugged, cut-off and buried) to ensure that no escape of contaminants can occur (UK Government, 2015). Well construction varies but wells are generally constructed with nested casings (normally steel piping) with the annular spacing between the casings sealed with cement to form a barrier between the wellbore and the environment. Sometimes the outer casing, or surface casing, is isolated from the rock it passes through by a cement barrier. In short, wells are constructed in a way to ensure that material cannot leak into them and then be transported either to the surface or to another formation through which the well passes (for example an aquifer) (Jackson, 2014).

However, a well's integrity may degrade through time, or factors during its construction may increase the risk of integrity failure (Watson and Bachu, 2009). Damage to either the cement or casing can produce pathways along or within the well (mechanisms summarised by: Lackey and Rajaram, 2018; Watson and Bachu, 2009)). Gases can flow through the wellbore, via the cement annuli matrix or interfaces between cement and casing, so called surface-casing vent flow (SCVF). In some instances, sustained casing pressure (SCP) within the well can force material out of the wellbore and allow gas migration (GM) through the material outside the well casing. There are several potential mechanisms of well integrity failure. Dusseault *et al.* (2000) argued that the cement used could suffer from shrinkage and cracking, these fractures would then continue to extend due to pressure from rising subsurface gases.

### 1.1.3 Fugitive Emissions from Wells

There is ample evidence that wells can lose integrity and that this can lead to fugitive emission of natural gas. Ingraffea *et al.* (2014) explored regulation violations in Pennsylvania where well integrity was reduced through cement and casing problems. The dataset used contained records from 41,381 wells sunk between 2000 and 2012 and looked at 75,505 compliance reports. Overall, the authors observed that 1.9% of all wells suffer a loss of integrity (weighted average of all well types across the state). However, geographical, temporal and extraction methods complicated the analysis with some groups, for example of the 2,714 unconventional wells in the North-East of the state, sunk after 2009, 9.18% had recorded incidences of integrity problems. Watson and Bachu (2009) analysed a dataset form Alberta Energy Resources Conser-

vation Board (A-ERCB), a group that regulates the 315,000 oil, gas and injection wells in Alberta, Canada. The A-ERCB requires testing of surface leakage from wells both during and after decommissioning. The dataset showed that 4.6% (3.6% SCVF; 0.6% GM) of the wells had recorded gas emissions.

The resulting emissions from these wells have mixed compositions. Gases from subsurface sources are dominated by methane, with smaller components of carbon dioxide and other short-alkanes (Etiope, 2015). Atmospheric compositions above active gas fields have been observed to have an increased methane component which was accompanied by an increase in propane and butane, strongly suggesting that all three gases originate from the gas extraction process (Caulton *et al.*, 2014; Karion *et al.*, 2013). Alongside release into the atmosphere, methane can enter groundwater (Cahill, Steelman, Forde, Kuloyo, Ruff, Mayer, Mayer, Strous, Ryan, A. Cherry, and Parker, 2017) and the bio-attenuation of the methane alongside syntrophic bacterial sulphate reduction may impact water quality (Van Stempvoort *et al.*, 2005). There has also been anecdotal evidence of flammable gas entering household water supplies at concentrations capable of ignition (Dusseault *et al.*, 2000).

### 1.1.4 Monitoring

Within the UK there has been little monitoring of legacy wells. Boothroyd *et al.* (2015) a survey of 102 abandoned wells, across four basins in the UK. The authors measured soil gas concentrations in the field, with three measurements around each well head and seven taken in neighbouring fields as controls. Of the 102 wells sampled, they observed that 30% had increased concentrations of methane compared with controls. It should also be noted that 39% had lower concentrations of methane compared with controls. Interestingly, the dataset suggested that the ages of the emitting wells had no impact on the recorded methane concentrations. Outside the UK there is evidence that abandoned and plugged wells can emit natural gases. A study of abandoned wells in Pennsylvania showed that plugged wells could still contribute to fugitive emissions, suggesting that leakage occurs from a relatively small number of 'high-emitters' which have a sustained release of gases (Kang *et al.*, 2016).

While monitoring of abandoned wells in the UK (or anywhere) is not currently completed, there is evidence to suggest some legacy wells can emit methane even when plugged. Watson and Bachu (2009), as discussed above, illustrated how a regulatory regime can monitor the decommissioning process but within their study measurements were taken within 60 days of drilling ceasing and before complete abandonment by

the operator, not as part of a long-term monitoring program. There are calls, such as those by Jackson (2014) and Davies *et al.* (2014), that a greater understanding of well integrity, especially in older wells, is needed and that a system of monitoring these wells should be in place.

## 1.2 Generation of Short-Chain Alkanes

Natural gas is a mixture of gases, predominantly composed of methane with smaller proportions of carbon dioxide and the short chain alkanes (SCAs) (Etiope, 2015). SCAs include methane, ethane, propane and butane. There are three groups of environmental processes that produce natural gas. These are driven by either biological activity or through geological processes.

### 1.2.1 Methanogensis

In anaerobic systems the final processes of microbial degradation of organic matter can generate methane. Under such conditions a microbial consortium degrades organic matter, firstly through fermentation generating single carbon compounds such as acetate ($CH_3OO$-) which can then be oxidised through anaerobic respiration. Methanogens can further reduce the carbon dioxide to methane when other oxidising agents (such as, iron (III), sulphate and nitrate) are scarce.

Methanogens are a diverse group of anaerobic archaea distributed through seven orders within the phylum Euryarchaeota (Barber, 2016). Biological methane production from wetlands is considered one of the largest sources of methane; the IPCC estimates a flux of 177-284 Tg/yr into the atmosphere (Ciais *et al.*, 2013). Methanogenesis also occurs in shallow geological strata, for example in shallow thermally immature shales (McIntosh *et al.*, 2008; Martini *et al.*, 1998). This can then degas when the rock is further uplifted, jointed or mined. Strata potentially exposed to sterilising temperatures and pressures during diagenesis can be reseeded with methanogenic archaea after uplift, through the percolation of meteoric water, leading to a secondary production of methane (for example: Schlegel *et al.*, 2011; Martini *et al.*, 1998). It should be noted that Strąpoć *et al.* (2011) observed that increasing thermal maturity appears to reduce the potential for methanogenesis with increasing maturity imposing a lower production ceiling.

There are three distinct routes of methanogenesis. However, each shares a final identical biochemical pathway where a methyl group ($CH_3$) is reduced to methane Figure

1.1 (Lessner, 2001). The difference between the routes is defined by how the organism provides a methyl group for the final reduction steps. One route is via the reduction of $CO_2$ with $H_2$ to a methyl group and $H_2O$. The other two routes involve the archaea consuming a 'pre-prepared' methyl group: either, methanol and methylamines are consumed and undergo dismutation (one substrate is oxidised while three reduced) or acetate is fermented. The majority of biogenic methane is produced through the latter pathway (Lessner, 2001). Here, acetate's C-C bond is cleaved, leaving a methyl group and $CO_2$. The methyl group is then reduced using the electrons generated from the oxidation of the $CO_2$ group.

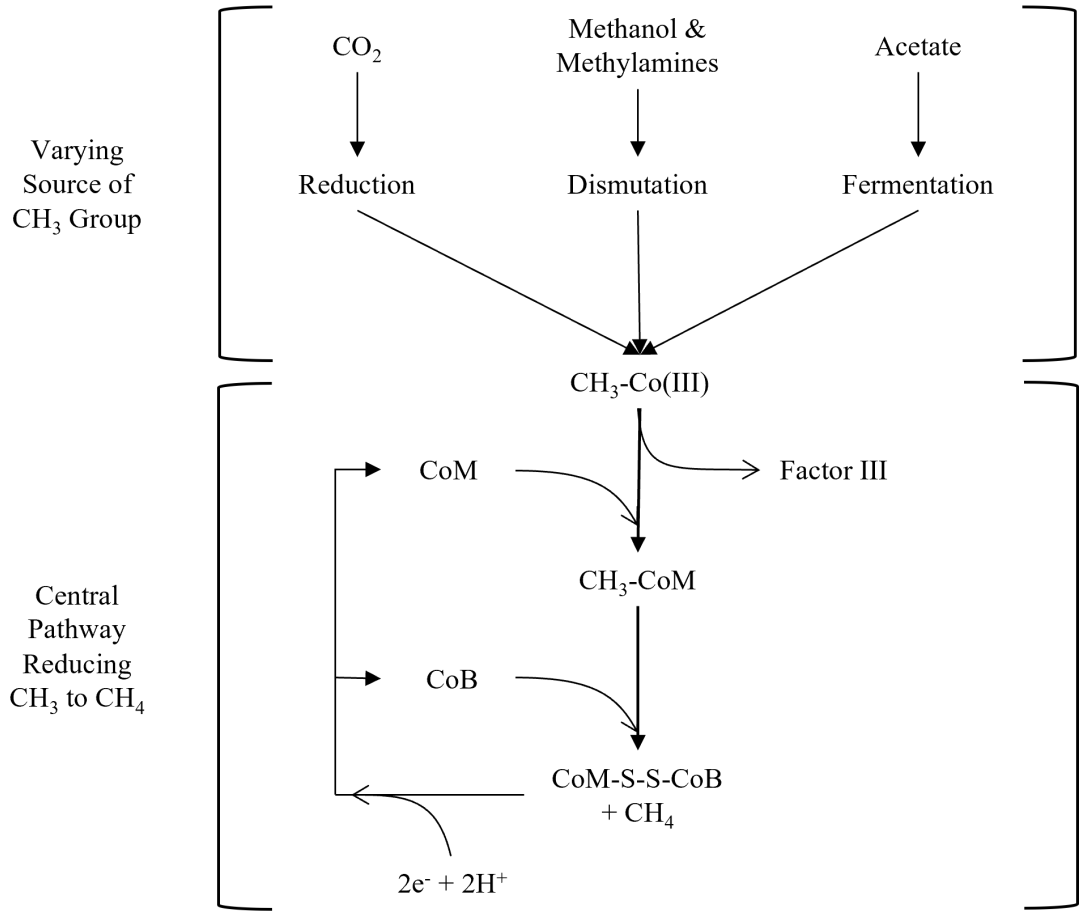



**Figure 1.1: Overview of the pathways in Methanogenesis.** The source of the methyl group reduced to $CH_4$ is variable depending upon the species. However, once the methyl group has been produced all methanogenic species have an identical pathway which completes the reduction. A methyltransferase moves the $CH_3$ group to co enzyme M (CoM). This is reduced by methyl co enzyme M reductase (MCR), producing $CH_4$ and binding CoM to co enzyme B (CoB). The two co factors are recovered by a heterodisulphide reductase. Figure adapted from Lessner (2001)

It is generally accepted that methanogenesis only occurs in anoxic environments, for example wetlands, paddy-fields or marine sediments. However, there is a small body of

research that suggests that methane, at low concentrations, can be produced in aerobic systems Gauthier *et al.* (2015); Angel *et al.* (2011); von Fischer and Hedin (2007). This is probably due to anaerobic microsites within the soil system facilitating anaerobic methanogenesis in an overall oxygenated environment.

### 1.2.2 Microbial Production of $C_2$+ Alkanes

Methanogenesis is exclusively the production of methane by archaea. While there is evidence that microbial processes can produce hydrocarbons, there is little in the literature to suggest that this occurs in the environment. Ethane, apparently of a microbial origin, has only been reported twice, potentially by methanogens in anoxic sediments (Oremland *et al.*, 1988) and in aquifers in a hydrocarbon producing area (Taylor *et al.*, 2000). Research into novel fuel sources has led to interest in the abilities of microbial processes to produce SCAs. Seemingly one of the best studied systems relies on photoautotrophic cyanobacteria, these systems produce hydrogen, ethanol and isoprene, not the core components of natural gas (for a review: Machado and Atsumi, 2012). Other approaches have used genetically modified *Escherichia coli* strains with novel metabolic/catabolic pathways capable of producing petrol-like mixtures with alkanes though, again, this did not produce natural gas Choi and Lee (2013). Overall, there does not currently appear to be strong body of evidence that shows naturally occurring $C_2$-$C_4$ synthesis by microbial processes in the subsurface or within soil and sediments.

### 1.2.3 Thermogenic Production

Some authors group thermogenic production with methanogenesis as a biotic source of SCAs (for example: Etiope, 2015). This is due to the starting carbon source being biological material undergoing sedimentation and lithification. However, the formation of gases is by a group of chemical reactions dominated by the thermal cracking of complex organic compounds, for example cellulose and lignin rather than biological activity. The formation of SCAs is part of wider petroleum (bitumen, natural gas and oil) generation.

At the point of deposition organic rich material contains complex biopolymers. Through the initial stages of lithification these decompose into kerogens. Kerogens are complex chemical mixtures, insoluble in organic solvents, which cannot be easily extracted from the sedimentary rock formed. With coal, the final mature rock is almost entirely kerogen formed from humic material. Kerogens, when exposed to increasing temperatures, in later stages of diagenesis, crack into oil and gas. Once produced, the lighter

hydrocarbon molecules migrate from the source rock into reservoirs due to pressure gradients and the increased volume of the products. This probably leads to fractionation of the petroleum products, with heavier, larger components left in the source rock (Seewald, 2003). Overall, the process of burial, and the associated increase in temperature, leads to the conversion of complex organic compounds to low mass hydrocarbons.

There are various factors that influence the extent and volume of the conversion to low mass hydrocarbons. The starting biological material has an impact upon the kerogen and subsequently the petroleum produced, due to variation in the H/C and O/C atomic ratios. Oils arise from algal and planktonic debris, whereas humic material (plant material) produces natural gas during diagenesis. Time and the temperature reached also impact catagenesis; oil deposits form sooner and at a lower temperature (approximately 50°C) than required for gas formation. However, as the thermal maturity of the strata increases so does the production of natural gas, with increasing maturity related to increased proportions of methane within the gases produced. Besides increased temperature and time the presence of other reactants appears to influence the production of natural gas (for a review: Seewald, 2003). The presence of water, for example, inhibits the production of bitumen and instead leads to increased production of oil Lewan (1997).

### 1.2.4 Abiotic Production

Abiotic reactions can produce methane and potentially other alkanes as part of petroleum production in the upper mantle. Critically, abiotic production pathways do not include either organic matter or biological processes. There are two predominant systems in which methane is produced by abiotic processes, either within high temperature magmatic systems or lower temperature gas-water-rock systems Etiope (2015). (As reviewed by Etiope and Lollar, 2013), magmatic systems involving high temperature reactions (500-1500°C) within the upper mantle can generate methane, as can late stage reactions within deep cooling magmas (<500-600°C). Reactions, involving gas-water-rock interactions, occur at lower temperatures (mostly under 100°C) and are frequently associated with igneous intrusions, for example, in volcanoes. These lower temperature reactions are thought to contribute more to the global methane budget than the magmatic systems. Cumulatively, however, abiotic methane production is a much smaller source than biotic and thermogenic sources (Etiope and Lollar, 2013).

## 1.3 Differentiation of Sources

The varying processes and starting materials lead to gas seeps having characteristics that can indicate their origins. Seeps have unique compositional and isotopic signatures (for a review see: Golding *et al.*, 2013). Furthermore, it is possible that a seep can be composed of gases from more than one process due to mixing and the action of multiple production pathways on the same starting material (Strąpoć *et al.*, 2011; Martini *et al.*, 1998): for example, methanogenesis within a mature, uplifted coalbed producing natural gas from both thermogenic cracking and methanogenesis.

There are several methods of determining the generation pathway(s) and the origin(s) of the natural gas. Gas wetness/dryness ratios can relate to the pathway and starting material; a 'wet' gas has a higher proportion of $C_2$+ alkanes to $CH_4$, this can be found using one of several equations, for example Equation 1.1 or Equation 1.2. Oil reservoirs tend to effuse wetter gases than natural gas sources, which in turn are wetter than microbial sources.

$$\frac{C_1}{C_2 + C_3} \tag{1.1}$$

Equation from Bernard *et al.* (1976). Where $C_n$ is the number of carbon atoms in the alkane.

$$\frac{C_2 + C_3 + C_4 + C_5}{C_1 + C_2 + C_3 + C_4 + C_5} \tag{1.2}$$

Equation from Golding *et al.* (2013). Where $C_n$ is the number of carbon atoms in the alkane.

Alongside proportions of gas species, the isotopic composition of natural gas can allude to its origins (for example Osborn and McIntosh, 2010). Biochemical processes thermodynamically favour lighter isotopes, enabling the use of carbon ($^{13}C/^{12}C$) and hydrogen ($^{2}H/^{1}H$) isotopes to characterise the source, with biological sources producing mixtures with a greater proportion of the lighter isotopes compared to geological processes. Isotopes are measured relative to a standard, with the shift from the standard being used to compared samples. For example, $^{1}H$ (deuterium or, D) is presented as $\delta$D per mille (‰) against Standard Mean Ocean Water (SMOW), calculated using Equation 1.3.

$$\delta_x = [\frac{(R_a)sample}{(R_a)standard} - 1]10^3 \quad (1.3)$$

Equation from Whiticar (1999). Where R is the isotope ratio (e.g. $^{13}C/^{12}C$)

A plot comparing a combination of two these metrics, wetness/dryness, $\delta^{13}C$ or $\delta D$ enables a source to be surmised. (for example Whiticar, 1999), plotted $\delta D$ against $\delta^{13}C$ to ascertain a potential source pathway. Using $\delta D$ and $\delta^{13}C$ also enables some distinction between methanogenic processes; $CO_2$ reduction clusters separately to the pathways that utilise acetate fermentation or dismutation reactions with a donated methyl group Strąpoć *et al.* (2011); Whiticar (1999). While these indicators are useful, the complex history of reserves can often make these basic measurements more complicated to interpret (Golding *et al.*, 2013). For example, secondary methane production, via biogenic pathways occurring after uplift, potentially alters both these indicators, masking the history of the reserve gas.

## 1.4 Detecting Alkane Flux Using Soil Microbiology

Changes in the microbial community can be used indirectly to measure gas flux, i.e. as an indicator of the gas's presence. Changes in the composition of the bacterial community should be related to changes in the environment, including the gaseous and SCA composition of the environment: therefore, communities with an increased exposure to alkanes should have a greater proportion of members that can exploit alkanes for oxidation. Methanotrophs and alkanotrophs can oxidise methane and $C_2$+ for growth and reproduction.

### 1.4.1 Oil and Gas Prospecting

Short chain alkanes can be detected using multiple methods. These include techniques which measure SCAs themselves using tools such as gas chromatography. Or they can be measured indirectly using another measurable quantity as an indicator of the alkanes' presence. The use of alkanotrophs to detect alkane seepage from the subsurface is not new: methane fluxes have detected using methanotrophs for several decades in the oil and gas industry (for example: Brisbane and Ladd, 1965). Above oil and gas reservoirs, gas seeps form due to gas density and sub-surface pressure, coupled with microfractures in overlying strata (Tedesco, 1995). The light hydrocarbons migrating into overlaying strata and soil lead to an observable increase in the abundance of methanotrophs and other alkanotrophs which can be used in prospecting efforts Liu

*et al.* (2016); Zhang *et al.* (2014); Wagner *et al.* (2002).

Compared to measuring soil gas fluxes or concentrations directly, microbial communities are thought to be more stable over time. As a result, measuring the composition of these communities acts as a time-based average, revealing the state of a gas flux over several months rather than the flux at that point in time (Environmental Bio Technologies Inc., 2017) (Environmental Bio Technologies Inc. 2017). Cahill *et al.* (2017) noted that a groundwater microbial community remained in a disturbed state, following a methane injection test, after 253 days. Furthermore, halo or apical anomalies, (Figure 1.2, in surface gas concentrations caused by the activity of microbial communities are also avoided.

Quantification of environmental alkanotroph populations can be completed using a variety of different methods. Until recently, most techniques used in hydrocarbon prospecting used culture-based techniques, for example the Microbial Prospection for Oil and Gas (MPOG) method (see Wagner *et al.*, 2002). Culture-based techniques attempt to separate the microbial communities from soil samples (making samples easier to handle and removing any other carbon substrates), suspending/plating them and exposing them to methane in one test and to alkanes and other hydrocarbons in a separate test (Wagner *et al.*, 2002; Tedesco, 1995). Samples with more colony forming units and faster rates of respiration are considered to be from areas with increased alkane concentrations, as the recovered microbial community is primed to use alkanes as a carbon source. These techniques claim to have success rates of 90% (Wagner *et al.*, 2002).

Newer methods are focussing upon molecular techniques to remove the bias of culture techniques, thus improving accuracy and precision. These use polymerase-chain reaction (PCR) assays to detect conserved genes encoding catalytically active subunits of the methane mono-oxygenase (MMO) and soluble di-iron mono-oxygenase (SDIMO) enzymes to quantify the methanotroph and alkanotroph populations (Zhang *et al.*, 2017; Liu *et al.*, 2016; Zhang *et al.*, 2014; Miqueletto *et al.*, 2011). PCR enables the rapid amplification of specific genes from a sample to the point where they can be detected. This requires the extraction of bacterial DNA from soils.

## 1.5 Potential Problems

There are numerous soil factors that can impact microbiology and community structure thus complicating their use as an indicator. Microbes are heterogeneously distributed

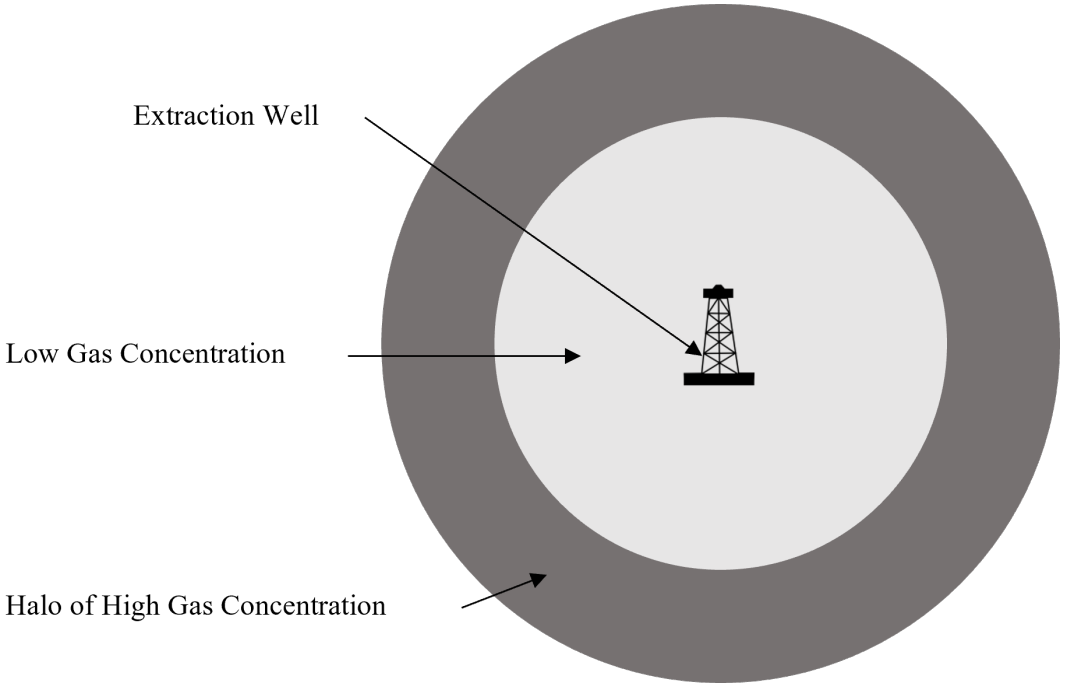



**Figure 1.2: Plan view of surface gas concentrations around an active oil or well head.** The area immediately around the well often shows a decrease in surface concentration or flux of gas due to the drop in subsurface pressure caused by hydrocarbon extraction through the borehole. This generates a halo of relatively high gas concentration. A similar effect may be generated by a body of alkanotrophs thriving on a high concentration of gas effusing from the subsurface and removing the gases before they reach the surface.

through soils. It is estimated that despite the vast number of microbial cells in soil, $10^7$-$10^{12}$ cells per gram, they only inhabit around $10^{-6}$ % of the soil's total surface area (Young *et al.*, 2008; Young and Crawford, 2004). Furthermore, within 1-5 g of soil there is a very high bacterial diversity, probably because of the array of microhabitats and the limited competition between cells due to the slow rates of nutrient diffusion, especially in unsaturated soils (Young *et al.*, 2008; Zhou *et al.*, 2002).

On larger spatial scales, changes in diversity and community composition are also visible. Changes in pH, metals, carbon content etc. can all influence the microbial community (Zhou *et al.*, 2002). For example in British soils, Griffiths *et al.* (2011) observed that $\alpha$-diversity increased but $\beta$-diversity decreased with increasing soil pH (approximate range 3.5 - 8.5), i.e. soil with higher pH had more diversity within one sample but the diversity between samples with similarly high pH fell. Soil pH is potentially such a key variable due to the impact it has on other edaphic factors such as organic carbon solubility (Andersson *et al.*, 2000), metal solubility and plant cover.

For methanotrophs, and (other) alkanotrophs, the rate of soil gas diffusion is important. Limited movement of soil gases will impact both the availability of oxygen and gaseous alkane substrates. Moisture content, soil structure and temperature all impact the rate of gas diffusion. Compacted soils, or those with a fine texture, have less pore space and poor pore connectivity, cumulatively leading to lower rates of diffusion and therefore lower rates of methane oxidation (Gebert *et al.*, 2011). Furthermore, smaller pore size, resulting either from compaction or soil texture, exacerbates the effects of moisture content upon the rate of gas diffusion and methane oxidation: smaller pore size increases the tendency of water to reduce the gas-filled pore space, thus reducing the effective diffusivity of gases.

Cumulatively, high diversity and community interactions with multiple edaphic factors may obscure any community shifts driven by the presence of specific metabolites. Robust sampling strategies coupled with the observation of other soil factors might enable interpretation of the data collected. However, experimental work exploring the interactions between soil habitat, alkanotrophs and gas diffusion may be needed to draw clear mechanistic conclusions.

### 1.5.1 Methanogenesis

Using methanotrophs as the only indicator of alkanes can generate potential problems. One of the largest is the presence of biogenic methanogenesis which could generate noise and hide, or exaggerate, a soil flux originating from the subsurface (Zhang *et al.*, 2014; Miqueletto *et al.*, 2011). This leads to the inability to relate methane to either a biogenic or thermogenic source. Therefore, to distinguish clearly between sources other indicators may be required. It may be possible, given the lack of biological SCA production, to use the presence of butane and propane oxidising genes to determine if the source is thermogenic.

### 1.5.2 DNA Extraction

During DNA extraction the soil is disaggregated and homogenised. A balance between this disaggregation and the DNA quality and yield must be found. Greater disaggregation will enable more thorough exploration of a soil sample, given there is a greater chance of cells being released from the soil structure and greater chance of recalcitrant cell lysis (e.g. spores). However, increasing physical disruption also increases the risk of DNA shearing and a subsequent loss of sample quality. Bürgmann *et al.* (2001) suggested that there was some unextractable DNA within soils, probably from cells

and sites within the soil that cannot be lysed or accessed with the current techniques.

It is important to recognise that no DNA extraction will represent the whole community. Furthermore, different protocols for DNA extraction will not only alter the quality and yield of the DNA, they also have the potential to alter the observable microbial diversity (İnceoğlu *et al.*, 2010). That said, the use of molecular methods to explore community composition is still a powerful tool compared to classical cultivation techniques that may only represent 1% of the total community.

### 1.5.3 Vertical Flux?

Using soil indicators and gas flux as an indicator of a sub-surface source relies on the assumption that the gases rise predominantly vertically. Tedesco (1995) described this as a 'chimney effect' with the vertical buoyancy of the gases dominating the overall movement of the gas on a reservoir scale. Etiope (2015) painted a more complex picture, suggesting that seepage associated with petroleum exploration is typically associated with rock permeability and movement along fractures and faults, therefore lateral migration along a fault is possible. It is plausible that the well casing will act as a vertical conduit along which the gas will migrate. However, some authors have demonstrated that hydrological impacts will produce a plume, with the gases being moved horizontally by the movement of groundwater and interactions with geological structures. Cahill *et al.* (2017), for example, demonstrated that a plume (>17 m) could form during a shallow (9 m) injection test, while their conceptual model suggests that well integrity failure in an aquifer could produce a plume leading to surface hotspots removed from the surface well-head. Plume formation and extent would depend on hydrological and geological conditions. Cumulatively, vertical flux on reservoir scales is considered possible and detectable; however, relating surface anomalies to specific wells may be difficult due to the possible deviation of seeps. As a result, any biomonitoring method used on a small spatial scale (tens of metres) will need to have a high degree of precision.

## 1.6 Methanotrophs

Soils can act as a sink for methane. Diffusion into soils is one of the two major sinks of atmospheric methane, estimated to be around 29 Tg/yr (Smith *et al.*, 2000). Microbial communities oxidise methane to $CO_2$ in both aerobic and anaerobic environments. However, while aerobic oxidation has been clearly observed in soil systems, anaerobic oxidation has only recently been described in terrestrial systems (Bannert *et al.*, 2012;

Blazewicz *et al.*, 2012; Gauthier *et al.*, 2015). As a result, most studies into anaerobic methane oxidation cover aquatic and marine environments.

Microbial methane oxidation in soils, sediments and marine environments is completed by methanotrophs. Methanotrophs are a specialist subset of methylotrophs, i.e. organisms that can oxidise single carbon compounds, for example methanol. Methanotrophs are generally considered to have an obligate life strategy, only having the capability to utilise methane as their sole carbon source for respiration. Furthermore, the enzymes required are only seen in the methanotroph genera. This makes them attractive indicators for natural gas flux. However, over the last 20 years organisms have been observed that can oxidise methane and a range of other substrates, including other SCAs. These facultative methanotrophs may be useful indicators of natural gas seepage indicating the presence of SCAs.

Aerobic methane oxidation is considered ubiquitous in oxygenated soil systems. Bacterial species exploit this ecological niche. methane is initially oxidised to methanol via methane mono-oxygenase (MMO) (Figure 1.3). A dehydrogenase oxidises the methanol to formaldehyde before formaldehyde is oxidised to formate and finally to $CO_2$ (Semrau *et al.*, 2010).

The first enzyme in this pathway, MMO, can either be soluble (sMMO), or a particulate form bound to the inner cell membrane (pMMO) (Figure 1.3). The two metalloenzymes are analogous, i.e. while they fulfil similar functions their protein sequences, tertiary structures and complexed metal ions differ (for reviews: Hakemian and Rosenzweig, 2007; Kopp and Lippard, 2002). The majority of methanotrophs possess pMMO (McDonald *et al.*, 2008); however, a small group possess both enzymes, with the sMMO enzyme produced at low copper to microbial biomass concentrations (Semrau *et al.*, 2013). Two genera, *Methylocella* and *Methyloferula*, are only capable of using sMMO (Vorobev *et al.*, 2011; Dedysh *et al.*, 2000).

The two enzymes have differing kinetics. Whilst methanotrophs are generally obligate methane users, the MMO enzymes are capable of co-metabolism where the enzyme oxidises substrates other than methane likely without any further metabolism possible within that microbe. That said, the sMMO has a broader substrate specificity compared to pMMO, capable of oxidising $C_1$-$C_8$ alkanes (van Beilen and Funhoff, 2007; Colby *et al.*, 1977).

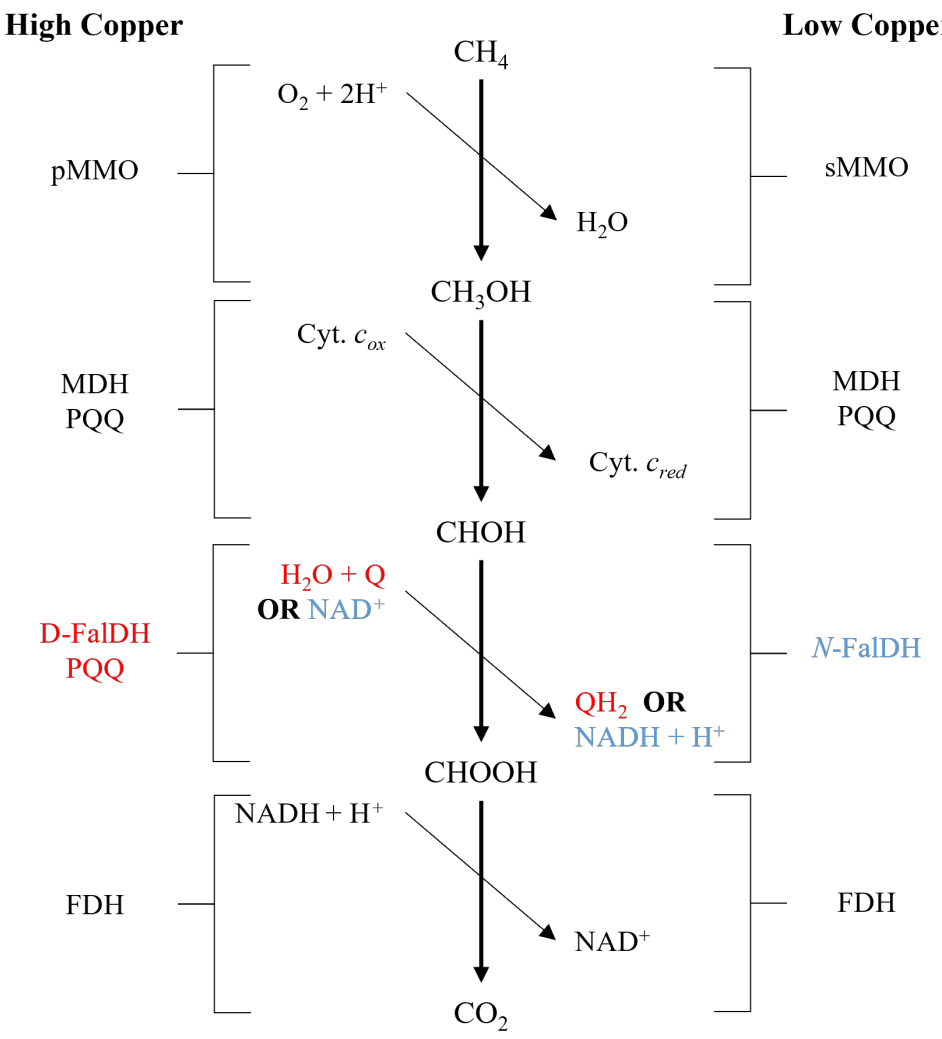



**Figure 1.3: Hypothetical outline of aerobic methane oxidation.** The enzymes involved can vary depending upon environmental copper concentration and the methanotroph. Methanotrophs with copies of all genes appear to be capable of switching phenotype depending upon the environmental copper concentration. The pathway occurs across the inner membrane, with MDH existing within the periplasm while pMMO and D-FalDH are bound to the inner membrane. Within the inner membrane a ubiquinone-ubiquinol cycle (Q cycle) occurs, generating H+ for the initial pMMO mediated oxidation of $CH_4$ (not shown). The substrates used in the oxidation of formaldehyde (CHOH) are colour coded to the enzyme which uses them. **Cyt. C**: Cytochrome-C., **D-FalDH**: dye/quinone-linked formaldehyde-dehydrogenase, **FDH**: Formate-dehydrogenase, **MDH**: methanol-dehydrogenase, **N-FalDH**: NAD linked formaldehyde-dehydrogenase, **pMMO**: particulate methane mono-oxygenase, **PQQ**: pyrroloquinoline quinone (co-factor) **sMMO**: soluble methane mono-oxygenase. Figure based upon: Semrau *et al.* (2010).

### 1.6.1 Obligate Methanotrophs

Aerobic methanotrophs were originally split into two groups, or types, depending upon their membrane structure: type I methanotrophs having more ordered bundles of paired membranes compared to disc-shaped vesicles seen in Type II (Davies and Whittenbury, 1970; Whittenbury *et al.*, 1970). . Further classification produced five genera *Methylomonas, Methylobacter, Methylococcus, Methylocystis* and *Methylosinus*, based on morphology and resting stage (Whittenbury *et al.*, 1970). Op den Camp *et al.* (2009) listed 16 distinctly described genera within the proteobacteria. The two taxonomic groups arrange methanotrophs into $\gamma$-proteobacteria (Type I) and $\alpha$-proteobacteria (Type II). The mechanism of carbon fixation differs between the groups, with Type I using the ribulose monophosphate pathway (RuMP) versus the serine cycle in Type II. A third type (Type X or Type Ib) was created to accommodate a subset of the Type I, $\gamma$-proteobacteria which expresses enzymes from the serine cycle despite predominantly using the RuMP pathway and which could grow at warmer temperatures (Hanson and Hanson, 1996). As observations of new phyla and genera have accumulated, some examples have been difficult or impossible to fit into this classification. This has led to suggestions that a newer method of classification may be required (Semrau *et al.*, 2010; Op den Camp *et al.*, 2009).

The pMMO enzyme is evolutionarily highly conserved and therefore can be used in phylogenetic analysis of communities (Holmes *et al.*, 1995). The catalytically active $\beta$-subunit, encoded by *pmoA*, is frequently used in molecular work. Using molecular tools instead of cultivation enables a greater diversity to be observed. Knief (2015) used *pmoA* to explore the diversity of and habitat of methanotrophs using the 15000 sequences available in GenBank, of which >3% were from cultured methanotrophs. Their work demonstrated that genera can have specific habitat preferences, for example upland soil clusters which decrease in prevalence in intensively managed soils (support by Knief *et al.*, 2005). Other genera were more generalist but could still contain habitat specific species.

That said, most methanotrophs are mesophilic and neutrophilic (Knief, 2015).. Some methanotrophs do have the ability to grow in more extreme environments, for example the thermoacidophilic species from the *Verrucomicrobia* phylum (Op den Camp *et al.*, 2009). Between the Proteobacteria Type I appears to be more competitive in low methane environments than Type II. Intriguingly, there is also some evidence that temperature selects for groups of methanotrophs. Low temperatures, between 0°C and 10°C appear to select for Type I methanotrophs: studies of soils and sediments from cold climates have repeatedly observed a dominance of Type I genera such as *Methy-*

*lobacter, Methylosarcina* and *Methylomonas* (He *et al.*, 2012; Martineau *et al.*, 2010; Yergeau *et al.*, 2010; Vecherskaya *et al.*, 1993). Increasing temperature leads to greater methane oxidation and a loss of Type I dominance: Börjesson *et al.* (2004) observed only Type I activity at temperatures between 3°C -10°C with Type II activity occurring alongside Type I at 2°C in samples from the same site. However, the mechanism(s) for this are unclear and Liebner and Wagner (2007) suggested that the evolutionary differences between Type I and Type II should not provide an advantage to either group in cold environments, observing that changes in community were more closely linked to methane concentration.

### 1.6.2 Facultative Methanotrophs

A small number of methanotrophs have a limited facultative lifestyle. Acetate can be utilised by several *Methylocystis* species and by *Methylocapsa* (Belova *et al.*, 2011; Im *et al.*, 2011; Dunfield *et al.*, 2010), (Belova et al. 2011; Im et al. 2011; Dunfield et al. 2010), both are classified as Type II methanotrophs. Im *et al.* (2011) also demonstrated that *Methylocystis* strain SB2 could grow on ethanol. An even greater shift is lifestyle strategy is seen with the *Methylocella* genus. These are facultative methanotrophs capable of growth on variety of multi–carbon compounds such as ethanol and acetate (Dedysh *et al.*, 2005).

Species within the *Methylocella* genus do not have the pMMO enzyme. Instead they can only express a version of sMMO. (Dedysh *et al.*, 2000) first isolated the genus (species *Methylocella palustri*) from Sphagnum peat bogs. *M. palustri* is an acidophilic, $\alpha$–proteobacterium and isolates possess enzymes specific to the serine pathway of carbon fixation. However, the *Methylocella* genus lacks the *pmoA* gene and the distinct membrane morphology seen in other methanotrophs from the Proteobacteria. Furthermore, 16S rDNA separates the genus from other Type II methanotrophs (*Methylocystis* or *Methylosinus*) with the nearest relative being *Beijerinckia indica* (similarity 96.5%) (Dedysh *et al.*, 2000). The *mmoX* gene also appears analogous to *mmoX* from other methanotrophs (Dedysh *et al.*, 1998) with later authors using specific primers for its detection in quantitative polymerase chain reaction assays (Rahman *et al.*, 2011a; Farhan Ul Haque *et al.*, 2018) as previous *mmoX* primers failed to amplify the version in *Methylocella*. Due to their facultative traits *Methylocella spp.* may possess a competitive advantage compared to other methanotrophs by being capable of growth on multiple components of natural gas. Farhan Ul Haque *et al.* (2018) observed that the *Methylocella* genus formed a large component of methanotroph communities around natural gas seeps. The need for a unique set of primers specific to *mmoX* in *Methy-*

*locella* and the facultative lifestyle of the genus may be useful when attempting to differentiate between gas sources.

### 1.6.3 Note on Ammonia Oxidisers

Ammonia oxidisers also oxidise methane to methanol in aerobic environments. These species do not appear to be capable of fully oxidising methane and therefore cannot generate ATP from this process. Instead the ammonia mono–oxygenase (AMO) appears homologous to pMMO with a similar copper containing active region allowing methane oxidation during periods of low ammonia concentration (Holmes *et al.*, 1995; Hyman *et al.*, 1988). As a result primers, for use in PCR, exist that can amplify both *amo* and *pmo* genes (Dumont and Murrell, 2005) and it is plausible that some primer sets may amplify both by mistake (Conrad, 2007).

### 1.6.4 Methanotrophs and Contaminated Land

Methanotrophs and the MMOs have received attention as potential biocatalyst tools for example for use in bioremediation (Holmes and Coleman, 2008). Methanotrophs and their methane mono-oxygenases have been reported to degrade halogenated hydrocarbons (Im and Semrau, 2011; Lee *et al.*, 2006). Methanotrophs, both Type I and II, have been isolated from variety of contaminated environments (McDonald *et al.*, 2006).That said, very little published work is available exploring the presence and diversity of methanotrophs in contaminated land sites. The majority of work with methanotrophs and MMOs appears to have focused upon groundwater and the treatment chlorinated compounds (Inoue *et al.*, 2020).

## 1.7 Alkantrophs

Alkanes present an inherent challenge for microbial degradation given their insolubility in water and the chemically stable C-H bond. That said, microbial degradation of alkanes is possible through a variety of biochemical pathways across several genera (van Beilen and Funhoff, 2007). Both eukaryotes and prokaryotes have genera capable of growth on alkanes under a range of environmental conditions. Examples of both filamentous fungi and yeasts are capable of growth on alkanes (review by: Van Beilen *et al.*, 2003). Furthermore, *Prototheca* spp., a genus of algae, have also been reported to oxidise C11-C17 *n*-alkanes through sub-terminal oxidation (Takimura *et al.*, 2014).

Typically, alkane oxidation will proceed through terminal oxidation to alcohol then

an aldehyde before entering $\beta$-oxidation generating acetyl-CoA for the citric cycle (Van Beilen *et al.*, 2003). The less common subterminal oxidation pathway produces a secondary-alcohol, which is converted to a ketone and then an ester which can broken down to a fatty acid (and another alcohol) before being uptaken by the $\beta$-oxidation pathway (Rojo, 2009). Methane oxidation does not follow these pathways and its carbon assimilation pathways are described in Chapter 1.6.1.

Both marine and terrestrial environments have a diverse range of microorganisms that degrade hydrocarbons, including alkanes (for reviews see: Rojo, 2009; Melcher *et al.*, 2002). Specialist hydrocarbon degraders, referred to as hydrocarbonclastic, have been reported and are described as having a restricted phenotype or growth when growing on sugars or amino-acids, for example *Alcanivorax* (Lai *et al.*, 2011; Yakimov *et al.*, 1998). Other organisms appear to have a more versatile metabolism and are capable of opportunistic growth on hydrocarbons and alkanes alongside other substrates (Tian *et al.*, 2019; Maugeri *et al.*, 2002).

Terrestrial sites with increased burdens of oil contamination see an increase in species capable of growth on alkanes and other hydrocarbons (Alrumman *et al.*, 2015). However, it appears that soil factors impact the development and final composition of bacterial communities degrading hydrocarbons (Bundy *et al.*, 2002). Furthermore, whilst the presence of hydrocarbons might stimulate some microorganisms (Bundy *et al.*, 2002) the toxicity can lead to a loss in both diversity and activity (Klimek *et al.*, 2016; Labud *et al.*, 2007).

The initial stage in the aerobic metabolism of oxygenases is catalysed by hydroxylases as part of mono-oxygenases or cytochrome P450 family of enzymes (van Beilen and Funhoff, 2007). Between and within each group there are unique enzymes each with a preferred substrate range, usually dependent upon alkane length and structure. It is common for alkane degraders to carry multiple enzymes with differing substrate preferences (van Beilen and Funhoff, 2007). Two large families of mono-oxygenases enzymes are discussed here, these contain multiple enzymes capable of reducing a variety of organic compounds including SCAs. Due to methanogenesis a subsurface seep may be obscured, therefore detecting organisms capable of growth on SCAs using these genes may be advantageous when differentiating between sources.

### 1.7.1 Soluble Di–Iron Mono–Oxygenases

Propane mono-oxygenase (PMO), butane mono-oxygenase (BMO) and sMMO are Soluble Di-Iron Mono-Oxygenases (SDIMO). These multi-component enzymes catal-

yse the addition of a hydroxyl group to organic compounds via a hydroxylase (with a di–iron centre) associated with NADH reductase and a coupling protein (Leahy *et al.*, 2003). There are six groups of SDIMO, classified by both their structure and substrate; Groups 1 and 2 are aromatic oxygenases, while groups 3-6 are aliphatic oxygenases (with some exceptions). Leahy *et al.* (2003) demonstrated that the different SDIMO groups are closely related, having probably been dispersed between a small number of genera through horizontal gene transfer. As a result, the hydroxylase protein is well conserved, especially within individual SDIMO groups. Furthermore, between the SDIMO groups, the hydroxylase's iron-ligand residue is well conserved.

Coleman *et al.* (2006) used the genes encoding the iron-ligand residue as a target in a nested PCR strategy. The primer sets generated have been used with environmental samples. Of particular interest is their use by Miqueletto *et al.* (2011) in an oil and gas prospection technique, highlighting potential use of SDIMO primer sets in environmental monitoring. As with the molecular techniques measuring the abundance of MMO, changes in SDIMO genetic abundance may be indicative of increased alkane concentrations. Specific primer sets for the hydroxylase component also exist for both PMO and BMO; the newer molecular hydrocarbon prospecting techniques have used the PMO set (for example Zhang *et al.*, 2017; Liu *et al.*, 2016).

It is possibly surprising that only one SDIMO, sMMO, appears to have the capability to oxidise methane. BMO, from the same SDIMO group, has been reported to be inhibited by methanol potentially due to the structure of the active site (Halsey *et al.*, 2006). It may also be due to the stability of methane and the resulting requirement for a high activation energy to oxidise it to methanol; Lipscomb (1994) suggested this could be the reason why sMMO has a broad range of substrates it can co-metabolise.

### 1.7.2 Copper Mono–Oxygenases

The copper membrane mono–oxygenases (CuMMO) include pMMO and AMO. Alongside these two enzymes authors have reported CuMMOs with roles in SCA oxidation, (for example Coleman *et al.*, 2012) described a hydrocarbon mono–oxygenase with $C_2$-$C_4$ activity. Furthermore, pMMO has been reported to co-metabolise substrates other than methane leading to the possibility that methanotrophs could have a role in bioremediation (Im and Semrau, 2011).

The structure of the CuMMO enzymes has reportedly been more difficult to ascertain due to their membrane bound nature (Hakemian and Rosenzweig, 2007). As a result, the precise nature of the catalytic subunit is unknown. From pMMO studies, these met-

alloenzymes probably utilise a catalytic copper centre within the $\beta$-subunit (Hakemian and Rosenzweig, 2007). Primer sets targeting a conserved region of *pmoA* have been published and widely used with environmental samples for the detection of methanotrophs (for example: Costello and Lidstrom, 1999; Holmes *et al.*, 1995). Detection of *pmoA* has also been used in phylogenetic analysis of environmental samples with results comparable to 16s RNA phylogenies. More universal *xmoA* primers for qPCR have recently been developed and used with environmental samples (Rochman *et al.*, 2020).

## 1.8 Overview of Thesis

Methane is a potential pollutant, notably as a potent greenhouse gas in the atmosphere. The rich legacy of onshore oil and gas exploration may contribute to the release of methane into the atmosphere from the subsurface. This release may become a more serious problem as legacy oil and gas wells age and potentially degrade. Furthermore, the potential for future onshore hydrocarbon extraction could continue to add to this legacy. As a result, techniques to monitor, locate and mitigate sources of methane emission need to be developed; these ideally need to be easy to deploy and to have low cost given the number of wells in the UK (and globally). This thesis explores methods to detect methane and natural gas flux through soils, focusing on using soil microbiology as an indicator.

### 1.8.1 Project Approach

In recent oil and gas prospecting techniques authors have used molecular tools to detect anomalies above known reservoirs. For example the abundance, diversity and ratios of *pmoA* and *prmA* (Zhang *et al.*, 2017; Liu *et al.*, 2016). These have detected differences between 'background' (unpolluted) soils and those experiencing micro-seepage. It is plausible that a similar approach, i.e. the use of PCR and qPCR, may work on smaller scales to monitor historic onshore oil and gas wells. Furthermore, these approaches may be better suited to these smaller scales than culture-based approaches due to increased accuracy and precision.

There are several published primer sets available for the amplification of functional genes involved in alkanotrophy. Methanotrophs can be detected using primers that target the genes encoding the first mono-oxygenases in the metabolic pathway of methane. Crucially, these genes and encoded enzymes are unique to methanotrophs. The majority of methanotrophs have the particulate form of MMO encoded by *pmoA*. The detection of sMMO, through the amplification/detection of *mmoX*, is also possible. Compared to *pmoA*, these primer sets are less inclusive given the limited prevalence of the enzyme in methanotrophs. However, the quantification of sMMO might highlight changes in methanotroph community composition between samples when considered alongside *pmoA*. To further explore methanotrophic community changes *Methylocella* specific primer sets can be used. The detection of other alkanotrophs using PCR, targeting either CuMMO or SDIMO related genes, may also aid the detection of oil and gas wells leaking natural gas. Using historic primer sets comes with caveats, principally the lack of published sequences at the time of their original design and pub-

lication, leading to primer sets which were never designed for, and cannot amplify, some taxonomic groups. Generally, such weaknesses can be considered during their use and when interpreting results. Overall, this project aims to detect soil community composition changes that may be related to increased gas flux using modern molecular techniques like PCR.

### 1.8.2 Key Hypothesis

The central hypothesis is that natural gas flux can be detected using changes in soil microbial community composition. Of particular focus are the response of methanotrophs and broader alkanotrophs to chages in natural gas flux. Furthermore, this flux can be detected at a spatial resolution that enables the identification of hotspots, potentially from oil or gas wells, at field scales (tens of metres). Finally, the source of the seep can be attributed to thermogenic rather than biogenic processes due to the presence of other SCAs, also detectable using changes in microbiology.

### 1.8.3 Thesis Objectives

This thesis sets out to develop and test methods to detect changes in microbial community composition in soils. Given the required spatial resolution these methods need to be precise. Secondly, as this is a monitoring tool the approach needs to be easily reproducible. The development of quantitative PCR and next generation sequencing tools are a primary focus. These are anticipated to meet the thesis aims better than historical culture-based methods. The methods are tested in real world sites to enable their improvement and demonstrate their potential. Work also explores the long-term changes in a soil's microbial community structure after a high methane exposure.

## 1.9 Project Layout

The thesis contains the following chapters. The Methods chapter (Section 2) lays out the final methods used throughout the project. The Preliminary Work (Section 3) chapter describes how those methods were developed for use in the subsequent chapters.

### 1.9.1 Preliminary Work

The thesis first lays out the preliminary work undertaken. This includes the selection of PCR primers and the development of PCR protocols used to amplify key genes expressed during the oxidation of alkanes by bacteria. Initial work exploring spatial variability in the field and lab equipment set up is also presented. Also discussed are challenges when handling and interpreting data from environmental samples. This chapter could be viewed as a continuation of the methods chapter, as it sets out and describes the method development process that produced the methods used throughout the remainder of the work.

### 1.9.2 Alpine Natural Gas Analogue Sites

Two field sites around gas seeps with a probable thermogenic origin were sampled. This chapter explores field sampling and data analysis for these sites. Crucially, the sites both had a well characterised macro-seep to 'ground' the sampling strategy and clearly test the central hypothesis. Furthermore, the presence of strong natural gas fluxes enabled an exploration of the use of facultative methanotrophs as indicator species.

### 1.9.3 Eakring – A UK Oil Field

A second site was sampled to further develop and test the approach. This site is an abandoned oil field which still has evidence of the wells at the surface. Well integrity was unknown, leading to the site providing a 'blind' test of the methods. The site also enabled an exploration of soil sampling depth.

### 1.9.4 Long Term Soil Priming Effects

Previous work suggested that microbiology could act as an indicator of the average flux over time. How long microbiology remained in a disturbed state was explored in this chapter. This also tested if a false positive signal could be detected from a site with a history of methane leakage that had subsequently ceased.

### 1.9.5 Gasworks Remediation Sites

Finally, the methods are explored as markers of natural attenuation in contaminated land sites as another potential use of soil biomonitoring.

# Chapter 2

# Methods

## 2.1 Field Methods & Sample Handling

Field sites need both *in-situ* characterisation and *ex-situ* analysis of samples collected. During sampling campaigns estimates of surface gas fluxes were also taken. The soil texture and site was described. Photos were taken if deemed necessary. Any samples collected were treated identically to preserve the samples until laboratory analysis could be carried out.

### 2.1.1 Gas Flux

Estimates of gas fluxes between the soil surface and the overlying atmosphere were made on site using a portable fluxmeter (West Systems, Italy – FI). This was completed before the sampling site was disturbed by either soil collection or compaction by people and equipment; such disturbances influence gas diffusion rates by altering the structure of the upper few cm of the soil. The fluxmeter could simultaneously analyse both methane and carbon dioxide fluxes from the soil surface. Fluxes were determined using an aluminium chamber approximately 3 L in volume covering a soil surface area of 315 $cm^2$; the chamber was firmly pressed into the surface to generate a seal. Within the chamber a paddle circulated air to ensure mixing of the chamber's atmosphere. A pump drew the atmosphere from the chamber into two instruments at 20 ml $s^{-1}$. Methane was detected in an infra-red (IR analyser) using a multipass cell and a tunable diode laser. Carbon dioxide was detected using a separate non-dispersive infrared gas analyser, a Licor LI-820.

A Bluetooth connection to a mobile phone enabled the operator to see the instrument

output in real time and change instrument settings, as required. A connection to a phone enabled the data collected to be named, annotated and for a set of GPS co-ordinates to be generated for the measurement being taken. While taking measurements, the fluxmeter pump was run for 30 s with the chamber off the ground to flush the instrument. After flushing, the chamber was immediately placed on the sample site. The software included generated an accumulation plot from the instrument providing the user with an in-field estimate. After three min, sampling was stopped and the data saved as a text file to be downloaded onto a computer for further analysis.

#### 2.1.1.1 Interpolation

To visualise gas fluxes across a site the data were interrogated by inverse distance weighted (IDW) interpolation using spatial packages within R in Table 2.1 (p=2 in all plots). Cross-validation of the IDW interpolation was used to produce the root mean squared error (RMSE) between the predicted and observed values. Interpolation was also used with normalised qPCR data which was generated using DNA from a single homogenised soil sample.

**Table 2.1: R packages used for the production of IDW plots.** Packages were also used in basic statistical analysis of plots (e.g. histograms, means, modes etc.) and to export the plots as a raster format for use in QGIS. Base R major version: 4.

| R Package | Version | Reference |
|---:|---:|---:|
| geoR | 1.8-1 | Ribeiro Jr *et al.* (2020) |
| gstat | 2.0-7 | Pebesma (2004) |
| nlme | 3.1-152 | Pinheiro *et al.* (2021) |
| raster | 3.4-10 | Hijmans (2021) |
| rgdal | 1.5-23 | Bivand *et al.* (2021) |
| sf | 0.9-8 | Pebesma (2018) |
| spm | 1.2.0 | Li (2019) |
| tmap | 3.3-1 | Tennekes (2018) |

### 2.1.2 Sample Collection

All experimental and survey work required a large element of laboratory investigation. The majority of work interrogating soil microbiology needed to be completed in the

lab. Sample collection, transport and storage had to be considered to ensure the validity of the samples and results generated from them. Field sampling methods were carefully designed to minimise contamination and sample degradation. Clean nitrile gloves were worn at all times when handling samples and new, clean sample containers were used to minimise contamination.

#### 2.1.2.1 Auger

Soil samples were collected from various depths. Sample depth depended upon the soil conditions, site history and the work being undertaken. Samples were collected using an open-face soil auger. The shaft, cutting face and tip were cleaned using disinfectant and rinsed in de-ionised water immediately before use. Once the required depth had been reached, the material within the auger was collected in a labelled plastic universal bottle, or, zip-lock, plastic bag from which the air was forced out and the bag sealed.

### 2.1.3 Transport from the Field

Samples were transported doubled bagged or with the sample container bagged. Where sample weight was not a concern, for example at local sites, zip-lock bags were preferred. Samples were doubled bagged to reduce the risk of contamination or loss of material (including water where gravimetric analysis was to be carried out). Samples were immediately placed in cool boxes containing ice-packs or similar. This stabilised the samples, ensuring they were not exposed to extremes of temperature or light that might impact the microbiology or moisture content. Samples were transported back to the lab., or the field lab., within 12 hours.

## 2.2 Soil and Gas Measurements

As discussed in Chapter 1.5 soil is highly heterogeneous. Only two key variables were objectively measured at every site: pH and moisture content. Other observations included the texture, colour, smell, and vegetation cover. Cumulatively, this detail provided enough background information on the soil variables that were likely to have a significant impact on methane flux and oxidation at the scale investigated.

### 2.2.1 pH Measurements

To measure pH, 20 g soil was mixed with equal parts by weight of deionised water (20 g) and mixed to produce a slurry. The use of water at a 1:1 ratio was selected due to

issues transporting and using reagents in field. Mixing began with vortexing or shaking for 30 seconds to promote disaggregation. In the laboratory the slurry was left for 30 min on a roller, in the field the slurry was shaken six times over 30 min (five-minute intervals). The slurry was left to equilibrate for an hour at room temperature before pH was measured. A Perkin-Elmer Probe was used in the laboratory. In the field a Hanna Probe was used. In both scenarios a three-point calibration of the pH probe, with buffers of pH 4, 7. and 10, was completed immediately before testing samples. Buffers were at the ambient temperature (same as the slurry). The pH probe was used when the slope of the calibration was between 92-102 %. The electrode was rinsed in DI water and blotted with tissue between samples.

### 2.2.2 Soil Moisture Content

The moisture content of the soil was determined gravimetrically. Soil was measured to 3dp into glass Wheaton® bottles of a known weight (also to 3dp). Ten grams of soil were used in each replicate. Water was driven off the sample by incubating it at 105°C overnight and the sample re-weighed. The change in mass was assumed to be caused by water loss. This was calculated as a percentage using the original mass of the sample.

The use of Wheaton® bottles was due to the presence of contaminants in some samples. A Wheaton® bottle more securely contained the sample than ceramic dishes. If the sample contained volatile contaminants, water was driven off at 40°C for two weeks to minimise the release of volatiles.

### 2.2.3 Gas Chromatography

Culture headspaces gases were sampled and analysed using Gas Chromatography. Headspace samples were stored in 3 ml clear, glass exetainers® (Labco Eng.) these were sealed with a plastic screw cap and self-healing rubber septa. These were evacuated using either a 20 ml syringe or vacuum pump. Single 1 ml samples for each time point, were drawn off using a 1 ml syringe and immediately stored in the Exetainer®.

Gas samples were analysed for methane concentration using gas chromatography. A Modus permanent gases and light hydrocarbon column was used in an Agilent 6890, with a flame ionisation detector. The sample was injected at 80°C, with a subsequent ramp of 40°C $min^{-1}$ to a final temperature of 200°C. The oven held this temperature for a further minute before completing the method. The column had a constant flow of He ( 29 ml $min^{-1}$). Methane was eluted at around 0.9 min. For each sample, 0.2 ml of

gas was injected. Three methane standards were used, 100000 ppm, 10000 ppm and 100 ppm. Standards were repeated every 12 samples.

Samples where carbon dioxide, and oxygen concentrations were also required were analysed on a different machine. An Agilent 7820A with a Restek® 80468-810 packed column and thermal conductivity detector. The carrier gas, He, was at a constant flow of 9 ml $min^{-1}$. The sample was injected at 23°C and held for three min, followed by a ramp of 3°C $min^{-1}$ to 40°C, followed by a second ramp to 90°C at 5°C $min^{-1}$.

## 2.3 Molecular Methods

Community composition was analysed using PCR and sequencing. Quantitative PCR (qPCR) was used to estimate the relative abundances of methanotrophs and alkanotrophs using 16S rRNA gene copy number as a proxy for total bacterial community size.

### 2.3.1 Sample Storage

Samples were immediately stored at 4°C or frozen at -18°C. DNA was extracted as soon as possible to ensure the sample were still representative of the environment. Samples stored at 4°C underwent DNA extraction within 48 hours. Samples were thoroughly mixed within their sample container to produce a single homogenous sample for use. Samples were handled in an aseptic manner to reduce contamination; this involved the use of a laminar flow cabinet and nitrile gloves. Between samples, gloves were replaced, the work bench wiped down with blue roll (moistened with methanol or 5% bleach) and any tools or glassware were rinsed or wiped down with methanol. Disposable weighing boats were used. Samples were stored at -80°C where long term storage was needed both as aliquoted soil samples and as DNA extracts.

### 2.3.2 DNA Extraction

DNA extraction was performed using a commercial kit: FastDNA™ SPIN (MP BioMedicals - US). As a commercial kit was used, the exact composition of the buffers and washes is unknown, but their purpose can be understood. The use of other extraction techniques, for example a phenol-chloroform extraction, would enable a complete understanding of the chemistry, but would generate problematic waste and require the use of fume hoods instead of a laminar flow hood.

Soil sample weight varied from 0.25 - 0.5 g depending on the soil encountered at sam-

pling sites. Smaller masses were used with material from drier sites or those suspected to have high inhibitor content (for example clays). To begin cell lysis and aid disaggregation of soil 1 ml of chaotropic agents which disrupt hydrogen bonds and buffer were added.

Samples were ball-beaten at 6 m $s^{-1}$ for 40 s (vertical speed). Ceramic, silica and glass beads of varying sizes are used to mechanically break down the sample during agitation in FastPrep (MP BioMedicals - US). The buffers, added before beating, stabilised the nucleic acids released by the process.

Immediately after ball-beating the homogenised mixture was centrifuged at 14,000 g for 10 min to pellet soil and cellular debris. The supernatant, consisting of soluble soil components, buffer, surfactants, and cellular material, was moved to fresh Eppendorf tubes. 250 $\mu$l protein precipitation solution was added to encourage the flocculation of proteins and cellular material that was not nucleic acid. The flocculated material was removed by centrifuging at 14,000 g for 5 min, after which the supernatant was moved to a clean 15 ml tube.

The supernatant was mixed for 2 min with 1 ml of binding matrix by repeated inversion; a 15 ml tube was used to improve mixing. The binding matrix was composed of silica beads suspended alongside chaotropic salts. Chaotropic agents promoted the binding of DNA to the silica beads over water. After mixing, the suspension was given three min for the silica beads to settle before excess supernatant was removed. The DNA, bound to the suspended silica beads in the binding matrix, was transferred to a spin filter. This was centrifuged at 14,000 g for 1 min leaving the silica beads (and the DNA) on the filter.

The bound DNA and silica matrix was resuspended in 500 $\mu$l of ethanol rich solution. This 'washed' the sample enabling further purification and removed inhibitor carryover from the soil. Ethanol is chaotropic, resulting in stronger bonding between the bead and DNA but weaker binding between the beads and other carbon residues. The wash buffer also had an unknown mixture of other compounds to aid removal of contaminants. The sample was centrifuged twice at 14,000 g for one minute to remove the wash buffer with the catch tube, which was emptied after each spin. To remove the ethanol, which would impact future analysis of the samples, the filter was left to dry for 5 min at room temperature with the lid open.

The DNA was stripped from the beads via resuspension in 90 $\mu$l ultra-pure water. This was incubated for five minutes at room temperature. Finally, the suspension was

centrifuged at 14,000 g for 1 min, with the eluent caught in a clean micro-centrifuge tube.

During DNA extraction three other steps were employed to improve yield and quality. Ethanol is a potent inhibitor of PCR but is needed to stabilise the DNA-silica binding during sample clean-up. A simple 'drying' step was extended to 5 min to maximise the loss of ethanol. Secondly, the volume of ultra-pure water used to elute the DNA from the silica binding matrix was kept as low as possible to maximise DNA concentration. Thirdly, the silica binding matrix was incubated at 55°C for 5 min to improve yields; this step was only used in extractions where a low mass of soil was used due to the presence of soil contamination.

### 2.3.3 DNA Quantification

A Qubit® (Invitrogen, SG) uses fluorimetry to measure the concentration of dsDNA. This was used to assess the quantity of DNA extracted from the raw sample. It was also used in plasmid production and sequencing workflows. The fluorescent dyes used here bind to intact dsDNA, therefore higher concentrations of DNA yield a greater fluorescent signal. Using a pair of standards, provided by the manufacturer, the concentration of DNA can be found. To ensure accuracy the assay was run in triplicate. A high specificity master mix of 1 part Qubit® reagent to 199 parts buffer is prepared. The high specificity assay has a quantification range of 10 – 100 $\rho$g $\mu$l$^{-1}$ (in the raw sample). When preparing standard series, 2 $\mu$l of plasmid stock were added to 198 $\mu$l of master mix. Qubit® standards were prepared by adding 10 $\mu$l standards to 190 $\mu$l master mix. Solutions were vortexed for 2-3 sand left at room temperature for five min. Fluorescence was measured within the Qubit® fluorometer.

### 2.3.4 Sample Dilution

To prepare DNA samples for qPCR assays of functional genes they were diluted with ultra-pure water. This aimed to overcome any PCR inhibitors carried over from the soil sample, see Chapter 3.6 for a complete rationale. To estimate the optimum dilution, a subset of samples from each site was tested for inhibition using the 16S rRNA qPCR assay. The 16S rRNA gene assay was selected as it amplified the most abundant bacterial gene within the sample, therefore offering the greatest chance of amplification, even with samples rich in potential inhibitors. DNA extracts were selected from across the field site using the weight of the soil used in the DNA extraction as a guide. The DNA extracts from larger soil masses were thought to contain more soil inhibitors than

those extracts from a soil sample with a lower soil mass. As a result, using extracts from both the smallest soil mass and the greatest provided an understanding of the range of inhibition within the extracts.

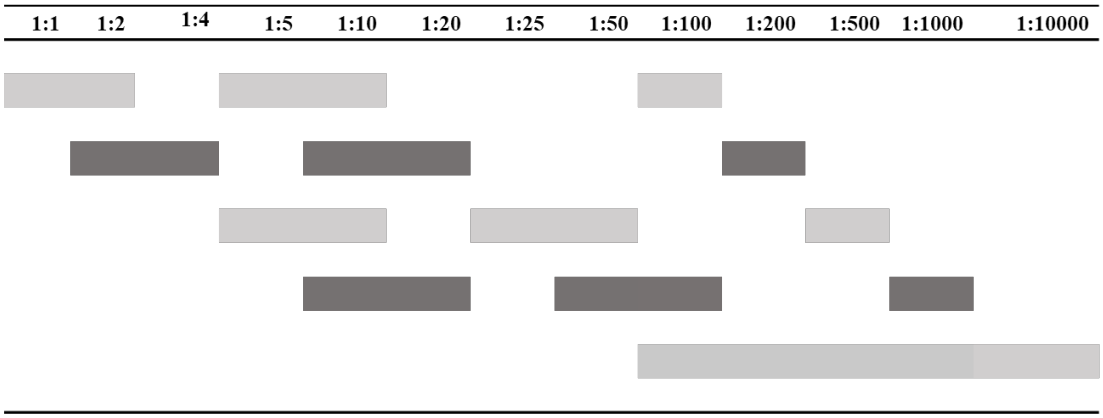



**Figure 2.1: Dilutions and the groupings used to calculate E values.** Dilutions are in the top row. The coloured horizontal bars indicate groups of five of dilutions which were used for efficiency calculations. These determined the appropriate dilution ranges for field samples

Selected DNA extracts were diluted down to 1:10,000 in 12 increments. Dilutions used are shown in Table 2.2. All 12 dilutions plus an undiluted aliquot of the DNA extract were used in the qPCR assay. Alongside the dilution series using experimental samples, a standard series using plasmids from cloning was also run for comparison and to act as a positive control in the assay. Once the assay was complete, efficiency values (E values) were calculated (from the quantification cycle ($C_q$) values) for 5 groups of dilutions, as shown in Figure 2.1 with each group covering a 100-fold dilution. An average E value for that dilution range was generated using all the samples undergoing the assay. This average E value was then compared to the E value for the standards, as it was assumed this would exhibit little to no inhibition. The group of dilutions with the closest E value to that of the standards was considered the optimum dilution to use, with the smallest dilution of the group performed on all the samples ready for analysis.

### 2.3.5 DNA Storage

DNA was stored at -20°C immediately after extraction. Samples requiring longer term storage before analysis were frozen at -80°C. To avoid repeated freeze-thaw impacting sample quality, samples were aliquoted enabling the original sample to be revisited. This also reduced the risk of contamination with a 'clean sample' being kept frozen for future use.

### 2.3.6 Eurofins Sequencing – Tubeseq

Genetic sequence information was needed to confirm PCR products and successful cloning of plasmids to create qPCR standards Chapter 2.3.7. Samples were used in the relevant PCR assay, this product was then ligated, transformed and grown before plasmid purification using the TA Cloning® kit (Invitrogen, US). This to produce a plasmid stock for use in qPCR as a standard. The template PCR product was ligated into a plasmid before competent *Escherichia coli* cells were transformed with the plasmid and grown.

Target template was produced using the relevant PCR as described in 2.3.8. A 1:1 molar ratio of PCR product (hereafter referred to as the insert) to plasmid vector (pCR® 2.1 Invitrogen, US) was used. Fresh PCR product (less than 24 hours old) was cleaned using Diffinity RapidTips® (Chiral Technologies, US) to remove salts, reagents, and dNTPs. The cleaned product's concentration was quantified using a qubit (Chapter 2.3.3. The required mass of insert calculated using Equation 2.1 allowing the correct volume of insert to be added to the ligation reaction.

$$ng = \frac{bp_n \times 50ng}{3931bp} \tag{2.1}$$

Equation to calculate the required mass of insert (in ng) needed to create a 1:1 ratio of insert to plasmid vector. $bp_n$ is the number of bases in the insert. The plasmid has 3931 bp.

The ligation reaction contained, 1 $\mu$l of T4 DNA ligase, 2 $\mu$l buffer and 2 $\mu$l pCR®2.1 plasmid vector (at 25 ng $\mu l^{-1}$). After the addition of insert, the reaction was made up to 10 $\mu$l using ultra-pure water. The reaction was mixed using the force of a pipette and incubated at room temperature for 30 min. This reaction mixture could then be stored at -20°C until transformation.

The circular plasmid containing the template to be cloned was transformed into TOP10 *E. coli* cells. Briefly, 25 $\mu$l of One Shot® Competent cells were thawed on ice before 2 $\mu$l of the ligation reaction (containing the plasmid) was added and stirred gently using the pipette tip. The mixture was left on ice for 30 minutes before being heat shocked at 42°C for 30 s. This was immediately transferred to ice and 250 $\mu$l of room temperature super optimal broth (SOB) added. The culture was shaken horizontally for an hour at 37°C at 225 rpm.

Plates of lysogeny broth (LB) were prepared with containing 100 $\mu$g $ml^{-1}$ ampicillin.

These were spread with 40 $\mu$l of X-gal 40 mg ml$^{-1}$ and allowed to dry and equilibrate for 30 min. X-gal was suspended in dimethyl sulfoxide (DMSO) to enable spreading. For each transformation two plates were prepared: one plate was spread with 75 $\mu$l of the cell culture while the second used the remainder. Plated cells were grown overnight at 37°C After growth, plates were kept in the fridge for at least two hours to promote the development of blue/white colonies.

If transformation of the plasmid into the competent cell is successful, cells grow on the plate as a white coloured colony. The plasmid vector contains an ampicillin resistant gene and *lacZ*$\alpha$. Cells without the plasmid vector cannot grow due to the presence of ampicillin. The *lacZ*$\alpha$ alone is a non-functional peptide but complements *lacZ* in *E. coli* forming a $\beta$-galactosidase. X-gal is a synthetic compound composed of indole and galactose; $\beta$-galactosidase attacks this compound and the product oxidises to form a blue product. Successful ligation and transformation will enable growth in the presence of ampicillin. However, correct ligation breaks the plasmid inside *lacZ*$\alpha$ stopping its expression, providing a visual indication of the success of ligation and transformation. Cells with plasmids are capable of growth in the presence of ampicillin but those lacking the insert express *lacZ*$\alpha$ off the plasmid, leading to blue colonies forming.

Up to ten white colonies were picked from the plate and grown overnight individually at 37°C in 5 ml LB with 100 $\mu$g ml$^{-1}$ ampicillin. Following overnight growth, plasmids were purified from the culture using a QIAprep Spin Miniprep kit (Qiagen, Hilden, DE). Cells were centrifuged at 5,000 rpm for 5 min and the supernatant discarded. The pellet was resuspended and lysed under alkaline conditions using a sodium hydroxide (NaOH) and sodium dodecyl sulphate (SDS) buffer, which includes RNase. Lysis is driven by the SDS (a surfactant), solubilising lipids and proteins. Alkaline conditions denature nucleic acids and proteins. Lysis under these conditions is time sensitive and over-exposure will permanently denture plasmids, so lysis is only allowed to proceed for 5 min.

The lysis buffer is neutralised using a high salt buffer, which leads to the precipitation of SDS, proteins, lipids and chromosomal DNA. Centrifugation, at 13,000 rpm for ten minutes pellets the debris enabling 800 $\mu$l supernatant to be transferred to a spin filter. The sample is spun down at 13,000 rpm for one minute and the flow-through discarded. The sample is then washed twice, first with a buffer to remove trace nuclease activity and finally with a chaotropic buffer to remove salts. Each wash is centrifuged through the filter at 13.000 g for one minute. After washing, the plasmid is eluted from the spin filter using 50 $\mu$l of 10 mM Tris-Cl: the filter is left in contact with the eluent for one

minute before centrifuging for one minute at 13,000 rpm.

The concentration of the resulting purified plasmid was determined using a Quibit Assay as described in Chapter 2.3.3. If the required concentration of 50-300 ng $\mu^{-1}$ of plasmid had been obtained, 15 $\mu$l of the plasmid stock plus 2 $\mu$l of the correct primer (either forward or reverse) were mixed in an Eppendorf. Two samples per plasmid were sent to Eurofins TubeSeq service for sequencing, one with the forward and one reverse primer. The resulting FASTA file, supplied by Eurofins, was aligned using BLASTn on the National Center for Biotechnology Information (2021) website to ascertain the sequence's closest known gene.

### 2.3.7 Polymerase Chain Reaction

The polymerase chain reaction (PCR) enabled the rapid production of large quantities of DNA from a small starting template. This allowed for the detection of key genes used in oxidation of methane and other alkanes. Accurate quantification of these genes could be achieved using quantitative PCR (qPCR). Each primer set had unique thermocycling conditions which required optimising; described here are the final reaction conditions used after Preliminary Work (Chapter 3). Considerations and steps taken to optimise reaction conditions are described in Chapter 3.1, 3.2 and 3.3.

#### 2.3.7.1 Primer Sequences

Using Dumont and Murrell (2005) as a starting point, potential primers were identified. The programs used for thermo-cycling in the original papers were compiled as a starting point for assay development. Primer selection rationale is covered in Chapter 3.3 sequences and original authors are listed in Table 2.2. Primers were ordered from Eurofins (Eurofins Scientific Group, DE), with equimolar degenerate primers.

**Table 2.2: Primer sequences used**. Primers used and their original reference. Primers were sourced from Eurofins with degenerate bases in equimolar concentrations.

| Target Gene | Primer Name | Primer Sequence | Reference |
|---|---|---|---|
| *pmoA* | A189f | 5'-GGNGACTGGGACTTCTGG-3' | Costello and Lidstrom (1999) |
| | mb661r | 5'-CCGGMGCAACGTCYTTACC-3' | |
| *mmoX* | mmoX206f | 5'-ATCGCBAARGAATAYGCSCG-3' | Hutchens *et al.* (2004) |
| | mmoX886r | 5'-ACCCANGGCTCGACYTTGAA-3' | |
| *Methylocella spp.* - mmoX | [1]mmoXLR | 5'-CCCAATCATCGCTGAAGGAGT-3' | [1]Rahman *et al.* (2011a) |
| | [2]mmoXLF | 5'-GAAGATTGG GGCGGCATCTG-3' | [2]Farhan Ul Haque *et al.* (2018) |
| | [2]mmoXLF2 | 5'-TGCGCGACGCCAARAAG-3' | |
| SDIMO – $\alpha$-subunit | NVC57 | 5'-CAGTCNGAYGARKCSCGNCAYAT-3' | Coleman *et al.* (2006) |
| | NVC66 | 5'-CCANCCNGGRTAYTTRTTYTCRAACCA-3' | |
| 16S rRNA | 341f | 5'- CCTACGGGAGGCAGCAG-3' | Juck *et al.* (2000) |
| | 543r | 5'-ATTACCGCGGCTGCTGG-3' | Nossa *et al.* (2010) |

#### 2.3.7.2 Polymerase Chain Reaction Protocol

All PCR was completed using a C1000 thermocycler (Bio-Rad, USA). Reaction chemistry is detailed in Table 2.3. The Q5® (New-England BioLabs, USA) polymerase was used in all PCR assays after method development was completed. A commercial master mix was used for reliability and reproducibility. Final PCR thermocycling conditions are listed in Chapter 2.3.8.

After PCR was completed the reaction product(s) were visualised using gel electrophoresis. Agarose gels contained either 2% agarose for products sized up to 750 bp, or 1.8% agarose for products between 750 and 1300 bp. TAE buffer was used to suspend the agarose. Gels were run at 60 V for 30-60 min depending on the size of gel and the agarose concentration (longer denser gels were run for longer time periods). Gels underwent electrophoresis horizontally in a mini-sub cell tank (Bio-Rad, USA). Either a 100 bp or 1 kb ladder were used, both obtained from the Quick-Load®range (New-England BioLabs, USA). Gel images were captured using a GelDoc XR+ system (Bio-Rad USA) and its associated ImageLab software (ver. 6.1).

**Table 2.3: Key components and their final concentration in PCR.** These are based on the datasheet for the Q5 polymerase used in this project. Most reactions were completed in 15 $\mu$l volumes using the volumes provided here. The buffer, polymerase and deoxynucleotides were all included within the master mix from the supplier.

| Component | | Concentration | Volume |
|---|---|---|---|
| Reaction Buffer | $Mg^{2+}$ | 2.0 mM | 7.5 µl |
| | Polymerase | 0.02 U/µl | |
| | Deoxynucleotides | 200 µM | |
| Primers | Forward | 0.5 µM | 0.3 µl |
| | Reverse | 0.5 µM | 0.5 µl |
| Template DNA | | <1000 ng | 0.5 µl |

#### 2.3.7.3 Quantitative PCR

Bio-Rad CFX96 system was used. This was controlled remotely, and the data analysed on a PC running CFX Manager™ (V1.2). The SSoAdvanced™ Universal Inhibitor-Tolerant SYBR® Green Supermix (Bio-Rad, USA) was used. Reactions were run in 15 $\mu$l volumes with final concentrations shown in Table 2.3. Thermocycling condi-

tions for each assay are described in Chapter 2.3.8. All qPCR assays were completed with a melt curve from 65°C to 98°Cin 0.5°C increments every 5 s. Automatic baseline selection was enabled but visual checks and adjustments were made if deemed necessary. Similarly, as recommend by Bio-Rad, outliers in the standard series were removed if obviously incorrect when compared to standards of the same concentration. If the standard series required multiple standards to be removed, the assay was rerun.

#### 2.3.7.4 Statistical Analysis

All samples, standards, and controls were analysed in triplicate. A qPCR assay can show large variability even when well-designed, for example through pipetting error or random variation between wells (Bustin *et al.*, 2009). As suggested by Bustin *et al.* (2009), confidence intervals (95%) were used when presenting the standard curve's error. Standard deviation was used with technical repeats where they are compared.

For sample comparison the qPCR data was normalised to produce the relative abundance of the functional gene in a community. Normalisation used the mean abundance of a functional gene and 16S rRNA to generate the relative abundance using Equation 2.2. The mean copy number of either the functional gene or 16S rRNA is displayed with ±2 standard deviations. Normalisation propagated the error from both the 16S rRNA and functional gene means. Therefore, the error of the relative abundance must be derived using Equation 2.3 and was displayed as part of the results.

The relative abundances could be used in interpolations to visualise changes across field sites. These were completed in R using an inverse distance weighted approach (P=2). Cross-validation was used to generated RMSE between predicted and observed values. Packages used in R, are in Table 2.1.

$$R = \frac{F}{B} \tag{2.2}$$

**Relative abundance estimation**

$$\frac{\delta R}{R} = \sqrt{\left(\frac{\delta B}{B}\right)^2 + \left(\frac{\delta F}{F})^2\right)} \tag{2.3}$$

**Estimating the error of the relative abundance**

**For Equations 2.2 and 2.3:** Where: R equals relative abundance, B is mean 16S rRNA copy number and F is the mean functional gene copy

number. Here, $\delta$ signifies the error or uncertainty of the mean; for B and F, 2 standard deviations are used.

#### 2.3.7.5 Overview of qPCR Standards - Production and Range

A stock plasmid containing the inserted template was used as a standard for each qPCR assay. Standards are used to generate a standard curve for estimating copy number, to enable an estimation of the assay's efficiency and to establish the usable range of that assay. The MIQE guidelines suggest a standard range covering at least three but preferably five to six orders of magnitude (Bustin *et al.*, 2009). Standards used covered a range of $10^2$ - $10^7$ copies $\mu l^{-1}$. However, due to the increasing stochastic error at low and high concentrations the $10^2$ or $10^7$ copies $\mu l^{-1}$ standards were sometimes excluded from the analysis due to the scatter within their $C_q$ values. This was assessed after each assay was completed.

Standards were either produced through cloning in-house or purchased from Eurofins (Eurofins Scientific Group, DE). When using Eurofins, the National Center for Biotechnology Information (2021) database was used to find a suitable sequence to be inserted into the plasmid. Three genes, preferably from different genera or species, were aligned using ClustalW in MEGA (ver. 10.2.5; Kumar *et al.* (2018). The alignment was queried for the primer motif. One sequence was trimmed between these motifs and uploaded to Eurofins, sequences used are in Appendix A.1. Eurofins' standard vector with ampicillin resistance was used with EcoRI restriction sites. These were supplied at masses up to 10 $\mu$g.

Standards were also produced in-house via cloning from the positive template used in the assay design and optimisation. The process of cloning a PCR product through to a purified plasmid is described in detail in Chapter 2.3.6.

The DNA concentration of the resulting eluent was estimated using a Qubit® (Invitrogen, SG) (2.3.3. The copy number of a gene (and the plasmid) has a fixed mass and can be estimated using Equation 2.4. Using the concentration provided by the Qubit® an accurate dilution of the plasmid stock can be performed, with Equation 2.5 used to determine the mass of the plasmid stock needed. Finally, a serial dilution was performed to produce the standard series.

$$MW = m(3931 + bp_n) \tag{2.4}$$

**Molecular weight of one gene copy.** M is the mass of one base pair: 650

Da. $bp_n$ is the number of bases in the functional gene. The plasmid has 3931 bp.

$$g = \frac{C_n \times MW}{N_A} \tag{2.5}$$

**Grams of plasmid stock needed for required copy number ($C_n$).** $N_A$ is Avogadro's constant.

### 2.3.8 Final PCR and qPCR Assays

These are the final assays developed through Chapter 3 and used in the subsequent chapters. PCR protocols typically yielded better results using a touchdown protocol, especially when working with complex samples like soils. However, the touchdown protocol (detailed in Chapter 3.3.2) was removed for some samples and 35 cycles at a fixed temperature were used. For all qPCR assays a melt curve from 65°C to 98°C was used to check for melt specificity.

#### 2.3.8.1 Methane Mono–Oxygeneases

*pmoA*: Amplicon Size 510 bp

**Table 2.4:** l

***pmoA* PCR Protocol** PCR steps are abbreviated to **D**: Denaturing **A**: Annealing **E**: Extension

| **Initial D** | **D** | **A** | **E** | **D** | **A** | **E** | **Final E** |
|---|---|---|---|---|---|---|---|
| 98°C | 98°C | 68°C | 72°C | 98°C | 58°C | 72°C | 72°C |
| | | -1°C | | | | | |
| 120 s | 10 s | 30 s | 25 s | 10 s | 30 s | 25 s | 300 s |
| | | Touchdown 10 Cycles | | | 20-30 Cycles | | |

**Table 2.5: *pmoA* qPCR Protocol** PCR steps are abbreviated to **D**: Denaturing **A**: Annealing **E**: Extension

| **Initial D** | **D** | **A** | **E** | **Final E** |
|---|---|---|---|---|
| 98°C | 98°C | 64°C | 72°C | 72°C |
| 120 s | 10 s | 30 s | 10 s | 120 s |
| | | 35 Cycles | | |

*mmoX*: Amplicon Size 720 b

**Table 2.6: *mmoX* PCR Protocol** PCR steps are abbreviated to **D**: Denaturing **A**: Annealing **E**: Extension

| **Initial D** | **D** | **A** | **E** | **D** | **A** | **E** | **Final E** |
|---|---|---|---|---|---|---|---|
| 98°C | 98°C | 72°C | 72°C | 98°C | 62°C | 72°C | 72°C |
| | | -1°C | | | | | |
| 120 s | 10 s | 30 s | 45 s | 10 s | 30 s | 45 s | 300 s |
| | | Touchdown 10 Cycles | | | 20-30 Cycles | | |

**Table 2.7: *mmoX* qPCR Protocol** PCR steps are abbreviated to **D**: Denaturing **A**: Annealing **E**: Extension

| **Initial D** | **D** | **A** | **E** | **Final E** |
|---|---|---|---|---|
| 98°C | 98°C | 62°C | 72°C | 72°C |
| 120 s | 10 s | 35 s | 45 s | 120 s |
| | | 35 Cycles | | |

#### 2.3.8.2 *Methylocella mmoX*

The same PCR protocol was used for both combinations of primers. More success was found when using the XLR/XLF2. Amplicon size: XLF: 450bp or, XLF2: 390 bp.

**Table 2.8: *Methylocella mmoX*: PCR Protocol** PCR steps are abbreviated to **D**: Denaturing **A**: Annealing **E**: Extension

| **Initial D** | **D** | **A** | **E** | **D** | **A** | **E** | **Final E** |
|---|---|---|---|---|---|---|---|
| 98°C | 98°C | 68°C | 72°C | 98°C | 65°C | 72°C | 72°C |
| | | -0.5°C | | | | | |
| 120 s | 10 s | 30 s | 30 s | 10 s | 30 s | 45 s | 300 s |
| | | Touchdown 10 Cycles | | | 20-30 Cycles | | |

**Table 2.9: *Methylocella mmoX*: qPCR Protocol** PCR steps are abbreviated to **D**: Denaturing **A**: Annealing **E**: Extension

| **Initial D** | **D** | **A** | **E** | **Final E** |
|---|---|---|---|---|
| 98°C | 98°C | 65°C | 72°C | 72°C |
| 120 s | 10 s | 20 s | 10 s | 120 s |
| | | 35 Cycles | | |

#### 2.3.8.3 Soluble Di–Iron Mono–Oxygeneases

Amplicon size 420bp. No reliable standard PCR protocol was developed due to Covid time constraints. Primer set was originally designed as a nested approach, but this was unlikely to work with qPCR and successful amplification was achieved with only the inner bracket on environmental samples.

**Table 2.10: SDIMO qPCR Protocol** PCR steps are abbreviated to **D**: Denaturing **A**: Annealing **E**: Extension

| Initial D | D | A | E | Final E |
|---|---|---|---|---|
| 98°C | 98°C | 54°C | 72°C | 72°C |
| 120 s | 10 s | 30 s | 10 s | 120 s |
| | | 35 Cycles | | |

#### 2.3.8.4 16S rRNA Gene

Amplicon size 223bp.

**Table 2.11: 16S rRNA: PCR Protocol** PCR steps are abbreviated to **D**: Denaturing **A**: Annealing **E**: Extension

| Initial D | D | A | E | Final E |
|---|---|---|---|---|
| 98°C | 98°C | 63°C | 72°C | 72°C |
| 120 s | 30 s | 30 s | 30 s | 300 s |
| | | 35 Cycles | | |

**Table 2.12: 16S rRNA: qPCR Protocol** PCR steps are abbreviated to **D**: Denaturing **A**: Annealing **E**: Extension

| Initial D | D | A | E | Final E |
|---|---|---|---|---|
| 98°C | 98°C | 64°C | 72°C | 72°C |
| 120 s | 10 s | 30 s | 10 s | 120 s |
| | | 35 Cycles | | |

### 2.3.9 Nanopore Sequencing

Amplicon sequencing targeting the 16S rRNA gene was used to gain insights into community composition. The Oxford Nanopore Technologies' (ONT) MinION and Flongle platforms were used. ONTs' technology uses nanopores mounted in an electrically resistive membrane which is housed in a flow cell. The dsDNA sample has a Rapid Sequencing Adapter (RAP) bound to it. This consists of a motor protein which is captured by a tether on the membrane. The motor protein splits the dsDNA and guides a ssDNA strand through the membrane's protein nanopores causing a change in the ionic potential across the pore. The membrane is mounted on a sensor chip which provides each pore with its own electrode enabling signal detection across the pore. This signal (a squiggle) can be interpreted and recorded by inbuilt software. This data can be converted into sequences ready for further processing and analysis.

The original MinION®chassis supported both the larger MinION® and Flongle flow cells. Flongles were used more during the project. Compared to the MinION® flow cells the Flongles have only a quarter of the pores (126 versus 512) which limits the sequencing depth during multiplexing as pores eventually fail irreparably during sequencing. However, the use of Flongles still provided adequate depth for the samples under analysis whilst providing better storage options (shelf-life was a concern). Price per Gb of data generated was predicted to be equal between flow cells.

#### 2.3.9.1 Wet Chemistry

A library was prepared for sequencing using ONTs' 16S barcoding kit (SQK-RAB204). This allows for 12 samples to be run simultaneously on one flow cell due to the addition of unique barcodes for each sample. Barcodes and 5' tags are added using modified 27F and 1492R primers producing a $\sim$1500 bp PCR product. Library preparation involved the PCR amplification of the 16S rRNA followed by clean-up and quantification of the samples.

The PCR reactions were set up in a laminar flow hood to minimise contamination. LongAmp® (New-England BioLabs, USA) polymerase, as recommended by ONT, was used. Kit primers, containing the barcodes, were used. Each sample's PCR run was set up in 30 $\mu$l volumes then split into three 10 $\mu$l aliquots to reduce primer bias. The volumes of reaction components are in Table 2.14. With the PCR protocol described in Table 2.15. A negative control reaction was set up containing all the barcode primers. Following PCR, the amplified product was visualised on a 1.8% agarose gel as described in Chapter 2.3.7.

**Table 2.13: Key components in the ONT PCR.** Volumes are for a 30 $\mu$l mix that would be split into three aliquots for thermal cycling.

| Component | Final Volume |
|---|---|
| LongAmp® Master Mix | 15 $\mu$l |
| Primer (with barcodes) | 3 $\mu$l |
| Ultra-Pure Water | 11 $\mu$l |
| DNA Sample | 1 $\mu$l |

**Table 2.14: PCR Protocol for ONT 16S RNA Barcoding Kit.** Using the recommend LongAmp® polymerase in 10 $\mu$l reaction volumes.

| Initial D | D | A | E | Final E |
|---|---|---|---|---|
| 95°C | 95°C | 55°C | 65°C | 65°C |
| 60 s | 20 s | 30 s | 120 s | 300s |
| | | 35 Cycles | | |

The PCR products were cleaned using ChargeSwitch® PCR clean-up kit (Invitrogen, US). This uses magnetic beads to separate DNA from the PCR reaction components. Care was taken to avoid drying the beads or introducing air bubbles to the suspensions. Each sample's PCR reactions are pooled in a clean 1.5 ml LoBind® microcentrifuge tube (Eppendorf, DE), totalling 30 $\mu$l. To this, 30 $\mu$l of purification buffer and 10 $\mu$l of the suspended magnetic beads were added. The mixture was left at room temperature for one minute to allow DNA to bind to the surface of the beads. The magnetic beads were separated from the suspension using a magnet and the supernatant discarded.

The pelleted beads and the bound DNA, were resuspended in 150 $\mu$l wash buffer. The beads were again pelleted using a magnet, the supernatant removed and the pellet resuspended in another 150 $\mu$l of wash buffer. The beads were pelleted again using a magnet and the supernatant removed.

The DNA was eluted by resuspending the pellet with 25 $\mu$l elution buffer (10 mM Tris-HCl, pH 8.5). The suspension was left at room temperature for one minute before the beads were pelleted using a magnet for one minute. The supernatant, now with the DNA in solution, was transferred to a clean LoBind® microcentrifuge tube.

The cleaned PCR amplicon was quantified using a HS Qubit®assay, as described in Section 2.3.3. All amplicons were pooled into a 10 $\mu$ll single sample. Samples were pooled to ensure equal masses of DNA from each sample were added to reduce bias during the nanopore analysis. Ideal total DNA mass in the pooled library was between 50-100 ng.

The pooled library was prepared for sequencing. ONTs' Rapid Sequencing Adapter (RAP) was attached: 1 $\mu$ll of RAP was add to the library and mixed by flicking the tube before being centrifuged and left at room temperature for five minutes. The RAP adhered to the tagged 5' end of the PCR product. The library was now ready for loading onto the ONT flow cell.

The flow cells were prepared for sequencing after being loaded into the MinION®chassis. Once loaded into the chassis the health of the nanopore membrane was checked using ONTs' software; this ensured there were adequate pores available on the membrane to complete analysis (for a Flongle >50 pores). The flow cell was flushed and primed with a mixture consisting of 177 $\mu$ll Flush Buffer and 3 $\mu$ll Flush Tether. Care was taken to ensure no air bubbles entered the flow cell. Air bubbles would have damaged the membrane leading to a loss of pores.

The library was prepared for loading by mixing with buffer and loading beads. 5.5 $\mu$ll of DNA library were added to 11 $\mu$ll Loading Beads and 13.5 $\mu$ll Sequencing Buffer. All 30 $\mu$ll of prepared library were loaded into the flow cell. Again, care was taken to ensure that air was not introduced. Loading Beads aided the movement of dsDNA to the pore for sequencing and improved the speed of sequencing.

#### 2.3.9.2 Data Analysis

Once the sample was loaded, the run was started immediately. The flow cell and chassis were managed by the MinKNOW software hosted on a MinIT® (ONT, UK) which in turn was interfaced over a network connection with a laptop. At the beginning of a run the software was informed which kit had been used. Runs were allowed to collected data for 24 hours, after which the number of usable pores had dropped to a point where sequencing had almost ceased (for a Flongle > 15 pores).

Live basecalling was enabled. This allowed inbuilt software, GUPPY (ONT, UK), to classify the squiggles generated during sequencing into nucleotide bases and output FASTQ files whilst the run continued. The basecalled FASTQ format carried both sequence ID information and quality scores. The files were compressed and split into

Pass and Fail folders. Samples were also split at this point by barcode (and therefore by sample).

Basecalled data could be called through either ONTs' own bioinformatics software (EPI2ME) or through an in-house pipeline. Frequently, the *16S Taxonomic Classification*[1] in EPI2ME was used to quickly verify if the data appeared to be usable, i.e. how many reads were generated and what diversity was present within the dataset. If usable, the data was moved into the in-house process with the aim of producing a .BIOM file format that could be used within the R environment for interrogation and graphing. The outline of this pipeline is shown in Figure 2.2.

Upon completion of the run and basecalling the data was concatenated in the R environment (Ver 4.0.5, R Core Team (2021). GUPPY could only write a limited number of records to one file before starting a new file. Therefore, for one barcode there were multiple FASTQ files, as shown below in Table 2.15. Within R, the `ShortReads` package (Ver. 1.48.0, Franklin and Mills (2003) was used to concatenate each barcode's FASTQ files into one file within a new directory. Finally, within the new directory all barcodes were merged into a single FASTQ file for QC visualisation.

Following concatenation, the FASTQ files were processed in a `conda` environment (python V3.9) within a Linux OS. Initial visualisation of data quality was performed using `FastQC` (Ver. 0.11.9, Andrews (2015) using the file containing all sequences from all barcodes. Reads were checked again for length, Q-score and over-represented sequences. Any irregularities were explored by examining each individual barcode's data to identify the barcode with potential quality issues.

Each barcode's respective FASTQ file was trimmed and filtered with `NanoFilt` (Ver. 2.6.0, De Coster *et al.* (2018). Sequences were filtered for reads with lengths between 500 and 1750 bp. The first 50 bp were removed owing to poor quality and to ensure all barcodes and adapted material had been removed. Finally, an average minimum Q-score of 10 was applied. Nanopore data generally has lower Q-scores than data produced by Illumina where the Q-score may be >40. The filtered barcodes were then examined again in `FastQC` to ensure this process had been successful.

[1]Multiple software revisions were used: 3.2.0; 2020.04.06 and 2021.03.05.

**Table 2.15: Example file concatenation from GUPPY into a merged file using R.** Merged files could then be used in the Python environment for quality control and further processing

| GUPPY Output | R Output |
|---|---|
| >Barcode01 | |
| >Barcode01_1.fastq.gz | |
| >Barcode01_2.fastq.gz | |
| >Barcode01_3.fastq.gz | |
| >Barcode01_4.fastq.gz | >Merged_Barcodes |
| >Barcode01_5.fastq.gz | >Barcode01_Merged.fastq.gz |
| ... | >Barcode02_Merged.fastq.gz |
| >Barcode02 | >Barcode02_Merged.fastq.gz |
| >Barcode02_1.fastq.gz | >Barcode03_Merged.fastq.gz |
| >Barcode02_2.fastq.gz | ... |
| >Barcode02_3.fastq.gz | |
| >Barcode02_4.fastq.gz | |
| >Barcode02_5.fastq.gz | |
| ... | |

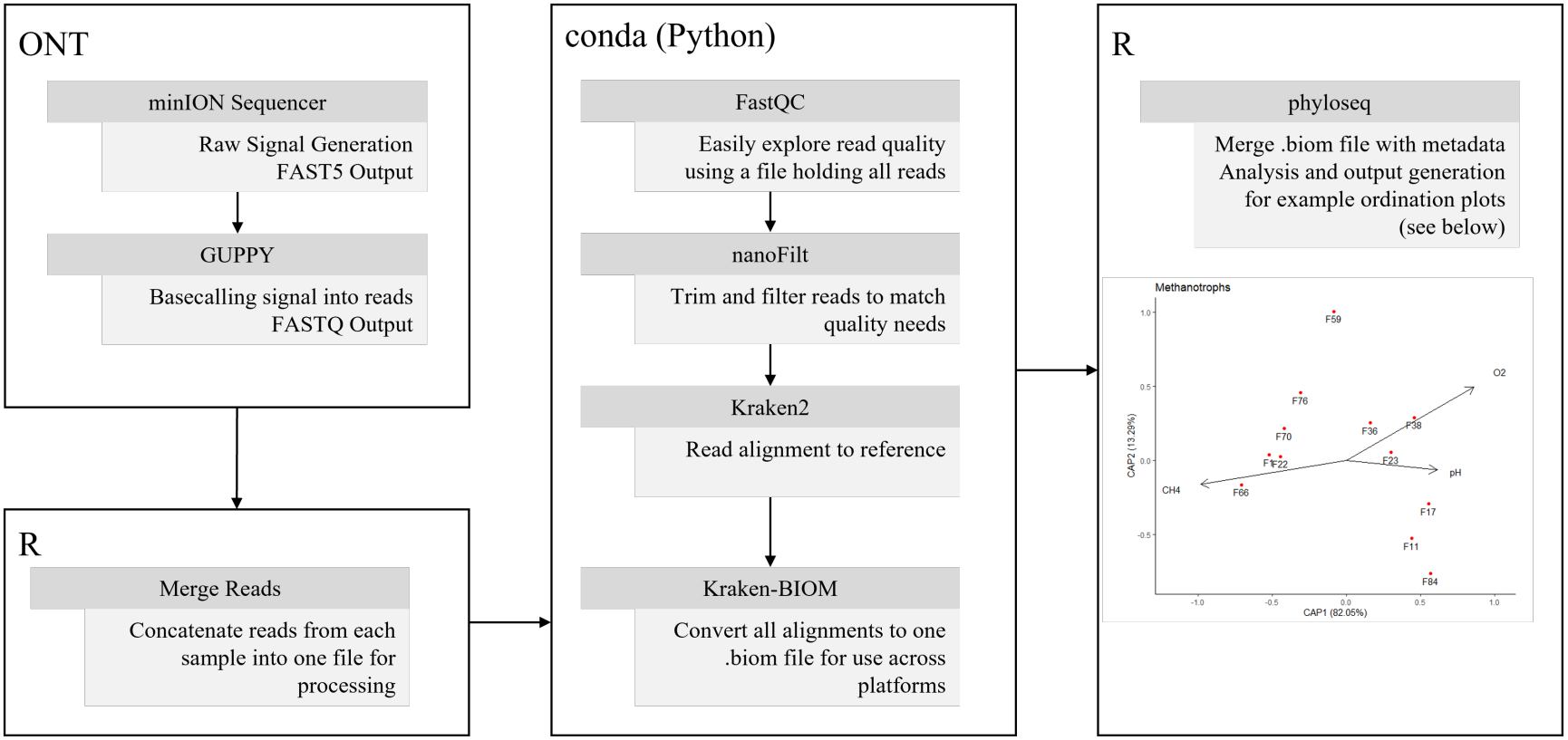



**Figure 2.2: Overview of Data Pipeline for ONT Sequencing Data.** Boxes represent different computing environments. As described in the text, data was pre-processed within the ONT platform producing multiple files of reads above a threshold quality separated by barcode. These were merged in R and moved into a `conda` environment for further QC and processing. The dark grey boxes in the `conda` environment detail the packages used. In `conda` the reads were aligned to a taxonomic database and converted to the BIOM format. The BIOM format output could be statistically analysed using packages within R. Notably, the `phyloseq` package and its integrations were used to analyse the data and draw inferences.

Filtered and trimmed sequences were assigned taxonomic labels using the `Kraken2` package (V2.1.2, Wood *et al.* (2019). This is a high-speed classifier which exactly aligns *k*-mers ($k$ = 31) rather than attempting inexact alignments of whole sequences to the reference dataset (Wood *et al.*, 2019). Every *k*-mers in the query sequence is mapped to the lowest common ancestor (LCA) that also contains the *k*-mers r in the reference dataset. This enables a tree of LCAs to be constructed. Each node is weighted by the number of *k*-mers associated with that taxon. The query sequence is assigned a taxonomy based on which branch of the tree has the highest weight. Query sequences without an exact *k*-mers alignment in the database are not classified. The reference database was the prebuilt MiniKraken (V1-2019) using the Refseq library (National Center for Biotechnology Information, 2021). The 'mini' version was selected due to its smaller RAM requirements. While V1 was used as it excluded the human genome and focussed upon bacteria, archaea and viral libraries.

After alignment, the output files for each barcode were reformatted and merged. The tab separated output format from `Kraken` was converted to the BIOM format (hdf5) using the `Kraken-Biom` package (V1.0.1). The Biological Observation Matrix (BIOM) format is a standardised file format allowing the movement of taxonomic data between different 'omics platforms (McDonald *et al.*, 2012). During conversion, all individual barcodes were merged into one data structure whilst retaining sample ID.

#### 2.3.9.3 Ecological Analysis

The single .BIOM format file, with all reads, was imported into the R environment. The `phyloseq` package (V1.34.0. McMurdie and Holmes (2013) can read the .BIOM format and uses this to build a data object in R for statistical analysis. During construction of the new R object, the .BIOM file was merged with a metadata (or mapping) file containing sample information, for example moisture and pH measurements. Once constructed, the data object could be manipulated to explore community composition and to plot graphical summaries. Data manipulation and visualisation was completed with tools from `phyloseq`, `vegan` (V2.5-7, Oksanen *et al.* (2020) and ggplot2 (V3.3.3, Wickham (2016). This thesis did not normalise reads for ecological analysis unless clearly stated in the chapter methods or results. Instead, reads were handled as absolute values, in future work the reads should normalised and analysed as relative.

Statistical methods for the community analysis were selected based upon reviews by Paliy and Shankar (2016) and Ramette (2007). Exploratory analysis highlighting changes in community composition used taxa with relative abundances greater than $1\times10^{-2}$ in stacked bar-charts. Exploration at the phylum and family taxonomic lev-

els was undertaken with the taxa grouped by unique taxa at that level. The use of a novel pipeline, the inherent error and sample size of reads collected using a Flongle meant using higher resolution taxonomic data was probably not reliable. Patterns at the family level would allow the hypothesis to be explored by investigating broad trends in community composition. Alongside analysis with all families a subset of the dataset containing the families Methylocystaceae, Methylococcaceae and Beijerinckiaceae was created to explore differences in methanotrophic guild.

Further exploratory analysis using an unconstrained Non-Metric Multi-Dimensional (NMDS) ordination was undertaken. This was selected due the ease of interpretation and the ability to quickly model, with significance values, the impact of abiotic variables onto the ordination. Furthermore, NMDS was recommended to be used with fragmented habitats like those seen at these field sites (Ramette, 2007)). To complete the analysis, Bray-Curtis dissimilarity distances were estimated using the distance function. Generated distances were employed in the ordinations using the ordinate function. The fit of the NMDS ordination was measured and plotted on a Shepard diagram to assess if the model represents the data and distances well. Environmental metadata was modelled onto the ordination using the `envfit` function within the vegan package. The `envfit` function returns vectors for the variables under test relative to the ordination axes. The function was requested to complete 999 random permutations to enable both an $r^2$ and $p$ value to be generated for the vectors.

To test the hypotheses, different approaches were used depending upon the data and the questions posed. Unconstrained ordination was employed to explore the separation of samples based upon abiotic data from the sample sites. Analysis of Similarity (ANOSIM) tests were used to test if groups were significantly dissimilar, these used Bray-Curtis distances metrics and were ran with 9999 permutations. The methods and rationale for these methods are in the chapters where they are used.

Constrained and unconstrained ordinations were used to understand the data as completely as possible. While a constrained ordination is ideal to test the hypotheses that microbial community composition is altered by the presence of a measured variable, it only allows variation related to the measured variables to be represented by the main axes. Including an unconstrained ordination enables the maximum variability within the biological dataset to be plotted. This ensures that if the community is poorly explained by the measured variables, it will be clear when comparing the two ordination approaches.

# Chapter 3

# Preliminary Work and Development

This chapter outlines the preliminary work which enabled the final methods to be developed. Sample size, experimental scale and the effect of the soil's constituent components upon methods had to be explored and defined. Furthermore, the impact of inhibitor carryover from soils on PCR and qPCR needed understanding to adapt protocols to the challenges posed. Inhibitors are contaminants that interact with DNA and PCR chemistry limiting the accuracy and sensitivity of the assay.

## 3.1 PCR Design and Reagent Selection

PCR is a cycle of reactions (Figure 3.1), driven by repeatedly alternating the temperature of the reaction. This aims to amplify a short section of DNA, the template, to a concentration where it can be detected. At a higher temperature, the template is denatured (Fig. 3.1: Denaturation) into two single strands of DNA. The temperature drops enabling the binding, or annealing (Fig. 3.1: Annealing), of two oligonucleotide primers with complementarity to sequences at the start of the template and at the end. At the beginning of the reaction re-hybridisation of the two parent strands during the annealing step is limited due to the excess concentrations of the primers in the reaction. The temperature ramps up to an intermediate point allowing the extension (Fig. 3.1: Extension) of the primed template using a DNA polymerase and free nucleotides in the reaction mixture. The reaction returns to the higher temperature and repeats, over 20-40 cycles during which the template is amplified until the reagents are exhausted, or, the generation of new product ceases as double-stranded DNA is forming instead of primed templates.

The reaction needs to be carefully designed to ensure accurate and consistent amplification of the template. The design of the primers is critical to ensure that the target is selected and amplified. Poor quality primer design leads to non-specific products being formed or, primer self-amplification. Primer stringency, i.e. the strength of the bond between primer and template, is strongly impacted by the temperature of the annealing step. The choice of polymerase also needs consideration; for example, would the use of a reagent with a proof-reading ability be useful in the assay? Proof-reading, or high-fidelity, enzymes include an exonuclease subunit; if an incorrect nucleotide is bound by the polymerase it is less likely to bind to the template strand and the reaction slows. In this slower reaction state, the loose end is more likely to interact with the exonuclease subunit leading to the faulty base being removed. Overall, the use of a proof-reading enzyme should reduce errors in DNA replication during PCR.

Another factor to be considered carefully is the template quality, often linked to the source and extraction method. In this thesis, most samples were extracted from soils. Soil extractions pose a series of interesting problems to PCR and other downstream techniques. Contamination by inhibitors, such as humic acids, can frequently be carried over and these can adversely affect the PCR reaction. As a result, high quality extraction procedures must be used and the consideration of additional reagents to negate the impact of these contaminants must be considered.

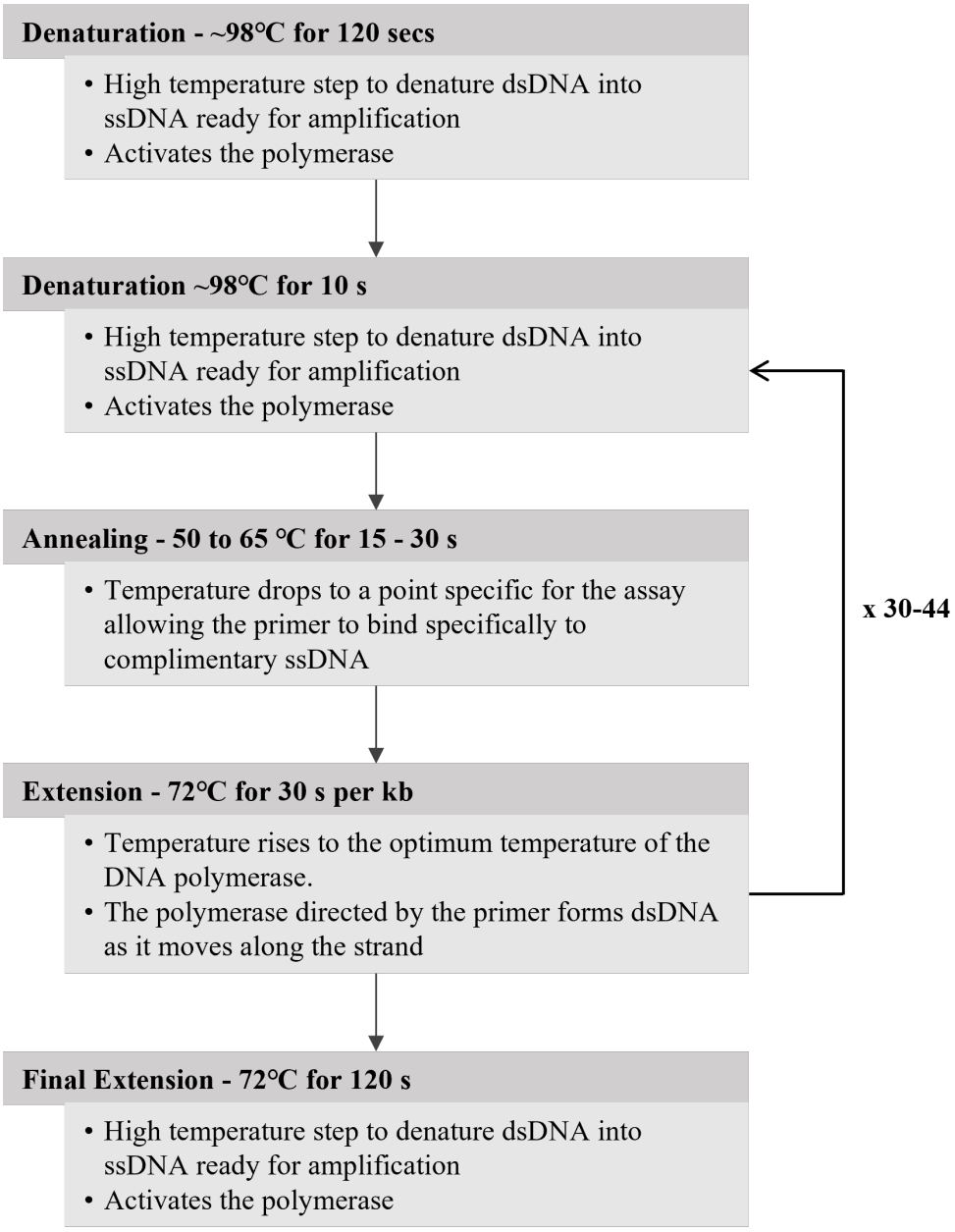



**Figure 3.1: Overview of the PCR reaction**. PCR is a cyclic series of reactions at alternating temperatures that enable the amplification of DNA. Each assay has specific temperatures, time, and number of cycles to maximise efficiency and success.

### 3.1.1 PCR Reagents and Equipment

For the reaction chemistry, the use of a supplier prepared master-mix was thought to likely result in optimal results, based on the potential time saving and the increased consistency of having the reagents prepared by a commercial company. The same enzyme (and master-mix) was used throughout the project to ensure reproducibility. Given both the source material for the template and the use of PCR products in cloning and sequencing a high-fidelity enzyme was selected.

Initial worked tested the iProof™ (Bio-Rad, USA) master-mix. Problems with non-specific amplification, no amplification and contamination persisted despite numerous alterations to thermo-cycling conditions and the use of different templates and primers. These problems were not limited to one template or one target gene. After five months of testing, it was decided this polymerase was not suitable for the project.

A second, high fidelity enzyme was selected: Q5® (New-England BioLabs, USA). This is a thermostable polymerase, with a 3'→5' exonuclease activity fused to Sso7d, a non-specific dsDNA binding protein first cloned from *Sulfolobus solfataricus* (Wang *et al.*, 2004). This protein improved the processivity of the polymerase, Sso7d stabilises the bond between polymerase and template, thus increasing the speed and robustness of the extension stage of the PCR reaction. Consistent amplification was achieved using a touchdown approach to optimise annealing temperatures (touchdown is described in more detail in Chapter 3.3.2). The enzyme worked well with templates from varying sources with different qualities.

## 3.2 Quantitative - Polymerase Chain Reaction

Quantitative-PCR (qPCR) proceeds identically to 'normal' PCR but measures the increase in PCR product every cycle. This enables an estimation of the starting concentration within the sample. Here, qPCR used an intercalating dye to measure the increase in each reaction's fluorescence at the end of every cycle's extension step. Using a standard curve, plotted using reactions with a known starting concentration of target, an estimate of the sample's starting concentration can be inferred using the cycle where the fluorescence of the PCR reaction exceeds the background signal (the $C_q$). Assays were designed and run using the *'Minimum Information for Publication of Quantitative Real-Time PCR Experiments'* (MIQE) (Taylor *et al.*, 2000) guidelines, which are considered best practice.

### 3.2.1 Reagents and Equipment

As with generic PCR, a commercially prepared master-mix was deemed most suitable. Given reliability of the Q5 enzyme a qPCR master-mix using a similar enzyme was sought. The SSoAdvanced enzyme was selected as it also used the Sso7d, a non-specific dsDNA binding protein that improved enzyme processivity. Furthermore, the master-mix selected contained a selection of components within the buffer that claimed to improve reaction efficiency when used with samples from a difficult matrix material (for example soils).

### 3.2.2 The Measure of qPCR Assays

A successful PCR reaction is typically designed to generate a clean single band of DNA, of the correct size when separated using gel electrophoresis. When analysing qPCR, the operator can use a variety of metrics generated during amplification and during product melt.

Throughout amplification, fluorescence within the reaction well will increase as ds-DNA concentration increases. If the assay is working and there is adequate template, the fluorescence will increase to a detectable level above the background. This point is recorded as the threshold cycle ($C_t$), or the quantification cycle ($C_q$). This metric enables an estimation of the starting concentration of the samples through a comparison with the standards. All qPCR runs should include at least four standards from which a standard curve can be derived (Taylor *et al.*, 2000). Standards are normally a serial dilution series of a plasmid stock (see Chapter 2.3.7.5) and their concentration expressed in $\log_{10}$; they should cover at least three orders of magnitude (ideally five or six) (Taylor *et al.*, 2000). This results in a log-linear trendline within the usable range of the assay. The slope of the trendline can be converted to an efficiency value, using Equation 3.1. An optimum assay will have a trendline with a coefficient of determination ($R^2$ value) of 0.99 and an efficiency of around 100%. Assays were considered usable if they had an $R^2 \geq 0.95$ and an efficiency (E value) of 90-110%.

$$E = 10^{\frac{1}{slope}} - 1 \tag{3.1}$$

Equation: qPCR efficiency calculation. Where E=Efficiency and the Slope is derived from the log-linear trendline plotted from the standards.

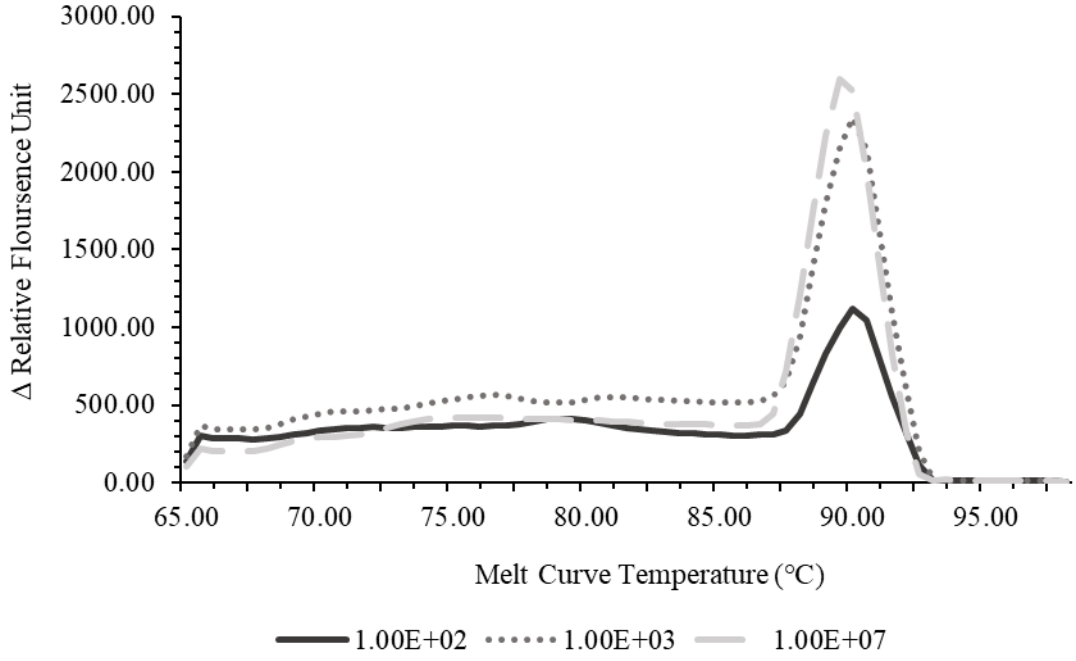



**Figure 3.2: Example derived melt curve, using selected standards.** As shown, a single amplified product will produce a strong single peak at a set temperature. A reaction with a low yield will produce a smaller peak; here both the $10^3$ and $10^7$ standards have amplified to exhaustion so have a similar sized melt peak. The $10^2$ standard amplified later and never plateaued resulting in a smaller peak. Assay: SSoA, XLF2/XLR *(mmoX)*.

Following amplification, the product is 'melted.' This is a quality control step in which the sample is heated incrementally by 0.5°C from 65°C to 98°C and the fluorescence measured after five seconds. As the temperature increases, fluorescence drops as the dsDNA denatures to ssDNA and the intercalated dye is released. The rate of fluorescence loss will rapidly increase as the melt temperature of the product is reached. The melt temperature is determined by the GC-content and length of the product. As a result, the melt stage can inform the specificity of the assay. To interpret this data the temperature is plotted against the change in florescence during the previous temperature increase, as shown in Figure 3.2 A successful assay will have a strong single peak representing a single specific product.

The use of melt peaks with complex soil communities was difficult. Often a single peak could be obtained but, as illustrated in Figure 3.3, the width of the peak could be relatively broad and the peak shifted up to 2°C compared to the standards. This resulted in the qPCR product being explored with gel electrophoresis like a standard PCR assay. A selection of wells was used to ensure a single product of the right size had been obtained in samples under investigation. Where there were concerns with

specificity, the product was sequenced.

The change in peak shape is not surprising. The complexity of soil communities makes it possible for several variants of the same gene to be present in a single sample. Small variations in length and GC content could feasibly impact the melt temperature. In a soil community this diversity in melt temperatures would 'smear' the peak despite the product forming a single band of the correct length when run on a gel.

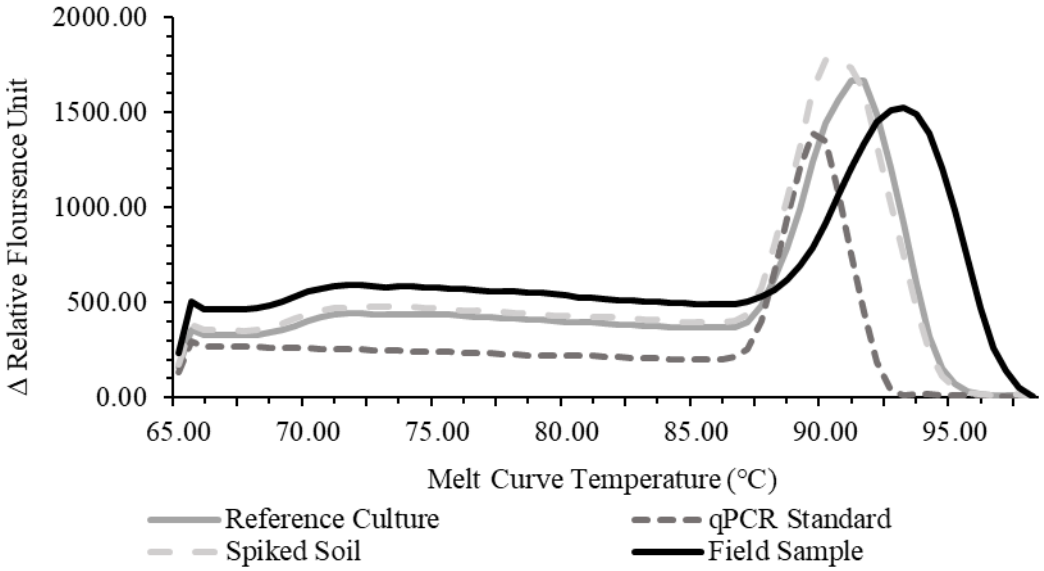



**Figure 3.3: Example melt curves produced in one qPCR assay.** With the exception of the field sample all the melts have subtly different peaks despite all the samples being amplified from the same reference strain. The field sample is broader and has a higher peak melt temperature (93°C to the other three samples. Assay: SSoA, XLF2/XLR *(mmoX)*.

### 3.2.3 Normalisation

Using a standard series enables the absolute quantification of gene copy-number in the sample. While this provides an accurate estimate of the copy number, direct comparison of two samples is not possible. Soil is highly heterogeneous, so every sample will behave differently when being assayed. Both the extent of inhibition and the efficiency of the DNA extraction will vary between samples. As a result, the data produced must be normalised to make it comparable. The approach used in this normalisation is described in Chapter 2.3.7.4.

The assays presented here target an ecological niche within soils and attempt to quantify it. Therefore, it is logical to attempt to express that community as a fraction of the whole bacterial population. To do this the 16S rRNA gene was selected to assess the

size of the bacterial community and to act as the gene with which to normalise. As a result, copy number is normally expressed as a ratio of the functional gene to 16S rRNA. This overcomes the problems of inhibition and extraction quality/bias as both the 16S rRNA assay and functional gene assay should have been equally affected. So, if a sample suffered from high inhibition and a low extraction quality then both the 16S rRNA and functional gene data will be negatively impacted; in theory these should be impacted to a similar extent. This ratio enables the direct comparison between samples or for them to be used alongside one another in other plots (e.g. interpolation).

Expression of a functional gene as an abundance has been used by previous authors. Farhan Ul Haque *et al.* (2018) used the assumption that that every bacterial cell has two copies of the 16S rRNA gene. The authors also used published whole genomes of *Methylocella silvestris (mmoX)* and *Methylosinus trichosporium (pmoA)* to estimate the number of copies of the functional genes in individual cells, with 2 copies of *pmoA* per cell and one copy of *mmoX* per cell. In the method used here, no adjustment for the number of gene copies per cell is made. Due to the unknown variability in copy number across members of a community it was not felt that applying these assumptions would aid or improve the accuracy of the data nor would it change the relationship observed between samples.

## 3.3 Desiging PCR Assays

As discussed previously, setting up a robust polymerase chain reaction (PCR) is a multi-faceted process. Once primers and polymerases have been selected, the conditions within the thermocycler during the assay must be optimised to ensure the assay is robust and replicable. Working with soils compounds the complexity of PCR reaction design by reducing template quantity, quality and introducing a variety of reaction inhibitors.

The core technique of this project employed PCR reactions; therefore, the design of these assays was important and requires explanation. All assay design and optimisation efforts followed a similar pathway, detailed in Figure 3.4. First, a positive template was sought, then a PCR reaction designed to amplify that template with the selected reagent and primers. This was used to prepare standards for qPCR; the qPCR assay was then optimised using the PCR assay as a starting point and with the MIQE guidelines as an indicator of best practice. The final assays used are described in the Chapter 2.3.8.

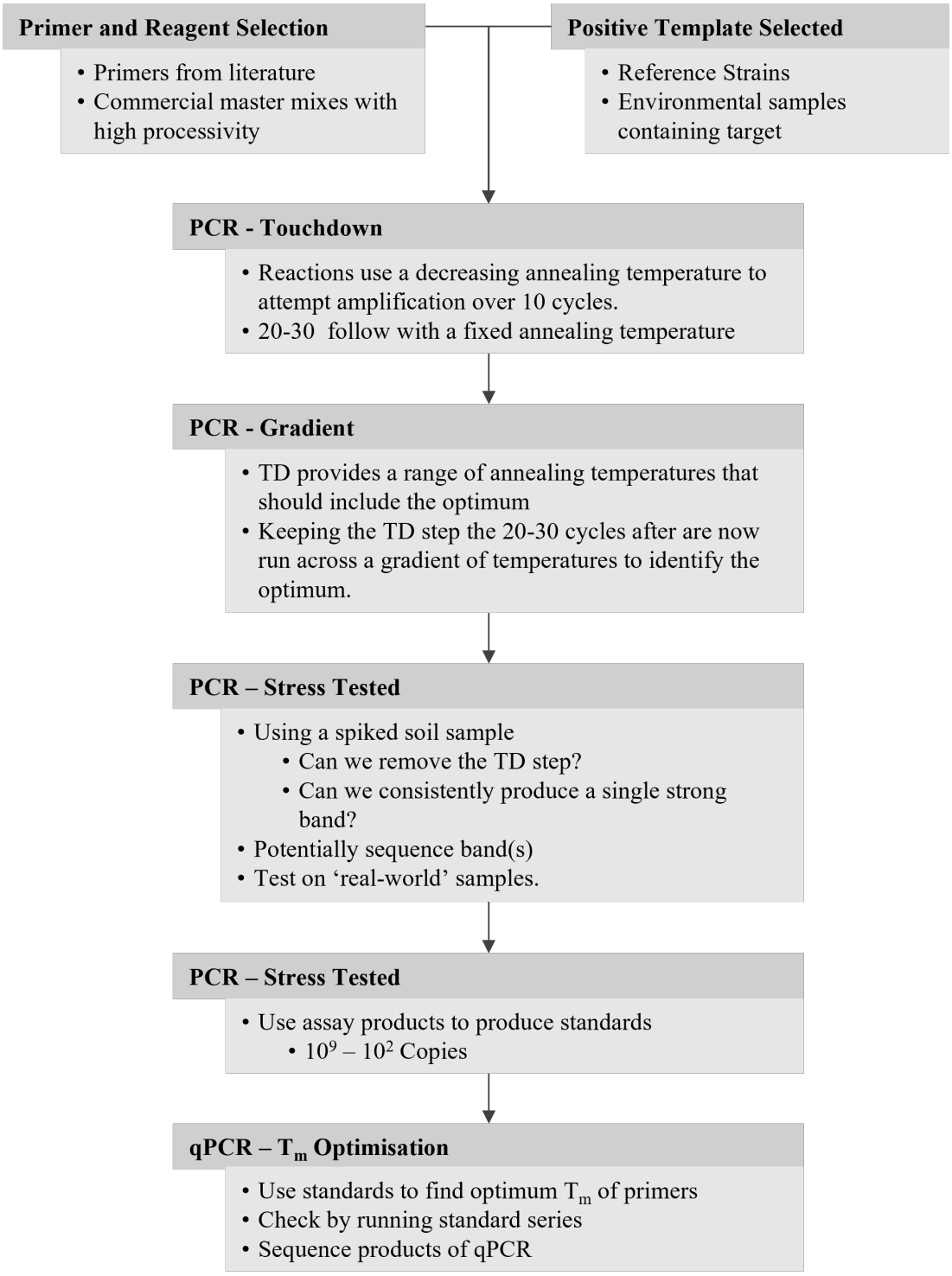



**Figure 3.4: Flowchart outlining steps in PCR assay design.** The selection of primers and regents is disc ussed in the main text. Positive template samples were from reference strains, from sites with a known methane seepage, or, from soil incubated in the presence of methane. These were then used to optimise the PCR, focussing upon finding the optimum annealing temperature to produce standards. The standards could then be used to optimise qPCR assays.

### 3.3.1 Primer Selection

Primers were selected from the literature and sequences for the primers used can be found in Chapter 2.3.7.1. The choice to use pre-published primer sets, rather than design new sets, was to save time and to use proven tools within new techniques. Selecting primer sets that had already been used in numerous published studies in a variety of scenarios meant that the work being completed during this project could focus on the surveying method, not the primers. Therefore, confidence in the results was stronger as the strengths and weaknesses of the primer sets being used had already been well described. Furthermore, the results collected could be interpreted with those weaknesses considered. Designing new primer sets, while potentially possible, would have been difficult given the complexity of soil communities. Furthermore, the use of new primer sets in new survey methods would have required extensive testing to explore their weaknesses.

With qPCR the length of the amplicon is also an important consideration. As discussed in the previous section, efficiency is a key metric for assessing the success and reliability of an assay. Longer target sequences have inherently lower probability of complete amplification, which has the effect of lowering the efficiency of the assay. Furthermore, longer templates are more likely to become fragmented during degradation (through storage, freeze-thaw or extraction).

#### 3.3.1.1 Methane Mono-Oxygenase Primers

As discussed in previous chapters, there are two distinct analogues of the enzyme, viz. particulate and soluble. These are targeted using different primer sets. Primer sequences are in Table 2.2.

The particulate methane mono-oxygenase enzyme is probably an evolutionary relative of ammonia mono-oxygenase (Holmes *et al.*, 1995). The similarities between the genes for the two enzymes has meant that primers can inadvertently amplify both. Bourne *et al.* (2001) demonstrated with A189f/A682r primer set (from: Holmes *et al.*, 1995) could do this. The remaining primer sets are more specific, which appears to limit their ability to amplify all known *pmoA* containing taxa.

The primer set initially developed by Costello and Lidstrom (1999) was used and is referred to as A189f /mb661. This amplifies a 510 bp product. The primer set was originally used to explore methanotrophs in lake sediments and was designed using the small number of sequences available from selected phyla, which may mean it will

not amplify more recently described clusters such as the upland cluster(s), NC10 or *Verrucomicrobia.*

Two primer sets were selected to amplify *mmoX* for the soluble methane mono-oxygenase. Auman and Lidstrom (2002) published a primer set, A166f / B1401r, designed using the sequences available on GenBank, in their work exploring fresh water sediments. Hutchens *et al.* (2004) also used available sequences to construct a new primer set, mmoX206f / mmoX886r, this time for work with organisms from methane rich cave environments. Both amplified a reference strain used to develop the assay, *Methylocystis heyeri*[1]. The primer set from Auman and Lidstrom (2002) had problems with unspecific amplification in soil samples; furthermore, it amplified a 1230 bp product. The mmoX206f / mmoX886r (referred to from here as: 206f / 886r) primer set amplified product from several samples and the shorter 720 bp product was likely to be more successfully amplified in qPCR.

A third primer set amplifying *mmoX* was also required specifically for *Methylocella spp.*. The unique evolutionary history of the gene means the primer sets designed by Hutchens *et al.* (2004); and Auman and Lidstrom (2002) cannot amplify it. Farhan Ul Haque *et al.* (2018) designed a qPCR primer set, mmoXLF2 and mmoXLR (XLF2 / XLR), which was used throughout the project.

### 3.3.2 Intial PCR Design

Positive templates were sought to speed the optimisation of PCR assay design. Reference strains, detailed above, were used for some assays. For the *pmoA* assay, soils were incubated for several weeks in a methane enriched headspace (detailed in Chapter 3.4) to increase the size of the methanotroph population All assays needed optimisation to overcome problems posed by soils through inhibition and to improve specificity. The annealing step is the most difficult to tune and has the greatest effect upon the assay's specificity after primer design.

Primers are designed to bind to complimentary ssDNA sequences during the annealing step (Figure 3.1). The optimum binding temperature depends upon the sequence of base-pairs and can be predicted *in silico*, although differing templates and reagents can affect this temperature. Finding the optimum binding temperature is the most protracted stage of assay design but is vital to ensure assay specificity and robustness.

[1]Sourced from Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ). February 2019. Catalogue number: DSM-16984.

The predicted melting temperature ($T_m$) served as a starting point for assay development. The $T_m$ is the temperature at which half the primers are annealed to the target amplicon. Touchdown PCR (TD-PCR) protocols with the predicted $T_m$ were used to identify the optimum annealing temperature for the project's reagents, equipment and starting material (soil). TD-PCR consists of two stages with distinct cycling conditions. In the first stage the annealing temperature was lowered every cycle by a fixed amount, normally 0.5 - 1.0°C for ten cycles. This was followed by a second cycling stage with a fixed annealing temperature for 20 - 30 cycles. Primer (and assay) specificity increases with annealing temperature because, as the number of matched bases between primer and template increases, so does their melt temperature. So, a perfectly compliment primer will remain bound to the DNA strand at a higher temperature than a poorly matched primer.

If the annealing temperature was too high even the most complementary binding was not possible. During TD-PCR the assay's annealing temperature starts too high for primer-template binding. As the annealing temperature drops it will pass through the optimum annealing temperature causing the target template to be amplified. The annealing temperature will drop further, and the assay becomes less specific with the primers potentially binding to sequences with a lower complementarity. However, the amplification of the target during the cycles around the optimum $T_m$ will have biased the reaction, increasing the probability of the target template being amplified in greater quantities than any poorly matched sequences.

The initial stages of PCR optimisation use TD-PCR to check if the primers can amplify the target. Using the predicted $T_m$ the range of annealing temperatures is selected for the first stage of cycling. The reagents selected for the project were known to increase the theoretical $T_m$ by up to 3°C, therefore a range of 10°C was used with the annealing temperature starting 7°C higher than the theoretical $T_m$. Through the second stage of cycling, the annealing temperature was set at the theoretical $T_m$. If the target could be amplified there would be two or three cycles of strong amplification at some point during the touchdown stage of cycling. This biases the reaction through the second stage of cycling producing a product that can be detected using gel-electrophoresis.

The first TD-PCR tests illustrated that the primers could amplify a product of the correct size from the positive template. The optimum annealing temperature was still unknown, but a possible range of temperatures had been identified. To identify the optimum annealing temperature a TD-PCR protocol was used but, during the second stage of cycling, the temperature gradient within the thermocycler held each reaction

at a different, fixed annealing temperature. The thermocycler allowed a gradient to be applied across the block, enabling the second half to be run with reactions at different temperatures to one another. These individual reactions, with differing annealing temperatures, were compared side by side after gel-electrophoresis to estimate which annealing temperature generated the most product. The reaction with the closest annealing temperature to the optimum should contain the brightest and cleanest band.

With a known annealing temperature, the PCR assay was stress-tested. This involved using a selection of starting samples and attempting to remove the TD stage of the protocol. If a reference strain had been used to this point, soils were spiked with the strain, left for 24-48 hrs and then the DNA extracted. The extract was used in the assay to test its specificity; often the presence of complex soil communities and inhibitors lead to the annealing temperature being raised by two or three degrees centigrade. If possible, soils with a greater fraction of clay were used as positive templates to further test the assay's robustness. A final test used soils from the field that should contain the template but potentially at low concentrations. Once the PCR assay was optimised it was used to produce standards for use in qPCR.

With a set of standards, optimisation of qPCR assays began, using the PCR assay as a starting point. Given the similarity between the polymerases, equipment and template, the optimum annealing temperature was not expected to change greatly. Therefore, running the qPCR on a narrow gradient of annealing temperatures, using a template extracted from soil, would ensure the annealing temperature was still valid. This was checked by ensuring that $C_q$ values between replicates was close, that the melt-curve only yielded one peak and finally visually checking the product using gel-electrophoresis. With the annealing temperature checked, a standard series, produced using the methods in Chapter 2.3.7.5, was run to check the efficiency. Problems with the efficiency could be overcome by re-checking the annealing temperature or by altering the duration of either the annealing and extension steps. Care was taken adjusting the duration of annealment and extension as this could reduce the sensitivity of the reaction. Finally, before those standards were used in assays with samples from field-sites or lab experiments, they were submitted for sequencing (Chapter 2.3.6) to ensure the standards were the correct target.

## 3.4 Microcosm Designs

A robust and reliable experimental system was needed to test the rate of methane oxidation by communities in test soils. To design these systems and to test/develop methods of measuring these communities a pilot study was set up. This was designed to highlight any shortcomings in the microcosm design, to identify optimum testing frequency and to test sampling methods both of the gas headspace and of the soil community.

Early molecular work struggled to amplify the MMO encoding genes (*pmoA* and *mmoX*). This was potentially due to a problem with the assays or because target genes were below the limit of detection (LoD). Probably, it was a combination of both. Microcosms provided soil samples with complex communities which had elevated densities of methanotrophs which could be used in method development. Finally, this work also explored how quickly the methanotroph community might change, demonstrating that DNA could be used to detect short term (days) changes in community function.

### 3.4.1 Methods

Airtight microcosms were constructed to assess how the rate of methanotrophy and changes in microbiology might be observed in the laboratory. Microcosm conditions were adapted from Dedysh and Dunfield (2011) who discuss optimum methanotroph growth (in liquid media) with suggestions for headspace composition and the ratio of head space to broth volumes. Three microcosms were sealed with an atmosphere comprising 15% methane and 85% normal laboratory air. One of the microcosms developed a leak and so could not be used. Three additional microcosms were controls, sealed with unaltered laboratory air. These were left on the laboratory bench for 17 days; periodically a sample of the headspace was analysed to assess methane concentration, oxygen consumption and carbon dioxide production. Once aerobic respiration had reduced to almost nil, the microcosms were disassembled and DNA samples from the soil extracted for analysis.

#### 3.4.1.1 Soil Sample Homogenisation

A sandy loam soil was collected from the surface of a pasture at the Sutton Bonington Campus of Nottingham University, UK, (SK50663,26484). The site had no known recent exposure to high concentrations of methane. The soil was air dried for three days at 20°C on a laboratory bench, then ground, sieved at 1 mm and mixed thoroughly. This was left for a further seven days in a glass beaker on the bench before being mixed again. Samples of this homogenised material were used in DNA extractions as a pre-

incubation control.

#### 3.4.1.2 Microcosm Construction

Each microcosm was a closed system, constructed within a clear volume 120 ml Wheaton® crimp top serum bottle. This was sealed with a self-healing butyl bung and aluminium crimp. Each glass bottle and seal were autoclaved prior to being used to construct the microcosm. Into each microcosm 34 ml (40 g) of the homogenised soil were added, along with 6 ml of water to give a final moisture content of 15%. The bottles were immediately sealed. To construct the methane enriched headspace, 13 ml of air were drawn from the sealed microcosms and immediately replaced with pure methane, to give a headspace concentration of 15% methane and 85% laboratory atmosphere. To ensure this had been achieved, an aliquot of the headspace gas was analysed with gas chromatography (Chapter 2.2.3) within an hour of the bottle being sealed and the methane being injected. The microcosms were left on a bench out of direct sunlight for the duration of the incubation within a temperature-controlled laboratory at 20°C.

#### 3.4.1.3 Headspace Analysis

Headspace samples were collected as described in Chapter 2.2.3. Results were pooled for each time point, with the average gas concentration and peak area used in comparisons. To estimate the point at which aerobic respiration ceased, a ratio between the $O_2$+$N_2$ peak and the $CO_2$ peak was used. It was assumed that the area of the two peaks would be related, with the area of the $CO_2$ peak increasing as the $O_2$+$N_2$ peak area decreased. Plotting the change in ratio against time enables the estimation of respiration period.

#### 3.4.1.4 Molecular Analysis

After 17 days of incubation, the microcosms were disassembled. DNA extractions used 0.5 g soil. One sample from each microcosm and three samples from the soil pre-incubation were used. The extracts were aliquoted and diluted 1:10 with ultra-pure water before being stored frozen at -20°C until use. Aliquots were used in PCR and qPCR assays for 16S rRNA and *pmoA*. Primers and the final assays used are presented in Chapter 2.3.

All data were analysed using MS Excel. Relative abundances of *pmoA* and *mmoX* were estimated using the 16S rRNA copy number as an indicator of total bacterial community size. All three replicates were pooled as an average. The propagated error of the relative abundance was estimated using two standard deviations ( Chapter 2.3.7.4).

Means were compared using an independent t-test, with a significant difference accepted when $p \leq 0.01$.

### 3.4.2 Results

#### 3.4.2.1 Methane Consumption

In microcosms with a methane headspace the methane concentration fell throughout the incubation. The greatest loss of methane was seen in the first 72 hrs (-38090 ppm) (Figure 3.5). Methane loss plateaued after 72 hrs and remained constant until between 240 and 288 hrs when another small drop in methane concentration occurred. No methane was detected in the control microcosms at any point.

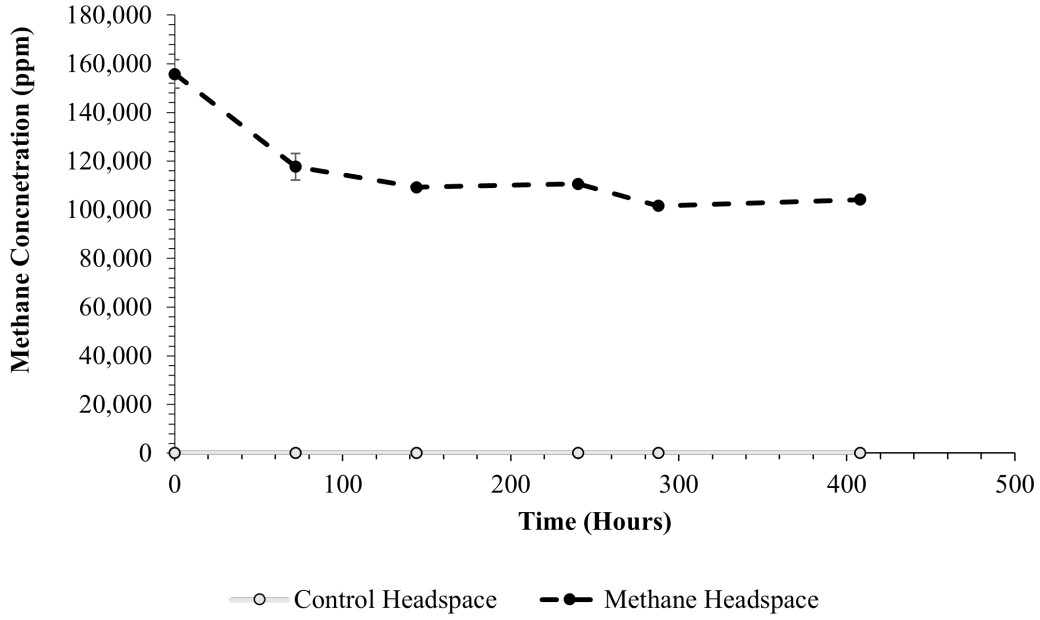



**Figure 3.5: Loss of methane from the microcosm headspace over the course of the incubations.** No methane was observed in the control headspace. Error bars are ±1SD.

#### 3.4.2.2 Respiration

After 72 hrs the aerobic respiration slowed under both microcosms (Figure 3.6). The ratio between the $O_2$ + $N_2$ peak and the $CO_2$ peak continued to fall throughout the incubation but after 150 hrs the ratio remained constant.

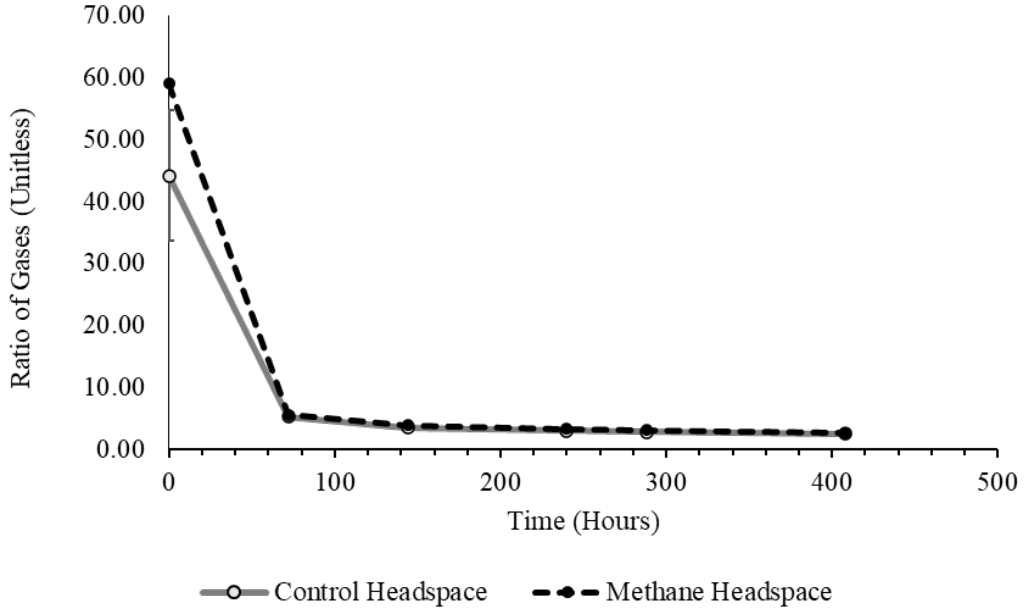



**Figure 3.6: Ratio of $O_2$ + $N_2$ peak area to $CO_2$ peak area over time.** The greatest shift in the ratio occurred in the first 72 hrs. After 150 hrs the change plateaued with only a small increase in $CO_2$ and decrease in Ratio of $O_2$ + $N_2$ peak areas. Error bars are ±1SD.

#### 3.4.2.3 Changes in *pmoA*

To assess if methane exposure had led to detectable populations of methanotrophs, PCR was used on soil DNA extracts. In the controls and microcosms with a non-enriched headspace, PCR yielded no detectable product, suggesting that *pmoA* copy number in these samples was below the LoD of the PCR assay. Conversely, the two microcosms with a methane enriched headspace did yield a PCR product of between 500 and 600 bp using the *pmoA* primer set, indicating the presence of a methanotrophic population. This enabled the optimisation of the qPCR protocol for *pmoA* and an estimation of the functional genes' relative abundance. The optimised qPCR protocol amplified *pmoA* from all three soil conditions (Figure 3.7). The soil under a methane headspace had a significantly greater relative abundance of *pmoA* than the control headspace ($p$<0.01). While the mean *pmoA* abundance was greater under the control headspace compared to the pre-incubation soil there was no significant difference between the two ($p$=0.08).

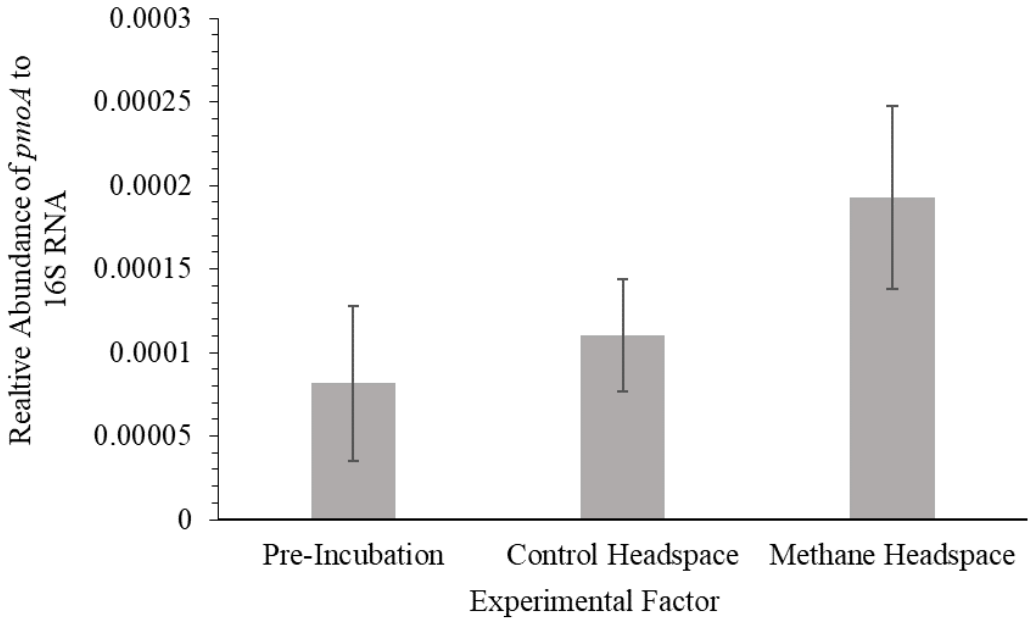



**Figure 3.7: Relative abundance of *pmoA* before and after incubation.** Error bars are the errors propagated from both the 16S rRNA and pmoA assays using ±2SD.

## 3.4.3 Discussion

### 3.4.3.1 Headspace Composition

The loss of methane appeared to be related to the ratio of $O_2$ + $N_2$ and the $CO_2$ peaks. This suggests that, as the microcosms reached a state with limited $O_2$ for respiration, the oxidation of methane slowed with the greatest loss of methane seen between time 0 and 72 hrs (Figure 3.5) followed by continued loss at a reduced rate till 150 hours had elapsed. This was expected given methanotrophy is an aerobic process. It was surprising, however, that the drop in O2 and therefore aerobic processes, occurred that quickly.

It is possible that the rapid loss in $O_2$ enabled anaerobic processes to begin, including methanogenesis. The control headspace never contained detectable concentrations of methane. That said, it is possible that small amounts of methane were being produced which were being oxidised by methanotrophs. This might explain why there was a small increase in the relative abundance of *pmoA*. If the incubation had continued, the difference between the pre-incubation soil and the controlled headspace soil may have increased to a statistically significant level.

Anaerobic oxidation of methane might have begun in the microcosms. The sudden and pronounced drop in methane concentration between the 240 hr and 288 hr analyses

(Figure 3.5) would suggest active removal or loss of methane. The drop was clearly seen in both sets of microcosms. It is possible this is due to operator error during sampling or analysis using the GC. That said, given methane loss did not continue to decline after 288 hrs it is unlikely that it is due to active anaerobic oxidation of the methane. Instead, it is more likely that methanotrophic capability of methane oxidation at low $O_2$ concentrations contributed to the loss of methane. Without further analysis of the DNA (sequencing and abundance of *mcrA*) or a repeat of the experiment with destructive sampling to observe community changes through time, it is not possible to explain definitively what occurred.

#### 3.4.3.2 Relative Abundance of *pmoA*

The significant increase in *pmoA* compared to the control and the pre-incubation soil was expected. The presence of a high methane concentration stimulated the methanotrophic population leading to population growth. The increased methanotrophic population probably oxidised methane, leading to the loss of methane from the headspace. As methane is not accessible as a metabolic substrate to most bacteria, the proportion of methanotrophs within the total bacterial community grew. That said, DNA evidence only suggests that there had been an increase in the methanotrophic population; it does not provide any information on whether the cells from which the DNA was extracted were still alive and active. The increased abundance of DNA, however, enabled the molecular methods under development to detect the *pmoA* and then enable optimisation of the PCR and qPCR protocols.

It was surprising to see that qPCR could detect *pmoA* but the standard PCR could not. It is possible that the qPCR was more sensitive to increases in dsDNA than the gel electrophoresis and the subsequent image capture.

### 3.4.4 Microcosm Conclusions and Future Work

Soils incubated under an enriched methane headspace saw an increased relative abundance of *pmoA*. The methane in the headspace was depleted while aerobic conditions exist. It is likely that methanotrophs, detected using *pmoA*, are contributing to this depletion.

This work demonstrates that changes in the abundance of methanotrophs are detectable in soils using qPCR but standard PCR was not able to detect *pmoA* in these samples. This work also supports the use of DNA as a marker of community changes occurring over a period of days. However, DNA does not allow an assessment of community

activity; for this the use of RNA is needed.

### 3.4.5 Project Development

This work strongly supports the use of DNA in detecting changes in methanotroph abundance. However, the rapid loss of aerobic conditions within microcosms was a surprise and suggests that changes need to be made to the sampling frequency, headspace composition and soil to headspace ratio. In future work, more sampling will have to occur in the first 72 hrs to explore methane oxidation rate in more detail. The use of 15% methane in the headspace appears to be too high: the low oxygen content in the headspace meant aerobic processes stopped, therefore limiting the amount of data collected on the rate of methane oxidation. Furthermore, using a lower concentration of methane which is more representative of the environment will increase the validity of the assay. Finally, the use of a smaller soil volume/mass within the microcosms will increase the period in which aerobic conditions persist by both reducing the size of the microbial community and increasing the headspace volume. This study also enabled the development of PCR and qPCR assays. It highlighted detection and specificity issues and enabled the testing of multiple primer sets on a complex soil community.

## 3.5 Field Testing – Hardstoft One

Hardstoft, Derbyshire, is reported to be the site of Britain's oldest onshore oil well. Hardstoft 1 (H1) was sunk vertically in 1915 to reduce Britain's reliance on imported oil and remained productive until 1952. Two other wells were sunk within a 300 m radius of H1 (UK Onshore Geophysical Library, 2018), but do not appear to have been as productive.

The site was selected as it was reported to be unburied and contaminated with material rising from the subsurface. As a result, the well head could be located at the surface and a sampling strategy designed to include it as a positive control. This allowed a thorough testing of sample collection, field methods (gas flux measurements) and molecular methods (qPCR). Furthermore, the site was also considered ideal for testing the hypothesis that increased methane and alkane fluxes could be detected at the field scale using changes in near-surface microbiology. This site enabled a thorough testing of the field-based methods. Overall the site aimed to test whether or not qPCR was sensitive enough to detect changes in the soil community over small field scales and to help identify the best spacing of sample grids at field sites.

It had been hoped that modelling the variability of the two functional genes and the 16S rRNA would inform future fieldwork. The models could have aided selection of the optimum grid spacing, to capture an ideal resolution of data. Sadly, the site proved to be too homogenous and not enough samples were collected to enable this. While the site did not manage to address all the original research questions it served as an excellent test bed for future work.

### 3.5.1 Methods

The original field design used a 40-45m transect starting at the oil-well. This was based upon the position of the well-head reported by UK Onshore Geophysical Library (2018) and permitted a transect running parallel to the contours, on grassland, (Figure 3.8). Samples were taken every 5 m along the transect, within a 1 $m^2$ area. The well head was incorrectly mapped, rendering the planned transect untenable. Instead, soil sampling was performed down a slope with a varied topsoil depth (limiting transect length and stopping soil collection at 30 m) (Figure 3.8 and Figure 3.9). The area around the well head was water-logged. In total, five points were sampled: one at the well head and then at 5, 10, 20 and 30 m along a transect laid out in a south westerly direction away from the well.

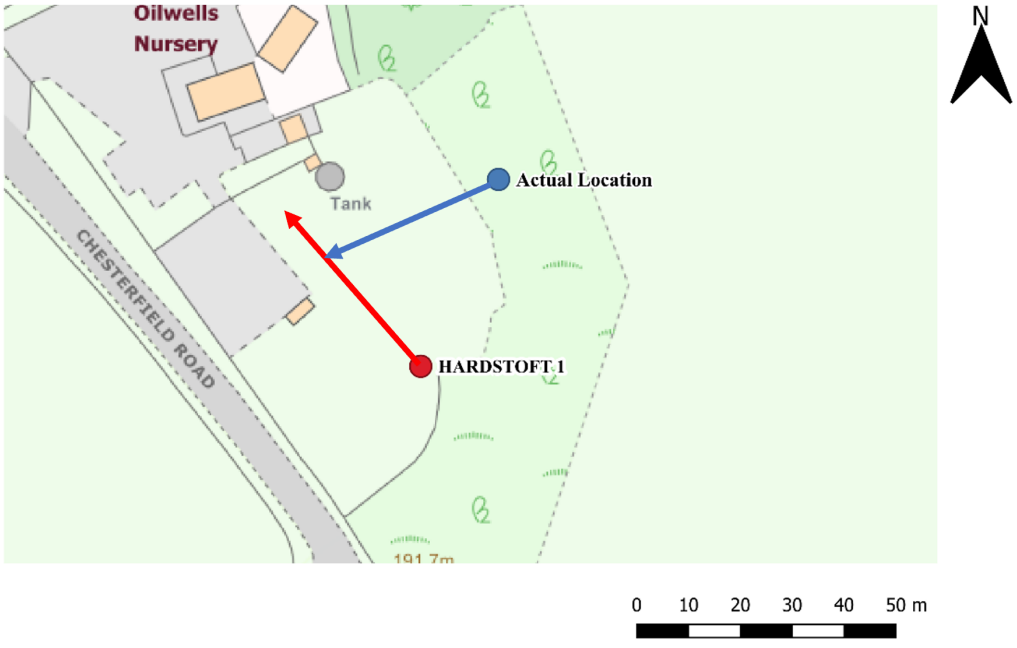



**Figure 3.8: Map of Hardstoft-1 Field Site.** The red point and transect were planned before the site visit but had to be changed as the oil-well had not been accurately mapped. Instead, the transect headed SW from the actual location of the well, as shown in blue. Basemap: OS MasterMap®, (EDINA Digimap October 2018)

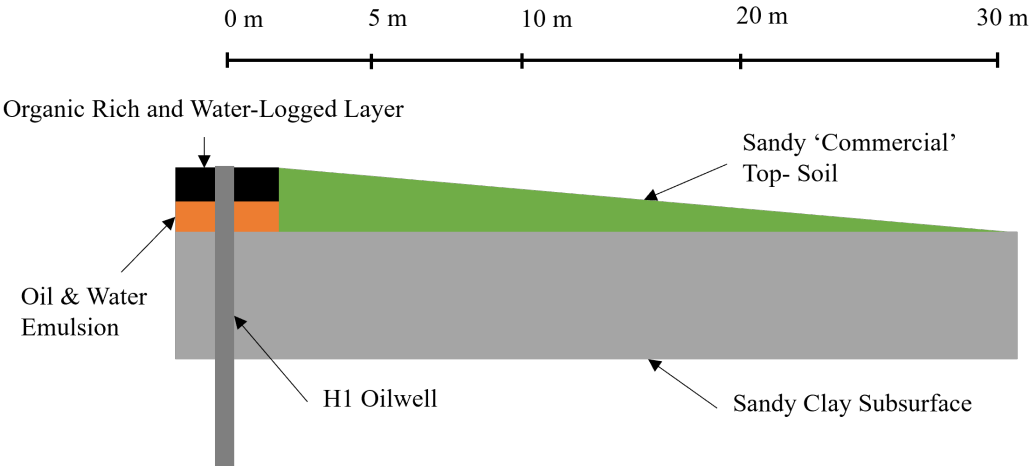



**Figure 3.9: Sketch (not to scale) of the soil boundaries along the site transect.** Sampling points along the transect are indicated by the bar at the top. The soil around the well-head was water-logged and consisted of at least three distinct layers. The commercial topsoil gradually became thinner down the transect, at 30 m the topsoil was too thin to collect a sample

#### 3.5.1.1 Site Characterisation

Pore gas samples were collected at a depth of ∼10 cm. Steel pipes were hammered into the ground and a 20 ml syringe was used to obtain two samples of pore gas. Each sample was then stored in an Exetainer® which had been evacuated prior to sampling. One sample was analysed using gas chromatography for pore gas composition (see Chapter 2.2.3 for methods). The second sample was used in prototype isotope analysis at the British Geological Survey to identify the isotopic signature of the carbon and hydrogen. Gas flux was estimated using a Portable Flux Meter (West System, IT), three measurements were taken at each transect sampling point. Three soil samples were collected at each point along the transect, from a depth of 25 cm. Soil pH and moisture content were determined in the lab using these samples (Chapter 2.2).

#### 3.5.1.2 Microbial Characterisation

Each point along the transect was sampled multiple times for molecular analysis. Soil underwent DNA extraction within 24 hours of sampling after storage at 4°C. Extracted DNA was used in qPCR assays to estimate the abundance of key functional genes. The soil layers around the well-head were treated separately given their obviously different physical characteristics. Methanotrophs were quantified using qPCR assays for *pmoA* and *mmoX*; the assays are described in depth in Chapter 2.3. To estimate the total bacterial community a 16S rRNA assay was used.

#### 3.5.1.3 Statistical Analysis

Sample replicate results, for both PCR and gas flux, were pooled to produce averages. The proportion of methanotrophs within the community, estimated using a ratio of the functional gene to the 16S rRNA copy number, was used when presenting the results. Propagated errors of the estimated relative abundance were generated using two standard deviation (Chapter 2.3.7.4). The impact of distance from the well on the relative abundance of functional genes was completed using a one-way ANOVA with subsequent Tukey tests were used to explore differences between groups.

### 3.5.2 Results

#### 3.5.2.1 Soil Environment

The soil around the well head was water-logged (42.8% soil water, dried for 3 weeks at 40°C) and heavily contaminated with crude oil. As sketched in Figure 3.9, the material around the well head consisted of three layers through the top 50 cm. The surface layer, around 15 cm in depth, was black, filled with partially decomposed matter and contained obvious evidence of oil. Below this was an orangey coloured layer that extended another 10-15 cm, this had no obvious structure, was an emulsion of water and oil and contained recalcitrant, partially decomposed, plant material. At ∼25-30 cm depth a clay rich, hydromorphic layer, with characteristic gleyic mottling, started and extended deeper than 50 cm. This layer was also contaminated with oil. The maintained grassland along the transect was a well-drained sandy topsoil overlaying a grey sandy-clay subsoil with red mottling suggesting periods of anaerobic conditions.

**Table 3.1: Soil pH values along transect and the SD of the value.** Soil collected was from 10cm depth. *At 0 m there was not enough material left to produce replicates of the pH and the value is based only on the surface material.

| Distance Along Transect (m) | pH | SD (to 1 DP) |
|---|---|---|
| 0* (-50 cm) | 6.93 | - |
| 0* (-10 cm) | 6.42 | - |
| 5 | 7.36 | 0.1 |
| 10 | 7.31 | 0.3 |
| 20 | 8.02 | 0.1 |

Soil moisture and pH were measured in the laboratory. These data were to contextualise any variability in the gas flux and gene abundance data. The soil pH at the well head (6.4) was lower than the rest of the transect, Table 3.1. The soil water content at the well head was higher than the rest of the transect (Figure 3.10).

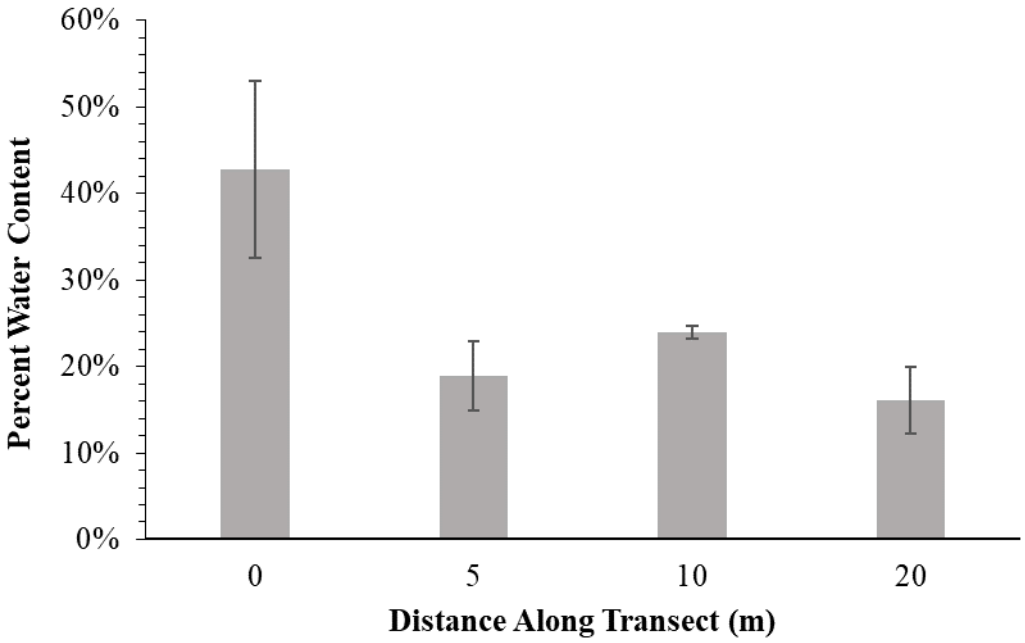



**Figure 3.10: Soil water percent along the transect at Hardstoft 1.** Determined gravimetrically by drying at 40°C for three weeks. Value for soil at 0 m uses data from the -10 and -50 cm soil layers. Error bars are ±2SD of the mean.

#### 3.5.2.2 Gas Flux

Methane efflux was observed at the well-head (1.6 ppm s-1 $R^2$=0.99). In addition to the relatively high flux, the meter also detected methane concentrations reaching 23,000 ppm. No other sample points had an observable methane flux but, as shown in Table 3.2, all sites had observable effluxes of $CO_2$. Pore gas samples were taken at the 5, 10 and 20 m points from a depth of ∼10 cm. These were analysed using gas chromatography; no detectable methane was observed but $CO_2$ was detected (average across all sites =827±239 ppm 2SD).

#### 3.5.2.3 Isotope Analysis

Gas samples from around the well-head were collected for isotope analysis to estimate the origin. Pore gas collection was impossible given the water-logged nature of the soil around the well. Instead, surface water samples were collected and injected into vacuum sealed bottles; the resulting degassed atmosphere was analysed. Alongside

**Table 3.2: $CO_2$ flux along the transect and $R^2$ value of the fitted trendline.** The $R^2$ is generated by fitting a linear trendline to the $CO_2$ concentration against time. Flux is shown to 1 DP.

| **Distance Along Transect** (m) | **$CO_2$ Flux (ppm $s^{-1}$)** | **$R^2$** |
|---|---|---|
| 0 | 0.6 | 0.99 |
| 5 | 1 | 0.99 |
| 10 | 0.9 | 0.99 |
| 20 | 0.6 | 0.99 |
| 30 | 0.8 | 0.99 |

these, samples of gas were collected from the accumulation chamber of the flux meter. Due to problems with the instrument only $\delta$D-$CH_4$ was measured; results are shown in Table 3.3. The increased error and marginally lower $\delta$D-$CH_4$ values for the air samples may be related to the method of collection.

**Table 3.3: Shift in hydrogen isotope (deuterium) of methane.** Samples were collected from surface water and from air within the accumulation chamber of the flux meter. Samples was analysed multiple times to generate an error value.

| **Sample Type** (m) | **Repeats** | **$\delta$D-$CH_4$ (‰)** | **Error (‰)** |
|---|---|---|---|
| Water | 6 | 322.8 | ±2.0 |
| Water | 7 | 319.0 | ±1.0 |
| Air | 3 | 285.8 | ±6.5 |
| Air | 3 | 299.8 | ±3.7 |

#### 3.5.2.4 Functional Gene Abundances

As shown in Figure 3.11, the 16S rRNA assay suggests that the two deeper soil layers around the well-head have a very small bacterial community compared to the other points along the transect. Distance from the well head was predicted to have a significant effect on the relative abundance of both *pmoA* ($F_{3,10}$=8.632, $p$<0.05) and *mmoX* ($F_{3,10}$=3.481, $p$<0.01). Post-hoc Tukey tests showed that within the soil layers around the well-head, the relative abundance of *pmoA* and *mmoX* was significantly higher than at all other points along the transect ($p$<0.05). Suggesting MMO possesing organism

comprise a larger proportion of the total community than at other point along the transect (Figure 3.12).

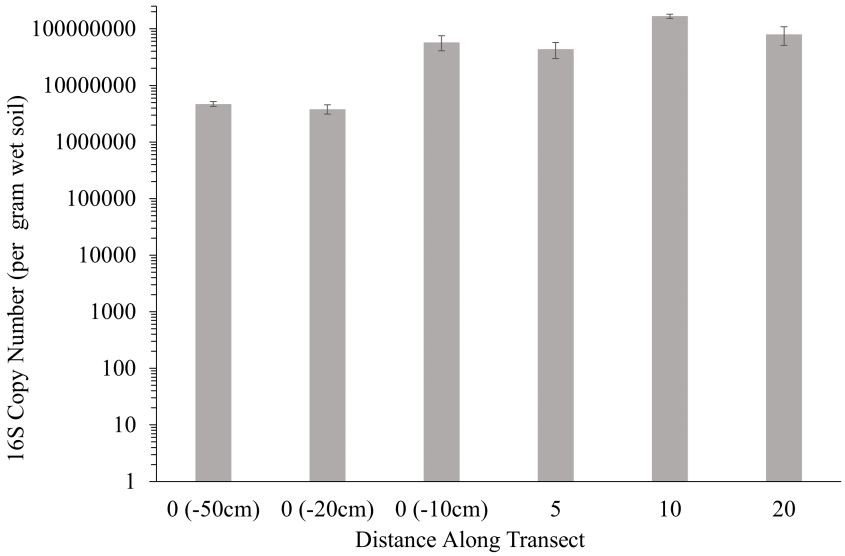



**Figure 3.11: 16S copy number at each sampling point along the transect.** The sampling point at the well head (0 m) has three depths, with an assay completed on each soil layer. The 16S copy number can be used as an estimate of total bacterial abundance. Error bars are ±1SD.

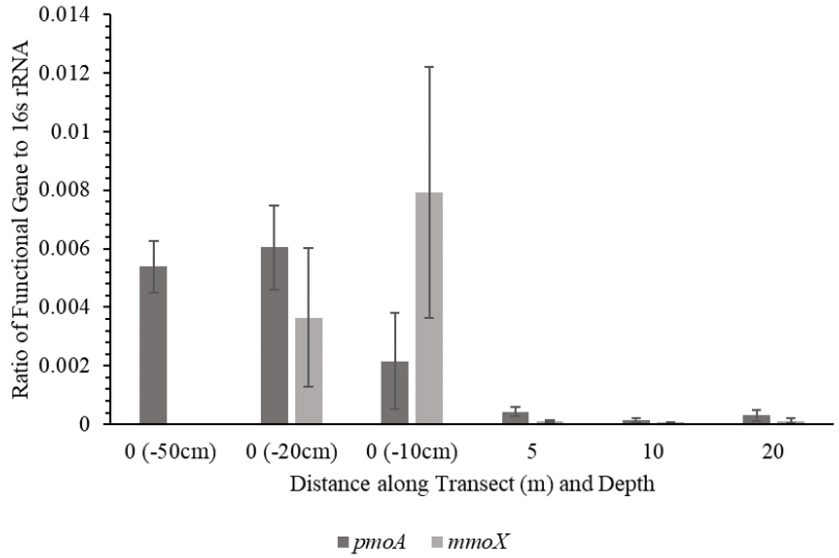



**Figure 3.12: Ratio of methane mono-oxygenase genes to 16S rRNA generated through qPCR assay of relative abundance.** The sample point around the well-head (0 m) was sampled and analysed at three different depths within each of the unique soil horizons. Error bars are the propagated errors generated using 1SD.

### 3.5.3 Discussion

A transect, originating from an abandoned oil-well, was used to assess field-scale variability of key indicators and factors in alkane emission and monitoring. The presence of crude oil and a water-logged sample point combined to make sampling and analysis problematic.

#### 3.5.3.1 Microbial Community

The DNA extractions were complicated due to the presence of oil pollution. However, the material extracted was used successfully in qPCR assays to estimate relative abundance of key genes, *pmoA* and *mmoX*, using 16S rRNA as a marker for total population size. Increased relative abundance of genes related to methane oxidation was observed around the well head, as shown in Figure 3.12, which was observed alongside increased methane flux.

The low bacterial abundance at soil depths of -50 and -20 cm observed around the wellhead, shown in Figure 3.11 is potentially explained by a number of factors. Hydrocarbon contamination provides an increased pool of carbon substrates for catabolism while also potentially being toxic (Alrumman *et al.*, 2015; Labud *et al.*, 2007). Given the age of the site it would be expected that microbial communities would contain members which are able to exploit these resources. Indeed, studies have shown that contaminated soils quickly develop hydrocarbon degrading communities (for example: Alrumman *et al.*, 2015; Margesin *et al.*, 2000) and the initial exposure can prime the soil community for future contamination events (Kauppi *et al.*, 2012; Greenwood *et al.*, 2009).

The characteristics of the soil have an impact upon the toxicity of contaminants and the overall impact of the contamination. Studies involving soils have repeatedly shown that soil type influences community adaptation. Soil physico-chemical factors can influence the pre-contamination community size, thus leading to different community responses following contamination (Alrumman *et al.*, 2015). Furthermore, the same contaminant does not necessarily select for the same hydrocarbon degraders in different sites, with no convergence seen in community composition between different soils (Bundy *et al.*, 2002). Soils with a greater cation exchange capacity, i.e. soils with increased clay or organic matter content, can have lower bioavailable fractions due to the increased absorption of contaminants to the soil matrix (Labud *et al.*, 2007; McEldowney *et al.*, 1993). The deepest soil layer around the well head, a hydromorphic, clay rich soil, could have absorbed significant quantities of contaminants, but there

was obvious evidence of free oil in the pore space suggesting that age and volume of exposure had negated any of the potential effects of soil texture.

Another consideration is the DNA extraction technique and sensitivity of the qPCR assay to contaminants carried over from the soil. It is plausible that the complex starting material has hindered these techniques and that the reduction in the in total 16S rRNA copy number simply represents a limitation of the laboratory techniques used. This does not downgrade the usefulness of the relative abundance results; it is plausible that all assays on the same sample were equally impacted, therefore the ratios between sites remain comparable. As a result, the dataset highlights that the use of relative abundance to compare sample points is probably the best approach for these samples. Other biological indicators can be used in contaminated sites to explore the microbial population. Total microbial biomass, enzyme activity (e.g. dehydrogenases or lipases) can all provide insights into the size and activity of a population (Margesin *et al.*, 2000).

#### 3.5.3.2 Methane Flux

The only point sampled with a detectable methane flux was immediately adjacent to the well-head. The methane flux was observed alongside the increased relative abundance in *pmoA* and *mmoX*, compared to samples from elsewhere on the transect. This suggests the two are related and supports the hypothesis that microbial community composition changes in response to natural gas flux. To further test this, the extraction of RNA from the soil for use in the qPCR assay could be completed. This should show if the methanotrophic population is actively oxidising methane. However, the high pollution levels will hamper work to achieve this.

The origin of the methane is not clear due to the lack of carbon isotope data. The waterlogged soil environment is almost certainly contributing to any methane flux through anaerobic reduction of organic matter. This is supported by the isotope analysis which, if compared to a diagrams from Whiticar (1999), suggests that a mix of both biogenic and thermogenic methane is possible. Isotope analysis allows a differentiation between sources due to the preference of biochemical process for lighter isotopes. The large variability within the air samples is probably due to the collection method; sampling from the accumulation chamber potentially allowed atmospheric methane contamination and atmospheric methane has a low negative $\delta$D-$CH_4$ shift due to photochemical and hydroxyl reactions within the atmosphere (Whiticar, 1999). While it is not clear whether this methane is, at least partly, of thermogenic origin this work supports the project's approach to use microbiology as an indicator of increased methane flux by

illustrating that increased flux does occur alongside an increased relative abundance of methanotrophs.

#### 3.5.3.3 Other Soil Variables

Alongside gas flux, pH and soil moisture content were also measured. Around the well-head the soil was both water-logged and had slightly lower pH. Here, a comparison between the grassland soil's pH and the pH around the well-head cannot be made given the substantial differences between soil structure, pollution levels and due to the design of the fieldwork. Interestingly though, soils with lower pH (6-6.5) have been associated with increased heterotrophic abundance even in the presence of contamination (Alrumman *et al.*, 2015). Despite the reducing environment around the well head, degradation of components within the crude oil is still possible.

The polluted and anaerobic nature of the site could be limiting soil fauna contributing to the lack of mixing between the top two layers. Earthworms can potentially accelerate the removal of organic contaminants (Dendooven *et al.*, 2011) but high concentrations of oil (>1%) are also toxic (Shakir Hanna and Weaver, 2002). The increased bacterial abundance in the uppermost layer (-10cm) could be due to the increased oxygen concentration and/or an increased prevalence of semi-decomposed organic matter and therefore increased substrates for bacterial catabolism alongside marginally reduced oil pollution levels, but this would require further work to corroborate.

### 3.5.4 Hardstoft – Conclusions and Future Work

Britain's oldest oil well, Hardstoft One, is currently a heavily polluted and waterlogged site. A transect of samples emanating from the well showed that the area around the well head contained both a high methane flux and a high relative abundance of methanotrophic genes within the soil community. The observation of an increased flux alongside an increased prevalence of methane oxidation genes supports the hypothesis that relative abundance can be used as an indicator of increased methane flux. Using an RNA extraction method would have enabled an assessment of the methanotrophic community activity, but the high contaminant burden and resulting low RNA recovery would probably have not allowed for a successful qPCR assay.

### 3.5.5 Project Development

The work presented here provided a series of unique challenges, illustrating potential weaknesses in the approach. The severity of contamination at the site was unexpected

and was unlikely to be encountered again but highlighted the problems caused by sample inhibition. The oil, high concentration of organic matter and high clay content of the soil impacted DNA recovery and led to inhibition of subsequent molecular assays. Template dilution allowed these problems to be overcome but caused the target's concentration to drop to near the limit of detection (LoD) of the assay. A method of ascertaining highest possible dilution and minimal dilution was sought leading to the development of the inhibition assay described previously.

It had been hoped that the site would illustrate field variability of the functional genes under investigation. To design survey strategies with the correct resolution, prior knowledge of spatial variability is invaluable. Through modelling, an estimation of spatial variation can be made which enables optimal grid spacing to be devised for accurate Kriging interpolation methods in future work. While the data would not have been directly applicable in all systems, it would have been a useful tool when considering future work. Sadly, the homogenous nature of the site meant too little variability was observed to enable this modelling.

## 3.6 Determining Optimum Sample Dilution

Even with a robust and reliable qPCR and PCR assay, problems arose from the sample under analysis. The study at Hardstoft One, a preliminary piece of field work, illustrated that sample inhibition from complex samples would quickly reduce the reliability, accuracy and precision of the assays being used. Initial attempts generated little or no PCR product from the DNA extracts originating from heavily polluted samples. To overcome this problem, the sample was diluted 1:10 with ultra-pure water. This enabled the amplification of template but limited the assay by decreasing the template concentration to near the LOD. While qPCR and PCR can theoretically amplify from a single copy, reliable amplification is rarely seen in practice, even using clean standards diluted to $10^2$ copies $\mu l^{-1}$ and, at this concentration, confidence in the $C_q$ value is limited. The limit of detection is pushed further when using environmental templates where the assay must compete with inhibitors such as those found in soil.

Soil is chemically and physically complex, formed through the chemical and physical weathering of underlying geology and the accumulation of organic matter. The largest constituent by volume is mineral matter, making up 40-60% of soil (White, 2006). Mineral material in soil can be broadly divided into three continuous groups, sand, silt and clay, depending upon particle size. Sand and silt are the recalcitrant remains of the parent material, predominantly made up of silicates. Clay minerals are largely

phyllosilicates, which have a large, charged, surface area and are therefore important in soil chemistry. Pore water and gas normally make up 20-50% and 10-25% respectively, followed by soil organic matter.

Soil organic matter (SOM) includes all organic material in the soil. This includes living biomass and organic debris. SOM includes humus, which must be chemically extracted from the soil matrix using a solvent. Humus is a complex, and poorly understood, mixture of organic compounds bound to the mineral fraction. It consists of covalently bonded aromatic and alkyl compounds, forming complex polymers with a large mass and surface area. Humic material is important in soil structure and formation. Phenolic and carboxyl functional groups are important in cation exchange while the complex polymers bind the mineral fraction together.

As reviewed by Wilson (1997), PCR inhibition occurs when a substance interferes with DNA and polymerase interaction. Inhibitors can outcompete polymerases; clays and humic substances, that can be co-extracted from soils, bind strongly to DNA thus stopping the polymerase from completing amplification. It is also possible that material carried over from the soil, such as phenolic compounds, can denature the polymerase. As a result, inhibition was a potential problem with all DNA extracts from soils. Diluting samples offered a solution but only where template concentration could be kept above the LoD.

A method to estimate the optimum dilution was devised. Initially, a published program was sought. The PREXCEL-Q (P-Q) program described by Gallup and Ackermann (2008) sought to improve the reproducibility and uniformity of qPCR assay design. It claimed to aid the user in finding a suitable dilution series for each sample and ensure a near perfect efficiency. Unfortunately, the program had not been maintained and was unavailable for use.

The method developed here enables a target to be amplified efficiently whilst minimising the risk of over dilution and therefore loss of amplification. Ideally, each environmental sample would be diluted by a sample-specific factor, depending upon the inhibitors present; however, this is not feasible due to time and cost restraints. Therefore, each set of samples, i.e. all those from the same site, were diluted equally. As each group of samples had similar soil characteristics, inhibition was assumed to be similar across that site. At sites where different soils, or different horizons, were sampled, the method could be altered to account for this.

### 3.6.1 Example Use

The first complete use of this approach was with soil samples from the French Alpine seeps ( Chapter 4). The field campaign generated 18 samples, three from a background area. The background soil had a greater proportion of organic matter. Samples underwent DNA extraction with the soil mass used in the extraction recorded. Three samples were selected for inhibition testing: one background sample plus the extracts generated from the greatest (0.43 g) and smallest soil mass (0.27 g). All three samples were diluted to the points shown in Figure 2.1 and used in a 16S rRNA qPCR assay alongside a standard series. The standard series demonstrated the assay had a usable range of $10^3$-$10^7$ copies $\mu l^{-1}$ (E = 101.7%, $R^2$ = 0.99). Generated $C_q$ values were grouped, as shown in Figure 2.1 and were plotted against the $\log_{10}$ of the dilution enabling E-values to be calculated, an example plot is in Figure 3.13.

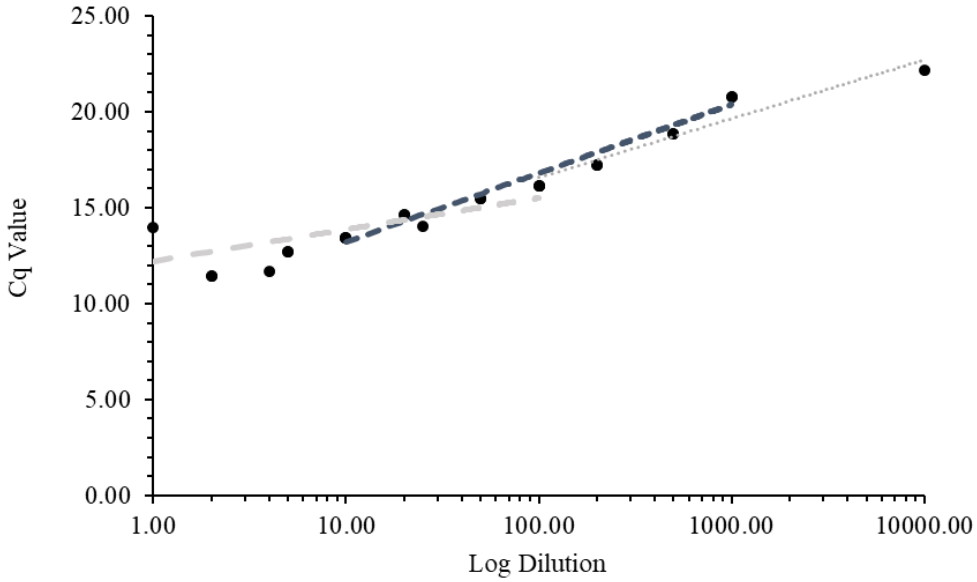



**Figure 3.13: Example plot of $C_q$ against dilution with three example trendlines**. Trendlines are generated using groups of five dilutions, described in Figure 2.1 (here only three dilution groups are shown for clarity). The legend refers to the smallest dilution used in the series to produce the trendline. This sample is showing inhibition; the 1:1 dilution does not fit the expected trend with a higher $C_q$ value (13.95) than the 1:10 dilution (13.43). The 1:1 trendline leads to an E-value of 305.2%. The 1:10 trendline has an acceptable E-value (E = 90.2%, $R^2$ = 0.97). The 1:100 (E = 111.5%, $R^2$ = 0.92) is probably exhibiting stochastic error due to low template concentration.

An average $C_q$ value for each dilution grouping was produced from the three samples and is displayed in Table 3.4. The 1:10 dilution group was selected given its similarity to the standards and that the average E-value ± CI best matched the desired range of E values (90%-110%). No one dilution would suit all three samples. The background

sample still showed some inhibition in the 1:10 dilution series (E = 115.2%, $R^2$ = 0.99), possibly due to the increased organic matter in the sample. Conversely the other two samples were potentially being over diluted if the 1:100 dilution was selected. Even with this method some subjective decision making is required; overall, however, the approach was considered a tool to better inform sample preparation and improve the likelihood of successful qPCR assays.

**Table 3.4: Average efficiency for each group of dilutions and the 95% CI of that value.** The average is produced from all three samples used in the test. The smallest dilution is the most concentrated sample used in that group of dilutions, the other dilutions group with it to generate E value are in Table 2.2. The efficiency of the standard series in included, as this was only run once no CI could be calculated.

| Smallest Dilution | Efficiency | CI (95%) |
|---|---|---|
| Standards | 101.68% | - |
| 1:1 | 211.63% | 96.56% |
| 1:2 | 144.35% | 40.18% |
| 1:5 | 114.25% | 5.26% |
| 1:10 | 104.45% | 14.60% |
| 1:100 | 113.40% | 6.40% |

### 3.6.2 Caveats of Approach

There are limitations with this approach. Firstly, it assumes that inhibition will impact all qPCR assays equally, for example the 16S rRNA assay and *pmoA* assay will be inhibited equally. Secondly, it uses an external standard as a benchmark of optimum efficiency; this involves a template well removed from the sample both in terms of quality and purity.

Other approaches to the problem could have been attempted. An internal recovery standard was a potential solution. This could have taken the form of a multiplex reaction, i.e. with two targets and amplicons in one reaction, for example the SPUD assay (Nolan *et al.*, 2006). however the issue of whether both reactions were equally inhibited remained. Burggraf and Olgemooller (2004), presented a method for reverse-transcriptase PCR (RT-PCR) using an internal standard, consisting of single-strand oligonucleotides, which would be amplified using the same reaction. This was a synthetic strand which contained only the primer and probe binding sites and could be

differentiated in the melt analysis. The authors did not believe this would be usable in a qPCR setting and after exploring the idea for this project it was rejected for the same reasons.

Overall, the simplicity and ease of performing a dilution series made it the most viable option. The two caveats are both problems encountered when trying to normalise the data, (see, Chapter 2.3.7.4 and 3.2.3) which is necessary to compare samples. Therefore, a dilution series was not introducing further considerations or assumptions when analysing the dataset. Sample dilution testing aims to reduce error within samples, ensuring target detection and strong amplification; the normalisation enables direct comparison between samples, but both make the same assumptions regarding equal inhibition between assays and comparability of an environmental (and inhibitor laden) sample with standard. As a result, the data must be acknowledged as only an estimated relative abundance, but this improves upon methods that do not attempt to correct for these problems.

## 3.7 Choosing Sequencing Platform

Oxford Nanopore Technology's platform was selected due to several key benefits. Firstly the portability of the unit, reagents and supporting hardware enabled the use at field sites were the transport of samples back to the laboratory in a stable state was not possible. Secondly, the long read length of amplified sequences carried inherent advantages for amplicon sequencing and subsequent taxonomic assignment. Use with longer DNA sequences, i.e. from more distance primer pairs, theoretically allowed for easy isolation and exploration of more variable regions of DNA between, or surrounding more conserved sites. This in turn allowed for a greater assessment of gene variability and therefore potentially community diversity.

Early studies exploring the reliability of the platform had demonstrated consistent results were possible even between laboratories. With ONT's first products Ip *et al.* (2015) demonstrated across five laboratories that with the same protocol and control target sequence 93% (median) of reads could be correctly assigned with 12% (median) reads considered an error. As reviewed by Wang *et al.* (2021), yield, read length and accuracy have climbed from 2015 to 2020. With the technology and chemistry in constant development the accuracy and reliability will likely improve, with improvements in software potentially allowing for reanalysis of previous sequencer runs. Indeed, during this project raw squiggles were reanalysed (base-called) as new software was released.

The use of newer and less developed technology carried disadvantages. The pipeline described in Chapter 2.3.9 is mostly comprised of packages that were developed for other sequencing platforms that generate shorter reads. Other platforms, for example Illumina, also have inherently lower error rates, with bases generally called with >99.9% (Q-Score=30) certainty compared to 90% with ONT (Q-Score=10). Whilst, longer reads might initially be thought to help counter this increased error rate instead longer reads merely carry more erroneous data further limiting the confidence in the dataset. Cumulatively the use of ONT sequencing led to difficulty processing the data as often packages where not designed to handle reads with such low accuracy or that are that long.

Overall, the challenges posed by using the ONT platform were considered acceptable given its benefits. Computational challenges only became truly apparent once datasets were being analysed, as did the difficulties inherent to the high error rate. That said, the ability to quickly sequence target functional genes in an environmental sample remained a strong attraction, without the disruption of Covid-19 it would have been an interesting addition to the thesis.

## 3.8 Chapter Findings

This chapter presents the work that laid the methodological foundation for the rest of the project. Methods to measure the rate of methanotrophy in soils were developed from the microcosm experiment. The Hardstoft One study was invaluable in developing qPCR assays. Furthermore, this work provides evidence that the use of DNA could detect shifts in methanotroph abundance, even over timescales of days. With the Hardstoft One study, the project's approach was further endorsed with the observations that changes in methanotroph abundance could be detected on field scales and that these shifts correlated to sub-surface methane fluxes. Finally, this work illustrates that, while the assays are sensitive, they are robust enough to work with difficult samples.

# Chapter 4

# Field Surveys of Methanotrophs Around Two Sub-Alpine Natural Gas, Macro-Seeps

## 4.1 Rationale

Sites were sought to study microbial community change in response to gas flux. Ideally, the site(s) would be an area(s) of thermogenic alkane seepage, increasing the likelihood that the methane would be accompanied with butane and propane. Increase oxidation rates of longer alkanes ($C_2$-$C_4$) were indicative of a hydrocarbon reservoir in traditional culture methods, with the ratio of methane to other alkanes being much larger in gas mixtures originating from biogenic sources (Brisbane and Ladd, 1965). The two sites selected for this work had well characterised fluxes of methane, with trace concentrations of ethane and propane (Gal *et al.*, 2018, 2017). Sampling at these sites aimed to compare changes in relative abundance of genes encoding for MMO with changes in alkane flux.

## 4.2 Introduction

Methane is a potent greenhouse gas, understanding and quantifying its release is important in monitoring and mitigating climate change. Methane is produced through three broad pathways: biogenic, thermogenic and abiotic (Etiope, 2015). The degradation of complex organic matter leads to methane production in both the biogenic and thermogenic pathways. Production of methane through thermogenic processes is associated with oil and gas development, leading to short chain alkanes accompanying methane in flux from the subsurface. The two gas sources can be distinguished by both C and H isotope data, and the wetness of the gas; biogenic methane is a very dry gas (low $C_1/C_2+C_3$) with a $\delta^{13}C$ less than -50‰(Etiope, 2015; Whiticar, 1999).

Detecting fluxes from soils can be achieved either directly where the methane is measured, or indirectly, where an indicator of the methane's presence is used. Microbial indicators of methane flux potentially have advantages compared to direct surface gas flux and pore gas measurements: flux is highly variable, changing in response to air pressure and soil moisture. Conversely, microbial community composition is thought to be much slower to change. For example, subsurface methane injection experiments by Cahill *et al.* (2017) demonstrated a microbial community could remain in a disturbed state for up to 253 days after exposure to methane. As a result, measurements of microbial community composition may be indicative of flux over previous months.

Methanotrophs are an excellent indicator species as the majority obligately grow on methane. As a result, an increased abundance should indicate an increased concentration of methane. Similarly, an increased abundance of butane and propane oxidising species could indicate an increased concentration of butane and propane. Using microbial indicators to distinguish a predominantly thermogenic from a biogenic source would have to rely on the shift in the gas wetness/dryness. Such a shift might be indicated by the increased abundance of alkanotrophs capable of oxidising $C_2$-$C_4$ alkanes, and a shift in the ratio of methanotrophs to these alkanotrophs.

Above oil and gas reservoirs micro-seepage of light hydrocarbons and alkanes occurs. Hydrocarbon degrading bacteria have been used as part of near surface oil and gas prospecting techniques for decades (Brisbane and Ladd, 1965). Oil and gas prospecting approaches used culture-based techniques to estimate methanotroph, and other alkanotroph, abundance. Wagner *et al.* (2002) described prospecting methods that suspended soil samples, before diluting and mixing them with nutrient solutions. These samples were split and exposed to either methane, or, an alkane mix. Cell growth estimated using a most probable number approach (MPN), alongside gas oxidation, en-

abled an enumeration of community members and activity. The data provided could be used to assess whether the soil community was above a hydrocarbon reservoir. Large cell numbers of propananotrophs and butanotrophs suggested a wetter reservoir and therefore a greater likelihood of oil than gas.

Elucidating the abundance of soil alkanotrophs and methanotrophs using molecular techniques should be possible. The use of quantitative PCR (qPCR) to amplify key enzymes in the oxidation of methane or $C_2$-$C_4$ alkanes has the potential be quicker and more accurate than culture-based techniques by removing time need for growth and removing the bias generated by counting only culturable species. The mono-oxygenases used in the initial oxidation of alkanes can be detected by qPCR. In methanotrophs there are two ioszymes of methane mono-oxygenase, a particulate (pMMO) and soluble (sMMO) form, both have well conserved catalytic subunits whose genes can be targeted in PCR and qPCR assays (Hutchens *et al.*, 2004; Costello and Lidstrom, 1999; Holmes *et al.*, 1995). Longer chain alkanes can be oxidised by a group the soluble di-iron mono-oxygenases (SDIMOs), again this group has a well conserved catalytic centre whose genes and have been detected in environmental samples using PCR (Cano *et al.*, 2013; Coleman *et al.*, 2006).

Attempts using molecular techniques in oil and gas prospecting have been published. Miqueletto *et al.* (2011) demonstrated increased diversity in SDIMO clone libraries from soils above hydrocarbon reservoirs but did not manage to quantify gene copy number using qPCR. Other authors have used abundance, diversity and ratios of *pmoA* and *prmA*, the catalytically active subunits of pMMO and propane mono-oxygenase respectively, to detect anomalies above known reservoirs and differences between background soils and those experiencing micro-seepage (Liu *et al.*, 2016; Zhang *et al.*, 2017).

A series of sites, with well described methane macro-seeps, are found in the French Alps. The source(s) of the methane seeps is unclear. Alongside methane, ethane and butane have been reported as trace constituents of the gas mixtures at some of the seeps (Gal *et al.*, 2018, 2017). Gal *et al.* (2019) explored the gas wetness at five seeps; based on this analysis, four are potentially from a thermogenic source. The fifth site, the Fontaine Ardente du Gua, was much drier and probably originates from a biogenic source within the outcropping black shales.

These sites offered an opportunity to explore if soil microbiology on a field scale can be used as an indicator of alkane gas flux. Molecular microbiology tools are tested to explore if they offer the necessary precision and accuracy on sites covering tens of

metres. Two sites were selected for this work. The drier seep, the Fontaine Ardente du Gua (hereon referred to as FA) and a wetter seep, La Rochasson (LR). The purpose of this work was to compare alkane flux measurements with the relative abundance of microbial communities that can oxidise alkanes. Furthermore, the impact of these seeps on the overall community composition was also explored using 16S rRNA sequencing. If microbial indicators can accurately identify areas of increased alkane flux, then the relative abundance of these indicators will be positively correlated with alkane flux. In turn, communities in proximity to seeps will have a different composition given the increased presence of community members that can utilise carbon sources provided by the seep.

### 4.2.1 Chapter Objectives

Cumulatively this Chapter aimed to test:

- That shifts in soil methanotrophic communities can be detected over sites only tens of metres across using qPCR and amplicon sequencing.
- That points within sites with increased methane flux will have a corresponding increase in methanotrophs.

## 4.3 Chapter Methods

Two sites in the French alps, with known macro-seeps, were sampled. One soil sample for each survey point was collected for analysis of the microbial community, alongside measurements of gas-flux and pore gas samples. At the sites, three points, well removed from the main seep, were also sampled for gas flux and community composition as background points. Each point sampled for soil had one sample taken, this was spilt for all subsequent analysis.

### 4.3.1 La Fontaine Ardent

The FA lies around 13 miles South of Grenoble. Here, a total of 15 points spaced in a grid with five metre spacing,, were sampled around a seep which could be ignited and remain lit, a further three 'background' points were also sampled (F83-F85) (Figure 4.1). The seep is thought to originate from surrounding argillaceous limestones (black shales) or mid-jurassic claystones (Gal *et al.*, 2017). The site had undergone significant disturbance, with the main seep originating from a borehole bringing the gases to the surface; this had been sunk in the early 1990s by the Bureau de Recherches

Géologiques et Minières (BGRM). At the surface, the gas emerged at several points across a concrete pad, with the main seep within a small stream of water. There was a thin layer of soil, gravel, and stone on top of the concrete pad. The area around the pad generally flat leading into a river along the North-West of the site (Figure 4.1. The soil is a gravel rich, clayey sand, with little organic matter, with no cracking at the surface. Soil samples were collected from a depth of 20 - 30 cm.

### 4.3.2 La Rochasson

The seep is on the Northern edge of Grenoble. The emission point was in the middle of a steep slope. The site had recently (circa 2016, Gal *et al.* (2019)) experienced a landslide moving material away from the seep. From LR-1 to LR-21 (Figure 4.2) the slip had exposed a heavily weathered shale or mudstone material which could be augered to a shallow depth. This fits with description of the seep being in a shale formations similar to those at FA (Gal *et al.*, 2019). Overall, ten samples were collected from around the seep whilst even spacing across a grid was attempted the steepness of the site made this difficult. Three background samples and two (LR-14 and 21) samples within the grid came from a forest soil, which presumably had once overlain the entire site. The gases reaching the surface could be ignited but only at a single sheltered point; this quickly extinguished itself.

### 4.3.3 Soil Sampling and Preservation

Soil, for community analysis, was sampled using an auger at a depth of 20 - 30 cm. At FA a regular grid with 5 m spacing was possible, at LR given sampling was more irregular due to the soil and slope. Between samples, the auger was cleaned using disinfectant and DI water. Soil was transferred into sterile 50 ml Falcon tubes, sealed and placed on ice. Samples were stored at -20°C once at the laboratory until DNA extraction could take place. Upon extraction the samples were kept frozen at -20°C and aliquoted to reduce the impact of freeze-thaw.

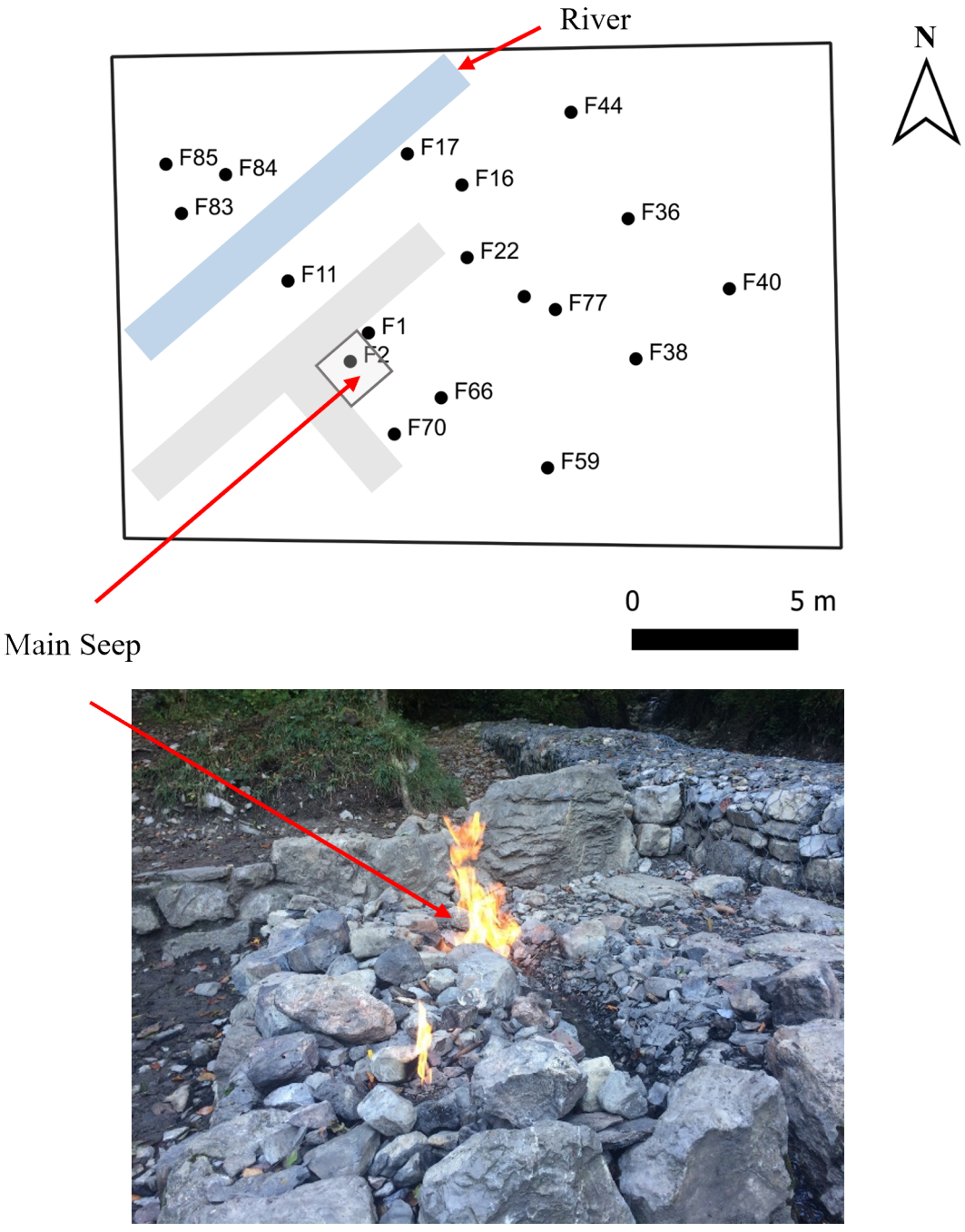



**Figure 4.1: Sketch of FA field site and photo of the main seep.** Points marked in the sketch are those used in soil sampling and subsequent microbial analysis. The seep could be lit and remain lit, there was larger seep at the centre of a concrete pad. There were four smaller points nearer the corners that could also be lit, one is shown lit in the photo.

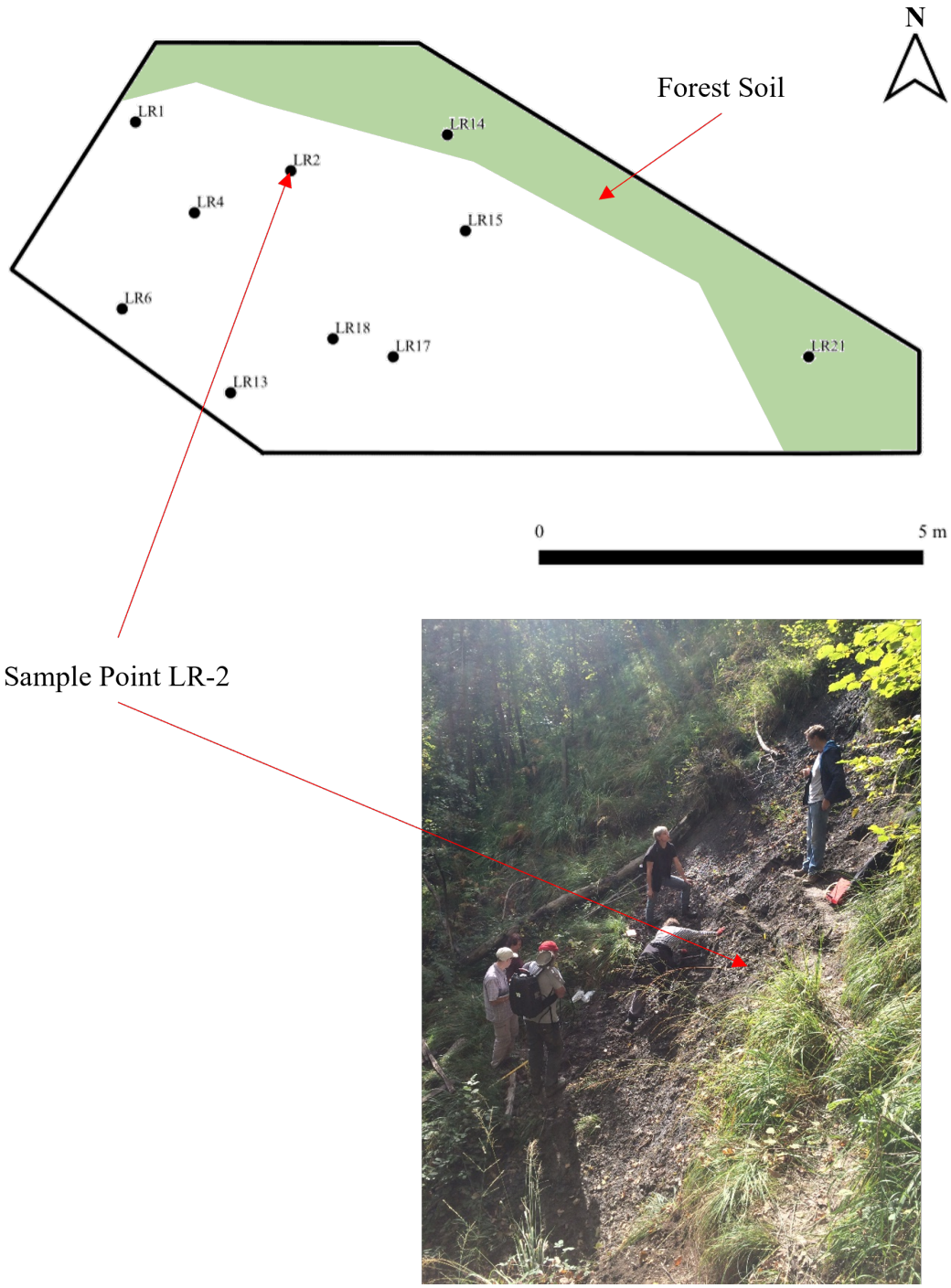



**Figure 4.2: Sketch of LR field site and photo of site.** Points marked in the sketch are those used in soil sampling and subsequent microbial analysis. The main seep is at LR-1 and would not remain lit. Photo is taken from LR-14 looking SW.

### 4.3.4 Soil Characterisation

Moisture and pH measurements were made at the points sampled for microbial analysis. Soil moisture was observed *in-situ* using a ML3-Theta Probe (Delta-T Devices, UK). Soil pH was measured in field using a handheld pH meter and probe (Hanna Instruments, UK) (2.2.1).

### 4.3.5 DNA Extraction and Analysis

Samples underwent DNA extraction within a week of collection (2.3.2). Samples were homogenised under sterile conditions before approximately 0.3 g of soil was processed. Each point sampled had one DNA extract, this was subsequently analysed in triplicate. To assess inhibitor carryover three samples from each site were used (Chapter 3.6). Relative abundance methanotrophy genes were estimated using the qPCR methods (2.3. Sequencing of the 16S rRNA gene using ONT's Flongle platform was completed on a subset of 12 samples FA (2.3.9).

### 4.3.6 Gas Sampling

Gas sampling was completed by the BGRM. Flux measurements were made at both sites across a regular 1 m grid with an Echo Soil Flux Portable instrument which uses a non-dispersive infra-red detector for $CO_2$ and $CH_4$. Pore gas samples were collected by attempting to drill a metal pipe to 100 cm depth and gas drawn off for analysis using gas chromatography for oxygen, methane, propane, butane and carbon dioxide. At FA 86 points were sampled for gas flux, with ten points also sampled for pore gas analysis. For LR 20 points were sampled for flux, while ten points were sampled for pore gas however the depth of sampling varied due to the lack of soil at the site.

### 4.3.7 Data Analysis

Collected data were analysed in R (Ver. 4.0.5). Interpolations of the pore gas data were sub-sampled to produce estimates of the pore gas concentrations at each sampling point; these estimates were used as the environmental metadata in the ordination analysis (detailed below). Points below the LoD of the qPCR assays ($\leq$100 copies $\mu l^{-1}$) were treated as zeros in the relative abundances when used for the IDW interpolation. Technical errors are presented and were propagated from two standard deviations (2SD) (Chapter 2.3.7.4). To test correlations between the gas data and the qPCR data a Pearson correlation test was used in R.

Bioinformatic and ecological analysis was completed as described in Section 2.3.9. Exploratory analysis using stacked bar-charts and NMDS was completed. A constrained ordination using a distance-based redundancy analysis (db-RDA) approach was employed. The use of db-RDA enabled absolute abundances to be used so no species were excluded in the ordination. A gap-statistic Tibshirani *et al.* (2001) was used on the Bray-Curtis Dissimilarity Distances (in R) to explore if there were distinct clusters. Soil pore gas concentrations of methane, oxygen and propane were estimated by sub-sampling the interpolation of methane pore gas concentrations (Chapter 2.1.1.1).

## 4.4 Results

### 4.4.1 Inhibition Tests

Three samples from each site were selected to test for PCR inhibition. For LR a 1:2 dilution was selected (mean ± 95% CI: E = 96.1 ± 13%, standard series E = 96.1%, $R^2$ = 0.99). For FA a 1:10 dilution was used (mean ± 95% CI: E = 104 ± 14%, standard series E = 101.7%, $R^2$ = 0.99).

### 4.4.2 Relative Abundance of Methanotrophs

Soil samples underwent DNA extraction and qPCR assays used to estimate gene copy number of *pmoA* and both *mmoX*. Data were normalised and used in interpolation plots, copy number as a percentage and efficiency values are in Table A.1 10.1 and Table A.2). All samples had detectable quantities of both *pmoA* and *mmoX* except a single background site at FA (FA-85) which had a *pmoA* abundance below the limit of detection, but which still had detectable copies of *mmoX*.

The plots of *pmoA* and *mmoX* (not *Methylocella spp. mmoX*) broadly showed similar patterns at FA (Figure 4.3). Both plots showed increases in abundance along the Southern edge of the site near the seep. The observed relative abundance of *pmoA* at FA was greatest 3 m South-East of the main seep (FA-66, $8.6\times10^{-1} \pm 5.0\times10^{-2}$%) (Figure 4.3.A). Background samples, above the LoD, had an average relative abundance of $3.0\times10^{-2} \pm 4.6\times10^{-2}$%. Estimated abundances of *mmoX* (not *Methylocella spp.*) indicated an increased relative abundance at FA1, a point well removed from FA-66. However, FA-66, 70 and 76 did have copies number above the background relative abundance.

The *Methylocella spp. mmoX* qPCR assay suggested a reversal of the trend seen with the other two genes at FA (Figure 4.3.C). Greater relative abundances were observed along the northern and eastern edges of the site, with a hotspot also seen at the background point FA-83 ($1.5x10^{-1} \pm 8.9x10^{-2}$ %). The copy number was below the LoD at FA-84.

At LR the predicted *Methylocella spp.* specific *mmoX* abundance also appeared to be the inverse of the other two genes, (Figure 4.4). Four points, LR-1, -2, -4 and -13, were below the LoD of the assay. The remainder of the sampling points also had low relative abundances. Even the point with the highest copy number, LR-21, was only observed to have a relative abundance of $2.0\times10^{-2} \pm 3.4\times10^{-3}$%.

Disparity was again observed between the hotspots within the *pmoA* and *mmoX* datasets at LR. For *pmoA*, a high relative abundance of 30.8 ± 12.9% was recorded at LR-13; this was driven by the lowest measured 16S rRNA copy number and the highest *pmoA* copy number seen at the site. This led to a bright hotspot at the southern edge of the site that obscures the relatively high (compare to FA) *pmoA* abundance seen across the site; for example, LR-4 in the darkest colour bracket still had a predicted relative abundance of $1.1 \pm 7.3\times10^{-1}$%. The background samples had an average relative abundance of *pmoA* of $8.3\times10^{-3} \pm 2.3\times10^{-4}$ %. For the mmoX assay, two points LR-2 and -18 had similarly high abundances of 1.0 ± 0.31% and 0.9 ± 0.5%, respectively. There was some agreement in the trends seen in *mmoX* and *pmoA*. Both heatmaps show a slightly warmer vertical area passing through LR-2 and -18. The 16S rRNA copy number was significantly lower in the exposed C horizon soil compared to the sample from the forest soil ($1.8\times10^{6}$ versus $6.65x10^{6}$ copies $\mu l^{-1}$: Students' T-test $p$=0.03)

### 4.4.3 Pore Gas Concentrations and Surface Flux

A total of 86 points at FA were sampled for gas flux by BGRM. Ten points also had pore gas samples collected at 100 cm depth. Surface flux was greatest at FA-65 (953.18 $m^{-2}$ $h^{-1}$) aligning well with the pore gas data which also indicated the highest concentrations of methane at FA-65 (81%) (Figure 4.5.A). Fewer points were sampled for propane, butane, and oxygen (Table A.3 Appendix A.2). No points had propane pore gas concentrations above the LoD.

At LR a total of 30 surface flux measurements were made with 10 pore gas sampling points. The landslip at LR reduced the depth of pore gas sampling to between 10 cm and 100 cm depth due to the remaining depth of material. Flux and pore gas concentrations correlarted to each other across the site (Figure 4.6) (Pearson: $r_{(8)}$ =

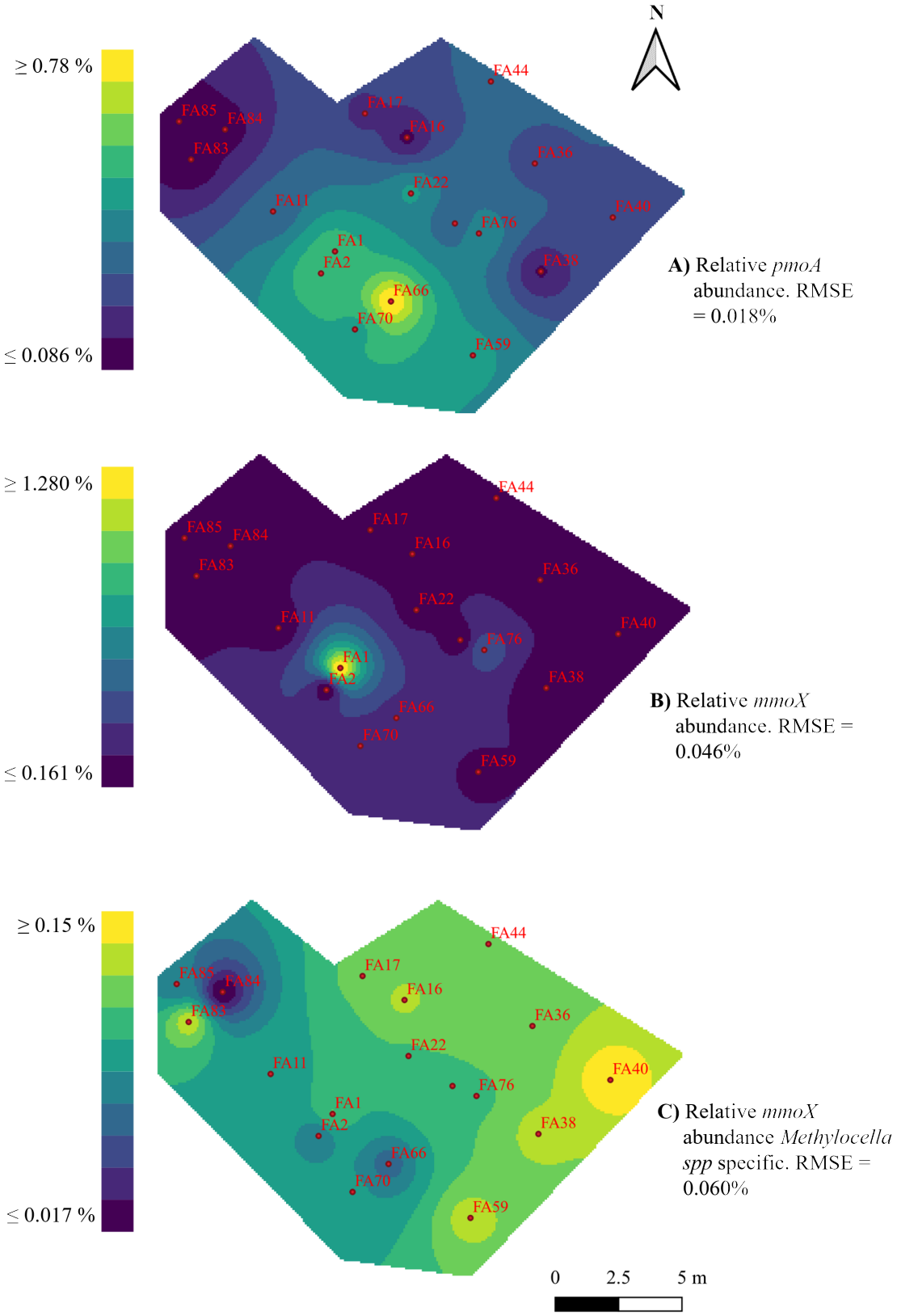



**Figure 4.3: Plots of estimated gene abundances at FA.** Plots generated using an IDW interpolation (p=2) and presented using 10 discrete, equal intervals for ease of interpretation. Points labelled are those sampled for soil and used in PCR.

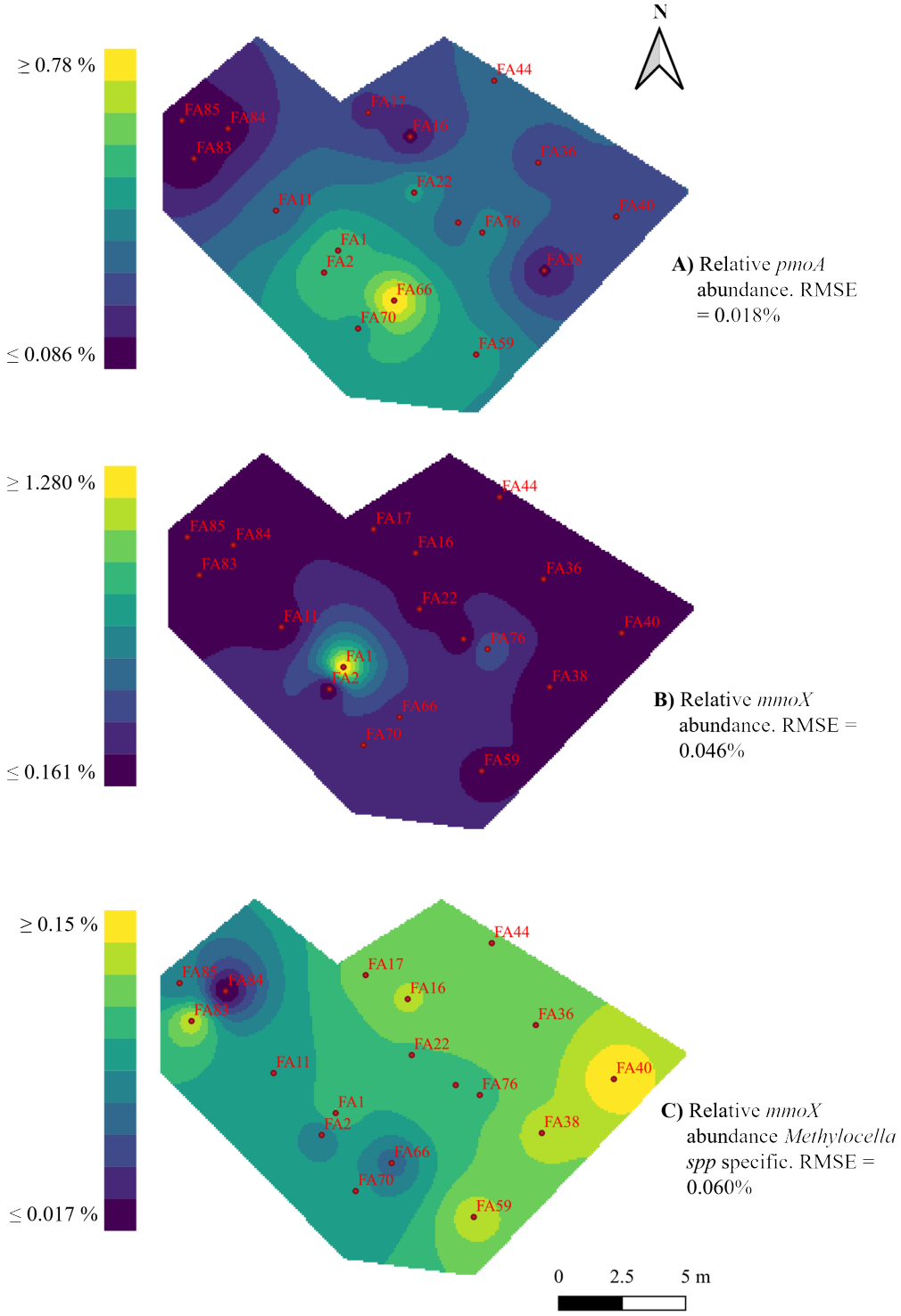



**Figure 4.4: Plots of estimated gene abundances at LR.** Plots generated using an IDW interpolation (p=2) and presented using 10 discrete, equal intervals for ease of interpretation. Points labelled on the plots are those sampled for soil and used in PCR. Background samples are not plotted here given their distance from the main site. The arrow on plot A) illustrates the approximate direction of slope from the seep at LR-1 to the base of the slope at LR-21.

0.98, $p$<0.01), but pore gas sampling was only possible at 10 cm depth. Agreement was seen at the south-west edge of the site between the flux at LR-12 (429.5 g $m^{-2}$ $h^{-1}$) and the pore gas at LR-7 (30% at 80 cm depth). These points were approximately 1.1 m apart. The remainder of the site showed little surface flux or pore gas methane.

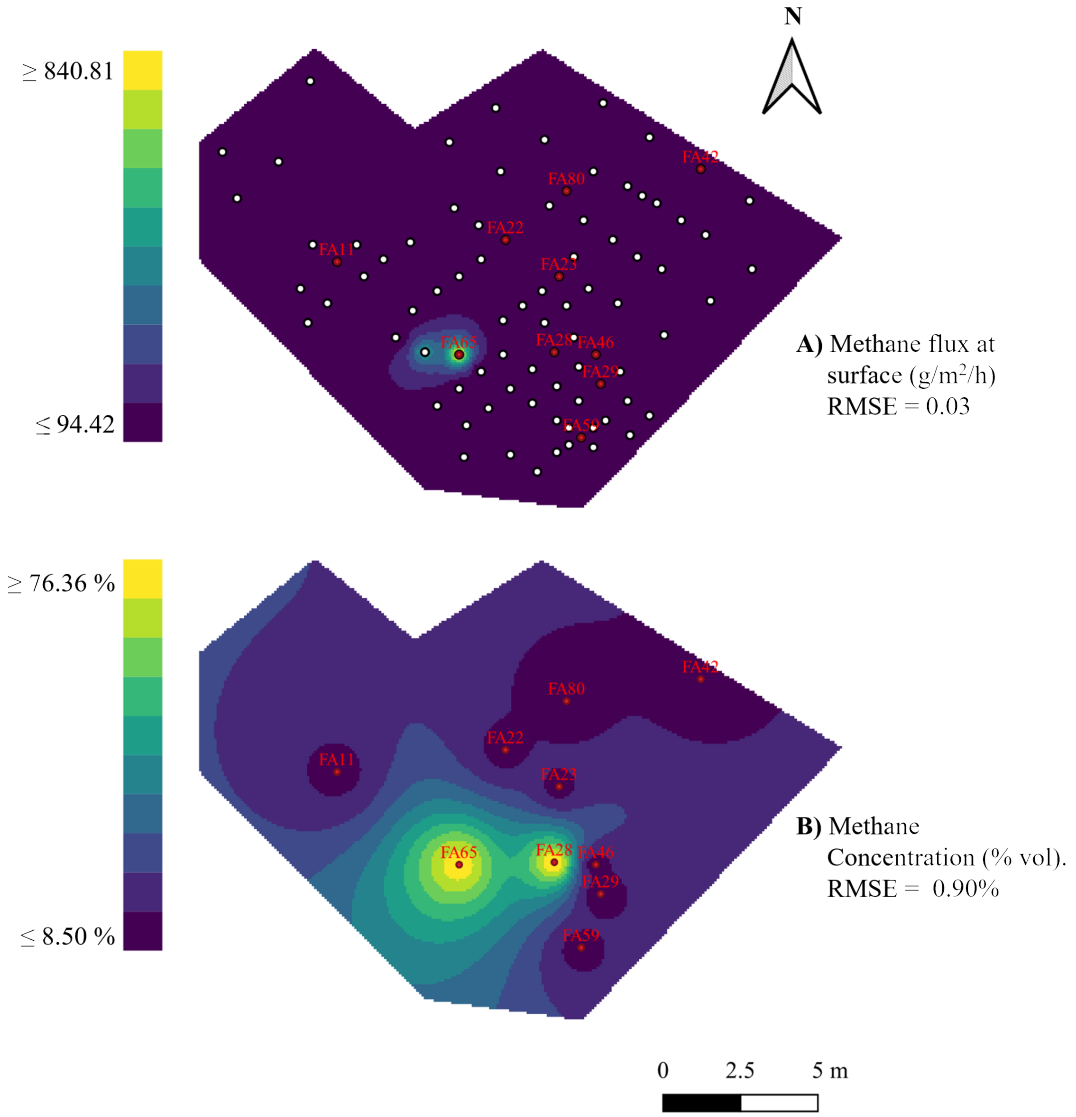



**Figure 4.5: Methane flux and pore gas concentrations at the Fontaine Ardent.** Plots generated using an IDW interpolation (p=2) and presented using 10 discrete, equal intervals for ease of interpretation White points (only shown in A) were sampled for flux only. Red points, (shown in both plots) were used for both pore gas and surface flux measurements.

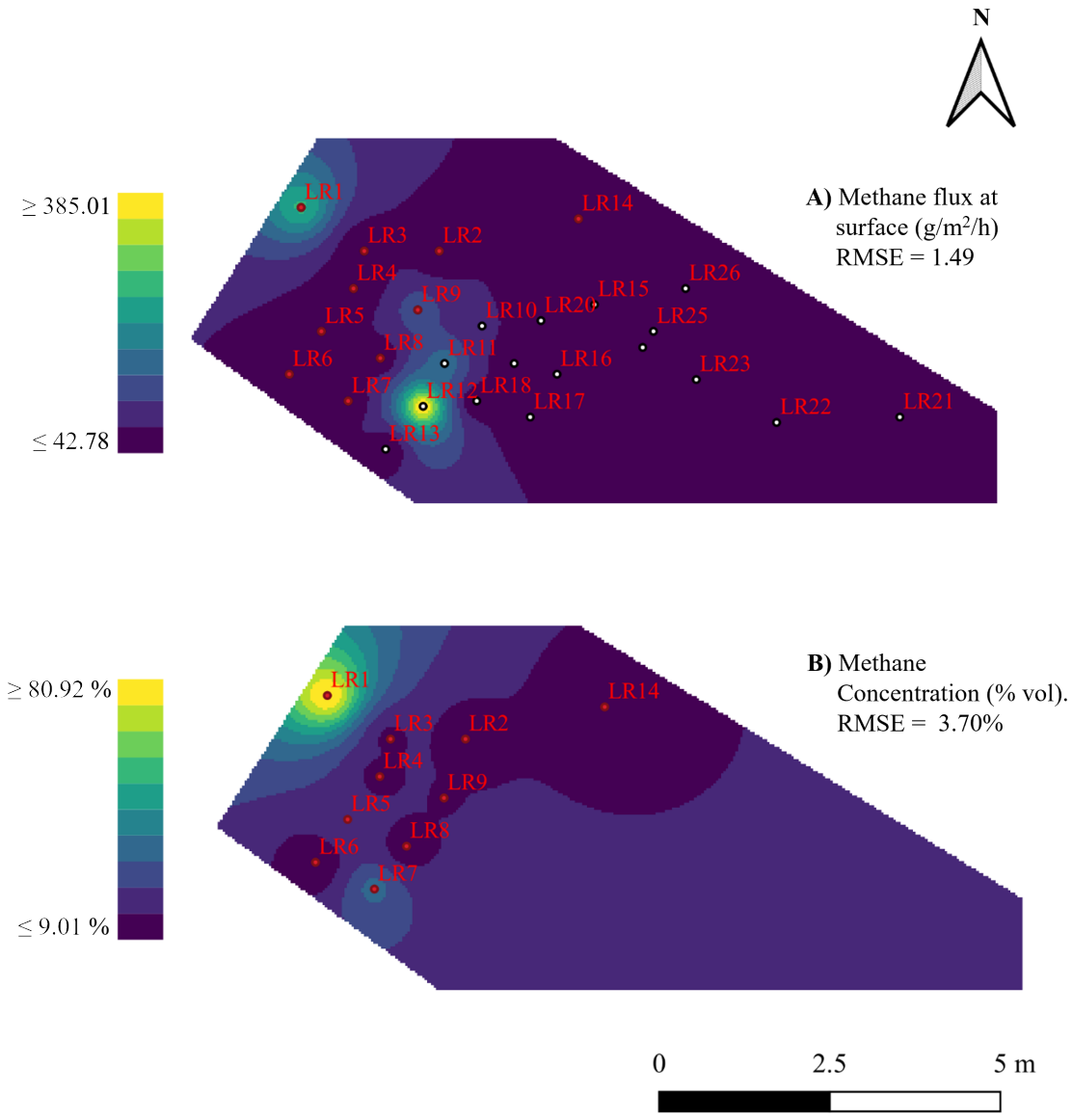



**Figure 4.6: Methane flux and pore gas concentrations at La Rochasson.** Plots generated using an IDW interpolation (p=2) and presented using 10 discrete, equal intervals for ease of interpretation White points (only shown in A) were sampled for flux only. Red points, (shown in both plots) were used for both pore gas and surface flux measurements.

### 4.4.4 Correlations between qPCR and Gas Results

Pearson rank correlation was used to test for a relationship between the relative abundance of functional genes and gas data. At FA the relative abundance of *pmoA* was significantly, positively correlated to $CH_4$ ($r_{(16)}$ = 0.68, $p$<0.05) and $CH_2H_6$ ($r_{(16)}$ = 0.64, $p$<0.05) pore gas concentrations. However, *pmoA* was significantly, negatively correlated to $O_{(2)}$ ($r_{(16)}$ = -0.62, $p$<0.01). No significant relationships were present for *mmoX* and either $CH_4$ or $CH_2H_6$. A significant, positive correlation was present between *Methylocella mmoX* and $O_{(2)}$ ($r_{(16)}$ = 0.49, $p$<0.05). Surface $CH_4$ flux did not significantly correlate with any of the genes. At LR no significant correlations were seen between any genes and pore gas concentrations or surface $CH_4$ flux.

### 4.4.5 Fontaine Ardent – 16S rRNA Sequencing

Twelve samples from FA were used in sequencing analyses. These were selected after qPCR and subsequent interpolation steps. The samples used covered the entire site, included a range of methanotroph abundances and incorporated one background sample. After quality steps and taxonomic alignment, 2697 unique taxa across the 12 samples were identified with between 16605 and 26371 reads for each sample, see Table A.4 (Appendix A.2) for further details on the dataset.

Proteobacteria were the dominant phylum at all sites (Figure 4.7.A). A notable difference between the sites at phylum level was the increase in Firmicutes at points FA-66 and 70, with a relative abundance of 0.36 and 0.34, respectively, versus the site mean at 0.15. At the family level (Figure 4.7.B) sites FA-1, 22, 59, 66, 70, and 76 had greater relative abundances of Beijerinckiaceae, with FA-66 and 70 also showing increase in Clostridiaceae and a loss of Comamondaceae compared to the other sites.

Initial NMDS ordination at the phylum level suggested differences between sample sites were present, (Figure 4.7.C). The `envfit` model suggested that all the abiotic factors, except moisture, had a significant impact upon the ordination's results ($p \leq 0.05$). No differences were clearly seen at the family or methanotroph level using NMDS. Based upon the NMDS results, the metadata for pH, $CH_4$ and $O_2$ concentrations were used in the db-RDA ordinations.

The db-RDA plots are shown in Figure 4.8. At both the phylum (Figure 4.8.A) and family (Figure 4.8.B) levels separation of FA-66 and 70 from the other samples was present along CAP1. This appeared to be related to methane and oxygen concentrations. Also, at the family level, FA-1 and 17 appear removed from the other sam-

ples, possibly due to the impacts of pH and oxygen. Scatter between the remaining samples appeared to be mostly related to the NMDS2 axis which was predicted to account for only a small amount of inter-sample variability. Within the subset of families containing mostly methanotrophs, the constrained ordination (Figure 4.8.C) showed no clear separation of samples being driven by the environmental factors under test (CAP1=20.8%, CAP2=4.8%). Further analysis using a gap statistic supported this with no strong or significant differences seen been *k*-values. Gap statistics and a clustering dendrogram are included in Figure A.1 (Appendix A.2). With all the db-RDA ordinations, when plotted on a scree plot (data not shown) the first unconstrained axes had greater eigenvalues than the second and third constrained axes.

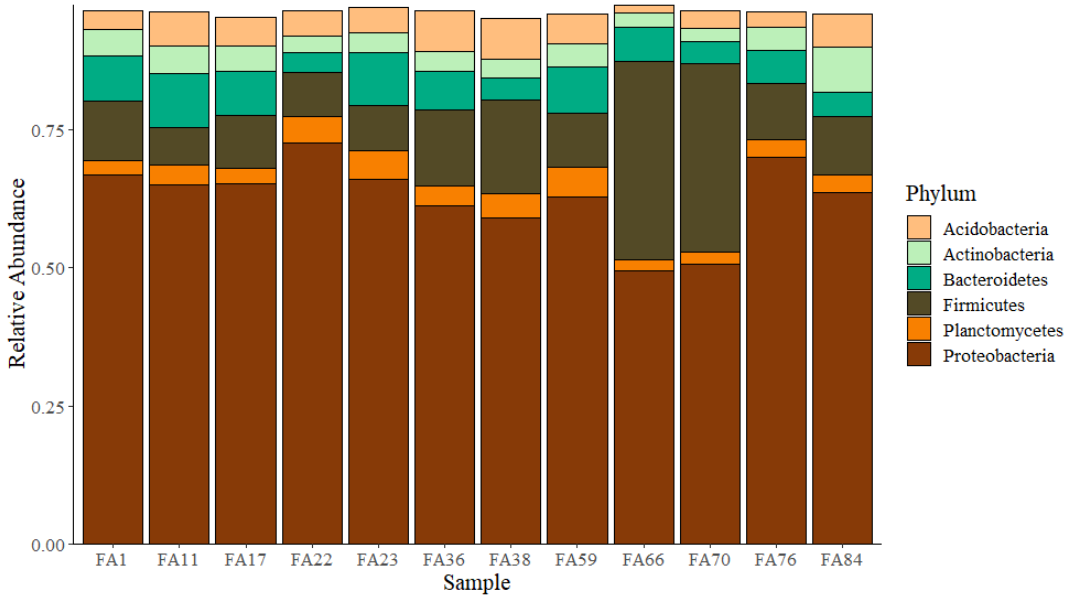



Fig:4.7.A.Bar plot of key phyla. Samples FA-66 and 70 appear to have a greater relative abundance of Firmicutes compared with other samples.

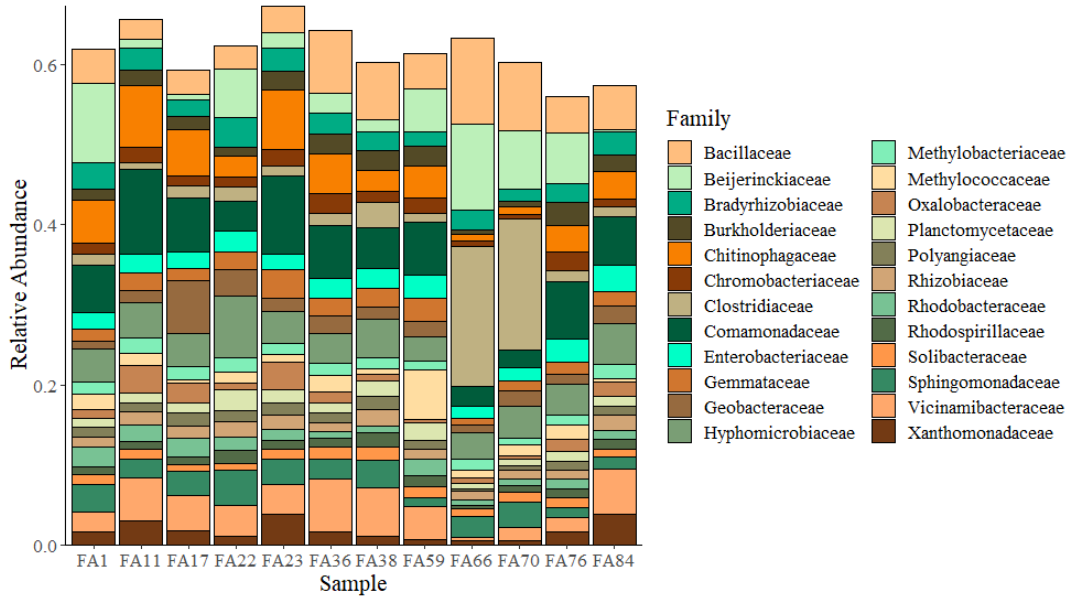



Fig:4.7.B Bar plot of the key families present in the samples. Samples F66 and 70 have notable increases in Clostridiaceae and decreases in Comamonadaceae.

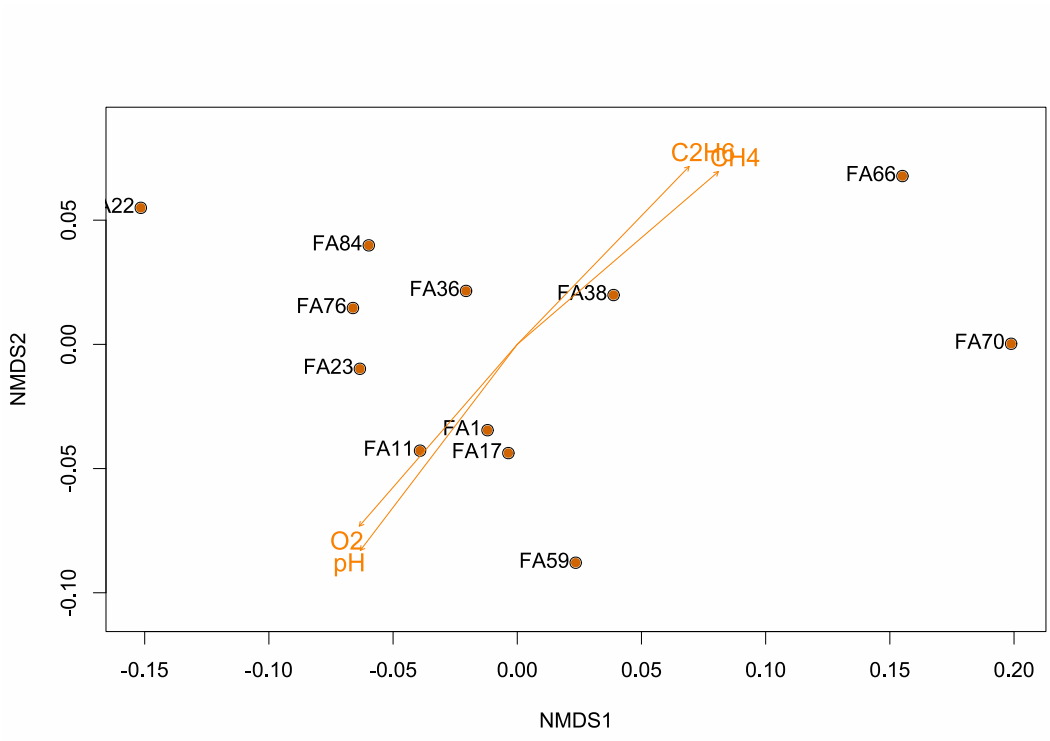
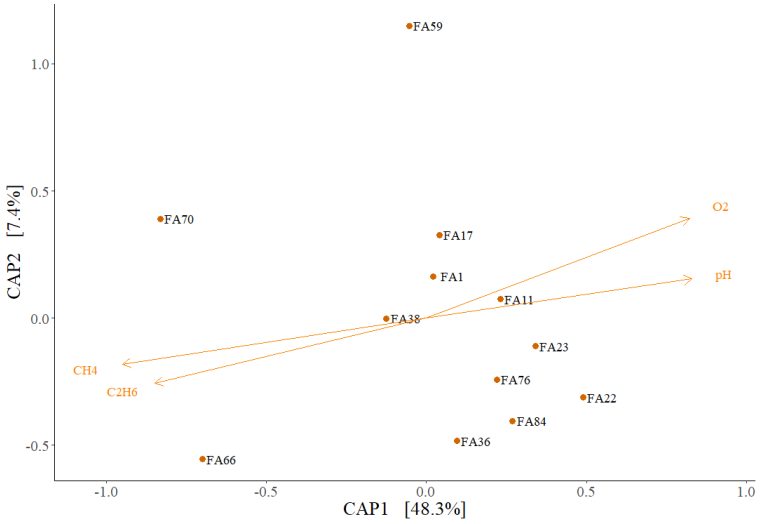
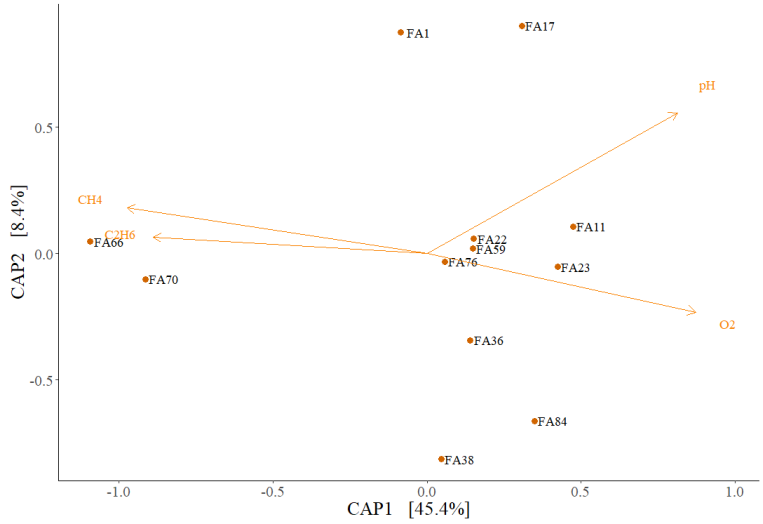



Fig:4.7.C NMDS analysis at the phylum level; environmental variables have been modelled relative to the ordination axes and plotted as vectors if $p<0.05$.

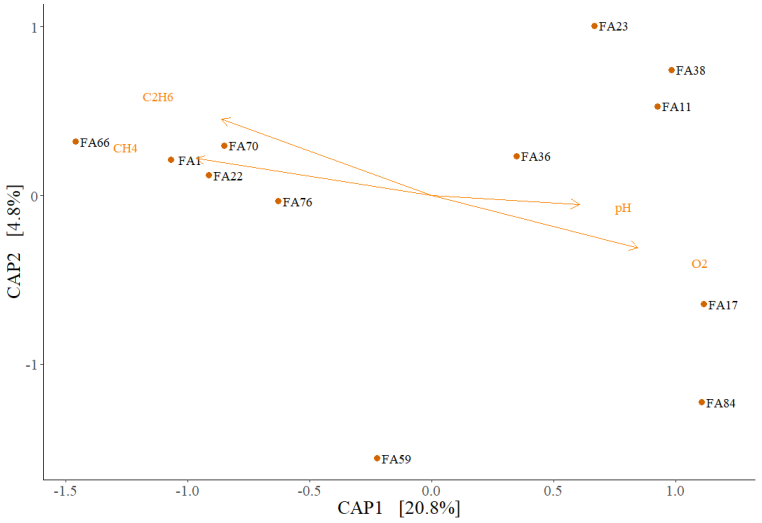
**Figure 4.7: Exploratory analysis of 16S rRNA sequencing data.** Plots use relative abundances, with taxa above $1x10^{-2}$ abundance included.




(a)




(b)




(c)

**Figure 4.8: db-RDA ordination plots using WA and Scaling 1.** Ordinations used absolute abundances, with a Bray-Curtis dissimilarity index. Percentages in square brackets represent the amount of variation explained by that axis.
**A)** Ordination at the phylum level. FA-66 & 70 appear to be separate from the main group (and each other) possibly due to changes in $CH_4$ and $O_2$.
**B)** Ordination at the family level. FA-66 & 70 are slightly removed, as are FA-1 & 17.
**C)** Ordination using families with known methanotrophs. No clear separation of samples was seen. Constrained variables appear to be poorly related to community variation with only a small, predicted impact.

## 4.5 Discussion

Two natural gas macro-seeps were sampled to compare the relative abundance of methanotrophs in response to subsurface methane fluxes. Both sites had surface fluxes of natural gas that were capable of being ignited. The use of qPCR was successful in elucidating relative abundance across both sites. This qPCR data and the gas dataset were explored for spatial patterns using interpolation.

At both sites high concentrations of methane flux were observed, at FA the maximum flux was within the ranges previously reported, the greatest point flux being 130 kg/m$^2$/day (5420 g/m$^2$/hr) (Gal *et al.*, 2017) however more recent studies have estimated a much lower maximum flux (17 g/m$^2$/hr) (Gal *et al.*, 2018) which might be due to the previously reported pulsing behaviour where flux at the site appears to rise and fall (Gal *et al.*, 2018). For LR the $CH_4$ flux was around double the maximum previously reported (Gal *et al.*, 2017). Methane pore gas concentrations were similar to previously reported values for both sites (Gal *et al.*, 2017).

### 4.5.1 Methane and Methanotroph Distribution

At both sites the highest pore methane concentration exceeded 80% by volume (Figure 4.4 and Figure 4.6). At FA, whilst there was general agreement between the pore gas and flux measurements, for example at FA-65, there was a notable disparity between the spatial size of the methane anomaly at depth versus the surface flux. This could be due to the lack of sampling points at depth enabling the high methane signal at FA-65 and 28 to influence a large area of the plot. Another potential explanation is the high relative abundance of *pmoA* which correlated with the increased methane concentration, particularly at FA-65 and across the south-western portion of the site. The methanotrophic community is potentially oxidising the methane before it reaches the surface therefore masking the high concentrations at depth.

The gas measurements at LR were more difficult to interpret and compare to the qPCR data. Varying depths of gas and soil sampling made it difficult to understand if a similar relationship existed between methane and methanotrophs. Similarities between surface flux and pore gas still existed, but, for example at LR-1, this can hardly be considered surprising given the pore gas was sampled at only 10 cm depth due to the lack of material above the bedrock.

The relative abundance of *pmoA* was high across the LR site (interpolated range = $5.2\times10^{-2}$ - $3.1\times10^{1}$ %). This is higher than abundances reported by other authors ex-

ploring sites above landfill and hydrocarbon reservoirs (using the same primer sets). Values up to 1.3×108 copies $g^{-1}$ dw at LR compared to $2.7\times10^6$ (Zhang *et al.*, 2017) or $5.6\times10^4$ (Liu *et al.*, 2016). Values from qPCR not normalised should be used with caution but it is interesting that LR had such high concentrations of *pmoA*.

Indeed, at LR-13 the relative abundance of *pmoA* was so high it could be thought to be erroneous (measured relative abundance 30.75±12.89%). However, the landslip had, in effect, reset the soil to an early stage of development where the 'soil' was little more than a weathered rock. There is surprisingly little published work that explores the microbiology of sites recovering from landslip incidents but Singh *et al.* (2001) observed that younger landslip sites had significantly lower microbial biomass and a lower nutrient status. This is mirrored at LR, where the mean 16S rRNA copy number in the exposed C horizon samples was significantly lower than the undisturbed forest soil samples. Again, these raw qPCR values should be used with caution. At LR, the area sampled had yet to be recolonized by plants and there was no evidence of moss or lichen growth. It is likely therefore, that carbon sources were limited to either the methane emitting from the seep, or the recalcitrant shale/mudstone material; methanotrophs could therefore thrive compared to other generalist microbial guilds.

It is not known why *mmoX* at LR does not show similar high abundances. Indeed, at both sites the differing patterns in abundance of *mmoX* and *pmoA* is surprising. While there is some broad agreement in the distribution of these genes (Figure 4.4), there are distinct hotspots for *mmoX* and *pmoA*. All known obligate methanotrophs have pMMO and a subset of these also possess sMMO (McDonald *et al.*, 2008; Hanson and Hanson, 1996). The switch from pMMO to sMMO is mediated by low copper concentrations (Semrau *et al.*, 2013; Csáki *et al.*, 2003). Taking FA-1 as an example, if this point had low copper concentration which limited pMMO use, methanotrophs capable of expressing sMMO would feasibly have an advantage. It is possible that the hotspot at FA-1 is being generated simply by the presence of more genera that have both genes due to a lower copper concentration. That said, without having measured copper concentrations and with a lack of high resolution 16S rRNA sequencing data it is impossible to state this hypothesis with certainty.

Similarly, the lack of agreement between *Methylocella*-specific *mmoX* and the other MMOs at both sites was unexpected. The qPCR data suggested this gene was marginally more prevalent at points removed from the main seep and, by extension,at points with high *pmoA* and *mmoX* (Figure 4.3 and Figure 4.4). Previous work has suggested that *Methylocella spp*. are capable of growth in a range of soil habitats and are found

at sites with both biogenic and thermogenic methane seeps (Farhan Ul Haque *et al.*, 2018; Rahman *et al.*, 2011a; Dedysh *et al.*, 2000). The results here supported prior work, with *Methylocella spp.* detected at both sites.

### 4.5.2 Fontaine Ardent Community Analysis

The 16S rRNA sequencing and subsequent ordinations revealed that the community composition varied across the site. These shifts in composition were related to changes in $CH_4$, $C_2H_6$, $O_2$ and pH. However, the first unconstrained axes of db-RDA had higher eigenvalues the than the second, third and fourth constrained axes suggesting that other factors, not measured, were impacting community composition. The inclusion of the NMDS ordination (Figure 4.7), suggests that similarities are still present when the ordination is unconstrained with FA-66 and 70 still separate at the phylum level.

Sampling points with a strong $CH_4$ and $C_2H_6$ pore gas signal (FA-66 and 70) were separated from the main group at both the Phylum and Family level (Figure 4.8). Differences at these two points, compared to other points, included increases in the relative abundance of Firmicutes seemingly driven by the increased presence of Clostridiaceae (Figure 4.7). Other changes at family level included increases in Beijerinckiaceae.

The lowest $O_2$ concentrations were predicted at FA-66 and 70 (5.6% and 8.9% respectively). Therefore, the increase in Clostridiaceae is potentially unsurprising given the number of obligate anaerobes within the family. The lower $O_2$ could be occurring due to either respiration or due to displacement by the seep gases as reported by Gal *et al.* (2018). The presence of increased $C_2H_6$ suggests that a seep, at least partially of thermogenic origin, is present. It is plausible that high methanotrophic activity has contributed to the low $O_2$ concentration. Methanotrophs have been observed to oxidise methane at $O_2$ concentrations as low as 1.1 ± 0.3% volume (with 0.08 ± 0.002% $CH_4$) (Henckel *et al.*, 2000). This observation is supported by Ren *et al.* (1997) who studied single strains. However, it should be stressed that in soil systems the rate of gas diffusion is vital, and this is impacted by soil structure and moisture content, the soil at FA was a clayey sand, and despite the presence of gravel this might have limit gas diffusion. Limited $O_2$ diffusion will limit the rate of methanotrophy; indeed Gebert *et al.* (2011) have argued that ingress of $O_2$ was more important than other soil physical factors such as pH.

The apparent lack of interaction between moisture and community distribution is therefore surprising. This is probably due to the way moisture analysis was completed. Here, the surface moisture was measured, not the moisture content of the sample from

which the DNA was extracted. Plus, measuring the moisture content at a single point in time is not necessarily representative of the moisture regime. For example, this site is known to flood. It is likely that the moisture measurements here are not suitable for an in-depth discussion and comparison with the predicted soil community.

The predicted impact of pH on community composition is unsurprising. Other authors have reported the impact of soil pH on bacterial diversity (Griffiths *et al.*, 2011) and the methanotrophic community composition (Knief *et al.*, 2003). Soil pH is frequently described as the 'master variable' with impacts soil chemistry and therefore biology. Overall, the site's pH was neutral to mildly alkaline with no outliers present (defined using 1.5 multiplied by the interquartile range). That said, FA-66 and 70 had the lowest pH (7.22 and 7.33 respectively) although the cause of this is hard to determine; if the seep was carrying a high carbon dioxide component this could have dissolved into the soil pore water forming carbonic acid resulting in a lower pH.

Attempts were made to explore the methanotrophic community at each sampling point. Three families known to contain methanotrophs (Methylocystaceae, Methylococcaceae and Beijerinckiaceae) were grouped and used in an ordination. The Beijerinckiaceae is the only family that does not contain wholly methanotrophs. Of the Beijerinckiaceae's six genera, it has three capable of methane oxidation, viz. *Methylocella, Methyloferula* and *Methylocapsa*, with the latter two being described as obligate methanotrophs (Dedysh *et al.*, 2016). Interestingly, the constrained ordination using this data sub-set did not suggest the FA-66 and 70 composition was different from the other sites (Figure 4.8.C). This was supported by an unconstrained NMDS ordination that also found no significant relationships between composition and the measured abiotic variables and no clear separation of samples. The gap-statistic suggested only one cluster was likely. Cumulatively this would indicate that, although those communities may have been in greater abundance at some points, the composition of methanotrophs, at a family level, did not alter significantly. It also potentially suggests that shifts in community composition are more likely to be on a continuous scale rather than a distinct shifts in the presence of methane flux.

Overall community composition was related to soil gases and pH. The exact mechanisms cannot be elucidated here, but increased $CH_4$ and $C_2H_6$ was accompanied by lower $O_2$ and pH. These abiotic factors were accompanied by shifts in the community composition at a phylum and family level. That said, there were no clear changes seen in the composition of the methanotrophic guild between sites. This is surprising when compared to the qPCR data; if *pmoA* relative abundances have increased compared to

*Methylocella mmoX* a shift within the community composition would have been seen. However, an increase in non-methanotrophic members of the Beijerinckiaceae might have masked this.

## 4.6 Conclusion

Work presented here explores the relationship between soil microbiology and alkane gas flux. Increases in obligate methanotrophs, detected using qPCR assays of the *pmoA* and *mmoX* genes, were linked to areas of increased pore gas methane concentrations at the Fontaine Ardent, supporting the hypothesis that methane fluxes and the relative abundances of methanotrophs are related and therefore could indicate where increased methane fluxes are present. Weaker relationships were seen at La Rochasson but the site was harder to interpret due to the loss of the overlying soil and the high relative abundance of *pmoA* across the site. Interactions between obligate and facultative methanotrophs remain difficult to explain; work presented here appears to suggest that *Methylocella spp*. are present at low abundances in soils while the obligate methanotrophs thrive or bloom in the presence of high methane concentrations. Sequencing data using 16S rRNA from the FA samples appears to support the qPCR results with shifts in community composition related to pore gases and pH. Interestingly, there was little change in the composition of the methanotrophic community across FA, suggesting that community composition remains stable.

## 4.7 Implications and Future Work

The data presented here supported the core hypothesises of the project. Firstly, variation in the relative abundances of methanotrophs can be detected across the spatial scales explored here using qPCR. Secondly, increases in the relative abundances of methanotrophs are potentially related to increases in natural gas flux. The use of qPCR to resolve spatial differences, and the ease of its data interpretation make it the more attractive technique for this style of surveying compared to using culture-based techniques. The work has highlighted some interesting questions on the interaction between facultative and obligate methanotroph groups that need further exploration. It also questions the potential use of *Methylocella spp*, as an indicator species of a thermogenic seep. It would appear that the use of ordination to explore shifts in community composition at these sites is the better tool when compared to cluster analysis as community changes did not appear to be in discontinuous.

# Chapter 5

# Field Survey of Methanotrophs Around a Decommissioned Oil Well at Dukes Wood

## 5.1 Rationale

A second field site was sought to further explore the relationship between natural gas fluxes and methanotrophs. An English oilfield, no longer in operation, was selected. Work by Boothroyd *et al.* (2015) had suggested elevated gas fluxes of methane might be present at some points in the oilfield. Furthermore, the original 'nodding donkeys' were still *in situ* enabling the easy location of well heads. Before the disruption caused by the Covid-19 pandemic, this work was designed to have a seasonal element as well, allowing the impact of seasonal temperature and changing water levels at the site to be explored. Unfortunately, due to travel restrictions and other risks this had to be abandoned. Instead, only one complete survey was undertaken, in February 2021 and the wet conditions made it impossible to collect pore gas samples. Despite these drawbacks the site still offered an excellent exploration of a UK oilfield and its soil microbiology.

## 5.2 Introduction

Methane is a potent greenhouse gas with a radiative forcing effect 25 times that of $CO_2$ (Rodhe, 1990). Human methane emissions are currently estimated to be around 360 Tg $yr^{-1}$, with fossil fuel production and transport accounting for approximately one third (Saunois *et al.*, 2020). These estimates include fugitive emissions, from sources such as leaking and damaged infrastructure involved in the abstraction and transport of fossil fuels.

The UK has a long legacy of oil and gas extraction. There are 2249 onshore oil and gas wells in the UK (UK Oil and Gas Authority, 2019) most now abandoned. Currently, there is no regulatory framework for the monitoring of these wells. Over time well infrastructure can degrade (Lackey and Rajaram, 2018; Watson and Bachu, 2009) and it is possible that the wells become a source of methane emission. In effect, allowing methane from the old reservoir to migrate toward the surface.

The Eakring oilfield, in North Nottinghamshire, was explored and exploited through the second world war and into the 1970s. The site overlies the Mercia mudstone group (Keuper marl), with with Mercia siltstones (Keuper waterstones) also outcropping within 500 m of the site (British Geological Survey, 1996). Today, the site at Dukes Wood maintains some surface infrastructure as a local attraction. Given the age of the decommissioned wells and the presence of the original pumpjacks the site offers an excellent opportunity to survey a UK site where increased natural gas fluxes through soils might be present.

Methane flux can be detected using direct methods where sampling attempts to measure the $CH_4$ flux or concentration. Indirect methods can also be used, where an indicator of methane's flux or concentration is used. Changes in soil microbiology is one method of indirectly measuring $CH_4$ in soils. Modern molecular techniques offer the ability to accurately assess the size of methanotroph populations in soils. Microbial indicators, detected using PCR based techniques, have already shown promise in distinguishing soils above hydrocarbon reservoirs (Zhang *et al.*, 2017, 2014; Miqueletto *et al.*, 2011). Previous work has used soil samples from 50 cm (Zhang *et al.*, 2017, 2014) and 70 cm (Miqueletto *et al.*, 2011). Sampling microbiology at greater depths might increase the sensitivity as $CH_4$ concentrations from a sub-surface seep should be higher while atmospheric diffusion from the atmosphere should be limited. However, with increasing depth available $O_2$ from the atmosphere is also expected to to decrease, thus limiting aerobic processes.

An survey of the soil methanotrophs around a well head at Dukes Wood (SK 67854 30163), Eakring was made. This set out to test the relationship between methane flux and methanotrophs in soil. Here, the depth of sampling was varied to assess if shallow and deep samples would detect different anomalies in the methanotrophic population. Soil $CH_4$ flux was attempted to compare and contrast with the microbial results.

### 5.2.1 Chapter Objectives

Cumulatively this Chapter aimed to test:

- That shifts in soil methanotrophic communities can be detected over sites only tens of metres across using qPCR and amplicon sequencing.
- That points within sites with increased methane flux will have a corresponding increase in methanotrophs.
- Increasing depth of soil microbial sampling will have impacts on the spatial size of any microbial community anomaly. With increasing soil depth improving the sensitivity of the survey approach by increasing the likelihood of detecting methane fluxes from a geological source.

### 5.2.2 Preliminary Work

Eakring was in part selected due to previous work by (Boothroyd *et al.*, 2015). Their original data reports two well locations at Eakring that had an increased relative flux of methane compared to the background. In October 2019, a preliminary survey was completed, this attempted to locate the Eakring 194 well in Dukes Wood. Due to poor GPS reception and dense vegetation this well was not found. Instead, two other wells at Dukes Wood were surveyed (Eakring-55 and Eakring-144) which still had the pumpjack present enabling their easy location. Wells and their relative locations are in Figure 5.1.

Briefly, a 25 m transect was sampled. Sampling points were 5 m apart, with the well-head located at 5m into the transect. Samples were collected at a depth of 15-20 cm. Obligate methanotrophy genes, *pmoA* and *mmoX* were estimated using qPCR at each point. No methane flux was detected along the transect, possibly due to the high water and clay content of the soil. Eakring 144 appeared to have an increased pmoA abundance immediately adjacent to the well with a decline potentially being present as distance from the well increased, as shown in Figure 5.2. It was decided that this well

would be further surveyed to assess methane flux and methanotroph abundance.

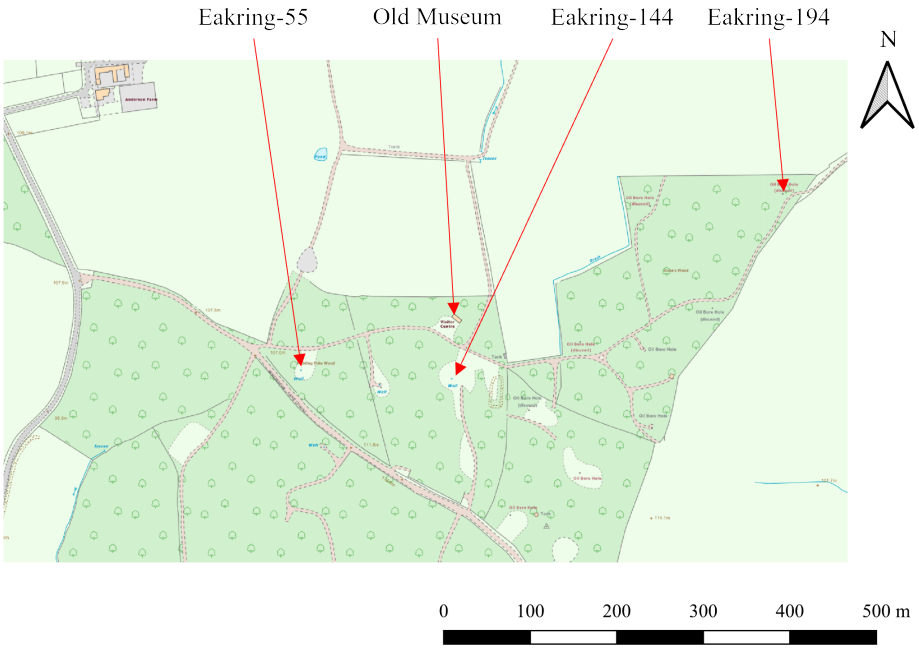



**Figure 5.1: Overview of the Dukes Wood Site.** The map covers 23 wells as reported by UKOGL (2018). Map centre is approximately SK 6788,6023. Well names are those used by the well operator. Basemap: OS MasterMap® Topography Layer, 1:1000. Downloaded from EDNIA DIGIMAP, 22 November 2019.

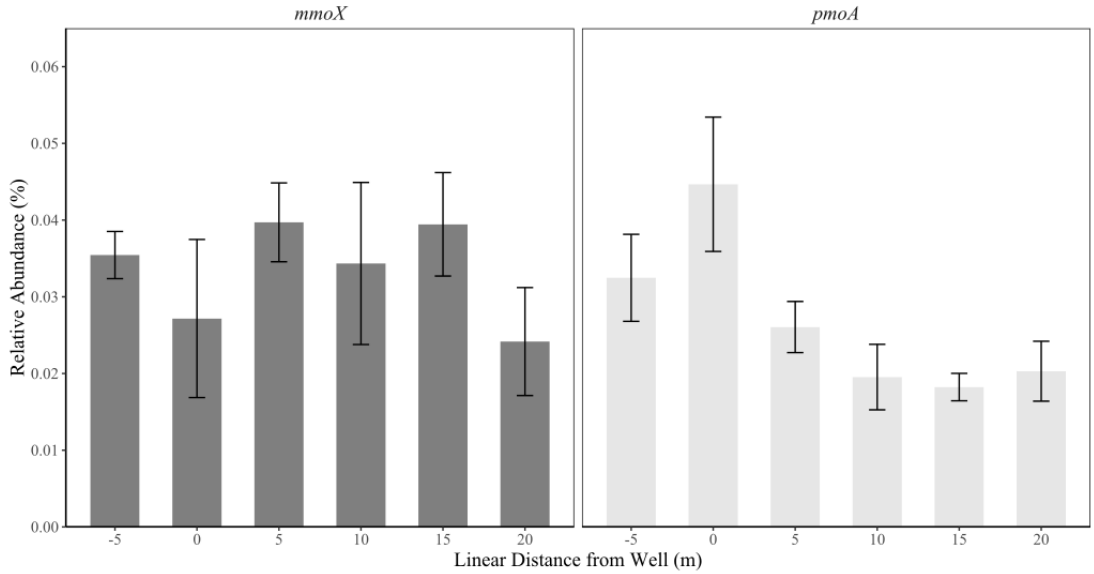



**Figure 5.2: Relative abundance of methane mono-oxygenase genes along a transect at Eakring 144.** The well was located at point 0. Error is the propagated error of technical repeats, using 2SD.

## 5.3 Methods

One well (Eakring 144) at Dukes Wood in the Eakring Oil Field, Nottingham UK, was surveyed for methanotroph abundance and methane flux. Fieldwork was completed in February 2021.

### 5.3.1 Survey Design

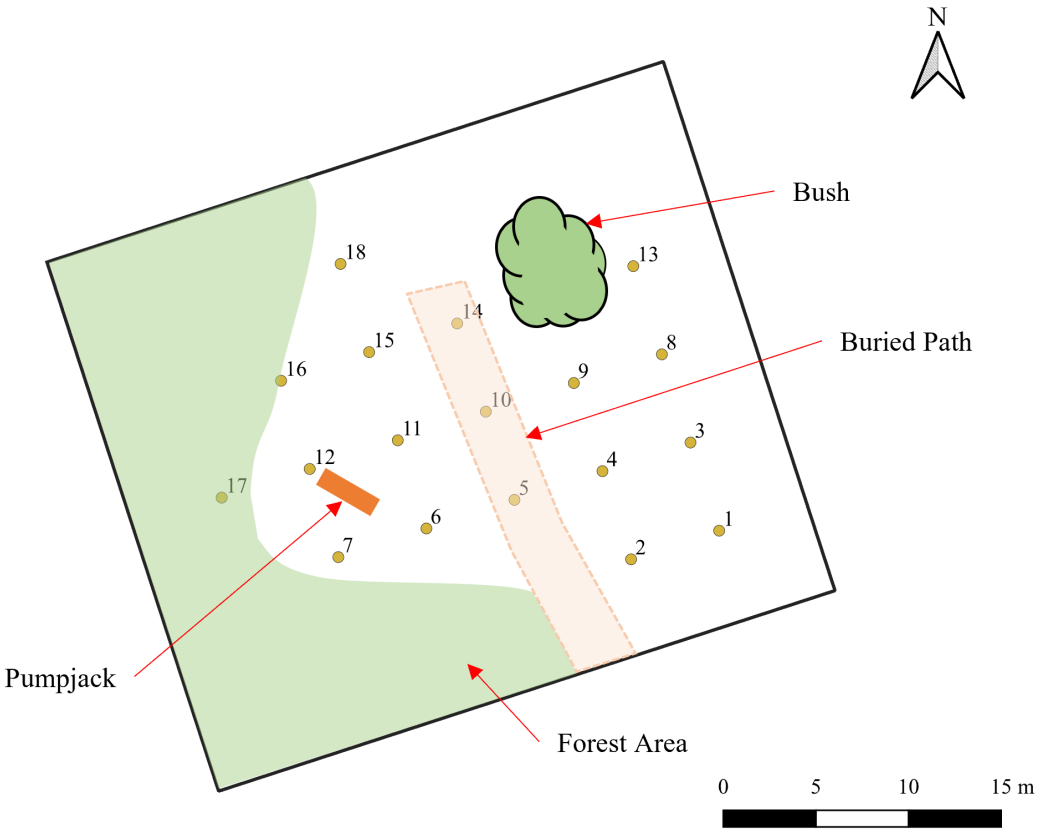



**Figure 5.3: Plan sketch of the survey area around Eakring 144.** A 5 m grid spacing was used. Point 12 was immediately next to the pumpjack at the top of the well. The site was bounded by woodland. What is thought to be a buried track was present under points 5, 10, 14. A bushy shrub stopped one sample being collected.

The well Eakring 144 was surveyed, a 5 m grid was set out, offset from magnetic North by 72°(Figure 5.3). All points were sampled for soil at 20 cm and at 80 cm depth. Samples were collected using an auger. Before augering, points were assessed for gas flux (Chapter 2.1). The well was situated in grassland surrounded by woods. A site picture is in Figure A.2 (Appendix A.3) The site sloped in a North-Westerly direction and was heavily waterlogged. Surface soil was a dense, red brown, clay rich material, this contained a few, small stones. Deeper samples were almost entirely clay with a

reddish matrix alongside grey to green streaks or layers present in samples suggesting the presence of reducing conditions.

### 5.3.2 Microbial Community

Collected soil was stored at 4°C for 24 hrs before used in DNA extractions. Soil from the two depths was handled separately, each sample was mixed aseptically and 0.3-0.4 g of soil sub-sampled, this was stored at -80°C until extraction. DNA extractions and inhibition testing were completed using the methods laid out in Chapter 2.3.2 and 3.6 respectively. Two methanotrophy genes were quantified using qPCR, *pmoA*, *mmoX*, and *Methylocella spp. mmoX,* with their relative abundance estimated using qPCR of the 16S rRNA gene. Also completed was sequencing of the 16S rRNA gene, using the ONT Flongle platform (Chapter 2.3.9). A subset of six samples from both soil depths (totalling 12 samples), attempting to follow the transect from the preliminary work 5.2.2, were used.

#### 5.3.2.1 Inhibition Tests

Two sets of inhibition tests were completed, one for the shallow samples and a second for the deeper samples. The design of the normal inhibition test does allow for the dilution grouping to predict efficiency from using 1:20 as a start point. Shallow samples were diluted 1:20 with an estimated E = 104.1% $R^2$ = 0.99 (standard series E = 109.7%, $R^2$ = 0.99). Deeper samples were diluted 1:25, A one in 1:20 dilution may have been sufficient, E = 102.2%, $R^2$ = 0.99 (standard series E = 103.0%, $R^2$ = 0.99), but a cautious approach was taken given the exceptionally high clay content of the starting material.

### 5.3.3 Soil Flux

At each point, prior to augering, soil $CO_2$ and $CH_4$ flux was estimated using a portable gas fluxmeter (West Systems) (Chapter 2.1.1). Sampling was completed over three minutes, using the accumulation of $CO_2$ to ensure adequate data were collected.

### 5.3.4 Data Analysis

Collected data were analysed in RStudio (Ver 1.4.1106) using R (Ver. 4.0.5). The qPCR and abiotic soil data from both depths were interpolated (Chapter 2.1.1). This enabled a visual comparison of gas flux to changes in the relative abundance of the three genes.

Bioinformatics and ecology data was handled using the pipeline previously described (Chapter 2.3.9). Statistical analysis was completed in R with the data grouped at Phyla and Family taxa levels. The methanotrophic guild was explored using the three families known to contain methanotrophs, Methylocystaceae, Methylococcaceae and Beijerinckiaceae.

Exploratory analysis was carried out using stacked bar charts and Non-Metric Multi-Dimensional (NMDS) ordination using Bray-Curtis dissimilarity measures. Abiotic soil factors were modelled onto the ordination using the envfit function in the vegan package. To test if the communities from the two depths were dissimilar an analysis of similarity (ANOSIM) was completed using Bray-Curtis distances and 9999 permutations, with significance accepted if $p \leq 0.05$.

## 5.4 Results

### 5.4.1 Soil Abiotic Factors

No methane flux was detected across the site. Carbon dioxide flux at the soil surface was detected and estimated at every sampling point. Estimated $CO_2$ flux was used an IDW interpolation to estimate flux across the site (Figure 5.4). The pH and moisture content were measured at every sampling point at both sampling depths. Moisture and pH observations were used in IDW interpolations (Figure 5.5). Interestingly the deeper samples had a lower mean percent moisture content (23.4%, range: 19.4-27.0%) compared to the surface samples (mean=31.2%, range: 23.4-52.6%). The deeper samples had a mean higher pH (8.23) than the shallow samples (7.54).

Weak relationships appeared to be present between high $CO_2$ flux, lower moisture areas and high pH through a diagonal running NE/SW, through points 7, 11 and 14. In the shallow samples a significant negative correlation was present between pH and moisture (Pearson coefficient = -0.54, $p<0.05$). No other significant correlations existed between $CO_2$ flux, pH and moisture either within the deep samples or between the deep and shallow samples.

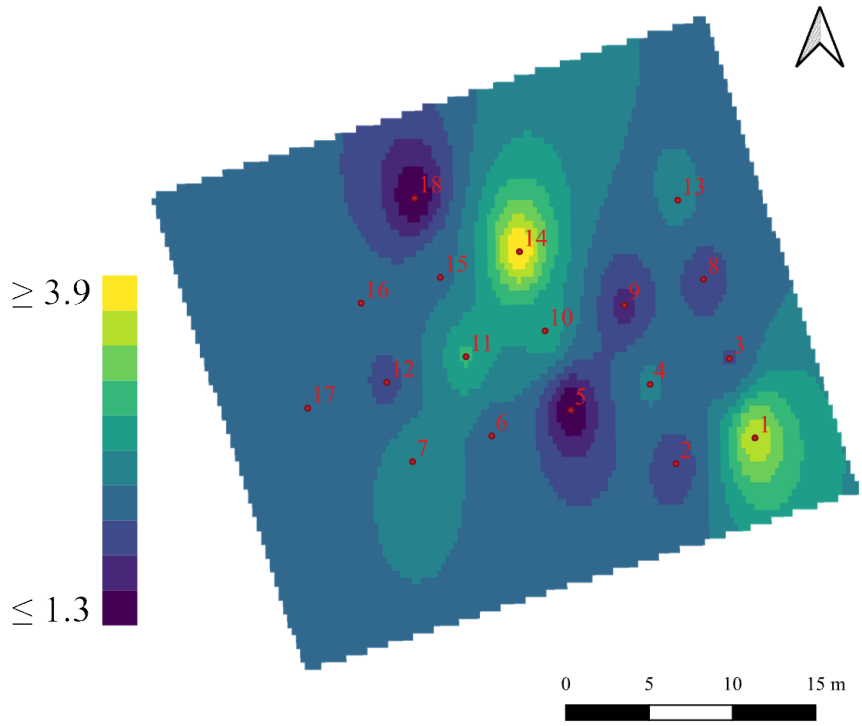



**Figure 5.4: Estimated $CO_2$ flux across the site, produced using an IDW interpolation**. Flux is shown in g $m^{-2}$ $h^{-1}$. P=2 and RMSE = 0.024. Expected source is at Point 12.

### 5.4.2 Distribution of Methanotrophs

Three methanotrophy genes, *pmoA, mmoX* and a *Methylocella spp. mmoX* were quantified using qPCR. Using the 16s rRNA gene copy number, also quantified using qPCR, as an estimate for the total community size, these methanotrophy genes are presented as percent relative abundance of the bacterial community as a whole. Efficiency and $R^2$ values for the qPCR assays are in Table 5.1.

**Table 5.1: Efficiency and $R^2$ values for the qPCR assays calculated from the slope of the standards.** Assays were deemed a success only if the efficiency value was between 90-110% and if the $R^2 \geq 0.95$.

| | | **E-Value** | **$R^2$** |
|---|---|---|---|
| Shallow (-20 cm) Samples | 16S rRNA | 103% | 0.99 |
| | *pmoA* | 102% | 0.99 |
| | *mmoX* | 93% | 0.99 |
| | *Methylocella mmoX* | 92% | 0.99 |
| Deep (-80 cm) Samples | 16S rRNA | 103% | 0.99 |
| | *pmoA* | 104% | 0.99 |
| | *mmoX* | 91% | 0.99 |
| | *Methylocella mmoX* | 99% | 0.99 |

The qPCR assays failed to amplify *mmoX* and *pmoA* in the majority of the -80 cm samples. No samples had *mmoX* greater than the negative control. Furthermore, for *pmoA* only ED-2 and ED-7 had amplifiable concentrations of the gene ($4.34 \times 10^{-4} \pm 2.11 \times 10^{-4}$ and $2.18 \times 10^{-4} \pm 5.54 \times 10^{-5}$). These two genes were not used in interpolations. For the *Methylocella spp. mmoX* all but four points had detectable copies (<LoD at ED -1, -4, -9 and -12). The points above the LoD were used in an interpolation (Figure 5.6).

The -20 cm samples had detectable concentrations of all three methanotrophy genes at most points. For *mmoX* only ES-18 was below the LoD. All other genes, at every -20 cm sampling point were above the LoD. There was a cluster of hotspots in obligate methanotrophs around points 2, 4 and 5.

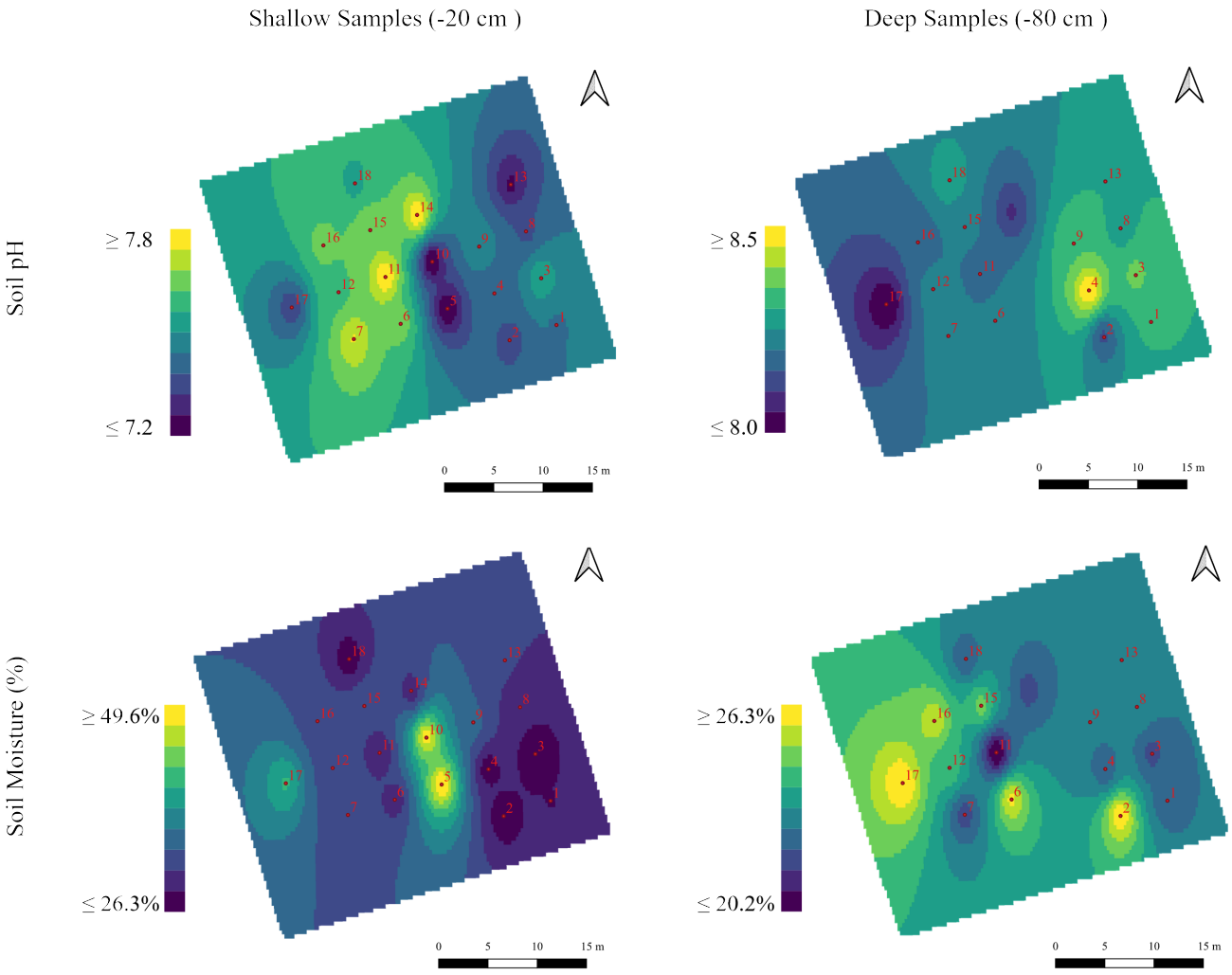



**Figure 5.5: Plots of soil pH and moisture at the two sampling depths**. Interpolations completed using IDW, p=2. Expected source is at Point 12.
**Shallow Samples RMSE:** pH: $1.47 \times 10^{-2}$: Moisture : $1.38 \times 10^{-4}$
**Deep Samples RMSE:** pH: $1.86 \times 10^{-2}$: Moisture: $1.54 \times 10^{-3}$

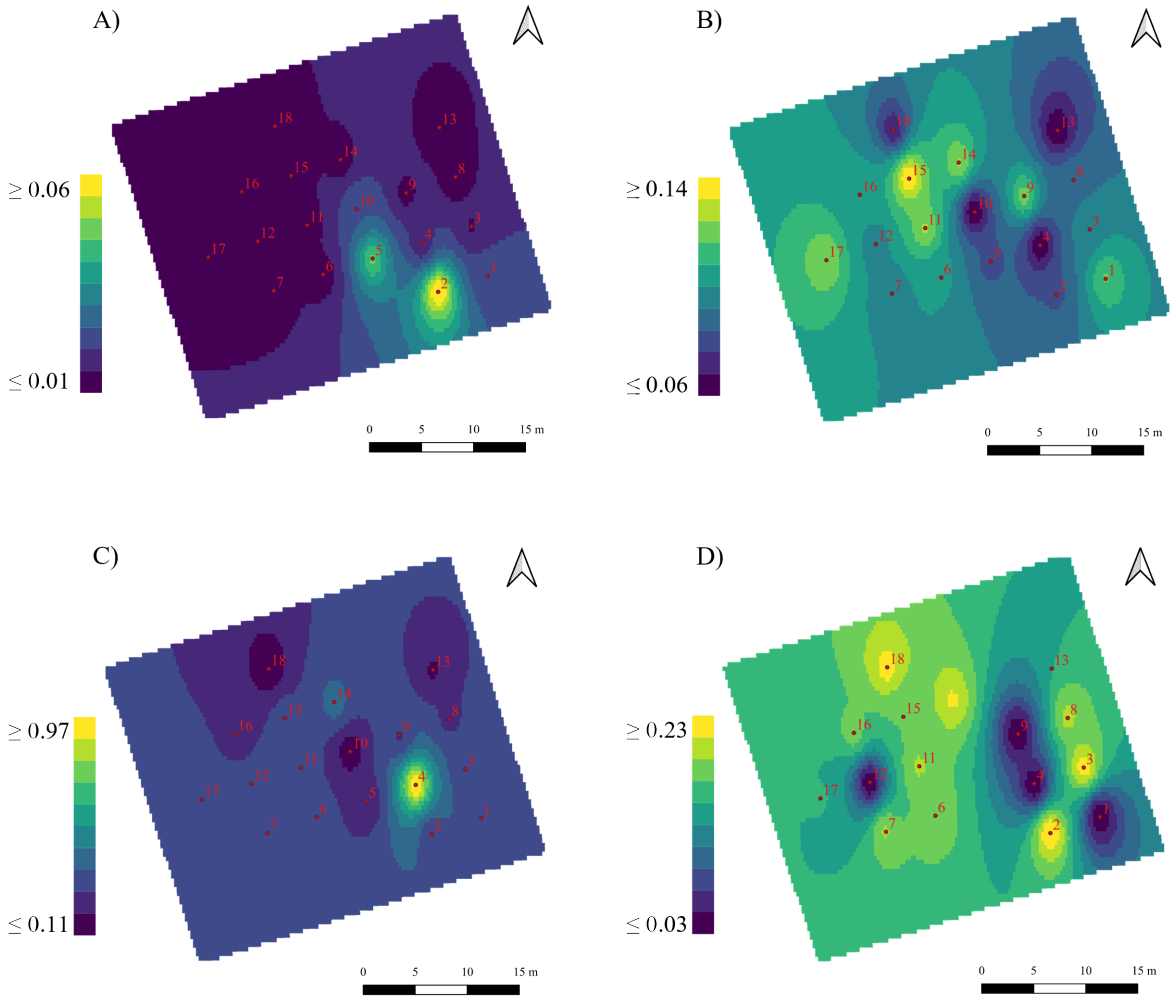



**Figure 5.6: Plots of estimated relative abundance of MMO related genes in percent.** Expected source is at Point 12. Estimations produced using IDW interpolation (p=2) and displayed using 10 equal intervals. Expected source is at Point 12. A – C used soil samples from -20 cm. While D used samples from -80 cm. Points labelled are those sampled for that depth. **A) *pmoA* RMSE:** $3.25 \times 10^{-4}$
**B) *Methylocella spp. mmoX* RMSE:** $3.31 \times 10^{-5}$
**C) *mmoX* RMSE:** $1.34 \times 10^{-2}$
**D) *Methylocella spp. mmoX* RMSE:** $4.60 \times 10^{-5}$

### 5.4.3 Community Composition

Analysis of community composition was completed by sequencing of the 16S rRNA gene. A subset of samples, forming a transect across the site, was used. Sampling points, listed in Table 5.2, ran from point 17 to point 8. As point 10 was not taken at the -80 cm it was substituted for point 7. Problems with amplification led to only 9 samples producing results, these generated 1817 unique taxa; read counts per sample are listed in Table 5.2.

**Table 5.2: Read counts and number of unique taxa at the family level.** * Denotes a sample that was excluded from downstream processing due to low read number or due to repeated problems with 16S rRNA gene amplification. Point 10 was only available at the -20 cm so Point 7 was used for the -80 cm sample. Sample ID shorthand is used throughout future plots, with: Shallow (ES) Deep (ED) samples followed by the sample points.

| | Sample ID | Read Number | Unique Family Taxa |
|---|---|---|---|
| Shallow (-20 cm) | ES-8 | 860 | 69 |
| | ES-9 | 2845 | 141 |
| | ES-10 | 2549 | 142 |
| | ES-11 | 10792 | 203 |
| | *ES-12 | 63 | 11 |
| | ES-17 | 18762 | 230 |
| Deep (-80 cm) | ED-7 | 1522 | 113 |
| | ED-8 | 1045 | 91 |
| | *ED-9 | – | – |
| | *ED-11 | – | – |
| | ED-12 | 7206 | 164 |
| | ED-17 | 2339 | 138 |

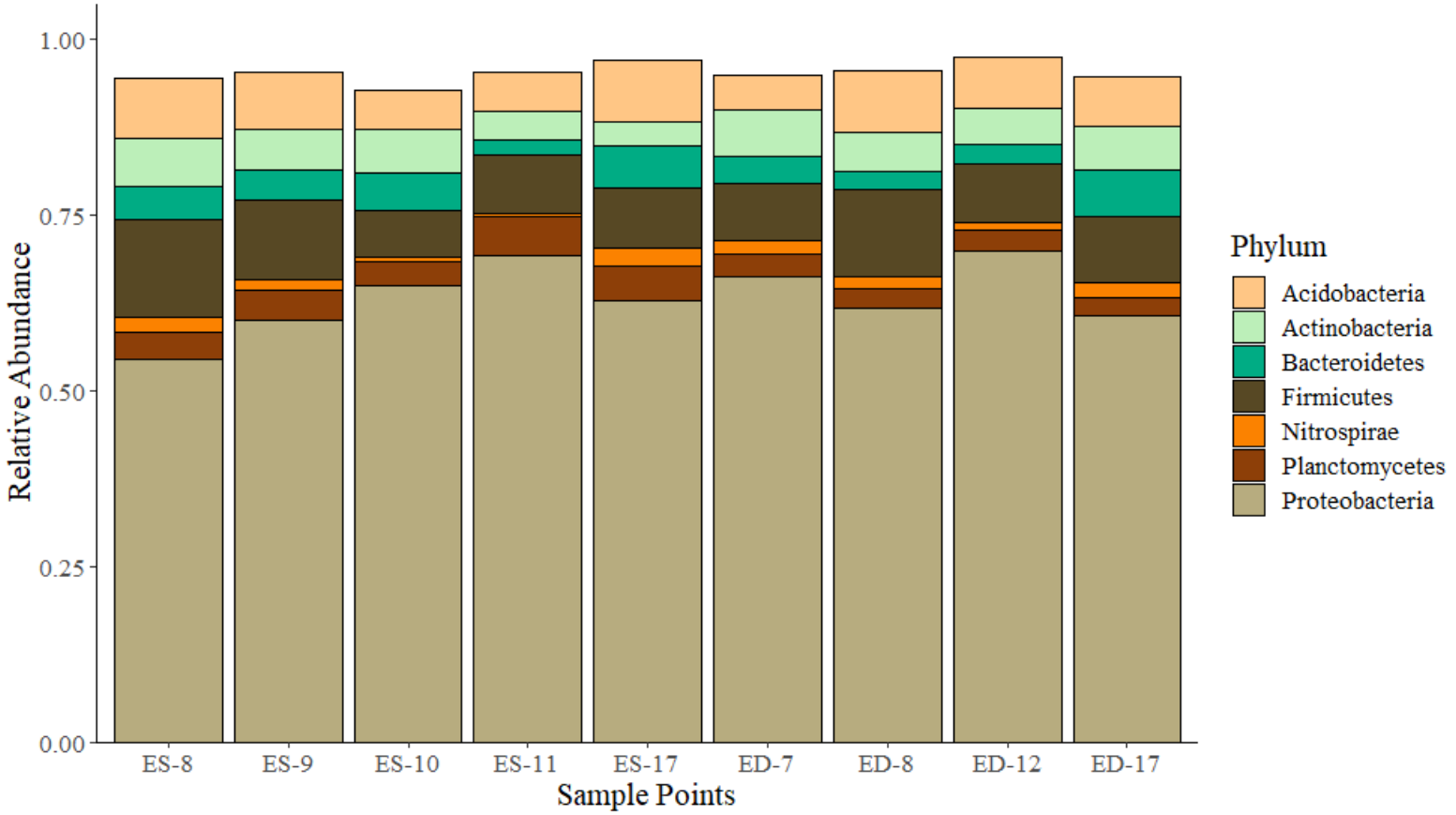



**Fig:5.7.A** –Bar plot of most abundant taxa grouped at the phylum level. No clear differences were seen between samples.

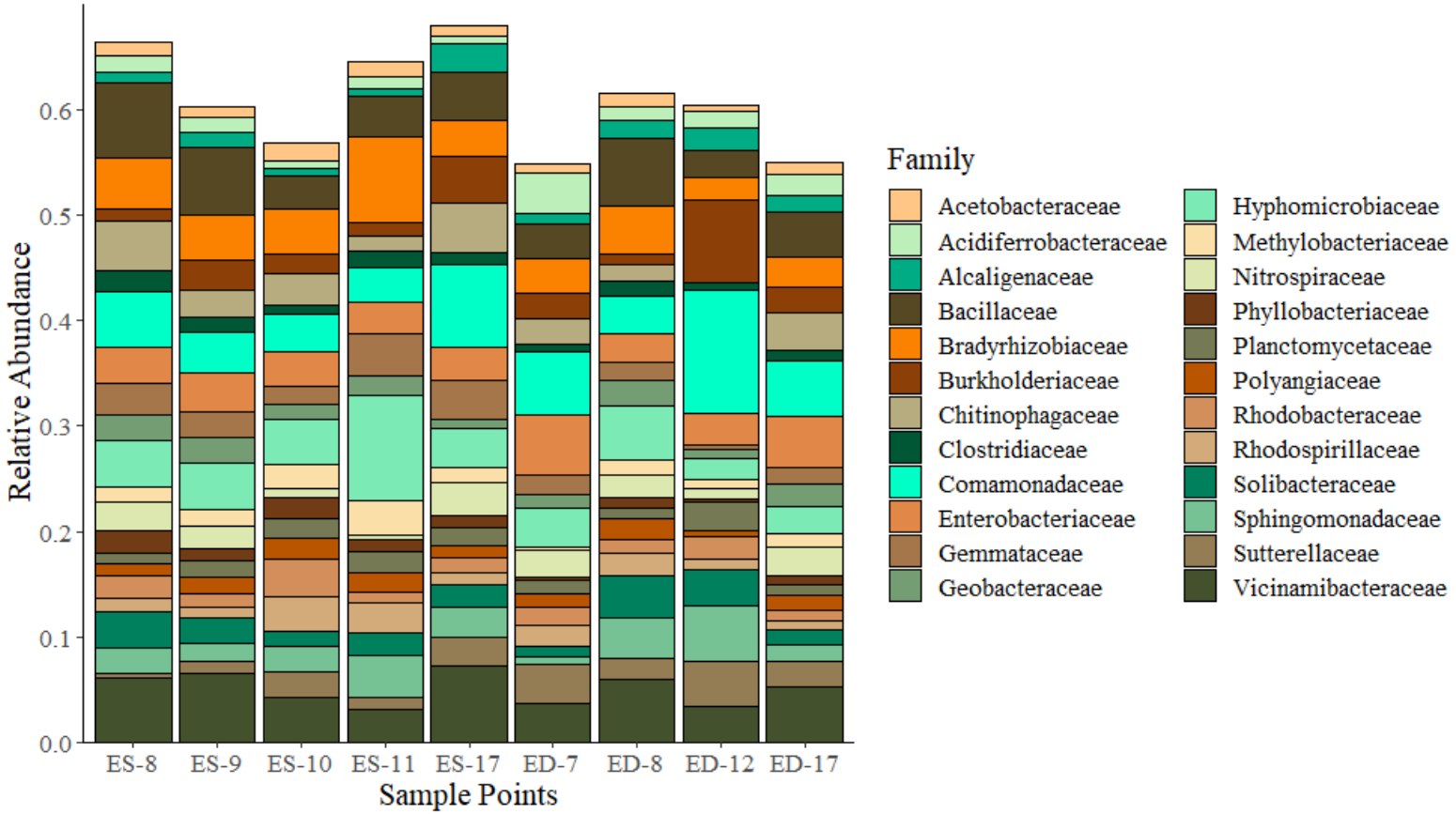



**Fig:5.7.B** –Bar plot of most abundant taxa grouped at the family level. No clear differences are present.

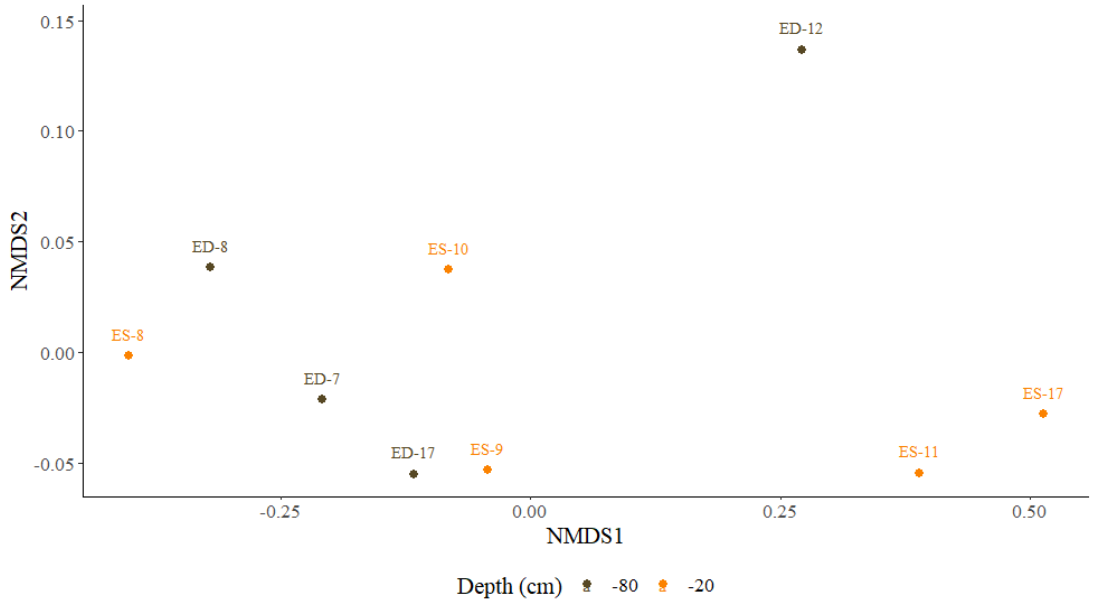



**Fig:5.7.C** – NMDS ordination at the family level using Bray-Curtis dissimilarities. Points are coloured code to depth of sampling.

**Figure 5.7: Exploratory plots of 16S rRNA abundances at the Phylum and Family level.** Boxplots show relative abundance of taxa with a relative abundance greater than $1x10^{-2}$.

Differences between the sampling points and the two depths were explored using boxplots and NMDS ordination. No consistent differences were seen between depths or across the site at either the phylum or family level. Proteobacteria was the most prevalent phylum in all samples (mean relative abundance = 0.63). At the family level, the NMDS ordination at the family level showed no separation, as shown in Figure 5.7.C. The `envfit` model did not predict any significant interactions with the measured abiotic soil factors. Furthermore, an ANOSIM testing for similarities between the two depths supported the null hypothesis, results are in Table A.6 (Appendix A.3).

For the methanotrophic guild, differences between samples were present. However, very low read numbers were obtained, with ES-8 having only 3 reads, while the maxim ES-11 having 54 reads (mean reads across all samples equals 17.1). Deeper samples appeared to have a smaller proportion of Beijerinckiaceae, as shown Figure 5.7.A. However, no clear separation was seen in the NMDS ordination in Figure 5.7.B, nor were the two depths significantly dissimilar when tested using an ANOSIM (R = 0.159, $p$=0.197).

**Fig:5.8.A**

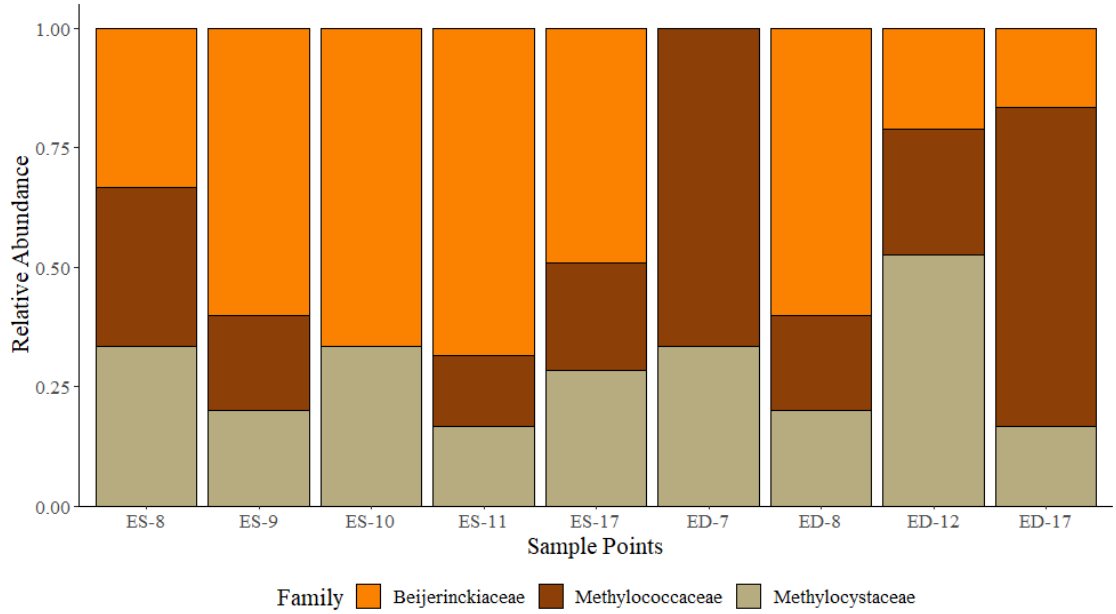



**Fig:5.8.B**

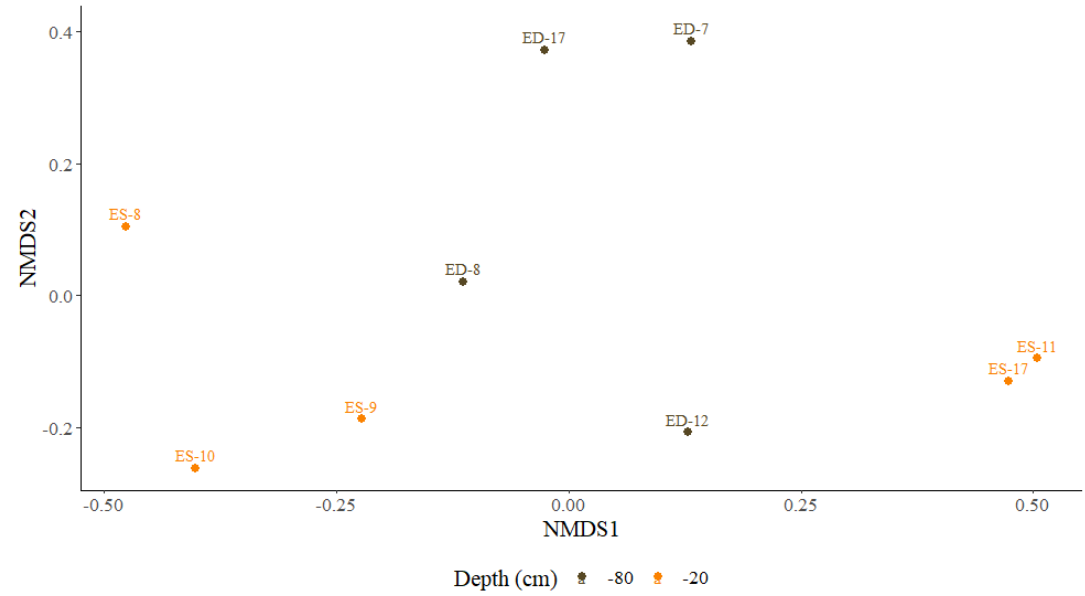



**Figure 5.8: Amplicon sequencing analysis for the methanotrophic guild.** **A)** Bar plot of relative abundance of the three methanotrophic families. **B)** NMDS using Bray-Curtis dissimilarities points are coloured by sampling depth. No separation was observed. Results should be interpreted with caution given the low number of recovered reads.

## 5.5 Discussion

A survey around a decommissioned oil well was completed to explore changes in microbial community composition and soil gas flux. In particular, changes between two soil depths were explored. No methane flux was detected at the soil surface. There were no clear spatial patterns in the microbial community, nor any significant differences between the two depths. The presence of the well head appeared to have little impact upon the soil microbiology.

### 5.5.1 Soil Factors and Gas Flux

The waterlogged nature of the site coupled with the high clay content limited the site's usefulness in testing the hypothesis. Given the evidence of gleying, the site was likely anaerobic for a several months of the year. Gleying is the reduction of iron to $Fe^{2+}$ under anaerobic conditions leading to iron's solubilisation and loss from the soil. Iron reduction and loss produces colour changes in the soil matrix, normally grey, blue or green mottling. The high clay content (and subsequent small particle size) likely inhibited gas movement during drier periods both upwards from the subsurface and down from the soil surface. As a result, changes in the microbial community between the two depths driven by loss of oxygen at depth were likely muted given the low oxygen regime at even shallow depths.

Even if methane flux from the subsurface was not present a methane flux from the wetter sections was expected to have been present. The anaerobic conditions were expected to have been ideal for microbial methanogenesis, with an anaerobic, reducing environment and an ideal pH range for mesophilic organisms (Barber, 2016). That said, the soil still had a strong red colour dominating the matrix suggesting the presence of iron oxides ($Fe^{3+}$). Bacteria that can make use of less reduced compounds, such as $Fe^{3+}$ or sulphates, can outcompete the methanogens given the greater reduction potential of their preferred terminal electron receptors resulting in a sequential reduction of compounds under anaerobic soil conditions (Barber, 2016; White, 2006). Both the anaerobic (Geobactereaceae and Clostridacea) and aerobic families are present at a relative abundance of greater than $10^{-2}$ (Figure 5.7.B). The Geobactereaceae are iron reducers suggesting was iron reduction was occurring in these sites. As a result, it is possible that the activity of methanogens was too small to detect with the fluxmeter used given the current soil conditions and given the fluxmeters LoD being $\sim$100 ppm.

There were subtle similarities between the $CO_2$ flux, soil moisture and pH. Areas closer to a neutral pH with a lower moisture appear to have a similar trend to areas with a

higher $CO_2$ flux (Figure 5.4 and Figure 5.5). A consistent finding in the literature is that an optimum moisture content is likely to enable more aerobic processes and therefore a greater release of $CO_2$ with extremes, either drought or waterlogging, leading to a drop in soil respiration (Rastogi *et al.*, 2002; Bowden *et al.*, 1998). However, the relationship between soil moisture and respiration has been shown to be more complex, with soil moisture content observed to not have an optimum range for soil respiration in some studies (for example: Lai *et al.*, 2012).

### 5.5.2 Methanotroph Distribution

The lack of methane flux made the comparison of methanotrophs to methane fluxes impossible. Plus, the high clay content likely also contributed to difficulties in detecting genes of interest using qPCR. DNA can sorb to the surface of clays (Gallori *et al.*, 1994; Khanna and Stotzky, 1992; Tebbe and Vahjen, 1993) thus limiting DNA recovery and subsequent detection. Difficulties due to the high clay content may have impacted the efficiency of the qPCR, but inability to detect either *mmoX* or *pmoA* in the majority of the -80 cm samples was surprising. High gas diffusion, and by extension high oxygen availability, is important for methanotrophy in soils (Bohn *et al.*, 2011; Gebert *et al.*, 2011; Rachor *et al.*, 2011). It is possible that the environment was simply too anoxic for too much of the year due to the high clay content and waterlogging to enable detectable aerobic methanotrophic growth at that depth.

The obligate methanotrophs were detectable in the shallower samples, at -20 cm depth (Figure 5.6). No clear spatial trends between moisture pH and the distribution of methanotrophs were present (Figure 5.5 and Figure 5.6). Within the obligate methanotrophs the greatest relative abundances were around points 2, 4 and 5. There was no increased abundance near or around the well head. This could be due to a lack of methane flux from the well to the surface at that point, or due to the high soil moisture limiting aerobic methanotroph activity at that point. Soil moisture at point 12 was 30.2%, previous studies have suggested an optimal soil moisture content of between 15 and 20% for methanotrophs (Visvanathan *et al.*, 1999). Given the high moisture and clay content, any methane effusing along the well could be taking preferential pathways through the soil and reaching the surface elsewhere.

The *Methylocella* specific *mmoX* was detectable in the deeper samples using qPCR. There were no obvious similarities between this taxa's distributions at the two sampling depths. The detection of *Methylocella spp.* suggests the facultative methanotroph occupies a different ecological niche compared to the obligate methanotrophs which were

not detectable at that depth. Methylocella spp. can metabolise acetate and short-chain alcohols (Dedysh *et al.*, 2005; Visvanathan *et al.*, 1999). These compounds are generated in anoxic environments by acetogenic and fermentative bacteria (Inglett *et al.*, 2005). Aerobic pockets or persistence through water logging events might be allowing these aerobic, facultative methanotrophs to persist.Rahman *et al.* (2011a) reported that the *mmoX* gene was detectable in a range of environments, including estuarine and lake sediments, supporting the observation of these genes in water-logged environments. Of note is the relative decrease of *Methylocella spp. mmoX* at points 2, 4 and 5, where the obligate methanotrophs are most abundant.

### 5.5.3 Amplicon Sequencing

The community across the site and at the two depths appeared homogenous. No significant dissimilarity was detected between the two depths using an ANOSIM. As discussed, it is likely that the waterlogged conditions and clay limited changes related to pore gas composition. Furthermore, while gleying increased at depth, the soil was not observed to have any other changes apart from decreasing moisture content at depth. Interestingly, low pore connectivity has been suggested to improve drivers of microbial diversity within soils by limiting competition (Carson *et al.*, 2010). High clay soils such as these seen here might have an inherently high diversity which the techniques used here struggled to detect.

The relative decrease in Beijerinckiaceae within the deeper samples appears to conflict with the qPCR data. However, the qPCR data only explored the *Methylocella* genus and the low read numbers for the three families weakens the reliability of any interpretations made.

Changes in the microbial community were expected with increasing soil depth. As depth increases, the availability of oxygen is anticipated to decline, and the composition of the matrix changes with an expected loss in soil organic matter (White, 2006). Previous studies have noted a loss of diversity with depth and reducing available organic matter in agricultural soils (Hao *et al.*, 2021; Fierer *et al.*, 2003).

### 5.5.4 Future Work

The data collected supports the null hypothesis, that sampling at increased soil depth does not improve survey sensitivity. Different field sites with a better characterised methane flux would increase chances of successfully drawing comparisons between methane flux and the abundance of methanotrophs at depth. Similarly, a site with

better drainage might enable an understanding of changes with depth and by extension changes related to oxygen availability in soils.

Further exploration of this site would require sampling during dryer periods and sampling other wells. Pore gas analysis and attempting to sample the anaerobic and aerobic interface during these periods would potentially improve the understanding of gas flux at the site and potentially reveal seasonal changes in the microbial community. The molecular techniques would need improvement to make them more robust in heavy clay soils, the addition of further purification steps or a means to dissociate DNA from clay surfaces during extraction and potentially using methods and modifications such as those described for bentonite clays (for example: Povedano-Priego *et al.*, 2021; Engel *et al.*, 2019) would hopefully achieve this.

## 5.6 Conclusions

The soil around an abandoned oil well was sampled for methane flux and changes in the methanotrophic community. No methane flux was detected nor any patterns in the spatial distribution of methanotrophs. This also limits the use of the site to test the hypothesis that deeper samples might improve the sensitivity of a molecular survey approach for methane flux.

## 5.7 Thesis Implications

While the chapter has not addressed the hypotheses it has still provided useful insights for the thesis. The issues with the clay rich soil highlighted that even if inhibition is controlled the impact may still be notable, with problems potentially occurring during DNA extraction, that limit recovery and subsequent detection. However, this site demonstrated the sensitivity of the molecular survey approach across field scales; interpolations were possible and the dataset has again highlighted the differences between methanotrophic guilds.

# Chapter 6

# Long Term Impact of a Methane Pulse Event on Soil Microbial Structure

## 6.1 Chapter Rationale

The presence of methane and other alkanes have the potential to alter the microbial community composition. However, an understanding of the scale and duration of change are poorly understood. It is therefore possible that shifts in community composition might be resulting from a historical disturbance, not a current subsurface flux. Work presented in this chapter aims to understand if historical methane exposure might pose a problem when monitoring decommissioned oil and gas wells with microbial indicators. Furthermore, it investigates the concept of microbiology being an indicator of average flux rather than a measure of flux at the time of sampling.

### 6.1.1 Note

This chapter was originally designed with a field element. Soils exposed to methane, and their corresponding controls, were left to age, potted, in the field. Sadly, these samples were inadvertently destroyed during farm work at the site. Other issues arose thanks to Covid-19, with unevenly spaced sampling points due to access problems.

## 6.2 Introduction

Soil microbial communities are a cornerstone in nutrient cycling and soil health. These communities are dynamic and respond rapidly to disturbance. Disturbances are events that result in a community change (Shade *et al.*, 2012). A disturbance can occur over three temporal patterns: a sudden change that persists over the long term (a press), a sudden short term event that dissipates (a pulse) or, an event which gradually accumulates over time (a ramp) (Lake, 2000).

Soil microbial communities can enter a new state post disturbance due to population mortality and shift in relative abundance. For example, contamination of soil is a disturbance that can provide both positive and negative selection pressures. The introduction of new substrates for metabolism can produce an advantage for the specialists (individual or consortium) that can degrade them. Conversely the presence of contaminants in toxic concentrations can elicit a strong negative selection pressure upon members of the community sensitive to that substance. For example, soils exposed to complex petroleum contamination undergo dramatic losses in microbial diversity but increases in the abundance of groups that can metabolise the substrate (van Dorst *et al.*, 2014; Flocco *et al.*, 2009).

In the presence of a natural gas flux, a disturbance, the community composition is expected to change. If the flux persists and is large enough it will cause an increase in the relative abundance of methanotrophs and alkanotrophs (Farhan Ul Haque *et al.*, 2019, 2018). However, the ability for soil microbial communities to recover from a methane disturbance is poorly understood. Therefore, it is plausible that the community might remain in a disturbed state for a prolonged period after the methane flux has ended. An attraction of using microbial indicators to detect for natural gas fluxes is the suggestion that it can act as an average indicator of flux over time.

Here, a soil microbial community was exposed to 2.5% methane repeatedly over several months. The soil was subsequently incubated for a further 18 months to allow the soil community to recover. This enabled the exploration of changes in community composition and the stability of the new community structure. Post exposure the soil community is predicted to have an increased relative abundance of methanotrophs. With increasing time since methane exposure, the relative abundance of methanotrophs is expected to decrease. Methanotrophic activity, measured using the rate of methane oxidation, is also anticipated to decline.

### 6.2.1 Chapter Objectives

Cumulatively this Chapter aimed to test:

- That exposure to a sustained methane pulse will alter a soil microbial community, pushing it into a disturbed state. This will also alter activity, with the rates of methanotropy expected to increase during the pulse.

- Once the methane pulse has ceased, the soil microbial community structure and methanotroph activity will return to a state similar to that of the pre-disturbed state,

## 6.3 Methods

Soil was collected, homogenised and then exposed to either methane or normal (control) atmosphere in laboratory mesocosms. The incubated soil was split into pots and left to age. Periodically, pots from both mesocosms were destructively sampled. Methanotroph abundance and methane oxidation were measured to assess changes in community composition and ability to oxidise methane. Figure 6.2 sets out the key components of the work and when samples were taken.

### 6.3.1 Soil Collection and Preparation

Soil was collected from a farm site at Sutton Bonington Campus, England (BNG 450653, 326476). The field had been used to graze sheep. A W-survey was completed to check there was no detectable methane flux from the site. Soil was a dark brown sand rich (coarse) loam which began to transition to a more gravel rich layer after about 25 cm depth. Soil was collected from 5-30 cm depth and left to dry at room temperature for eight weeks. The dried soil was sieved through a 1cm sieve. This served to mix the material whilst still leaving some aggregate structure. Small aggregates were required to maintain soil diversity and improve gas diffusion. Sieving removed much of the gravel.

### 6.3.2 Initial Mesocosms

Two mesocosms were built, a control flushed with normal atmosphere and one flushed with 2.5% methane. Homogenised soil was rewetted and mixed in a cement mixture. Two litres of water were added to 10 kg of soil. This was then packed, gently, into the two large mesocosm systems ready for gassing, with 10 kg of soil in each mescocosm.

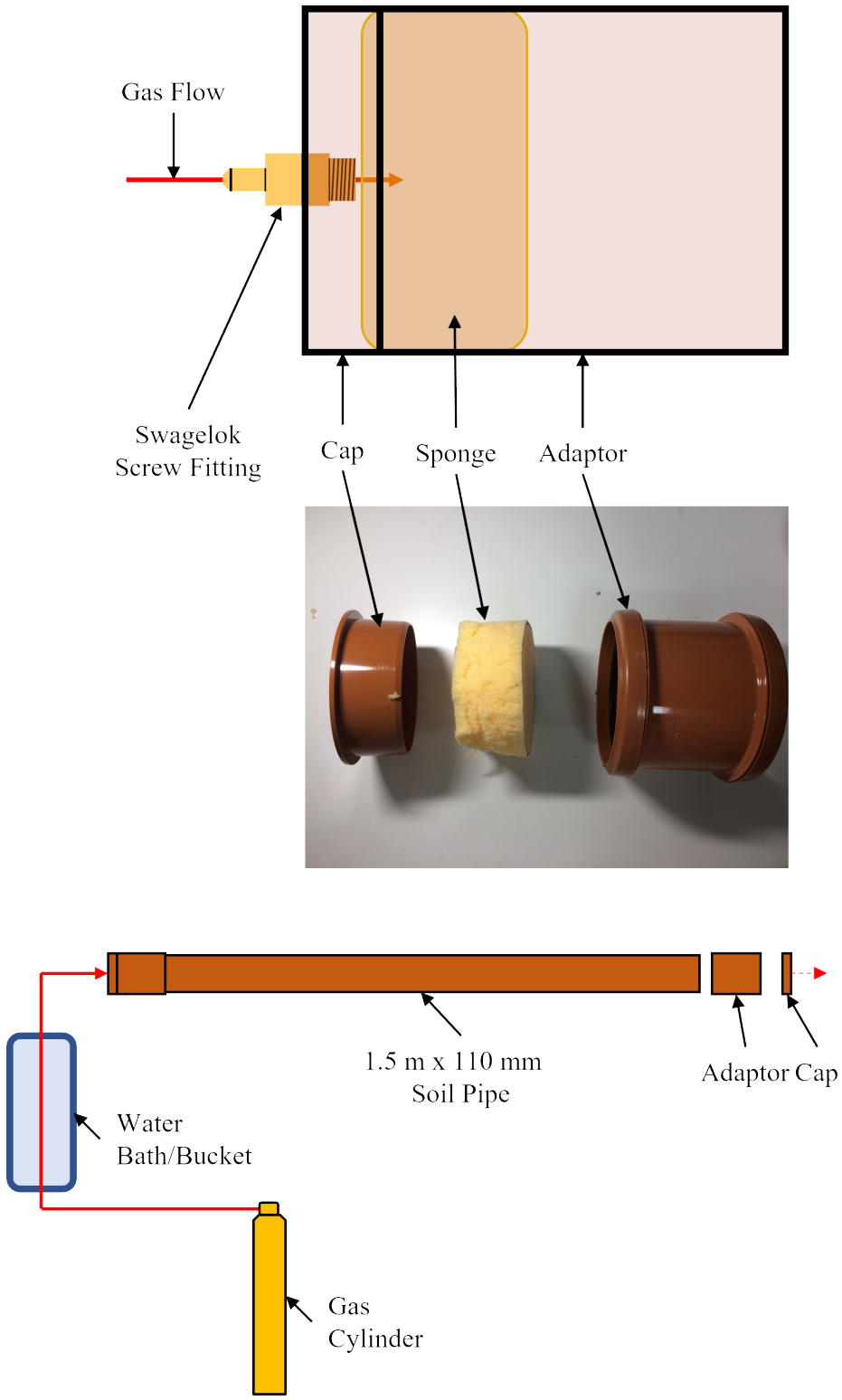



**Figure 6.1: Components of the mesocosms end caps and adaptor.** The sponge was moistened and packed into the cap. This was then pushed into the adaptor that fitted on the packed soil pipe. The internal seals within the adaptor were coated with silicone grease to improve the gas tightness of the system.

Mesocosms were constructed from 1500 x 110 mm soil pipe, sealed using adapters and caps (Figure 6.1), the integrated seals were coated with silicone grease. The caps were drilled, and a Swagelok screw fitting inserted and secured from either side. To improve the seal, rubber was fitted between the plastic cap and the fitting on both sides with silicone sealant applied as a glue. Inside the cap a moistened sponge was fitted, this aimed to stop soil blocking the inlet/outlets. The added moisture protected the soil from desiccation from the dry gases being flushed.

The systems were sealed using a Swagelok valve at each end. Three sampling ports were drilled into the side of the system equidistant apart and sealed using 13 mm butyl rubber stoppers coated in silicon grease. These enabled checks of gas concentration within the systems. Once constructed, to assess the ability for the mesocosms to withhold gas, both systems were pressurised to  250 kPa using a pump with a pressure gauge and held at that pressure for one minute whilst the system was inspected for leaks.

Gas was vented into the system from one end. A pressurised 2.5% methane in air mix (Calgaz, UK) was used for one system, while laboratory atmosphere was pumped into the second. Both gas streams were run through a narrow rubber tube coiled in bucket of room temperature water to attempt to keep gas and soil temperatures constant. Gas flow was measured using a bubble trap, the entire volume of the soil system (  12 L) was displaced over 12 to 13 mins.

Both mesocosms were flushed every two to three days for four weeks. For a final four weeks the mesocosm were flushed every seven days. This aimed to simulate a pulse event that gradually tapered off. The normal atmosphere mesocosm was flushed and sealed before the methane mesocosm. Once completed, mesocosms were dismantled individually and the soil mixed again. Mixed soil was potted and sampled for DNA extraction. Each pot contained 300 g of soil plus a piece of glass filter paper in the bottom to stop water logging but avoid soil loss. Pots were kept in the laboratory, out of direct light in a large plastic container.

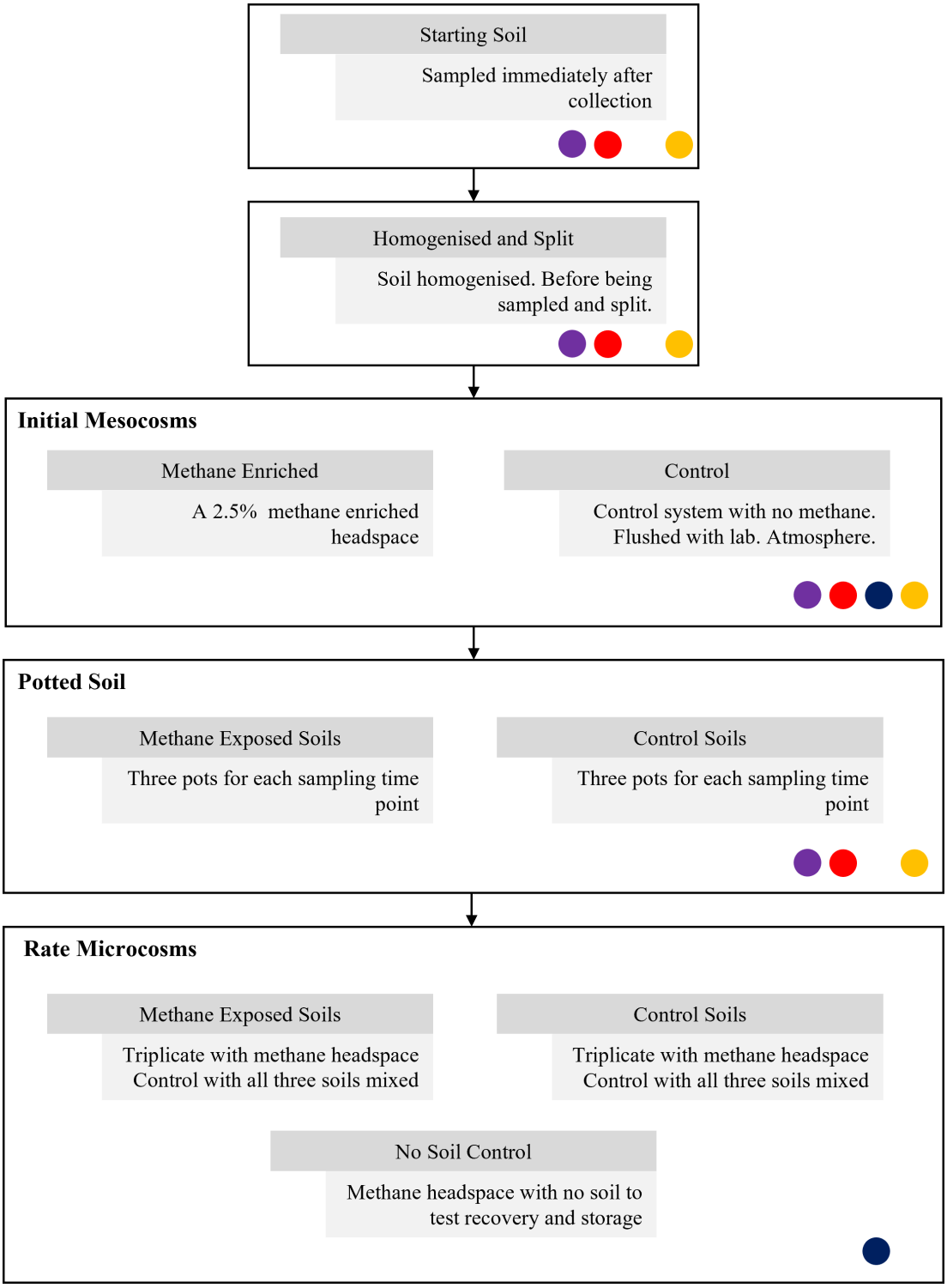



**Figure 6.2: Key sections of the experimental work and the samples collected at that point.** Coloured circles represent the samples collected at that point. **Purple:** Abiotic soil factors such as pH and moisture content. **Red:** DNA exacts used for 16S rRNA gene sequencing. **Blue:** Gas sampling for chromatography. **Orange:** DNA extracts collected for qPCR. The same potted soils sampled for DNA extracts were also used for the rate microcosm experiments for each time point.

### 6.3.3 Rate Microcosm Design

Immediately after mesocosm dismantlement, rate of methane oxidation was measured. Measurement of methane oxidation rates was repeated after 5, 9, 12 and 18 months using the potted soil. For each time point a total of nine microcosms were set up. Triplicate microcosms, with a methane enriched headspace (at 2.5%), were prepared for soil from both mesocosms. Three controls were set up, two with soil and one without soil. The controls with soil were not enriched with methane and instead tested if methane was being produced within the microcosms. The soil in these controls was an equal mix from all three pots being sampled. The control without soil acted as a recovery standard, testing if the microcosm design, sampling, or storage of gas samples might lead to gas loss and error within the data.

Each microcosm consisted of 40 g of soil at 15±1% moisture. Soil moisture content was determined gravimetrically, before being corrected to 15% with the addition of DI water. Rewetted soil was sealed in a 125 ml glass Wheaton® bottle with a self-healing butyl-rubber stopper and left for 48 hours to enable the water to distribute evenly. All glassware and butyl-rubber stoppers were autoclaved prior to use. Microcosms were sealed using metal crimps to seal the butyl stopper. To test the seal a vacuum was maintained for 30 seconds, to achieve this one needle was pushed through the stopper and attached to a vacuum pump with a pressure gauge. Once a vacuum had been established and held for 30 seconds, a second needle was pushed through the butyl stopper (Figure 6.3), to begin flushing the headspace. The second needle was attached to either a gas bag of 2.5% methane (remainder made up of normal atmosphere) or left open to allow atmosphere to enter the system. Both needles were kept in place for 15 seconds before the negative pressure needle was removed. The second needle was left in place for 30 seconds to enable the pressure to equilibrate with atmosphere before it was removed. The methane mixture was from the same source as the mixture used in the initial mesocosms.

Immediately after flushing a 1 ml sample of headspace was collected using a needle and syringe. Headspace samples were stored in 3 ml clear, glass Exetainers® (Labco UK), these had been evacuated of atmosphere prior to use. Headspace sampling continued for a week, with sampling becoming less frequent as the microcosm progressed. Gas samples were analysed for methane concentration using gas chromatography (Chapter 2.2.3). The microcosms were stored out of direct sunlight at room temperature ( 20°C).

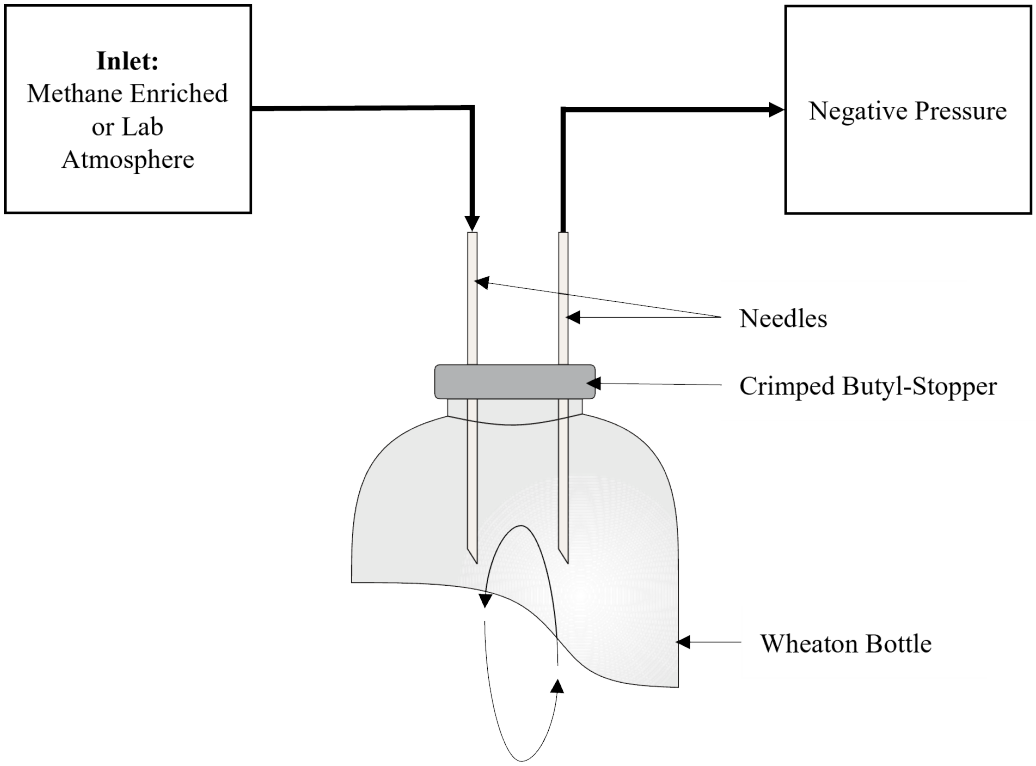



**Figure 6.3: Visual description of gas flushing during the setup of microcoms.** Negative pressure was induced via a vacuum pump attached to one needle to test for leaks. Without its removal a second needle linked to the gas under test was introduced and the microcosm flushed for 15 seconds. Finally the needle driving the negative pressure was removed and the microcosm's headspace allowed to equlibriate to amtopsheric pressure from the source gas.

### 6.3.4 Molecular Ecology

Soil samples and DNA extracts were collected throughout the experimental work. Samples from the original soil, the newly homogenised soil, and the soils collected immediately after the dismantling of the mesocosms were taken. At each sampling point three samples were collected. At the point of microcosm setup, the potted soil undergoing destructive sampling was also used for DNA extractions. DNA samples were stored at -80°C for the duration of the investigation for all qPCR assays to be completed simultaneously to ensure comparability between datasets. DNA extraction, inhibition tests, and qPCR assays were completed as described in Chapter 2.3.

Three soil samples were used to test for PCR inhibition, one sample was from homogenised soil prior to the mesocosm. Based on the inhibition tests the samples were diluted 1:25 with ultra-pure water. As all soils were from the same starting material they were diluted equally. Tests suggested the two homogenised soils had greater problems with inhibition. Using all three extracts, a linear model suggested that a 1:90

dilution would achieve optimal efficiency (linear trendline: y = -0.0047x + 1.4244, $R^2$ = 0.71). Using the trendline the 1:25 dilution was predicted to yield an average efficiency of 130% this was considered the best compromise between efficiency and over-dilution.

Soil DNA extracts were also used in 16S rRNA gene sequencing. Extracts were pooled by time point and mesocosm factor, for example all three samples originally flushed with methane in the mesocosm were pooled to produce one sample for sequencing. In total 12 samples were produced, two for each time point (one from the methane mesocosm the second from the control mesocosm).

### 6.3.5 Data Analysis

At each sampling time point there were three pots of soil for each starting mesocosm. Each pot was sampled for DNA and used for an individual rate microcosm. Overall, each pot was seen as an individual community that had developed since the initial mesocsom. Extracted DNA for each timepoint and treatment was pooled for sequencing.

Data from the molecular biology methods (qPCR and amplicon sequencing) was handled using the methods described in Chapter 2.3. Differences and interactions seen between the relative abundances, observed using qPCR, were explored using a two-way multivariate analysis of variance (MANOVA) with R's base statistics functions. Homogeneity of variances was tested using Levene's Test. Where the MANOVA detected significant differences, a one-way ANOVA and subsequent Tukey Tests were used to further test difference within factors. For all tests, significance was accepted at $p$<0 .05.

Preparation and analysis of 16S rRNA gene data was completed using the pipeline described in Chapter 2.3.9. To test if differences between mesocosm factor (methane or control) and time points, an ANOSIM was completed using the `vegan` package. The ANOSIM excluded the baseline samples.

Data from the GC analysis of headspace samples were used to estimate the maximum rate of methane oxidation. Data were normalised as a percentage of the starting value before being inverted by subtracting the percentage from 100, i.e. the value at T0 = 0. To estimate maximum rate a logistic growth curve was fitted in R using Equation 6.1. Equation from Rockwood (2015).

$$N_t = \frac{K}{1 + (\frac{K - N_0}{N_0})e^{-rt}} \tag{6.1}$$

**Logistic equation used to estimate maximum oxidation rate.** $N_0$ gives the starting methane concentration. The maximum predicted rate of oxidation is given by $r$. The asymptote, or carrying capacity of the system, is denoted by $K$ which was fixed at 100. The values of $N_0$ and $r$ were estimated by fitting the model to the experimental data using R. The solution could then be plotted and compared to the original data, with the predicted concentration $(N_t)$ plotted against time.

In R, a non-constrained logistic growth model was first fitted using the growthcurve package. The fitted, $K$, $r$ and $N_0$ values were then used in a constrained model. The non-linear statistics function (from the base `stats` package) with the port algorithm was used enabling an upper bound to be fitted. Here, the asymptote equalled 100 as the dataset was a percentage. The final model was plotted to check accuracy and the r value, with its associated error, were used for comparison between datasets.

## 6.4 Results

Soil was exposed in mesocosms to either methane or normal atmosphere. Subsequently the soil was left to age in the lab and periodically sampled. A total of five time points were sampled plus samples immediately after soil flushing mesocosm. Also analysed were two baseline samples, one of the soils immediately after being collected from the field and a second sample of the soil after being homogenised.

### 6.4.1 Relative Abundance of Methanotrophs

Two methanotrophy genes were quantified using qPCR, efficiency and $R^2$ values of the assays are in Table A.7 (Appendix A.4). Relative abundances of *pmoA, mmoX* and *Methylocella spp.* specific *mmoX* (Figure 6.4). A two-way MANOVA, excluding the baseline samples, was used to explore changes in the relative abundance of genes through time and between mesocosm factor. The results are summarised in Table 6.1. If a significant difference was predicted, a further one-way ANOVA was used to explore inter-factor effects.

For the control soil, the final two time points had *pmoA* copy numbers below the LoD of the assay (<100 copies $\mu l^{-1}$). The relative abundance of *pmoA* in the methane meso-

cosm was significantly different to the control mesocosm soils ($F_{(1,4)} = 32.88$, $p<0.001$). The methane mesocosm had a greater relative abundance of *pmoA* (Figure 6.4).

Time was also predicted to have a significant impact upon *pmoA* ($F_{(4,24)} = 656.65$, $p=0.006$). Within the methane mesocosm soils, a one-way ANOVA suggested that time elapsed since the mesocosm potting was still significant ($F_{(1,4)} = 3.75$, $p=0.041$). While *pmoA* does appear to decline through time (Figure 6.4), no significant difference between time points was detected using Tukey's Tests. For the control mesocosm, time was also predicted, by a one-way ANOVA, to be a significant factor ($F_{(4,10)} = 10.75$, $p=0.001$). Tukey's Tests repeatedly showed that at 12 and 18 months, the mean relative abundance of *pmoA* was significantly different to earlier time points. The *p* values from the Tukey's tests are in Table A.8 (Appendix A.4).

For both *mmoX* and *Methylocella mmoX* significant differences between mesocosms were predicted in a two-way MANOVA, (Table 6.1). From the means, (Figure 6.4), the control mesocosm's soil had greater abundances of both *mmoX* genes compared to the methane flushed mesocosm. Pot age and the mesocosm factor had a significant impact upon the relative abundances of the *mmoX* gene. Further one-way ANOVAs suggested that time since potting only had a notable impact upon the methane mesocosm soils ($F_{(4,10)} = 4.48$, $p=0.025$) not the control soils ($F_{(4,10)} = 1.24$, $p=0.356$). The *Methylocella spp. mmoX* was not significantly impacted by pot age.

**Table 6.1: Results of the two-way MANOVA exploring the impact of soil age and differences between mesocosm gassing.** MANOVA excluded the baseline soil. Reported *p* values are the rounded significance level reported, here a result was held to be significant if $p < 0.05$. p values are rounded to 3 dp, * indicates significance. The F column represents the test statistic (df = 4, 10).

| Gene | Level | *F* | *p* |
|---|---|---|---|
| *pmoA* | Time | 4.66 | 0.006* |
| | Methane or Control | 656.65 | <0.001* |
| *mmoX* | Time | 3.56 | 0.020* |
| | Methane or Control | 32.88 | <0.001* |
| *Methylocella spp. mmox* | Time | 0.63 | 0.646 |
| | Methane or Control | 78.08 | <0.001* |

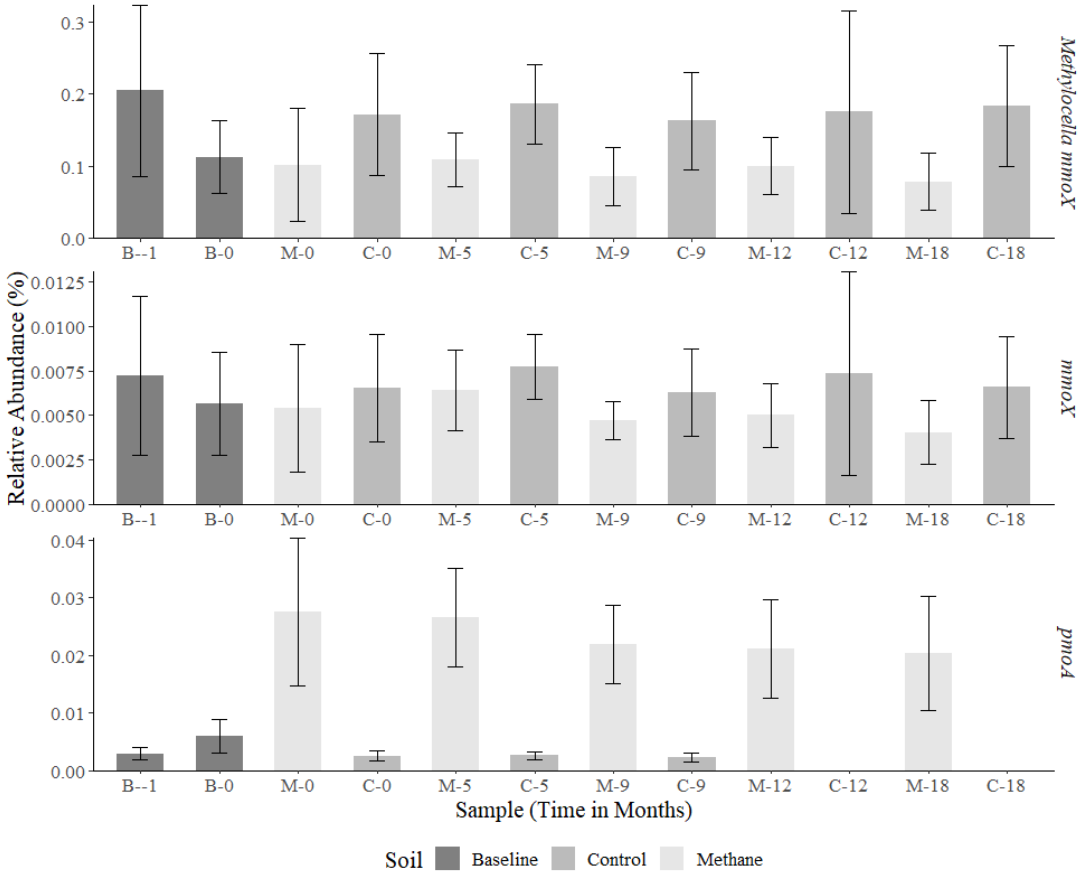



**Figure 6.4: Relative Abundances, expressed as percentages, for the two methanotrophy genes detected using qPCR.** Data uses a sample from each of the three pots from each starting mesocosm for each time point. Time (in months) elapsed from the deconstruction of the mesocosms is shown as the number on the x-axis label. Error bars are the propagated error derived from ±2 SD

### 6.4.2 Community Composition

Soil DNA extracts were pooled by mesocosm factor and time, producing 12 samples for 16S rRNA amplicon sequencing. Post quality steps and alignment, a total of 307511 reads were available for data analysis. These covered 2641 unique taxa. Reads per sample and unique taxa at the family level, are in Table A.9 (Appendix A.4).

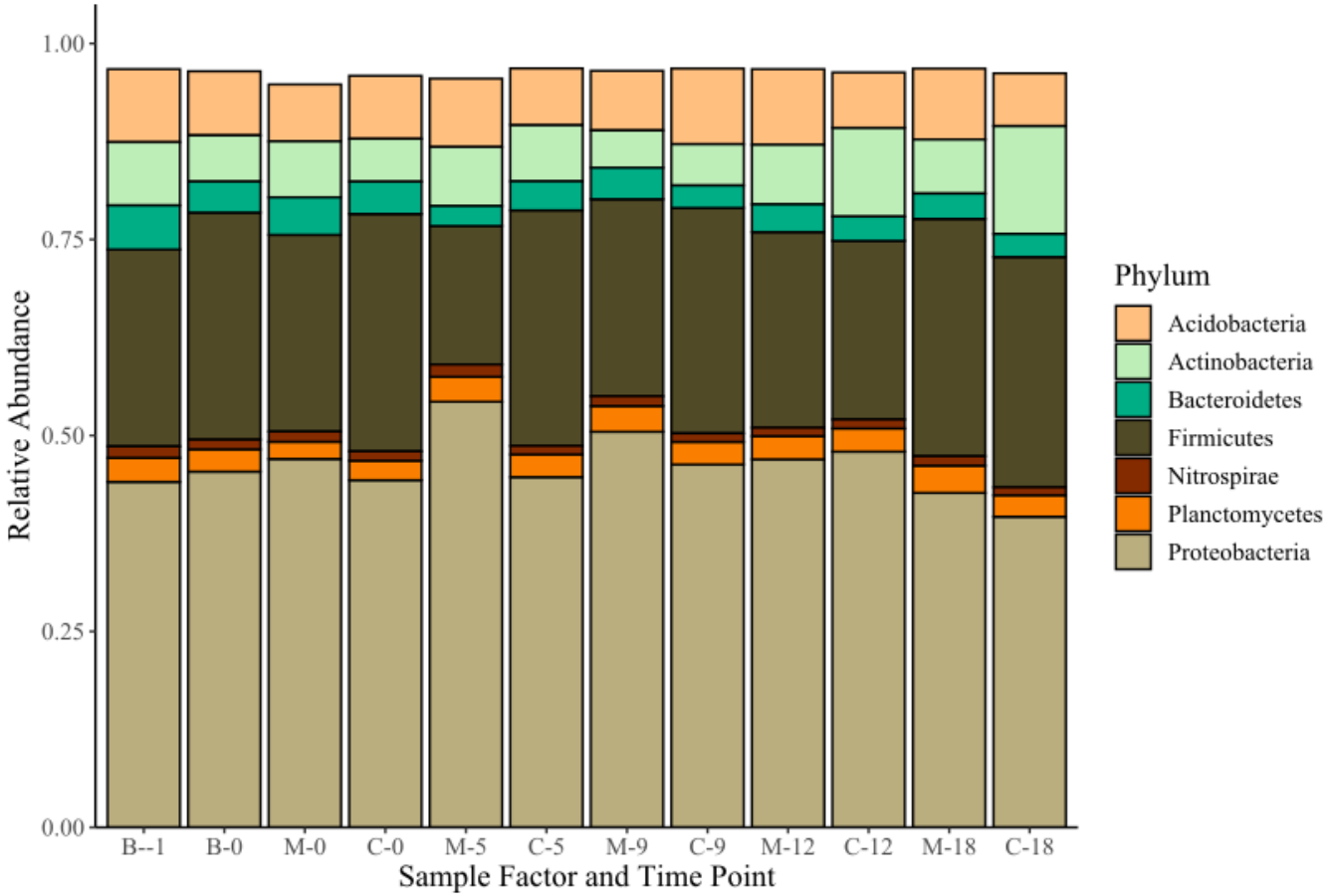



Fig:6.5.A Plot of most abundant phyla. There is little variation between the samples. M-5 shows some difference, notably an increase in the proteobacteria and a loss of Firmicutes.

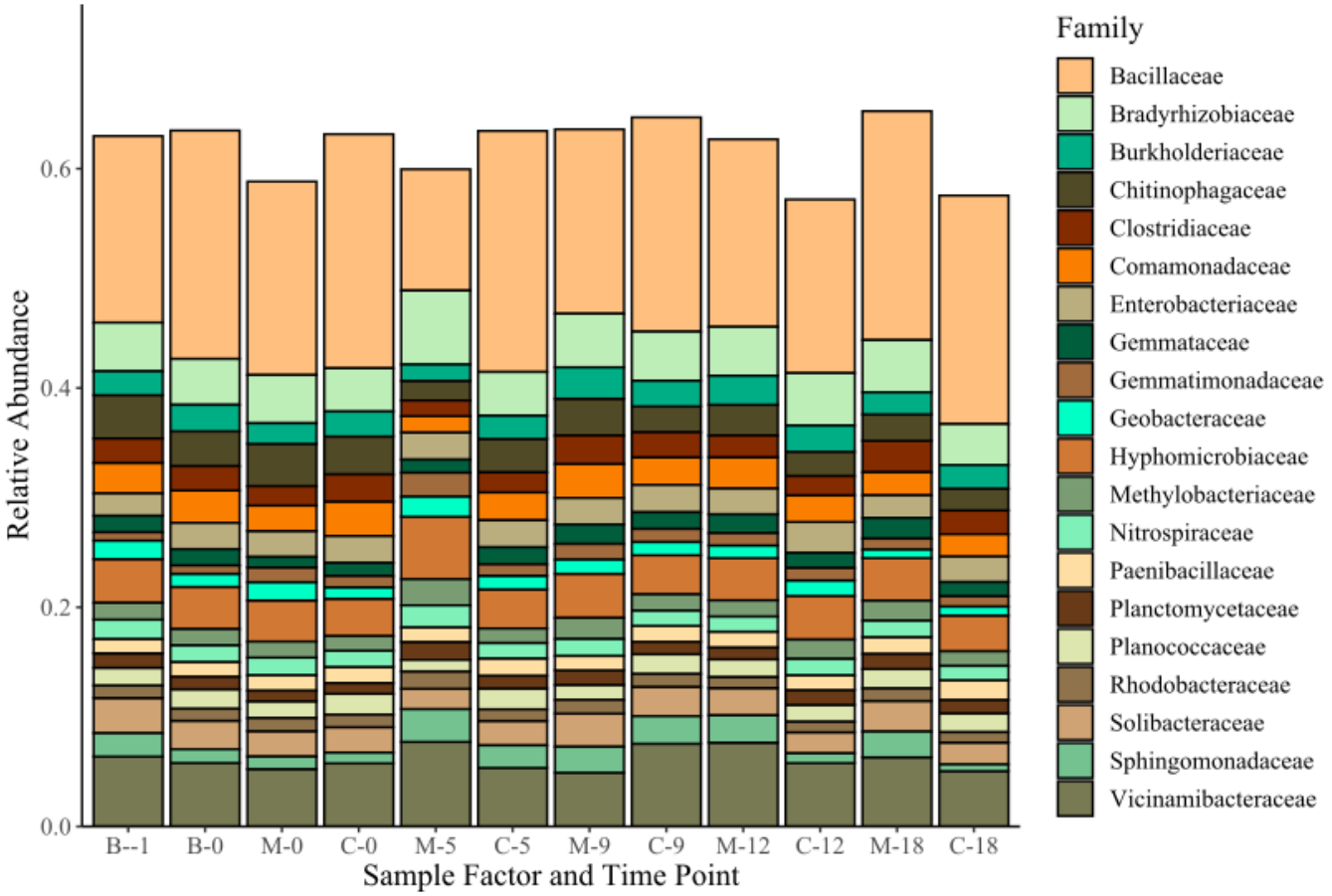



Fig:6.5.B Plot of the most abundant families. Little variation is present with broadly similar relative abundances across all samples. M-5 has a slightly different composition.

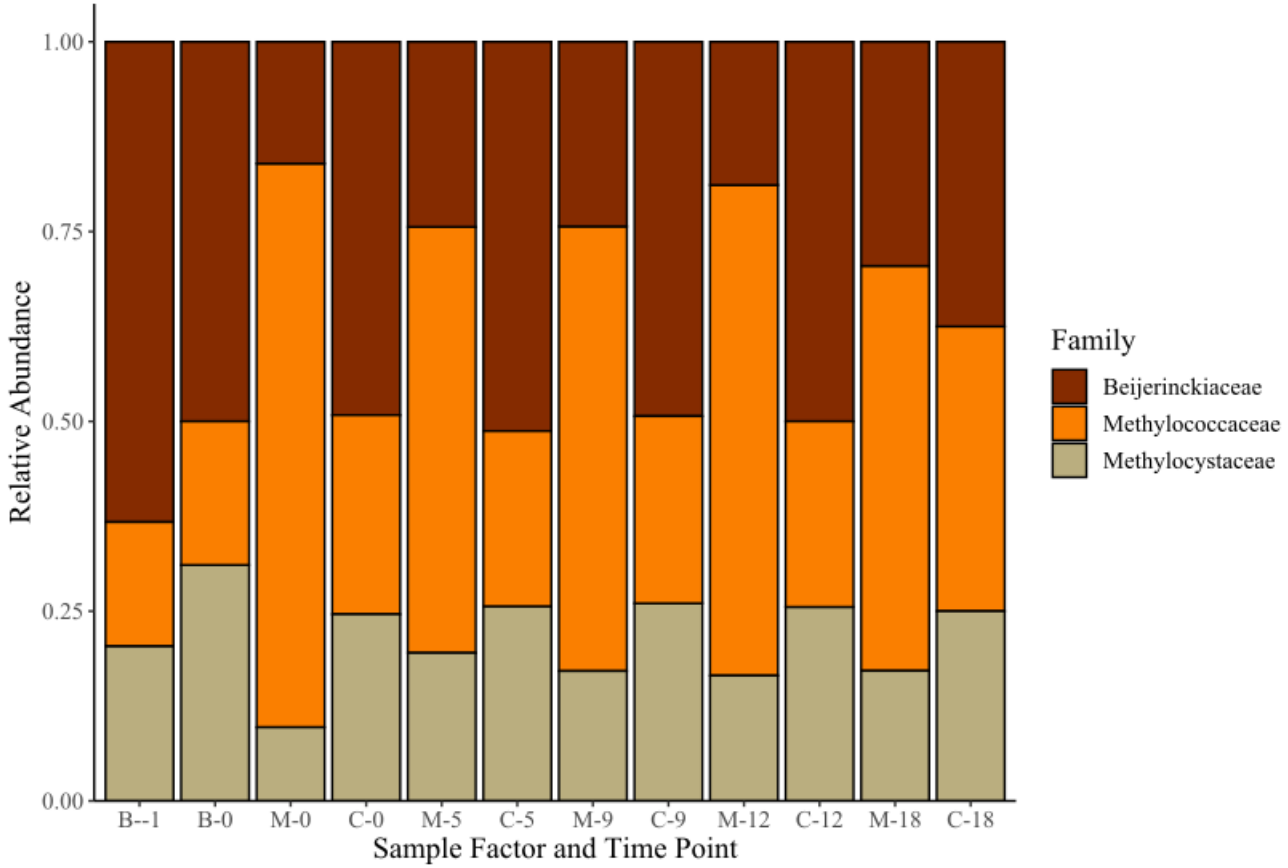



Fig:6.5.C Plot of relative abundance of the three families know to contain methanotrophs. The Methylococcaceae shows clear increases in the methane mesocosm soils compared to the control and baseline.

**Figure 6.5: Relative abundances at different taxa level from the baseline and mesocosm soils.** Plots A and B use taxa with a relative abundance above 1x10-2. Sample lists the factor and the time point, B–1 = Baseline pre-homogenisation, B-0 = Baseline post-homogenisation. M = Methane Mesocosm, C = Control Mesocosm. Number after M and C is the number of months since mesocosm deconstruction.

Stacked bar plots (Figure 6.5.A) showed little variation at the phyla level. A lack of dissimilarity between the samples was seen in NMDS ordination, shown in (Figure 6.6), with no clear separation between the mesocosm factors or time points. Proteobacteria was the main constituent in all samples (mean±SE = 0.46±0.01), followed by Firmicutes (mean±SE = 0.27±0.01), (Figure 6.5.A). At the family level community composition also appeared similar between sites, (Figure 6.5.B). Of the three families known to include methanotrophs none had a relative abundance greater than $10^{-2}$ and are not seen (Figure 6.5.B). One exception is M-5 which does appear different in its community composition at both taxa levels and which clusters separately on the NMDS. The lack of separation between mesocosm treatments was supported by ANOSIM tests, with no significance dissimilarity observed at either taxa levels (Table A.10 Appendix A.4). However, the ANOSIM did suggest weak, significant dissimilarities at the Phyla level through time (R = 0.33. $p$=0.036).

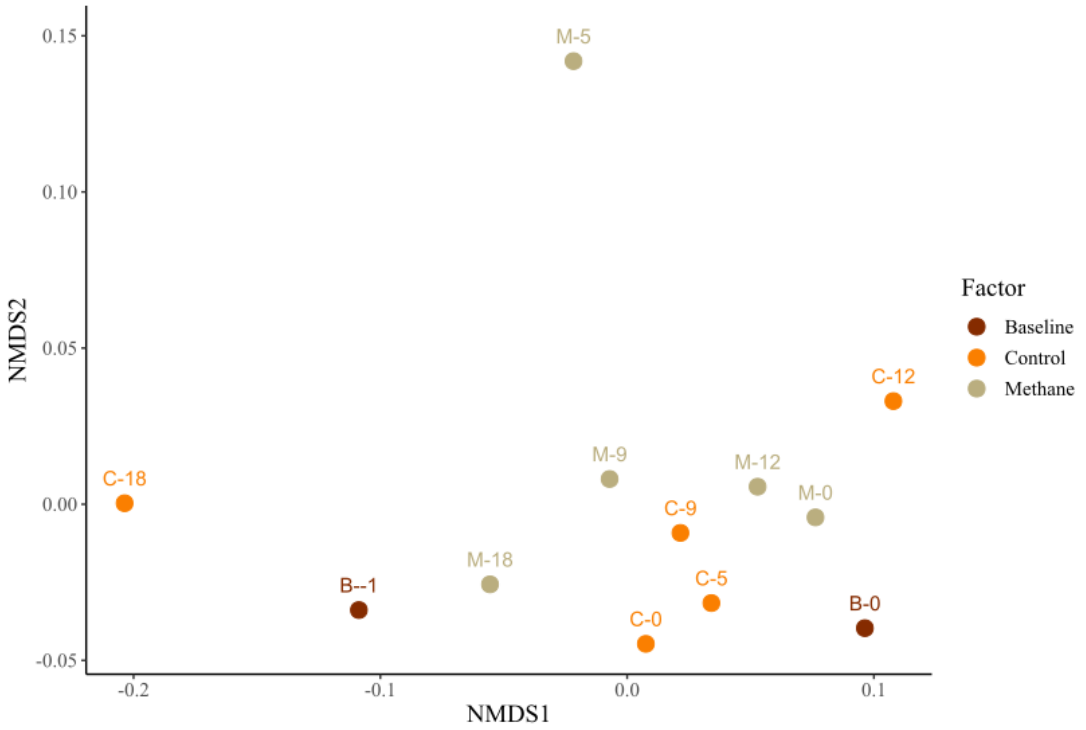



**Figure 6.6: NMDS Ordination using the 16S rRNA sequencing data grouped at the family level.** No clear separation of the samples, by either mesocosm factor or time point is seen. If M-5 is treated as an outlier and removed no impact is seen in the ordination. Ordination uses Bray-Curtis dissimilarity distances.

Differences between the three families known to contain methanotrophs were more apparent (Figure 6.5.C). The methane mesocosm soils contained a greater mean relative abundance of Methylococcaceae, 0.61±0.04 compared to 0.27±0.03 in the control soils. An ANOSIM suggested that strong significant dissimilarities existed between the

methanotrophic guild between the two mesocosm factors (R = 0.78, $p$=0.008). These differences can be seen with in the NMDS (Figure 6.7), where the two mesocosms communities clearly separated. The ANOSIM did not detect significant differences related to time (R = -0.38, $p$=0.929).

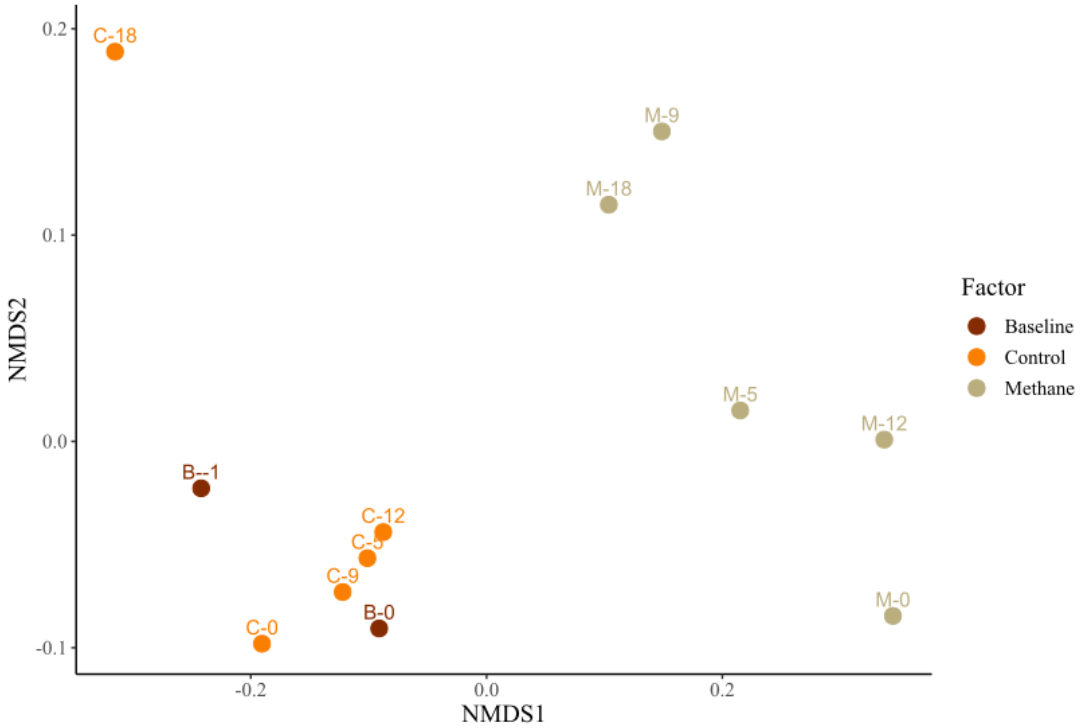



**Figure 6.7: NMDS Ordination using the 16S rRNA sequencing data using only the Methanotroph containing families.** Samples separate clearly by mesocosm factor. Baseline samples cluster with the Control samples. Time does not appear to cause any separation of the samples. Ordination uses Bray-Curtis dissimilarity distances.

### 6.4.3 Methane Oxidation Rate Analysis

A total of 45 microcosms were setup over the course of 18 months to test methane oxidation rates. There were 15 controls, no methane production was detected in the microcosms with soil and only minor losses in methane concentration were detected the final time points for the recovery microcosm. These losses were only seen after the point at which all methane had been removed from the test microcosm. Maximum methane oxidation rates were determined using a logistic model. The generated maximum methane oxidation rate (r values), plotted in Figure 6.8: showed a general decline over time within the soils from the methane mesocosms, while the rate within the control soils appeared to have an increasing trend leading to similar rates after 18 months.

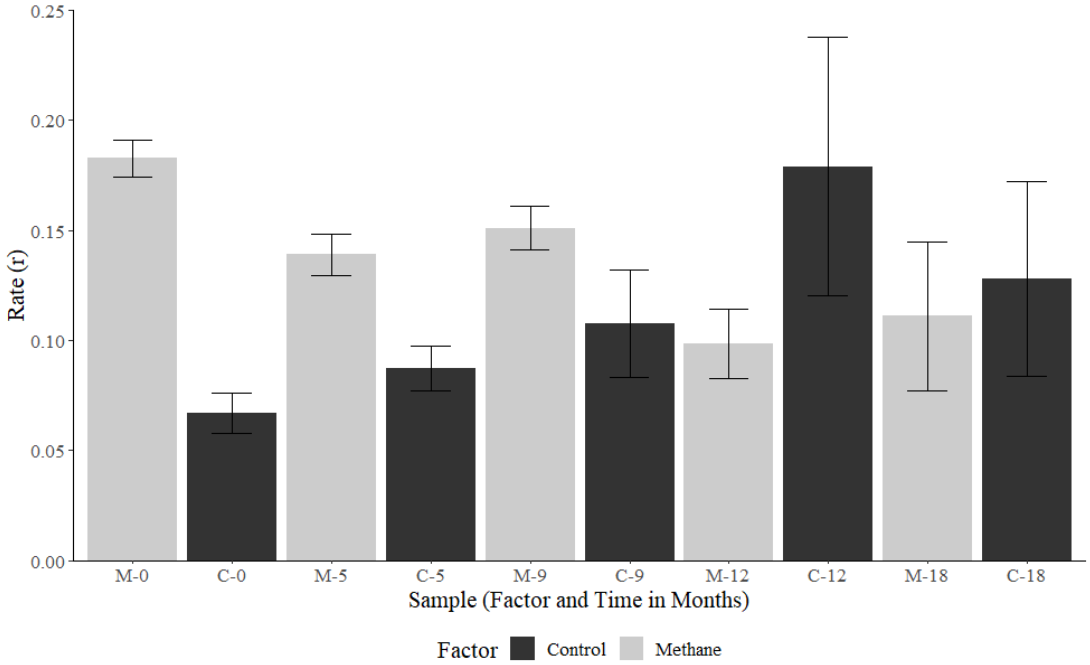



**Figure 6.8: Estimated maximum methane oxidation rates (r) for soils as they aged after mesocosm incubation.** Rate is estimated using a logistic growth model using relative concentrations and is therefore unitless.It should be noted that r is theoretically negative but due to transformation to enable estimation it is presented as positive here. Error presented is ± 1SE as produced by the logistic model using three rate microcosms for each time point and mesocosm soil.

Whilst no changes in the rate were observed, shifts in the time taken to completely oxidise methane were present. As the soils from the methane mesocosm aged, the time taken for methane to be oxidised to the limit of detection (LoD) of 100 ppm increased (Figure 6.9). At 0 months (i.e. immediately after the mesocosm) methane concentration declined to the LoD after 36 hours, whereas after 18 months it took between 84 and 96 hours. For the control mesocosm soils all replicates reached the LoD between 96 – 120 hours after incubation started.

### 6.4.4 Soil Abiotic Factors

During pot destruction, pH and moisture content were measured. Moisture content declined from around 15% to 2%. The pH showed some variability but remained between 6.7 and 7.4 for all pots through the experiment. Data is in Table A.11 (Appendix A.4).

Fig:6.9.A

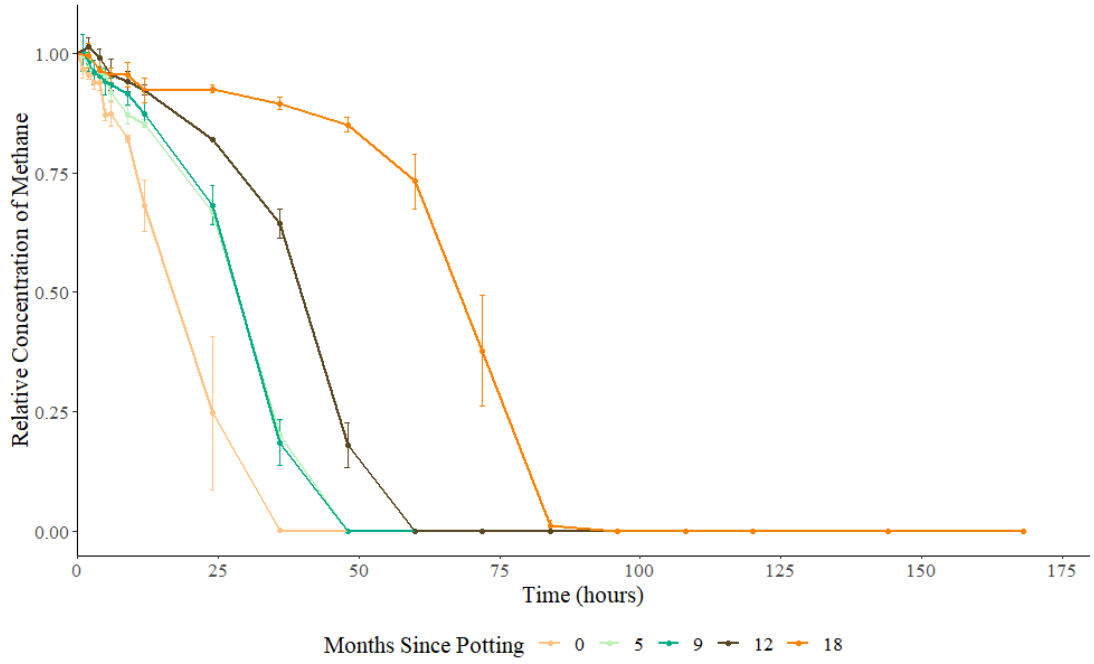



Fig:6.9.B

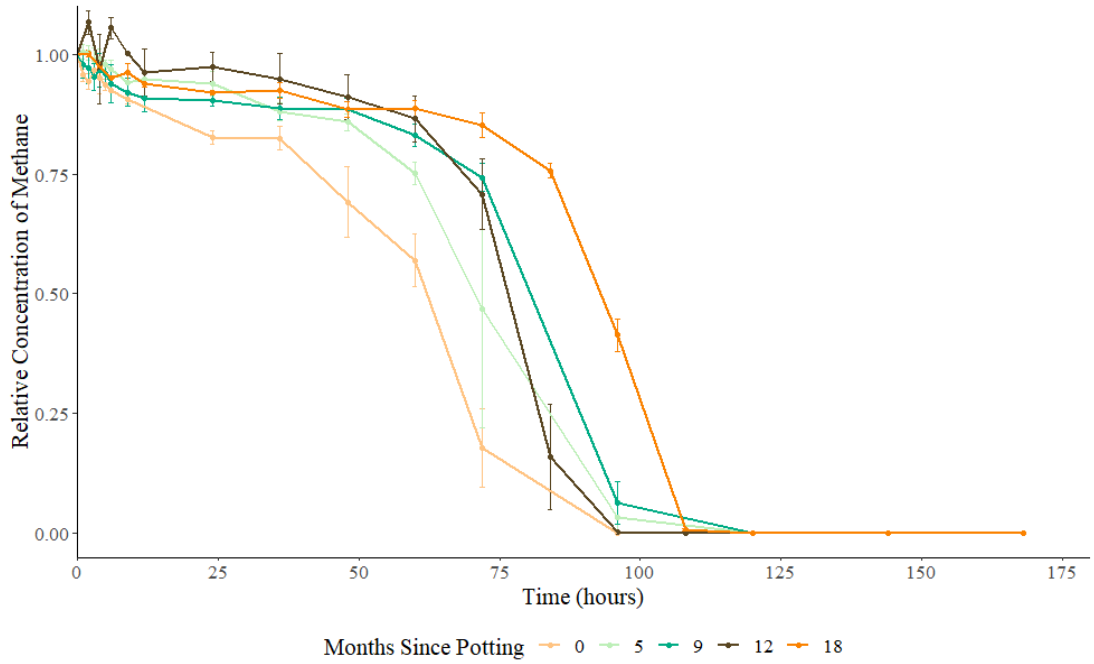



**Figure 6.9: Relative methane concentrations against time from microcosm headspaces.** Methane oxidation rates were measured six times as the soil from the intial mesocosms aged. **A)** The methane mesocosm soils, which have a clear increase in the time taken for methane to decline to the LoD. **B)** The control mesocosm soils, throughout the ageing these soils took between 96 and 120 hours for the methane to decline to the LoD. Error bars are ± 1 SE from headspace measurements of the triplicate rate microcosms.

## 6.5 Discussion

Soil microbial community stability in response to increased methane flux was explored. Community structure appeared stable with only the taxa known to contain methanotrophs showing changes in response to the methane disturbance. Once disturbed, the methanotrophic community structure did not return to the state observed in the baseline soil.

### 6.5.1 Methanotroph Abundances

The relative abundance of the three methanotrophy genes, measured using qPCR, responded unexpectedly. It was thought that increased concentration of, or exposure to, methane would promote increases in the relative abundance for all three genes. However, compared to the control, significant increases were only observed in the relative abundance of *pmoA* within the methane mesocosm soils. Prior work has also observed increases in *pmoA* abundance after incubation of soil in methane-enriched headspaces (Shiau *et al.*, 2018; Ho *et al.*, 2011). The abundance of both *mmoX* gene targets were significantly lower in the methane mesocosm soils compared to the control (Figure 6.4). No published studies exploring the changes of all three genes in a soil environment have been found to compare this result against.

Within the 16S rRNA gene sequencing data, the methane mesocosm soils saw an increase in Methylococcaceae family compared to the control and baseline soils with a relative loss of the other two methanotroph containing families (Figure 6.5.C). This also supports the increased relative abundance of *pmoA* seen in the qPCR data with obligate methanotrophs, those in Methylococcaceae and Methylocystaceae, all possessing the *pmoA* gene. The greater abundance of *pmoA* compared to *Methylocella mmoX* has been observed in soils and sediments where a strong methane flux has been observed without other alkanes (Farhan Ul Haque *et al.*, 2018).

The Methylococcaceae family only has methanotrophic genera. These are classified as the Type I methanotrophs, being $\gamma$-proteobacteria which utilise the ribulose monophosphate pathway (RuMP) for carbon fixation (Hanson and Hanson, 1996). It has been suggested that the use of soil microcosms might lead to Type I methanotrophs becoming the most abundant methanotrophic guild potentially due to Type I being more responsive to high methane concentrations (Shiau *et al.*, 2018). The Type II methanotrophs utilise the serine pathway for carbon fixation and their members are from the Methylocystaceae family (Hanson and Hanson, 1996). While Beijerinckiaceae are $\alpha$-proteobacteria and use the serine pathway, differences in cell structure lead them to

being classified as separate from the classical Type II genera (Semrau *et al.*, 2010).

There is conflicting evidence over the habitat preferences of Type I methanotrophs. Type I methanotrophs have been described as having a competitive advantage under high oxygen and low methane conditions (Amaral and Knowles, 1995). Conversely, in hydromorphic soils Type II appeared to be capable of growth at lower methane concentrations than Type I (Knief *et al.*, 2006).Henckel *et al.* (2000) observed that in incubations under a variety of $CH_4$:$O_2$ mixing ratios, the Type I population had a greater initial contribution to methane oxidation and increased in size faster than Type II. However, their population remained smaller than Type II and as incubation time increased Type II's activity increased as well.

Here, Type I developed a larger population in an oxygen and methane rich environment. The Methylocystaceae family, and Type II methanotrophs, appeared to maintain a stable relative population size despite the increase in $CH_4$. This work suggests that Type I is more responsive to this pulsed $CH_4$ event than Type II, when in an oxygen rich environment. Previous work has suggested $O_2$ limited environments have promoted the Type II methanotrophs. Macalady *et al.* (2002) reported that in rice paddies during flooded periods the Type I:II ratio decreased to the point where Type II was in greater abundance than Type I. Ho *et al.* (2013) in their review summarised Type I organisms using competitive-stress-ruderal (CSR) life strategy approach as a competitor to competitor-ruderal group which rapidly adapts to increased methane and are the more active component of the methanotroph community. While Type II populations are more generally more stable and capable of adapting to stress. These proposed strategies fit well with the data presented here, where a relative increase in Type I was seen whilst Type II maintained a similar relative population size.

That said, the usefulness of such broad taxonomic groupings has been called into question when discussing the ecological niches of methanotrophs (Knief *et al.*, 2006), and there are calls to find more suitable classification strategies (Semrau *et al.*, 2010; Op den Camp *et al.*, 2009).

Taxa within both the $\alpha$- and $\gamma$-proteobacteria have both pMMO and sMMO. It is therefore surprising that an increase in the relative copy number of *mmoX* was not seen. It suggests that the possession of a *mmoX*, and therefore sMMO, does not always confer an advantage. It is possible that if the soil had been copper limited or nitrogen rich that sMMO expression would have conferred a competitive edge, given the proposed copper catalytic centre in pMMO (Semrau *et al.*, 2010; Hakemian and Rosenzweig, 2007) and the potential for pMMO to be inhibited by ammonium and nitrate (Nyerges

and Stein, 2009).

Overall, this and previous work suggests that Type I methanotrophs thrive in environments where stress is limited. The data here, supports the possibility that methanotrophs capable of utilizing both pMMO and sMMO exist in a separate ecological niche to methanotrophs with only pMMO. Further work is needed to better understand the competitive advantage possessed by methanotrophs with sMMO. A greater understanding of how the differing physiology of the Type I and Type II produces alternative life strategies is also needed.

### 6.5.2 Methanotrophs and Wider Community Changes

Changes in methanotrophy genes through time, estimated using qPCR, were more complex to interpret. The relative abundance of *Methylocella mmoX* was not predicted to have an interaction with time. That said, significant differences were predicted for both *pmoA* and *mmoX* in the mesocosms by a MANOVA. Subsequent one-way ANOVAs supported the significant interactions between time and *pmoA* or *mmoX* for each soil factor, except for interactions between *mmoX* and time in the control soils. The interaction between *pmoA* and time, in the control soils, should be interpreted cautiously given the impact of *pmoA* being below the LoD of the assay for the final two time points. Despite the predicted significance by the MANOVA and one-way ANOVA, differences between individual time points were not detected, making a predicted trend hard to determine. Tentatively it would appear that in both mesocosm soils, there was a gradual decline in pmoA over the course of the experiment. No other clear trend in methanotrophy genes can be drawn from the dataset.

The general stability of the methanotrophy community through time is supported by the 16S rRNA gene sequencing. Whilst a clear change in methanotrophic community is present between the control and methane mesocosms (Figure 6.5 and Figure 6.7), no obvious change as the soils aged was present. This was supported by ANOSIM tests which only predicted significant differences between the mesocosms not between time points.

Taken together, exposure to a methane pulse was related to an increase in *pmoA* and the methanotrophs known to possess the gene. Once in this new disturbed state the community appeared stable. The *pmoA* did appear to decline, however the relative abundances of the methanotroph containing families appeared to remain the same.

Besides the methanotrophs, the 16S rRNA gene sequencing suggested that microbial

community composition was stable and unchanging in the presence of a methane disturbance. At both the phylum and family level no clear changes in the relative abundance of taxa were seen, (Figure 6.5). This lack of differences was supported by the NMDS ordination at the family level, (Figure 6.6), with no separation of samples caused by mesocosm factor or by time. Furthermore, the ANOSIM tests predicted no significant dissimilarities, between soil factors at either the phyla or family grouping. The dissimilarity between phyla at different time points was possibly driven by the gradual desiccation of the potted soils but there was no clear trend.

The lack of community change in response to methane flushing was surprising. While methane oxidation and methanotroph proliferation is independent of other aerobic microorganisms and is generally considered and obligate lifestyle (see reviews of methane oxidation pathways: Semrau *et al.*, 2010; Conrad, 2007), there is evidence of interactions between methanotrophs and other heterotrophs.

Previous work has suggested that methanotrophs can influence the wider bacterial community and vice-versa (review by Ho *et al.*, 2016). Methanotrophs are at the bottom of the microbial food chain, capable of fixing a gaseous carbon resource into readily accessible organic compounds (van der Ha *et al.*, 2013). Similarly, the bacterial community surrounding methanotrophs can remove waste products that might inhibit further methane oxidation. Indeed, Krause *et al.* (2017) demonstrated, in a model community, that methanol removal by partners was accompanied by changes in methanotroph gene expression. This was mediated by an unknown soluble compound, suggesting intra-community communication. Intra-community signalling, mediated by volatile compounds, promoting methanotroph growth, has also been reported in a model system (Veraart *et al.*, 2018). However, the relatively small size of the methanotrophic community might not have had a strong impact in an already nutrient rich environment. It is also possible that impacts are only felt in small portions of the soil community with spatial separation limiting any impact methanotrophs can have.

Work exploring temporal changes in methanotroph community structure have often focussed on wetland systems. Studies around rice paddies, with seasonal changes in water level and plant cover, have reported changes in the ratio and activity of Type I:II (for example: Macalady *et al.*, 2002; Ma *et al.*, 2013). Work on more stable environmental systems, like groundwater has observed that a methanotroph community can remain in a perturbed state for at least 250 days after methane injection (Cahill *et al.*, 2017). In the work here, the stability of storage conditions likely contributed to the stability of the microbial community. Unlike field systems, with climatic changes, and

plant and invertebrate colonisation the soil was held in an almost static state. The lack of biological activity likely limited community changes in response to competition, predation or other stresses.

### 6.5.3 Methane Oxidation Rates

Methane oxidation rates in the two mesocosm soils appeared to converge through time (Figure 6.8). The lag time for methane oxidation rate to reach maximum and for the methane to reach the LoD (100 ppm) (Figure 6.9) also showed convergence. Immediately after dismantling mesocosms, the methane incubated soils oxidised the methane to the LoD up to 60 hours before the control soils. This difference decreased, with the final time point, at 18 months, being approximately 12 hours.

The shorter lag phase immediately after dismantling the mesocosms was likely due to both increased community size and activity. However, despite the lag in methane oxidation increased with pot age, no reduction in population size was seen in the sequencing dataset. It is plausible that the PCR amplified dead bacterial cell DNA or dormant community members, while active methanotroph numbers had gradually declined. Furthermore, while rate declined in the methane mesocosm soils no significant change was seen with the methanotrophy genes (measured using qPCR), further supporting the concept that PCR was amplifying DNA from dead cells or cells that had entered a dormant or low activity state. As time from the initial mesocosm increased the immediately viable methanotrophic population size fell and lag time increased.

The increasing maximum methane oxidation rate through time in the control soils is harder to explain. Given the stability of the laboratory conditions, it might have been due to heterotrophic microbes entering a dormant state, allowing methanotrophs to compete more effectively for oxygen as the soil was rewetted and disturbed. With a sudden improvement in environmental conditions and a high methane resource the methanotrophs responded faster than other members of the population.

The gradual desiccation of the potted soils could be forcing methanotrophs into a dormant stage or causing mortality. Previous work has shown that soil desiccation can cause a significant drop in methane oxidation rates accompanied by a drop in pmoA copy number, but these recovered after a few days (Fierer *et al.*, 2003).

Teasing apart the interaction between soil factors and community activity is difficult. Previous laboratory and field studies have shown that soil factors can impact soil oxidation rate. Decreasing soil porosity, altered through either soil texture or compaction,

has been shown to have a negative impact on methane oxidation rates (Bohn *et al.*, 2011; Kightley *et al.*, 1995)). Similarly, there appears to be an ideal moisture regime for methanotrophic activity (Visvanathan *et al.*, 1999). These soil factors will impact gas diffusion rates, thus limiting both oxygen and methane availability.

Here, the rate of maximum rate of methane oxidation appeared independent of community size, when measured using PCR. The lag time to reach this rate increased as time from the initial methane exposure increased but was not related to changes in methanotroph community size or composition. Cumulatively, methane oxidation could not be clearly linked to the community metrics measured here. This suggests that the methanotroph community is potentially impacted by other soil factors or that the use of DNA and PCR to assess potential community function is not ideal.

### 6.5.4 Community Stability

A community's response to a disturbance can be framed using the terms resistance and resilience which cumulatively describe a community's temporal stability. Using the definitions from Shade *et al.* (2012) and Allison and Martiny (2008), resistance is the ability for a community to remain unchanged in the face of a disturbance while resilience is the ability for a community to recover to its previous state after a disturbance. In this work, the majority of bacterial community demonstrated a resistance and overall stability to the methane pulse event with no shifts in composition observed. That said the methanotrophic community did enter a disturbed state, with changes in relative composition and changes in methane oxidation rate. The change was seen the primarily in the methanotrophs possessing pMMO. Once the disturbance was over, after dismantling the mesocosms, the methanotrophs showed a low resilience and did not appear to return to the pre-disturbed state, however changes in the methane oxidation rate did occur. This would suggest that the activity of the disturbed community did return to the pre-disturbed state, possibly due to mortality or dormancy of the methanotrophic community which was not detected using PCR given it can amplify DNA from inactive cells.

### 6.5.5 Future Work

The difficulty linking community activity to size is unsurprising. The use of RNA as an indicator of methanotroph activity and survival is likely to be more accurate. Further work with soils in a field environment is needed to test if the abundance of methanotrophs remains as stable over time as observed here. This had been planned

but unfortunately the samples were lost. A less stable soil environment might promote the loss of methanotrophs after a pulse event if other aerobic microbial processes and predation continue and an enlarged methanotrophic community cannot compete without a high flux of methane to support it.

## 6.6 Conclusions

A soil microbial community was exposed to methane for several months before being left to age. Exposure to methane increased the pmoA copy number and increased the proportion of Type I methanotrophs in the methanotrophic guild. Once perturbed the soil microbial community appeared to remain in an altered state for 18 months compared to both the control and baseline soils; however, the activity of the community declined, with the lag time increasing and predicted methane oxidation rate decreasing. The stability of the methanotroph population was unexpected but might in part be explained by the PCR assays amplifying DNA from dead or dormant cells. Further work with soils in a field environment is needed to better explore if dead cellular DNA remains amplifiable in a less stable environment.

## 6.7 Thesis Implications

The apparent stability and longevity of the disturbed methanotroph community after a pulse event is potentially concerning. If the community does not return to an unaltered state after the flux has ended, then using qPCR to survey or monitor methane fluxes might be misleading as the flux might have ceased but it does support the idea the microbial indicators offer an insight into the flux over time. That said, agreement with previous chapters regarding changes in *mmoX* are interesting and pose interesting questions for future work on the advantages of possessing two methane-mono-oxygenases.

# Chapter 7

# Relative Abundance of Mono-Oxygenase Genes in Contaminated Land Sites at Former Gasworks

## 7.1 Rationale

The techniques developed throughout the thesis potentially have uses outside of monitoring methane and natural gas seeps. The mono-oxygenases are a broad group of enzymes with the potential to interact with a variety of organic substrates. Some of these are persistent environmental contaminants, for example PAHs. To test if these assays could be used in bioremediation studies samples were collected from contaminated land sites to see if the mono-oxygenases were a part of the microbial community degrading the contaminants present. This is hoped to broaden the use of the techniques developed here and potentially showcase a novel use with a commercial application.

## 7.2 Introduction

Contamination of soils is prevalent at sites throughout the UK thanks to a rich industrial heritage. These sites have a range of organic compounds that may pose a risk to health and the environment, for example polycyclic-aromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBs). Frequently, these compounds are persistent in the environment thanks to a stable chemical structure whilst also being hydrophobic and lipophilic thus promoting their partitioning out of the aqueous phase into cell membranes or soil aggregates (Jones and de Voogt, 1999). Furthermore, their inherent toxicity can limit growth and activity of microorganisms that might degrade them (Klimek *et al.*, 2016; Labud *et al.*, 2007). Sites used for the gasification of coal and the storage of gas are frequently contaminated with both organic and inorganic compounds (Baylis and Allenby, 2006). Previous research on gas works soils has demonstrated that soil microbial communities are heterogenous across the sites and that there is a relationship between the contaminant burden and the community present (Ferguson *et al.*, 2007) but little further work exploring the microbial ecology of these sites has been completed.

Recalcitrant organic contaminants, for example PAHs, can be catabolised by microbial consortia (Janbandhu and Fulekar, 2011; Johnsen *et al.*, 2005). This is despite the potential toxicity and low solubility of the starting compounds and the subsequent metabolites. Degradation can either be for energy and biomass production, to enable excretion from the cell, or as a result of co-metabolism (Johnsen *et al.*, 2005). During co-metabolism, the compound is transformed inadvertently by enzymes evolved to interact with a similar compound, no advantage is gained by the microbe involved. Microbial consortia can degrade compounds thorough co-metabolism in a stepwise fashion, with compounds being absorbed, altered, and secreted as waste repeatedly by multiple unique microorganisms.

The mono-oxygenase enzymes are a broad group involved in the oxidation of a variety of organic compounds. Particulate methane mono-oxygenase (pMMO) is a copper metalloenzyme bound to the cytoplasmic membrane in a small number of specialist taxa (methanotrophs). Similarly, the soluble methane mono-oxygenase (sMMO) has a small distribution between taxa but uses a di-iron centre to catalyse the oxidation of methane. Whilst these two enzymes have evolved with a specific substrate, there is evidence that they can oxidise other compounds (Pandey *et al.*, 2014; Semrau, 2011). The sMMO is part of a broader group of enzymes, the soluble di-iron mono-oxygenases (SDIMOs) which can oxidise a variety of organic compounds and have been reported

in 1,4 dioxane contaminated groundwater (Li *et al.*, 2013). The SDIMO are grouped by their preferred substrates and structure (Leahy *et al.*, 2003).

It is suggested that the relative abundance of mono-oxygenase genes might increase within microbial communities co-metabolising organic contaminants given their wide substrate range. However, the concentration of contaminants will potentially reach a point where toxic effects will reduce the overall community size. The mono-oxygenase enzymes could enable the oxidation of stable organic compounds, thus assisting in the natural attenuation of contaminated land sites. While unable to grow on these compounds, the presence of other 'waste' metabolites being produced organisms within the community, could support the organisms possessing these mono-oxygenases and vice versa. A better understanding of these soil communities and the organisms within them could help better inform monitoring of natural attenuation or bioremediation strategies, with the mono-oxygenase genes acting as a bioindicator of these processes.

### 7.2.1 Chapter Objectives

Cumulatively this Chapter aimed to test that:

- Soils with increased organic contamination burden will have have an increased abundance of microbes with mono-oxygenase enzymes.
- There will be a ceiling where contanimant burden becomes toxic and there will be a loss of microbes with mono-oxygenase enzymes..

## 7.3 Methods

Samples of soils and 'made ground' with a range of contaminant burdens were collected for chemical and biological analysis. In total 50 samples were used, a spectrum of organic and inorganic contaminants were quantified along with the relative abundance of three mono-oxygenase genes.

### 7.3.1 Sampling Sites

A set of 50 samples with a gradient from agricultural land through to grossly contaminated were sought. Three former gasworks sites undergoing remediation were visited and 40 samples were collected from soil heaps, newly excavated trenches or from the surface soil. At these gasworks, samples from material that could be reused, material that would need treating and material that would need to be disposed of were collected

**Table 7.1: Description of the subjective grading used when collecting sample material.** At most contaminated land sites, the material had been sorted into spoil depending upon the action needed. At one contaminated site material was collected before sorting into spoil and there was little to clearly distinguish the contaminant burden, these samples were graded a 0 and described as unknown.

| Grade | Description of Burden |
|---|---|
| 0 | Burden unknown |
| 1 | Reusable or Background Material |
| 2 | Borderline Contaminated |
| 3 | Treatable Material |
| 4 | Material For Disposal |

to produce a spectrum of samples with varying burdens. Alongside these samples, ten were collected from various sites thought to have a lower contamination burden, these samples were from roadsides, car parks, public parks, and arable fields. Samples were graded from 0 to 4 during collection dependent upon their suspected contamination burden; this was often informed by consultants on the ground and an understanding of the materials' sources. These gradings are given in Table 7.1.

Approximately 2 kg was collected for each sample. These were transported in coolboxes and frozen at -20°C as quickly as possible. Samples were partially defrosted for sub-sampling for DNA extraction and refrozen before shipping in coolboxes for chemical analysis.

### 7.3.2 Sample Analysis

To assess the biological community, samples were removed from the freezer and between 0.3-0.5 g of material were subsampled to produce one sample for biological analysis per soil sample. Immediately after sub-sampling DNA extraction was completed following the protocol described in Chapter 2.3.2. Due to the potential toxicity of the samples the DNA was extracted in a HEPA filtered cabinet; this increased the risk of contamination and additional kit blanks were completed. Extracts were used in qPCR to estimate the relative abundance of *pmoA*, *mmoX* and the SDIMO $\alpha$-subunit gene (from here referred to as SDIMO). Relative abundances used the copy number the 16S rRNA gene as a proxy for the total bacterial community size, this was also

estimated using qPCR. Details for the qPCR assays can be found in Chapter 2.3.7. The remainder of the sample was sent to an external laboratory, ALS Environmental Services, for chemical analysis; analytes and techniques are in Table A.12 (Appendix A.5)). A total of 97 data points for each sample were quantified by ALS; from these the total aliphatic, total aromatic and total PAH concentrations were used, along with concentrations of a selection of metals which were above limit of detection in the majority of samples.

### 7.3.3 Data Analysis

To attempt to separate lethal from the sub-lethal interactions, samples were sorted into four sub-groups by their potential toxicity. This was determined using the polycyclic aromatic hydrocarbon (PAH) concentrations as set out in the Safe 4-Use Levels (S4ULs) (Nathanail *et al.*, 2015). An initial group using the threshold concentrations for land being used as a public park with 6% soil organic matter (SOM) was produced with all samples containing at least one PAH above the thresholds being excluded. The PAH threshold concentrations from this S4UL were multiplied by 0.5, 2, and 10 to produce a further three sets of threshold values. Each group was produced from all 50 samples; therefore some samples were in all four groups and the grouping with the highest threshold values contained the largest number of samples. This approach also aimed to enable some distinction of the inhibitory effects of the starting material on the qPCR assay.

Collected data was tested for normality using the Shapiro-Wilk test. To assess interactions between the chemical and biological analyses a correlation was completed in R (Ver. 4.0.5, R Core Team (2021)). Non-normal data were correlated using Spearman's rank correlation and visualised using a correlogram (`corrplot`, Ver. 0.90, (Wei and Simko, 2021)). Given the filtering by PAH concentration correlations using only PAHs and the biological analysis was completed. A second series of correlations including the inorganic analytes and grouping the organics into either, aliphatics, aromatics or PAHs was also completed.

## 7.4 Results

Soils and made ground from contaminated and uncontaminated sites were collected for chemical and biological analysis. Fifteen samples, from the gasworks sites, where it was unclear as to their history and contamination burden, were graded a zero.

### 7.4.1 Community Recovery

To assess the recovery and detection of the bacterial community the 16S rRNA gene was quantified using qPCR. Boxplots of copy number plotted against the sample grades are in Figure 7.1. Increasing contamination appeared correlated to lower 16S rRNA gene copy number. Sample material not from the Grade 1 (background/reusable material) had low (>$10^6$) copy numbers of the 16S rRNA gene. The additional kit blanks had no copies of the 16S rRNA gene above the limit of detection

Low recovery and detection of the soil community led to difficulty quantifying the functional genes. All functional genes were detected in the background sample set. There were four samples from the gas works sites which had no detectable copies, or the number of copies was below of the limit of detection for all four of the functional genes. While all bar four sites had copies of the SDIMO $\alpha$-subunit gene multiple sites had only detectable copies of either *pmoA* or *mmoX*. The E-values and $R^2$ values for the qPCR assays are in Table A.13 (Appendix A.5).

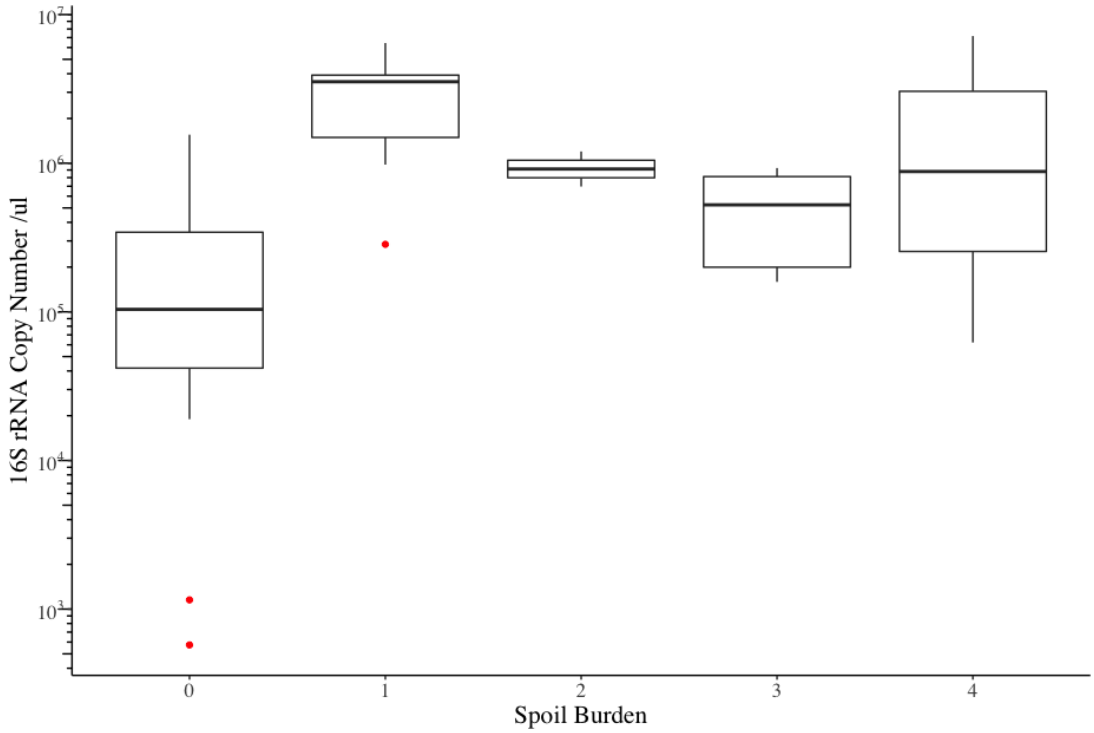



**Figure 7.1: Boxplots of 16S rRNA copy number as determined using qPCR.** Samples are separated by their estimated contaminant burden as described in Table 7.1. Group 0 contains samples whose background or characteristics were not well enough known to classify.

### 7.4.2 Sample Groups

A total of 50 samples covering a range of contamination burdens were collected and analysed. Four filter groups, each starting with all 50 samples, were produced by filtering samples out that exceeded a multiple of the PAH threshold concentrations given in S4UL suitable for a Public Park with 6% SOM. The number of samples for each group is given in Table 7.2.

**Table 7.2:** Number of samples in each filter group. Groups were generated using a multiple of the PAH concentrations given in the S4UL. The 'Grouping' column is the multiple used on the threshold concentration of the PAH. The largest group contained 46 samples meaning 4 samples were never used in the subsequent analysis.

| Filter Grouping | Sample Number |
|---|---|
| 0.5 | 21 |
| 1.0 | 29 |
| 2.0 | 35 |
| 10.0 | 46 |

### 7.4.3 Chemical Analysis

A large range in the concentrations of both organic and inorganic analytes was seen in the samples. One of the ten background samples had higher than expected concentrations of some analytes, with it carrying the highest chloride concentration of all the samples (1760 mg/kg). The concentrations of PAHs prior to filtering their range covered five to six orders of magnitude (mg/kg) (Figure 7.1).

### 7.4.4 Correlations

Correlations between the biological and chemical analyses were explored. Shapiro-Wilk tests indicated a non-normal distribution leading to the use of a Spearman's Rank correlation. The *pmoA* and SDIMO genes showed similar significant, negative correlations with the 16 individual PAHs, with increasing PAH threshold and more samples these relationships weakened (Figure 7.2). The obligate *mmoX* gene showed few significant interactions with no trends persisting through the four groups of threshold concentrations.

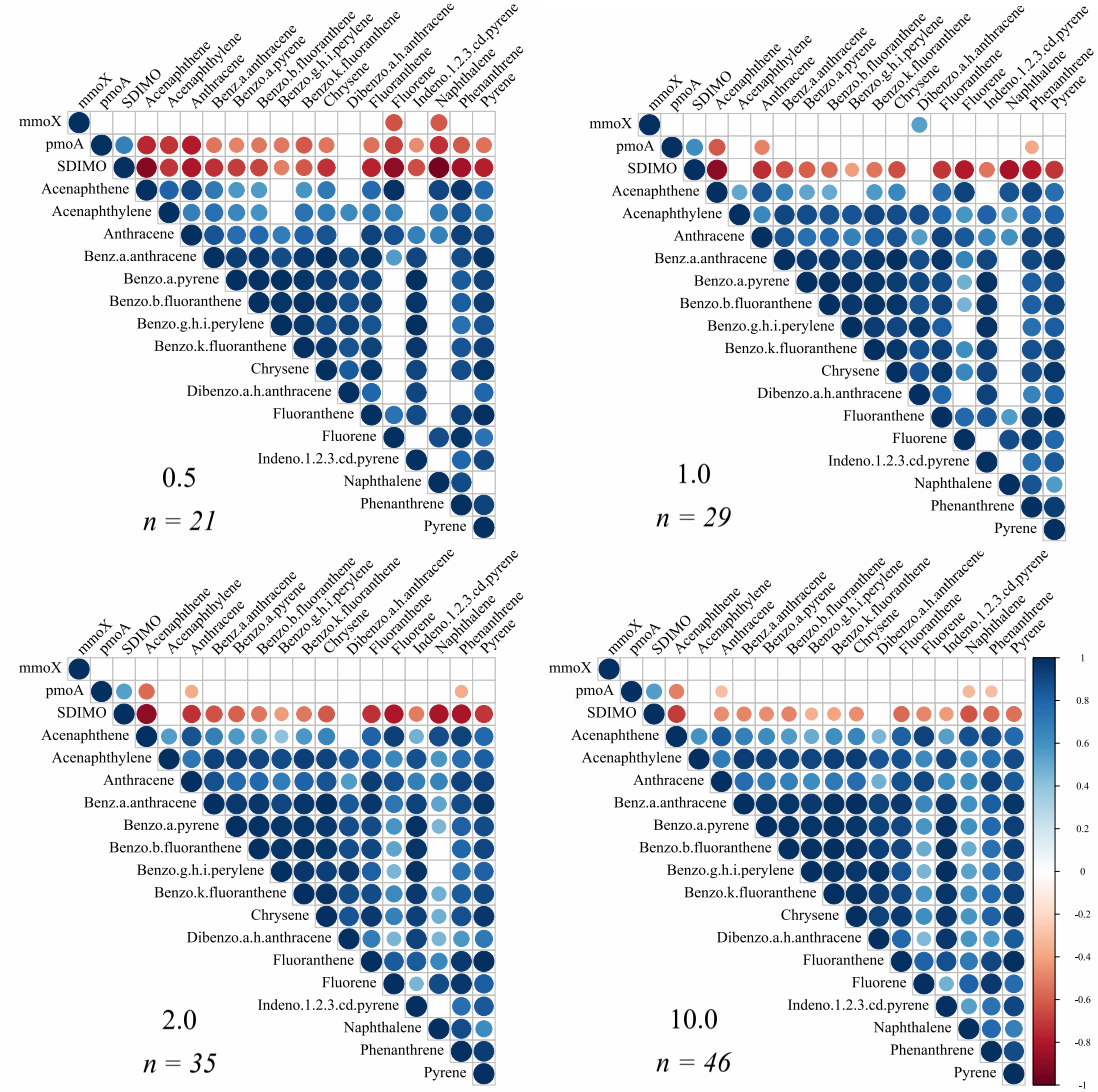



**Figure 7.2: Correlograms for each filter group against all 16 PAHs.** Blue is positive correlation, red negative, with colour value related to strength of correlation. Grouped by multiple of PAH threshold concentration. Only significant correlations are shown ($p < 0.05$)

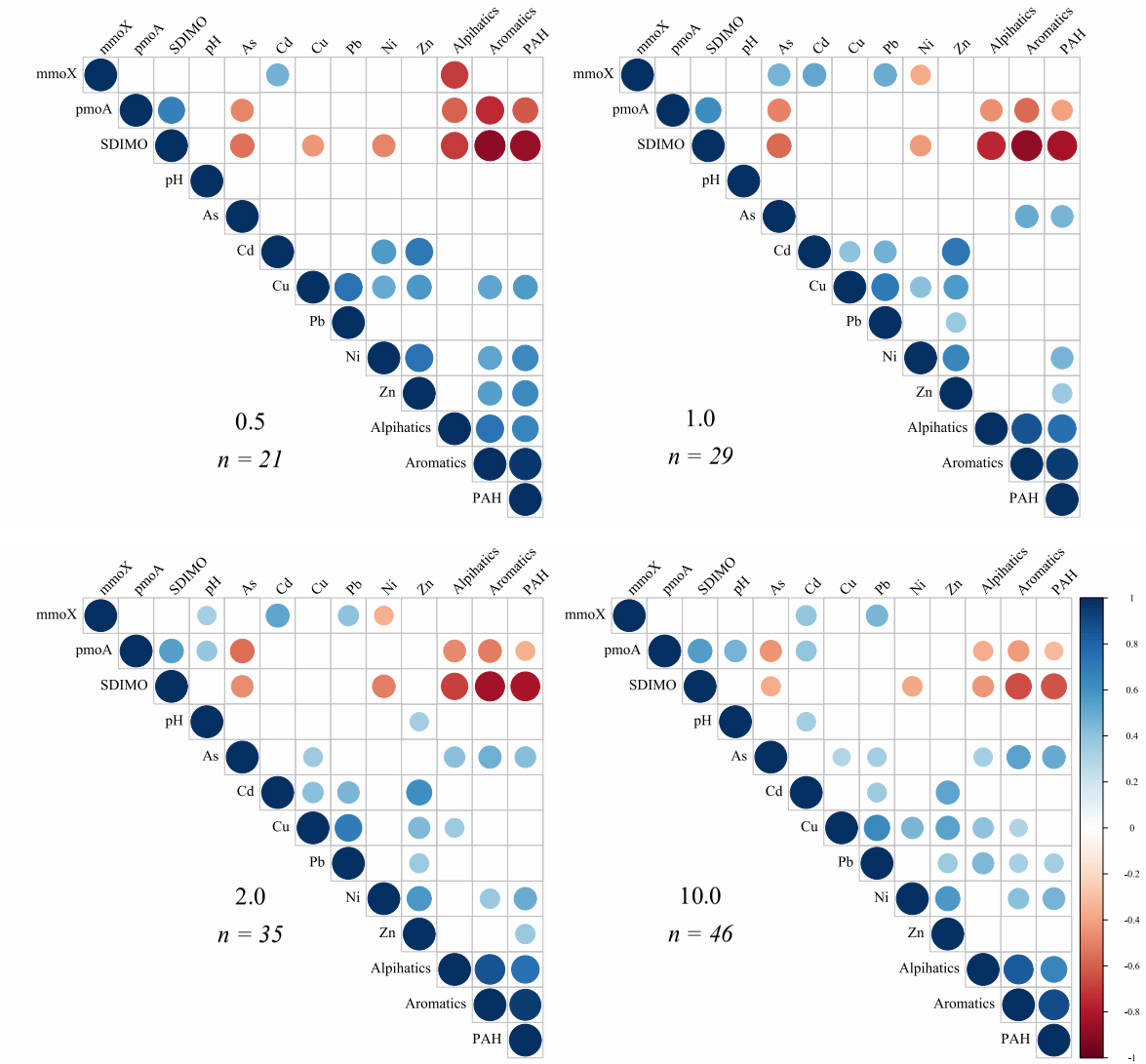



**Figure 7.3: Correlograms for each filter group against a selection of soil analytes.** Blue is positive correlation, red negative, with colour value related to strength of correlation. Grouped by multiple of PAH threshold concentration. Only significant correlations are shown ($p < 0.05$).

The second set of correlations including a broader group of analytes (Figure 7.3), showed a similar trend. The *pmoA* and SDIMO genes both showed a significant negative correlation with the organic analytes and arsenic. While the *mmoX* gene showed a negative correlation with the aliphatics in the 0.5 group but no correlation in the other groups. The *mmoX* gene was also predicted to have a significant positive correlation with cadmium and lead.

## 7.5 Discussion

To explore the interaction(s) between contaminants and the distribution of mono-oxygenase genes 50 samples of soil and made ground were collected. These were from a range of backgrounds covering a gradient of contaminant burdens. Filtering the data by PAH concentrations to try and separate sub-lethal effects and qPCR inhibition appears to have had limited success.

### 7.5.1 Community Recovery

The ability to extract and amplify DNA to explore the microbial community appeared to be hampered by contaminant burden (Figure 7.1). The difficulty amplifying the 16S rRNA gene indicated an issue in either the extraction or qPCR. It is likely it was a combination of both factors. The background samples amplified as expected with 16S rRNA gene copy numbers in excess of $10^6$ copies $\mu l^{-1}$. Low amplification could feasibly be due to DNA binding to either organics or clay during both extraction and amplification (Wilson, 1997). Conversely, it is also possible that these sites had an inherently small bacterial population due to the toxicity of the matrix. Given the apparent relationship between low amplification and high contamination it could be either, or both, of these scenarios.

The difficulty recovering the bacterial community makes drawing strong conclusions difficult. As the PAH thresholds were increased, enabling a larger sample size to be used, the number and strength of correlations decreases. This is likely due to a loss of data quality owing to a loss in community recovery which weakens the ability to draw out correlations. Samples with a greater toxicity often possessed copies of the mono-oxygenase genes below the limit of detection. Therefore overall, whilst the number of samples increases, the ability for that data to answer the questions posed decreased as it was impossible to tell if the loss in community size was due to toxic effects or due to technical limitations in the laboratory.

### 7.5.2 Metals

Arsenic consistently displayed a negative correlation through all filtering groups with *pmoA* and SDIMO- genes (Figure 7.3). Arsenate (As[V]) is well known toxin; the ion can displace phosphate in a variety of biochemical reactions while arsenite (As[III]) is also known to bind to thiols and interfere with biochemical reactions (Hughes 2002). Why significant negative correlation between arsenic and *mmoX* was not observed is unknown.

As *mmoX* is a specific SDMIO the two were expected to have similar responses to soil contaminants. As a result the striking differences in correlations with metals is hard to explain. Inoue *et al.* (2020) demonstrated that some strains possessing 1,4-Dioxane degrading enzymes (an SDIMO) had differing responses to metals. Their work demonstrated that Cu(II) had the most inhibitory effect, while Ni(II) and Zn(II) also inhibited activity. Pornwongthong *et al.* (2014) only reported a significant impact of copper on 1-4-dioxane degrading SDIMOs, while Cd(II) and Zn(II) had no significant impact upon activity. Here nickel was seen to negatively correlate with both SDIMO and *mmoX*. While copper was only predicted to negatively correlated with SDIMO in the filter group with the lowest PAH concentrations. It is possible the *mmoX* possessing microorganisms have distinct traits for dealing with metal induced stress but no literature was found to support this.

An increased concentration of metals has been repeatedly observed changing soil microbial community composition and total bacterial community size ((as reviewed by: Tang *et al.*, 2019)). Reductions in total community size, or microbial biomass, has been suggested as a useful bioindicator for metal contamination (Zhang *et al.*, 2010). This work did not explore the broader community composition and using relative abundances to draw correlations deliberately avoids using total bacterial size (Chapters 2.3.7.4 and 3.2.3). But, a loss of microbial biomass aligns well with the low 16S rRNA gene copy number detected in the more contaminated soils (Figure 7.1). If a relatively uniform loss in community size occurred due to metal, then the use of relative abundances would not display any meaningful correlations as the percentage of the bacterial community with the MMO genes would not alter greatly.

However, a uniform loss of the entire community was not expected. Berg *et al.* (2012), using 16S rRNA gene sequencing, reported increases in relative abundance for some taxa in the presence of copper, however a negative correlation in the relative abundance of $\gamma$-proteobacteria (containing the Type I methanotrophs) was also reported. Here, it had been expected that the combination of increased bioavailable metals and readily

available organic substrates would promote certain taxa. Copper was expected to negatively correlate to *mmoX*, given the inhibition of SDIMOs by Cu(II) and the copper mediated switch to the expression of *mmoX* in low copper to biomass environments (Semrau *et al.*, 2013; Csáki *et al.*, 2003). Therefore, low copper conditions might have been expected to drive an increased relative abundance of *mmoX*.

Alongside changes in community size and composition, metals been correlated to losses in enzyme activity (Kızılkaya *et al.*, 2004) (Kızılkaya et al. 2004). Zhang et al. (2010) reported that increasing bioavailable heavy metals concentration were linked to a decreased activity of a variety of enzymes, however those relationships were not as clear when using total metal content (the data here used total metal content).

Cumulatively, the presence of metal contamination likely contributed to losses in total bacterial size. This likely obscured any changes in relative abundances whilst also making the detection of MMO more difficult. The presence of metals in these soils might also limit any degradation of organic contaminants by reducing the activity of the bacterial population. The predicted correlations between metals and *mmoX* did not appear to agree with the broader SDIMOs.

### 7.5.3 Organics

The lack of significant positive correlations between the mono-oxygenase genes and the organic analytes suggests the hypothesis was wrong. It was thought that bacterial communities possessing SDIMO, pMMO, or sMMO would respond positively to increased organics until the concentration of those organics reached a toxic threshold. Here, even in the sample group with the strictest PAH thresholds, no positive correlation was seen. This could have been due to synergistic effects of both the organic and inorganic compounds in the soil eliciting a strongly toxic effect. Interestingly, the *mmoX* again was predicted to have very different correlations with compared to the broader SDIMOs.

Here the relative abundances of SDIMO were significantly and negatively correlated with the organic contaminants (Figures 7.2 and 7.3). The initial oxidation of aromatic rings such as those in PAHs is usually driven by dioxygenases ((as reviewed by: Andreoni and Gianfreda, 2007)). Dioxygenases lead to two oxygen atoms being added to the substrate, whilst mono-oxygenases only add one. The phenol and toluene-benzene mono-oxygenases are notable exceptions, both SDIMOs (Leahy *et al.*, 2003). While the initial stages of aromatic structures may not always involve the mono-oxygenases it was thought that the broad substrate specificity might lead to them being involved in

later stages of biodegradation.

Previous authors have observed potential biodegradation of contaminants at former gasworks. Ferguson *et al.* (2007) observed that soil and groundwater samples had numerous heterotrophic bacteria of which some appeared to grow on naphthalene. A mass-balance approach at a gasworks site, measuring the change in benzene, toluene, ethyl-benzene, xylene (BTEX) and PAHs along a groundwater plume, also suggested that biodegradation was taking place via anaerobic iron and sulphate reduction as well as potential for adsorption also along the plume (Bockelmann *et al.*, 2001).

Soils can develop bacterial communities capable of degrading complex organic contaminant mixtures (for example: Greenwood *et al.*, 2009) and Hamamura *et al.* (2006)). Indeed some authors have reported a stimulatory impact, Bundy *et al.* (2002) observed an initial spike in microbial biomass after diesel addition. However the addition of complex hydrocarbons to soils can lead to a loss in microbial diversity and activity (Klimek *et al.*, 2016; Labud *et al.*, 2007). This can lead to a community dominated by members that can make use of the substrates present, for example the Psedomonas genus in oil contaminated soils (van Dorst *et al.*, 2014). However, the same contaminant mixture can elicit a different response in microbial communities in different soils, suggesting that soil type is also an important factor in the potential degradation of organic compounds (Hamamura *et al.*, 2006).

The dataset suggests that the microbial community at these gasworks did not contain mono-oxygenases. Furthermore, it suggests these enzymes are not part of the communities that might be biodegrading contaminants in these soils and made ground. Organisms possessing mono-oxygenases might not be able to cope with the toxicity of the organic compounds present at these sites, in these soils.

### 7.5.4 Future Work

This work suggests that the mono-oxygenases are not abundant in the microbial communities at former gasworks sites. That said, the low community recovery and detection limits the certainty of this statement. It is plausible that the problems faced were due to an inherently high toxicity resulting in a low community size, the filtering approach using different PAH concentrations enabled a slightly better exploration of the data providing more confidence in the correlations seen. However, future work needs to better extract and amplify the DNA from these soils and made ground, this might be achievable by pooling and concentrating DNA or using different extraction techniques. Given the issues with low community recovery focussing on sites with a lower total

burden or a less complex mixture of contaminants might help better understand the ecology of contaminated sites.

If possible, more samples including samples from material where attenuation is known to be occurring would help develop understanding of these communities. For example, the inclusion of samples from soils from various sites undergoing composting to reduce contamination might highlight which microbial guilds predominate under different mixtures of contaminants. Another approach might explore the community, before, during and after composting.

Exploration of other genes involved in natural attenuation might enable a greater understanding of the communities at these sites. As touched upon, the dioxygenases might be group of genes worth exploring in future work. A broader approach using 16S rRNA gene sequencing to explore overall shifts in community composition might also highlight future avenues for research.

## 7.6 Conclusions

Given the lack of significant positive correlations between the organic contaminants and these genes the null hypothesis must be accepted. Organisms possessing these mono-oxygenases do not appear to be a part of the communities degrading these contaminants. This meant that an exploration of a toxic threshold was not explored as no stimulation was seen in any of the filter groups. Future work should explore other metabolic pathways, sites with a lower contaminant burden, and attempt to improve community recovery and detection.

### 7.6.1 Commercial Implications

This work was sponsored by National Grid Property Holdings. It had been hoped this work might highlight some key microbial indicators of biodegradation. These could then be further explored in future research. Unfortunately, it would appear that at these sites the mono-oxygenases are not suitable indicators. As discussed, an exploration of soils with a lower or less complex burden might develop a better understanding of biodegradation. Whilst this work suggests that in-situ biodegradation involving the mono-oxygenase genes is unlikely to occur at these sites, it does provide guidance for future work to look at other metabolic pathways and/or to focus upon sites with much lower contaminant burdens.

# Chapter 8

# Data Synthesis and Discussion – Comparing Relative Abundances of Methanotrophy Genes

## 8.1 Rationale

This chapter aims to draw together the datasets presented in the previous chapters and analyse them as a whole. This hoped to tease out any interactions and correlations between the MMO genes. Of particular interest was how the MMO genes appear to consistently show spatial separation of their relative abundances. Plus relative abundance was seen to differing responses when in the presence of a methane pulse event. This suggested that organisms possessing these genes might have differing life strategies.

## 8.2 Introduction

Methanotrophs are considered to have either obligate or facultative lifestyles. The majority of methanotrophs possess the enzyme particulate methane mono-oxygenase (pMMO) (McDonald *et al.*, 2008). A small number possess both pMMO and a soluble version (sMMO). Methanotrophs with pMMO are generally considered obligate, i.e. they can only grow on methane. The conserved catalytic subunit of both these enzymes, *pmoA* for pMMO and *mmoX* for sMMO, can be detected and quantified in samples using qPCR with a distinct primer set for each gene. Two genera, Methylocystis possessing both pMMO and sMMO and Methylocapsa, possessing only pMMO, go against this trend with slow growth on acetate and ethanol reported likely as a survival strategy in low methane conditions (Belova *et al.*, 2011; Im *et al.*, 2011; Dunfield *et al.*, 2010).

Two genera, both from the Beijerinckiaceae family, *Methylocella* and *Methyloferula*, only possess sMMO (Vorobev *et al.*, 2011; Dedysh *et al.*, 2000). *Methylocella spp.* can grow on a variety of substrates not just methane, i.e. they are facultative. It is possible that the facultative methanotrophs have developed methane oxidation as a new metabolic strategy. In other words, the facultative methanotrophs evolved a novel sMMO or acquired it through horizontal gene transfer from an obligate methanotroph (Leahy *et al.*, 2003). Whichever means is correct, the *mmoX* gene within *Methylocella* requires a distinct PCR primer set to quantify it accurately (Farhan Ul Haque *et al.*, 2018; Rahman *et al.*, 2011a). The *mmoX* primer set for the obligate methanotrophs does not amplify it. This allows for some distinction between the obligate and facultative methanotrophs, and their relative abundance, using qPCR.

It is assumed that possessing a trait such as methane oxidation, provides a competitive advantage in certain environments. Therefore, the ability for *Methylocella* to oxidise methane alongside other carbon compounds appears to be a clear advantageous trait against methanotrophs that can only grow on methane. However, the advantage of possessing two unique analogous enzymes, the pMMO and sMMO, is unclear although it is possible that the differing kinetics and metalloenzymes might enable a greater versatility for those organisms which possess them.

A consistent theme in the data collected throughout the thesis is the separation of the MMO genes. While all three primer sets usually detected their target gene in a single sample, one primer target is more prevalent with the other two appearing as a small proportion of the bacterial population. Indeed, in Chapter 6, it appeared that a relative increase in *pmoA* was linked to a relative decrease in both versions of the

*mmoX*. Here, the data collected throughout the thesis is brought together in one data set. This is used to explore the relationship between the relative abundances of facultative methanotrophs (in this chapter considered to be represented by *Methylocella mmoX*) and obligate methanotrophs (in this chapter considered to be represented by the summed relative abundances of *pmoA* and *mmoX*). The relationships of the *pmoA* and 'obligate *mmoX*' are also considered separately.

### 8.2.1 Chapter Objectives

This chapter aimed to explore:

- The spatial separation and of the MMO genes using the data collected through the thesis.
- To assess the implications this has on our understanding of the obligate and faculative lifestyles seen in methanotrophs.

## 8.3 Data Analysis

To explore the potential relationships between obligate and facultative methanotrophs the data from the preceding chapters will be used; this includes the Chapter 4, 5, and 6 data. The Contaminated Land work (Chapter 7) will be excluded due to the difficulties interpreting results. Data generated from the preliminary work has also been excluded to ensure all results are derived from the same assays and techniques.

The data was pooled producing 98 samples. Two correlations were completed: first all functional genes were treated separately. Secondly, the obligate *mmoX* and *pmoA* genes were summed to give an 'obligate' group. Using both primers sets limited the risk of methanotrophs being missed through a lack of coverage within on primer set. The obligate group underwent a correlation to the facultative *Methylocella mmoX* specific gene. Relative abundances were tested for normality using Shaprio-Wilk test.

### 8.3.1 Distribution

A large range in the relative abundance of *pmoA* was observed from $0\text{-}3x10^{-1}$. The highest abundances were seen in the La Rochasson (LR) samples and, as a result, this site was excluded from further analysis. The removal of LR also removed the predicted outliers from the facultative *mmoX* dataset (Figure 8.1). After removal of LR a total of 86 samples were available. The Shaprio-Wilk test suggested that only

the facultative (*Methylocella mmoX*) was normally distributed; as a result Spearman's Rank Correlation was used for the correlations.

Fig:8.1.A

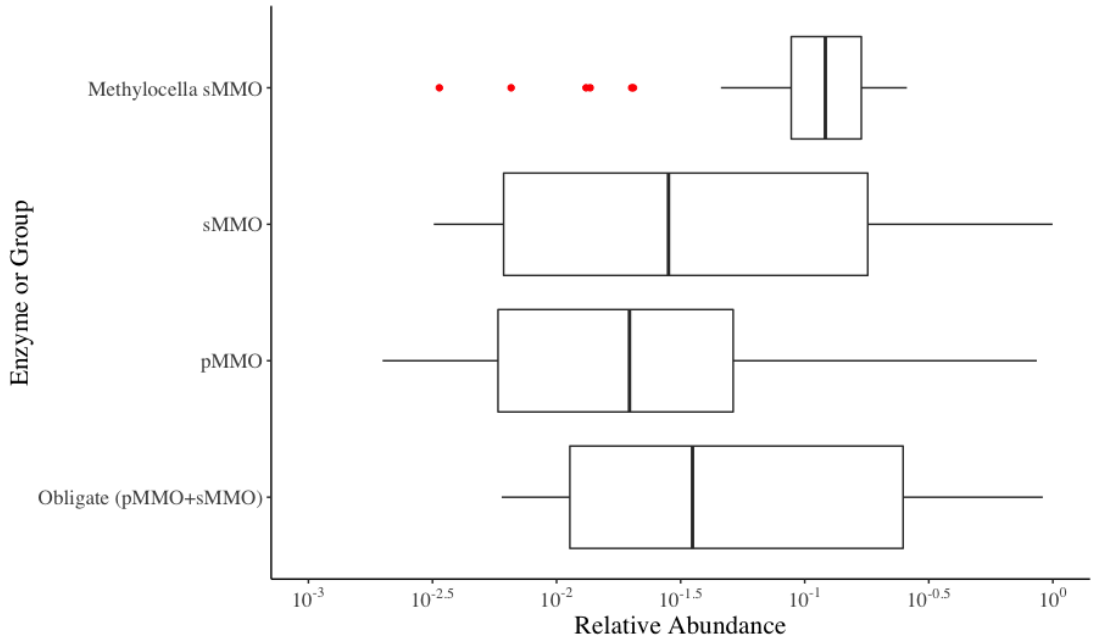



Fig:8.1.B

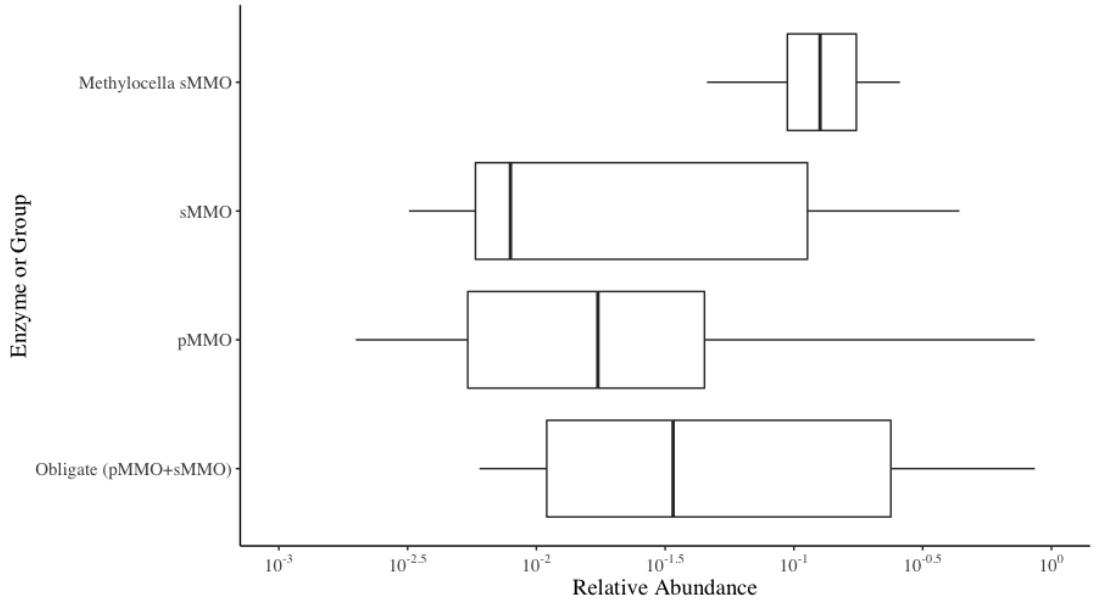



**Figure 8.1: Relative abundances of pMMO (*pmoA*), sMMO (*mmoX*) and *Methylocella* sMMO to 16S rRNA copy number.** The Obligate group is the sum of the *pmoA*, *mmoX* and is only an estimate as some species will have been counted twice. Furthermore, this plot does not include sites where the gene was below the LoD of the assay. **A)** shows the distribution before La Rochasson was removed. Red dots indicate outliers, as predicted using the inter-quartile range. **B)** Shows the distribution after La Rochasson was removed.

### 8.3.2 Correlations

Results for both sets of correlations are shown in Table 8.1. The summed relative abundance of obligate genes were negatively correlated with the abundances of facultative *Methylocella mmoX*. When correlated as three distinct assays (*pmoA*, *mmoX* and *Methylocella mmoX*) a significant negative correlation persisted between *pmoA* and *Methylocella mmoX*. There existed a positive correlation between the relative abundance of obligate *mmoX* and *pmoA*.

**Table 8.1: Summary of the Spearman Rank Correlations for the three primer targets.** Facultative is *Methylocella mmoX*. $r_s$ : Spearmans Rank Coefficent to three significant figures. Significance was accepted if $p$<0.05

| | $r_s$ | $p$ |
|---|---|---|
| Obligate ∼ *Methylocella mmoX* | -0.365 | <0.001 |
| *pmoA* ∼ *Methylocella mmoX* | -0.346 | =0.001 |
| *mmoX* ∼ *Methylocella mmoX* | -0.185 | 0.085 |
| *pmoA* ∼ *mmoX* | 0.508 | <0.001 |

## 8.4 Discussion

A cumulative analysis using the relative abundances of the methanotrophy genes was completed. All data was generated during the work on this thesis. The resulting correlations suggest that the relative abundance of the obligate methanotrophs is negatively correlated with the Methyolocella, a genus of facultative methanotrophs.

### 8.4.1 Caveats

Before further consideration of the data developed here some caveats should be presented. Firstly, only one facultative methanotroph genus, *Methyelocella*, has been included. This has been detected using a primer set that is specific for the *Methyelocella mmoX*. The primer set for obligate *mmoX* appears unable to amplify the *Methylocella mmoX*. However, it is possible these *Methylocella* specific primers might inadvertently amplify the obligate gene, or vice versa, therefore it is possible that some overlap has occurred. There are three sequences recorded on the NCBI database (2021) for *Methyloferula mmoX*. Using *in-silico* PCR with these three sequences suggests that

the primers for *Methylocella mmoX* (Chapter 2.3.7.1) will not amplify the *Methyloferula mmoX* gene. Similarly, the primers used for the obligate *mmoX* also did not amplify the *Methyloferula mmoX* gene.

Another note of caution is the overlap with *pmoA* and *mmoX*. By combining the results from both assays, some organisms will have been counted twice; however, by doing this any organisms missed by one primer set are more likely to be counted. Here, the correlation and significance between the *pmoA* and *Methylocella mmoX* generated a similar result as the summed obligates (*pmoA+mmoX*) and *Methylocella mmoX*. It is interesting that the obligate mmoX was not significantly correlated with the *Methylocella mmoX*; however, this fits with the repeated observations in this thesis that the relative abundance of *pmoA* is more responsive to the presence of methane seen in Chapter 4 (Fontaine Ardent Data) and Chapter 6.

The removal of the La Rochasson dataset was done with caution. Whilst these were extreme observations, they were still 'real world' and therefore, whilst unusual, they still represented an extant field community that could further develop understanding of community dynamics. However, it could be argued that the material was not soil (as discussed in Chapter 4.3.2 and 4.5.1) and the increased from a mean relative abundance of *pmoA* from $7.2 \times 10^{-2}$ (without LR) to $6.8 \times 10^{-1}$ (with LR), thus skewing the data severely.

### 8.4.2 Competition Based Interactions

The negative correlation observed between the relative abundance of *pmoA* and *Methylocella mmoX* methanotrophs could be due to competition between the two groups for resources. As seen in Chapter 4 and 6 *pmoA* tends to increase in relative abundance when the environment has a greater methane flux/concentration. This increase is seen without an increase in the relative abundance of *Methylocella mmoX*. Using the bioinformatics dataset from Chapter 6 , at a family level, the methane mesocosm saw an overall increase in the raw counts of methanotrophs but this was almost entirely from Type I group which is thought to be exclusively obligate methanotrophs.

It is possible the pMMO possessing organisms can exploit the resource more readily and/or efficiently. Lee *et al.* (2006), when exploring pollution degradation by MMOs, observed sMMO to have a lower affinity and specificity compared to pMMO, whilst having a higher $V_{max}$. The authors went on to hypothesise that despite sMMO being faster at methane oxidation, the lower affinity leads to the enzyme interacting more frequently with co-metabolites producing a less efficient process overall. This was

supported in subsequent work by Yoon and Semrau (2008) Y; however, they observed pMMO to have the higher $V_{max}$ at lower temperatures (20°C). This might have enabled the obligate pMMO possessing methanotrophs to outcompete the *Methylocella spp.* thus enabling a relative increase in the obligate population. This is supported by both the correlations presented here, the heatmaps from the Fontaine Ardent (Figure 4.3) and qPCR results from Chapter 6 (Figure 6.4).

For the Eakring data, there was no clear evidence of soil methane flux. Only *Methyolcella mmoX* was observed in the deeper samples. Given the combination of low oxygen diffusion and limited methane concentration it is unsurprising that the aerobic methanotrophs were not present, but it highlights a scenario where the *Methylocella* genus had a competitive advantage compared to the obligate methanotrophs. The ability of *Methylocella spp.* to grow on a selection of substrates allows them to persist in an environment where the obligate methanotrophs struggle. It would have been interesting to better characterise the pore gases, fluxes, and seasonal changes to understand why the obligate methanotrophs were not present in such a wet environment.

The data collected through the thesis paints a picture of a low abundance widespread *Methylocella* population. This population is present in most soils sampled at a low relative abundance but is present in a variety of soil types and reducing conditions, this is supported by Rahman *et al.* (2011a). The *Methylocella* population does not rapidly increase in the presence of methane. The lack of relative abundance change appears linked to an increased relative abundance of the obligate methanotrophs which are possibly better competing for the methane. It should be noted that even though these methanotroph populations have been detected, there is no data on their activity as DNA, not RNA, was extracted and analysed from these samples.

The relationship between *pmoA* and *Methylocella mmoX* has been described in a previous study. Farhan Ul Haque *et al.* (2018) reported that *pmoA* formed a larger proportion of the methanotroph community at sites with a biogenic methane source. However, at sites where a thermogenic source was present the *Methylocella* specific *mmoX* comprised 60-85% of the methanotroph population. The authors suggest that, at seeps with other alkanes present, *Methylocella spp.* are more competitive compared to the obligate methanotrophs as they have a greater pool of resources from the seep. Subsequent work observed, using 16S rRNA gene sequencing, that at natural gas sites the *Methylocella* genus had a greater relative abundance at a site with higher ethane and propane composition (Farhan Ul Haque *et al.*, 2019). The authors suggested that metabolic versatility of the *Methylocella* genus provided a competitive advantage at thermogenic

sites. Interestingly this study also observed *Methylococcus* at greater than 5% relative abundance, a far higher abundance than any data collected in this thesis using 16S rRNA gene sequencing.

### 8.4.3 Limited Competition Interactions

A simple competition for resources model is potentially unsuitable with microbial communities in soil systems. Spatial separation and a lack of connectivity between micro-habitats (<1 mm size) raise questions about the feasibility of direct competition and potentially explain how such large diversity can be seen in a soil environment (Zhou *et al.*, 2002; Young *et al.*, 2008). It is possible to review the data already presented in way that precludes direct competition between methanotrophic guilds.

Moisture content, pore characteristics (size, shape etc.) and gas diffusion are inextricably linked (White, 2006; Killham, 1994). At La Fontaine Ardent, concentrations of both $O_2$ and $CH_4$ were predicted to have a significant interaction with community composition (Figure 4.8). Data on porosity and bulk density were not collected, but the soil porosity, and therefore gas diffusion, is likely to have been improved by the presence of gravel despite the clay and sand rich soil matrix. Overall, at FA, sampling sites with a higher relative abundance of *pmoA* also had higher $CH_4$ concentrations. This was coupled with lower $O_2$ concentrations. Therefore, it is plausible that obligate methanotrophs were blooming and rapidly exhausting the $O_2$ supply.

By adapting the distinctions from Langer *et al.* (2004) the obligate methanotrophs can be described as a zymogenous population, or r-strategist. Here, zymogenous describes a bacterial community whose activity and population size are transient in the soil; whilst they may not disappear from the habitat when sub-optimal conditions are encountered their population will crash, and the remainder enter a near dormant state. For obligate methanotrophs their population blooms with the presence of methane while shifts in their population composition possibly being driven by factors like copper concentrations and pH (Semrau *et al.*, 2013; Knief *et al.*, 2005) (e.g. FA-1 versus FA-66 – Figure 4.3).

The facultative *Methylocella* genus could be described as autochthonous population or k-strategist. An autochthonous population is one which maintains a constant presence in the soil habitat with a relatively stable population size, these organisms might use a diverse range of complex substrates. There is a more or less uniform relative abundance of *Methylocella mmoX* within all datasets. This is demonstrated in Figure 8.1 where the gene has a small range in relative abundances compared to the obligate

genes. Intriguingly it has a very low population size, for example *Methylocella's* family (Beijerinckiaceae) has never had a relative abundance of greater than 1% in any bioinformatics dataset. As observed here, this relative population size appears to be inversely correlated to that of the obligate methanotrophs.

It is suggested that *Methylocella spp.* are not competing for methane; instead, a small population is consistently present being capable of growth in a wide array of soil conditions on a selection of carbon substrates including methane. Furthermore, these two populations could be spatially separated and not directly competing. Under this hypothesis the only interaction potentially seen in this thesis is at FA, where the limited availability of $O_2$ at FA-66 might be a driver of competition. It is possible that growth on methane is merely a survival strategy for *Methylocella spp.* in sub-optimal conditions where other preferred substrates are limited. Studies have demonstrated that the presence of acetate downregulates the expression of *mmoX* in *Methylocella* (Rahman *et al.*, 2011b; Theisen *et al.*, 2005). Interestingly the opposite is seen in the *Methylocystis* genus where acetate utilisation is thought to be a survival strategy (Belova *et al.*, 2011; Im *et al.*, 2011).

The data from LR supports the concept the *Methylocella* genus is not competing for methane. Here, a landslip has probably reduced the soil community size and the labile nutrient pool (Chapter 4.3.2 and 4.5.1). Fast growing populations of obligate methanotrophs were present, exploiting the readily available methane. While *Methylocella spp.* could (and were expected to) be using methane, their proposed slower life strategy has meant the population had not grown at the rate of obligate methanotrophs; indeed, their population was below the LoD of the assays used here at the head of the landslip. Again, at LR the overall range in relative abundances of *Methylocella mmoX* is small compared to the other MMO genes studied (Figure 4.4).

Ho *et al.* (2013) in their review used the competitive-stress-ruderal (CSR) framework to classify and describe the life strategies of methanotrophs. This does not completely align with the description presented above but there is broad agreement, the authors split the methanotrophs by Type (i.e. Type I and II). *Methylocella spp.* are Type II methanotrophs. Interestingly, *Methylocystis* and *Methylocapsa* (both also having facultative traits) are also Type II. The Type I were described as responding rapidly to substrate availability and being competitors or competitor-ruderals, similar to the zymogenous definition. The Type II were stress tolerant, stress tolerant-ruderals or stress tolerant-competitors, persisting in inactive states and becoming active after disturbances. This does not completely fit with the description offered here as it has been

assumed that whilst at low relative abundances *Methylocella* (stress-tolerant only) was still active, but no data was gathered in this thesis to measure *Methylocella spp.* activity.

A further weakness in this description is the kinetics of the MMOs since, normally, the high affinity enzyme would be associated with the slower autochthonous population (Killham, 1994). As discussed above (Chapter 8.4.2), the pMMO is the higher affinity enzyme. In the model proposed here the zymogenous population uses the pMMO. Despite the *Methylocella* genus having the metabolic tools to use methane, those tools might not allow its population to grow as quickly as the more constrained obligate methanotrophs using *pmoA*. This was seen in the LTE Chapter where the relative abundance of *pmoA* and relative population size of the Type I methanotrophs increased compared to *Methylocella mmoX* and the Beijerinckiaceae.

A second weakness in this hypothesis is the persistence of *pmoA* seen in the Chapter 6. If the *pmoA* possessing population is more transient, with population boom and collapse dependent upon methane flux then the population and its activity should have fallen dramatically once flushing ended. As discussed in that chapter, it is possible that the incubation in the laboratory contributed to the stability of the microbial population. This might have artificially extended the presence of the *pmoA* possessing methanotrophs and needs further exploration using samples stored in the field.

## 8.5 Conclusions

One recurring question throughout the thesis has revolved around the disconnect between the relative abundances of *pmoA* and obligate *mmoX*. In the FA dataset and the LTE *pmoA* and *mmoX* appeared to respond independently. The analysis here suggests that the two genes are significantly and positively correlated; however, this relationship could be driven by the impact of obligate *mmoX* never being present without *pmoA* therefore they will always be related.

A significant, negative correlation between the abundance of *Methylocella* genus of facultative methanotrophs and the obligate methanotrophs was seen. The mechanism driving this relationship is not clear, but it could be due to either direct competition and/or differing life strategies. Further work is needed to better understand these relationships and the advantages for possessing *pmoA* and *mmoX*.

This analysis still does not address why some organisms possess both enzymes, why

distinct hotspots are present (FA) or why the relative abundance of *mmoX* did not increase in the LTE soils. The evolutionary need for two analogous enzymes could be to improve versatility but no evidence was collected in this thesis to explore this.

# Chapter 9

# Discussion

## 9.1 Introduction

An in-depth analysis and discussion of the data collected throughout the thesis is presented in Section 8. Here a final review of the data is presented, how it addresses the original hypotheses and objectives. Finally, suggestions for future work are made.

### 9.1.1 Community Change

Changes in the soil methanotroph community were seen with changes in methane flux. A consistent theme throughout the thesis is an increase in the relative abundance of *pmoA* in response to increased methane concentrations, observed in both Chapters 4 and 6. That said, it was unusual for the methanotrophic community to be more than 1% of the total bacterial community, as seen in the bioinformatics data and estimated by qPCR. Only La Rochasson saw bacterial communities with a methanotrophic proportion greater than 5%. Estimates of the size of the methanotrophic community from previous work seem limited but Farhan Ul Haque *et al.* (2019) also reported methanotrophs making up more than 5% of populations around a natural gas seep. As discussed in Chapter 8, the increase in *pmoA* appeared to be negatively correlated to *Methylocella mmoX* abundance. The amplicon sequencing data from Chapter 6 suggested that methane has a limited impact on the broader microbial community; shifts in composition were limited to families involved in methane oxidation. The change in methanotroph abundance and composition persisted for the remainder of the experiment.

### 9.1.2 Detection Using qPCR

The precision of qPCR has allowed spatial analysis of soil communities across tens of metres to be completed. Soil samples posed challenges for qPCR during both DNA extraction and detection. To improve the robustness of the approach, a means of estimating and correcting for qPCR inhibition was developed during preliminary work and described in Chapter 3. Data handling techniques using relative abundances based on 16S rRNA gene copy number, interpolations and error propagation were trialed with the aim of making the analysis as robust and repeatable as possible.

### 9.1.3 Amplicon Sequencing

To further understand the soil community as a whole, 16S rRNA amplicon sequencing techniques were used. Given the need for PCR in the setup of this analysis, inhibition estimation was useful. The use of the Oxford Nanopore Technology platform presented challenges in the data analysis that required a custom pipeline. That said, the work presented demonstrates the potential of the ONT platform in exploring soil microbial communities.

### 9.1.4 Broader Literature

This work continues from the culture-based methods for indirectly detecting natural gas flux using methanotrophs for hydrocarbon prospecting (Wagner *et al.*, 2002; Tedesco, 1995; Brisbane and Ladd, 1965). More recent work has started using molecular techniques, like those developed here; however, these have been over larger spatial scales and included propane oxidation and soluble di-iron mono-oxygenase genes, as well as *pmoA* (Zhang *et al.*, 2017; Liu *et al.*, 2016; Zhang *et al.*, 2014; Miqueletto *et al.*, 2011). These studies have also seen increases in alkanotrophy genes and changes in diversity in samples taken above hydrocarbon reservoirs. Work on smaller scales, for example at natural gas seeps, has also seen similar results with molecular tools. Changes in the community composition of methanotrophs altering depending upon the source of the methane emission and with an active methanotrophic community present (Farhan Ul Haque *et al.*, 2019, 2018). However, it was interesting to observe the ubiquity and stability of the *Methylocella* genus in the soils sampled, even in those soils without an obvious alkane flux and the populations apparent stability in the presence of a strong methane flux. Rahman *et al.* (2011a) also observed a widespread environmental distribution of *Methylocella*.

## 9.2 Further Work

The thesis has highlighted gaps in the understanding of methanotrophs, and their ecology. Further work improving methods and developing new sequencing approaches would also improve both understanding of these communities and any techniques used to monitor natural gas emission from terrestrial sources. If time had allowed more fieldwork at sites with a known thermogenic seep would have been completed. This work would have been completed in a similar manner to the other field sites, with a grid survey and spatial analysis using gene abundances, gas fluxes, pore gas concentrations and amplicon sequencing. This would aim to better understand community composition and the interactions between *pmoA* and *Methylocella mmoX*, differences in Type I and II and gain more information of the reliability of this approach in detecting and monitoring methane seeps. As a follow on, more information on soil properties and their interactions on methane oxidation and methanotrophs would have been developed to attempt to develop an improved understanding of how soil functions as a sink for methane.

### 9.2.1 Ecology

Throughout the thesis, questions surrounding the composition of methanotrophic communities have been addressed. For example, the apparent differences in the ecological niches of Type I and II was seen in the Chapter 6, with Type I being more responsive to the simulated pulse event. Also highlighted in that chapter was a lack of understanding of why some methanotrophs possess both pMMO and sMMO. Further work to attempt to understand the dynamics within methanotrophic communities is needed, both within the obligate methanotrophs and between the obligate and facultative methanotrophs. It would be interesting to use well characterised soils, seeded with a known methanotrophic community in microcosms, to explore how the starting community changes in response to the soil characteristics. Of particular interest would be pH, temperature and oxygen conditions and then copper, ammonium and nitrate concentrations. Hopefully, these would encourage the divergence of communities best adapted to the conditions posed by the microcosm and therefore the advantages of key traits could be characterised. Instead of using soil, an inert substrate mimicking soil, like rock wool, might enable even greater control by allowing the control of labile carbon sources. Both approaches might shed insight into the competition or lack of competition between these methanotrophs and provide opportunity to test the ecological models discussed in Chapter 8.

Future work at thermogenic sites needs to include other alkanotrophs, in particular propane and butane oxidisers. The measurement of these groups might aid the differentiation of biogenic to thermogenic seeps. Attempts were made during the preliminary work to use the primer sets developed by Cano et al. (2013) but due to time constraints (primarily Covid related) problems with PCR specificity in spiked soil samples were never overcome.

### 9.2.2 Method Improvements

Further development of the methods used to detect genes could be made. A repeated issue was DNA extraction and the co-extraction of inhibitors. Clays, humic acids and ethanol can also interact with DNA and the polymerase in PCR (Wilson, 1997). Steps to better control these problems could be explored, for example the use of polyethylene glycol (PEG) precipitation to further purify the DNA by precipitating it out and resuspending in clean solute in an effort to further limit any soil carryover. Work exploring how soil type and extraction techniques interact with qPCR precision and accuracy is needed to further develop the techniques used here. Furthermore, due to time and cost only one sample was taken for each sample point in the spatial survey grid: from this only one DNA sample was extracted. This potentially limits our understanding of bacterial diversity and may fail to account for errors arising during DNA extraction. Whilst the preliminary work attempt to best understand the sampling requirements it would be ideal to increase the number of DNA samples used in the qPCR.

To help meet the targets for a proposed 30% reduction in methane emission from COP26, the techniques used here could feasibly be applied to landfill sites which are major artificial sources of methane. This would help to monitor the emission of methane with reasonable precision and possibly develop a better understanding of how caps might be engineered to promote the greatest possible methanotrophic activity.

The data analysis for bioinformatics is another area of improvement. The choice of packages and the overall pipeline was carefully considered. The normalisation of the read data for subsequent analysis in Bray-Curtis would improve the relaibilty of the results and limit any random error being produced during sample handling and sequencing. As packages and tools for the ONT platform improve, so will the confidence in the results improve. A second area of improvement is the development of more specific environmental datasets for taxonomic assignment. Both of these areas are continuously developing and it would be interesting to revisit the data in the future and see if new analysis would provide a greater insight into community changes. Similarly,

amplicon sequencing of functional genes, for example *pmoA*, had been planned and would have provided a detailed insight of methanotroph community structure. Completing this analysis might assist in resolving the ongoing questions about Type I and Type II methanotrophs and the possession of both sMMO and pMMO.

## 9.3 Conclusions

The thesis has demonstrated that using qPCR to detect changes in soil methanotrophs is possible across tens of metres. Using qPCR to detect methanotrophs provides enough resolution to predict changes across tens of metres, this is most clearly demonstrated in Chapter 4. Improvements to sample handling and data analysis have been proposed and tested, these include methods to overcome inhibition. This supports the first half of the thesis's hypothesis. Changes in the relative abundance of soil methanotrophs appear related to changes in methane concentration. Using pMMO, detected using *pmoA* abundance, as an indicator of flux over other genes might be more useful when dealing with biogenic seeps given its consistent increase in the presence of methane. Further work is needed to confirm the use of these techniques at sites with a thermogenic seep to confirm if microbiology can predict the source of the seep accurately. Finally, broader work is needed to better understand the ecology and interactions of the methanotrophs.

# Appendix A

# Supplementary Information and Data

# A.1 Methods

## A.1.1 Plasmid Inserts Sent to Eurofins for qPCR Standards

### A.1.1.1 16S rRNA gene

Sequence taken from 315- 525 bases.

| ***Methylomonas rubra*** | **GenBank: AF150807.1** |
|---|---|

CGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATG
GGCGAAAGCCTGATCCAGCAATACCGCGTGTGTGAAGAAGGCCTGAG
GGTTGTAAAGCACTTTCAATAGGAAGGAATACCTACCGGTTAATACCC
GGTAGACTGACATTACCTATACAAGAAGCACCGGCTAACTCCGTGCCA
GCAGCCGCGGTAATACGGAGG

### A.1.1.2 *pmoA*

Complete Sequence Used.

| **Uncultured *Methylocystis sp.*** | **GenBank: MN394651.1** |
|---|---|

GGCGACTGGGACTTCTGGGTTGACTGGAAGGATCGTCGTATGTGGCCG
ACGGTCGTGCCGATTCTGGGCGTGACCTTCTGCGCGGCGTCGCAGGCG
TTCTGGTGGGTGAACTTCAAGCTGCCGTTTGGCGCTGTGTTCGCGGCTC
TTGGCCTTCTGATTGGCGAGTGGATCAACCGCTACGTCAACTTCTGGGG
CTGGACCTATTTCCCGATCAGCCTTGTGTTCCCGTCCGCTCTGATCGTT
CCGGCGATCTGGCTTGACGTGATCCTGCTTCTGTCGGGCTCCTATGTGA
TCACGGCGATTGTCGGTTCGCTGGGCTGGGGTCTTTTGTTCTATCCGAA
CAACTGGCCGGCGATTGCGGCGTTCCATCAGGCGACCGAGCAGCATGG
CCAGCTGATGACGCTGGCGGATCTGATCGGCTTCCACTTCGTCCGCAC
GTCGATGCCGGAATACATCCGCATGGTCGAGCGCGGCACGCTGCGCAC
CTTCGGTAAAGACGTTGCTCCGG

#### A.1.1.3 *mmoX*

Sequence taken from 720-1450 bases.

| ***Methylocystis sp.* M** | **GenBank: U81594.1** |
|---|---|

AAGGTCATCGCCAAGGAATACGCCCGTATGGAGGCGGCCAAGGACGA
GCGCCAGTTCGGCACTCTGCTCGACGGCCTCACCCGGCTCGGCGCCGG
CAACAAAGTGCATCCGCGCTGGGGCGAGACCATGAAAGTGATCTCGA
ACTTCCTCGAGGTCGGCGAGTATAACGCCATCGCCGCTTCGGCCATGC
TGTGGGACTCCGCCACCGCCGCCGAGCAGAAGAATGGCTATCTCGCGC
AGGTGCTCGACGAAATTCGTCATACGCATCAGTGCGCTTTCATCAATC
ACTATTATTCCAAGCATTATCACGATCCGGCCGGCCACAATGACGCCC
GTCGCACGCGCGCGATCGGCCCGCTGTGGAAAGGCATGAAGCGCGTCT
TCGCCGACGGCTTCATCTCCGGCGACGCCGTGGAGTGCTCGGTCAATC
TGCAGTTGGTCGGCGAAGCCTGCTTCACCAATCCGCTGATCGTCGCCGT
CACCGAATGGGCTTCGGCCAATGGCGACGAGATCACGCCGACGGTGTT
CCTGTCGGTGGAGACGGACGAGCTGCGCCATATGGCGAATGGCTATCA
GACGGTGGTGTCGATCGCCAATGATCCGGCGGCGGCCAAATATCTCAA
CACCGATCTCAACAACGCCTTCTGGACGCAGCAGAAGTATTTCACGCC
GGCCCTCGGCTATCTGTTCGAATACGGCTCCAAGTTCAAGGTCGAGCC
TTGGGTGAAGAC

#### A.1.1.4 SDIMO $\alpha$-subunit

Sequence taken from 3601 to 4067 bases.

| **Uncultured *Pseudonocardia sp.* clone** | **GenBank: MK370037.1** |
|---|---|

TCATGAACTCGGTGCAGTCAGACGAGGCGCGCCATATCAACAACGGCT
ACGCGACTCTCCTGTATCTACTTCAGGAGCCGGAGAACGCCCCACTAC
TGGAGCAGGACATCCAGCAGATGTTCTGGACCGTGCACGCGTTCGTCG
ACGCCTTCATGGGAATTCTCGTCGAGTACGCCCCGATTGATGCCACGG
ATCCAGAGAGCTGGACAGCGAAGTGGGACCGGTGGGTCCATGATGAC
TACTACCGTTCCTACATCGTCAATCTCGGGAAGCTCGGCCTGAAGATC
CCCGAGTCGATGTTCAAGCGCGCTCGGGAGCGCATCGCCGCGGACTAC
CATCACAAGGTCGCCATCGGAGTGTGGGCCTCCTGGCCCTTCCACTACT
ACAAGTCGGGAACCTCGGCGAGAAGGACTACGCCTGGTTCGAGAGCA
AGTACCCGGGCTGGAACGAGAAGTACGG

## A.2 Field Studies of Sub-Alpine Seeps:

**Table A.1: Relative abundance, estimated using qPCR, of methanotrophy genes**. These are expressed as percentages rounded to 2 d.p.

| | Gene | | |
|---|---|---|---|
| **Sample Point** | ***mmoX*** | ***pmoA*** | ***Methylocella mmoX*** |
| FA1 | 1.42% | 0.59% | 0.12% |
| FA2 | 0.04% | 0.60% | 0.07% |
| FA11 | 0.06% | 0.27% | 0.10% |
| FA16 | 0.04% | 0.07% | 0.14% |
| FA17 | 0.06% | 0.15% | 0.13% |
| FA22 | 0.11% | 0.45% | 0.12% |
| FA23 | 0.07% | 0.26% | 0.12% |
| FA36 | 0.04% | 0.23% | 0.13% |
| FA38 | 0.03% | 0.08% | 0.15% |
| FA40 | 0.05% | 0.21% | 0.17% |
| FA44 | 0.03% | 0.27% | 0.13% |
| FA59 | 0.14% | 0.45% | 0.15% |
| FA66 | 0.25% | 0.86% | 0.06% |
| FA70 | 0.17% | 0.50% | 0.09% |
| FA76 | 0.44% | 0.42% | 0.12% |
| FA83 | 0.03% | 0.01% | 0.15% |
| FA84 | 0.03% | 0.05% | 0.00% |
| FA85 | 0.02% | 0.00% | 0.07% |
| LR1 | 0.12% | 0.47% | 0.01% |
| LR2 | 1.00% | 6.33% | 0.01% |
| LR4 | 0.39% | 1.05% | 0.01% |
| LR6 | 0.12% | 8.07% | 0.02% |
| LR13 | 0.16% | 30.75% | 0.01% |
| LR14 | 0.00% | 0.05% | 0.00% |
| LR15 | 0.21% | 4.13% | 0.01% |
| LR17 | 0.24% | 0.67% | 0.01% |
| LR18 | 0.86% | 7.82% | 0.01% |
| LR21 | 0.08% | 1.68% | 0.03% |

**Table A.2: Efficiency and $R^2$values for the qPCR assays calculated from the slope of the standards.** Assays were deemed a success only if the efficiency value was between 90-110% and if the $R^2 \geq 0.95$

| | Gene | E-Value | $R^2$ |
|---|---|---|---|
| La Fontaine Ardent | 16S rRNA | 95.00% | 0.99 |
| | *pmoA* | 97.00% | 0.99 |
| | *mmoX* | 90.00% | 0.99 |
| | *Methylocella mmoX* | 98.00% | 0.99 |
| La Rochasson | 16S rRNA | 97.00% | 0.98 |
| | *pmoA* | 98.00% | 0.99 |
| | *mmoX* | 92.00% | 0.99 |
| | *Methylocella mmoX* | 90.00% | 0.99 |

**Table A.3: Pore gas measurements for FA.** Samples were collected and analysed by BGRM. $C_2H_6$ was observed from samples in lab while the others were measured in field. Values are reported as percent volume of pore gas. These values were used in IDW interpolation to provide estimate values for use in ordination with the 16S rRNA gene data.

| Point ID | $O_2$ | $CH_4$ | $C_2H_6$ |
|---|---|---|---|
| FA11 | 15.20% | 5.84% | 0.00% |
| FA22 | 16.90% | 5.00% | 0.00% |
| FA28 | 17.70% | 5.64% | 0.04% |
| FA42 | 21.00% | 85.00% | 0% |
| FA59 | 20.80% | 0.80% | 0.00% |
| FA65 | 12.50% | 0.60% | 0.06% |
| FA80 | 16.00% | 4.30% | 0% |

**Table A.4: Summary values for the 16S rRNA sequencing samples.** Only a subset of the Fontaine Ardent samples could be sequenced. The total number of reads and unique taxa, at the family level, classified by the in-house pipeline are presented here. Both the total number of unique families and number families above $1x10^{-3}$ relative abundance (RA) are presented here.

| | | Unique Family Taxa | |
|---|---|---|---|
| **Sample ID** | **Read Number** | **Total** | **RA > $1x10^{-3}$** |
| FA-1 | 18892 | 245 | 126 |
| FA-11 | 20107 | 251 | 126 |
| FA-17 | 18873 | 249 | 126 |
| FA-22 | 26371 | 251 | 126 |
| FA-23 | 22217 | 248 | 126 |
| FA-36 | 22925 | 251 | 126 |
| FA-38 | 20894 | 242 | 126 |
| FA-59 | 16605 | 240 | 126 |
| FA-66 | 23890 | 257 | 125 |
| FA-70 | 19186 | 245 | 126 |
| FA-76 | 22172 | 267 | 126 |
| FA-84 | 23628 | 252 | 126 |

**Table A.5: Reported results for the `envfit` model for the Phyla NMDS.** Vector coordinates given for the two ordination axes (fitted to NMDS ordination). Model was permutated 500 times to produce $r^2$ and *p* values. Results are plotted in main text in Fig. 5.8.C.

| | **NMSD1** | **NMSD2** | **$r^2$** | ***p*** |
|---|---|---|---|---|
| pH | -0.60613 | -0.79537 | 0.7072 | 0.006 |
| Moisture | 0.9999 | 0.01394 | 0.0826 | 0.691 |
| $CH_4$ | 0.75873 | 0.65141 | 0.741 | 0.008 |
| $O_2$ | -0.65608 | -0.75469 | 0.6111 | 0.022 |
| $C_2H_6$ | 0.69503 | 0.71898 | 0.6447 | 0.016 |

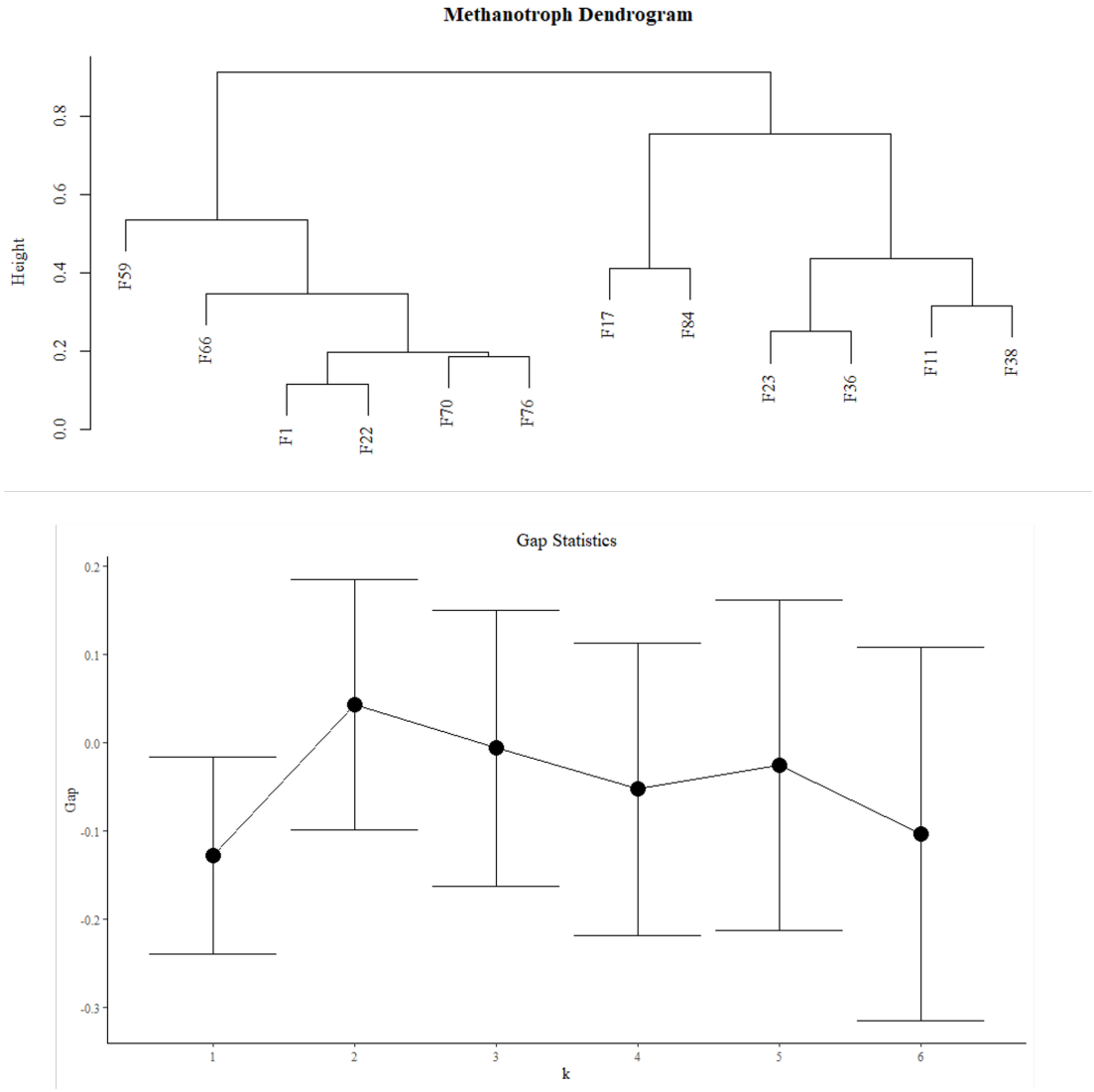



**Figure A.1: Clustering Dendrogram and Gap Statistics for 16S rRNA Methanotroph Families.** While the dendrogram suggests that two clusters might be present the gap statistic suggests that only one cluster (k) is likely given the overlap in error between k=1 and k=2 (error is standard error). Dendrogram was plotted using Bray-Curtis dissimilarity distances. Gap statistic was generated using 500 bootstraps on the same distances (k max = 6).

## A.3 Field Survey at Dukes Wood Eakring

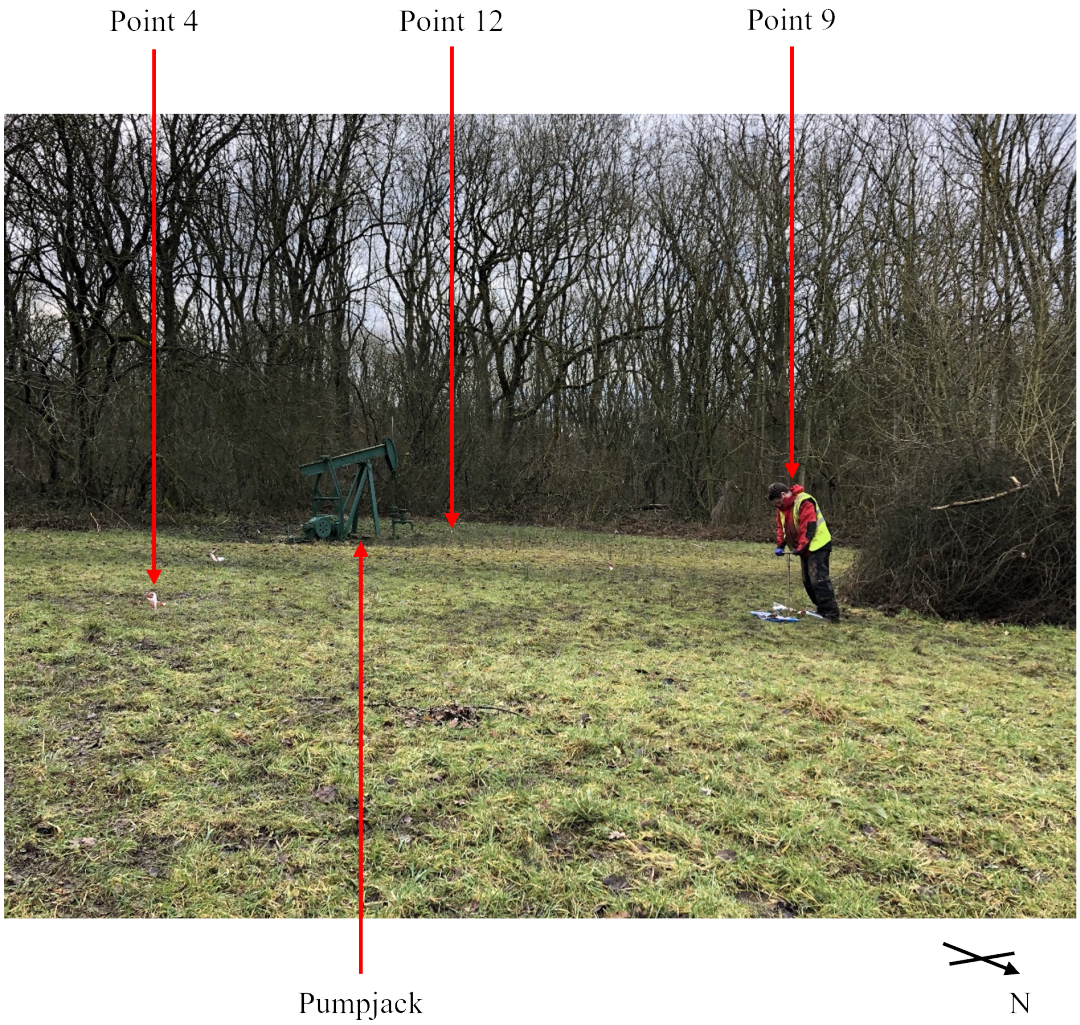



**Figure A.2: Site picture, taken at the North-East edge between points 3 and 8.** North arrow is approximate. Simon Gregory is augering the deeper sample from Point 9. The pumpjacks are clearly preserved at the site. Photo was taken during the February 2021 sampling.

**Table A.6: ANOSIM results testing for dissimilarity between depths using the amplicon sequencing dataset at three taxa levels.** All samples were not significantly dissimilar ($p$>0.05), therefore the null hypothesis that the communities at different depths are similar must be accepted. Values are rounded to three significant figures. Note, a negative R value is numerically possible but not expected, this suggests that some samples were more similar between depths rather than within depths!

| | **R** | ***p*** |
|---|---|---|
| Phyla | -0.206 | 0.971 |
| Family | -0.188 | 0.935 |
| Methanotrophic Guild | 0.159 | 0.197 |

## A.4 Impacts of a Methane Pulse Event on Soil Communities

**Table A.7:** Efficiency and $R^2$ values for the qPCR assays calculated from the slope of the standards. Due to the number of samples, two runs for each qPCR assay were completed, for ease the samples were split equally by the starting mesocosm. Assays were deemed a success only if the efficiency value was between 90-110% and if the $R^2$ $\geq$0.95

| | Gene | E-Value | $R^2$ |
|---|---|---|---|
| Control Mesocosm | 16S rRNA | 98% | 0.99 |
| | *pmoA* | 97% | 0.99 |
| | *mmoX* | 91% | 0.99 |
| | *Methylocella mmoX* | 96% | 0.99 |
| Methane Mesocosm | 16S rRNA | 109% | 0.98 |
| | *pmoA* | 103% | 0.99 |
| | *mmoX* | 93% | 0.96 |
| | *Methylocella mmoX* | 93% | 0.99 |

**Table A.8: *p*-values from Tukey's Test for pairwise comparisons of *pmoA* relative abundance between different time points.** Only the control soils displayed significant changes through time in *pmoA* relative abundance. Significance was accepted when $p$<0.05, values are rounded to 3 dp, * denotes a significant difference.

| | ***p*-Value** | |
|---|---|---|
| **Paired Time Points** | **Control** | **Methane** |
| 5-0 | 0.444 | 0.987 |
| 9-0 | 0.721 | 0.174 |
| 12-0 | 0.071 | 0.106 |
| 18-0 | 0.071 | 0.098 |
| 9-5 | 0.985 | 0.339 |
| 12-5 | *0.005 | 0.217 |
| 12-9 | *0.005 | 0.202 |
| 18-9 | *0.010 | 0.997 |
| 18-12 | *0.010 | 0.995 |

**Table A.9: Summary values for the 16S rRNA sequencing samples.** DNA extracts for each time point and mesocosm soil were pooled for sequencing. The total number of reads and unique taxa at the family level are presented here.

| Sample ID | Read Number | Unique Family Taxa |
|---|---|---|
| Baseline-1 | 19246 | 249 |
| Baseline-0 | 31477 | 264 |
| Methane-0 | 29594 | 267 |
| Methane-5 | 23438 | 248 |
| Methane-9 | 23774 | 243 |
| Methane-12 | 29408 | 258 |
| Methane-18 | 20932 | 239 |
| Control-0 | 24143 | 247 |
| Control-5 | 27709 | 265 |
| Control-9 | 26573 | 247 |
| Control-12 | 33558 | 270 |
| Control-18 | 17659 | 244 |

**Table A.10: ANOSIM results using the 16S rRNA sequencing data.** Baseline samples were excluded from comparisons. A total of 6 ANOSIM tests were completed one for each taxonomic grouping for either Time or Mesocosm factor. Where $R \geq 0$ the null hypothesis must be accepted (there is no similarity between groups). Significance was accepted when $p<0.05$, values are rounded to 3 dp, * denotes a significant difference.

| | Comparison | R | *p* |
|---|---|---|---|
| Phyla Level | Time | 0.330 | *0.036 |
| | Mesocosm Factor | 0.036 | 0.889 |
| Family Level | Time | 0.240 | 0.064 |
| | Mesocosm Factor | -0.028 | 0.585 |
| Methanotrophs | Time | -0.380 | 0.929 |
| | Mesocosm Factor | 0.780 | *0.008 |

**Table A.11: Abiotic soil factors measured during pot destruction.** Values are the average of the pots being destructively sampled for that time point.

| Sample ID | pH | Moisture |
|---|---|---|
| Methane-0 | 6.73 | 15.40% |
| Methane-5 | 6.79 | 10.20% |
| Methane-9 | 7.34 | 7.50% |
| Methane-12 | 6.97 | 6.00% |
| Methane-18 | 6.68 | 2.20% |
| Control-0 | 6.73 | 15.70% |
| Control-5 | 6.98 | 12.10% |
| Control-9 | 6.73 | 8.90% |
| Control-12 | 6.79 | 5.80% |
| Control-18 | 6.70 | 2.20% |

## A.5 Exploration of MMO in Soils from Contaminated Land

**Table A.12: Analytes and assays performed by ALS environmental services.**

| Analyte | Method Used |
|---|---|
| Ammonia Exchangeable as $NH_4$ | Titration |
| Arsenic | Inductively Coupled Plasma Optical Emission Spectroscopy (ICP-OES) |
| Boron Water Soluble | IRIS |
| BTEX | Gas Chromatography - Mass Spectrometry (GC-MS) |
| Cadmium | ICP-OES |
| Chloride Soluble | Spectrophotometric |
| Chromium | ICP-OES |
| Copper | ICP-OES |
| Cyanide Complex | Segmented flow analysis (SFA) |
| Cyanide Free | SFA |
| Cyanide Total | SFA |
| EPH Total (DRO) C10-40 | GC |
| Lead | ICP-OES |
| Loss on Ignition | Gravimetric |
| Mercury | ICP-OES |
| Moisture Content | Gravimetric |
| Nickel | ICP-OES |
| PAH Speciated 16 | GC-MS |
| Particle Size Distribution % of Stones >10mm | Gravimetric |
| pH Value | Meter |
| Phenols Total of 8 Speciated | High Performance Liquid Chromatography (HPLC) |
| Selenium | ICP-OES |
| Sulphate (Soluble 2:1 Extract) | Spectrophotometric |
| Sulphur Elemental | HPLC |
| Thiocyanate | SFA |
| Zinc | ICP-OES |

**Table A.13: Efficiency and $R^2$ values for the qPCR assays calculated from the slope of the standards.** Due to the number of samples the assay had to be split across three plates, the same reaction mix was used for all three plates. Plates were kept on ice until they could there was space in the thermocycler. Assays were deemed a success only if the efficiency value was between 90-110% and if the $R^2 \geq 0.95$.

| | Gene | E-Value | $R^2$ |
|---|---|---|---|
| Group 1 | 16S rRNA | 107% | 0.99 |
| | *pmoA* | 95% | 0.99 |
| | *mmoX* | 90% | 0.99 |
| | SDIMO | 93% | 0.99 |
| Group 2 | 16S rRNA | 105% | 0.99 |
| | *pmoA* | 96% | 0.99 |
| | *mmoX* | 96% | 0.99 |
| | SDIMO | 92% | 0.99 |
| Group 3 | 16S rRNA | 104% | 0.98 |
| | *pmoA* | 96% | 0.99 |
| | *mmoX* | 91% | 0.96 |
| | SDIMO | 92% | 0.99 |

# Bibliography

S. D. Allison and J. B. H. Martiny. Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences*, 105(Supplement 1):11512–11519, 2008. doi: 10.1073/pnas.0801925105. URL https://www.pnas.org/content/pnas/105/Supplement_1/11512.full.pdf.

S. A. Alrumman, D. B. Standing, and G. I. Paton. Effects of hydrocarbon contamination on soil microbial community and enzyme activity. *Journal of King Saud University - Science*, 27(1):31–41, 2015. ISSN 1018-3647. doi: https://doi.org/10.1016/j.jksus.2014.10.001. URL http://www.sciencedirect.com/science/article/pii/S1018364714000792.

R. A. Alvarez, S. W. Pacala, J. J. Winebrake, W. L. Chameides, and S. P. Hamburg. Greater focus needed on methane leakage from natural gas infrastructure. *Proceedings of the National Academy of Sciences*, 109(17):6435–6440, 2012. doi: 10.1073/pnas.1202407109. URL https://www.pnas.org/content/pnas/109/17/6435.full.pdf.

J. A. Amaral and R. Knowles. Growth of methanotrophs in methane and oxygen counter gradients. *FEMS Microbiology Letters*, 126(3):215–220, 1995. doi: doi:10.1111/j.1574-6968.1995.tb07421.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1574-6968.1995.tb07421.x.

S. Andersson, S. I. Nilsson, and P. Saetre. Leaching of dissolved organic carbon (doc) and dissolved organic nitrogen (don) in mor humus as affected by temperature and ph. *Soil Biology and Biochemistry*, 32(1):1–10, 2000. ISSN 0038-0717. doi: https://doi.org/10.1016/S0038-0717(99)00103-0. URL http://www.sciencedirect.com/science/article/pii/S0038071799001030.

V. Andreoni and L. Gianfreda. Bioremediation and monitoring of aromatic-polluted habitats. *Applied Microbiology and Biotechnology*, 76(2):287–308, 2007. ISSN

1432-0614. doi: 10.1007/s00253-007-1018-5. URL https://doi.org/10.1007/s00253-007-1018-5.

S. Andrews. Fastqc: A quality control tool for high throughput sequence data - ver. 0.11.9, 2015. URL http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

R. Angel, D. Matthies, and R. Conrad. Activation of methanogenesis in arid biological soil crusts despite the presence of oxygen. *PLOS ONE*, 6(5):e20453, 2011. doi: 10.1371/journal.pone.0020453. URL https://doi.org/10.1371/journal.pone.0020453.

A. J. Auman and M. E. Lidstrom. Analysis of smmo containing type i methanotrophs in lake washington sediment. *Environmental Microbiology*, 4(9):517–524, 2002. doi: doi:10.1046/j.1462-2920.2002.00323.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1462-2920.2002.00323.x.

A. Bannert, C. Bogen, J. Esperschütz, A. Koubová, F. Buegger, D. Fischer, V. Radl, R. Fuß, A. Chro?áková, D. Elhottová, M. imek, and M. Schloter. Anaerobic oxidation of methane in grassland soils used for cattle husbandry. *Biogeosciences*, 9(10):3891–3899, 2012. ISSN 1726-4189. doi: 10.5194/bg-9-3891-2012. URL https://www.biogeosciences.net/9/3891/2012/.

R. D. Barber. Methanogenesis: Ecology. *eLS*, pages 1–9, 2016. ISSN 9780470015902. doi: 10.1002/9780470015902.a0000475.pub3. URL http://dx.doi.org/10.1002/9780470015902.a0000475.pub3.

J. Baylis and D. Allenby. Gasworks remediation in the uk – a remediation contractor's perspective. In *10th IAEG*, volume 63. The Geological Society of London, 2006.

S. E. Belova, M. Baani, N. E. Suzina, P. L. E. Bodelier, W. Liesack, and S. N. Dedysh. Acetate utilization as a survival strategy of peat-inhabiting methylocystis spp. *Environmental Microbiology Reports*, 3(1):36–46, 2011. ISSN 1758-2229. doi: https://doi.org/10.1111/j.1758-2229.2010.00180.x. URL https://sfamjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1758-2229.2010.00180.x.

J. Berg, K. K. Brandt, W. A. Al-Soud, P. E. Holm, L. H. Hansen, S. J. Sørensen, and O. Nybroe. Selection for cu-tolerant bacterial communities with altered composition, but unaltered richness, via long-term cu exposure. *Applied and Environmental Microbiology*, 78(20):7438–7446, 2012. doi: doi:10.1128/AEM.01071-12. URL https://journals.asm.org/doi/abs/10.1128/AEM.01071-12.

B. B. Bernard, J. M. Brooks, and W. M. Sackett. Natural gas seepage in the gulf of mexico. *Earth and Planetary Science Letters*, 31(1):48–54, 1976. ISSN 0012-821X. doi: https://doi.org/10.1016/0012-821X(76)90095-9. URL http://www.sciencedirect.com/science/article/pii/0012821X76900959.

R. Bivand, T. Keitt, and B. Rowlingson. rgdal: Bindings for the 'geospatial' data abstraction library - ver. 1.5-23, 2021. URL https://CRAN.R-project.org/package=rgdal.

S. J. Blazewicz, D. G. Petersen, M. P. Waldrop, and M. K. Firestone. Anaerobic oxidation of methane in tropical and boreal soils: Ecological significance in terrestrial methane cycling. *Journal of Geophysical Research: Biogeosciences*, 117(G2):G02033, 2012. ISSN 2156-2202. doi: 10.1029/2011JG001864. URL http://dx.doi.org/10.1029/2011JG001864.

A. Bockelmann, T. Ptak, and G. Teutsch. An analytical quantification of mass fluxes and natural attenuation rate constants at a former gasworks site. *Journal of Contaminant Hydrology*, 53(3):429–453, 2001. ISSN 0169-7722. doi: https://doi.org/10.1016/S0169-7722(01)00177-2. URL https://www.sciencedirect.com/science/article/pii/S0169772201001772.

S. Bohn, P. Brunke, J. Gebert, and J. Jager. Improving the aeration of critical fine-grained landfill top cover material by vegetation to increase the microbial methane oxidation efficiency. *Waste Management*, 31(5):854–863, 2011. ISSN 0956-053X. doi: https://doi.org/10.1016/j.wasman.2010.11.009. URL http://www.sciencedirect.com/science/article/pii/S0956053X10005738.

E. Bomberg. Shale we drill? discourse dynamics in uk fracking debates. *Journal of Environmental Policy I & Planning*, 19(1):72–88, 2017. ISSN 1523-908X. doi: 10.1080/1523908X.2015.1053111. URL https://doi.org/10.1080/1523908X.2015.1053111.

I. M. Boothroyd, S. Almond, S. M. Qassim, F. Worrall, and R. J. Davies. Fugitive emissions of methane from abandoned, decommissioned oil and gas wells. *The Science of the Total Environment*, 2015. doi: 10.1016/j.scitotenv.2015.12.096.

D. G. Bourne, I. R. McDonald, and J. C. Murrell. Comparison of pmoa pcr primer sets as tools for investigating methanotroph diversity in three danish soils. *Applied and Environmental Microbiology*, 67(9):3802–3809, 2001. doi: 10.1128/aem.67.9.3802-3809.2001. URL http://aem.asm.org/content/67/9/3802.abstract.

R. D. Bowden, K. M. Newkirk, and G. M. Rullo. Carbon dioxide and methane fluxes by a forest soil under laboratory-controlled moisture and temperature conditions. *Soil Biology and Biochemistry*, 30(12):1591–1597, 1998. ISSN 0038-0717. doi: https://doi.org/10.1016/S0038-0717(97)00228-9. URL https://www.sciencedirect.com/science/article/pii/S0038071797002289.

P. G. Brisbane and J. N. Ladd. The role of microorganisms in petroleum exploration. *Annual Review of Microbiology*, 19(1):351–364, 1965. doi: 10.1146/annurev.mi.19.100165.002031. URL https://www.annualreviews.org/doi/abs/10.1146/annurev.mi.19.100165.002031.

British Geological Survey. Sheet 113 - Ollerton, solid and drift, 1996.

J. G. Bundy, G. I. Paton, and C. D. Campbell. Microbial communities in different soil types do not converge after diesel contamination. *Journal of Applied Microbiology*, 92(2):276–288, 2002. ISSN 1364-5072. doi: 10.1046/j.1365-2672.2002.01528.x. URL https://doi.org/10.1046/j.1365-2672.2002.01528.x.

S. Burggraf and B. Olgemooller. Simple technique for internal control of real-time amplification assays. *Clinical Chemistry*, 50(5):819–825, 2004. ISSN 0009-9147. doi: 10.1373/clinchem.2003.027961. URL https://doi.org/10.1373/clinchem.2003.027961.

S. A. Bustin, V. Benes, J. A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele, and C. T. Wittwer. The miqe guidelines: Minimum information for publication of quantitative real-time pcr experiments. *Clinical Chemistry*, 55(4):611–622, 2009. doi: 10.1373/clinchem.2008.112797. URL http://clinchem.aaccjnls.org/content/clinchem/55/4/611.full.pdf.

G. Börjesson, I. Sundh, and B. Svensson. Microbial oxidation of $CH_4$ at different temperatures in landfill cover soils. *FEMS Microbiology Ecology*, 48(3):305–312, 2004. ISSN 0168-6496. doi: 10.1016/j.femsec.2004.02.006. URL http://dx.doi.org/10.1016/j.femsec.2004.02.006.

H. Bürgmann, M. Pesaro, F. Widmer, and J. Zeyer. A strategy for optimizing quality and quantity of dna extracted from soil. *Journal of Microbiological Methods*, 45 (1):7–20, 2001. ISSN 0167-7012. doi: https://doi.org/10.1016/S0167-7012(01)00213-5. URL http://www.sciencedirect.com/science/article/pii/S0167701201002135.

A. Cahill, C. Steelman, O. Forde, O. Kuloyo, S. E. Ruff, B. Mayer, U. Mayer, M. Strous, M. Ryan, J. A. Cherry, and B. Parker. Mobility and persistence of methane in groundwater in a controlled-release field experiment. *Nature Geoscience*, 10, 2017. doi: 10.1038/ngeo2919.

R. Cano, C. Kitts, and B. J. Chan. Pcr primers for the detection of propane and methane oxidising organisms, 2013.

J. K. Carson, V. Gonzalez-Quiñones, D. V. Murphy, C. Hinz, J. A. Shaw, and D. B. Gleeson. Low pore connectivity increases bacterial diversity in soil. *Applied and environmental microbiology*, 76(12):3936–3942, 2010. ISSN 1098-5336 0099-2240. doi: 10.1128/AEM.03085-09. URL https://pubmed.ncbi.nlm.nih.gov/20418420https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2893478/.

D. R. Caulton, P. B. Shepson, R. L. Santoro, J. P. Sparks, R. W. Howarth, A. R. Ingraffea, M. O. L. Cambaliza, C. Sweeney, A. Karion, K. J. Davis, B. H. Stirm, S. A. Montzka, and B. R. Miller. Toward a better understanding and quantification of methane emissions from shale gas development. *Proceedings of the National Academy of Sciences*, 111(17):6237–6242, 2014. doi: 10.1073/pnas.1316546111. URL https://www.pnas.org/content/pnas/111/17/6237.full.pdf.

Y. J. Choi and S. Y. Lee. Microbial production of short-chain alkanes. *Nature*, 502 (7472):571–574, 2013. ISSN 1476-4687. doi: 10.1038/nature12536. URL https://doi.org/10.1038/nature12536.

P. Ciais, C. Sabine, G. Bala, L. Bopp, V. Brovkin, J. Canadell, A. Chhabra, R. DeFries, J. Galloway, M. Heimann, C. Jones, C. Le Quéré, R. B. Myneni, S. Piao, and P. Thornton. *Carbon and Other Biogeochemical Cycles*, book section 6. Cambridge University Press, Cambridge UK and New York USA, 2013.

J. Colby, D. I. Stirling, and H. Dalton. The soluble methane mono-oxygenase of methylococcus capsulatus (bath). its ability to oxygenate n-alkanes, n-alkenes, ethers, and alicyclic, aromatic and heterocyclic compounds. *The Biochemical journal*, 165(2):395–402, 1977. ISSN 0264-6021 1470-8728. doi: 10.1042/bj1650395. URL https://pubmed.ncbi.nlm.nih.gov/411486https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1164912/.

N. V. Coleman, N. B. Bui, and A. J. Holmes. Soluble di-iron monooxygenase gene diversity in soils, sediments and ethene enrichments. *Environmental Mi-*

*crobiology*, 8(7):1228–1239, 2006. ISSN 1462-2920. doi: 10.1111/j.1462-2920.2006.01015.x. URL http:https://doi.org/10.1111/j.1462-2920.2006.01015.x.

N. V. Coleman, N. B. Le, M. A. Ly, H. E. Ogawa, V. McCarl, N. L. Wilson, and A. J. Holmes. Hydrocarbon monooxygenase in mycobacterium: recombinant expression of a member of the ammonia monooxygenase superfamily. *The ISME Journal*, 6 (1):171–182, 2012. ISSN 1751-7370. doi: 10.1038/ismej.2011.98. URL https://doi.org/10.1038/ismej.2011.98.

R. Conrad. *Microbial Ecology of Methanogens and Methanotrophs*, volume 96, pages 1–63. Academic Press, 2007. ISBN 0065-2113. doi: https://doi.org/10.1016/S0065-2113(07)96005-8. URL http://www.sciencedirect.com/science/article/pii/S0065211307960058.

A. M. Costello and M. E. Lidstrom. Molecular characterization of functional and phylogenetic genes from natural populations of methanotrophs in lake sediments. *Applied and Environmental Microbiology*, 65(11):5066–5074, 1999. URL http://aem.asm.org/content/65/11/5066.abstract.

R. Csáki, L. Bodrossy, J. Klem, J. C. Murrell, and K. L. Kovács. Genes involved in the copper-dependent regulation of soluble methane monooxygenase of *Methylococcus capsulatus* (bath): cloning, sequencing and mutational analysis. *Microbiology*, 149(7):1785–1795, 2003. doi: doi:10.1099/mic.0.26061-0. URL http://mic.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.26061-0.

R. J. Davies, S. Almond, R. S. Ward, R. B. Jackson, C. Adams, F. Worrall, L. G. Herringshaw, J. G. Gluyas, and M. A. Whitehead. Oil and gas wells and their integrity: Implications for shale and unconventional resource exploitation. *Marine and Petroleum Geology*, 56(Supplement C):239–254, 2014. ISSN 0264-8172. doi: https://doi.org/10.1016/j.marpetgeo.2014.03.001. URL http://www.sciencedirect.com/science/article/pii/S0264817214000609.

S. L. Davies and R. Whittenbury. Fine structure of methane and other hydrocarbon-utilizing bacteria. *Journal of General Microbiology*, 61(2):227–232, 1970. ISSN 1350-0872. doi: https://doi.org/10.1099/00221287-61-2-227. URL https://www.microbiologyresearch.org/content/journal/micro/10.1099/00221287-61-2-227.

W. De Coster, S. D'Hert, D. T. Schultz, M. Cruts, and C. Van Broeckhoven. Nanopack: visualizing and processing long-read sequencing data. *Bioinformatics*, 34(15):2666–2669, 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty149. URL https://doi.org/10.1093/bioinformatics/bty149.

S. N. Dedysh and P. F. Dunfield. *Chapter Three - Facultative and Obligate Methanotrophs: How to Identify and Differentiate Them*, volume 495, pages 31–44. Academic Press, 2011. ISBN 0076-6879. doi: https://doi.org/10.1016/B978-0-12-386905-0.00003-6. URL http://www.sciencedirect.com/science/article/pii/B9780123869050000036.

S. N. Dedysh, N. S. Panikov, W. Liesack, R. Großkopf, J. Zhou, and J. M. Tiedje. Isolation of acidophilic methane-oxidizing bacteria from northern peat wetlands. *Science*, 282(5387):281–284, 1998. doi: 10.1126/science.282.5387.281. URL https://science.sciencemag.org/content/sci/282/5387/281.full.pdf.

S. N. Dedysh, W. Liesack, V. N. Khmelenina, N. E. Suzina, Y. A. Trotsenko, J. D. Semrau, A. M. Bares, N. S. Panikov, and J. M. Tiedje. *Methylocella palustris* gen. nov., sp. nov., a new methane-oxidizing acidophilic bacterium from peat bogs, representing a novel subtype of serine-pathway methanotrophs. *International Journal of Systematic and Evolutionary Microbiology*, 50(3):955–969, 2000. doi: doi:10.1099/00207713-50-3-955. URL http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-50-3-955.

S. N. Dedysh, C. Knief, and P. F. Dunfield. *Methylocella* species are facultatively methanotrophic. *Journal of Bacteriology*, 187(13):4665–4670, 2005. ISSN 0021-9193 1098-5530. doi: 10.1128/JB.187.13.4665-4670.2005. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1151763/.

S. N. Dedysh, E. S. Haupt, and P. F. Dunfield. Emended description of the family beijerinckiaceae and transfer of the genera chelatococcus and camelimonas to the family chelatococcaceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology*, 66(8):3177–3182, 2016. ISSN 1466-5034. doi: https://doi.org/10.1099/ijsem.0.001167. URL https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.001167.

L. Dendooven, D. Alvarez-Bernal, and S. M. Contreras-Ramos. Earthworms, a means to accelerate removal of hydrocarbons (pahs) from soil? a mini-review. *Pedobiologia*, 54:S187–S192, 2011. ISSN 0031-4056. doi: https://doi.org/10.1016/j.

pedobi.2011.08.006. URL http://www.sciencedirect.com/science/article/pii/S0031405611000874.

E. Dlugokencky and NOAA/GML. Global $CH_4$ monthly means, 2022. URL gml.noaa.gov/ccgg/trends_ch4/.

M. G. Dumont and J. C. Murrell. *Community?Level Analysis: Key Genes of Aerobic Methane Oxidation*, volume 397, pages 413–427. Academic Press, 2005. ISBN 0076-6879. doi: https://doi.org/10.1016/S0076-6879(05)97025-0. URL http://www.sciencedirect.com/science/article/pii/S0076687905970250.

P. F. Dunfield, S. E. Belova, Vorob, apos, A. V. ev, S. L. Cornish, and S. N. Dedysh. Methylocapsa aurea sp. nov., a facultative methanotroph possessing a particulate methane monooxygenase, and emended description of the genus methylocapsa. *International Journal of Systematic and Evolutionary Microbiology*, 60(11):2659–2664, 2010. ISSN 1466-5034. doi: https://doi.org/10.1099/ijs.0.020149-0. URL https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.020149-0.

M. Dusseault, M. Gray, and P. Nawrocki. Why oilwells leak: Cement behavior and long-term consequences - spe 64733. In *SPE International Oil and Gas Conference and Exhibition*. Society of Petroleum Engineers Inc., 2000. doi: 10.2118/64733-MS.

K. Engel, S. Coyotzi, M. A. Vachon, J. R. McKelvie, J. D. Neufeld, and H. Tamaki. Validating dna extraction protocols for bentonite clay. *mSphere*, 4(5):e00334–19, 2019. doi: doi:10.1128/mSphere.00334-19. URL https://journals.asm.org/doi/abs/10.1128/mSphere.00334-19.

Environmental Bio Technologies Inc., 2017. URL http://www.e-b-t.com/.

G. Etiope. *Natural Gas Seepage*. Springer, Cham, Switzerland, 2015.

G. Etiope and B. S. Lollar. Abiotic methane on earth. *Reviews of Geophysics*, 51(2):276–299, 2013. ISSN 1944-9208. doi: 10.1002/rog.20011. URL http:https://doi.org/10.1002/rog.20011.

M. Farhan Ul Haque, A. T. Crombie, S. A. Ensminger, C. Baciu, and J. C. Murrell. Facultative methanotrophs are abundant at terrestrial natural gas seeps. *Microbiome*, 6(1):118, 2018. ISSN 2049-2618. doi: 10.1186/s40168-018-0500-x. URL https://doi.org/10.1186/s40168-018-0500-x.

M. Farhan Ul Haque, A. T. Crombie, and J. C. Murrell. Novel facultative methylocella strains are active methane consumers at terrestrial natural gas seeps. *Microbiome*, 7 (1):134, 2019. ISSN 2049-2618. doi: 10.1186/s40168-019-0741-3. URL https://doi.org/10.1186/s40168-019-0741-3.

A. Ferguson, W. Huang, K. Lawson, R. Doherty, O. Gibert, K. Dickson, A. Whiteley, L. Kulakov, I. Thompson, R. Kalin, and M. Larkin. Microbial analysis of soil and groundwater from a gasworks site and comparison with a sequenced biological reactive barrier remediation process. *Journal of Applied Microbiology*, 102(5): 1227–1238, 2007. ISSN 1364-5072. doi: https://doi.org/10.1111/j.1365-2672.2007.03398.x. URL https://sfamjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2672.2007.03398.x.

N. Fierer, J. P. Schimel, and P. A. Holden. Variations in microbial community composition through two soil depth profiles. *Soil Biology and Biochemistry*, 35(1):167–176, 2003. ISSN 0038-0717. doi: https://doi.org/10.1016/S0038-0717(02)00251-1. URL https://www.sciencedirect.com/science/article/pii/S0038071702002511.

C. G. Flocco, N. C. M. Gomes, W. Mac Cormack, and K. Smalla. Occurrence and diversity of naphthalene dioxygenase genes in soil microbial communities from the maritime antarctic. *Environmental Microbiology*, 11(3): 700–714, 2009. ISSN 1462-2912. doi: 10.1111/j.1462-2920.2008.01858.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1462-2920.2008.01858.x.

R. B. Franklin and A. L. Mills. Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern virginia agricultural field. *FEMS Microbiology Ecology*, 44(3):335–346, 2003. ISSN 0168-6496. doi: https://doi.org/10.1016/S0168-6496(03)00074-6. URL http://www.sciencedirect.com/science/article/pii/S0168649603000746.

F. Gal, W. Kloppmann, E. Proust, G. Bentivegna, P. Defossez, B. Mayer, and E. C. Gaucher. Natural $CH_4$ gas seeps in the french alps: Characteristics, typology and contribution to $CH_4$ natural emissions to the atmosphere. *Energy Procedia*, 114: 3020–3032, 2017. ISSN 1876-6102. doi: https://doi.org/10.1016/j.egypro.2017.03.1430. URL http://www.sciencedirect.com/science/article/pii/S1876610217316156.

F. Gal, W. Kloppmann, E. Proust, and P. Humez. Gas concentration and flow rate measurements as part of methane baseline assessment: Case of the fontaine ardente

gas seep, isère, france. *Applied Geochemistry*, 95:158–171, 2018. ISSN 0883-2927. doi: https://doi.org/10.1016/j.apgeochem.2018.05.019. URL http://www.sciencedirect.com/science/article/pii/S0883292718301355.

F. Gal, E. Proust, and W. Kloppmann. Towards a better knowledge of natural methane releases in the french alps: A field approach. *Geofluids*, 2019, 2019. doi: 10.1155/2019/6487162. URL https://doi.org/10.1155/2019/6487162.

E. Gallori, M. Bazzicalupo, L. Dal Canto, R. Fani, P. Nannipieri, C. Vettori, and G. Stotzky. Transformation of bacillus subtilis by dna bound on clay in non-sterile soil. *FEMS Microbiology Ecology*, 15(1-2):119–126, 1994. ISSN 0168-6496. doi: 10.1111/j.1574-6941.1994.tb00236.x. URL https://doi.org/10.1111/j.1574-6941.1994.tb00236.x.

J. M. Gallup and M. R. Ackermann. The 'prexcel-q method' for qpcr. *International journal of biomedical science : IJBS*, 4(4):273–293, 2008. ISSN 1550-9702 1555-2810. URL https://www.ncbi.nlm.nih.gov/pubmed/19759920https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2744046/.

M. Gauthier, R. L. Bradley, and M. Šimek. More evidence that anaerobic oxidation of methane is prevalent in soils: Is it time to upgrade our biogeochemical models? *Soil Biology and Biochemistry*, 80:167–174, 2015. doi: 10.1016/j.soilbio.2014.10.009.

J. Gebert, A. Groengroeft, and E.-M. Pfeiffer. Relevance of soil physical properties for the microbial oxidation of methane in landfill covers. *Soil Biology and Biochemistry*, 43(9):1759–1767, 2011. ISSN 0038-0717. doi: https://doi.org/10.1016/j.soilbio.2010.07.004. URL http://www.sciencedirect.com/science/article/pii/S0038071710002518.

S. D. Golding, C. J. Boreham, and J. S. Esterle. Stable isotope geochemistry of coal bed and shale gas and related production waters: A review. *International Journal of Coal Geology*, 120:24–40, 2013. ISSN 0166-5162. doi: https://doi.org/10.1016/j.coal.2013.09.001. URL http://www.sciencedirect.com/science/article/pii/S0166516213002061.

P. F. Greenwood, S. Wibrow, S. J. George, and M. Tibbett. Hydrocarbon biodegradation and soil microbial community response to repeated oil exposure. *Organic Geochemistry*, 40(3):293–300, 2009. ISSN 0146-6380. doi: https://doi.org/10.1016/j.orggeochem.2008.12.009. URL http://www.sciencedirect.com/science/article/pii/S0146638008003811.

R. I. Griffiths, B. C. Thomson, P. James, T. Bell, M. Bailey, and A. S. Whiteley. The bacterial biogeography of british soils. *Environmental Microbiology*, 13 (6):1642–1654, 2011. ISSN 1462-2912. doi: 10.1111/j.1462-2920.2011.02480.x. URL https://sfamjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1462-2920.2011.02480.x.

A. S. Hakemian and A. C. Rosenzweig. The biochemistry of methane oxidation. *Annual Review of Biochemistry*, 76(1):223–241, 2007. ISSN 0066-4154. doi: 10.1146/annurev.biochem.76.061505.175355. URL https://doi.org/10.1146/annurev.biochem.76.061505.175355.

K. H. Halsey, L. A. Sayavedra-Soto, P. J. Bottomley, and D. J. Arp. Site-directed amino acid substitutions in the hydroxylase alpha subunit of butane monooxygenase from *Pseudomonas butanovora*: Implications for substrates knocking at the gate. *Journal of bacteriology*, 188(13):4962–4969, 2006. ISSN 0021-9193 1098-5530. doi: 10.1128/JB.00280-06. URL https://pubmed.ncbi.nlm.nih.gov/16788204https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1482983/.

N. Hamamura, S. H. Olson, D. M. Ward, and W. P. Inskeep. Microbial population dynamics associated with crude-oil biodegradation in diverse soils. *Applied and Environmental Microbiology*, 72(9):6316–6324, 2006. ISSN 0099-2240 1098-5336. doi: 10.1128/AEM.01015-06. URL https://www.ncbi.nlm.nih.gov/pubmed/16957258https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1563594/.

R. S. Hanson and T. E. Hanson. Methanotrophic bacteria. *Microbiological Reviews*, 60(2):439–71, 1996. URL http://mmbr.asm.org/content/60/2/439.abstract.

J. Hao, Y. N. Chai, L. D. Lopes, R. A. Ordóñez, E. E. Wright, S. Archontoulis, and D. P. Schachtman. The effects of soil depth on the structure of microbial communities in agricultural soils in iowa, usa. *Applied and environmental microbiology*, 87 (4):e02673–20, 2021. ISSN 1098-5336 0099-2240. doi: 10.1128/AEM.02673-20. URL https://pubmed.ncbi.nlm.nih.gov/33310710https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7851703/.

R. He, M. J. Wooller, J. W. Pohlman, J. Quensen, J. M. Tiedje, and M. B. Leigh. Shifts in identity and activity of methanotrophs in arctic lake sediments in response to temperature changes. *Applied and Environmental Microbiology*, 78(13):4715–4723,

2012. doi: 10.1128/aem.00853-12. URL http://aem.asm.org/content/78/13/4715.abstract.

T. Henckel, U. Jäckel, S. Schnell, and R. Conrad. Molecular analyses of novel methanotrophic communities in forest soil that oxidize atmospheric methane. *Applied and Environmental Microbiology*, 66(5):1801–1808, 2000. doi: 10.1128/aem.66.5.1801-1808.2000. URL https://aem.asm.org/content/aem/66/5/1801.full.pdf.

R. J. Hijmans. raster: Geographic data analysis and modeling - ver. 3.4-13, 2021.

A. Ho, C. Lüke, and P. Frenzel. Recovery of methanotrophs from disturbance: population dynamics, evenness and functioning. *The ISME Journal*, 5(4):750–758, 2011. ISSN 1751-7370. doi: 10.1038/ismej.2010.163. URL https://doi.org/10.1038/ismej.2010.163.

A. Ho, F.-M. Kerckhof, C. Luke, A. Reim, S. Krause, N. Boon, and P. L. E. Bodelier. Conceptualizing functional traits and ecological characteristics of methane-oxidizing bacteria as life strategies. *Environmental Microbiology Reports*, 5(3):335–345, 2013. doi: doi:10.1111/j.1758-2229.2012.00370.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1758-2229.2012.00370.x.

A. Ho, R. Angel, A. J. Veraart, A. Daebeler, Z. Jia, S. Y. Kim, F.-M. Kerckhof, N. Boon, and P. L. E. Bodelier. Biotic interactions in microbial communities as modulators of biogeochemical processes: Methanotrophy as a model system. *Frontiers in Microbiology*, 7(1285), 2016. ISSN 1664-302X. doi: 10.3389/fmicb.2016.01285. URL https://www.frontiersin.org/article/10.3389/fmicb.2016.01285.

A. J. Holmes and N. V. Coleman. Evolutionary ecology and multidisciplinary approaches to prospecting for monooxygenases as biocatalysts. *Antonie van Leeuwenhoek*, 94(1):75–84, 2008. ISSN 1572-9699. doi: 10.1007/s10482-008-9227-1. URL https://doi.org/10.1007/s10482-008-9227-1.

A. J. Holmes, A. Costello, M. E. Lidstrom, and J. C. Murrell. Evidence that particulate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. *FEMS Microbiology Letters*, 132(3):203–208, 1995. ISSN 1574-6968. doi: 10.1111/j.1574-6968.1995.tb07834.x. URL http:https://doi.org/10.1111/j.1574-6968.1995.tb07834.x.

R. Howarth, A. Ingraffea, and T. Engelder. Should fracking stop? *Nature*, 477(7364): 271–275, 2011. doi: 10.1038/477271a.

E. Hutchens, S. Radajewski, M. G. Dumont, I. R. McDonald, and J. C. Murrell. Analysis of methanotrophic bacteria in movile cave by stable isotope probing. *Environmental Microbiology*, 6(2):111–120, 2004. doi: doi:10.1046/j.1462-2920.2003.00543.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1462-2920.2003.00543.x.

M. R. Hyman, I. B. Murton, and D. J. Arp. Interaction of ammonia monooxygenase from nitrosomonas europaea with alkanes, alkenes, and alkynes. *Applied and Environmental Microbiology*, 54(12):3187–3190, 1988. ISSN 0099-2240 1098-5336. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC204451/.

J. Im and J. D. Semrau. Pollutant degradation by a *Methylocystis* strain sb2 grown on ethanol: bioremediation via facultative methanotrophy. *FEMS Microbiology Letters*, 318(2):137–142, 2011. ISSN 0378-1097. doi: 10.1111/j.1574-6968.2011.02249.x. URL https://doi.org/10.1111/j.1574-6968.2011.02249.x.

J. Im, S.-W. Lee, S. Yoon, A. A. DiSpirito, and J. D. Semrau. Characterization of a novel facultative *Methylocystis* species capable of growth on methane, acetate and ethanol. *Environmental Microbiology Reports*, 3(2): 174–181, 2011. ISSN 1758-2229. doi: 10.1111/j.1758-2229.2010.00204.x. URL https://sfamjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1758-2229.2010.00204.x.

P. W. Inglett, K. R. Reddy, and R. Corstanje. *Anaerobic Soils*, pages 72–78. Elsevier, Oxford, 2005. ISBN 978-0-12-348530-4. doi: https://doi.org/10.1016/B0-12-348530-4/00178-8. URL http://www.sciencedirect.com/science/article/pii/B0123485304001788.

A. R. Ingraffea, M. T. Wells, R. L. Santoro, and S. B. C. Shonkoff. Assessment and risk analysis of casing and cement impairment in oil and gas wells in pennsylvania, 2000 2012. *Proceedings of the National Academy of Sciences*, 111(30):10955–10960, 2014. doi: 10.1073/pnas.1323422111. URL https://www.pnas.org/content/pnas/111/30/10955.full.pdf.

D. Inoue, T. Tsunoda, K. Sawada, N. Yamamoto, K. Sei, and M. Ike. Stimulatory and inhibitory effects of metals on 1,4-dioxane degradation by four different 1,4-dioxane-degrading bacteria. *Chemosphere*, 238:124606, 2020. ISSN 0045-6535.

doi: https://doi.org/10.1016/j.chemosphere.2019.124606. URL https://www.sciencedirect.com/science/article/pii/S0045653519318302.

C. L. C. Ip, M. Loose, J. R. Tyson, M. de Cesare, B. L. Brown, M. Jain, R. M. Leggett, D. A. Eccles, V. Zalunin, J. M. Urban, P. Piazza, R. J. Bowden, B. Paten, S. Mwaigwisya, E. M. Batty, J. T. Simpson, T. P. Snutch, E. Birney, D. Buck, S. Goodwin, H. J. Jansen, J. O'Grady, and H. E. Olsen. Minion analysis and reference consortium: Phase 1 data release and analysis. *F1000Res*, 4:1075, 2015. ISSN 2046-1402 (Print) 2046-1402. doi: 10.12688/f1000research.7201.1.

R. B. Jackson. The integrity of oil and gas wells. *Proceedings of the National Academy of Sciences*, 111(30):10902–10903, 2014. doi: 10.1073/pnas.1410786111. URL https://www.pnas.org/content/pnas/111/30/10902.full.pdf.

A. Janbandhu and M. H. Fulekar. Biodegradation of phenanthrene using adapted microbial consortium isolated from petrochemical contaminated environment. *Journal of Hazardous Materials*, 187(1-3):333–340, 2011. ISSN 0304-3894. doi: http://dx.doi.org/10.1016/j.jhazmat.2011.01.034. URL //www.sciencedirect.com/science/article/pii/S0304389411000562.

R. Jaspal and B. Nerlich. Fracking in the uk press: Threat dynamics in an unfolding debate. *Public Understanding of Science*, 23(3):348–363, 2014. doi: 10.1177/0963662513498835. URL http://journals.sagepub.com/doi/abs/10.1177/0963662513498835.

A. R. Johnsen, L. Y. Wick, and H. Harms. Principles of microbial pah-degradation in soil. *Environmental Pollution*, 133(1):71–84, 2005. ISSN 0269-7491. doi: http://dx.doi.org/10.1016/j.envpol.2004.04.015. URL http://www.sciencedirect.com/science/article/pii/S0269749104001587.

K. C. Jones and P. de Voogt. Persistent organic pollutants (pops): state of the science. *Environmental Pollution*, 100(1):209–221, 1999. ISSN 0269-7491. doi: https://doi.org/10.1016/S0269-7491(99)00098-6. URL https://www.sciencedirect.com/science/article/pii/S0269749199000986.

D. Juck, T. Charles, L. Whyte, and C. Greer. Polyphasic microbial community analysis of petroleum hydrocarbon-contaminated soils from two northern canadian communities. *FEMS Microbiology Ecology*, 33(3):241–249, 2000. ISSN 0168-6496. doi: 10.1111/j.1574-6941.2000.tb00746.x. URL https://doi.org/10.1111/j.1574-6941.2000.tb00746.x.

M. Kang, S. Christian, M. A. Celia, D. L. Mauzerall, M. Bill, A. R. Miller, Y. Chen, M. E. Conrad, T. H. Darrah, and R. B. Jackson. Identification and characterization of high methane-emitting abandoned oil and gas wells. *Proceedings of the National Academy of Sciences*, 113(48):13636–13641, 2016. doi: 10.1073/pnas.1605913113. URL https://www.pnas.org/content/pnas/113/48/13636.full.pdf.

A. Karion, C. Sweeney, G. Pétron, G. Frost, R. Michael Hardesty, J. Kofler, B. R. Miller, T. Newberger, S. Wolter, R. Banta, A. Brewer, E. Dlugokencky, P. Lang, S. A. Montzka, R. Schnell, P. Tans, M. Trainer, R. Zamora, and S. Conley. Methane emissions estimate from airborne measurements over a western united states natural gas field. *Geophysical Research Letters*, 40(16):4393–4397, 2013. ISSN 0094-8276. doi: 10.1002/grl.50811. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.50811.

S. Kauppi, M. Romantschuk, R. Strömmer, and A. Sinkkonen. Natural attenuation is enhanced in previously contaminated and coniferous forest soils. *Environmental Science and Pollution Research*, 19(1):53–63, 2012. ISSN 1614-7499. doi: 10.1007/s11356-011-0528-y. URL https://doi.org/10.1007/s11356-011-0528-y.

M. Khanna and G. Stotzky. Transformation of bacillus subtilis by dna bound on montmorillonite and effect of dnase on the transforming ability of bound dna. *Applied and Environmental Microbiology*, 58(6):1930–1939, 1992. doi: doi:10.1128/aem.58.6.1930-1939.1992. URL https://journals.asm.org/doi/abs/10.1128/aem.58.6.1930-1939.1992.

D. Kightley, D. B. Nedwell, and M. Cooper. Capacity for methane oxidation in landfill cover soils measured in laboratory-scale soil microcosms. *Applied and Environmental Microbiology*, 61(2):592–601, 1995. URL http://aem.asm.org/content/61/2/592.abstract.

K. Killham. *Soil Ecology*. Cambridge University Press, Cambridge, UK, 1st edition, 1994.

R. Kızılkaya, T. Aşkın, B. Bayraklı, and M. Sağlam. Microbiological characteristics of soils contaminated with heavy metals. *European Journal of Soil Biology*, 40(2):95–102, 2004. ISSN 1164-5563. doi: https://doi.org/10.1016/j.ejsobi.2004.10.002. URL https://www.sciencedirect.com/science/article/pii/S1164556304000214.

B. Klimek, A. Sitarz, M. Choczycski, and M. Niklińska. The effects of heavy metals and total petroleum hydrocarbons on soil bacterial activity and functional diversity in the upper silesia industrial region (poland). *Water, Air, & Soil Pollution*, 227 (8):265, 2016. ISSN 1573-2932. doi: 10.1007/s11270-016-2966-0. URL http://dx.doi.org/10.1007/s11270-016-2966-0.

C. Knief. Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on pmoa as molecular marker. *Frontiers in Microbiology*, 6:1346, 2015. ISSN 1664-302X. doi: 10.3389/fmicb.2015.01346. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4678205/.

C. Knief, A. Lipski, and P. F. Dunfield. Diversity and activity of methanotrophic bacteria in different upland soils. *Applied and environmental microbiology*, 69 (11):6703–6714, 2003. ISSN 0099-2240 1098-5336. doi: 10.1128/AEM.69.11.6703-6714.2003. URL https://www.ncbi.nlm.nih.gov/pubmed/14602631https://www.ncbi.nlm.nih.gov/pmc/PMC262299/.

C. Knief, S. Vanitchung, N. W. Harvey, R. Conrad, P. F. Dunfield, and A. Chidthaisong. Diversity of methanotrophic bacteria in tropical upland soils under different land uses. *Applied and environmental microbiology*, 71(7):3826–3831, 2005. ISSN 0099-2240 1098-5336. doi: 10.1128/AEM.71.7.3826-3831.2005. URL https://www.ncbi.nlm.nih.gov/pubmed/16000794https://www.ncbi.nlm.nih.gov/pmc/PMC1169035/.

C. Knief, S. Kolb, P. L. E. Bodelier, A. Lipski, and P. F. Dunfield. The active methanotrophic community in hydromorphic soils changes in response to changing methane concentration. *Environmental Microbiology*, 8(2):321–333, 2006. ISSN 1462-2912. doi: https://doi.org/10.1111/j.1462-2920.2005.00898.x. URL https://sfamjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1462-2920.2005.00898.x.

D. A. Kopp and S. J. Lippard. Soluble methane monooxygenase: activation of dioxygen and methane. *Current Opinion in Chemical Biology*, 6(5):568–576, 2002. ISSN 1367-5931. doi: https://doi.org/10.1016/S1367-5931(02)00366-6. URL http://www.sciencedirect.com/science/article/pii/S1367593102003666.

S. M. B. Krause, T. Johnson, Y. Samadhi Karunaratne, Y. Fu, D. A. C. Beck, L. Chistoserdova, and M. E. Lidstrom. Lanthanide-dependent cross-feeding of

methane-derived carbon is linked by microbial community interactions. *Proceedings of the National Academy of Sciences*, 114(2):358–363, 2017. doi: 10.1073/pnas.1619871114. URL https://www.pnas.org/content/pnas/114/2/358.full.pdf.

S. Kumar, G. Stecher, M. Li, C. Knyaz, and K. Tamura. Mega x: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35:1547–1549, 2018.

V. Labud, C. Garcia, and T. Hernandez. Effect of hydrocarbon pollution on the microbial properties of a sandy and a clay soil. *Chemosphere*, 66(10):1863–1871, 2007. ISSN 0045-6535. doi: https://doi.org/10.1016/j.chemosphere.2006.08.021. URL http://www.sciencedirect.com/science/article/pii/S0045653506011507.

G. Lackey and H. Rajaram. Modeling gas migration, sustained casing pressure, and surface casing vent flow in onshore oil and gas wells. *Water Resources Research*, 55(1):298–323, 2018. ISSN 0043-1397. doi: 10.1029/2018wr024066. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR024066.

L. Lai, X. Zhao, L. Jiang, Y. Wang, L. Luo, Y. Zheng, X. Chen, and G. M. Rimmington. Soil respiration in different agricultural and natural ecosystems in an arid region. *PLOS ONE*, 7(10):e48011, 2012. doi: 10.1371/journal.pone.0048011. URL https://doi.org/10.1371/journal.pone.0048011.

Q. Lai, L. Wang, Y. Liu, Y. Fu, H. Zhong, B. Wang, L. Chen, J. Wang, F. Sun, and Z. Shao. *Alcanivorax pacificus* sp. nov., isolated from a deep-sea pyrene-degrading consortium. *International Journal of Systematic and Evolutionary Microbiology*, 61(6):1370–1374, 2011. ISSN 1466-5034. doi: https://doi.org/10.1099/ijs.0.022368-0. URL https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.022368-0.

P. Lake. Disturbance, patchiness, and diversity in streams. *Journal of the North American Benthological Society*, 19(4):573–592, 2000. doi: 10.2307/1468118.

U. Langer, L. Böhme, and F. Böhme. Classification of soil microorganisms based on growth properties: a critical view of some commonly used terms. *Journal of Plant Nutrition and Soil Science*, 167(3):267–269, 2004. ISSN 1436-8730. doi: https://doi.org/10.1002/jpln.200421362. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jpln.200421362.

J. G. Leahy, P. J. Batchelor, and S. M. Morcomb. Evolution of the soluble di-iron monooxygenases. *FEMS Microbiology Reviews*, 27(4):449–479, 2003. ISSN 1574-6976. doi: 10.1016/S0168-6445(03)00023-8. URL http:https://doi.org/10.1016/S0168-6445(03)00023-8.

S.-W. Lee, D. R. Keeney, D.-H. Lim, A. A. Dispirito, and J. D. Semrau. Mixed pollutant degradation by *Methylosinus trichosporium* ob3b expressing either soluble or particulate methane monooxygenase: Can the tortoise beat the hare? *Applied and Environmental Microbiology*, 72(12):7503–7509, 2006. doi: doi:10.1128/AEM.01604-06. URL https://journals.asm.org/doi/abs/10.1128/AEM.01604-06.

D. J. Lessner. Methanogenesis biochemistry. *eLS*, 2001. ISSN 9780470015902. doi: 10.1002/9780470015902.a0000573.pub2. URL http://dx.doi.org/10.1002/9780470015902.a0000573.pub2.

M. D. Lewan. Experiments on the role of water in petroleum formation. *Geochimica et Cosmochimica Acta*, 61(17):3691–3723, 1997. ISSN 0016-7037. doi: https://doi.org/10.1016/S0016-7037(97)00176-2. URL http://www.sciencedirect.com/science/article/pii/S0016703797001762.

J. Li. spm: Spatial predictive modeling - ver. 1.2.0, 2019. URL https://CRAN.R-project.org/package=spm.

M. Li, J. Mathieu, Y. Yang, S. Fiorenza, Y. Deng, Z. He, J. Zhou, and P. J. J. Alvarez. Widespread distribution of soluble di-iron monooxygenase (sdimo) genes in arctic groundwater impacted by 1,4-dioxane. *Environmental Science & Technology*, 47(17):9950–9958, 2013. ISSN 0013-936X. doi: 10.1021/es402228x. URL https://doi.org/10.1021/es402228x.

S. Liebner and D. Wagner. Abundance, distribution and potential activity of methane oxidizing bacteria in permafrost soils from the lena delta, siberia. *Environmental Microbiology*, 9(1):107–117, 2007. ISSN 1462-2920. doi: 10.1111/j.1462-2920.2006.01120.x. URL http://dx.doi.org/10.1111/j.1462-2920.2006.01120.x.

J. D. Lipscomb. Biochemistry of the soluble methane monooxygenase. *Annual review of microbiology*, 48:371–399, 1994. ISSN 0066-4227. doi: 10.1146/annurev.mi.48.100194.002103. URL http://europepmc.org/abstract/MED/7826011https://doi.org/10.1146/annurev.mi.48.100194.002103.

Y.-C. Liu, Z. He, S. Zhang, M.-Y. Yin, Z. Ning, and C.-Y. Zhang. Abundance and diversity of methanotrophs and propanotrophs in soils above yangxin oil reservoir, china. *Geomicrobiology Journal*, 33(8):661–670, 2016. ISSN 0149-0451. doi: 10.1080/01490451.2015.1076542. URL https://doi.org/10.1080/01490451.2015.1076542.

K. Ma, R. Conrad, and Y. Lu. Dry/wet cycles change the activity and population dynamics of methanotrophs in rice field soil. *Applied and Environmental Microbiology*, 79(16):4932–4939, 2013. doi: doi:10.1128/AEM.00850-13. URL https://journals.asm.org/doi/abs/10.1128/AEM.00850-13.

J. L. Macalady, A. M. S. McMillan, A. F. Dickens, S. C. Tyler, and K. M. Scow. Population dynamics of type i and ii methanotrophic bacteria in rice soils. *Environmental Microbiology*, 4(3):148–157, 2002. ISSN 1462-2912. doi: https://doi.org/10.1046/j.1462-2920.2002.00278.x. URL https://sfamjournals.onlinelibrary.wiley.com/doi/abs/10.1046/j.1462-2920.2002.00278.x.

I. M. P. Machado and S. Atsumi. Cyanobacterial biofuel production. *Journal of Biotechnology*, 162(1):50–56, 2012. ISSN 0168-1656. doi: https://doi.org/10.1016/j.jbiotec.2012.03.005. URL http://www.sciencedirect.com/science/article/pii/S0168165612001721.

R. Margesin, A. Zimmerbauer, and F. Schinner. Monitoring of bioremediation by soil biological activities. *Chemosphere*, 40(4):339–346, 2000. ISSN 0045-6535. doi: https://doi.org/10.1016/S0045-6535(99)00218-0. URL http://www.sciencedirect.com/science/article/pii/S0045653599002180.

C. Martineau, L. G. Whyte, and C. W. Greer. Stable isotope probing analysis of the diversity and activity of methanotrophic bacteria in soils from the canadian high arctic. *Applied and Environmental Microbiology*, 76(17):5773–5784, 2010. doi: 10.1128/aem.03094-09. URL http://aem.asm.org/content/76/17/5773.abstract.

A. M. Martini, L. M. Walter, J. M. Budai, T. C. W. Ku, C. J. Kaiser, and M. Schoell. Genetic and temporal relations between formation waters and biogenic methane: Upper devonian antrim shale, michigan basin, usa. *Geochimica et Cosmochimica Acta*, 62(10):1699–1720, 1998. ISSN 0016-7037. doi: https://doi.org/10.1016/S0016-7037(98)00090-8. URL http://www.sciencedirect.com/science/article/pii/S0016703798000908.

T. L. Maugeri, C. Gugliandolo, D. Caccamo, and E. Stackebrandt. Three novel halotolerant and thermophilic *Geobacillus* strains from shallow marine vents. *Systematic and Applied Microbiology*, 25(3):450–455, 2002. ISSN 0723-2020. doi: https://doi.org/10.1078/0723-2020-00119. URL https://www.sciencedirect.com/science/article/pii/S0723202004701319.

D. McDonald, J. C. Clemente, J. Kuczynski, J. R. Rideout, J. Stombaugh, D. Wendel, A. Wilke, S. Huse, J. Hufnagle, F. Meyer, R. Knight, and J. G. Caporaso. The biological observation matrix (biom) format or: how i learned to stop worrying and love the ome-ome. *GigaScience*, 1(1), 2012. ISSN 2047-217X. doi: 10.1186/2047-217x-1-7. URL https://doi.org/10.1186/2047-217X-1-7.

I. R. McDonald, C. B. Miguez, G. Rogge, D. Bourque, K. D. Wendlandt, D. Groleau, and J. C. Murrell. Diversity of soluble methane monooxygenase-containing methanotrophs isolated from polluted environments. *FEMS Microbiology Letters*, 255(2):225–232, 2006. ISSN 0378-1097. doi: 10.1111/j.1574-6968.2005.00090.x. URL https://doi.org/10.1111/j.1574-6968.2005.00090.x.

I. R. McDonald, L. Bodrossy, Y. Chen, and J. C. Murrell. Molecular ecology techniques for the study of aerobic methanotrophs. *Applied and Environmental Microbiology*, 74(5):1305–1315, 2008. ISSN 0099-2240 1098-5336. doi: 10.1128/AEM.02233-07. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2258629/.

S. McEldowney, D. Hardman, and S. Waite. *Pollution: ecology and biotreatment*. Longman Scientific & Technical, 1993. ISBN 9780582086555. URL https://books.google.co.uk/books?id=tvxRAAAAMAAJ.

J. McIntosh, A. Martini, S. Petsch, R. Huang, and K. Nüsslein. Biogeochemistry of the forest city basin coalbed methane play. *International Journal of Coal Geology*, 76(1):111–118, 2008. ISSN 0166-5162. doi: https://doi.org/10.1016/j.coal.2008.03.004. URL http://www.sciencedirect.com/science/article/pii/S0166516208000499.

P. J. McMurdie and S. Holmes. phyloseq: An r package for reproducible interactive analysis and graphics of microbiome census data. *PLOS ONE*, 8(4):e61217, 2013. doi: 10.1371/journal.pone.0061217. URL https://doi.org/10.1371/journal.pone.0061217.

R. J. Melcher, S. E. Apitz, and B. B. Hemmingsen. Impact of irradiation and polycyclic aromatic hydrocarbon spiking on microbial populations in marine sediment

for future aging and biodegradability studies. *Applied and Environmental Microbiology*, 68(6):2858–2868, 2002. doi: 10.1128/AEM.68.6.2858-2868.2002. URL https://doi.org/10.1128/AEM.68.6.2858-2868.2002.

P. B. Miqueletto, F. D. Andreote, A. C. F. Dias, J. C. Ferreira, E. V. dos Santos Neto, and V. M. de Oliveira. Cultivation-independent methods applied to the microbial prospection of oil and gas in soil from a sedimentary basin in brazil. *AMB Express*, 1:35–35, 2011. ISSN 2191-0855. doi: 10.1186/2191-0855-1-35. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3282667/.

P. Nathanail, C. McCaffrey, A. Gillett, R. Ogden, and J. Nathanail. The lqm/cieh s4uls for human health risk assessement. Report, 2015.

National Center for Biotechnology Information. National library of medicine (us), national center for biotechnology information, 2021. URL https://www.ncbi.nlm.nih.gov/.

T. Nolan, R. E. Hands, W. Ogunkolade, and S. A. Bustin. Spud: A quantitative pcr assay for the detection of inhibitors in nucleic acid preparations. *Analytical Biochemistry*, 351(2):308–310, 2006. ISSN 0003-2697. doi: https://doi.org/10.1016/j.ab.2006.01.051. URL http://www.sciencedirect.com/science/article/pii/S0003269706000856.

C. W. Nossa, W. E. Oberdorf, L. Yang, J. A. Aas, B. J. Paster, T. Z. Desantis, E. L. Brodie, D. Malamud, M. A. Poles, and Z. Pei. Design of 16s rrna gene primers for 454 pyrosequencing of the human foregut microbiome. *World journal of gastroenterology*, 16(33):4135–4144, 2010. ISSN 2219-2840 1007-9327. doi: 10.3748/wjg.v16.i33.4135. URL https://pubmed.ncbi.nlm.nih.gov/20806429https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2932916/.

G. Nyerges and L. Y. Stein. Ammonia cometabolism and product inhibition vary considerably among species of methanotrophic bacteria. *FEMS Microbiology Letters*, 297(1):131–136, 2009. ISSN 0378-1097. doi: 10.1111/j.1574-6968.2009.01674.x. URL https://doi.org/10.1111/j.1574-6968.2009.01674.x.

J. Oksanen, F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. vegan: Community ecology package - ver. 2.5-7, 2020. URL https://CRAN.R-project.org/package=vegan.

H. J. M. Op den Camp, T. Islam, M. B. Stott, H. R. Harhangi, A. Hynes, S. Schouten, M. S. M. Jetten, N.-K. Birkeland, A. Pol, and P. F. Dunfield. Environmental, genomic and taxonomic perspectives on methanotrophic verrucomicrobia. *Environmental Microbiology Reports*, 1(5):293–306, 2009. ISSN 1758-2229. doi: 10.1111/j.1758-2229.2009.00022.x. URL http://dx.doi.org/10.1111/j.1758-2229.2009.00022.x.

R. S. Oremland, M. J. Whiticar, F. E. Strohmaier, and R. P. Kiene. Bacterial ethane formation from reduced, ethylated sulfur compounds in anoxic sediments. *Geochimica et Cosmochimica Acta*, 52(7):1895–1904, 1988. ISSN 0016-7037. doi: https://doi.org/10.1016/0016-7037(88)90013-0. URL http://www.sciencedirect.com/science/article/pii/0016703788900130.

S. G. Osborn and J. C. McIntosh. Chemical and isotopic tracers of the contribution of microbial gas in devonian organic-rich shales and reservoir sandstones, northern appalachian basin. *Applied Geochemistry*, 25(3):456–471, 2010. ISSN 0883-2927. doi: https://doi.org/10.1016/j.apgeochem.2010.01.001. URL http://www.sciencedirect.com/science/article/pii/S0883292710000065.

O. Paliy and V. Shankar. Application of multivariate statistical techniques in microbial ecology. *Molecular Ecology*, 25(5):1032–1057, 2016. ISSN 0962-1083. doi: https://doi.org/10.1111/mec.13536. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13536.

V. C. Pandey, J. S. Singh, D. P. Singh, and R. P. Singh. Methanotrophs: promising bacteria for environmental remediation. *International Journal of Environmental Science and Technology*, 11(1):241–250, 2014. ISSN 1735-2630. doi: 10.1007/s13762-013-0387-9. URL https://doi.org/10.1007/s13762-013-0387-9.

E. Pebesma. Simple features for r: Standardized support for spatial vector data. *The R Journal*, 10(1):439–446, 2018. doi: https://doi.org/10.32614/RJ-2018-009.

E. J. Pebesma. Multivariable geostatistics in s: the gstat package. *Computers & Geosciences*, 30(7):683–691, 2004. ISSN 0098-3004. doi: https://doi.org/10.1016/j.cageo.2004.03.012. URL https://www.sciencedirect.com/science/article/pii/S0098300404000676.

J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R. C. Team. nlme: Linear and nonlinear mixed effects models ver. 3.1-152, 2021. URL https://CRAN.R-project.org/package=nlme.

P. Pornwongthong, A. Mulchandani, P. B. Gedalanga, and S. Mahendra. Transition metals and organic ligands influence biodegradation of 1,4-dioxane. *Applied Biochemistry and Biotechnology*, 173(1):291–306, 2014. ISSN 1559-0291. doi: 10.1007/s12010-014-0841-2. URL https://doi.org/10.1007/s12010-014-0841-2.

C. Povedano-Priego, F. Jroundi, M. Lopez-Fernandez, R. Shrestha, R. Spanek, I. Martín-Sánchez, M. V. Villar, A. evc?, M. Dopson, and M. L. Merroun. Deciphering indigenous bacteria in compacted bentonite through a novel and efficient dna extraction method: Insights into biogeochemical processes within the deep geological disposal of nuclear waste concept. *Journal of Hazardous Materials*, 408:124600, 2021. ISSN 0304-3894. doi: https://doi.org/10.1016/j.jhazmat.2020.124600. URL https://www.sciencedirect.com/science/article/pii/S0304389420325905.

R Core Team. R: A language and environment for statistical computing, 2021. URL https://www.R-project.org/.

I. Rachor, J. Gebert, A. Gröngröft, and E.-M. Pfeiffer. Assessment of the methane oxidation capacity of compacted soils intended for use as landfill cover materials. *Waste Management*, 31(5):833–842, 2011. ISSN 0956-053X. doi: https://doi.org/10.1016/j.wasman.2010.10.006. URL http://www.sciencedirect.com/science/article/pii/S0956053X10005301.

M. T. Rahman, A. Crombie, Y. Chen, N. Stralis-Pavese, L. Bodrossy, P. Meir, N. P. McNamara, and J. C. Murrell. Environmental distribution and abundance of the facultative methanotroph *Methylocella*. *The ISME Journal*, 5:1061–1066, 2011a. doi: 10.1038/ismej.2010.190https://www.nature.com/articles/ismej2010190#supplementary-information. URL http://dx.doi.org/10.1038/ismej.2010.190.

M. T. Rahman, A. Crombie, H. Moussard, Y. Chen, and J. C. Murrell. Acetate repression of methane oxidation by supplemental *Methylocella silvestris* in a peat soil microcosm. *Applied and Environmental Microbiology*, 77(12):4234–4236, 2011b. doi: doi:10.1128/AEM.02902-10. URL https://journals.asm.org/doi/abs/10.1128/AEM.02902-10%XMethylocellaspp.arefacultativemethanotrophsthatgrowonmethaneandmulticarbonsubstratsuchasacetate.Acetaterepressestranscriptionofmethanemonooxygenasec protect\unhbox\voidb@x\bgroup\def, {Methylocellasilvestris}\let\futurelet\@let@

token\let\itshapeMethylocellasilvestris\
egroupinlaboratoryculture.DNAstable-isotopeprobing(DNA-SIP)
using13C-methaneand12C-acetate,carriedoutwithMethylocella-spikedpe
showedthatacetatealsorepressedmethaneoxidationby\
protect\unhbox\voidb@x\bgroup\def,{Methylocella}\
let\futurelet\@let@token\let\itshapeMethylocella\
egroupinenvironmentalsamples.

A. Ramette. Multivariate analyses in microbial ecology. *FEMS Microbiology Ecology*, 62(2):142–160, 2007. ISSN 0168-6496. doi: 10.1111/j.1574-6941.2007.00375.x. URL https://doi.org/10.1111/j.1574-6941.2007.00375.x.

M. Rastogi, S. Singh, and H. Pathak. Emission of carbon dioxide from soil. *Current Science*, 82(5):510–517, 2002. ISSN 00113891. URL http://www.jstor.org/stable/24105957.

T. Ren, J. A. Amaral, and R. Knowles. The response of methane consumption by pure cultures of methanotrophic bacteria to oxygen. *Canadian Journal of Microbiology*, 43(10):925–928, 1997. doi: 10.1139/m97-133. URL https://cdnsciencepub.com/doi/abs/10.1139/m97-133.

P. J. Ribeiro Jr, P. J. Diggle, M. Schlather, R. Bivand, and B. Ripley. geor: Analysis of geostatistical data ver. 1.8-1, 2020. URL https://CRAN.R-project.org/package=geoR.

F. F. Rochman, M. Kwon, R. Khadka, I. Tamas, A. A. Lopez-Jauregui, A. Sheremet, A. V. Smirnova, R. R. Malmstrom, S. Yoon, T. Woyke, P. F. Dunfield, and T. J. Verbeke. Novel copper-containing membrane monooxygenases (cummos) encoded by alkane-utilizing betaproteobacteria. *The ISME Journal*, 14(3):714–726, 2020. ISSN 1751-7370. doi: 10.1038/s41396-019-0561-2. URL https://doi.org/10.1038/s41396-019-0561-2.

L. L. Rockwood. *Introduction to Population Ecology*. John Wiley & Sons, Incorporated, Hoboken, UNITED KINGDOM, 2015. ISBN 9781118947562. URL http://ebookcentral.proquest.com/lib/nottingham/detail.action?docID=1895786.

H. Rodhe. A comparison of the contribution of various gases to the greenhouse effect. *Science*, 248(4960):1217–1219, 1990. doi: 10.1126/science.248.4960.1217. URL http://science.sciencemag.org/content/sci/248/4960/1217.full.pdf.

F. Rojo. Degradation of alkanes by bacteria. *Environmental Microbiology*, 11(10): 2477–2490, 2009. ISSN 1462-2912. doi: https://doi.org/10.1111/j.1462-2920.2009.01948.x. URL https://sfamjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1462-2920.2009.01948.x.

M. Saunois, A. R. Stavert, B. Poulter, P. Bousquet, J. G. Canadell, R. B. Jackson, P. A. Raymond, E. J. Dlugokencky, S. Houweling, P. K. Patra, P. Ciais, V. K. Arora, D. Bastviken, P. Bergamaschi, D. R. Blake, G. Brailsford, L. Bruhwiler, K. M. Carlson, M. Carrol, S. Castaldi, N. Chandra, C. Crevoisier, P. M. Crill, K. Covey, C. L. Curry, G. Etiope, C. Frankenberg, N. Gedney, M. I. Hegglin, L. Höglund-Isaksson, G. Hugelius, M. Ishizawa, A. Ito, G. Janssens-Maenhout, K. M. Jensen, F. Joos, T. Kleinen, P. B. Krummel, R. L. Langenfelds, G. G. Laruelle, L. Liu, T. Machida, S. Maksyutov, K. C. McDonald, J. McNorton, P. A. Miller, J. R. Melton, I. Morino, J. Müller, F. Murguia-Flores, V. Naik, Y. Niwa, S. Noce, S. O'Doherty, R. J. Parker, C. Peng, S. Peng, G. P. Peters, C. Prigent, R. Prinn, M. Ramonet, P. Regnier, W. J. Riley, J. A. Rosentreter, A. Segers, I. J. Simpson, H. Shi, S. J. Smith, L. P. Steele, B. F. Thornton, H. Tian, Y. Tohjima, F. N. Tubiello, A. Tsuruta, N. Viovy, A. Voulgarakis, T. S. Weber, M. van Weele, G. R. van der Werf, R. F. Weiss, D. Worthy, D. Wunch, Y. Yin, Y. Yoshida, W. Zhang, Z. Zhang, Y. Zhao, B. Zheng, Q. Zhu, Q. Zhu, and Q. Zhuang. The global methane budget 2000–2017. *Earth Syst. Sci. Data*, 12(3):1561–1623, 2020. ISSN 1866-3516. doi: 10.5194/essd-12-1561-2020. URL https://essd.copernicus.org/articles/12/1561/2020/.

M. E. Schlegel, J. C. McIntosh, B. L. Bates, M. F. Kirk, and A. M. Martini. Comparison of fluid geochemistry and microbiology of multiple organic-rich reservoirs in the illinois basin, usa: Evidence for controls on methanogenesis and microbial transport. *Geochimica et Cosmochimica Acta*, 75(7):1903–1919, 2011. ISSN 0016-7037. doi: https://doi.org/10.1016/j.gca.2011.01.016. URL http://www.sciencedirect.com/science/article/pii/S001670371100024X.

J. S. Seewald. Organicinorganic interactions in petroleum-producing sedimentary basins. *Nature*, 426(6964):327–333, 2003. ISSN 1476-4687. doi: 10.1038/nature02132. URL https://doi.org/10.1038/nature02132.

J. Semrau. Bioremediation via methanotrophy: Overview of recent findings and suggestions for future research. *Frontiers in Microbiology*, 2(209), 2011. ISSN 1664-302X. doi: 10.3389/fmicb.2011.00209. URL https://www.frontiersin.org/article/10.3389/fmicb.2011.00209.

J. D. Semrau, A. A. DiSpirito, and S. Yoon. Methanotrophs and copper. *FEMS*

*Microbiology Reviews*, 34(4):496–531, 2010. ISSN 1574-6976. doi: 10.1111/j.1574-6976.2010.00212.x. URL http://dx.doi.org/10.1111/j.1574-6976.2010.00212.x.

J. D. Semrau, S. Jagadevan, A. A. DiSpirito, A. Khalifa, J. Scanlan, B. H. Bergman, B. C. Freemeier, B. S. Baral, N. L. Bandow, A. Vorobev, D. H. Haft, S. Vuilleumier, and J. C. Murrell. Methanobactin and mmod work in concert to act as the copper-switch in methanotrophs. *Environmental Microbiology*, 15(11):3077–3086, 2013. ISSN 1462-2920. doi: 10.1111/1462-2920.12150. URL http://dx.doi.org/10.1111/1462-2920.12150.

A. Shade, H. Peter, S. Allison, D. Baho, M. Berga, H. Buergmann, D. Huber, S. Langenheder, J. Lennon, J. Martiny, K. Matulich, T. Schmidt, and J. Handelsman. Fundamentals of microbial community resistance and resilience. *Frontiers in Microbiology*, 3(417), 2012. ISSN 1664-302X. doi: 10.3389/fmicb.2012.00417. URL https://www.frontiersin.org/article/10.3389/fmicb.2012.00417.

S. H. Shakir Hanna and R. W. Weaver. Earthworm survival in oil contaminated soil. *Plant and Soil*, 240(1):127–132, 2002. ISSN 1573-5036. doi: 10.1023/A:1015816315477. URL https://doi.org/10.1023/A:1015816315477.

Y.-J. Shiau, Y. Cai, Z. Jia, C.-L. Chen, and C.-Y. Chiu. Phylogenetically distinct methanotrophs modulate methane oxidation in rice paddies across taiwan. *Soil Biology and Biochemistry*, 124:59–69, 2018. ISSN 0038-0717. doi: https://doi.org/10.1016/j.soilbio.2018.05.025. URL http://www.sciencedirect.com/science/article/pii/S0038071718301810.

K. P. Singh, T. N. Mandal, and S. K. Tripathi. Patterns of restoration of soil physciochemical properties and microbial biomass in different landslide sites in the sal forest ecosystem of nepal himalaya. *Ecological Engineering*, 17(4):385–401, 2001. ISSN 0925-8574. doi: https://doi.org/10.1016/S0925-8574(00)00162-2. URL https://www.sciencedirect.com/science/article/pii/S0925857400001622.

K. A. Smith, K. E. Dobbie, B. C. Ball, L. R. Bakken, B. K. Sitaula, S. Hansen, R. Brumme, W. Borken, S. Christensen, A. Priemé, D. Fowler, J. A. Macdonald, U. Skiba, L. Klemedtsson, A. Kasimir-Klemedtsson, A. Degórska, and P. Orlanski. Oxidation of atmospheric methane in northern european soils, comparison with other ecosystems, and uncertainties in the global terrestrial sink. *Global Change*

*Biology*, 6(7):791–803, 2000. ISSN 1365-2486. doi: 10.1046/j.1365-2486.2000.00356.x. URL http://dx.doi.org/10.1046/j.1365-2486.2000.00356.x.

D. Strąpoć, M. Mastalerz, K. Dawson, J. Macalady, A. V. Callaghan, B. Wawrik, C. Turich, and M. Ashby. Biogeochemistry of microbial coal-bed methane. *Annual Review of Earth and Planetary Sciences*, 39(1):617–656, 2011. doi: 10.1146/annurev-earth-040610-133343. URL https://www.annualreviews.org/doi/abs/10.1146/annurev-earth-040610-133343.

Y. Takimura, E. Sakuradani, Y. Natsume, T. Miyake, J. Ogawa, and S. Shimizu. Achlorophyllous alga *Prototheca zopfii* oxidizes n-alkanes with different carbon-chain lengths through a unique subterminal oxidation pathway. *Journal of Bioscience and Bioengineering*, 117(3):275–277, 2014. ISSN 1389-1723. doi: https://doi.org/10.1016/j.jbiosc.2013.09.004. URL https://www.sciencedirect.com/science/article/pii/S1389172313003253.

J. Tang, J. Zhang, L. Ren, Y. Zhou, J. Gao, L. Luo, Y. Yang, Q. Peng, H. Huang, and A. Chen. Diagnosis of soil contamination using microbiological indices: A review on heavy metal pollution. *Journal of Environmental Management*, 242:121–130, 2019. ISSN 0301-4797. doi: https://doi.org/10.1016/j.jenvman.2019.04.061. URL https://www.sciencedirect.com/science/article/pii/S0301479719305286.

S. W. Taylor, S. B. Lollar, and I. Wassenaar. Bacteriogenic ethane in near-surface aquifers:? implications for leaking hydrocarbon well bores. *Environmental Science & Technology*, 34(22):4727–4732, 2000. ISSN 0013-936X. doi: 10.1021/es001066x. URL https://doi.org/10.1021/es001066x.

C. C. Tebbe and W. Vahjen. Interference of humic acids and dna extracted directly from soil in detection and transformation of recombinant dna from bacteria and a yeast. *Applied and Environmental Microbiology*, 59(8):2657–2665, 1993. doi: doi:10.1128/aem.59.8.2657-2665.1993. URL https://journals.asm.org/doi/abs/10.1128/aem.59.8.2657-2665.1993.

S. A. Tedesco. *Surface Geochemistry in Petroleum Exploration*. Chapman and Hall, New York (US), 1995.

M. Tennekes. tmap: Thematic maps in r - ver. 3.3-1. *Journal of Statistical Software*, 1(6), 2018. doi: http://dx.doi.org/10.18637/jss.v084.i06.

URL https://www.jstatsoft.org/v084/i06http://dx.doi.org/10.18637/jss.v084.i06.

A. R. Theisen, M. H. Ali, S. Radajewski, M. G. Dumont, P. F. Dunfield, I. R. McDonald, S. N. Dedysh, C. B. Miguez, and J. C. Murrell. Regulation of methane oxidation in the facultative methanotroph methylocella silvestris bl2. *Molecular Microbiology*, 58(3):682–692, 2005. ISSN 0950-382X. doi: https://doi.org/10.1111/j.1365-2958.2005.04861.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2005.04861.x.

X. Tian, X. Wang, S. Peng, Z. Wang, R. Zhou, and H. Tian. Isolation, screening, and crude oil degradation characteristics of hydrocarbons-degrading bacteria for treatment of oily wastewater. *Water Science and Technology*, 78(12):2626–2638, 2019. ISSN 0273-1223. doi: 10.2166/wst.2019.025. URL https://doi.org/10.2166/wst.2019.025.

R. Tibshirani, G. Walther, and T. Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423, 2001. ISSN 1369-7412. doi: https://doi.org/10.1111/1467-9868.00293. URL https://doi.org/10.1111/1467-9868.00293.

UK Government. Onshore oil and gas exploration in the uk: regulation and best practice, 2015.

UK Oil and Gas Authority, 2019. URL https://www.ogauthority.co.uk/media/6108/c-pdf-landwells-24-october-2019.xlsx.

UK Onshore Geophysical Library, 2018. URL https://ukogl.org.uk/.

J. B. van Beilen and E. G. Funhoff. Alkane hydroxylases involved in microbial alkane degradation. *Applied Microbiology and Biotechnology*, 74(1):13–21, 2007. ISSN 1432-0614. doi: 10.1007/s00253-006-0748-0. URL https://doi.org/10.1007/s00253-006-0748-0.

J. B. Van Beilen, Z. Li, W. A. Duetz, T. H. Smits, and B. Witholt. Diversity of alkane hydroxylase systems in the environment. *Oil Gas Science and Technology - Rev. IFP*, 58(4):427–440, 2003. URL https://doi.org/10.2516/ogst:2003026.

D. van der Ha, I. Vanwonterghem, S. Hoefman, P. De Vos, and N. Boon. Selection of associated heterotrophs by methane-oxidizing bacteria at different copper

concentrations. *Antonie van Leeuwenhoek*, 103(3):527–537, 2013. ISSN 1572-9699. doi: 10.1007/s10482-012-9835-7. URL https://doi.org/10.1007/s10482-012-9835-7.

J. van Dorst, S. D. Siciliano, T. Winsley, I. Snape, and B. C. Ferrari. Bacterial targets as potential indicators of diesel fuel toxicity in subantarctic soils. *Applied and Environmental Microbiology*, 80(13):4021–4033, 2014. ISSN 1098-5336-0099-2240. doi: 10.1128/AEM.03939-13. URL https://www.ncbi.nlm.nih.gov/pubmed/24771028.

D. Van Stempvoort, H. Maathuis, E. Jaworski, B. Mayer, and K. Rich. Oxidation of fugitive methane in ground water linked to bacterial sulfate reduction. *Ground Water*, 43:187–199, 2005.

M. S. Vecherskaya, V. F. Galchenko, E. N. Sokolova, and V. A. Samarkin. Activity and species composition of aerobic methanotrophic communities in tundra soils. *Current Microbiology*, 27(3):181–184, 1993. ISSN 1432-0991. doi: 10.1007/bf01576018. URL https://doi.org/10.1007/BF01576018.

A. J. Veraart, P. Garbeva, F. van Beersum, A. Ho, C. A. Hordijk, M. Meima-Franke, A. J. Zweers, and P. L. E. Bodelier. Living apart together bacterial volatiles influence methanotrophic growth and activity. *The ISME Journal*, 12(4):1163–1166, 2018. ISSN 1751-7370. doi: 10.1038/s41396-018-0055-7. URL https://doi.org/10.1038/s41396-018-0055-7.

C. Visvanathan, D. Pokhrel, W. Cheimchaisri, J. Hettiaratchi, and J. Wu. Methanotrophic activities in tropical landfill cover soils: effects of temperature, moisture content and methane concentration. *Waste Management & Research*, 17(4):313–323, 1999. doi: 10.1177/0734242x9901700408. URL http://journals.sagepub.com/doi/abs/10.1177/0734242X9901700408.

J. C. von Fischer and L. O. Hedin. Controls on soil methane fluxes: Tests of biophysical mechanisms using stable isotope tracers. *Global Biogeochemical Cycles*, 21(2):GB2007, 2007. ISSN 1944-9224. doi: 10.1029/2006GB002687. URL http://dx.doi.org/10.1029/2006GB002687.

A. V. Vorobev, M. Baani, N. V. Doronina, A. L. Brady, W. Liesack, P. F. Dunfield, and S. N. Dedysh. *Methyloferula stellata* gen. nov., sp. nov., an acidophilic, obligately methanotrophic bacterium that possesses only a soluble methane monooxygenase. *International Journal of Systematic and Evolutionary Microbiology*, 61(10):2456–2463, 2011. doi: doi:10.1099/ijs.0.028118-0.

URL http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.028118-0.

M. Wagner, M. Wagner, J. Piske, and R. Smit. *Case Histories of Microbial Prospection for Oil and Gas Onshore and Offshore in northwest Europe*. SEG Geophysical References Series: Surface Exploration Case Histories: Applications of Geochemistry, Magnetics, and Remote Sensing. The American Association of Petroleum Geologists and the Society of Exploration Geophysicists, 2002. ISBN 0-89181-055-2.

Y. Wang, D. E. Prosen, L. Mei, J. C. Sullivan, M. Finney, and P. B. Vander Horn. A novel strategy to engineer dna polymerases for enhanced processivity and improved performance *in vitro*. *Nucleic Acids Research*, 32(3): 1197–1207, 2004. ISSN 1362-4962-0305-1048. doi: 10.1093/nar/gkh271. URL https://www.ncbi.nlm.nih.gov/pubmed/14973201https://www.ncbi.nlm.nih.gov/pmc/PMC373405/.

Y. Wang, Y. Zhao, A. Bollas, Y. Wang, and K. F. Au. Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, 39(11):1348–1365, 2021. ISSN 1546-1696. doi: 10.1038/s41587-021-01108-x. URL https://doi.org/10.1038/s41587-021-01108-x.

T. Watson and S. Bachu. Evaluation of the potential for gas and $CO_2$ leakage along wellbores. *SPE Drilling & Completion*, 24:115–126, 2009. doi: 10.2118/106817-MS.

T. Wei and V. Simko. R package 'corrplot': Visualization of a correlation matrix - ver: 0.90, 2021.

R. E. White. *Principles and Practice of Soil Science*. Blackwell Publishing, Malden, MA, USA, 4th edition, 2006.

M. J. Whiticar. Carbon and hydrogen isotope systematics of bacterial formation and oxidation of methane. *Chemical Geology*, 161(1):291–314, 1999. ISSN 0009-2541. doi: https://doi.org/10.1016/S0009-2541(99)00092-3. URL http://www.sciencedirect.com/science/article/pii/S0009254199000923.

R. Whittenbury, K. C. Phillips, and J. F. Wilkinson. Enrichment, isolation and some properties of methane-utilizing bacteria. *Microbiology*, 61(2):205–218, 1970. doi: doi:10.1099/00221287-61-2-205. URL http://mic.microbiologyresearch.org/content/journal/micro/10.1099/00221287-61-2-205.

H. Wickham. ggplot2: Elegant graphics for data analysis - ver. 3.3.3, 2016. URL https://ggplot2.tidyverse.org.

I. G. Wilson. Inhibition and facilitation of nucleic acid amplification. *Applied and Environmental Microbiology*, 63(10):3741–3751, 1997. ISSN 0099-2240 1098-5336. URL https://pubmed.ncbi.nlm.nih.gov/9327537https://www.ncbi.nlm.nih.gov/pmc/articles/PMC168683/.

D. E. Wood, J. Lu, and B. Langmead. Improved metagenomic analysis with kraken 2. *Genome Biology*, 20(1):257, 2019. ISSN 1474-760X. doi: 10.1186/s13059-019-1891-0. URL https://doi.org/10.1186/s13059-019-1891-0.

M. M. Yakimov, P. N. Golyshin, S. Lang, E. R. B. Moore, W.-R. Abraham, H. Lünsdorf, and K. N. Timmis. Alcanivorax borkumensis gen. nov., sp. nov., a new, hydrocarbon-degrading and surfactant-producing marine bacterium. *International Journal of Systematic and Evolutionary Microbiology*, 48(2):339–348, 1998. ISSN 1466-5034. doi: https://doi.org/10.1099/00207713-48-2-339. URL https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-48-2-339.

E. Yergeau, H. Hogues, L. G. Whyte, and C. W. Greer. The functional potential of high arctic permafrost revealed by metagenomic sequencing, qpcr and microarray analyses. *The ISME Journal*, 4:1206–1214, 2010. doi: 10.1038/ismej.2010.41https://www.nature.com/articles/ismej201041#supplementary-information. URL http://dx.doi.org/10.1038/ismej.2010.41.

S. Yoon and J. D. Semrau. Measurement and modeling of multiple substrate oxidation by methanotrophs at 20°c. *FEMS Microbiology Letters*, 287(2):156–162, 2008. ISSN 0378-1097. doi: 10.1111/j.1574-6968.2008.01314.x. URL https://doi.org/10.1111/j.1574-6968.2008.01314.x.

I. M. Young and J. W. Crawford. Interactions and self-organization in the soil-microbe complex. *Science*, 304(5677):1634–1637, 2004. doi: 10.1126/science.1097394. URL https://science.sciencemag.org/content/sci/304/5677/1634.full.pdf.

I. M. Young, J. W. Crawford, N. Nunan, W. Otten, and A. Spiers. *Microbial Distribution in Soils: Physics and Scaling*, volume 100, book section 4, pages 81–121. Academic Press, 2008. ISBN 0065-2113. doi: https://doi.org/10.1016/S0065-2113(08)

00604-4. URL http://www.sciencedirect.com/science/article/pii/S0065211308006044.

C.-Y. Zhang, Z. He, S. Zhang, M.-Y. Yin, Z. Ning, and Y.-C. Liu. Quantitative significance of functional genes of methanotrophs and propanotrophs in soil above oil and gas fields, china. *Journal of Petroleum Science and Engineering*, 120:170–176, 2014. doi: 10.1016/j.petrol.2014.06.012.

C.-Y. Zhang, Z. He, S. Zhang, M.-Y. Yin, Z. Ning, and Y.-C. Liu. A dna-based analysis of a microbial technique for the prospecting of oil and gas applied to a known oil field, china. *Geomicrobiology Journal*, 34(1):63–70, 2017. doi: 10.1080/01490451.2016.1139641.

F.-P. Zhang, C.-F. Li, L.-G. Tong, L.-X. Yue, P. Li, Y.-J. Ciren, and C.-G. Cao. Response of microbial characteristics to heavy metal pollution of mining soils in central tibet, china. *Applied Soil Ecology*, 45(3):144–151, 2010. ISSN 0929-1393. doi: https://doi.org/10.1016/j.apsoil.2010.03.006. URL https://www.sciencedirect.com/science/article/pii/S092913931000048X.

J. Zhou, B. Xia, D. S. Treves, L. Y. Wu, T. L. Marsh, R. V. O'Neill, A. V. Palumbo, and J. M. Tiedje. Spatial and resource factors influencing high microbial diversity in soil. *Applied and Environmental Microbiology*, 68(1):326–334, 2002. ISSN 0099-2240 1098-5336. doi: 10.1128/aem.68.1.326-334.2002. URL https://pubmed.ncbi.nlm.nih.gov/11772642https://www.ncbi.nlm.nih.gov/pmc/articles/PMC126564/.

İnceoğlu, E. F. Hoogwout, P. Hill, and J. D. van Elsas. Effect of dna extraction method on the apparent microbial diversity of soil. *Applied and Environmental Microbiology*, 76(10):3378–3382, 2010. doi: 10.1128/aem.02715-09. URL https://aem.asm.org/content/aem/76/10/3378.full.pdf.