Investigation into the Potential Application of Microbial Enhanced Oil Recovery on Unconventional Oil: A Field Specific Approach

Sean Henry Goodman

200533353

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

December 2017

Abstract

A substantial amount of the world's recoverable oil reserves comprise unconventional resources. However great difficulty has been encountered in recovering oil lower than 22° API. Therefore, advanced methods of Enhanced oil recovery (EOR) such as microbial enhanced oil recovery (MEOR) have been employed to increase the amount of recovered residual oil. MEOR involves the use of bacteria and their metabolic products to alter the oil properties or rock permeability within a reservoir in order to promote the flow of oil. Although MEOR has been trialled in the past with mixed outcomes, its feasibility on heavier oils has not been fully demonstrated.

The aim of this study was to show that MEOR can be successfully applied to unconventional oil fields to increase oil production. Using both genomic and microbiologically applied petroleum engineering techniques, it was possible to target and isolate key indigenous microorganisms with MEOR potential from the reservoir of interest. In this study we have identified an indigenous microorganism (*Bacillus licheniformis* Bi10) that was capable of enhancing heavy oil recovery. This strain was applied to field specific microcosms and the effect of this microorganism was compared to variant inoculate, showing improved recovery beyond levels shown by previous MEOR related bacteria (Additional Oil Recovery- 11.8%). Furthermore, we also confirmed that the use of biosurfactant *lichenysin* alone was not as effective in MEOR compared to viable cell treatment, and hypothesized that a dual mechanism of action approach may be taking place within the microcosm, of both bio-plugging and wettability alteration. The interfacial tension of biosurfactant produced by the

Bi10 isolate also showed a substantial decrease in wettability calculations, to < 5 mNm^{-1}, lower than any other bacterial surfactants have been shown in heavy oil environments Comparative genomics also revealed key genetic variations between this and similar MEOR strains that could hold the key to its increased potential for future MEOR strategies.

The results presented in this thesis were part of an ERDF project, involving academic and industrial partner, BiSN Laboratory Services, on fundamental and applied aspects of microbial enhanced oil recovery in heavy oilfield environments, which was funded to improve the understanding of MEOR and its processes in these unconventional oil environments.

Acknowledgements

There are a number of people that I would like to thank and without whom the completion of this thesis would not be possible.

Firstly, I need to thank my academic supervisor, Dr Heather E Allison, for giving me this fantastic opportunity and having continual faith in me throughout the last four years. Not only have you supported me in my academic studies and helped to develop me as a scientist, but also often provided me with personal advice and helped me through some dark times, acting as my Microbiological parent. It is difficult to sum up in a short paragraph how much you have helped me and I will be ever in your debt (within reason).

I would also like to thank my secondary supervisor Prof Alan McCarthy and my assessors Dr Mal Horsburgh and Dr Peter Miller for their interest, advice and all that I have learned from them over my time at the University of Liverpool. The members of Lab H, both past and present have provided me with so much help and guidance during my PhD, alongside many humorous moments to keep me smiling along the way. I must particularly thank the likes of Anshul Gupta, Rhys Williams, Daniele Parisi, Barbara Franke, Bruno Manso, Chris Hill, Breno Salgado, Stuart McEwen, and Charlotte Chong (and of course Paddy & Duffy) for the continual support provided to me and the many nights of drinking that have helped to push me through to the other side of the formidable PhD tunnel.

I must also thank my industrial supervisor Andrew Millar and the managing director, Paul Carragher at BiSN Laboratory services, without would not have

achieved this momentous feat. I have learned so much about a new industry, working side by side with the company on industrial projects alongside PhD. I wish you both, the employees and the company itself all the best for your venture in the US.

A big thank you also goes to my Mother, Sandra Davies, who I cannot praise enough for your love, care and support (mental, physical and financial, and for putting a roof over my head when my funding ran out. I do not know what I would do without you and am blessed to have you as my mother, particularly for those many nights spent proofreading my thesis and taking me to LFC matches when I needed a break. And to my extended family also, of which there are too many to name, thank you for your well wishes and support.

And then we come to my wonderful fiancé, the soon to also be doctor, Sarah Long (although not a 'real doctor'). Thank you for putting up with my moans and complains, the evenings when I buried my head in books and research papers and also the continual love that you have provided me. Thank you, alongside Sandra, for nursing me through to a healthy state again, something which I did not think I would not recover from. You can rest easy now knowing that it is almost all out of the way.

Finally, I would like to dedicate this thesis to those who we have unfortunately lost during my time as a PhD student: My beloved Uncle Kevin, who was always there to encourage me, listen to my complaints and make me laugh with your hilarious southern ways. My grandmother, Joan Goodman, for the years of care

she gave to me as a child, continually encouraging me to pursue my dreams and aspirations and always taking an interest in my studies, even though you had no idea what I actually did; and last but certainly not least David Gardiner, you are always missed. I hope that I have made you all proud; you are all never far from my thoughts. YNWA

Contents

Investigation into the Potential Application of Microbial Enhanced Oil Recovery on Unconventional Oil

List of figures

List of tables

List of measurements

%(v/v)	percentage volume/volume
%(w/v)	percentage mass/volume
°C	Degrees Celsius
AOR	Additional Oil Recovery
bbl	Barrel (159 L)
BBL	Billion barrels
cfu	colony forming unit
cP	Centipoise
ft	Feet
g L^{-1}, mg L^{-1}	grams per litre, milligrams per litre , micrograms per litre
g, mg, µg, ng	Gram, milligram, microgram, nanogram
km, m, cm , mm	metre, centimeter,
L, mL, µL	Litre, mililitre, microlitre
mD	Milidarcy
mi	Mile
MMbbl	Megabarrel (Million barrels)
mNm-1	Millinewton/meter
ppm	parts per million
psi	Pounds per square inch
rpm	Revolutions per minute
V	Volume
wt.	Mass

List of abbreviations

API	American Petroleum Index
ATCC	American Type Culture Collection
CCA	Canonical correlation analysis
cDNA	Complementary DNA
CMC	Critical Micelle Concentration
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
EOR	Enhanced Oil Recovery
EPS	Extracellular Polymeric Substance
FW	Formation water
IFT	Interfacial tension
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG Orthologous groups
LB	Luria broth
Lch	Lichenysin
Lic	Lichenysin
MEOR	Microbial EOR
MIC	Microbial influenced corrosion
NGS	Next Generation Sequencing
NRP	Non ribosomal peptide
NRPS	Non ribosomal peptide synthetase
OOIP	Oil in place
OUT	Operational Taxonomic Unit
OWC	Oil-water Contact
PCoA	Principle coordinate analysis
PCR	Polymerase Chain Reaction
pH	Hydronium ion concentration; pH=-log[H+]
PV	Pore volume
PW	Production water
qPCR	Quantitative PCR
RDP	Ribosomal Database Project
RNA	Ribonucleic acid
rRNA	ribosomal RNA
SAC	Surface Active Compound
SRB	Sulphate Reducing Bacteria
Srf	Surfactin
SRP	Sulphate Reducing Prokaryote
VFA	Volatile Fatty Acid

Chapter 1

Introduction

1.1 Background

Crude oil is currently the world's principal form of non-renewable energy, with an estimated 90 million barrels used worldwide each day in 2012 *(*EIA*, 2013)*. It is a naturally occurring energy source found in formation rock or permeable subsurface rock that contains a mixture of hydrocarbons and numerous organic elements such as nitrogen, hydrogen and sulphur (Redman *et al.,* 2012*).* The type of oil varies depending upon its relative density. Heavier oils are defined as having an API (American Petroleum Index measurement of viscosity) less than 22° and contain higher numbers of carbon atoms per molecule. Over half of the oil available for recovery still remains in the ground, and following current practices, heavy/extra heavy oils may never be reclaimed.

Although the demand for oil is continually increasing, the availability of this non-renewable source is in rapid decline. There is now a great need to recover some, if not all of this trapped oil and postpone the impending 'peak' oil crisis for at least 100 years (Figure 1.1) (Hubbert, 1956). Furthermore, although there is a global drive to increase the promotion and use of renewable energy sources, the lack of implementable technologies that can easily and comprehensively produce quantities of energy similar to that of fossil fuels, signifies the importance of accessing these currently unrecoverable energy reserves.

The recovery of oil from field deposits is performed in three different phases. In the *primary phase*, oil is displaced from the rock by a progressive drop in reservoir pressure, caused by opening the well at the surface. As little as 10% of the oil in a well can be recovered via primary production (Sen, 2008a). The *secondary phase* recovery of oil is achieved by forcing either gas or water (water flooding) into the well to increase the reservoir pressure, pushing the oil out where it can be collected. Although secondary production processes substantially increase oil recovery, the efficiency only allows for between 15 and 40% of the total oil to be recovered. This percentage is even lower for heavier oils *(*Sen*,* 2008*;* Tunio *et al,* 2011*)*. More recently, newer *tertiary phase* recovery methods have been developed, known as Thermal and Chemical Enhanced Oil Recovery (EOR) (Hosseini-Nasab et al., 2016; Lu *et al*., 2014) . These methods include the use of steam or water injection of surfactants, gases, polymers and solvents, each acting to improve oil recovery (Latil, 1980).

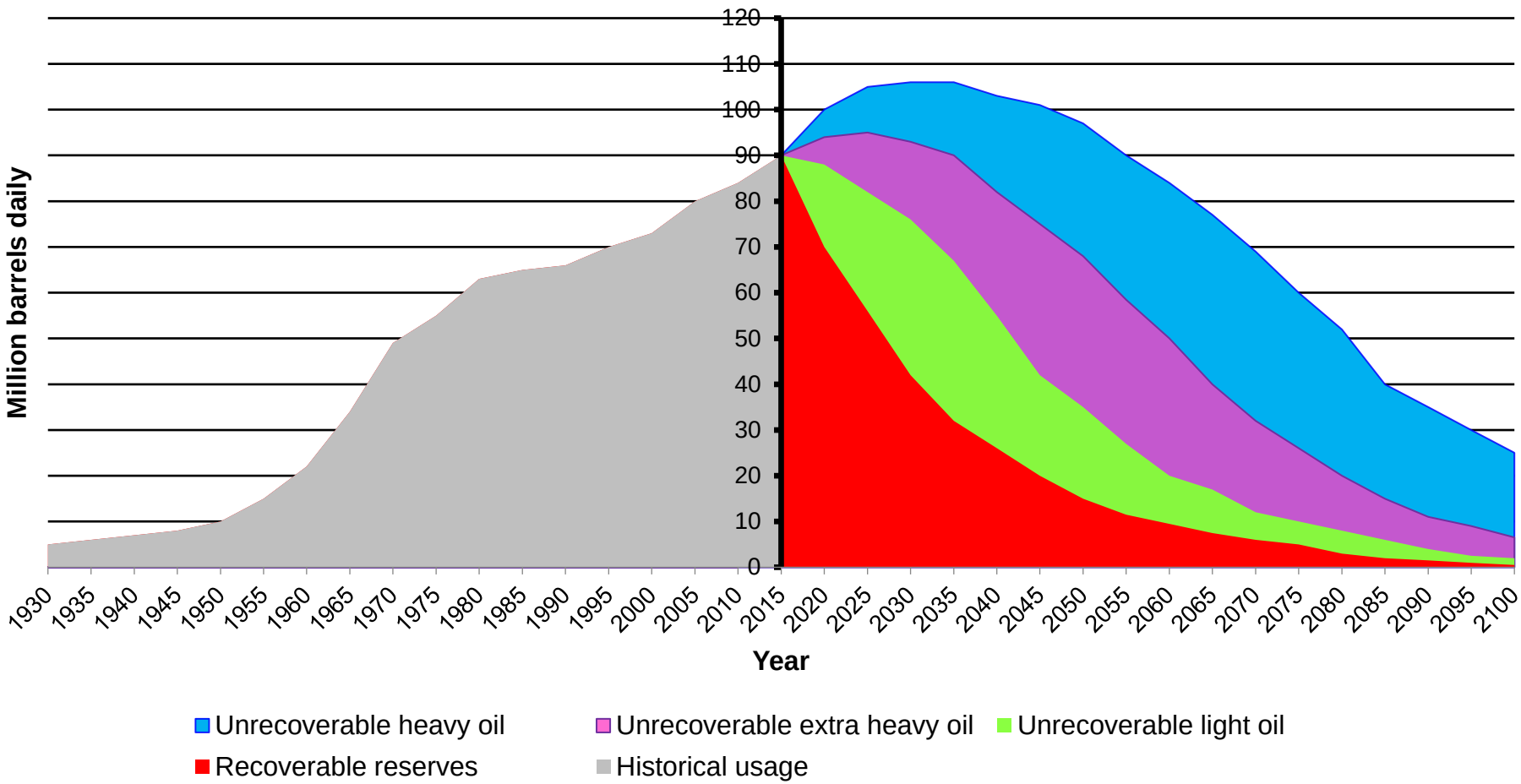


Figure 1.1 Historical worldwide oil production compared to approximate future oil reserve usage. Data amended from BP World Energy Review, 2015.

Although these tertiary phase methods of EOR are currently in use, their high cost, associated difficulties and expenses with chemical waste disposal have made the oil industry search for alternative approaches *(*Bordoloi & Konwar*,* 2008*;* Green & Willhite*,* 1998*)*. These EOR methods have only been shown to increase oil recovery by a maximum of 30% in light oils (Delamaide *et al* , 2014). Furthermore, alongside the disadvantages mentioned above, the chemical EOR methods are often hindered by a reduction in activity in the hostile conditions associated with oil reservoirs (e.g. pH, temperature, pressure) (Shukla, 2017). However, if it was possible to use biological materials that were similar in activity to the chemical EOR processes, but that were specific to the oilfield and could be produced *in situ*, this would theoretically increase the amount of oil that could be recovered, whilst reducing production/transportation and waste removal expenditures.

1.2 Microbial Enhanced Oil Recovery

Microbial enhanced oil recovery (MEOR) is an alternative tertiary recovery strategy that uses microbial populations and their products to increase oil recovery from low producing reservoirs after primary and secondary extraction methods have been exhausted. This technique manipulates many well-documented microbial metabolic products (*e.g*. biopolymers, biosurfactants, acids, biomass, biogas) to alter the properties of the oil flow and reservoir to promote the movement of oil in the well. The idea of using microbes in oil production was first described in the early twentieth century, by both Beckman (1926) and Zobell *(*1947*).* Although it is one of the older EOR technologies, the method has not been as successfully documented as other EOR

techniques. This is primarily due to differing degrees of success in the laboratory, lack of data and failures occurring in field trial scale-up (Rosnes *et al.*, 1991; Sheehy, 1991b). It is only in the past decade, as knowledge of bacterial metabolic products and processes have improved, that MEOR has been considered to be an effective and economical method to enhance oil production *(*Karimi *et al.,* 2012*)*.

There are two classifications of MEOR: *Bioaugmentation* and *Biostimulation*. In the process of bioaugmentative MEOR, allochthonous bacteria that can survive the extreme conditions of the reservoir (pH 4-9; salinity ~0.15-15%; temperature ~20-130ºC; oxygen presence and pressure ~1000-50,000 psi) are added to a well along with appropriate nutrients to aid their growth and establishment (*ex-situ* microbial growth). Biostimulation however, involves the encouragement of autochthonous (indigenous) bacterial growth within the reservoir through the addition of specific nutrients. These microorganisms then alter the physiological or geochemical features of the reservoir through metabolite production. This nutrient injection based approach has often been preferred, as indigenous bacteria are better equipped to survive these diverse environmental parameters. However, although in theory, the precise stimulation of only the microorganisms required might be a more successful approach, the real world ability to restrict growth stimulation of only 'beneficial' microorganisms can be a challenging process, and getting it wrong can result in negative effects on the reservoir, such as souring and microbial influenced corrosion (Dong *et al*., 2015; Jackson *et al*., 2010) .

1.3 Microbial Mechanisms of Oil Recovery

There are many ways in which bacteria can alter the properties of the oil and rock including, but not exclusive to, selective plugging and reduction in interfacial tension or wettability (Table 1.1). The most significant and commonly investigated mechanism is the reduction in interfacial tension (IFT) between oil and water, which has shown great potential for enhancing oil recovery (Biria *et al.,* 2007*;* Fang *et al.,* 2007*;* Amani *et al.,* 2010). IFT refers to the contracting force between two states of matter, but in this specific case refers to the oil-water or oil-rock boundaries. By reducing the IFT, capillary and viscous forces are reduced, favouring mobilisation of the oil *(*Nobakht, 2007*).*

Alternatively, the alteration of wettability, the selective plugging of rock pores or channels in the wells (Gray *et al*. 2008*)* and the dissolution of rock *(Jasjfi et al., 1995)* are other mechanisms of action that can also substantially increase oil production. Most of these mechanisms involve the production of cellular biomass or microbial products. These bioproducts act in similar ways to effectors of chemical EOR (Bryant & Lockhart, 2000*),* but unlike effectors of CEOR, bioproducts of MEOR are both environmentally and economically beneficial when compared to their chemical counterparts. Microbial products create fewer chemical disposal issues, hence reducing capital expenditure in clean-up costs. Some studies have also shown that microbial methods can result in a higher incremental oil recovery than the equivalent chemical effectors (Terry, 2001*).*

1.3.1 Biosurfactant-mediated alteration of interfacial tension

The literature on MEOR suggests that although there are a number of ways in which additional oil recovery can be achieved, some mechanisms may play more critical roles than others (Figure 1.2). The most significant of all mechanisms is that of the reduction in IFT between oil and water, which has shown great potential for enhancing oil recovery *(Biria et al, 2007)*. Biosurfactants, similar to that of chemically produced surfactants (surface-active-agents), are biological amphiphilic compounds produced by a variety of microorganisms. Biosurfactants contain hydrophilic and hydrophobic components and so can be both water-soluble and water-insoluble. They act to reduce the interfacial tension between the water and oil interface, allowing their separation as emulsions in the liquid *(Youssef et al, 2004)*.

Biosurfactant efficiency is measured by using the Critical Micelle Concentration (CMC), the concentration at which an emulsion is formed. Historically, biosurfactant use had been centred on the degradation of hydrocarbons. However, with the recent discovery of several biosurfactants that can substantially lower IFT (Maier & Soberon-Chavez, 2000), the focus has been switched to interfacial force reduction as the preferred method of oil recovery (Banat *et al,* 2010*;* Simpson *et al,* 2011*)*. It has been suggested that interfacial tension must be reduced to between 0.001 and 0.01 mNm^{-1} to achieve optimum recovery *(*Lake*,* 1989*;* Brown*,* 1986*)*.

Table 1.1 Applications of microbial products in an oil well and the bacterial strains that produce them (Based on Rashedi et al., 2012)

Microbial Product	Application in MEOR	Microbial Genera Driving MEOR
Biosurfactant e.g. Rhamnolipid, cyclic lipopeptide *Surfactin*	Emulsion formation and reduction of interfacial tension	*Bacillus, Pseudomonas*
Biomass/Biofilm Production	Plugging of well pores by accumulation and wettability alteration	*Xanthomonas, Pseudomonas, Bacillus*
Polymers e.g. Xantham gum, Levan	Viscosity reduction to improve sweep efficiency. Also used as plugging agent	*Bacillus, Leuconostoc, Corynebacterium, Brevibacterium, Xanthomonas*
Solvents e.g. Acetone, butanol, Ethanol	Viscosity reduction and removal of long chain hydrocarbons. Also potential use for IFT reduction and rock dissolution	*Clostridium, Zymomonas, Klebsiella*
Acids e.g. Acetic acid, lactic acid, butyric acid	Formation rock dissolution	*Clostridium, Enterobacter,*
Gases e.g. Carbon dioxide, hydrogen, methane	Re-pressurisation of well and reduction of oil viscosity	*Clostriduium, Enterobacter, Methanobacterium*

The most common biosurfactants, which are used in MEOR, are low molecular weight glycolipids (rhamnolipids), lipopeptides and trehalose lipids *(*Wang *et al,* 2007*)*. These lipids, produced by a range of bacteria (primarily *Pseudomonas* and *Bacillus spp.* strains), have been proven to be more effective than synthetic surfactants, increasing oil recovery in some cases by up to 75% *(*Herman *et al,* 1997*;* Maugdalya *et al* 2004*;* Youssef *et al,* 2007*).* Biosurfactants increase oil recovery by both IFT reduction, but also the alteration of rock surfaces to a more water wet state. By covering the surface of the rock in biosurfactant, ion bonding between surfactants and oil can alter the wettability, releasing more oil from the surfaces and improving the capillary number (ratio of viscous force to surface tension force). The combination of these two mechanisms will enhance oil recovery substantially. Rhamnolipid producing *P. aeruginosa* and lipopeptide producing *B. subtilis* strains have been shown to have the greatest beneficial effect on MEOR, improving residual oil recovery by 50% and 25%, respectively (Amani *et al,* 2010*;* Fang *et al,* 2007*).* Furthermore, certain strains of *Bacillus,* principally *P. mojavensis* JF-2, have been shown to reduce IFT to less than 0.016 mNm^{-1}, recovering up to 40% additional oil from sandstone cores (McInerney *et al,* 2004). The stimulation of *in situ* bacteria has shown vastly improved recovery outcomes

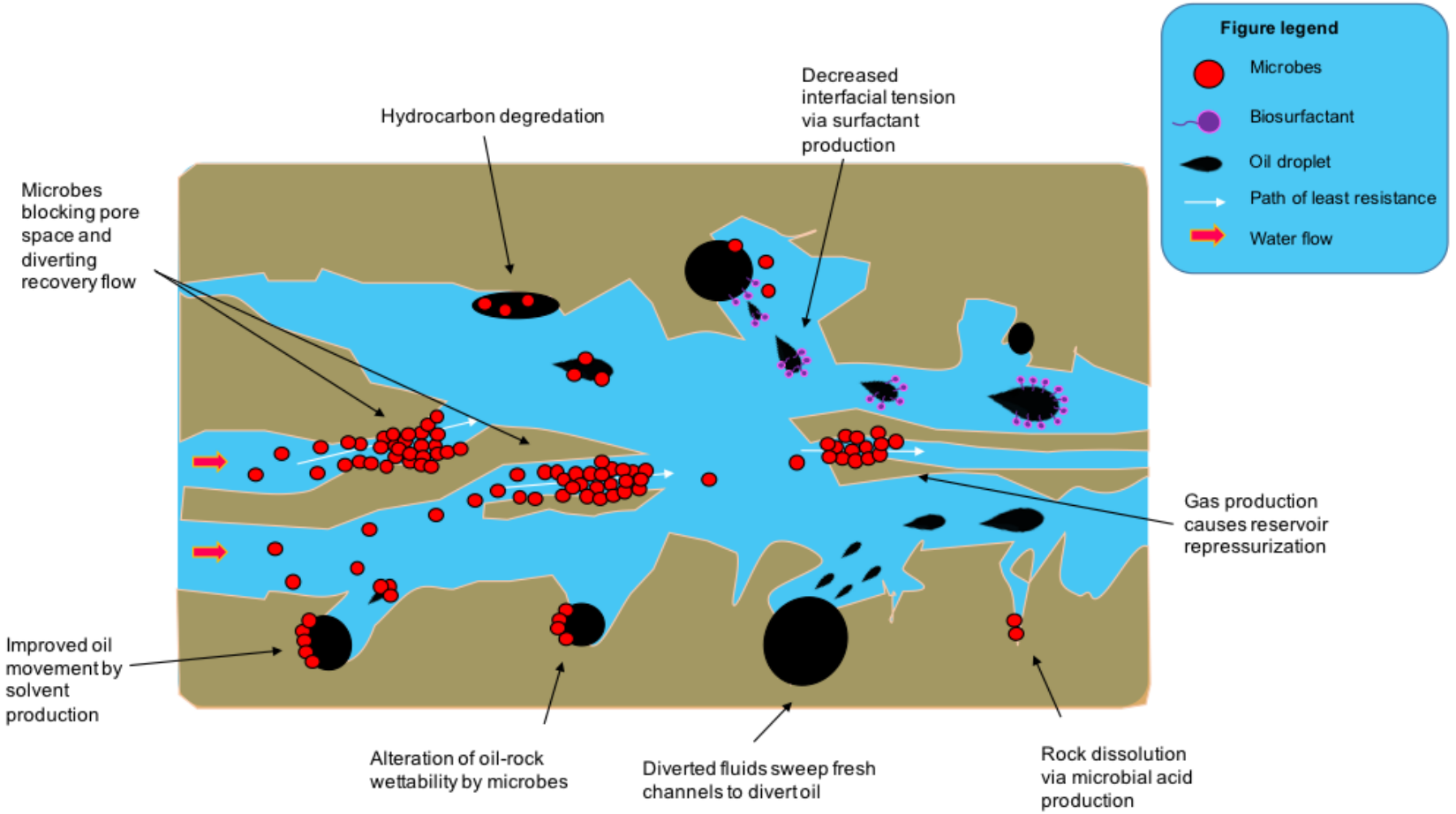


Figure 1.2 Hypothesized mechanisms of action of MEOR. Bacterial suspension acts upon the oil via surfactant production changing surface tension, blocking of rock pore by biofilm & altering of the rock permeability

when compared to bacterial injection, demonstrating biosurfactant production as one of the most feasible mechanism to increase oil recovery *(*Youssef *et al,* 2012*;* Gudina *et al,* 2012*).* This may be due to the fact that cells are also thought to play a role in IFT reduction (Kowalewski *et al.*, 2006).

1.3.2 Biomass, Biofilm and Biopolymer selective plugging

Another pivotal mechanism to successful MEOR is selective plugging by production of cellular biomass. During secondary recovery, injected water will travel through the reservoir via the path of least resistance, usually areas of high permeability. This biomass is responsible for the selective plugging and alters this flow by blocking highly permeable rock zones and the reservoir pore throat, the narrow passageways between rock/sand particles. As such, water will travel via differential routes, usually areas of the reservoir that still have large amounts of entrapped oil. This increases the sweep efficiency of the remaining crude oil *(*Suthar*,* 2009*)*.

Additionally, some bacteria are capable of producing a biofilm using a protective substance to aid with selective well plugging. This protective substance, known as an exopolysaccharide (EPS), acts to adhere the cells to a surface and surround them in a polymeric substance for protection (O'toole *et al,* 2000). This biofilm then acts to block the pores in the rock even further, thus allowing the water to move freely into the less permeable regions *(Vu et al, 2009)*. *Pseudomonas, Bacillus* and *Xanthamonas spp*. are the predominant species of bacteria that are known to aid in MEOR by biofilm production or biomass accumulation *(*McInerney & Sublette, 2002*)*. The EPS fixes the

microorganisms to the surface of the rock enabling the bacteria to form a resilient film. EPS is also more impervious to water than the microbial biomass and so has a greater plugging effect.

Some bacteria are known for their production of insoluble biopolymers, such as Xantham gum (produced by *Xanthamonas spp.*). These polymers beneficially impact MEOR in a manner similar to that of biomass/biofilms, causing selective plugging of the pore spaces to improve the flow of water to regions of untouched oil. Biopolymer production is used in combination with water flooding, where water is pushed up through injection wells, by-passing more permeable, oil-depleted areas of the well, in a similar mechanism to plugging by biomass and biofilm production. Research from Chinese oil fields have shown that *Corynebacterium, Brevibacterium, Leuconostoc, Pseudomonas* and some *Xanthamonas* strains are responsible for the production of biopolymers *(Wang et al, 1991;Nagase et al, 2002).* Biopolymers can also increase the viscosity of the water, which acts to enhance its sweep efficiency *(Akit-Ramsey et al, 1989)*.

1.3.3 Solvents

It has been suggested that the production of solvents, such as butanol or acetone, could be a useful mechanism for MEOR. Solvents, produced *in situ*, have been shown to dissolve in oil causing a reduction in oil viscosity, improving oil mobility *(*Youssef, Duncan & McInerney, 2005) and removing long chained hydrocarbons from the space between two grains in the rock, or pore throat. However, as discussed in *Grey et al, 2008*, such a significant

increase in viscosity would be needed (> 50%) for a minimal gain in oil recovery, meaning tremendous amounts of solvent would need to be produced for a nominal 6% increase in oil recovery. Recently, solvents have been shown to dissolve some forms of rock, which could increase rock permeability and porosity and has also been used as a co-surfactant to decrease IFT (*Bordoloi & Konwar, 2008)*. However, the insufficient data available regarding the use and benefits of solvent-based MEOR leaves much doubt to whether solvent use is actually an effective treatment strategy.

1.3.4 Acids

During the process of anaerobic sugar fermentation, some bacteria are capable of forming acids, *e.g.* acetic, butyric and lactic acids (McInerney *et al,* 2005*)*. Acid production can cause the dissolution of carbonates within the formation rock, which enhances rock porosity and permeability. Adkins *et al*, (1992) demonstrated that significant amounts of oil could be recovered following the dissolution of carbonate rock by an unknown halophilic, acid-producing bacterium. Carbonate dissolution can be quantified by increases in calcium ions (Ca^{2+}), as calcium is released from $CaCO_3$ present in the rock as it is dissolved.

1.3.5 Gases

Bacterial production of gases such as CH_4 and CO_2 from carbohydrate fermentation (usually molasses in MEOR *in vitro* models) provides two main advantageous mechanisms by which oil recovery can be improved. Primarily, the gas production can cause the re-pressurization of the well. The build-up in

pressure is caused by gas blockage of pore spaces, with the resulting pressure mobilising the oil *(*Bryant & Douglas, 1988). Biogas can also reduce oil viscosity by dissolving in the oil. As with some of the alternate MEOR mechanisms such as acid and solvent production, little data in the literature is available regarding the impact of biogas production in MEOR on increasing oil recovery. However, it is apparent that strains of *Clostridium, Enterobacter, Desulfovibrio spp.* and some methanogens are responsible for the production of biogas *(*Behlugil & Mehmetoglu, 2002). However, biogas would only be useful as an oil recovering agent in under-saturated (or dissolved gas) reservoirs, where pressure maintenance is vital for oil production.

1.3.6 Hydrocarbon Degradation

Until recently, hydrocarbon degradation has been suggested to be one of the more promising methods of MEOR. Microbes have been shown to be able to degrade heavier oil fractions by using them as a metabolic carbon source *(*Wentzel *et al,* 2007*)*. This breakdown causes the oil to become less viscous, more mobile and lighter, which also means the oil becomes more valuable. A range of bacteria have been found that are able to degrade hydrocarbons. Both *P. aeuroginosa* and *P. fluorescens* can degrade large carbon chained n-alkanes *(*Hasanuzzaman *et al,* 2007*)*, with some strains of *Bacillus spp*. able to breakdown shorter chained alkanes *(*Wang *et al,* 2006). Furthermore, Gudina *et al,* (2012) showed for the first time that some indigenous strains of *Bacillus spp.* had the ability to degrade large oil fractions under anaerobic conditions, suggesting they are well adapted to oil reservoir conditions. Looking at the current trend in research, it appears that hydrocarbon

degradation, alongside biosurfactant production, may be a route to successfully increasing oil recovery.

1.4 Characteristics of oil reservoirs for MEOR suitability

Formation of fossil fuels, in particular oils, is a process that takes thousands of years from start to finish. To convert organic materials to petroleum, an immense amount of heat and compression must be applied to the sedimentary layers containing these materials, which are primarily buried organisms. These organisms then undergo the chemical processes known as diagenesis (sediment alteration) and catagenesis(chemical bond breaking of organics). These reactions are primarily driven by microorganisms and transform simple organic matter into complex hydrocarbon compounds (Libes, 2009). Due to the conditions needed, petroleum formation usually occurs at extreme depths. This can be anywhere from < 500 m to > 5 km below the surface of the Earth, dependent on the nature of the reservoir (Ehrenberg & Nadeau, 2005). These depths are the primary features that determine the high temperatures and high pressures in oil reservoirs, alongside the differing rock geologies and biochemistries within a formation, and can drive differences in microbial community structure (Shelton *et al*., 2016). Therefore, the petrophysical parameters of a reservoir are a major consideration when assessing an oilfield for potential MEOR treatments. These petrophysical parameters could distinctly affect the growth of any microorganisms used in the process.

1.4.1 Temperature

Temperature has been previously shown to be the greatest regulating factor altering both microbial survival and community composition, with bacteria categorized into three groupings: psychrophilic (< 25°C), mesophilic (25-40°C), thermophilic (> 40°C). Most reservoirs are characterized by temperatures that fall between the high end of mesophilic and low end of thermophilic; however this can vary, even within a reservoir itself. Optimum microorganism survival is usually between 15-65°C, but for metabolic activity, a range of 30-45°C is more favourable (Hallenbeck, 2012), particularly for the production of metabolites, including biosurfactant (Souza *et al.*, 2017; Zhao *et al*., 2015).

1.4.2 Pressure

As reservoir depth increases, not only does temperature increase, but the pressure that the reservoir is under also increases. Therefore, pressure is hypothesized to be another critical factor in determining the suitability of microbial strains to a reservoir for MEOR. Yet until recently, the evidence available regarding high pressure environments on microorganisms was inadequate. Schedler *et al*, (2014) demonstrated that high pressures (> 15 MPa) slowed the metabolism and hydrocarbon degradation potential of two indigenous reservoir bacterial strains. This shows that higher pressures may have a role in reducing microbial metabolic activity and must be accounted for when evaluating microbial treatment strategies.

1.4.3 Brine salinity and pH

Salinity is also a parameter that varies significantly across oilfields and production waters, globally. It is estimated that formation or production water brines can range in salinity from 100 mg L^{-1} to over 200 g L^{-1}. Furthermore, extremely high salinities (> 200000 ppm) have been shown to have a detrimental impact on interfacial tension and viscosity reduction (Nmegbu, 2014). pH is also a factor that may affect the growth of microorganisms in a reservoir. Optimal pH ranges for the survival of microorganisms is found to be between 5.0 and 8.0, ranges that are commonly found in most oil reservoirs. However, pH in a reservoir may be reduced *in situ*, compared to pH measured from a sample at atmospheric pressures. The lower the pH, the greater the impact will be on the production of metabolites, which could have a negative effect in the promotion of MEOR. Furthermore, at pH < 6.0, the activity of some metabolic products, *e.g.* surfactants, are also greatly reduced (Elazzazy, Abdelmoneim, & Almaghrabi, 2015). This is due to acid denaturation, precipitation of the surfactants and ionic demulsification, which can render these metabolites ineffective (Daaou & Bendedouch, 2012).

1.4.4 Nutrients

In addition to physiological factors, nutrient types that are added to a reservoir, for both bio-stimulative and bio-augmentative MEOR, are of great importance to the MEOR process. Due to diagenesis, reservoirs are usually under a reduced redox potential (Siegert *et al*, 2013). This indicates that most electron acceptors are depleted, particularly oxygen and nitrate, resulting in an anaerobic environment. There are however electron donors available, in the

form of hydrogen and volatile fatty acids (Noha Youssef, Elshahed, & McInerney, 2009). There is also the potential of the microbial biodegradation of oil to utilise the oil itself as a carbon source for growth. However this is considered difficult to achieve under strict anaerobic conditions. Apart from any added nutrients from the formation water that may be passed through the reservoir, the oil itself is the only nutrient available to the endogenous microorganisms. Therefore, it is vital to correctly chose a nutrient system to inject during MEOR. A lack of accessible carbon is usually overcome by adding a fermentable substrate, usually a sugar. In previous laboratory and field trials sugar cane molasses has been the frequent nutrient of choice. This is primarily due to the fact it is substantially cheaper than other purified carbohydrates. Molasses also contains fundamental vitamins and minerals that promote microbial growth (Maudgalya, Knapp, & McInerney, 2007). In addition to molasses, other VFA's or nitrogen sources are also usually added to promote growth of numerous bacterial species. However, whatever approach is taken, it is imperative to consider the microbial species and bioproducts that are desired, as their production may be dependent on the substrates present.

1.4.5 Microbial communities associated with oilfield environments

The microbiome of an oil reservoir is complex in composition. For many years deep oil reservoirs were considered to be harsh environments that were unfavourable for microbial survival (Augustinovic *et al*., 2012). However, following on from the preliminary discovery of sulphate reducing bacteria in American subsurface oilfield waters, we have now accepted that these

ecosystems harbour a vast variety of environmentally adapted anaerobic and facultative microorganisms (Bastin *et al..,*1926; Nazina *et al*., 2006). Numerous strains of sulphate reducing bacteria, methanogens, iron reducers and fermentative bacteria have been identified from oilfield microbial communities (Magot *et al.*, 2000; Youssef *et al*., 2009). The majority of identified oilfield isolates are obligate anaerobes, with very few aerobes present. Although reservoirs can be very different both petrophysically or geographically, the populations and microorganisms are often similar in terms of identity, with similar species colonizing numerous reservoirs (Lewin *et al*., 2014). Though a number of petrophysical parameters such as temperature, pressure, salinity and porosity affect the exact abundances of each species present, in higher temperature reservoirs, *Firmicutes* and thermophilic bacteria/archaea usually predominate. Conversely at lower temperatures, *Proteobacteria* and methanogenic bacteria are usually the predominant phyla (L. Y. Wang *et al*., 2012)

Considering these unsympathetic conditions that encapsulate the reservoir, it is critical to accurately identify microbial populations in order to make an assessment of what microorganisms are suitable candidates for MEOR. To come to this conclusion, we first must evaluate the previously discussed physiological parameters of the reservoir. In combination with this petrophysical and geochemical evaluation, we can also delve deeper into the prokaryotic communities using a variety of culture dependent and independent identification methods.

1.5 Microbial sampling and monitoring of oil reservoirs

The sampling of the oilfield environments is crucial to the study of microbial communities. In order to achieve accurate representation, sampling strategies must be carefully considered. The easiest way to sample directly from a reservoir is to collect formation water post production, directly from the production pipelines. This can be detrimental to the accurate microbial community assessment for two main reasons. Firstly, as production has commenced previously, organisms associated with the pipeline itself, not necessarily those making up the microbiota of the reservoir will contaminate the sampling. Additionally microorganisms may be introduced through human contact, drilling mud or pipeline. Secondly, during production, the flooding of reservoirs with either formation or recycled production waters can have dynamic impacts on the microbiota of the sample. By passing either collected sea water or produced water through the reservoir, it is possible to influence the microbial community composition from the produced water, increasing the relative abundancies of aerobic bacteria than would normally be expected (Gao *et al*., 2015; Zhang *et al*., 2012). Although aseptic techniques are often attempted to reduce contamination, the use of aseptic techniques is more difficult in practice in a producing reservoir.

One way to sidestep this issue is by the direct subsurface sampling of rock core. To sample the subsurface, rock core material must be drilled down to the required depth and lifted to the surface (Kieft, 2014). Although this process is cost inefficient due to the machinery needed to drill and recover core fragments, it is one method of marginally reducing community contamination.

However, careful core depressurization and aseptic handling must be applied in order to stop the introduction of contaminants or cell lysis during post collection sample handling (Kotlar *et al.*, 2011; Tsesmetzis *et al.*, 2016). Other methods of subsurface sampling include the assessment of drilling muds and the use of high pressure single phase sampling bottles. These methods however have also been shown to introduce large abundancies of surface related organisms or are cost ineffective and labour intensive so tend to be less frequently used (Krüger *et al*., 2016; Spark *et al*., 2000).

Once samples have been collected, the next step is to quantify microbial numbers and assess phylogenetic groupings. Until the turn of the millennia, cultivation of microorganisms via enrichment of liquid/solid phase biosystems or media based growth were the only ways of analysing bacteria from environmental samples. These enrichment methods however selected for types of bacteria that were capable of growing in the growth media and conditions tested. Culturing introduces a significant amount of bias, as the specific conditions used *in vitro* will not support the growth of all microorganisms, and may actually be quite selective for only a few species. As we know only a fraction of bacterial diversity is cultivable under laboratory conditions, a more robust method was needed to analyse microbial community structure, not just that of oilfields (Amann *et al.*, 1995). It was at this point that the microbiological community turned to a novel DNA sequencing approach, that of ‘metagenomics’, a study that would revolutionize microbiology as we know it (Handelsman *et al.*, 1998).

1.5.1 Cultivating oilfield microorganisms

The cultivation of individual strains from oilfields can be a challenging procedure, primarily because only a minority (< 1%) of bacteria can be cultivated (Amann *et al.*, 1995; Saxena, 2015). This is substantially reduced in extreme habitats (e.g. increased pressure, high temperatures, high salinities), due to the difficulty in mimicking environmental conditions or distinct symbiotic relationships (Kennedy *et al.*, 2010; Stewart, 2012). As a result, much of the environmental microbial population analyses have been conducted using culture independent methods, for instance 16S rRNA gene sequence analysis, denaturing gradient gel electrophoresis (DGGE), temperature gradient gel electrophoresis (TGGE) and population metagenomics. Yet these practices only enable the identification and analysis of genetic characteristics, rather than the phenotypic traits that the strains may exhibit. Therefore, for the practice of MEOR to be proved effective in the laboratory or field, MEOR related microorganisms must be successfully isolated from the field.

It has previously been reported in a number of review articles that microorganisms of the phyla *Firmicutes, Thermotogae, Deferribacteres* and *Proteobacteria* (harbouring a vast majority of the sulphate reducers) are the most commonly isolated strains from oilfield environments (Frank *et al.*, 2016; Magot, 2005; Song *et al.*, 2017; Wentzel *et al.*, 2013). Furthermore, some of these bacterial taxa, such as *Petrotoga* of the phylum *Thermotogae,* have only been isolated from oil reservoirs, showing their adaptive nature to the diverse and sometimes hostile conditions (Magot, 2005).

1.5.2 Molecular techniques for microbial monitoring

With the ongoing advancement of molecular technologies, it is becoming increasingly important and easier to study microbial communities in relation to abundance of organisms, their relative abundamce and bacterial community changes that happen in response to environmental perturbations, without the need for cultivation. The use of quantitative DNA amplification techniques and high throughput sequencing methods now allow for the detection of the most abundant microbes, but also other less prevalent microbes. Some of these microbes are taken from extreme or low biomass environments, where they are not easily cultivable or detectable using culture-dependent methods. Furthermore, Bowman *et al*., (2012) and Urbieta *et al.* (2015) have shown that it is possible to sequence nucleic acids from diverse and extreme environments, with their respective investigations into microbial diversity in glacial and acidic ecosystems using 16S amplicon sequencing. Limited data was previously available on unconventional oil reservoir diversity, with published data only emerging over the last few years (Cluff *et al.*, 2014; Fahrenfeld *et al*., 2017). But with the implementation and reduced cost of next generation sequencing, it is only a matter of time before genomic studies become the primary tool for the study of microbial oilfield populations.

1.5.2.1 Polymerase Chain Reaction

PCR is commonly used throughout all branches of life sciences. Developed by Kary Mullis in 1983 (an invention for which he would later receive the Nobel Prize), PCR has become a dominant technology in improving microbial diagnostics in both healthcare and environmental settings (Mullis, 1990). It is

now possible to detect the sequences of specific types of microorganisms from a mixed community or within non-relevant nucleic acids, without having to independently isolate the organism of interest, using the universal 16S rRNA gene present in all bacterial species. Furthermore, there is an inherent difficulty in extracting large quantities of cellular biomass and nucleic acids from environmental samples, due to the lower biomass content found in environmental samples compared to pure microbial cultures (Amann *et al*., 1995). This is a particular problem in oilfield systems. PCR provides the ability to 'amplify' nucleotides. This means we can create multiple copies of DNA from very minute starting quantities by replicating the DNA under differing cycles of heating.

To identify specific sequences, it is paramount to design oligonucleotide primers that match the sequence of interest. These short oligonucleotides (usually < 30bp) are designed to recognise matching nucleotides at two ends of a DNA strand and allow replication to commence. Therefore, theoretically only sequences matching these primers will bind to the DNA and replicate. Many distinct primers are used in microbial monitoring, particularly in oilfield analysis. These primers relate to specific genes of bacteria that have a function in the oilfield, such as sulphate reduction (*DSR* gene) and methanogenesis (*mcrA*) that are responsible for energy conservation and microbial metabolism (Smith, Castro, & Ogram, 2007; Wagner *et al.,* 1998). Yet for microbial community analysis, it is not these specific metabolic genes that are of most importance. Instead we look to the highly conserved regions

in the 16S ribosomal gene for characterization of microbial and archaeal taxonomy.

1.5.2.2 Real time PCR

Although standard PCR is seen as an essential tool, it actually falls short for accurate microbial monitoring. PCR is unable to successfully quantify bacterial numbers based on starting material due to its amplification processes. Furthermore, PCR requires gel visualization to analyse the end product of the reaction. It is through this limitation that the advent of real time PCR or quantitative PCR (qPCR) has emerged. qPCR utilizes a fluorescent based reporter (either dye or probe) that increases proportionally with the amount of amplified product at the end of each cycle. Once compared to known premade standards, the quantities of total DNA can be calculated.

1.5.2.3 The 16S rRNA gene

The 16S ribosomal RNA gene has played a prominent role in recent times for the study of microbial diversity in relation to prokaryotic taxonomy and phylogeny. The structure of the ribosome consists of a 70S unit, which in turn comprises a 50S and 30S subunit. It is within this 30S subunit that the 16S rRNA (~1600bp) is located, alongside 21 ribosomal proteins (Wilson & Doudna Cate, 2012). The 50S subunit is composed of 5S and 23S. It is within the ribosome that translation of mRNA takes place and is integral to protein synthesis. The 16S rRNA gene itself comprises nine hypervariable regions (V1-V9), all of which display variant sequence differences between species types (Van De Peer *et al.*, 1996). In between each of these hypervariable regions are ‘conserved’ areas. This gene target is of great use in microbiology

due to its presence in almost all bacteria and suitable size for bioinformatic analysis. Yet most importantly, the conserved regions of the 16S gene are well preserved both in an evolutionary and temporal sense, whilst the hypervariable regions show species specific diversity (Janda & Abbott, 2007). Therefore, high similarity is seen between prokaryotic sequences that are of related subgroups, from phyla to species. This makes it an accurate tool for identifying novel species of bacteria, classifying species into absolute groups and assessing evolutionary distance compared to lineage. Furthermore, it is the highly conserved regions that make the 16S gene a suitable target for primer design for PCR.

1.5.3 PCR related bias

Despite the benefits of PCR and its wide spread use, there are however some common limitations to the technique. These primarily centre around the introduction of inherent bias within the primer specificity and template amplification, which result in the inaccurate amplification of individual templates (McDonald *et al*., 2016). This has particularly been seen to be problematic in all forms of PCR, yet its effect more carefully scrutinized for the use of primers specific to SSU rRNA (Diallo *et al.*, 2008; Sipos *et al*., 2007). Primarily, the majority of introduced bias arises from preferential amplification of specific sequences (Polz & Cavanaugh, 1998). Additionally, a further bias ascends from sequence annotation and database curation. PCR primers are designed based on sequences identified previously and assembled in specific gene databases. Therefore, PCR is only as specific as the genomic database that the sequence information is taken from. Therefore, novel or previously

unidentified microorganisms may not be identified due to dissimilarities in catalogued sequences. Additionally, the introduction of primer degeneracy to overcome selective amplification, does not create oligonucleotides degenerate enough to capture all diversity (Brooks *et al*., 2015). These 'universal' primers also result in differing primer binding energies and thus less specific template binding. For these reasons, 'universal' primers have been suggested to miss almost half of the diversity in SSU rRNA sequenced samples (Hong *et al.*, 2009). Finally, the use of nucleotide extracts from certain sample types, such as heavy crude oil samples, can reduce PCR efficiency due to the presence of PCR inhibitors such as heavy metals and polysaccharides (Schrader *et al.*, 2012).

To overcome these issues, it has been suggested that reconstruction of 16S rRNA from complete metagenome sequencing may be a suitable methodology (Venter *et al*., 2004). Although this approach has been shown to remove some bias associated with primer design and PCR, such as removal of chimeras, the fact that for environments of a large size, datasets would be vast and difficult and time consuming to fully analyse (Fan *et al.*, 2012). Therefore the process of PCR must either be improved to overcome these issues or novel approaches developed to bypass PCR entirely. This would lead to improved identification of complete diversity and accurate genomic monitoring.

1.7 Advantages and disadvantages of MEOR as a treatment strategy

Microbial Enhanced Oil Recovery has been seen as a beneficial method of increasing oil recovery because of its numerous advantages when compared to more conventional methods of oil extraction. Although some tertiary methods, such as chemical and thermal oil recovery, have improved the amount of residual oil extracted from reservoirs, their detrimental environmental effects and astronomical economical outlay has made them unsustainable and impractical for continued EOR use. MEOR is economically viable due to the cost of setting up the process. Nutrients, bacteria and injection fluids are relatively inexpensive compared to alternate tertiary methods, as there is no need for large quantities of chemicals. Furthermore, the cost of adapting existing reservoirs to accommodate MEOR processes is cost effective and does not require extensive modification. Due to the exponential replicative potential of bacteria, there is a continuous production once down the well, with their effect not decreasing after a prolonged period of use, as would be the case with other EOR methods *(*Lazar *et el, 2007).* The majority of recent MEOR processes use *in situ* produced (indigenous) microbes. This utilizes the pre-existing microbes that are already down the well, removing the cost of engineering, producing and applying bacteria to a well that are produced externally. At present, it is estimated that 81% of MEOR trials have yielded a positive increase in oil recovery, with no studies reporting a decline in production.

In terms of environmental advantages, MEOR uses bacteria and natural and biodegradable products. This removes the painstaking task of disposing

dangerous and toxic chemicals from wells, which could later cause a problem with accumulation in the well or surrounding ecosystem. Additionally, MEOR does not use energy intensive processes, as is the case with techniques such as Steam Assisted Gravity Drainage, that use superheated steam to mobilize trapped oil. Processes like these use enormous amounts of fresh water and have a need for an incredibly high-energy input. In contrast, MEOR has the potential to significantly lower the carbon footprint.

Despite the success of MEOR, some negatives have been found. Although the MEOR theory has been around for a long time, until recently, little was actually known regarding its mechanisms. A field that is poorly understood makes it even harder to apply its principles. A major difficulty, as with any process, is targeting the bacteria or nutrients to a specific zone of the well. It has been very difficult to find and isolate microbes that are specific to certain well conditions (pH, salinity, pressure), especially with variation being so diverse amongst wells on the same oil field (Sen, 2008a). This is why indigenous bacteria appear to be the way forward in MEOR. However, even with this method, complete optimization of down well bacteria is very difficult to achieve. As conditions are so varied, even a slight difference in one factor between the reservoir and laboratory conditions could affect the growth and abundance of certain microbes, making it a potentially unpredictable process *(Mokhtab, 2006)*. When it comes to scaling up to field trials, many negative results are seen here due to this factor, making it very hard for MEOR to reach its full potential. Finally, reservoir plugging of rock in unwanted locations, bacterial disruption of crude oil composition and increased hydrogen sulfide

provide other major issues that need addressing *(Almeida et al, 2004; Wolicka & Borkowski, 2012)*.

The main downfall to MEOR is the quality of literature available. The majority of data is based on laboratory-based trials, rather than field-testing. It is therefore very difficult to assess the effect that could be had with field trials, particularly considering the heterogeneity of oil wells. The real value can only be assessed once the technique has been successfully applied to a wide range of oil fields. This is the predominant reason why so few patents have been accepted on the area to date, and why MEOR has not been fully accepted amongst the Oil Industry as a whole.

1.8 MEOR case studies

Although there is considerable data in relation to MEOR trials, the majority of this is only in the form of in-vitro or bioreactor experiments (Brown, 2010). This is suprising considering it has been a technique that was first hypothesized in the early 20th century. However, the transference from laboratory to field scale is a difficult and expensive process. If any parameters in field trials are not anticipated or selected carefully, a catastrophic and economically costly failure could ensue. Therefore, it is beneficial to carefully modify the experimental parameters in the laboratory before taking it to the field. Of the numerous laboratory trials, the largest increases in residual oil recovery are around 20%. This is representative of predominantly lighter oils, with again minimal literature available on the effect of the strategy on unconventional oils. However, the only way to successfully asses the effect of MEOR as a recovery

technique is with its trial in the field. Despite this, some field trials have taken place over the years.

The majority of microbial field trials undertaken before the 1990's were primarily carried out on gas related microbial production (Awan, Tiegland, & Kleppe, 2008) But post 90's, of the oil based field data we do have, there has been a balanced mixture of success and failure. The majority of these have been evaluated by Lazar, Petrisor, & Yen (2007) and Maudgalya *et al*. (2007), in which a global review of the successes and failures of MEOR has been performed. In the majority of field trials discussed here, a recovery of approximately 10-35% was achieved. However, approximately two thirds of described trials did show MEOR failure, with the treatment producing no more than 2%, a level which is recorded as uneconomical for production (Brown *et al.*, 2000). Previously for heavy oils, the only trials available were preliminary studies or pilot tests, which showed only some success (Davidson, 1988). Nonetheless, as of 2015, an efficacious field trial in China looking at heavy oil recoveries based on indigenous stimulation, showed a 100% increase in total recovery of a low-producing reservoir (Chai *et al.*, 2015). This trial, although one of very few, goes a long way to showing the feasibility of MEOR may be satisfactory. But to prove this, far more research and field trial data is needed. Yet with the recent focus shifting towards the production of heavier oils to stop the prophesized oil shortage, we expect more field trials to be undertaken in the coming years.

1.9 Future direction of MEOR

Due to the lack of effective field related data, it is apparent that more research is needed in the area of microbial enhanced oil recovery (*Xu & Lu, 2011*). However, from what we have discovered, the principal way forward appears to be in the use of indigenous bacteria that act to alter reservoir properties, rather than the insertion of generic microorganisms' downhole, as previously used in some MEOR trials. The fact that only certain bacteria can withstand the potentially harsh environmental characteristics of the reservoirs, exhibits the importance of finding a species that still has full metabolic capabilities under these conditions. It may be beneficial to assess how specific reservoirs affect microbial populations depending on changes in features such as pH, pressure, salinity and temperature, all factors which cause great diversity between oil wells. If a link can be made between relative quantity of a microbe, its mechanisms of action and a reservoir's physical/geochemical properties, it may make it easier to influence the effectiveness of microbes present in the reservoir.

1.9.1 Feasibility of MEOR on unconventional oils

A large proportion of the work being undertaken is with lighter oils, with an API of more than 22°. Oils this heavy, are usually easier to mobilize and extract, but with light oil resources rapidly diminishing and heavier more unconventional oils making up approximately 65% of total reserves worldwide *(*Dianilovic *et al, 2012)*, it is now important to exploit MEOR for heavy oil recovery. Previous studies have suggested that it is not possible to use MEOR for oils heavier than 15° or 500 CP (pressure coefficient), but if oil production

is to continue to meet demand, this is an area of research that must be considered further.

A continuous challenge facing the industry with regards to MEOR is the lack of data regarding genes involved in specific processes. It has therefore been suggested that genomic analysis of the relative microbial abundance and

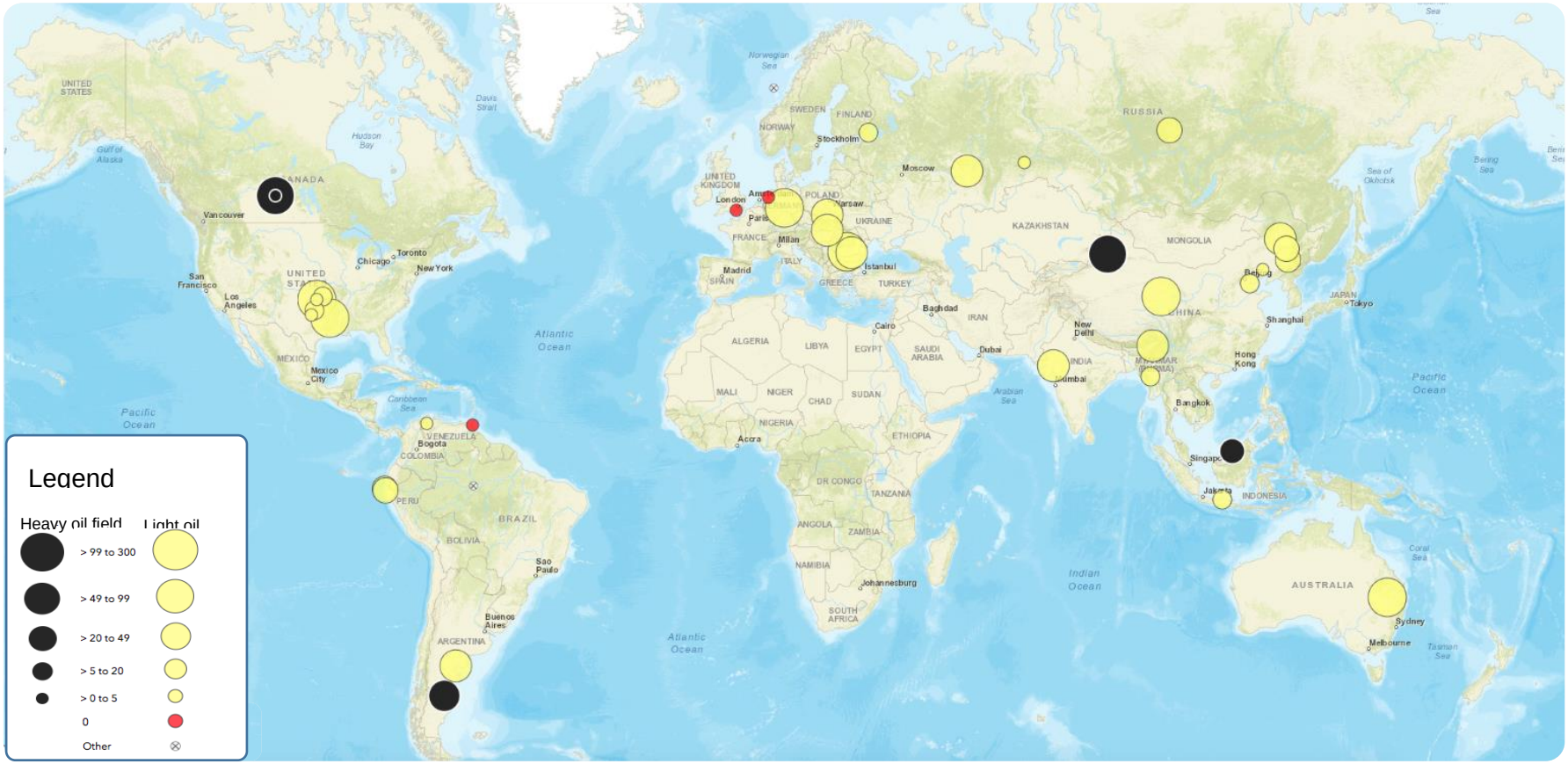


Figure 1.3 Map of worldwide MEOR field trials. Map depicts all MEOR field trials found in the literature that provided an AOR% figure for total oilfield recoveries. MEOR trials shown as depicted in legend. Black circles represent heavy oil fields. Yellow circles represent light oil. Red circles represent failed field trials. Size correlates to the % increase in AOR. Complete list of all recorded field trials given in Appendix B.

diversity of genes involved in certain MEOR processes could lead to the understanding of how and why MEOR works and how genes are differentially regulated once the MEOR processes commence. Similar studies have taken place in other diverse ecosystems (Venter *et al*, 2004). This would remove the need for individual culturing of microbes present in the reservoir to assess their metabolic activities.

A genomic database of microorganisms found in oil reservoirs, comparing differing petrophysical parameters, including mechanisms of action at a gene level would be beneficial to the research of MEOR technologies. Recently, the Research Institute of the Petroleum Industry (RIPI) and Fundação para a Ciência e a Tecnologia (BioSurfDB) have announced that they have independently initiated the collection, preservation and cataloguing of strains and genomes of microbes that play a role in oil biotechnology (Oliveira *et al.*, 2015). Although this is more in-depth than just MEOR related processes, the aim is to collect microbes present in many different oil samples and compare their presence and functionality (both positive and negative) to assess their roles in the oil industry. This is the beginning of a widespread microbial identification scheme that has great potential for improving the knowledge on microbial variation and uses in oil reservoirs.

1.10 Project aims and objectives

The initial aim is to test the feasibility of oil recovery on unconventional or heavy resources below 15º API. At the outset of the project, very little laboratory and field data was currently available in the literature regarding the effects of MEOR

on unconventional oilfields. This project will examine the effects that microbes can have on enhancing oil recovery and their mechanisms of action, at simulated oilfield conditions. It will also assess which specific microorganisms are most active in the process of MEOR, including those microorganisms directly isolated from the chosen oil reservoir. A range of molecular, culture dependent and bioinformatic methodologies will then be utilized to isolate a number of bacteria that may be beneficial for MEOR treatment. We will then attempt to determine the nature of the process by which certain microorganisms can increase oil recovery and their mechanisms of action, with a specific focus on metabolite production. The activity and oil recovery potential will be assessed against other known MEOR microbial strains. For any active metabolites identified, purification, characterization and genomic studies will be employed, alongside the testing of metabolite activity, again at oilfield simulated conditions.

Chapter 2

General Materials and Methods

2.1 Bacterial strains and culturing conditions

The growth media elements were obtained from Sigma and Fisher Scientific, with the exception of the pure cane molasses (Holland & Barrett, UK) used for reservoir simulated growth media.

All commercial strains (Table 2.1) used in the study were grown on Luria Bertani (LB) agar plates and cultured in LB media at 37°C with shaking (220 rpm). All strains were stored at -80°C in 50% (V/V) glycerol.

Table 2.1 Commercial bacterial strains

Bacterial strain	**Source**	**Reference relating to species**
Bacillus licheniformis Bi10 (Isolate Bi10)	This study: Bentley Oilfield, UK	N/A
Bacillus mojavensis JF-2	University of Oklahoma, USA	(Folmsbee *et al.*, 2004)
Bacillus licheniformis ATCC14580 (DSM13)	DSMZ, GER	(Gibson, 1944)
Pseudomonas aeruginosa PA01	University of Liverpool, UK	(Holloway *et al.*, 1979)
Escherichia coli DH5α	University of Liverpool, UK	(Taylor *et al*, 1993)

2.2 Buffers, solutions and enzymes

Product name	Components
Chloroform isoamyl alcohol	Ratio of 24:1
CTAB	10% (w/v) CTAB in 0.7 M NaCl; 240 mM potassium phosphate buffer; pH 8
Ethidium bromide	10 mg ml^{-1}
Glycogen	20 mg ml^{-1}
Luria Bertani (LB) agar	(Per L): 1% (w/v) Tryptone,; 0.5% (w/v) Yeast extract; 0.1% (w/v) NaCl; 1.5 (w/v) ; Agar; pH 7.0 ±0.2 at 25°C
Luria Bertani (LB) broth	(Per L): 1% (w/v) Tryptone; 0.5%(w/v) Yeast extract; 0.1 (w/v) NaCl; pH 7.0 ±0.2 at 25°C
Marine Broth (MB)	Difco Marine broth 2216
Minimal Salt media (MS)	(Per L): 0.35% (w/v) Na_2HPO_4; 0.1%(w/v) KH_2PO_4, 0.05 (w/v) (NH_4) 2 SO_4; 0.01% (w/v) $MgCl_2$. 6 H_2O; 0.005% (w/v) $Ca(NO_3)$2. 4 H_2O; 0.01% (w/v) Trace Element Solution SL-4; pH 7.0 ±0.2 at 25°C
Nutrient Broth (NB)	(Per L): 0.6% (w/v) Peptone, 0.1% (w/v) Beef extract, 0.6% (w/v) yeast extract, 0.5% (w/v) NaCl, pH 7.0 ±0.2 at 25°C
PEG	30% (wt/vol) polyethylene glycol 6000; 1.6 M NaCl
Phenol:chloroform: isoamyl alcohol	Ratio of: 25:24:1; equilibrated with 10 mM Tris pH 7.8
Phosphate buffered saline (PBS)	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4
Polymerase (Taq:Phusion)	5,000 units ml-1 : 2,000 U ml-1
Oilfield Media (OFM)	(Per L): 4% (w/v) Molasses; 5.5%(w/v) $NaNO_3$; 0.05 (w/v) K_2HPO_4; 3% (w/v) NaCl; 0.4% (w/v) EDTA
RNase Inhibitor	40 U/µl
Tris acetate EDTA (TAE) buffer	40 mM Tris acetate; 1 mM EDTA pH 8.0
Tris EDTA (TE) buffer	10 mM Tris-HCl; 1 mM EDTA; pH 8.0
Turbo DNase	2 U μl^{-1}
Xcite Formation Water (XFW)	See Table 2.2

2.3 Composition of synthetic reservoir formation water

To simulate reservoir conditions in the microcosms, chemically comparable formation water was needed. Chemical analysis was performed on produced water samples from the reservoir and the recipe for artificial formation water brine was created based upon the chemical composition of the Bentley produced water (Table 2.2). Total volume of 1L made from combining solutions A, B, C, E and VFAs (100 mL) into 1L duran with stirring. Formation water underwent gas flushing, or sparging, with nitrogen, to remove dissolved gasses (e.g. oxygen) present within the water before autoclave sterilization.

2.4 Sample collection

Production water (water injected into the oil well to effect secondary production) and rock core (taken from drilled core material) were taken from the Bentley oil reservoir (Block 9/3b, 59°56′N 1°34′E). 100 ml of produced water samples were collected under anoxic conditions in the field, using a sterile collection system (Xcite, UK) with 0.25% (w/v) L-cysteine added as a reducing agent. The samples were then passed through a 0.22 µm filter kit (BiSN Laboratory Services, UK) alongside 0.1 volumes of RNA Later (Qiagen, UK). Stoppers were fitted to either end of the filter unit and shipped to the laboratory at -20 ºC. A sample of production water effluent (50-100 ml) was also collected in an anaerobic container for culturing purposes (Conducted by Excite Energy Ltd. field technicians, Bentley Oilield, UK). Core samples were ground down using a sterilized pestle & mortar and collected into sterile 50 mL centrifuge tubes and stored at -20 ºC (For further information see Section 3.2.1). All samples were stored until further use (See Fig 2.2, pg 53).

Table. 2.2 Composition of replicated artificial reservoir formation water

Formation Water*		**Concentration**	
Solution A	NaCl	38.071 g L^{-1}	879 ml L^{-1}
	$SrCl_2$	0.313 g L^{-1}	
	$CaCl_2$	3.840 g L^{-1}	
	$MgCl_2$	6.187 g L^{-1}	
	KCl	5.337 g L^{-1}	
	Na_2SO4	0.097 g L^{-1}	
	$BaCl_2$	0.358 g L^{-1}	
	EDTA	0.900 g L^{-1}	
Solution B	$CuSO_4.5H_2O$	(5.108 g L^{-1}) (3.606 g L^{-1})	1 ml L^{-1}
	$ZnSO_4.7H_2O$		
Solution C	$BaCl_2.2H_2O$	(7.186 g L^{-1})	10 ml L^{-1}
	$MnCl_2.4H_2O$	(3.171 g L^{-1})	
	$AlCl_3$	(1.110 g L^{-1})	
	LiCL	(0.549 g L^{-1})	
	$FeCl_3$	(6.680 g L^{-1})	
Solution E	NaH_2PO_4	(3.843 g L^{-1})	10 ml L^{-1}

*****NB FW made up into 1L volume less than total required due to addition of VFA's

Volatile fatty acids (VFA)+	**Concentration**	
$C_3H_5NaO_2$	0.002 g L^{-1}	100 ml L^{-1}
HCOONa	0.191 g L^{-1}	
$C_2H_3NaO_2$	0.352 g L^{-1}	

+ NB VFA's calculated for total amount in 1L added to FW

2.5 Screening of isolates for biosurfactant producing strains

In order to assess the total amount of biosurfactant produced, all isolates were cultured in LB media under anaerobic and aerobic conditions at and incubated at 40°C under static conditions. For anaerobic culturing, cultures were sparged with nitrogen to remove dissolved gas and placed into an anaerobic sealed container. Samples from each culture were measured for surfactant production by assessing oil spreading potential *(*Morikawa *et al.,* 2000*)* and haemolysis on blood agar (Mulligan *et al.*, 1984*)*. Surfactants have the capability to reduce interfacial tension. By placing a droplet of biosurfactant containing supernatant or culture onto light crude oil, a zone of clearing is formed in the oil droplet. This can be visualized to qualitatively assess the strength of the biosurfactant. Using similar features of the biosurfactant, the breakdown of red blood cells on agar, or haemolysis, also indicates surfactant strength (Carrillo *et al.*, 1996). For evaluating the biofilm potential of each strain, 1 mL cell culture was added to a glass cuvette and incubated for 7 days, before undergoing crystal violet staining (Chapter 4.2.4). Although this method does reveal total biofilm formation, it can only be used as a basic assessment as crystal violet is a stain of both live and dead cells, so cannot be used for viability. Detailed information on each assay is available in Chapter 4.2.

2.6 Growth kinetics

The rate of growth for the bacterial strains and isolates was assessed by measuring changes in both optical density (OD_{600}) of broth culture and also by viable counting. Colony counts were made on agar plates, with appropriate serial dilutions made at each time point.

2.7 Bacterial DNA Extraction

Bacterial cultures of strains or produced water samples were collected (50 mL) and centrifuged at 13000 RPM for 15 min using Sorvell Lynx 6000 centrifuge (ThermoScientific, MA, USA). DNA extraction was carried out using the method described by Griffiths *et al* (2000), with exception that 1 µg μl^{-1} glycogen was added prior to the PEG precipitation stage to increase nucleic acid precipitation efficiency. DNA quantification was performed using Qubit 3.0 Fluorometer according to the manufacturer's instructions (Life Technologies, UK). Samples were stored at -80°C.

2.8 Bacterial RNA extraction

Bacterial cultures of strains or produced water samples were collected (50 ml-1 L) and centrifuged at 13000 RPM for 15 min. Dual DNA/RNA extraction was carried out using the method described by Griffiths *et al* (2000) with exception that 1 µg μl^{-1} glycogen was added prior to the PEG precipitation stage to increase nucleic acid precipitation efficiency. To remove contaminating DNA, samples were treated with Turbo DNase (Life Technologies, UK) as described by the manufacturer. RNA quantification was performed using Qubit 3.0 Fluorometer according to the manufacturer's instructions (Life Technologies, UK). DNA quantification was also assessed to rule out contamination. Samples were stored at -20°C with 40 U of RNasin (Promega, UK) added.

2.9 Polymerase Chain Reaction (PCR)

A polymerase chain reaction (PCR) was performed using either Taq polymerase (NEB, USA) or in cases where PCR products would be sequenced, a high fidelity, proof reading enzyme Phusion polymerase (NEB, USA) was used following the manufacturer's instructions (Table 2.3). A total of 1-2 µl (< 250 ng) of template DNA (in either 20 or 50µl reaction mix, respectively) from genomic extractions was used for each PCR, using Mastercycler Pro thermocycler (Eppendorf, UK). Routine thermal cycling conditions using Taq polymerase are as follows: initial denaturation (1 min at 95°C), followed by 35 cycles of: denaturation (30 seconds at 95°C), annealing (1 min at 50-68°C, dependent on primer Ta) and extension (1 min kb^{-1} at 72°C). The final step is a final elongation phase for 5 min at 72°C and then the reaction can be held at 4°C. If Phusion was used, all conditions remained the same except for the following: initial denaturation at 98°C, denaturation (cycle) for 10 sec at 98°C, annealing for 30 sec and extension for 30 sec/kb.

Table 2.3 Components of Phusion PCR mastermix

Component	Working concentration
Nuclease free H_2O	Variable
5x Phusion HF (GC) Buffer/Standard Reaction Buffer	1 U (or 1 x)
Phusion polymerase or Taq polymerase	1 U per 50 µl reaction 1.25 U per 50 µl
dNTPs (10 mM)	200 µM
Forward Primer (10 µM)	0.5 µM
Reverse Primer (10 µM)	0.5 µM
Template DNA	< 250 ng

*****N.B. For primer sequences, see Table 2.4

The Ta (Annealing temperature) for PCR is dependent on the primer combination used in the assay, product length and primer melting temperature (Tm). Therefore, the Ta will be different for every primer set used. For each of the primer sets, an optimal Ta was identified by undertaking PCR at a range of temperatures, typically 3-5 °C below primer Tm (Table 2.4).

Table 2.4 Primer sequences

#	**Primer name**	**Primer sequences (5-3 orientation)**	**Gene target**	**Annealing temperature (Ta)**	**Reference**
1	pA	AGAGTTTGATCCTGGCTCAG	16S rRNA	55°C	(Edwards *et al.*, 1989)
2	pH	AAGGAGGTGATCCAGCCGCA			
3	515f	GTGCCAGCMGCCGCGGTAA	16S rRNA (V4 region)	50°C	(Caporaso *et al*., 2012)
4	806r	GGACTACVSGGGTATCTAAT			
5	L02988	CAACAATAGGGTTTACGACCTC	H1 mtDNA	56°C	(Brotherton *et al*., 2013)
6	H03017	AACGAACCTTTAATAGCGGCTG			

2.9.1 16S rRNA amplicon sequencing library preparation

Libraries for sequencing were prepared by myself using the MiSeq earth microbiome primer set 515f-806r. PCR amplification included 5 ng μl^{-1} of template DNA with 0.2 µM barcoded forward primer 515f

(AATGATACGGCGACCACCGAGATCTACAC-TATGGTAATT-GT-GTGCCAGCMGCCGCGGTAA) and 0.2 µM barcoded reverse primer 806r (CAAGCAGAAGACGGCATACGAGAT-NNNNNNNNNNNN-AGTCAGTCAG-CC-GGACTACHVGGGTWTCTAAT) under PCR conditions described (Table 2.4) (Caporoso *et al.,* 2012*),* using 7 different barcode/index sequences. Primer set 515f-806r targets the V4 variable region of 16S rDNA for amplification of bacteria. Products for each sample were gel purified to remove primers, dNTP's and primer dimers by separating them on a 1% agarose gel and using a QIAquick Gel Extraction Kit (Qiagen, UK). Libraries for sequencing were prepared by myself using the Nextera XT library preparation kit (Illumina) following manufacturer's instructions. Indexes 1 to 6 were used for each sample, respectively. The clean-up process was carried out using AmPure XP beads (Beckman Coulter) as per manufacturer's instructions. Part of the quality control was done by analysing the resultant DNA on a 1% (w/v) TAE agarose gel to assess the products for appropriate sizes (~ 300 bp). Samples from the 7 barcoded libraries were then quantified using a Qubit fluorometer and pooled at the recommended concentrations for the sequencing provider. The completed amplicon libraries were then sequenced on the Illumina MiSeq sequencer platform by the Centre for Genomic Research, Liverpool, UK.

2.9.2 Quantitative PCR (qPCR)

To quantify total number of copies of the 16S rRNA gene in the samples Bacterial small subunit (SSU) rRNA genes were amplified with the universal V4 primer set (Caporoso *et al.,* 2012) 515F and 806R (Table 2.4). Standards

were created for each primer set and serial dilutions of total copy number made (10^2-10^7). All samples, standards and negative controls were assayed in triplicate, with three technical replicates undertaken for each individual sample. Standards from 10^8- 10^2 gene copies were created by Paul Brotherton (BiSN Laboratory Services, UK). The identity of the amplicon was confirmed by size via agarose gel electrophoresis and melting curve analysis.

2.9.3 Primer design

Multiple primer pairs were assessed for the amplification of fragments of the SrfA operon (Table 7.1). All primers were designed using Primer3 (Untergasser *et al*., 2012) and IDT Oligo Analyzer (Owczarzy *et al*., 2008) and ordered from Eurofins MWG, UK. Primer sequences were approximately 18-24 bp in length and amplified a fragment of size 250-400 bp. The regions to be amplified were at the start and end of the *LchA* gene cluster, the gene responsible for encoding the Lichenysin lipopeptidde. It was therefore chosen that one amplified fragment should include the start codon of the *LchAA* gene and the other placed at the end of the *LchAD* gene. Each primer pair was evaluated for optimal annealing temperatures and conditions and the best performing primer sets taken forward to amplicon gel purification and Sanger sequencing to confirm the amplicon sequence was correct.

2.10 Agarose gel electrophoresis

DNA and RNA products were separated by electrophoresis using a 1% (w/v) agarose in Tris Acetate EDTA (TAE) buffer containing 0.5 µg μl^{-1} ethidium bromide (Fisher Scientific, UK). Loading dye buffer (0.3 volumes; NEB, UK)

was added to the sample (Gel Loading Dye NEB, USA). The molecular weights of the products were assessed in comparison to 5 µl HyperLadder™ 1kb (Bioline, UK). A constant voltage of 100 V was applied for 45-90 min for electrophoresis. Gels were visualised under UV light at a wavelength of 302 nm.

2.10.1 Gel excision of DNA

Agarasoe gels were visualised under UV light at 365 nm and the appropriate bands excised from the gel using a sterile scalpel. The gel band was then processed using the MinElute gel extraction kit (Qiagen, UK) as described by the manufacturer. DNA was then eluted into 50 µl of ddH_2O.

2.11 Sequencing process

Next generation sequencing was conducted by Molecular Cloning Laboratories (CA, USA) and and the Centre for Genomic Research, (University of Liverpool, UK) using MiSeq V2 Chemistry (Illumina, CA, USA). A total of at least 50 ng of genomic DNA was required for sequencing. Sanger sequencing was conducted by Source Bioscience (Nottingham, UK). For 16S rRNA data analysis, remaining primers and chimeric sequences were removed within the BIONMeta pipeline, using an extra cleaning and reference based search for removal of chimeras matching known sequences identified from NCBI.

2.11.1 Sequence data analysis

All amplicon datasets were run through the BionMeta pipeline (DGI, Arnhaus, Denmark) following standard parameters and taxonomy identified using the RDP database (Version 11.5, accessed 24/11/2016). All failed or low quality sequences and chimeras were removed using the BionMeta pipeline. Database searches for amplicon identification were performed using the Basic Local Alignment Search Tool (Altschul, S.F. *et al,* 1990).

2.12 Crude biosurfactant extraction

Biosurfactant-producing bacterial strains were cultivated in synthetic molasses media (500 ml) at 37°C. After 24 h incubation without shaking, the supernatant was separated from the bacterial cells via centrifugation (13,000 x g) for 20 mins followed by filtration (0.2 µm), before the cell free supernatant was acidified using HCL, to first a pH of 4.0 for 6 hours and then a pH of 2.0 overnight. Centrifugation (13000 x g) was used to remove the precipitated surfactant. Extraction of the supernatant was performed with chloroform: methanol (2:1 v/v) and then recovered by rotary vacuum evaporation at 80°C.

2.13 Statistical analysis

Statistical analysis of the data was conducted using R or Minitab to produce mean, standard deviation (SD) and standard error of mean (SEM) values. Graphical figures were produced using Microsoft office, R and Graphpad Prism.

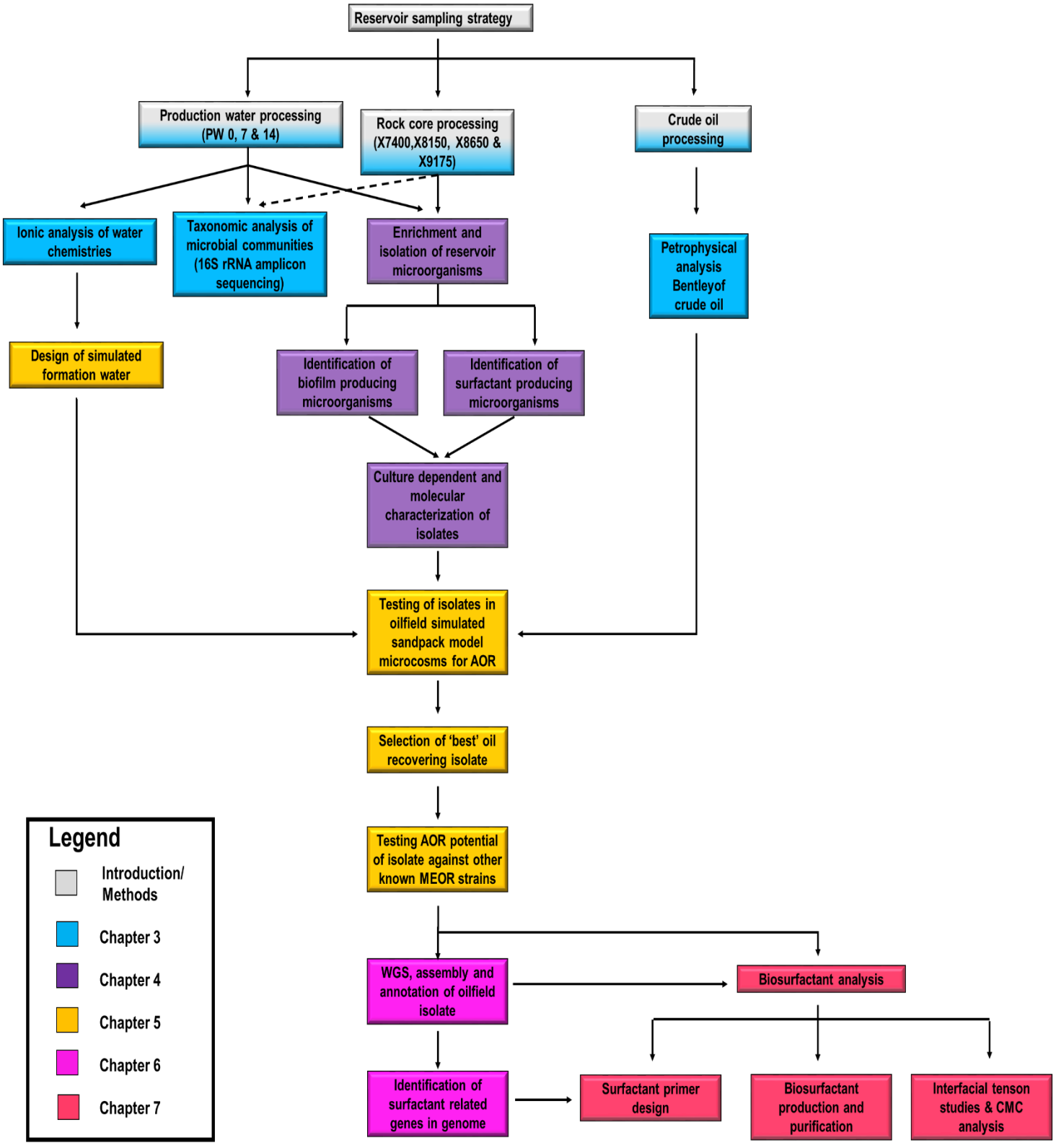


Figure 2.2 Flow diagram to show methodology of planned experimental work to be conducted in this thesis

Chapter 3

Petrophysical and Taxonomic Characteristics of an Offshore Unconventional Oil Field

3.1 Introduction

3.1.1 Background

To fully comprehend the oil recovery potential that an oil reservoir has, we must first be able to understand how the changes in petro-physical and microbial community composition impact upon the reservoir and its behaviour. Petro-physical and microbial community composition are strongly linked to oil field prosperity. Microorganisms were identified to flourish in oilfield communities as early as 1922 (Von Wolzogen-Kuhr & van der Vlugt, 1922; Bastin, 1926) and have been shown to be responsible for driving various oilfield processes, of both beneficial and detrimental nature (Davidova, Hicks, Fedorak, & Suflita, 2001*;* Wentzel *et al*, 2007). Yet the activity of a microbial community within an oil well is regulated by environmental and physical constraints. It is therefore imperative to base the determination of microorganism suitability for oil recovery enhancement (MEOR) upon reservoir temperature, pH, porosity, crude oil type and water chemistries, to yield the most beneficial effects for oil recovery from a specific reservoir.

3.1.2 Geology and geomorphology of reservoir for this study, the Bentley Oilfield

The work conducted throughout this thesis was centred on an offshore heavy oil (API of <20°API) reservoir. This reservoir is known as reservoir 9/3b-7, or

the Bentley Oilfield. It is situated in the North Sea, 53 miles off the coast of the Scottish Shetland Isles' (Fig. 3.1), in an area spanning ~77 sq km. This region is identified as block 9/3b (Fig. 3.1). The Bentley field was originally discovered in 1977 by Amoco Exploration Ltd. and subsequenctly leased to Xcite Energy on 1st October 2003 on a 26 year license. The Bentey Oilfield is the only reservoir located in block 9/3b and has a total surface area of ~52 mi^2, under an average water depth of ~370 ft. The oil cap itself is situated approximately 3600 ft below sea level.

Geologically, the Bentley Oilfield is classified as an unconsolidated reservoir consisting primarily of *Dornoch* sandstone. The oil cap was identified to be in the upper part of the *Dornoch* formation, which is from the late *Palaeocene* era and moving into Eocene Balder formation tuffs, which enclose the reservoir (Fig 3.2). There is also a shelf of a sandier base of *Dornoch* lithology that overlays the *Lista* shale. This shale consists of a composite, dislocated channel of loose and uncemented multidarcy fluvial sands (Brennan *et al,* 2014). It is this geological characteristic that gives the reservoir its high permeability and porosity characteristics, due to differences in grain size under less compaction, which allows water to pass through with ease (Avseth, P. *et al* 1998). The hydrocarbon pocket is bound by a four-way dip closure, an area where the stratigraphic contours diverge in all four directions.

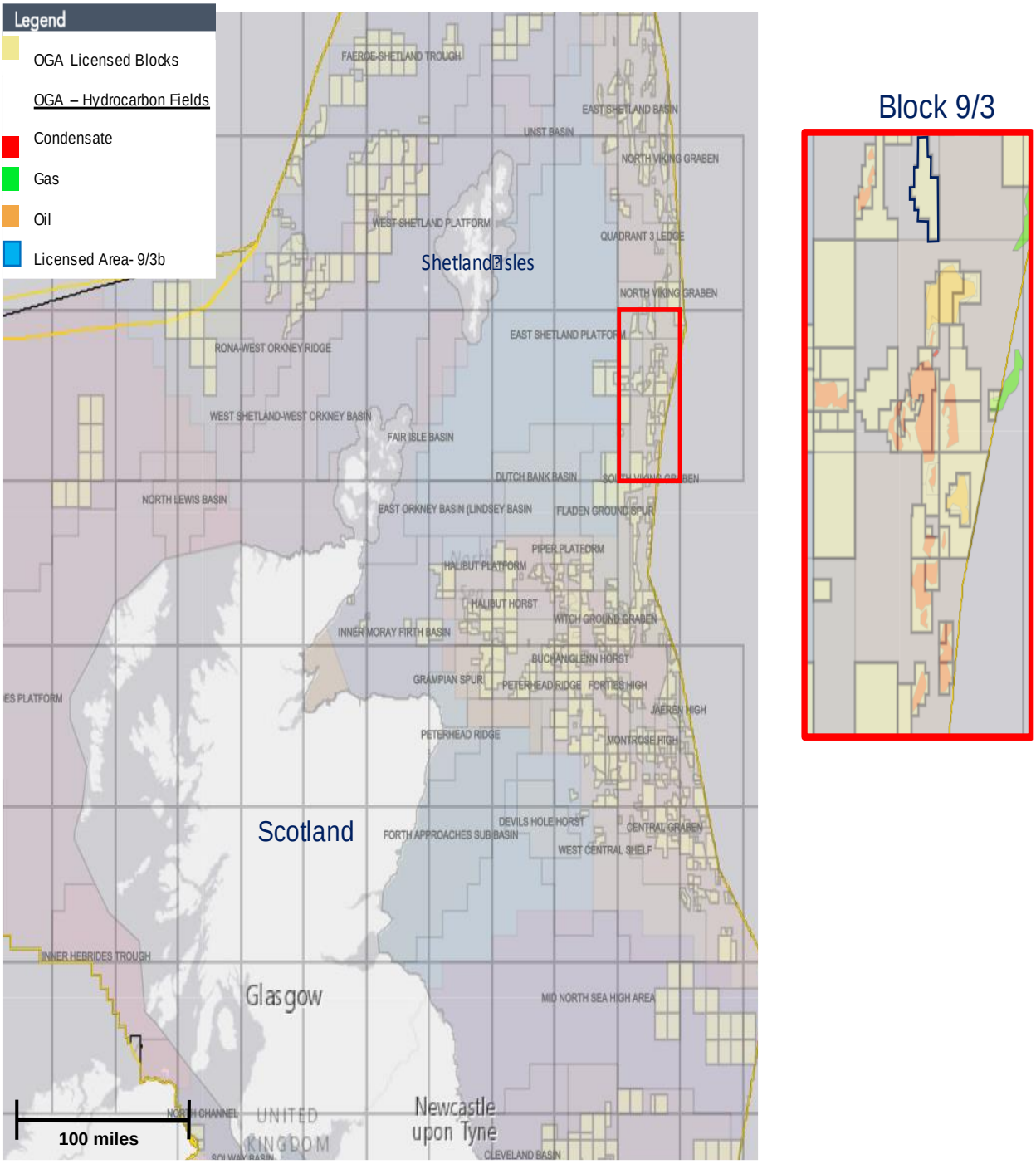


Figure 3.1. Geographic information system (GIS) map showing location of Bentley Oil Field (9/3b) and surrounding North Sea offshore reservoirs.Map area shows the distribution of oil reservoirs around the coast of Scotland, UK. Gas and oil reservoirs highlighet in green and orange respectively, with reservoirs producing predominantly gas but some hydrocarbon liquid (condensate reservoirs) shown in red.(Images obtained from Oil and Gas Authority, 2016. Accessed at: ogauthority.maps.arcgis.com/)

The reservoir is positioned above a water aquifer with a volume in excess of 13 bbls. The aquifer provides natural oilwell pressurization, and the aquifer may also assist in oil production via a bottom or peripheral water drive (Ehlig-Economides, Chan, & Spath, 1996). A water drive involves the influx of formation water into the reservoir through the aquifer, allowing for the natural flow of water through to the production wells.

3.1.3 Reservoir rock characterization

The ability of microorganisms to survive in oil reservoirs or subsurface environments is heavily dependent on the characteristics of that subsurface environment (Magot *et al.,* 2000). Until recently, it was believed that reservoirs were environments that were too hostile for the survival of microorganisms. However, through the advancement of both microbial detection and drilling technologies and the need for operators to understand their reservoirs' ecologies, it is now appreciated that an assortment of microorganisms are present in vast quantities in all oil reservoirs (McInerney *et al,* 2007).

The average temperature of well 9/3-b was measured to be at 40°C (104°F). This is classified at the upper temperature limit for a mesophilic reservoir (20°C - 45°C), but within the optimal temperature range for both microbial proliferation and metabolism. Pressure within the well is also an important consideration for the optimisation of microbial growth. Although the amount of literature available on pressure and its impact on microbial behaviours is limited, the suggestion that pressures above 10,000 psi, decrease the growth

rate of microorganisms and negatively impact microbial adaptation to additional environmental conditions is well accepted (Ravot *et al*., 1997). Reservoir rock porosity and permeability has not been previously shown to affect bacterial growth; however, it may play a role in the flow of nutrient. Throughout the formation strata of the Bentley field, rock porosity was estimated to fall between the ranges of 35-41%. This was dependent on the location and depth of assessment, with the areas of sandstone enclosing the oil cap and the deeper shale being of a higher porosity (Figure 3.2). Permeability was consistently measured at ~7.5 Darcy throughout the hydrocarbon basin. Pressure was calculated to be at ~1800 PSI at the oil water contact point (~3700ft) once initial oil breakthrough had been achieved. In comparison, standard producing wells of this depth, under normal hydrostatic pressure (0.465 psi/ft), would be estimated to have a pressure of 1700 psi.

3.1.4 Reservoir exploration and total oil in place

In 2012, after the successful drilling of two vertical wells 9/03b-5 and 9/03b-6, demonstrating oil flow was achievable, it was decided that the oilfield parameters would be analysed to better prepare for the next phase, production. Two horizontal extended test wells (Fig. 3.2) were drilled to form a planar dual lateral well (RPS Energy, 2009). This process, known as multilateral well drilling, is economically beneficial to reservoirs of short depth, as only a single horizontal reservoir is needed, as opposed to numerous vertical boreholes (Joshi, 2003). Furthermore, due to the presence of an

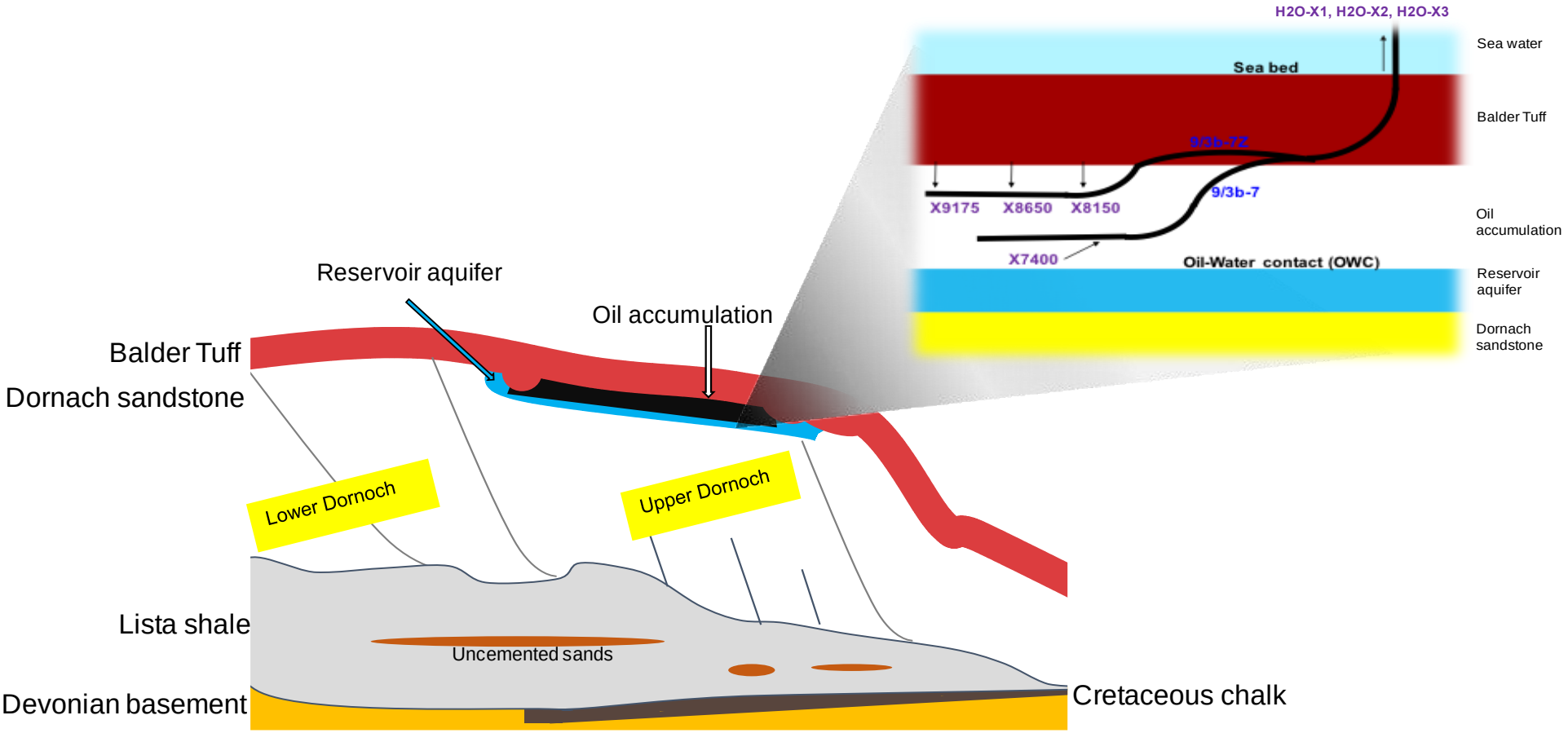


Figure 3.2 Schematic of Bentley Oilfield. Cross sectional layout of block 9/3b showing geological features of reservoir. Text denotes sampling points from wells 9/3b-7 & 9/3b-7Z.

aquifer underlying the oil phase, it was more feasible to use multilateral well drilling to ensure that secondary recovery (through thermal injection) would likely be more successful. With vertical wells, steam may enter the aquifer rather than the hydrocarbon pocket, thus reducing the recovery efficiency (Elyasi, 2016). The drilling of planar lateral wells resulted in the construction of well 9/03b-7 and its interconnected multi-lateral extended well, 9/03b-7Z (Fig. 3.2). 9/03b-7 was initially drilled as a lower lateral well, 60ft above the point of oil/water contact, to ensure water break-through. 9/03b-7Z was split off at ~7000 ft to form an upper lateral well, geosteered 10 ft below the hydrocarbon cap.

It is estimated that at a base case, this oil field would be able to readily produce 204 MMbbls of heavy oil (90% production certainty) via primary and secondary production. An additional 113 MMbbls of heavy oil (10-50% production certainty) is postulated to be available for further recovery by tertiary methods. Furthermore, ~48 MMbbls are also available from contingent reserves of 9/03b. Bentley is currently one of the largest remaining assets in the North Sea, surpassing its unconventional hydrocarbon producing neighbors, Bressay and Kraken (Wood-McKenszie, 2015). Base flow assessments demonstrated that the current test production rate plateaus at 3500 bbl/d. It has been hypothesized, however, that this rate will increase to above 15,000 bbl/d when production commences (Xcite Energy, 2014). Based on current lifetime assessments, it is estimated that the extent of the Bentley Field reserves could allow for a production profile that extends beyond 35 years, with the use of Enhanced Oil Recovery techniques.

3.1.5 Surfactant producing microorganisms

Surfactants and the microorganisms that produce them are paramount to the theory and practicality of MEOR (Chapter 1.3.1). Whether surfactants function by altering surface tension or through assistance in hydrocarbon biodegradation or biofilm formation, surfactant production is considered to be a vital mechanism in increasing the volumes of hydrocarbon that can be produce (Banat *et al.*, 2000; Saha & Bhaskara Rao, 2016). It is therefore imperative to understand the genes and pathways involved in surfactant production within specific environments. This knowledge will provide information on autochthonous microorganisms that are likely involved in surfactant related processes and provide potential for the identification and characterization of new genes. With the identification and characterisation of surfactant production mechanisms in mind, it is becoming increasingly important to create databases with reference to specific subsets of genes. Although many gene databases already exist, such as Uniprot, CAZymes and NCBI, most are too specific to certain research fields or lack accurate manual curation and constant updating. Therefore, the creation of a surfactant specific database, BioSurf (Oliveira *et al*., 2015), aims to overcome these challenges and provide a reference for genes that may be involved in the production of surface active compounds (SACs), and each of the associated microorganisms that are capable of producing these surfactants.

3.1.6 Research hypothesis, aims and objectives

The aim of this chapter was to successfully characterise the petrophysical and chemical properties of the Bentley Oilfield reservoir, as well as the microbiological communities that reside in the rock core and production waters. With this data, it will be possible to assess community differentiation based on petrophysical parameters, such as depth, water chemistries and sample type. The secondary aim of this chapter is to attempt to identify any potential surfactant producing microorganisms that may reside in the reservoir, using functional prediction based on taxonomy. These results will act as a guide for future culturing and isolation work to be conducted in Chapter 4. The data within this chapter tested the hypothesis that surfactant producing microorganisms are present in both produced water and rock core samples taken from an offshore heavy oil reservoir.

3.2 Methods

3.2.1 Sampling and sample locations

The samples used in this study were collected to aide with outlining the petrophysical and microbial parameters of the pre-production reservoir (Figure 3.2). Formation water samples were collected from a single sampling valve at the reservoirs well head. Three replicative samples were taken at three selected time points from one well head at 7 day intervals (commenced 12 September 2012 from well 9/03b-7). Each sample underwent a flushing process to remove foreign objects and residual oil that may have amassed over time and were then collected using a filter system, as mentioned previously (Chapter 2.4). Rock core material (X) was also collected for analysis. Four segments of rock core from variable depths (~ 7400ft, 8250ft, 8650ft & 9175ft) were collected from one bore hole of two lateral developmental test wells (Figure 3.2). Rock core sample X7400 was collected from test well 9/03b-7. The remaining three core samples (X8150, X8650, X9175) were collected from the second phase of the multilateral test well, 9/03b-7Z. Samples of ~ 15-20 cm in length, were stored at 4°C for transportation by the producer (Xcite Ltd.) and processed upon arrival in the laboratory (~ 2 days). On arrival, 3 cm from each end of the core pieces were removed to reduce contamination using an autoclaved saw and hammer and chisel. The carved inner core was ground down with a sterilized pestle and mortar to form the basis of material for DNA analysis. Three individual replicates of produced water and rock core were taken and pooled after relevant processing was undertaken (e.g. nucleic acid extraction).

3.2.2 Geochemical analysis

Produced water samples (~400 mL) were sent to Exova Ltd (UK) for chemical ion analysis. Chemical analysis of each sample was undertaken in September 2014. The pH of each produced water sample was measured on a 50 mL allocation of the remaining water using a fully calibrated Accumet AR50 pH meter. Total % salinity was calculated using the following formula based on geochemical ion concentrations.

$$\boldsymbol{Salinity\ (ppt) = 0.0018066\ \mathrm{x}\ Cl^{-}\ (mg/L)}$$

The physical properties of oil were also measured in the laboratory. Oil samples recovered from the reservoir were sent in 80 L capped drums and aliquoted into 500 mL beakers. Viscosity was calculated using a Brookfield DV-1 Viscometer, at a high viscosity setting. A thermal plate was used to heat the oil to 40 °C and the viscosity subsequently measured.

3.2.3 Data analysis

Post data processing all taxonomic output data, collated via BIONmeta (See Chapter 2.11.1) , was passed through the MEGAN program (Huson *et al*, 2016) to group taxonomic classifications and produce figures of publication standard. Beta analysis was conducted using both MEGAN and QIIME, using a sampling depth of 215,000 reads. Canonical correlation analysis (CCA) and starplots were undertaken using the statistical environment of R.

3.2.4 Function from taxonomy

In addition to phylogeny, function was also assessed from taxonomic data. This was achieved by passing BIONmeta output data through the Tax4Fun

pipeline (Aßhauer *et al.*, 2015). Functionality was based on the KEGG Orthology Database at a three-tier level and aligned using the RDP database (RDP 11.4: Downloaded 20/7/2015).

3.2.5 Metagenetic library preparation for amplicon sequencing

Total DNA extraction and amplification of 16S rRNA genes was conducted as stated in Chapters 2.9 and 2.11. Three replicate samples of extracted DNA were taken for each sample type and underwent initial amplification before being pooled prior to addition of barcodes and sequencing adapters. Pooled replicates were then amplified using Illumina sequencing primers, barcoded in accordance to sample type (Barcodes 1-7 were used) and with sequencing adapter sequences present in primers. Once amplification was completed, amplicons were purified and pooled before being sent for sequencing on MiSeq V2 (Chapter 2.11).

3.3 Results

3.3.1 Petrophysical and geochemical characterization of Reservoir 9/3B-7Z

Three water samples were taken from a single reservoir at three different time intervals related to sampling date (PW0:T0, PW7:T7, PW14:T14 days) and analysed for inorganic chemicals and metal ions. Variations in geochemical concentrations across these time points are listed in Table 3.1.

Water samples collected from reservoir 9/3B-7Z were classified as hypersaline, based on chloride ion concentration relative to water density and gravity. The salinity was approximately 5.6% (5.4-5.8%) with a Cl^- concentration of 30.1 to 31.9 g L^{-1}, a concentration of almost two times that of standard sea water. Between sampling points, differences can be seen in ionic concentrations (Fig. 3.3). The ions Na^+, K^+, Ca^{2+} and Cl^- were present at concentrations > 500 mg L^{-1}. Both Ba^{2+} and Ca^{2+}, were initially found at higher concentrations in PW0, containing concentrations 2.5 times that found in PW7 (Fig. 3.3D, Fig. 3.4). The concentrations of Ba^{2+} (41.7 – 43.2 mg L^{-1}) and Ca^{2+} (1140 – 1310 mg L^{-1}) remain comparatively constant between samples PW7 and PW14. Total iron, K^+, and bicarbonate ions are all present in high concentrations in PW0 (Table 3.1, Fig. 3.3 D). However, the concentrations are reduced over the period when samples PW7 and PW14 are taken, with the concentration of iron being reduced by over 10 fold (Table 3.1, Fig. 3.4). Other ion concentrations, such as Cl^- and BO_3^{3-}, remained generally constant between sampling points (Table 3.1). The produced water samples PW7 and

PW14 possessed levels below the equipment detection limit (0.1 mg L^{-1}) for SO_4^{2-}, HCO_3^- and Al^{3+} so were not recorded in

Table 3.1 Ion analysis of aqueous produced water samples from reservoir 9/3B-7Z

Geochemical parameter	PW 0 $T_{0\ days}$ (mg L^{-1})	PW 7 $T_{7\ days}$ (mg L^{-1})	PW 14 $T_{14\ days}$ (mg L^{-1})
Acetate ($C_2H_3O_2$)	62	22.4	23.2
Aluminium (Al^{3+})	N/A *	Undetectable +	Undetectable
Barium (Ba^{2+})	14	41.7	43.2
Bicarbonate (HCO_3^-)	530	N/A	N/A
Boron (BO_3^{3-})	14	13.2	12.9
Calcium (Ca^{2+})	155	1140	1310
Chloride (Cl^-)	31908	30100	31200
Copper (Cu^{2+})	N/A	14	22.3
Dissolved Iron (Fe^{2+})	N/A	1.42	0.85
Formate (HCO^{2-})	37	Undetectable	Undetectable
Lithium (Li^+)	1	1.04	1.01
Magnesium (Mg^{2+})	819	456	511
Manganese (Mn^{2+})	N/A	1.04	1.01
Phosphorus (HPO_4^{2-})	N/A	0.9	0.1
Potassium (K^+)	2519	493	485
Sodium (Na^+)	15897	20100	20500
Strontium (Sr^{2+})	119	105	107

Sulphate ($SO4^{2-}$)	56	Undetectable	Undetectable
Total Iron (Fe^{2+}, Fe^{3+})	23	1.42	0.852
Zinc (Zn^{2})	N/A	0.85	0.74
Salinity (%)	5.8	5.4	5.6
pH	7.50	7.38	7.44

* N/A – Value was not measured on selected sample due to ionic measurements available at testing house.
+ Undetectable – Value was below detection limit of equipmePW0

Table 3.1. Fe^{2+}, Al^{3+}, HPO_4^{2-}, Cu^{2+}, Zn^{2+} and Mn^{2+} were not measured in PW0 nor was HCO_3^- in samples PW7 and PW14 (Table 3.1). The pH of the produced water was neutral across all three samples. pH was, however, marginally elevated in PW 0 compared to the latter samples. (Table 3.1). Alongside the sampled production water, produced oil was also collected from the reservoir for petrophysical testing (Table 3.2). The viscosity of live oil (627 cP) was substantially lower than that of dead oil (1405 cP) when analysed in the laboratory. However, by increasing the temperature to match that of oilfield parameters to 40 °C, viscosity was reduced to 709 cP, levels almost comparable to that of what is found in the live oil at reservoir conditions.

3.3.2 Microbial diversity of oilfield reservoir samples

To assess the diversity of the microbial community within the oil reservoir, metagenetics was used to generate an accurate depiction of the community. Water samples previously collected for petrophysical investigation (PW0, PW7

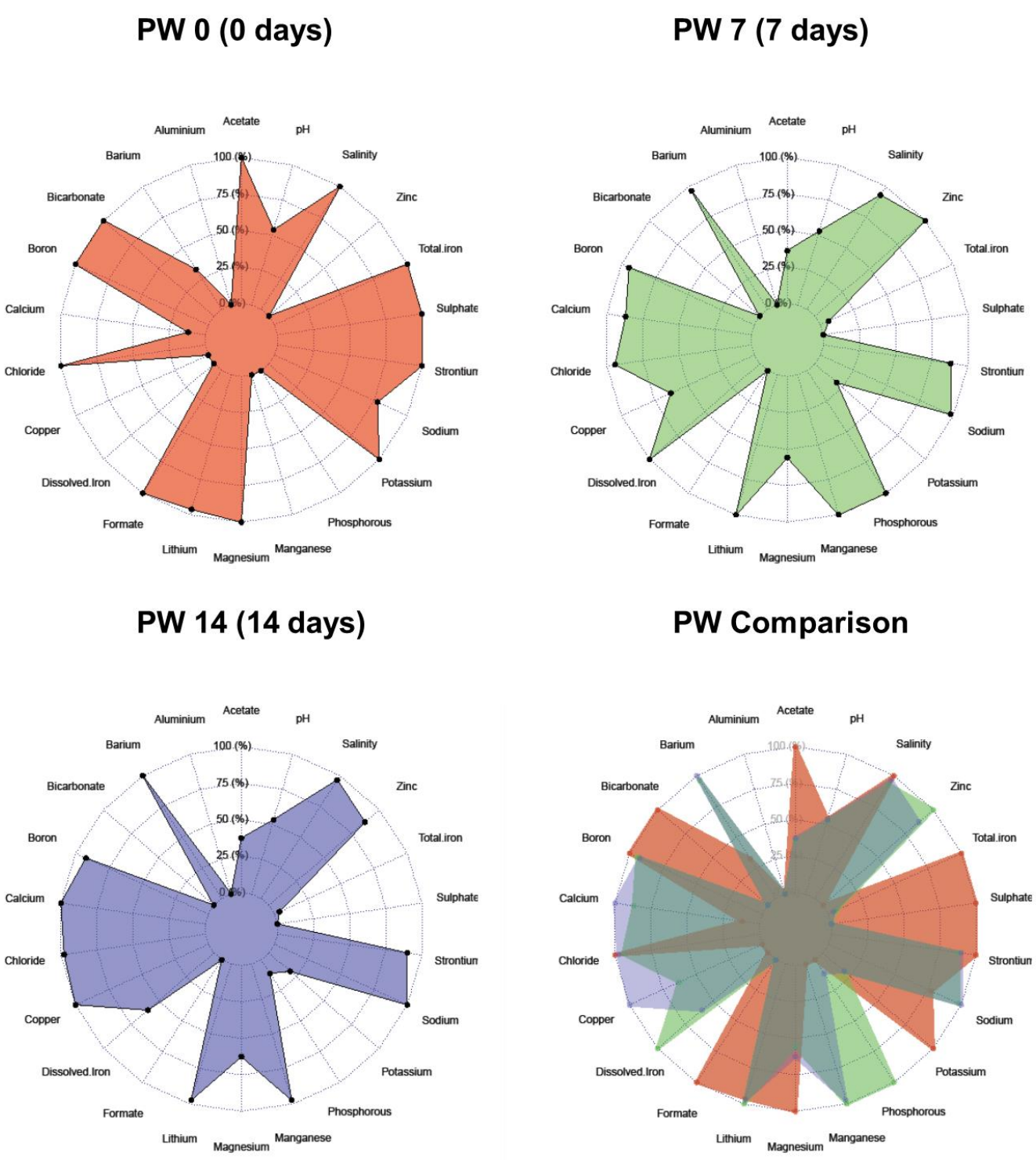


Figure 3.3 Comparison of PW geochemical and physiological characteristics over 14-day sampling period. Total ion concentration, pH and salinity shown for all three PW sampling points (PW0, PW7 & PW14). A) Total ion concentrations and petrophysical parameters of PW0. B) Total ion concentration and petrophysical parameters of PW7. C) Total ion concentrations and petrophysical parameters of PW14. D) Comparison of total ion concentration and petrophysical parameters of produced water samples from 0, 7 & 14 days .

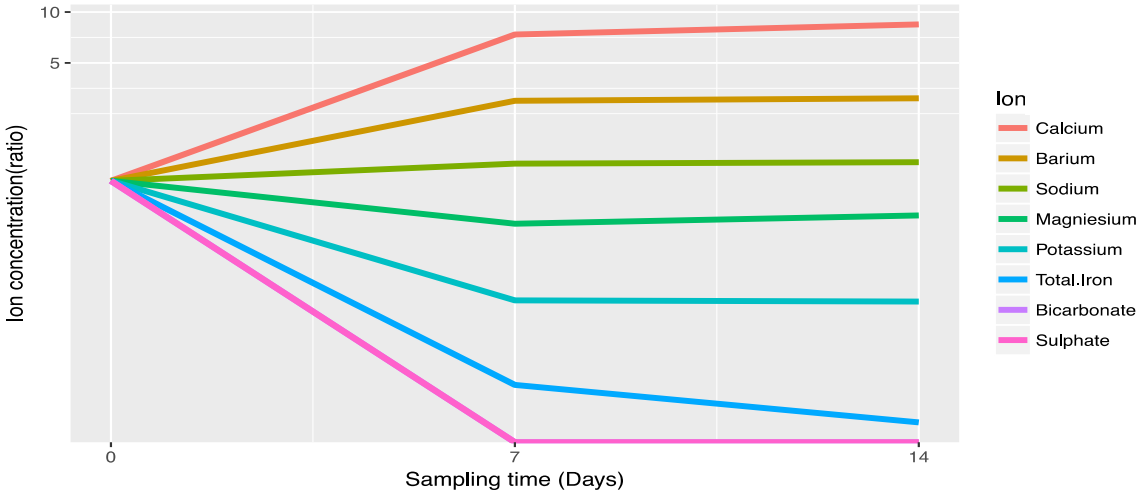


Figure 3.4 Comparison of key ionic variations in produced water samples over 14 day period of sampling. Total ion concentration converted to fold changes between 0-7 (PW0 – PW7) and 7-14 (PW7 – PW14) day sample points. Concentration change only shown for ions where significant difference is noted in one or both sampling time frames.

Table 3.2 Petrophysical parameters of oil samples collected from reservoir 9/3B-7Z

Petrophysical parameters		
Hydrocarbon properties		
Hydrocarbon concentration		Unknown
API Specific gravity		12° 0.9861
Viscosity (live oil)¥		627 cP
Viscosity (dead oil)‡	27 °C 40 °C	1405 cP 709 cP
Density		984 kg m^3

¥ Oil containing specific concentration of dissolved gas found at depth and pressure conditions

‡ Low pressure crude oil containing no dissolved gas or oil that has lost volatile componen

& PW14), alongside four samples of rock core derived from differing depths (X7400, X8150, X8650 & X9175), underwent DNA extraction, 16S rRNA amplicon PCR and were analysed using amplicon sequencing on the Illumina Miseq platform. The sample information and resultant DNA concentrations are listed in Table 3.3.

The taxonomic variances amongst the oilfield communities were identified based on 16S rRNA gene sequencing obtained from the seven sample points. A total of 3,004,255 reads were sequenced. After analysis and quality filtering, 2,798,611 reads were assigned to bacterial taxonomic groups, with 2290 reads assigned to archaeal taxa. Between 4.5-11.7% of sequences did not match at 90% similarity or identify to a bacterial taxon lower than the rank of phylum (Table 3.4). From the assigned sequences, in all four core samples (X7400, X8150, X8650 and X9175), the dominating phyla consisted of *Proteobacteria* (93.9%, 87.13%, 92.2% 61.76%, respectively; Fig 3.5). *Proteobacteria* also dominated PW0 produced water sample (61.76%). In the remaining two water samples (PW7 & PW14), members of the *Firmicutes* were in greatest abundance. Of the sequences obtained from sample X7400, alongside sequences belonging to the phyla *Proteobacteria,* phyla of *Firmicutes* (3.83%) *Actinobacteria* (1.89%) and *Bacteroidetes* (0.57%) accounted for the remaining proportion of the population. The remaining 17 phyla accounted for <1% total community. At a class level, members of the *Gammaproteobacteria* account for 91.18% and the *Betaproteobacteria* 1.12% of the total *Proteobacteria*. *Bacilli* (2.86%) and *Clostridia* (0.97%) comprise the majority of *Firmicutes* in this rock core sample.

Table 3.3 Reservoir sampling information of seven samples (PW, n=3; X, n = 4) and concentration of total extracted nucleic acid

Sample point	Sample location	Sample type	Sampling information	DNA concentration (ng μl^{-1}) ± SD
PW0	9/03b-7	Produced Water (PW)	Day: 0	99.01 ± 7.17
PW7	9/03b-7	Produced Water (PW)	Day: 7	112.89 ± 4.01
PW14	9/03b-7	Produced Water (PW)	Day: 14	102.30 ± 4.48
X7400	9/03b-7	Reservoir rock core	Depth: 7400 ft	13.39 ± 3.95
X8150	9/03b-7Z	Reservoir rock core	Depth: 8150 ft	13.84 ± 6.30
X8650	9/03b-7Z	Reservoir rock core	Depth: 8650 ft	20.28 ± 5.19
X9175	9/03b-7Z	Reservoir rock core	Depth: 9170 ft	17.68 ± 7.17

Table 3.4 Output statistics for BIONMeta pipeline from 16S rRNA gene sequencing data analysis

Sample	X7400	X8150	X8650	X9175	PW0	PW7	PW14
Total original input reads (Read count)	630696	614529	357870	510801	232424	382203	275732
Total reads with similarities <90 % (Read count)	46943	34672	22987	22194	17230	32695	28923
Similarities < 90% (%)	8.0%	6.0%	6.9%	4.5%	8.0%	9.4%	11.7%
Total reads mapped to a unique phylum (Reads count)	583753	579857	334883	488607	215194	349508	246809

The communities (99.4%) associated with samples X8150 and X9175 were comprised of the following phyla: *Firmicutes* (4.87% and 5.02%, respectively), Cyanobacteria (0.97% and 0.6%, respectively), *Actinobacteria* (0.92% and 0.97%, respectively), *Bacteroidetes* (0.7% and 0.59%, respectively) and, as already determined in sample X7400, the most abundant *Proteobacteria* were the *Gammaproteobacteria* (89.7 and 89.8%, respectively). However, *Alphaproteobacteria* (0.73 and 0.76%, respectively), *Betaproteobacteria* (1.22 and 1.36%, respectively) and *Epsilonproteobacteria* (1.03 and1.22%, respectively) were also present in ample numbers. Members of the *Bacilli* (3.93% and 3.91%, respectively) and *Clostridia* (0.97 and 1.13%, respectively) were the only *Firmicutes* identified in totals greater than 100 reads (Fig 3.6).

Together with *Proteobacteria*, the community structure of X8650 comprised in the majority of members of the phyla *Firmicutes* (7.03%), *Actinobacteria* (2.29%), *Bacteroidetes* (1.38%), *Cyanobacteria* (0.94%) and *Fusobacteria* (0.29%). At a class level for X8650, *Proteobacteria* was very similar in make up to that of X8150 and X9175 for each class, but the total abundance of *Proteobacteria* was lower than in the other rock core samples. The main differentiation was that *Gammaproteobacteria* (82.74%) was of a lower relative abundance in X8650 and *Alphaproteobacteria* (1.21%), *Betaproteobacteria* (1.88%) and *Epsilonproteobacteria* (2.28%) were present in greater abundancies. Members of the *Bacilli* (5.52%) and *Clostridia* (1.58%) were also increased in sample X8650 compared to other rock samples.

The three produced water sampling points harboured much more bacteria that the rock core samples. The water samples harboured the same four main phyla present in all the core samples. Alongside *Proteobacteria (*19.15-61.66%), *Firmicutes* (33.33-78.54%), *Actinobacteria* (0.54-2.11%) and *Bacteroidetes* (0.51-1.33%) predominated. *Cyanobacter* (0.38%) and *Acidobacteria* (0.19%) comprised the rest of the major phyla of PW0. Whilst only *Acidobacteria* (0.92%) accounted for the remaining principal phyla for sample PW7 (Fig 3.5).

Community member shifts were identified in the produced waters over the 14 days of sampling. From day 0 (PW0) to day 7 (PW7) there was a shift in the microbial communities from *Proteobacteria* to *Firmicutes*. At a class level

(Figure 3.6), the produced water samples were dominated by members of the *Gammaproteobacteria* (48.28%) at day 0. The population of PW7, although now transformed to a population of predominantly *Firmicutes*, was more diversified. The *Firmicutes* comprised of *Clostridia* (40.94%) and *Bacilli* (38.73%). Furthermore, *Betaproteobacteria* numbers increased (5.98%-9.15%), with the *Gammaproteobacteria* population diminishing (6.99%). By day 14 *Bacilli* (48.62%) now dominated the community and *Firmicutes* population, with *Clostridia* only accounting for 18.29%. *Gammaproteobacteria* numbers had additionally ascended (23.59%).

Cluster analysis using the Bray-Curtis dissimilarity calculation showed considerable variation between the microbial communities of the three PW and four X samples together with similar coordination using principal coordinate (PC1=98.62%variation). Bray Curtis dissimilarity is one of the most commonly used non-phylogenetic β-diversity analyses in microbial and ecological biodiversity. Although this method only quantifies compositional dissimilarity based on total abundance, rather than genetic relationship, the stringency remains high for samples where certain taxonomy are not present (null values), an issue that was encountered with these datasets (Jovel *et al.,* 2016; Wang *et al.*, 2016).

The three PW samples are more varied, with PW0 less related to PW7 and PW14. PC2 (1.27% variation) demonstrates the closeness of samples X8150 and X9175, with all core samples related in their own clade (Fig. 3.7 A). This indicates that at a genus level, although some similarity can be seen between

the rock core samples, samples X8150 and X9175 share the most taxonomic similarity compared to the other samples (X7450, X8650, PW0, PW7 & PW14). With regards to the PW samples, PW7 and PW14 are the more similar samples, with PW0 regarded as a dissimilar water sample, more closely related to the rock core samples. This is further illustrated in the UPGMA tree, with core samples having closer distances to each other than the PW samples, which are less related and have high variation (Fig 3.7 B).

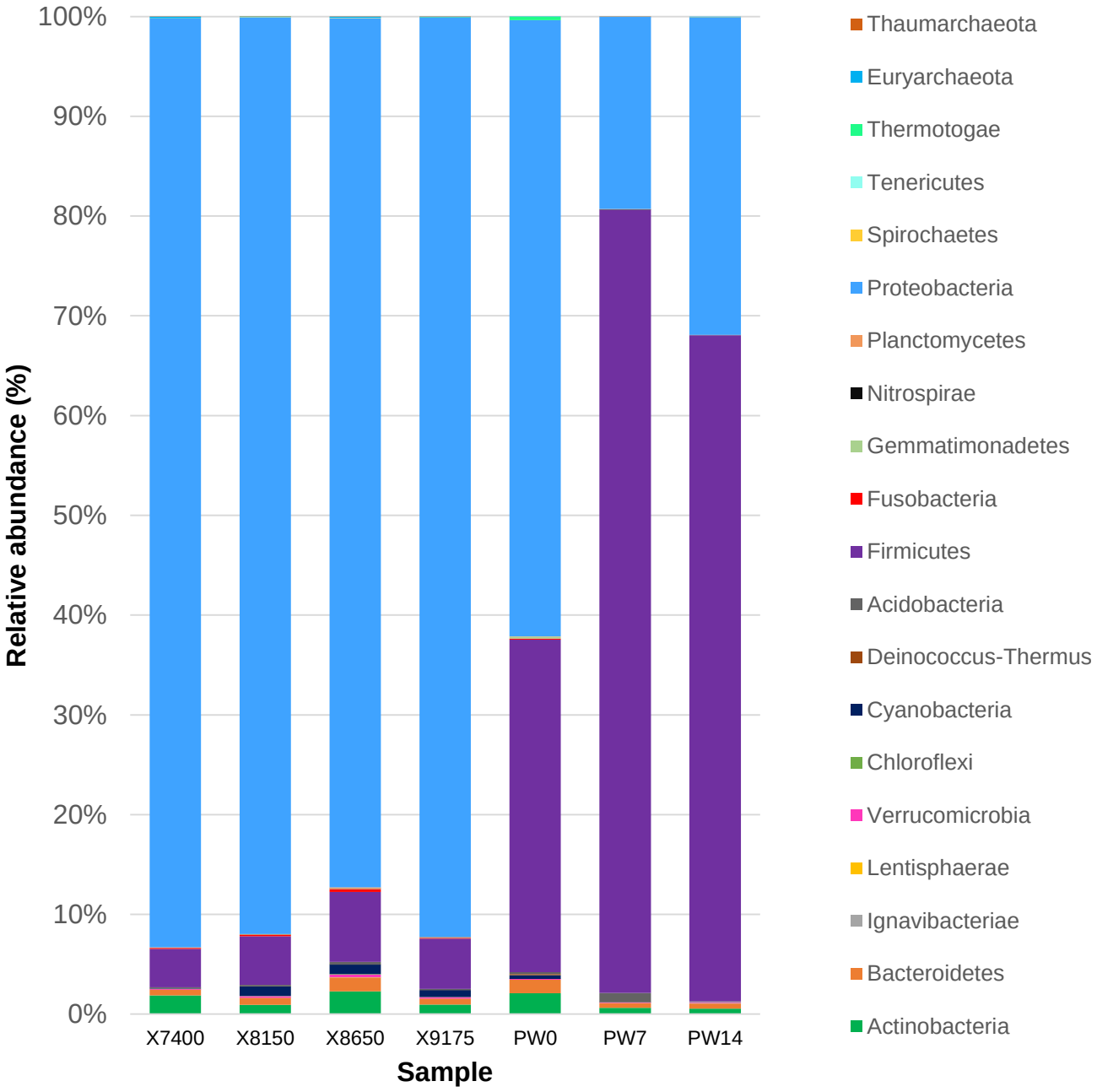


Figure 3.5 Microbial characterization of oilfield samples via 16S rRNA gene library sequencing. Abundance (%) of represented phyla from the 16S rRNA gene PCR amplification of rock core (X#) and produced water (PW#) samples. BIONmeta was used to classify and characterize reads obtained and RDP database used as a reference database for all samples.

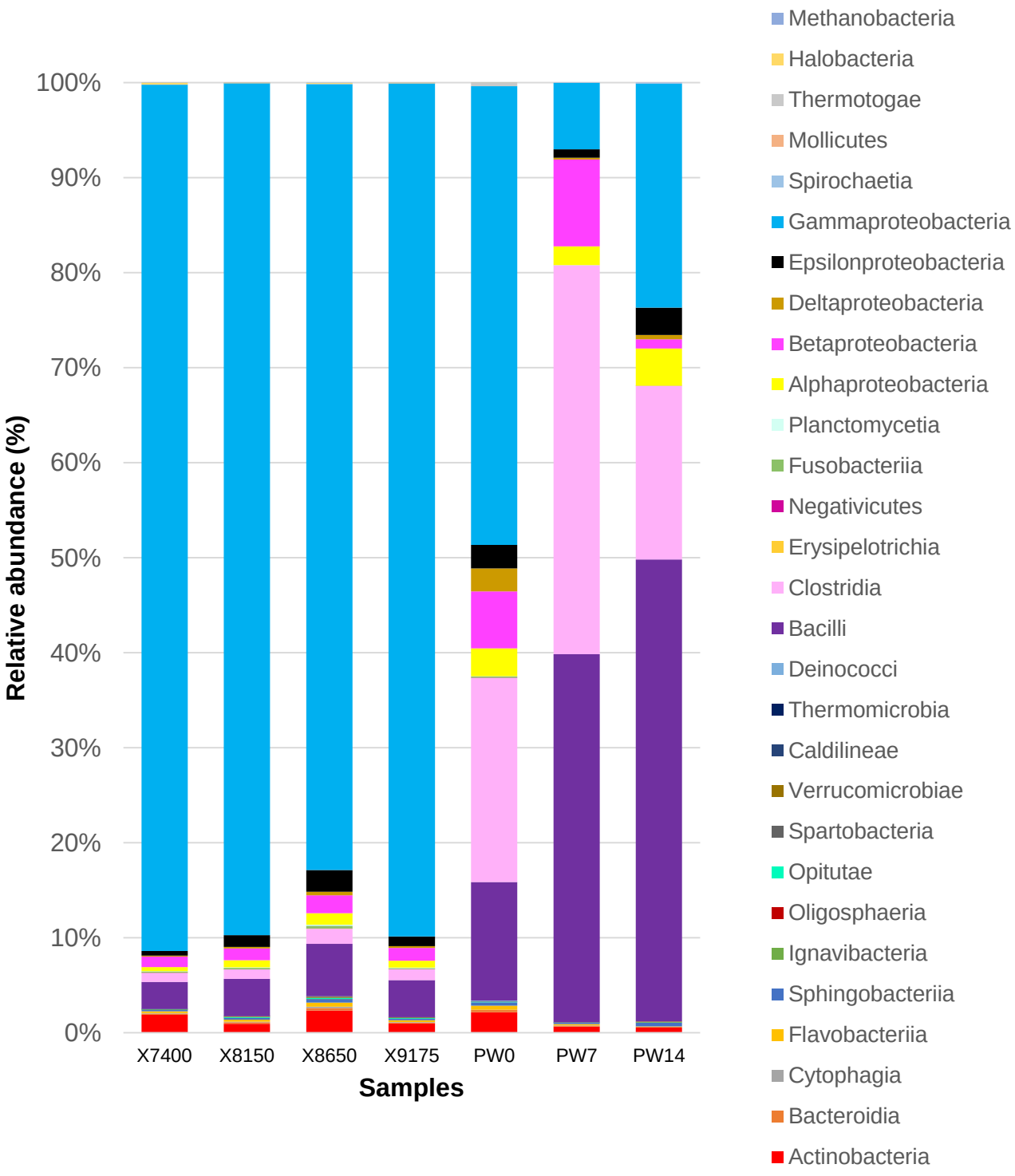


Figure 3.6 Microbial characterization of oilfield samples via 16S rRNA gene library sequencing. Abundance (%) of represented class from the 16S rRNA gene PCR amplification of rock core (X#) and produced water (PW#) samples. BIONmeta was used to classify and characterize reads obtained and RDP database used as a reference database for all samples.

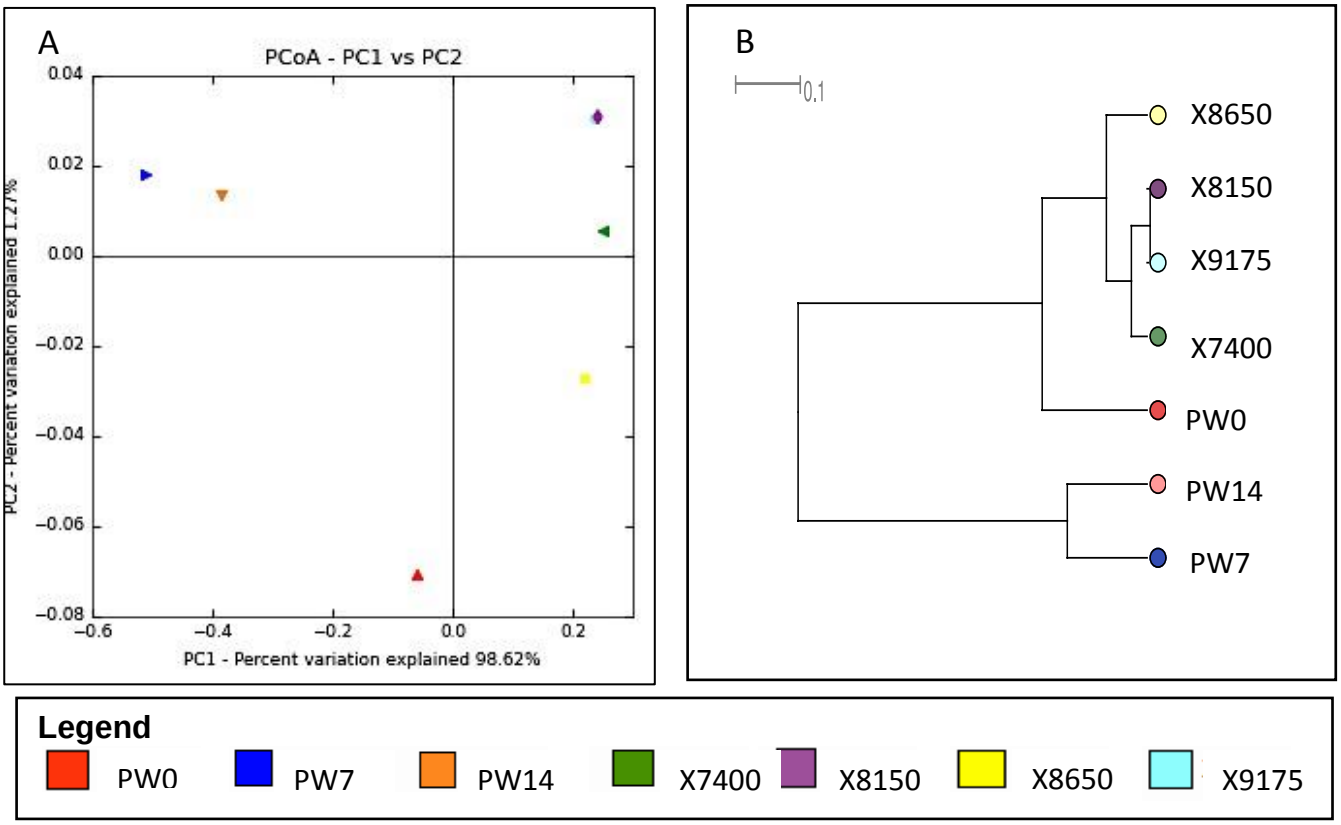


Figure 3.7 Principal coordinate analysis (PCoA) of microbial community for comparison of oilfield samples. A) 2D PCoA plot of PW *(n* = 3) and X (*n* = 4) samples using Bray-Curtis algorithm and an even sampling depth of 215000 reads to assess sample similarity. B) UPGMA tree of microbiological population structure of seven oilfield samples (PW, *n* = 3; X, *n* = 4) using Bray-Curtis distances of dissimilarity algorithm, conducted using QIIME and BION-Meta output file.

However, PW0 appears to share more similarity to core samples than the other PW samples. The relationship of the distribution of microbial taxa amongst samples and environmental parameters of the Bentley field were revealed by CCA analysis (Fig. 3.8 A & B). At both the phylum and genus levels, four constrained environmental parameters were identified as the explanatory factors for the variations in microbial communities ($p < 0.05$). These parameters were oilfield depth and concentrations of boron, acetate and barium ions. All other oilfield characteristics had no identifiable correlation with the variation of taxa. CCA analyses revealed that bacteria of the phyla *Lentisphaerae* and *Tenericutes* were positively correlated with depth. The presence of barium was significantly associated with changes on the x axis (CCA1).

3.3.3. Function predicted from the metagenomes of oil formation rock core and produced water samples

In addition to phylogenetic community structure, the functional capabilities of the identified microorganisms were also assessed, based upon bacterial taxonomy. A mean total abundance at a *genus level,* of the pooled four core and three PW samples, was calculated. The predicted functionality of the bacterial communities in produced water and rock core samples were assigned by the Tax4Fun package in R (Fig. 3.10). The effect of the community structure on gene functionality by assessing the likely function of the dominant bacterial genera (relative abundance > 0.5 %) from both sample types were also examined. The variances in the outcomes of these analyses between core and PW samples were evident (Fig. 3.9). The most

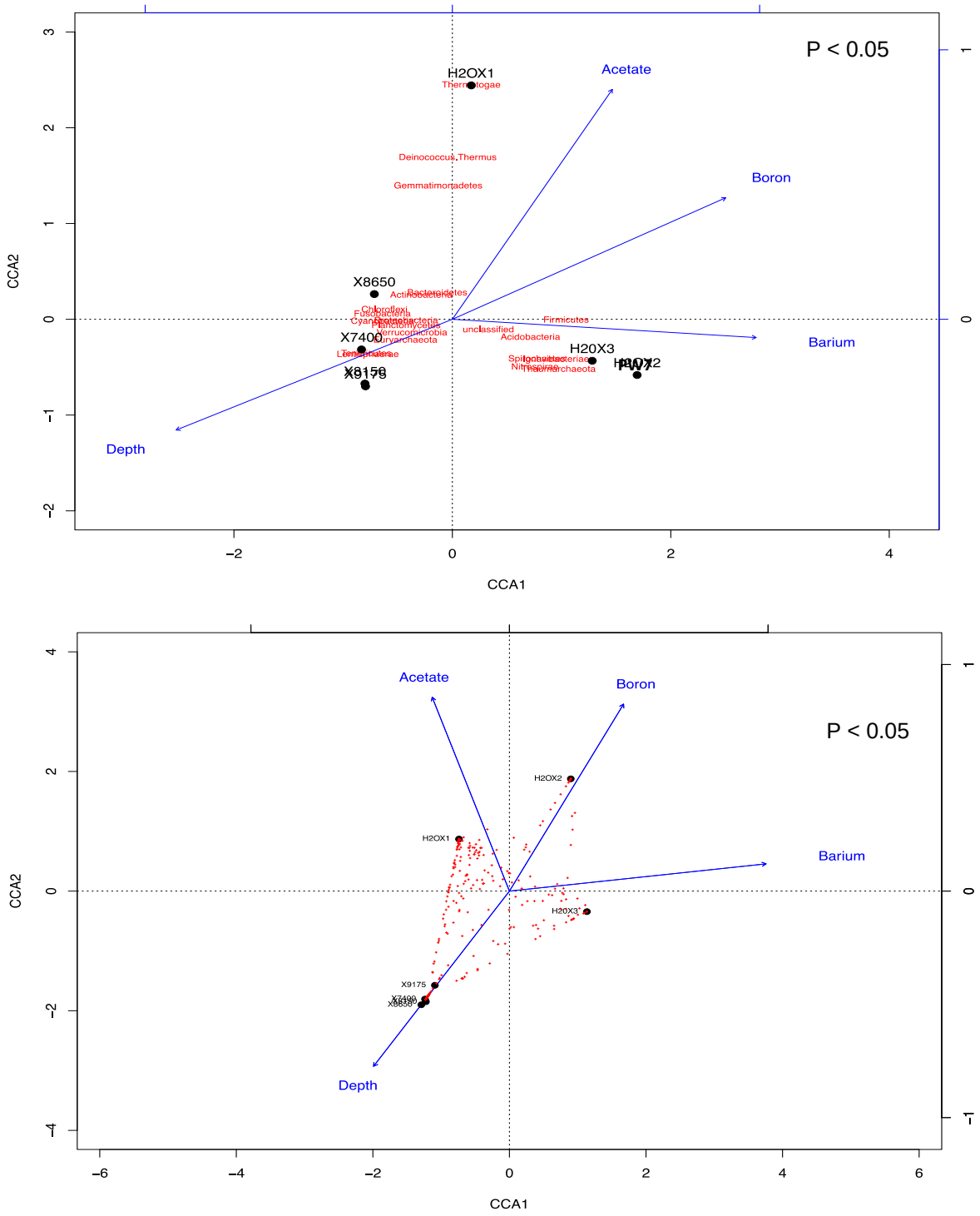


Figure 3.8 CCA Biplot of petrochemical and geographical features related to microorganisms present for PW and X samples. CCA shows only statistically significant effects of environmental features on samples. CCA conducted on BIONMeta output taxonomic data and ran on R ('Vegan' package). A) CCA plot comparing phyla classification of all 7 samples. B) CCA plot comparing genera classification. Red points represent individual groups of microorganisms. Black points represent sampling points. Blue lines signify explanatory CCA variable and directionality

abundant genera across all four X samples is that of *Xanthomonas,* accounting for ~78% of the total assigned reads. Yet for PW, members of the *Amphibacillus* predominate all three samples (PW0, 7 & 14), but at only ~29%. Members of the *Vibrio, Pseudomonas, Halomonas, Acinetobacter* and Arcobacter are the only other dominant mean genera in Core (>1%) but only constitute a further 10% of the total population. For PW samples (average of *n* = 3), although *Xanthamonas* is the second most abundant genus, it is present at four times lower abundance (19%) than in the core samples (77%). This data can distinguish the differences between the samples of a sessile and planktonic nature.

The assessment of gene functionality, predicted from taxonomic analysis, identified some possible differentiation between the orthologous genes in the Core and PW samples. The principal KEGG Orthology (KO) category was *Metabolism* (60.71-61.51%). However only three (carbohydrate, amino acid and energy metabolism) of the top five pathways were derived from the *Metabolism* category. The remaining two categories were both associated with *Environmental information processing* (membrane transport and signal transduction).

Of the 28 second- tier assigned KO pathways (Fig. 3.10), 14 were associated with statistically significant variation between Core and PW samples ($p < 0.05$). The most notable variations were increases in the PW samples in pathways associated with biosurfactant synthesis, membrane transportation and sugar metabolism. The predicted gene variation was distinctively

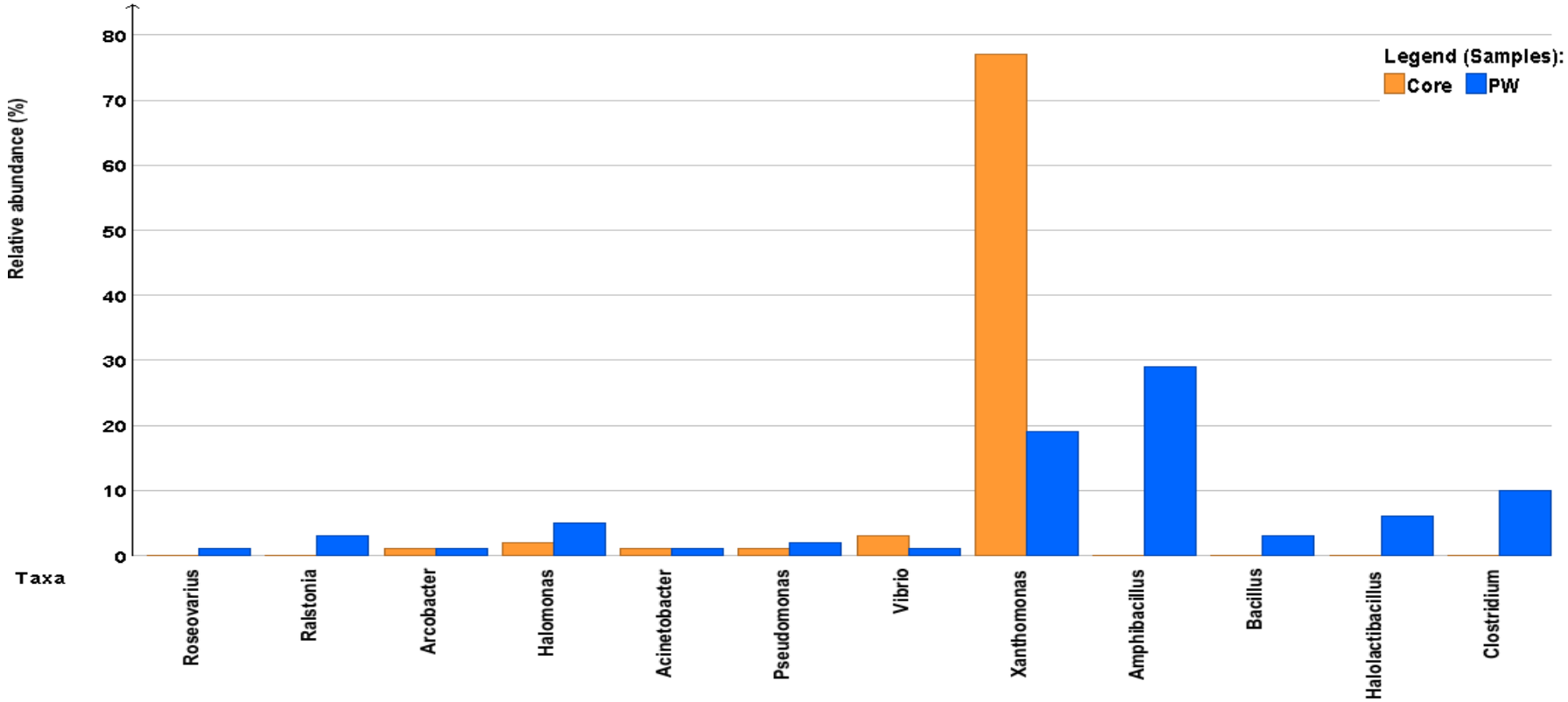


Figure 3.9 Bacterial distributions of most abundant genera from rock core and produced water samples. The relative abundance of all water (PW) and rock core (X) samples were averaged to create a total average abundance combined produced water & for rock core. Abundance of relative genera > 5% average abundance in rock and water samples are shown. Taxonomic analysis conducted using BIONMeta output from Section 3.3.2.

elevated in the core sample in pathways relating to *Metabolism of cofactors or vitamins and other amino acids*, *Replication and repair pathways* and *Glycan biosynthesis*. However, the pathway with the highest distribution of genes in rock core samples compared to water samples was that of *'Other'* (Diff = 0.61%, P <0.01). This ontological grouping though is an amalgamation of multiple genetic pathways that do not fit into specific bacterial categories, such as Human disease pathways (*Oncological, Neurodegenerative, Endocrine and Cardiovascular disease),* Auto immunity, other organismal systems (*Cardiovascular, Ageing, Development)* and *Drug development*.

The pathway with the greatest significant variation between Core (3.08%) and water samples (3.99%) was that of the *Metabolism of terpenoids and polyketides* (Diff = - 0.91%, p < 0.05). This grouping had a variation of twice that of the next nearest KO groups, *Carbohydrate metabolism* (Diff = - 0.42%) and *Membrane transport* (Diff = - 0.50%), in which water samples had the higher distribution of genes. Interestingly, the pathway of terpenoid and polyketide metabolism is directly linked to the formation of surface active compounds, such as lipopeptides and rhamnolipids, two key biological drivers of microbial enhanced oil recovery (Chapter 1.3). Therefore, the next step was to identify specific genes in these functional groups that were involved in these pathways (Fig. 3.11).

Within the class of *Terpenoids and polyketides metabolism*, both the highest abundance and greatest variation between samples was observed in the third-tier KEGG class of *Nonribosomal peptide structures* (Diff = -0.66%, p < 0.05).

The abundance of this class was twice that in the PW samples than in the core samples (Fig. 3.11). The KO the classes with the next most significant variation between the PW and Core samples were the *Sidophore group, Terpenoid backbone synthesis* and *Limonene and pinene degradation* pathways. All classes were identified at a higher abundance in PW, yet the differences between PW and X samples were not as great as for *Nonribosomal peptide structures*. No orthologous classes had a significantly higher average abundance in core samples than when compared to produced water samples.

3.3.3.1 Identification of reservoir taxa related to surfactant production

Using the BioSurf database (Oliveira *et al*., 2015), it was possible to create a phylogenetic tree based on bacterial species identified from the Bentley oilfield, which are likely to produce biosurfactants (Fig. 3.12.). BiosurfDB is a curated database consisting of a plethora of information on organisms, genomes, metabolic pathways and functional proteins associated with biodegradation or biosurfactant production. This is the only known database that specifically focuses on surface active compounds, particularly those isolated or used in oilfield systems. Comparison of BIONmeta and BiosurfDB outputs identified 80 species in the Core and PW samples that were capable to produce surface active compounds or carried genes encoding active domains that shared a high similarity with surfactant production genes. Of these taxa, only one, *Euryarcheota,* was identified. Of the remaining bacteria, 6 phyla were identified, with *Actinobacter*, *Proteobacteria* and *Firmicutes* dominanting. Within the phyla of *Proteobacteria,* the largest number of species identified were that of *Gammaproteobacteria*. From these identified species,

the majority of surfactants produced were not of an NRP class. Within the *Gammaproteobacteria*, 8 taxa could be classified, as opposed to only 3 in the *Firmicutes* and a single in the *Actinobacter*. The *Firmicutes* identified were all of the genus *Bacillus (B.subtilis, B.licheniformis & B. smithii).* However, the *Gammaproteobacteria* were more diversified, consisting of members of the *Vibrio, Serratia* and *Pseudomonas* genera.

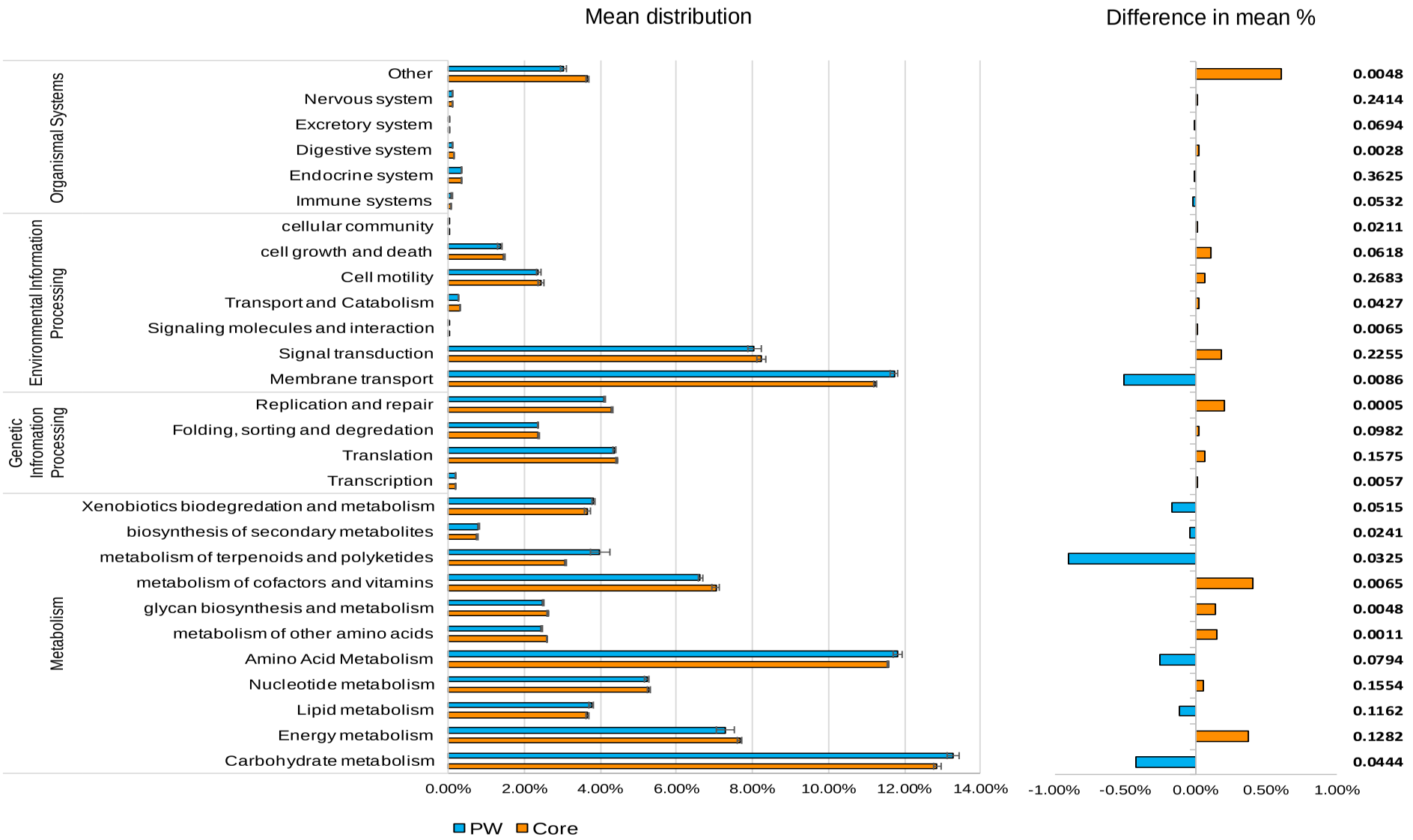


Figure 3.10 Mean distribution and difference of predicted functional genes based on taxonomy between production water (Blue) and drilled rock core (orange) samples. Averages taxanomic abundencies between rock core (*n* = 4) and production water (*n* = 3) samples calculated and functional taxonomy assessed based on mean abundance. Functionality calculated by Tax4Fun and categorized into relevant KO group. Total % distribution shown on left hand side, with % difference of means and p value (2-sample T-test) shown on the right side of the figure.

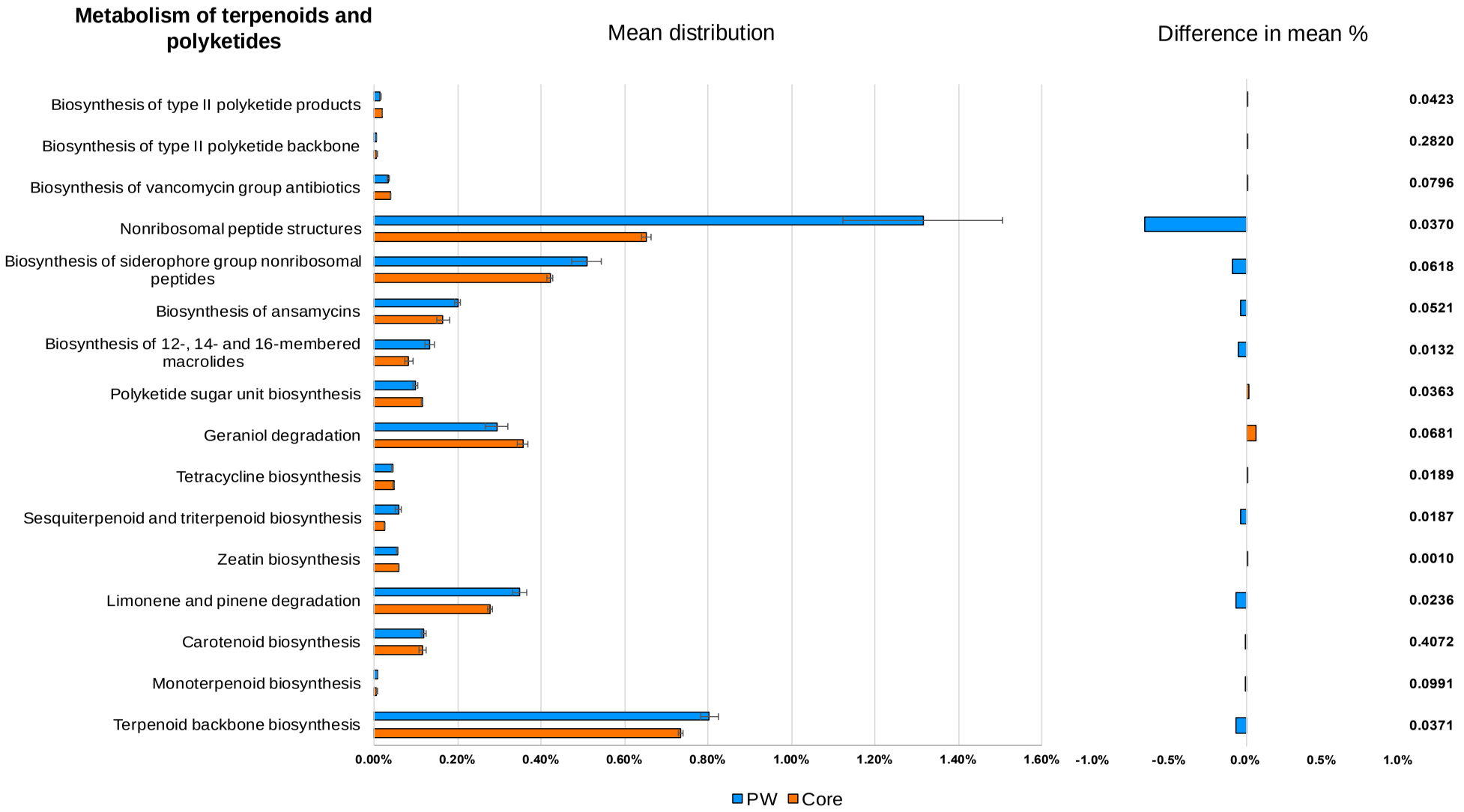


Figure 3.11 Mean distribution and difference of predicted functional genes for metabolism of terpenoids and polyketides between production water (Blue) and drilled rock core (orange) samples. Averages taxanomic abundencies between rock core (*n* = 4) and production water (*n* = 3) samples calculated and functional taxonomy assessed based on mean abundance. Functionality calculated by Tax4Fun and categorized into relevant KO group. Total % distribution shown on left hand side, with % difference of means and p value (2-sample T-test) shown on the right side of the figure.

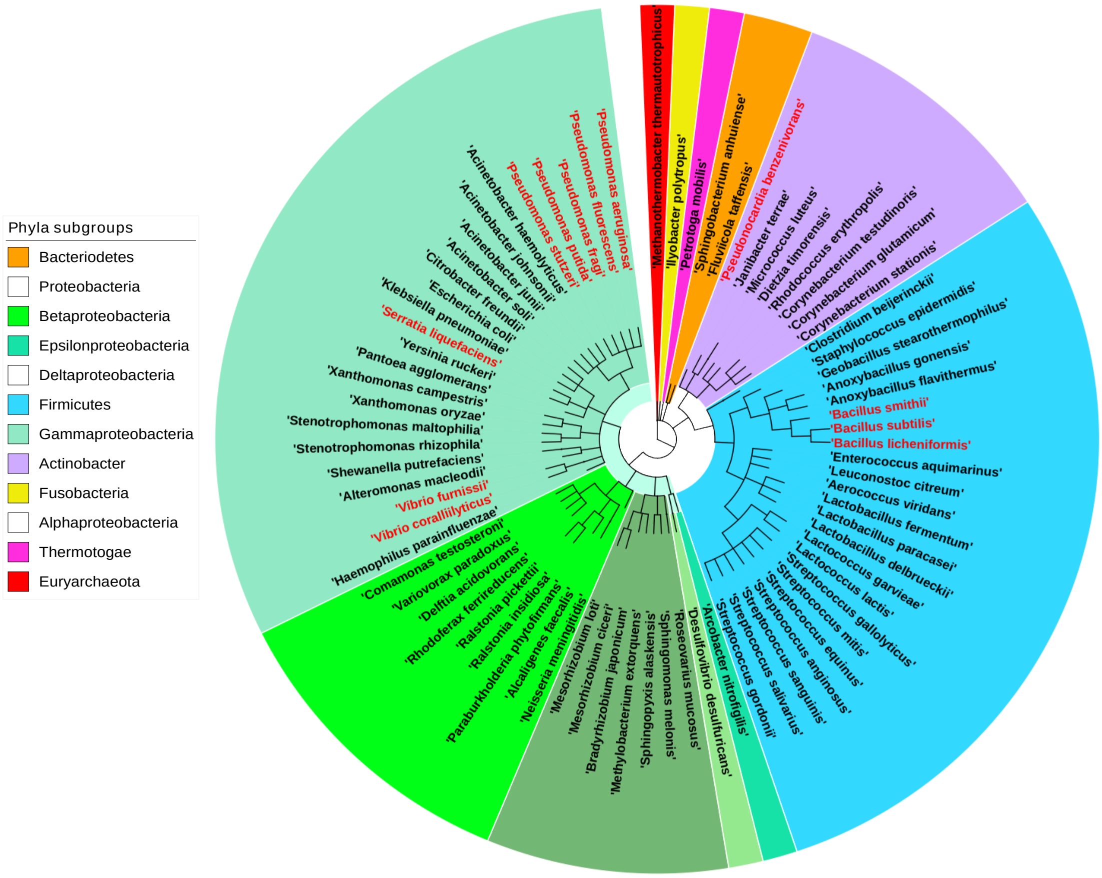


Figure 3.12 Phylogenetic tree of microorganisms identified from PW and Core samples that are known surfactant producers. Producers of SACs identified from functionality based on 16S rRNA taxonomy. Output from Tax4Fun analysed with BiosurfDB and NCBI to identify SAC producers. Red text distinguishes microorganisms that are producers of nonribosomal peptide related surfactants from those that are producers of any other SACs (Black text). Phyla and class subgroups are shown in the figure legend. Phylogenetic tree drawn using Interactive Tree of Life (iTOL, *https://itol.embl.de/*)

3.4 Discussion

Oilfields are complex environments with individual qualities that make them unique (Yang *et al.*, 2007). However, for microorganisms to thrive, certain parameters must be present for life to thrive. It is universally recognized that temperature, pH and salinity are key variables in promoting microbial growth (Michel Magot *et al*., 2000), however, nutrient availability is often overlooked. The chemical and petrophysical parameters of reservoir 9/3B-7Z were investigated in this study. At first glance it can be seen that the majority of ionic concentrations of this reservoir fall in the midpoint of chemical characteristics that are expected of most offshore reservoirs (Alley *et al.*, 2011; Fakhru'l-Razi *et al*., 2009). Despite this, the concentrations of barium, iron and sulphate are all at the lower end or below normal PW parameters. This is in keeping with the belief that when the majority of electron acceptors commonly found in surface waters are absent (Table 3.1), the primary metabolic processes undertaken by the microorganisms are sulphate reduction, methanogenesis and acetogenesis (Magot *et al*., 2000; Singh *et al*, 2014). This reduction of sulphate for energy development could provide a reasoning behind the sudden decrease in sulphate concentration after 24 hours.

From analyses of the ion concentrations of the production water samples, it is apparent that there are some significant variations between the chemical characteristics of the three samples. Primarily, the salinity of samples collected over the 14 day period are all relatively consistent. Furthermore, the reduction in sulphate and total iron concentrations suggest either the presence of low starting concentrations of these ions or the occurrence of sulphate reduction

before time point 0, which further decreases over the subsequent time points. The decreases in acetate and bicarbonate are likely attributed to the temperature changes of the reservoir during water flooding. A decrease in temperature correlates to decrease in microbial metabolic activity of the active community in the oilfield (Ijiri *et al.*, 2013) . The pH of the produced water remains relatively unchanged over the sample points. The pH of 7.38-7.5 ascertained for the PW samples is in the range of general pH of produced water according to the literature (7.0 - 8.8) (Al-Haleem *et al.*, 2010; Hanor, 1994), still in the optimal neutral range for microbial survival and production of extracellular metabolic products (Garrett *et al*., 2008). Reservoirs with a pH of below 7.0 are classified as acidic and usually have higher corrosion rates than alkaline reservoirs, thus providing major issues for the oil production company (Ogden, 2000).

The taxonomic analyses identified a diverse microbial community in reservoir 9/3B-7Z. A total of 31 phyla and 454 genera were identified from the seven samples. This finding differs to the current available literature and shows a higher number of individual taxa present than previously found. This is particularly true for unconventional and offshore reservoirs, where less than 50 individual bacterial OTUs have been identified from samples (Albokari *et al.*, 2015; Ren *et al*., 2011), with members of the *Bacillus, Betaproteobacteria* and *Gammaproteobacteria classes* predominating. The diversity between the two sample types from the Bentley oilfield, produced water and rock core, is evident. As expected from what we have seen in the literature (Giovannelli *et al*., 2016; Lavalleur *et al*., 2013), in samples associated with the water phase,

members of the *Firmicutes phyla (Clostridia* and *Bacilli*-related microorganisms*)* and *Alpha, Beta* and *Delta Proteobacteria* are the most abundant taxa. Similarly, members of the *Gammaproteobacteria* and *Actinobacteria* are found in far greater abundancies in the phase associated with rock/oil, particularly the genera of *Xanthomonas* of *Gammaproteobacteria* phyla. This is primarily because a large proportion of microorganisms are unable to tolerate the toxicity and hydrophobicity that is associated with the oil phase. The bacteria that do survive when in contact with crude oil, particularly members of the *Xanthomonas* in this reservoir*,* are able to degrade hydrocarbons and use them as a carbon source (Abbasian *et al*., 2016; Li *et al*., 2017; Wang *et al*., 2014). Although some members of the *Firmicutes* can survive and proliferate at the oil/rock interface via biofilm formation, this is not common due to reservoir characteristics in general (high pressures and wettability) and hence why they are present in higher abundancies in produced water samples than the rock core (Gittel *et al.*, 2009). Analysis of the two sampling types independently show that the differences between samples of the same origin are not strikingly obvious. Intra-sample variation within the rock and core groups is not as great as the variation between the groups, yet there are still some well-defined differentiations, particularly within members of the phyla *Proteobacteria, Actinobacteria* and *Firmicutes*. Within the rock core samples, depth of the sample was identified as having minimal impact upon the makeup of the microbial consortia. The two samples taken at 8150 ft and 9175 ft, were almost identical. However, the sample from the depth in between these two points (X8650) shows a lower abundance of *Proteobacteria,* yet higher concentrations of *Firmicutes* and *Actinobacteria.*

The population of X7400 was also more varied. At the taxonomic level of class, there is much deviation of microbiota between samples, with no visible association to depth. Although the metadata is related to total depths, the fact that the samples were taken from unilateral wells indicates that true vertical depth cannot be assumed. Therefore, differences in taxonomy could be due to reservoir heterogeneity by alterations in rock type through the cross section of the reservoir or alterations in conditions post drilling that allow the integration and cross-over of communities. Slight alterations in temperature, permeability or porosity may lead to titanic variations of the total microbiome and therefore makes the rock samples problematic to compare. The produced water samples show a more obvious variation between them (PW0, PW7 &PW14). As time from the commencement of sampling increases, members of the *Bacillus* genus also increases. However, bacteria related to the classes of *Gamma-, Delta-proteobacteria* and *Clostridia* show sequential fluctuation. This could primarily be due to the introduction of production waters and drilling related sediments which, either via the addition of new microorganisms or alteration of nutrients, can affect individual species growth and colonisation (Youssef *et al*., 2009).

The taxonomic differentiation between sites, or β-diversity, associated with the type of the sample from this reservoir was greater between rock and production water samples. This increase in bray-curtis metric indicates that there is an alteration in community composition between microbiomes that thrive off surface adherence, compared to free-floating and displaced planktonic microorganisms (Nyyssönen *et al*., 2014). No relationship was

identified between sampling depth and differentiation of microbial communities at a genus level, however *Lentisphaerae* and *Tenericutes* did positively correlate to the two lowest depth core sampling points. This however has not been previously observed in the literature. Conversely, within water samples, β-diversity was increased with sampling time. The latter two samples show similar communities compared to the initial sampling point. Investigations into the deep reservoir environment have shown that well development processes, such as flooding and drilling can distinctly affect microbial communities, particularly those microbial communities in produced water samples (Hirsch *et al.*, 1995; Zhang *et al*., 2012). This is believed to be caused by the fact that the creation of a reservoir, involving drilling processes and water flooding, can alter reservoir properties and introduce non-native nutrients and microorganisms into the environment, causing microbial contamination by exogenous microorganisms or invoking the growth low abundant species. Although this data shows the relative community composition within the Bentley Field, the lack of replicative data means that the results must be viewed with caution. Due to the cost of sequencing, it was not possible to undertake the next generation sequencing of three separate replicates for each sample type. Instead, three individual samples (of both rock core and produced water) were DNA extracted, pooled and amplified using the 16S barcoded primers before sequencing. It was hoped this would provide a more representative analysis, whilst not becoming too expensive for sequencing.

Alongside the assessment of the reservoir microbiome it is imperative to recognize the effect of the microbiota on oilfield recoveries, particularly in

relation to MEOR. However, despite significant improvements in the understanding of microorganism mediated oilfield processes, very little is known about microbial mechanisms that drive MEOR processes, particularly at a genomic level (Cai *et al*., 2015).

The more we learn about the processes of MEOR at a microbial level, the more we can make future predictions on functional enzymes and genes that may be vital to the oil recovery processes. This is further aided by the improvements in nucleotide sequencing and bioinformatics analysis. For example, it is now apparent that the process of surface mediated tension reduction is driven by surfactant production and hydrocarbon degradation is driven by hydroxylation. From investigation of these processes, the literature identifies that for certain microorganisms, these processes are mediated by specific gene groups. Surfactant production is mediated by surfactin related genes (SRF) in *Bacillus* and hydrocarbon degradation mediated by alkane hydroxylase genes (AlkB) (Liu *et al.*. Therefore, it is possible to categorize these metabolic processes via gene carriage and the other genes encoding subsequent pathways involved in surface tension alteration and hydrocarbon degradation. By utilising Taxonomy for Functionality (Tax4Fun), a programme that uses KEGG orthologous groups, this categorisation process can easily be achieved from taxonomic data, rather than the complicated and bioinformatically challenging procedure of complete metagenomic analyses (Aßhauer *et al*., 2015; Lindgreen *et al.*, 2016). It also provides the capability to compare differences in pathways from differing sample types. The highest abundance of genes was related

to pathways of *carbohydrate and amino acid metabolism* and *transportation and signal transduction*. Due to the lack of sugars and amino acids available in a reservoir it would be easy to assume that these pathways would not be carried by the indigenous bacteria. However, due to the limited available alternate nutrients (such as VFA's or carbohydrates) in this oil reservoir, it would seem that the microbes have evolved novel and effective systems to synthesise and transport nutrients into the cell and also utilise what minimal nutrients are available (Nie *et al*., 2016).

Significant differences in many pathways can be seen between produced water and rock core samples. These differences are primarily within pathways of metabolism (*e.g.* carbohydrate, amino acid and polyketide/terpenoid), replication and repair, transport (e.g. membrane, signalling) and glycan biosynthesis. Metabolism plays a vital role in both rock core and produced water sample types. Production waters primarily appear to contain more available nutrients and amino acids and this is potentially why we see an increase in the abundance of genes related to these in the water microbiome. However, to survive, if carbohydrates are limited, as mentioned previously, several bacterial genera are capable of degrading hydrocarbons to utilizable carbon form. Therefore, due to the presence of oil throughout the system, degradation will be continually happening across the field, in both the water, oil and rock strata. It is for this reason, primarily, that the difference in lipid metabolism is non-significant between both samples, as hydroxylation related genes are present in many adapted bacteria to aid in the breakdown of lipid substances for energy (Cao *et al.*, 2009). Interestingly, the orthologous

category with the largest difference between water and rock samples was that of metabolism of terpenoids and polyketdies. A few categories within this group have been associated with MEOR, particularly in relation to biosurfactant production and hydrocarbon degradation. Of the various biosurfactant compounds, rhamnolipids and lipopeptides are the most well characterized and their production has been associated with the greatest reductions in IFT and increases in recovery of residual oil (Maudgalya *et al*., 2005; McInerney *et al*., 1990). Lipopeptide based surfactants are primarily produced as a consequence of non-ribosomal peptide biosynthesis, catalyzed by a large multi-enzymatic complex consisting of four subunits, which make up the operon known as surfactin synthetase *(SrfA*) (Roongsawang *et al.*, 2011). Consequently the KEGG classifications of non-ribosomal peptides and their biosynthesis was of particular interest to this project. There is a significantly higher distribution of structural genes associated with the synthesis of non-ribosomal peptides in the produced water samples. These NRP surfactant producing bacteria have been previously shown to be more likely to produce surfactants when growing in the nutrient rich production waters or when bound to immiscible substrates, such as hydrocarbons, from where they are removed from the reservoir during water flooding. This may be why microorganisms with NRP production capabilities are less common within samples from the rock or sedimentary cores (Desai & Banat, 1997; Kurata *et al*., 2016).

However, if we specifically look at the number of non-ribosomal peptide producing microorganisms in relation to known surfactant producers, we

reveal a different image. Of a total of 80 known biosurfactant producers, only 12 where identified from the amplicon dataset to be of a non-ribosomal nature (Martínez-Núñez & López, 2016). These non-ribosomally synthesized biosurfactant producers include strains in the *Vibrio, Bacillus* and *Pseudomonas* genera. These three genera have all been previously utilized in microbial enhanced oil recovery processes and have shown to positively modify IFT (Banat, 1995; Eskandari *et al*., 2009; Hu *et al.*, 2015). Furthermore, between the two sample types, these genera of bacteria where all present in moderate abundances. Organisms of the genera *Vibrio (*X: 0.6-5.5%, PW: 0.25-2.3%) accounted for ~ 2% and *Bacillus* (X: 0.8-1.5%, PW: 0.85-5.7%) and *Pseudomonas (*X: 1.1-2.8, PW: 0.6-3.3%) accounted for ~ 1.8% of the total reservoir microbiome. Therefore, bacteria of these genera would be suitable target organisms for isolation and culturing, if bio augmentative MEOR was to be undertaken using an indigenous strain or if biostimulation of autochthonous microbes of the reservoir community was to be undertaken.

The identification of these microorganisms from the Bentley Oilfield provide a starting point for the potential identification of MEOR related isolates from this environment. However, due to the issues surrounding PCR bias (Section 1.5.3 & Appendix A), the data presented here cannot be guaranteed to be an accurate representation of the exact abundancies of this oilfield community. A proportion of microorganisms may have been under or over represented, incorrectly categorized or missed completely from the analysis. To overcome these issues, alternative sequencing techniques, such as community metagenomics or the direct sequencing of SSU rRNA, could instead be utilized

to provide a more accurate representation of the microbial community. Furthermore, with the decrease in sequencing costs and continual improvements in computational power, bioinformatics and sample preparation, these afore mentioned methods will become more mainstream techniques within microbial ecology.

Chapter 4

Isolation and Identification of Biosurfactant-producing Bacteria from Oilfield Environments

4.1 Introduction

4.1.1 Background

Advancements in the identification and ability to cultivate microbial communities present in extreme environments over the last two decades has allowed us to better understand the role that these microbes play within their ecosystem (Fakruddin & Mannan, 2013; Rampelotto, 2010; Vester *et al.*, 2015). Even though the existence of microbes in hostile environments has been postulated since the early 1900's (Bastin *et al.*, 1926; Lortet, 1892), the lack of physical proof within the literature had cast doubts over the validity of these claims (Magot, 2005). It is only in recent years that conclusive physical evidence has been provided regarding the cultivation of microbes from these adverse habitats (Gaytán *et al*., 2015; Pulschen *et al*., 2017). This evidence, combined with the microbiological similarities found across similar environments worldwide, further substantiates that some microorganisms are capable of surviving extreme physiological conditions, such as temperature, pressure and salinity.

Traditionally, the microbial life inhabiting environmental ecosystems has been an important factor for indicating the differences in environmental variables, for example ion concentration and gas production (Moberly,*et al., 2016; Ramos-Padrón et al., 2011)*. Yet more recently, a number of indigenous

microorganisms have proved to be useful for biotechnological purposes. Rather than the use of conventional metabolic products and enzymes that are produced by mesophilic microbes, for processes ranging from food production to medicine, metabolic products produced by extremophilic organisms are already suited to the inimical conditions (e.g. temperature, pressure) present in many industrial processes (Coker, 2016). This is especially true for industrial processes related to oil, such as MEOR (Gao *et al*., 2016; Hakima & Singleton, 2017; Zhang, *et al*., 2012). As discussed previously, oilfields are diverse environments, with much variation seen in petrophysical and geochemical conditions across the globe. Organisms that are suited to these downhole geochemical/ petrophysical conditions have a better chance of survival and being metabolically active, thus stand a good chance of improving incremental oil production. The failures of MEOR, particularly in the field, have centred around the use of microorganisms that were unable to endure extreme conditions, particularly temperature and pressure (Sen, 2008a; Sheehy, 1991a). To overcome this, it has been seen as an important requirement to use indigenous species, through either isolation or biostimulation (Chapter 1.2), that will have improved prospects of survival in the reservoir (Castorena-Cortés *et al.*, 2012; Rabiei *et al.*, 2013).

4.1.2 Indigenous microbes for MEOR treatment

Both biostimulation and *ex-situ* fermentation of indigenous isolates are commonly used MEOR strategies (Chapter 1.2). However, as pointed out by Lazar *et al.* (2007), nutrient injection can lead to a loss of injectivity, or flow from the injection side, due to increased microbial loads and biofilm production

leading to well bore plugging at the injector. Furthermore, as with all nutrient injection, targeted growth of MEOR microorganisms is difficult to achieve, and the proliferation of undesirable microorganisms, such as sulphate reducing bacteria (SRB), biofilm formers and acid producers (Kip & van Veen, 2015) are just as likely to respond to nutrient injection(i.e. nitrate injection). In some scenarios, this could lead to microbial induced corrosion (MIC) and reservoir souring (Gray *et al.*, 2008). Both of these phenomena are commercially detrimental and economically costly issues, making nutrient injection an unnecessary risk that most producers will not take. With these potential complications in mind, and a high risk of MIC due to the vast numbers of *Pseudomonas sp.* and SRB related *Deltaproteobacteria* (Abdolahi *et al.*, 2014; Zarasvand & Rai, 2016; Li *et al*., 2016) identified from the taxonomic data (Fig 3.5, Chapter 3.3.2), it was decided that isolation and enrichment of indigenous microorganisms was the appropriate MEOR approach to use for the Bentley field.

4.1.3 Cultivation of organisms by exploitation of MEOR characteristics

The ability to degrade oil is fundamental for bacterial survival in oil reservoirs as oil is an abundant C source in an otherwise C depleted environment. A number of oilfield bacteria, such as *Petrotoga* and representatives of the *Pseudomonas* and *Acinetobacter* genera, are able to degrade hydrocarbons and long chain alkenes (Kostka *et al*., 2011; Purwasena *et al.*, 2009). Not only is this useful from an MEOR perspective, but it is also proposed that this process plays an important role in the freeing of fixed carbon and VFA's for use by other microorganisms. This complex nutrient web is one factor that

drives and regulates the oilfield microbiome. To successfully achieve the cultivation of a range of oilfield species, the nutritional network should be replicated in the laboratory, primarily by enrichment with oil from the selected field.

Of the cultivable organisms found in oilfield environments, a high proportion have phylogenetic and metabolic traits, alongside hydrocarbon degradation, that make them suitable for MEOR treatment strategies (Safdel *et al.*, 2017). One of these traits is the production of biosurfactants/bioemulsifiers (Liu *et al.*, 2013; Youssef *et al*., 2004a). As discussed previously (Chapter 1), these surface active compounds (SACs) are important for a number of processes such as motility, biofilm formation, host colonisation and nutrient release, all of which result in the alteration of the interfacial properties of oil. In environments where carbon availability is reduced or entrapped in insoluble sources, biosurfactants are capable of emulsification or micellurisation and orchestrating adhesion/deadhesion between the oil and cells, in order to access the scarce nutrient (Sen, 2010). However, adhesion is only of interest in cases of cell-bound, surface-active compounds, rather than secreted biosurfactants. The emulsification and micellurisation are of the most significance for MEOR. Increasing the surface area and availability of the insoluble oil makes it more susceptible to degradation by biodegrative enzymes produced by the biosurfactant strain and other microbes, due to increased microbial access (Parthipan *et al*., 2017; Wijesekara *et al.*, 2017). For this reason, the enrichment of biosurfactant producers has been improved by the use of hydrophobic substrates, such as crude oil or polyaromatic

hydrocarbons (PAHs), as a sole carbon source (Patowary *et al.*, 2017; Rahman *et al.*, 2002). Thus, biosurfactant production and hydrocarbon degradation are often intertwined.

4.1.4 Research hypothesis, aims and objectives

The aim of the work reported in this chapter was to successfully isolate microorganisms from oilfield rock core and production waters via the use of enrichment cultures. The data within this chapter tested the hypothesis that a range of microorganisms, including previously identified surfactant producing microorganisms, could be successfully isolated from an oilfield environment and their interfacial tension properties assessed for MEOR purposes. The surfactant production and biofilm forming potential of all isolates was then investigated; with all strains that possessed measurable traits that might be useful to MEOR being taxonomically characterised.

4.2 Methods

4.2.1 Source of bacterial isolates

Strains were isolated from the unfiltered produced water and core samples described in Chapter 3.2.1, taken from the Bentley Oil Field, North Sea. These samples were used to initiate enrichment cultures following the modified method of McInerney *et al*. (1983; 2007) , from which individual colonies could be picked and used for screening.

4.2.2 Isolation and culture based identification of bacterial isolates from produced water and drilled core samples

To isolate bacteria from the collected formation water and production core samples, 5% (w/v) of each individual unfiltered PW sample (25 mL) and core sample (25g) was added to Erlenmeyer flasks containing 475 mL of modified mineral salts media, supplemented with 1% (w/v) Bentley Oilfield Heavy crude oil. Flasks were incubated anaerobically using an anaerobic container and sachets, or anaerobic cabinet at 40°C and 220 rpm for a total of 24 h. After 24 h, the OD_{600} of each culture was measured and diluted down to an OD_{600} of 1.0 absorbance unit. Culture (1 mL) was then added to 99 mL of each of the respective media (starting OD_{600} of ~ 0.01) and incubated for 7 days. Each media was assessed in triplicate. The media's used were; NB, MB, MS, XFW, 0.5% OFM, 2% OFM, 4% OFM, 6% OFM (See Chapter 2.2). At selected time points of 24 h, 48 h and 168 h, 1 mL of culture was taken and serial diluted from 10^{-4} to 10^{-7}. This was followed by the spreading of diluted culture (100 μL) onto LB agar plates and incubation at 37°C overnight. Colonies with different morphologies were picked and streaked separately onto LB agar and

incubated again to obtain pure isolates. At these three time points, the OD_{600} of the culture (100 μL) was also measured spectrophotometrically to assess bacterial growth rates and variation of growth between cultures. Growth of the cultures in NB, MB, MS and XFW was assessed, and a sterile vial of each respective media was used as the spectrophotometric blank. However, due to the optical density of the OFM media, any cellular growth had to be assessed after centrifugation (10 mins at 12000 x g) and resuspension of the cells in NB. Following 2 rounds of streak plate purification, a single colony was picked, added to the appropriate growth media (500 μL) and stored in 40% (v/v) glycerol solution at -80°C.

4.2.3 Screening of isolates for biosurfactant producing potential

4.2.3.1 Hemolysis assay

A hemolysis assay was the first screening test for biosurfactant production potential of bacterial isolates. Due to the erythrocyte-lysing capabilities that surfactants have, biosurfactant producing strains are capable of causing clearing zones, or halos, around bacterial colonies on blood agar. A single colony of each isolate was inoculated into 20 mL NB media and incubated at 40°C with shaking at 220 rpm for 24 h. Blood agar plates contained 40 g L^{-1} Blood Agar Base (Sigma Aldrich, US) was supplemented with 5% (w/v) sheep blood. Individual colonies of each isolate were streaked onto the blood agar plates and incubated for 48 h at 37°C (Mulligan, Cooper, & Neufeld, 1984). The clearing zone post incubation was then measured and assessed for hemolytic activity, based on the halo diameter: (-) growth with no halo formation or partial hemolysis, (+) complete hemolysis with diameter of < 1

cm, (++) complete hemolysis with diameter of > 1 cm but < 3 cm, (+++) hemolysis with a diameter of > 3 cm.

4.2.3.2 Oil spreading assay

The oil spreading assay accompanied the hemolysis screening assay to confirm biosurfactant production. To a sterile petri dish (9 cm diameter), distilled water (50 mL) was added followed by a 20 μL drop of light oil (West Firsby, UK, API 38°) to the surface of the water. Before dispersion of the oil occurred, culture (10 μL) was added to the surface and the resulting clearing zone measured (Youssef *et al*., 2004b): (-) no clearing zone, (+) clearing zone < 1 cm, (++) clearing zone >1 cm but <3 cm, (+++) clearing zone > 3 cm.

4.2.4 Screening of isolates for biofilm producing potential

To screen the bacterial isolates for biofilm production, the retention of crystal violet by biofilms or extracellular polysaccharide production due to the bacterial growth on a plastic surface was measured method based on that described by O'Toole & Kolter (1998). Briefly, single colonies of each isolate were used to inoculate NB broth and cultured overnight at 37°C at 220 RPM. The next day the culture was diluted 1/100 with fresh NB into sterile cuvettes. Cuvettes were then sealed and statically incubated at 37°C for 48 h. Biomass of the resulting biofilm was determined by adding 200 µL of 0.1%(w/v) crystal violet to each cuvette. The cuvettes were incubated at room temperature for 30 min followed by a rinse with distilled water to remove any cells not attached to the wall of the cuvette and excess dye. Ethanol (70% (v/v), 300 µL) was added to each well to de-stain the crystal violet stained biofilm and resuspend

the attached cells. The destained biofilm was incubated at room temperature overnight. The cuvettes were then read at an absorbance of 570 nm to determine the total abundance of biofilm or biofilm like extracellular material produced.

4.2.5 Identification of bacteria by morphological and physiological characteristics

4.2.5.1 Colony morphology

To assess colony morphology, an overnight culture grown in NB, derived from a single colony, was serially diluted (10^{-3},10^{-4},10^{-5},10^{-6}) and spread (100 mL) onto LB agar plates.

4.2.5.2 Gram stain

Gram stains were performed as per standard microbiological protocol (Bartholomew & Mittwer, 1952) on overnight LB broth cultures of each isolate to aid with bacterial identification.

4.2.5.3 Biochemical testing

To give indications on the genera of bacteria that were isolated, catalase and oxidase biochemical tests were performed on chosen surfactant producing isolates identified from the screening assays. For oxidase testing, 1 mL of an overnight bacterial culture in NB media was added to a single Oxidase strip (Sigma-Aldrich, US). A blue colour after 1-5 minutes indicated the presence of the cytochrome C enzyme. The test for catalase activity was based on the ability of the bacterial isolates to breakdown H_2O_2 into O_2 and H_2O. A single loop of the same culture used for the oxidase tests was added to a glass slide.

A single drop of 3% (v/v) H_2O_2 was added to the slide, and the production of vigorous bubbling indicated the production of catalase.

4.2.5.4 Anaerobic growth

To examine growth of the isolates in the complete absence of oxygen, each isolate was streaked onto LB agar and 4% OFM agar and incubated in an anaerobic container for a maximum of 48 h at 40°C.

4.2.6 Identification of bacteria by PCR analysis of 16S rRNA gene

Each of the chosen isolates was grown at 37°C with shaking at 220 rpm for 12 h and underwent genomic DNA (gDNA) extraction using a modified Griffiths *et al.* (2000) method (Chapter 2.9) that also removed contaminating RNA from the extract via RNAse A digestion. PCR amplification was then conducted on the DNA using V4 16S primers (Caporaso *et al*., 2012), using Phusion high fidelity enzyme (NEB, USA). A protocol using an initial denaturation of 94°C (2 min) was used before 30x cycles of the following conditions: 94°C denaturation (10 sec), 50°C annealing (20 sec) and 72°C elongation (20 sec); followed by a final elongation of 72°C (5 min), using Mastercycler Pro thermocycler (Eppendorf, UK). Amplicons were loaded onto a 1% TAE agarose gel (15 ul) and bands viewed, excised and purified as specified in the General Methods (Chapter 2.10). Post cleaning, amplicons were sent to SBS Sanger Sequencing Service (Source Bioscience, UK) for sequencing. Database searches for amplicon identification were performed using the Basic Local Alignment Search Tool against the non-redundant database for 16S rRNA gene sequences (Altschul, S.F. *et al,* 1990).

4.3 Results

4.3.1 Enrichment and isolation of bacterial isolates from produced water and rock core samples of the Bentley Oilfield

For all of the media types used for enrichment, bacterial growth via an increase in OD was observed for both the PW and RC samples after 24 h (Fig. 4.1). This increase was particularly apparent in the media, supplemented with sugar cane molasses (0.5-6% w/v), which resulted in OD_{600} >1.0 after 24 h. Furthermore, in the enrichment cultures where OFM was used, the OD_{600} increased continuously from 0-48 h, with some exceptions; all concentrations of OFM enrichments for X8650 and 6% OFM for X8150 and X1975, where no significant changes were seen. Growth of the enriched bacteria in XFW and MS media both showed the smallest increase in OD over the 168 h period (OD_{600} < 0.5). Enrichment cultures in MB and NB media predominantly showed minute increases (average OD_{600} < 0.2) to 48 h. The differences in error bars show some variation within and between samples (0.005-0.6 SD), particularly where a large spike in OD value is seen (Fig. 4.1). However, the majority of standard deviations were below 0.1 SD.

From the dilution plating of the 7 different oilfield samples, the media combinations at each of the three respective time points (24h, 48h & 168h) produced a total of 309 individual isolates. Of these, 160 were isolated from RC samples (X7400:33, X8150:25, X8650:64, X9175:37) and 149 from the PW samples (H20-X1:67, H20-X2:51, H20-X3:31). With a sizeable isolate

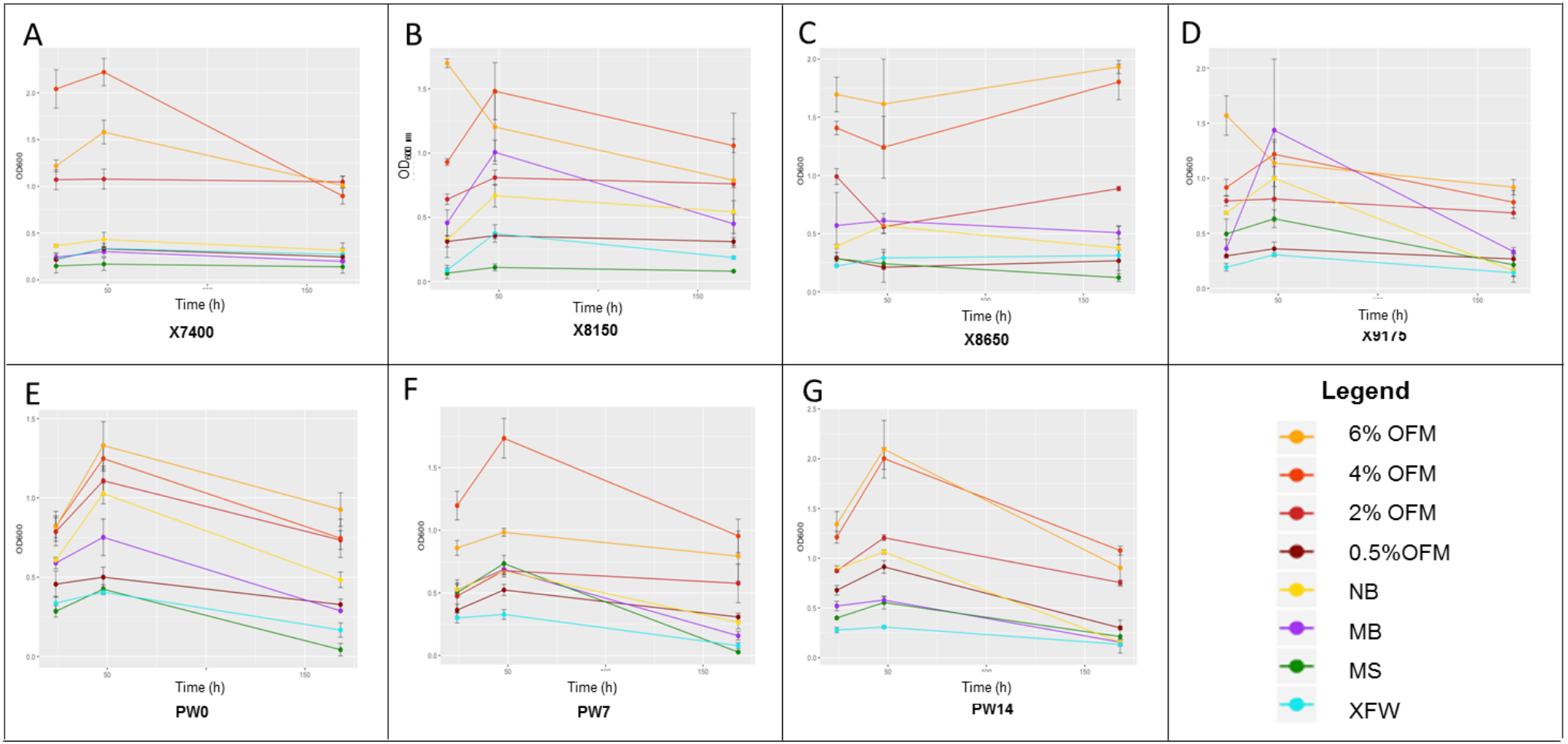


Figure 4.1 Variation in optical density of enrichment cultures for seven oilfield environments using differing media components over a 168 h period. OD_{600} shown during growth of microbial oilfield enrichments, with measurements taken at 24h, 48h @ 168h. Panels A-D represent enrichment cultures of rock core (x) samples of labelled depth. A) Enrichment of X7400 B) Enrichment of X81500 C) Enrichment of X8650 D) Enrichment of X9175. Panels E-G represent produced water (PW) samples taken at 7 day intervals (0, 7 & 14 days).Figure legend shown on left hands side relates to the growth nutrient and relative concentrations (if signified): Oilfield medium (OFM), Nutrient broth, Marine broth, Minimal salt media (MS) Bentley formation water (XFW). E) Enrichment of PW0 F) Enrichment of PW7 G) Enrichment of PW14. Mean OD_{600} and standard deviation shown for each time point (n=3).

library available for screening, it was now possible to identify isolates with desired traits: biosurfactant production and biofilm formation.

4.3.2 Screening of strain library for potential biosurfactant and biofilm producing isolates.

All 309 isolates underwent a rigorous screening process as described in Chapter 4.2 to determine their ability to produce biosurfactants. This process involved two assays: 1) the measurement of the diameter of oil spread and 2) bacterial colony hemolysis of blood agar, both assays where scored on a 4-point scale (-, +, ++, +++). The total number of '+' from the two assays, were then calculated for each isolate and scored out of a total of 6. Of the total 160 RC isolates, only three isolates recorded a total of 6 on the surfactant scoring chart; C10, C19 & C36 (figure 4.2). A further seven isolates scored 5; C13, C15, C16, C20, C82, C98 & C108. No PW isolates recorded a score of above 4 on the surfactant scoring chart.

Each of the isolates was also assessed for biofilm production. After the reading of OD_{570}, isolates were scored, with values ranging from ~ 0.0 – 1.5 . The highest scoring isolates were again predominantly from the rock core samples, with only one of the top seven biofilm producers coming from the PW samples (Fig. 4.3). However, the majority of the high scoring biofilm producers exhibited a low surfactant score (< 4) and so were not used for further analysis. The isolate C75, which was the best biofilm producer, was chosen for further analysis, despite only recording a surfactant production score of 1. C75 was

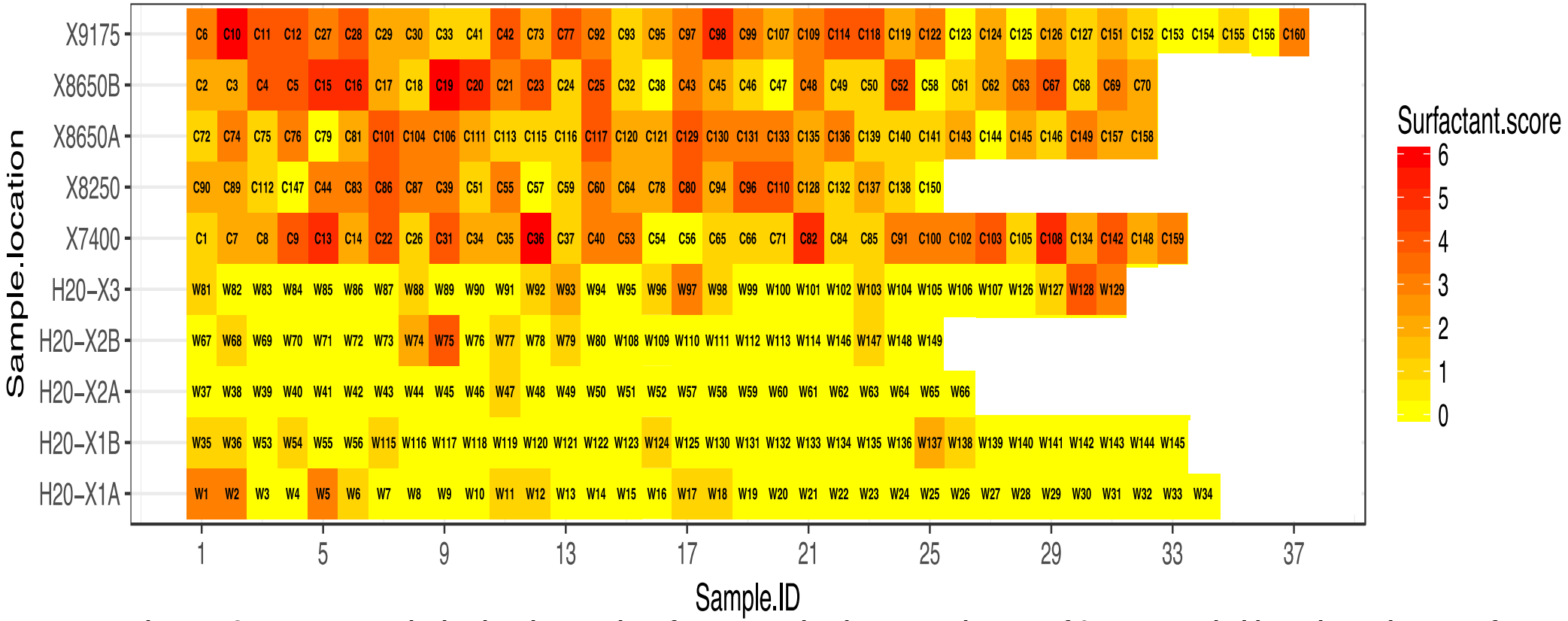


Figure 4.2 Heat map analysis showing total surfactant production. A total score of 6 was awarded based on: a) score of 0-3 in oil spreading diameter potential with: (-) no clearing zone, (+) clearing zone < 1cm, (++) clearing zone >1cm but <3cm, (+++) clearing zone > 3cm. b) score of 0-3 in haemolytic clearing diameter assay with: (-) growth with no halo formation or partial haemolysis, (+) complete haemolysis with diameter of < 1cm, (++) complete haemolysis with diameter of > 1cm but < 3cm, (+++) haemolysis with a diameter of > 3cm. Higher surfactant score represented by darker colour.

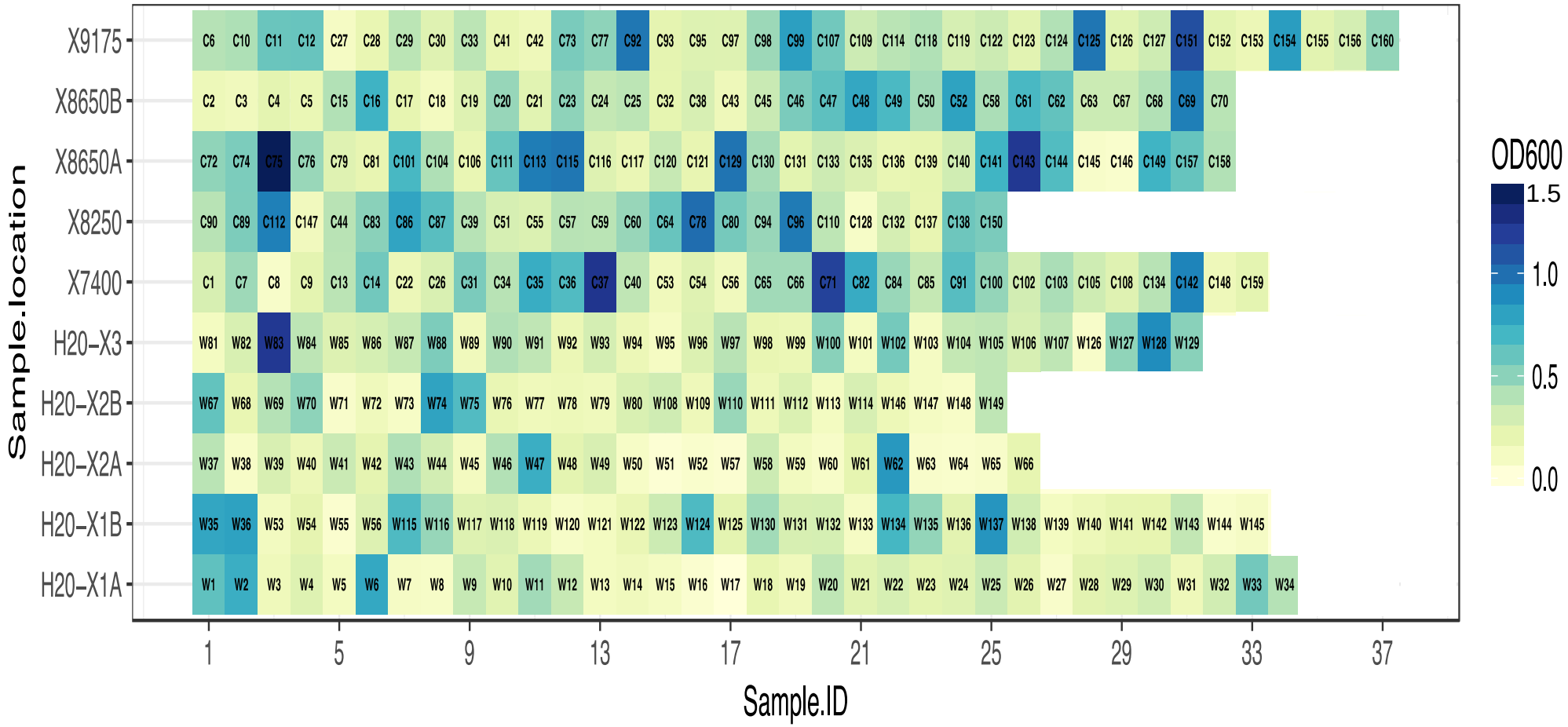


Figure 4.3 Heat map analysis showing total biofilm production. Biofilm production assessed at OD_{570} nm absorbance after overnight incubation. Each row represents sample location was isolated from. Higher OD represented by darker colours.

chosen to be used as a non-surfactant producing control, to assess the effects of biofilm formation on MEOR in comparison to surfactant production .

A total of eleven isolates isolates (C10, C13, C15, C16, C19, C20, C36, C75, C82, C98 & C108) were chosen for identification and molecular and biological characterization. Each isolate underwent biochemical testing, morphological colony assessment and Gram staining, alongside 16S rRNA amplicon sequencing analysis of the V4 region. Of these eleven isolates, only C15 and C98 were unable to proliferate in OFM media. Therefore, nine of the eleven isolates were characterized based on morphology, physiology and molecular identification methods.

4.3.3 Identification and characterization of isolates

All isolates meeting the criteria for further characterization were associated with RC samples, with four isolates from sample X7400, four from sample X8650 and one from sample X9175. Initial colonial morphological characteristics were identified from the growth on LB agar after growth from NB, MS and 4%OFM. All isolates were also assessed using biochemical methods for physical traits, such as catalase and oxidase production using the testing kits (Sigma-Aldrich, UK) and cellular structures observed following Gram staining (Table 4.1). Due to the nature of the oilfield environment that the bacterial strains were isolated from and its lack of oxygen within the system, it was crucial that all isolates were able to grow in the absence of oxygen. The anaerobic growth profile is given in Table 4.1.

4.3.3.1 Taxonomic assignment via 16S rRNA sequencing and phylogenetic analysis of unknown bacterial species

Alongside the physiological characterization of the nine isolates (C10, C13, C16, C19, C20, C36, C75, C82 & C108) each strain was subjected to gDNA extraction and PCR amplification of the 16S rRNA gene to enable genus and species assignment. The amplification was confirmed by agarose gel electrophoresis (Fig. 4.4). Of the nine isolates, two were identified as having a > 99.9% similarity to that of *Bacillus licheniformis* (100%), two of *Bacillus subtilis* (100%), two *Pseudomonas aeruginosa* (99%), one *Escherichia coli* (99%), and two *Acinetobacter sp.* (99%) (Table 4.1).

Table 4. 1 Identified bacterial isolates with potential for use in Microbial Enhanced Oil Recovery

Isolate strain name		C10	C13	C16	C19	C20	C36	C75	C82	C108
Isolate location		X9175	X7400	X8650	X8650	X8650	X7400	X8650	X7400	X7400
Colony morphology		Irregular, large, opaque, dull	Irregular, large, opaque, dull	Round, medium, wavy, light blue	Irregular, large, opaque, dull	Round, small, opaque, yellow	Irregular, large, opaque, dull	Round, small, cream, convex	Round, medium, wavy, light blue	Round, small, opaque, white
Growth Media	NB	+ [a]	+	+	+	+	+	+	+	+
	MS	+	+	+	+	+	+	+	+	+
	4%OFM	+	+	+	+	+	+	+	+	+
Catalase		+	+	+	+	+	+	+	+	+
Oxidase		+	+	+	+	-	+	-	+	-
Gram stain		+	+	- [b]	+	-	+	-	-	-
Cell shape		rod	rod	rod	rod	rod	rod	rod	rod	rod
Anaerobic growth		+	+	-	+	+	+	+	+	+
Hemolytic activity		+++	+++	+++	+++	++	+++	+	+++	+++
Oil spreading activity		+++	++	++	+++	+++	+++	-	++	++
Biofilm Production		0.4136	0.3753	0.7207	0.243	0.4722	0.6671	1.412	0.7566	0.2383
Bacterial species (% similarity)		*Bacillus licheniformis* (100%) [c]	*Bacillus subtilis* (100%)	*Pseudomonas aeruginosa* (99%)	*Bacillus subtilis* (100%)	*Acinetobacter sp.* (100%)	*Bacillus licheniformis* (99%)	*Escherichia coli (99%)*	*Pseudomonas aeruginosa* (99%)	*Acinetobacter junii* (99%)

a) (+) indicates positive result or growth of individual colonies
b) (–) indicates negative result or no growth
c) similarity scores based on NCBI blastn alignment search (available at https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch)

4.4 Discussion

A total of 309 isolates were identified from samples taken from the produced water and rock core of the Bentley Oilfield, UK. Isolates were identified from enrichment cultures that were initially supplemented with 1% (w/v) Bentley crude oil as a sole carbon source in the MS media. This was an important feature of the enrichments as the majority of oilfield bacteria only have oil constituents available as a carbon source in the extreme environment of an oil field, (Udgire *et al.*, 2015). This is apparent in the Bentley field, where acetate, bicarbonate and formate are present in very low/ negligible concentrations by the third sampling time point after water flooding of the reservoir. Therefore, the oil field isolates must utilize the oil or its constituents for growth. Furthermore, a number of biosurfactant producers are able to degrade hydrocarbons, also. By providing the bacterial enrichment cultures with crude oil, some isolates are able to augment other strains by breaking down or emulsifying the long hydrocarbon chains and providing a readily available source of energy for themselves and other commensal microorganisms to utilize. Of the 309 isolates, 10 were determined to have a convincing potential for biosurfactant production based upon the stringent screening system from hemolysis and oil spreading assays, that has previously been reported in the literature (Abdelhafiz *et al.*, 2017; Youssef *et al*., 2004a) . It was determined that only isolates that scored above 5 would be considered as strong candidates for biosurfactant mediated MEOR. This is primarily due to the fact that any scoring lower than this cut off, must not have performed strongly in one or both of the two tests. Preferentially, only isolates scoring 6 would be

used, but due to the low number of isolates scoring this, it was decided to also include any that scored one level below maximum on the scoring system.

The hemolysis assay has been a long accepted universal method of screening for biosurfactant production due to its ease of use (Banat, 1993; Ghojavand *et al*., 2008; Mulligan *et al*., 1984). However, the high likelihood of excluding biosurfactants that are of a weak hemolytic nature and its poor correlation between hemolysis and surface tension make it an unreliable test to use on its own. (Youssef *et al*., 2004b) In recent years, a number of studies, particularly from marine environments, have categorized microbial biosurfactants into two groups, those able to lyse erythrocytes and those unable to lyse erythrocytes, (Das *et al.*, 2008; Dhasayan *et al.*, 2015). It is therefore not unfeasible to suggest that some of the biosurfactants in the latter category may have been missed by only using the established hemolytic assay. However, it is still a rapid and easily operationalized biosurfactant identification method, still evidently used by many differential research fields (Augustine *et al.*, 2017; Kumar *et al.*, 2014; Woźniak-Karczewska *et al*., 2017). Therefore, as previously suggested, this method, coupled with an assay that directly associates biosurfactant production with IFT reduction, provides an effective tool for the screening of potential biosurfactant producers (Satpute *et al.,* 2008). This was further supported by the fact that only two of the 309 isolates (0.65%) in this study lacked hemolytic potential but showed oil dispersion.

All of the isolates identified as top surfactant producers exhibited a low or average score for biofilm production. This was determined based on the OD_{570}

of previous studies using a modified method based on O'Toole & Kolter (1998), and the level of biofilm formation they identified. It has been shown that weak biofilm formers produced an OD_{570} of below 0.5 units with strong biofilm formers associated with an OD_{570} of 0.8 units or above (Hossain & Uddin, 2014; Pathak *et al*., 2012; Tram *et al*., 2013) . This OD_{570} value was therefore used as a cut off for our screening protocol. With all isolates having scored under this threshold, it was decided to use the isolate with the highest OD_{570} value (C75), which incidentally had one of the lowest biosurfactant production scores in the screen of <3, in the assumption that it had the produced the greatest quantity of biofilm. It would therefore be possible to compare the effects of all of the biosurfactant producing strains to a known biofilm producer that did not produce an effective biosurfactant. This allows the possibility to compare the effect of a bio-plugging MEOR mechanism of an indigenous isolate against surface tension altering isolates. This strain, C75, was added to the list of isolates to be characterized and will be used in further studies (Chapter 5) to compare the effects of strong biofilm producing strains to that of moderate/weak biofilm producers that produced an active biosurfactant.

Although used to identify the potential biofilm producers in this study, crystal violet staining is not an accurate method for biofilm quantification. A major drawback to the crystal violet staining of biolfilms is that it stains both the extracellular polymeric substances and both the live and dead cells that are attached to the surface, or in this case cuvette wall (Mclandsborough, 2015) . This therefore is detrimental when assessing a more mature biofilm, as will not quantify only the active portion. However, for oilfield MEOR purposes, the aim

is to assess the use of biofilm for the blocking of pore space and diverting of flow. This therefore makes the need to assess only active cells less important, but increases interest in any cellular material that may cause a plugging effect. Furthermore, due to the short term analysis and lack of continual nutrient availability in this assay, mature biofilms would be unlikely to form. Therefore, the total quantity of dead cells would in all likelihood be reduced in this screening protocol.

The sequencing of the 16S rRNA gene identified isolates of 4 genera of bacteria. Within these genera, six individual species were recognized, however one of these could not be assigned to a speciation level within the *Acinetobacter* genus. The second isolate within this genus was identified via phylogenetic analysis and showed a 99% homology to *A.junii.* Numerous species of *Acinetobacter, a* common marine and aquatic bacteria, have been previously isolated from a number of diverse oilfield environments (Dong *et al*., 2016; Sarma *et al.*, 2004). The genus of *Acinetobacter* has shown some promise in relation to biosurfactant production for MEOR in recent years. Although not as common a research focus as members of the *Bacillus* and *Pseudomonas* due to the lack of biochemical profiling of produced biosurfactants, it is apparent from the literature that *Acinetobacter sp.* can produce a biosurfactant of a lipopeptide nature (Bao *et al*., 2014; Chen *et al.*, 2012). However, Zou *et al* (2014) exhibited that an indigenous oilfield isolate of *A. baylyi* was capable of reducing IFT to below 35 mN/m. This information, coupled with the high abundance (~ 0.5-1%) of *Acinebacter sp.* identified from

the oilfield samples (Chapter 3) and high surfactant assay scores, indicates that these isolates could be a good target for MEOR treatment.

The remaining species identified were that of *Escherichia coli, Pseudomonas aeruginosa, Bacillus licheniformis* and *Bacillus subtilis.* Within the Bentley field, the total number of *E. coli* derived sequence reads identified in X8650 (Chapter 3.3) was exceptionally low (>0.0001). This therefore suggests that there is a possibility that the *E. coli* isolate may have actually been a contamination introduced through either laboratory processing or from the culturing/enrichment process. However, the possibility also remains that this isolate may have been a true positive from sample X8650. This is further supported by the literature and experimental data (shown in Appendix A), where 16S rRNA gene primers were shown to significantly underestimate the abundance of bacteria of the class *Gammaproteobacteria,* particularly that of *Escherichia spp.* in both clinical and environmental samples (Jovel *et al*., 2016; Rajendhran & Gunaskaran, 2011).

Species of *Bacillus* and *Pseudomonas* are well characterized biosurfactant producers and have been isolated from numerous oilfield environments and used in MEOR trials previously with varying degrees of success (Daryasafar *et al.,*, 2016; Simpson *et al.,*, 2011; Youssef *et al.*, 2009; Zhao *et al*., 2015). Furthermore, *Bacillus* strains, including that of the *B. licheniformis* spp. have been isolated from heavy oilfields for MEOR purposes. However, these bacteria were only tested for their oil degradation potential rather than biosurfactant production (Al-Sayegh *et al*., 2015; García-Alcántara *et al.,*

2016). This study is the first to successfully document the isolation of a biosurfactant producing *B. licheniformis* strain from an oilfield with an API < 22°. Any *Bacillus* strains that have been tested for biosurfactant production and IFT reduction in heavy oils, used isolates that were identified from light oil reservoirs (API > 22°) or non-reservoir related environments (Joshi *et al.*, 2015; Pérez-Armendriz *et al.*, 2013). As such, the potential of these non-indigenous strains to survive and adapt to the extreme oil-field environments is substantially reduced, compared to that of indigenous strains that have been shown to thrive in these conditions.

Chapter 5

Characterizing the MEOR capabilities of biosurfactant and biofilm producing oilfield isolates using a Bentley Oilfield-simulating bioreactor microcosm

5.1 Introduction

5.1.1 Background

Since the 1950's biotechnological processes have been utilised in the oil industry for many practices, ranging from biocontrol of sulphate reducing bacteria in upstream oil production, to the denitrification and biocatalysis of petroleum products in the refinery process. However, with oil reserves rapidly dwindling and other forms of renewable energies not as readily available for energy production in the short term, the most important aspect of energy research is that of Microbial Enhanced Oil Recovery. MEOR, a tertiary oil recovery process, uses microorganisms and their metabolic products to increase the production of oil via a number of mechanisms (Youssef *et al.*, 2009). This method of oil recovery has been utilised ever since the early 1960's with varying degrees of success. However, most of our current understanding of MEOR relates to its use in laboratory-based models, with only a handful of field trials achieving substantial increases in oil recovery (References shown in Table 5.1).

5.1.2 Laboratory studies of MEOR

To understand the positive effects of any given treatment strategy on a system it is vital to test various treatments and measure their outcomes in the laboratory before moving forward to 'field' trials. This strategy applies to MEOR treatments, too. Accurately identifying both the benefits and pitfalls of the MEOR process in the laboratory makes it easier to design a treatment strategy that can be used in larger-scale, field trials. Therefore, one of the main factors that may have contributed to the failure of previous MEOR field trials was the lack of prior controlled, laboratory testing and thus limited knowledge of how simple perturbations might alter the microbial responses in the oil well. Though the majority of field MEOR trails have tried to use a biostimulative approach, stimulating the indigenous (authochonous) organisms already present in the well, this study has instead focussed on a bioaugmentative approach to MEOR (Chapter 1.2). Nevertheless, there have been a number of research groups that have carried out the preliminary lab work and shown vast increases in additional oil recovery *in vitro*, utilising a variety of microbial mechanisms (Banat, 1993; Qazi *et al.*, 2013; Suthar *et al., 2009)*.

In vitro MEOR microcosm work, using both sand-pack columns and core flooding analysis, can be observed using injected bacterial strains and purified biosurfactants, enabling recovery of over one third more additional oil (Makkar & Cameotra, 1999; McInerney *et al.*, 2005; Kanna *et al.,* 2016; Souayeh *et al*., 2014). Yet very few of these trials have actually gone on to lead to successful increases in oil recovery at a field reservoir level (McInerney *et al*., 2005; Youssef *et al*., 2007). This is likely to be primarily due to the discrepancies

between laboratory and field conditions, particularly that of pressure and temperature, which were not accurately matched in the laboratory (Table 5.1). Microorganisms that thrive in certain laboratory conditions were used in many of the *in vitro* studies, and these strains may not be suited to the harsher conditions of the extreme environments down hole, thus reducing the amount of oil recovered (Banat, 1993; Gudiña *et al*., 2013; Lazar *et al*., 1992; Wang *et al.*, 2008) .

Due to the diverse nature of an oil reservoir and heterogeneity between wells, it has proven an almost impossible task in the past to replicate oilfield conditions, a feature that Gray *et al.* (2008) suggested would be the downfall of MEOR. However, as expressed by both Maudgalya *et al.* (2007) and Safdel *et al.* (2017), the identification and matching of reservoir characteristics and the successful development of a laboratory models matching the oilfield parameters holds the key to the success, or failure, of MEOR.

5.1.3 Additional oil recovered by the use of microorganisms in reservoir simulated porous micromodels.

It is apparent that when oilfield parameters are ignored and testing of treatments are conducted under standard lab conditions (Table 5.1), higher recoveries are seen (Banat, 1993; Makkar & Cameotra, 1997; Qazi *et al*., 2013). However, with testing conditions not mimicking those of the reservoir this as we know does not bode well for the accuracy of the microcosm simulations, and will in all likelihood lead to failures in the field as we have seen many times before (Maudgalya *et al*., 2007). However, of the studies that

do mimic reservoir parameters, the most commonly simulated parameters are that of temperature, crude oil density and gravity and FW brine composition, with only a limited number of trials replicating additional parameters such as substrate porosity and well pressure. Yet the model systems where all oilfield conditions were replicated (*Pressure, Crude oil, Rock porosity, FW brine, Temperature*), yielded less oil recovery compared to the less exacting laboratory studies (Dastgheib *et al*., 2008; Suthar *et al*., 2009).

Most of the *in vitro* studies conducted have used sand pack porous micromodels. Although the use of core floods provides a more accurate estimate of a microorganisms MEOR potential, the increased expense in time and money of core flood experiments, alongside the difficulty of core aquisition, make the use of core floods an impractical process for large scale MEOR screening projects, like the one described in this thesis. Furthermore, it is apparent from the literature that very little research has been conducted with regards to heavy or unconventional oil fields. Although heavy oil is in a large abundance on the planet, due to difficulties in its mobilisation and refinery, heavy oil has not been the focus of research initiatives. However, with rapidly decreasing amounts of conventional oils and no obvious renewable energy technology ready to replace the use of fossil fuels, it is plausible to conceive that heavy oil MEOR could be the defining process to stave off the impending energy crisis.

Table 5.1 Reservoir simulated conditions and additional oil recovery in porous micromodel systems by bacterial species.

Microorganism	*Reservoir simulated features*	Model system / inoculum	Additional oil recovery (%)	Oil type	Reference
Bacillus sp. AB-2	-*	Sand pack column/ bacteria & nutrient	95	Light oil	(Banat, 1993)
B. subtilis MTCC 2423	-	Sand-packed column/ purified biosurfactant	60	Light oil	(Makkar & Cameotra, 1997)
B. licheniformis BNP29	-	Sandstone core/biomass & nutrient medium (selective plugging)	15	Light oil	(Yakimov *et al*., 1997)
B. brevis sp.	*N/A*	Sand packed column/consortia in nutrient medium	16.5	Light oil	(Almeida *et al*., 2004)
B. mojavensis JF-2	-	Berea sandstone core/crude purified biosurfactant	10–40	Light oil	(Mcinerney *et al*., 2004)
B. subtilis sp.	-	Glass etched flow micromodels/bacteria & nutrient medium	30	Light oil	(Soudmand-asli *et al*., 2007)
P. aeruginosa sp.	*Porosity, Temperature*	Sand pack column/ crude biosurfactant	30	N/A	(Bordoloi & Konwar, 2008)
B. licheniformis AC01	*Pressure, Crude oil, Rock porosity, FW brine, Temperature*	Sand packed column/ bacteria & nutrient medium	22	Light oil	(Dastgheib *et al.*, 2008)
B. licheniformis AC01	*Pressure, Crude oil, Rock porosity, FW brine, Temperature*	Sand packed column/ bioemulsifier	< 1	Light oil	(Dastgheib *et al.*, 2008)
B. licheniformis TT42	*FW brine, Pressure*	Sand-packed column/crude biosurfactant	35	Synthetic	(Suthar *et al.*, 2008)
B. licheniformis K125	*FW brine, Pressure*	Sand-packed column/crude bioemulsifier	43	Synthetic	(Suthar *et al.*, 2008)
B. mojavensis JF-2	*FW brine, Pressure*	Sand-packed column/crude bioemulsifier	29	Synthetic	(Suthar *et al.*, 2008)
B. subtilis 20B	*FW brine, Rock porosity*	Sand-packed column/crude biosurfactant	25–33	Light oil	(Joshi *et al*., 2008)
B. subtilis 20B	*Crude oil, FW brine, Rock porosity*	Glass packed column/crude biosurfactant	30	Light oil	(Joshi, *et al.*, 2008)
B. licheniformis TT33	*Pressure, Crude oil, FW brine, Temperature*	Sand packed column/microbial biomass in nutrient medium (selective plugging)	25–32	Heavy oil	(Suthar *et al.*, 2009)
Bacillus sp.	*Crude oil, temperature*	Glass etched micromodels/	13		(Gao, 2011)
E. sakazakii/B.subtilis fusion	*Pressure, FW brine, Temperature*	Sand pack column & sandstone core/ engineered bacteria & nutrient	17-25	N/A	(Xu & Lu, 2011)
B. licheniformis sp.	*Temperature, Crude oil*	Sand pack column/ bacteria & nutrient	6-25	Light oil	(Gudiña *et al*., 2013)
B. licheniformis sp.	*Temperature, Crude oiil*	Sand pack column/ bacteria & nutrient	15-17	Heavy oil	(Gudiña *et al*., 2013)
B.licheniformis R1	*FW brine, Rock porosity*	Sand pack column/ crude biosurfactant	32	N/A	(Joshi *et al*., 2013)
Fusarium sp. BS-8	*FW brine*	Sand pack column/ crude biosurfactant	46	Light oil	(Qazi *et al*., 2013)
B. subtillis W19	*Crude oil, FW brine, Porosity*	Berea sandstone core/ crude biosurfactant	15	Light oil	(Souayeh *et al*., 2014)
C. albicans sp.	*Crude oil, FW brine*	Sand pack column/ selective plugging	9	Light oil	(El-Sheshtawy, *et al.*, 2016)
B. subtilis MTCC2422	*N/A*	Sand pack column/ bacteria & nutrient	9	Synthetic	(Rajesh Kanna *et al.,* 2016)
B. licheniformis ATCC 14580	*Crude oil, FW brine*	Sand pack column/ crude biosurfactant	17	Light oil	(El-Sheshtawy, *et al.*, 2016)

*Signifies no petrophysical or geochemical characteristics were simulated in the model system.

5.1.4 Research hypothesis, aims and objectives

The primary aim of this chapter was to successfully create a reproducible microcosm model to simulate the reservoir characteristics of the Bentley Oilfield. Once created and optimized, these porous micromodels will be used to investigate the potential of reservoir isolates previously identified in Chapter 4 to drive additional oil recovery, in comparison to other known MEOR stimulating microorganisms. The potential of the reservoir isolates to drive MEOR in our simulated reservoirs will be done through analysis of bioreactor models and application of qPCR, establishing whether surfactant mediated IFT alteration is the predominant MEOR mechanism.

This chapter investigated two distinctive hypotheses. Firstly, it was hypothesised that any isolates from the heavy oilfield environment (Bentley Oilfield) that were capable of producing SACs would increase residual oil recovery. Secondly, the hypothesis that a dual mechanisms of action was taking place within the bioreactor was also tested.

5.2 Methods

5.2.1 Microorganisms

Alongside the strains listed in Table 2.1, bacterial isolates C10, C13, C16, C19, C20, C36, C75, C82 & C108 isolated in Chapter 4 were used in the work detailed in this chapter. Isolate C10 was utilised for the majority of experimental work for the remainder of this thesis and to reflect future uploading to the Genbank database, it is now referred to as strain Bi10.

5.2.2 Culture conditions

All strains were cultured from single colonies picked from LB agar plates and inoculated into fresh LB broth and incubated overnight as stated in Chapter 2. Cultures were then subcultured (1:4 or 1:40) into 4% OFM (at appropriate OD_{600} as stated in the results section) and incubated at 40ºC.

5.2.3 Reservoir microcosm construction

Microcosms were constructed according to the methodologies of (Gudiña *et al.*, 2012; Suthar *et al.*, 2008). All valves, fittings and components were purchased from Hamlet (Hamlet Group, Isreal) and rated to Letlock certified standards at 3000 psi and 80°C. Bioreactor model systems were manually constructed in the BiSN Oil Tools (UK) workshop and assessment of stability under high pressures conducted using gas flow systems (Figure 5.1).

Each microcosm was matched to formation specific conditions for this particular reservoir and was subjected to conditions mimicking primary/secondary and tertiary recovery. This was achieved by flooding approximately four pore volumes of formation water into each reactor to simulate water flooding of the reservoir. In order to evaluate the effect of

MEOR each microcosm trial was carried out in triplicate. The total pore volume (PV) of all sand pack columns was between 47-57 ml, with porosity in the range of 57%- 69%. Each bioreactor was flooded with 44-56 ml of heavy oil, equating to approximately 86-98% of the total sand pack saturation. Each sand packed microcosm was inoculated and left to incubate for a 14 day period at 40 ºC, after which the microcosms were then recovered by water flooding. Additional oil recovery was calculated.

Additional oil recovery was calculated using the equation:

$$\boldsymbol{AOR}\ (\%) = \frac{\boldsymbol{TRmf - TRwf}}{\boldsymbol{TRwf}} \times \mathbf{100}$$

AOR = Additional oil recovered

TRmf = Total recovery after microbial flooding including secondary recovery

TRwf = Total recovery after water flooding from control recoveries

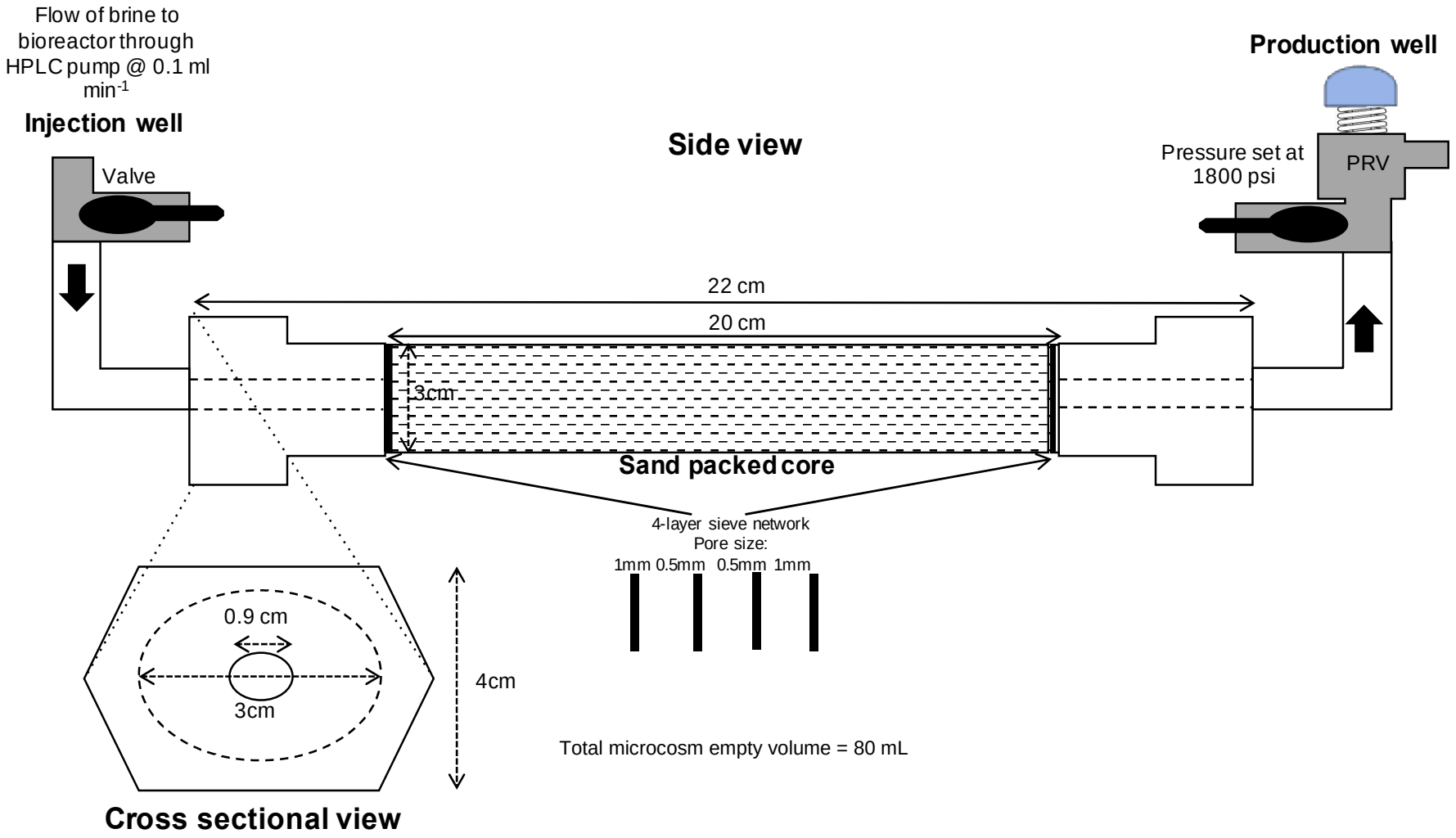


Figure 5.1 Diagrammatic sketch of sand pack model microcosm construction

5.2.3.1 Preliminary construction of sand-pack columns

Horizontal steel columns, manufactured by BiSN Oiltools, were assembled as shown in Chapter 2.12. To replicate bioreactor conditions, it was important to trial many combinations of differing sand grain sizes to be able to match the resulting column pore size and microcosm porosity to the reservoir porosity of 40%. Sand particles of three different mesh sizes were used: A= 20-30 (500-850 µM), B= 40-100 (150-425 µM), C= 50-70 (212-325 µM), provided by VWR, UK. The addition of these particles at different ratios were then assessed for total reservoir porosity (Fig. 5.2). Of the eight sand mixtures trialled (Fig. 5.2A), a ratio of 75%B:25%C produced a porosity that was the best match to the actual reservoir porosity (40% ± 1).

As stated in the introduction, tertiary recoveries are usually undertaken on a depleted oilfield after the oil cut (quantity of oil to total production liquid volume) has passed below 1%. It was therefore critical to assess what volume of FW would need to be passed through the bioreactor to reach this value. The oil cut was measured by flooding FW through the column in 10 mL increments and assessing the total oil recovery weight (Fig. 5.2b). After secondary recovery of 60-70 mL from the bioreactor, the model reservoir was in the classification zone of mature (Parshall, 2012). After 100 mL of water flooding, the model bioreactor reached >95% water cut. It was determined that the reservoir became fully depleted after 130 mL had been flooded through the bioreactor, with the oil cut reaching 0.4%. However, after 140 mL, the oil cut value remained consistent at 0.18%, with no further oil recovered after

additional flooding with up to another 50 mL. The decision was therefore made to use 100 mL as the total volume of secondary recovery in each of the microcosms, with a further 50 mL of bacterial culture and nutrient or nutrient alone inoculated for the tertiary recovery phase. This inoculation forced out a small fraction of additional oil (~0.8g) taking the total recovered oil volume after inoculation to above the 95% water cut threshold.

5.2.4 Chemical treatment of *B. licheniformis Bi10* cells for surfactant isolation and cell removal

Bi10 cells were cultured in 4% OFM for 24 hr at 37ºC with shaking at 220 rpm. Cells were removed from the Bi10 culture by centrifugation at 13000 g for 20 min and the supernatant collected. This supernatant was then treated with formalin treatment using equal parts formalin (37% w/v) solution to water and heated in a formalin kettle for 1 hr. The cell free supernatant was then assessed for microbial growth. Only those samples that contained no viable cells as determined by growth on agar plates or Gram/ spore staining under the light microscope, but demonstrated oil spreading ability were used in the bioreactors.

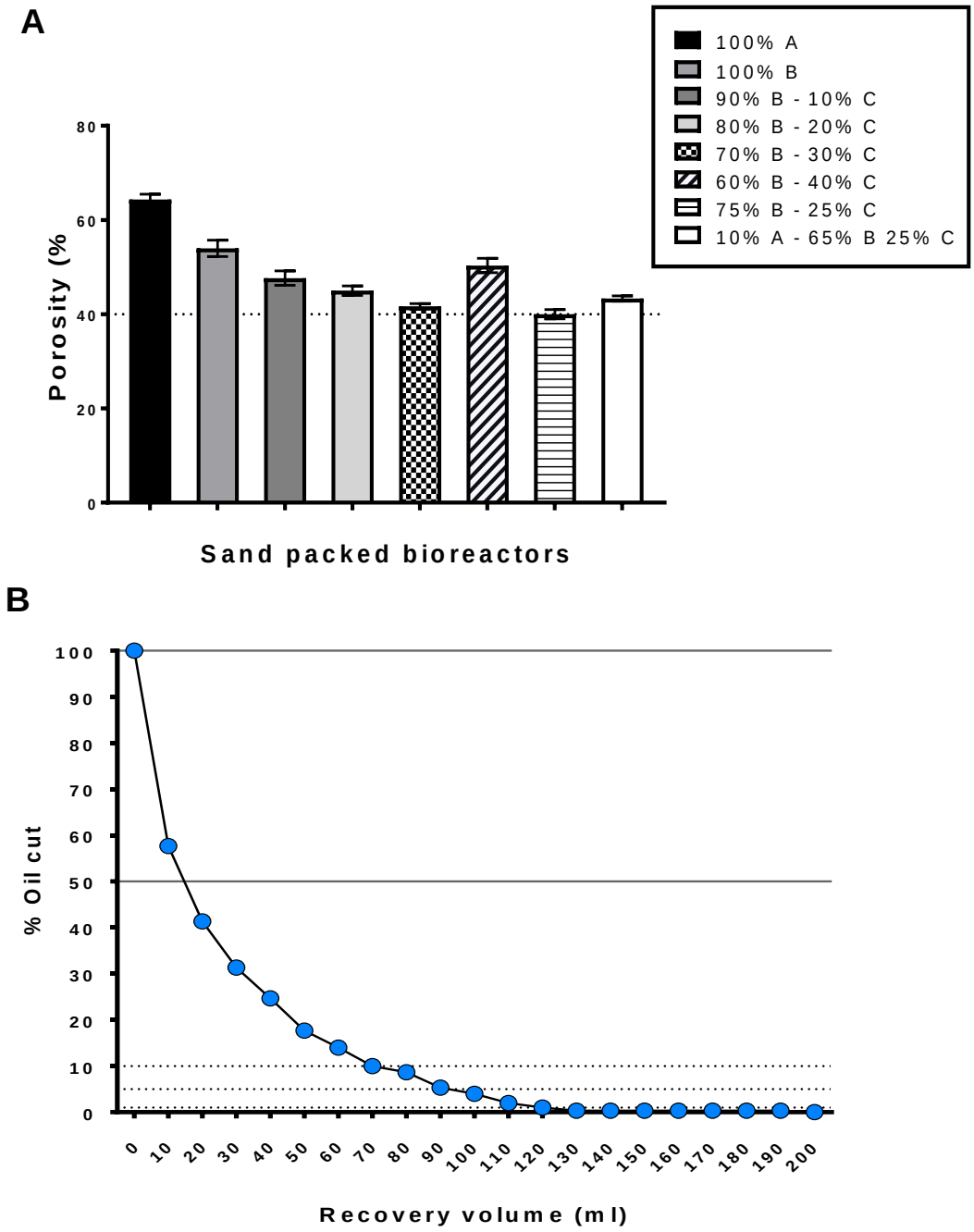


Figure 5.2. Preliminary assessment of bioreactor parameters to replicate reservoir recovery characteristics. Figure 5.2a shows the differing ratios of sand mesh sizes added to the bioreactor to achieve porosity values of around 40%. Figure 5.2b illustrates the total oil cut after secondary water flooding, based on the average of three individual bioreactors.

5.2.5 Heat treatment of *B. licheniformis Bi10* cells for spore isolation

To induce sporulation, Bi10 cells were cultured using standard culturing techniques into 4% OFM and incubated for 24hr at 37ºC with shaking at 220 rpm. After incubation, 100 mL flasks of inoculated 4% OFM was heat treated at differing temperatures to allow for spore survival using a waterbath. Heat treatment of 30 minutes at temperatures ranging from 60-90ºC, increasing in 5ºC increments, was tested. Both LB dilution plating and Schaffer-Fulton spore stains were conducted to test for bacterial remnants, germination and spore survival. Identified spore suspensions were inoculated into a bioreactor. Table 5.2 demonstrates that at temperatures below 75ºC, both spores and cells were detected. However, at temperatures above 80ºC, no cell or spore survival was identified. Therefore, the minimum temperature for cell killing and spore survival is ~75ºC. This temperature was used to create the spore only inoculum.

Table 5.2 Heat treatment of Bi10 cells at a range of temperatures to show survival of cells and spores in 4%OFM nutrient.

Temperature	50ºC	55ºC	60ºC	65ºC	70ºC	75ºC	80ºC	85ºC	90ºC
Cells	+	+	+	+	+	-	-	-	-
Spores	+	+	+	+	+	+	+	-	-

5.2.6 Quantitative PCR of biomass added to and extracted from bioreactor microcosms

16S rRNA gene copies were quantitated using qPCR of the inoculated media (cell number added) and effluent (cell number in the bioreactor) for each treatment strategy as stated in Chapter 2, using 515f – 806r primer sets and

recommended PCR conditions. qPCR assays were ran on the Eco Real-time PCR system (Illumina, US). All data analysis was conducted using EcoStudy software V5.0 (Illumina, US).

5.2.6.1 Identification and removal of PCR inhibitors via qPCR analysis of human mitochondrial DNA

To investigate the effect of PCR inhibitors on our DNA extractions from the molasses based media and bioreactor recoveries, a non-bacterial DNA template of the H1 mtDNA, previously prepared by Dr. Paul Brotherton (Brotherton *et al*., 2013) of known concentration was used as a PCR standard to check for reduced amplification. The synthesised template standard and primers used related to the hypervariable region 1 of human mitochondrial DNA (mtDNA). Reactions were set up using the protocol and reaction conditions given in Endicott *et al*.(2009): 2X KAPA SYBR FAST qpcr Master Mix (Kapa Biosytsmes, US), 0.2 µM mtDNAf & mtDNAr primers (Table 2.4, Chapter 2.9), 1 µL mtDNA standard (1000 copies), produced by Dr. Paul Brotherton, BiSN, UK. To this mix, 1 µL of pre/post treatment DNA extract (50 mL biomass extracted into 50 µL) for each treatment strategy was added and qPCR performed. Pre-made standards of human mtDNA from10^2-10^7 were used for quantification. For each sample, two assays were set up: i) A standard PCR reaction and ii) a PCR reaction with BSA solution added (prepared from crystallized BSA (Sigma-Aldrich, US) with DNA/RNAase free ddH_2O to achieve final concentration of 200 ng μL^{-1}).

5.3 Results

5.3.1 Application of biosurfactant producing isolates from an unconventional reservoir for the enhancement of oil recovery

To prove the feasibility of MEOR in a formation specific scenario, sand packed microcosms were set up matching the petrophysical parameters of the Bentley Oifield (as described in Chapter 2). Bacterial strains isolated from the enrichment of Bentley oilfield samples (Chapter 4) were subject to a replicated tertiary oil recovery protocol. The pore volume (PV) of the columns, or total volume once packed, ranged from 50 to 52 mL, possessing a porosity of 42% ± 0.8. The original oil in place (OOIP) accounted for 86.5% (~45 mL), with initial water saturation accounting for the remaining 13.5% (~ 7 mL). After secondary recovery water flooding, ~ 65% of the oil remained in the microcosm. The introduction of all eight isolates to their respective bioreactors resulted in a significant increase in additional oil recovery (AOR) when compared to the negative control (FW flooding, only) (Fig. 5.2). The highest recoveries where from isolates C10 (11.85%) and C36 (10.53%), both previously identified as strains of *B .licheniformis.* However, C10 produced a significantly higher recovery than all the other isolates. The lowest recovery was provided by the *E. coli* isolate (C75), which was only resulted in 5.1% AOR. As C10 resulted in the highest AOR, this isolate was characterised further, and is now referred to as strain Bi10. Due to the industrial nature of the project, the sponsor company decided they wanted the strain to be associated with their company in case of future work away from this project. This was the reasoning behind the name change.

5.3.2 Reduction of starting cell concentration in bioreactor experiments

Although isolate Bi10 facilitated the highest AOR, the concentration of cells used in the microcosm was impractical to replicate at the oil reservoir. To identify if a reduction in added cell numbers would still effect oil recovery, the starting concentration of Bi10 cells in the inoculum was reduced from 10^8 to 10^5 and AOR determined for each concentration (Fig. 5.3). Bi10 was inoculated using varied serial dilutions (OD_{600} 0.0001-0.1) into different sand-packed reactors that were incubated for 14 days. Due to cell death and sporulation over the 24h incubation period prior to inoculation, cell volume at the initial testing parameter was reduced in comparison to that of the sample that was inoculated into the bioreactor at of OD_{600} 0.1 without incubation. The removal of the 24h incubation time prior to inoculation yielded a significant decrease in the addition oil recovery (AOR) from 11.85% ± 0.4 to 10.91%± 0.3. Reducing cell concentration by a further 10 fold showed no significant variation in AOR (11.79% ±0.2). However, a reduction by 100 fold or higher did significantly reduce AOR to below 10%. A volume of less than 2 x 10^6 cells mL^{-1} shows a substantial reduction in the microbial enhanced oil recovery potential of the Bi10 isolate. Therefore, it is recommended that future experimental work with isolate Bi10 should be carried out using a starting OD_{600} of 0.01 (~ 7 x 10^6 cells mL^{-1}).

5.3.3 Additional oil recovery of biosurfactant producing isolate in comparison to other known MEOR utilized bacteria

To evaluate the effectiveness of oil recovery by the Bi10 isolate in comparison to other known MEOR related strains, the AOR of Bi10 was assessed against

other known bacteria, conditions and apposite controls (for experimental detail see Section 5.2). For all Bi10 conditions, bacterial cultures were set up as previously identified by the preliminary work above, using a starting bacterial concentration of ~ 7 x 10^6 cells mL^{-1}. A total of eight sand-pack bioreactors were set up for each of the variable treatment strategies, simulating reservoir conditions. The PV of the columns ranged from 47-57 mL respectively, with porosity measured between 39.17 and 47.5% (Table 5.3). Original oil in place (OOIP) accounted for between 44 - 56 ml (86.5-98% initial oil saturation). From the previous bioreactor experiment, Bi10 cells accounted for an 11.79% ±0.2 increase in AOR. This was significantly higher than the recoveries of the two non-microbial controls, FW (0.417% ± 0.644) and Nutrient flooding (2.3% ± 0.913).

Bioreactors treated with components produced by Bi10 cultures (Bi10 spores or surfactant, only) both increased AOR in comparison to the controls. Although, Bi10 spores facilitated an AOR of a similar percentage to Bi10 cells (11.214% ±0.190), statistically both spores and surfactant treatments resulted in an AOR that was significantly reduced ($p < 0.05$) when compared to the Bi10 cell mediated AOR (Fig. 5.4). Bi10 surfactant resulted in an AOR of only 8.535% ±0.227.

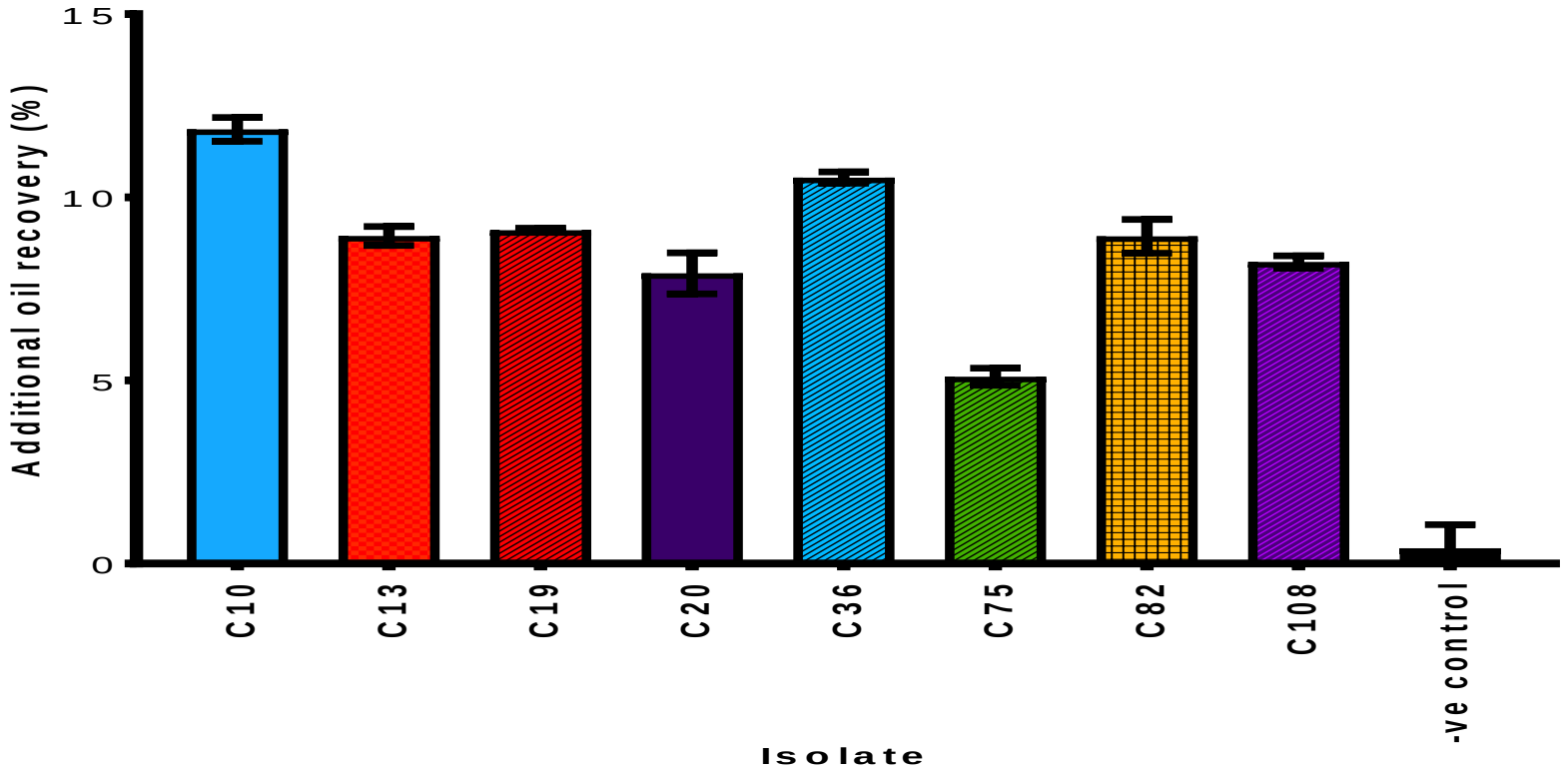


Figure 5.3 Comparisons of AOR in sand-pack bioreactors from Bentley Oilfield isolates. Values are shown as mean AOR (± SEM, *n*=3 replicates). Colours represent bacterial species as follows: *Bacillus licheniformis* (Blue), *Bacillus subtilis* (Red), *Acinetobacter spp*. (Purple), E*scherichia coli,* (Green) *. Pseudomonas aeruginosa (*Yellow)

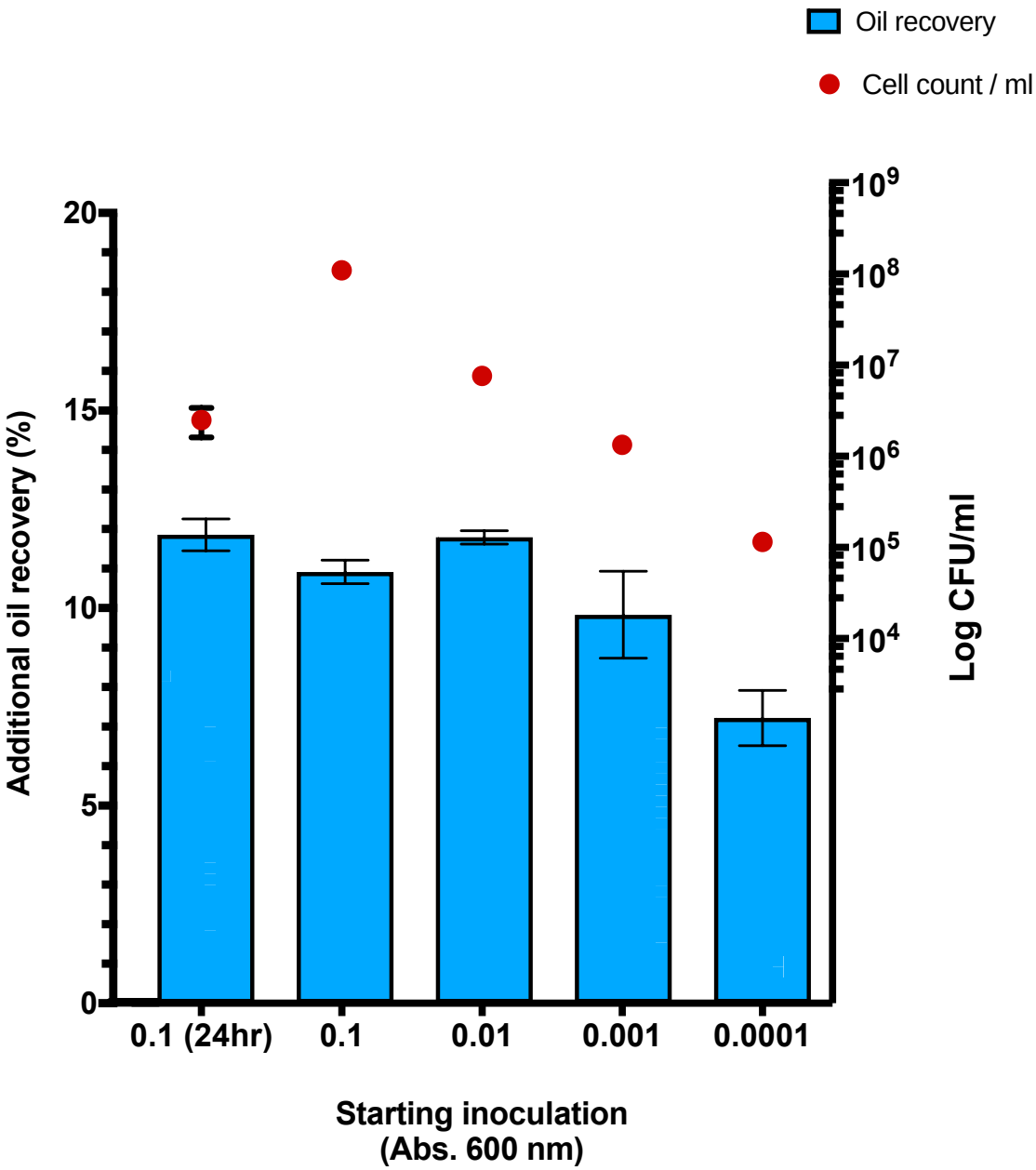


Figure 5.4. Effect of reduction in cell numbers on AOR of *B.licheniformis Bi10* cells. Concentration of cells mL^{-1} for each reduced absorbance and additional oil recovery. Values are shown as mean AOR and logCFU mL^{-1} (± SD, *n*=3 replicates).

Of the type strains that have been characterised as effecting MEOR (Michael J Mcinerney *et al.*, 1990; El-Sheshtawy *et al.*, 2016., 1998; Zhao *et al*., 2015), all three mediated a significantly higher amount of oil than the FW flooding control. The highest recovery associated with the type strains was that effected by *B. mojavensis JF2,* which mediated an AOR of 10.431 (±0.576), followed by *B.licheniformis ATCC14580* produced *an AOR of 10.076 (*±0.389) and *P.aeruginosa* PA01 mediating an AOR of 7.26% (±1.238). All five microbial treatment strategies recovered a significantly lower amount of oil than that of the Bi10 cell MEOR treatment (Fig 5.4).

5.3.4 Microbial analysis of bioreactor pre-recovery inoculum and post recovery effluent

To assess the changes in bacterial numbers in the microcosm after MEOR treatment, qPCR analyses was performed on total DNA extracted from the inoculum and post recovery effluent to determine the 16S rRNA gene copy number at varying times during the MEOR process (Fig. 5.5A). This data showed no significant reduction in 16S rRNA gene copy number after treatment in the Bi10 treated bioreactor microcosms (4×10^5 ±201528 mL^{-1} reduction). After treatment with only FW flooding or 4% OFM no significant difference in 16S rRNA gene copy number was apparent. However, the treatment of *P. aeruginosa* PA01 showed an increase of 1.2×10^6 mL^{-1} (±1478913), but a paired t-test showed that this difference was not statistically significant.

Table 5.3 Summary of results obtained from sand-pack bioreactors for unconventional oil recovery

	Inoculum							
	Bi10	**FW ***	**Nutrient only**	**Bi10 spores*ƒ***	**Bi10 Surfactant§**	***P. aeruginosa PA01***	***B. mojavensis* JF2**	***B. licheniformis* ATCC14580**
Bioreactor	PS1	PS2	PS4	PS6	PS3	PS5	PS8	PS7
Total Empty Volume (mL)	80	80	80	80	82	83	80	80
Pore Volume (mL)	52	51	49	52	48	57	47	53
% Porosity	43.33%	42.50%	40.83%	43.33%	40.00%	47.50%	39.17%	44.17%
Oil in Place (mL)	45	50	44	45	46	56	44	45
Initial Oil Saturation	86.50%	98.00%	89.79%	86.50%	95.80%	89.47%	93.62%	86.5%
Initial Water Saturation	13.50%	2.00%	10.21%	13.50%	4.20%	10.53%	6.38%	13.50%
TRwf ɘ	15.257	17.551	16.913	14.859	16.815	16.377	13.059	16.913
TRmf ∂	17.066	17.624	17.304	16.512	18.25	17.566	14.47	16.9974
AOR ℑ %	**11.854 ± 0.563**	**0.412 ± 0.644**	**2.3 ±0.913**	**11.171 ± 0.061**	**8.535 ± 0.227**	**7.26 ± 0.644**	**10.768 ± 0.885**	**9.989 ±0.713**

* FW= Sterile Formation water inoculated into microcosm and shut in for 14 days replicating continued water flooding

§ BI10 Surfactant = Cells/spores filtered and killed using formaldehyde vapour

ƒ BI10 Spores = Using heat treatment, viable cells were killed off so only heat resistant spores survive

ɘ Total recovery after water flood flow

∂ Total recovery after MEOR flow

ℑ Additional oil recovery

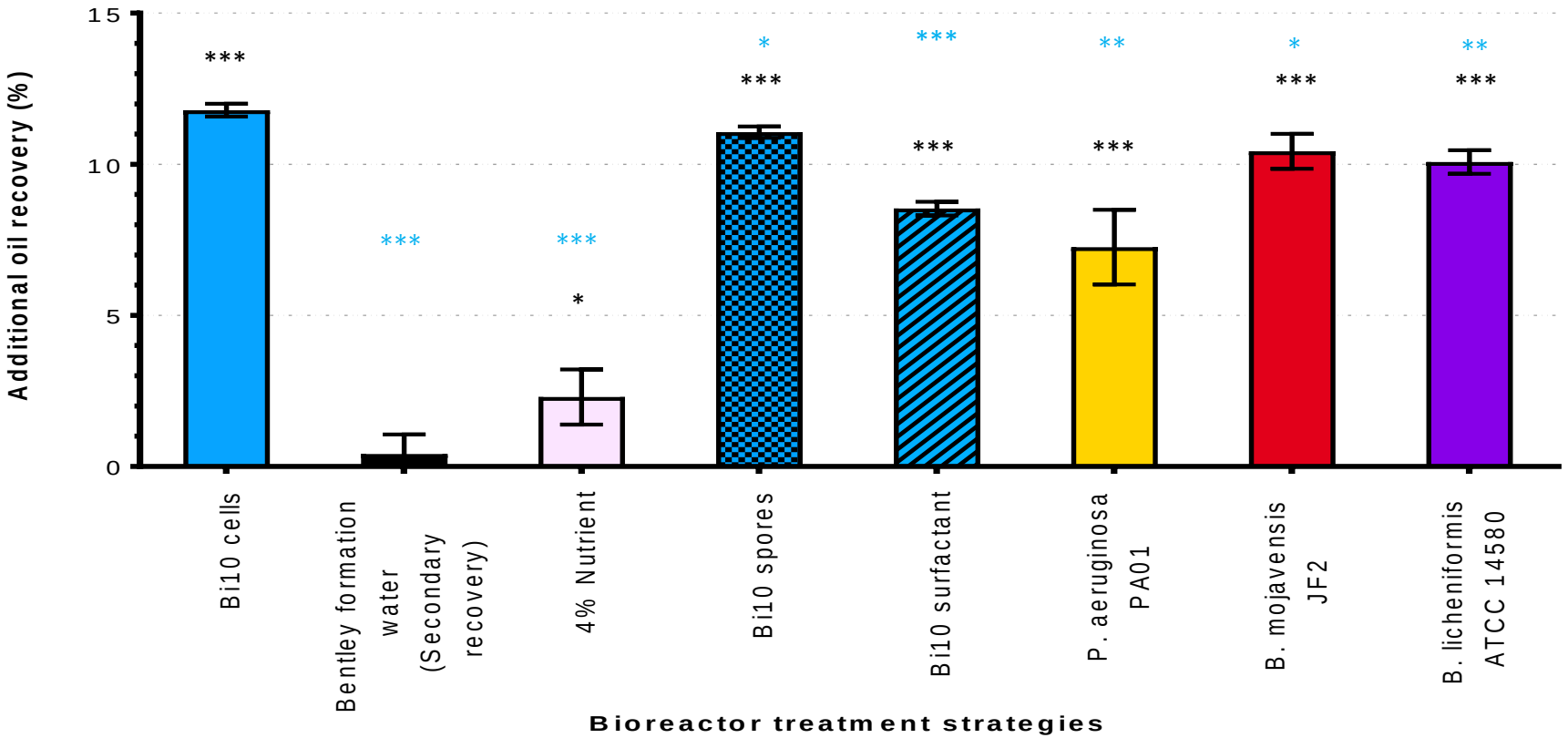


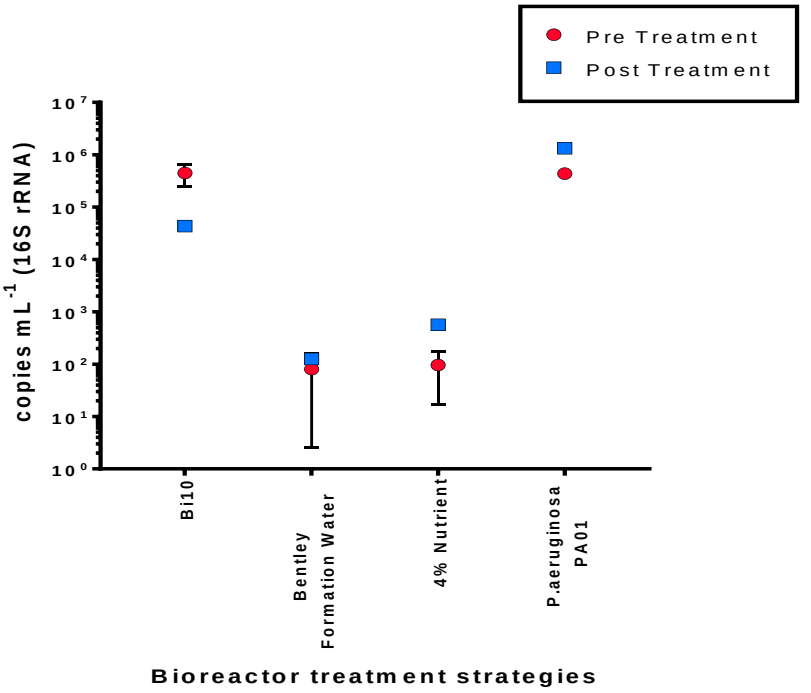
Figure 5.5 . Additional oil recovery (AOR) by variant MEOR treatment strategies. Total amount of additional residual oil recovered by several differing inoculate from an unconventional oil microcosms. Values shown as mean additional oil recovery ± SD. Statistical analysis compares AOR of treatment strategy against AOR of Bi10 cells (* p <0.05, ** p <0.01, *** p < 0.001). Statistical analysis compares AOR of treatment strategies against FW control (* p <0.05, ** p <0.01, *** p < 0.001).

Alongside the qPCR data, for the initial two bioreactor treatment strategies (Bi10 cells & 4% nutrient flooding) serially diluted colony counts were also performed before and after treatments (Figure 5.5b). This was implemented primarily due to previous issues in the laboratory with PCR inhibition from oilfield and concentrated nutrient samples. In comparison to the qPCR data, these counts revealed some interesting results. For Bi10, at both pre and post treatment CFU mL^{-1} was significantly increased in comparison to that of the 16S qPCR data, by 1.22 units and 0.9 log units respectively. However, no significant difference was seen for the nutrient inoculation at pre or post treatment conditions (Fig, 5B).

5.3.4.1 Improved qPCR amplification by adaptation of conditions

To determine the effects of amplification reduction by the presence of PCR inhibitors, pre-treatment culture and post treatment effluent recoveries were spiked with a known copy number of human mtDNA and underwent qPCR analysis. Samples were assayed in triplicate with one assay undergoing qPCR with standard reagents and the other with the addition of BSA, a known treatment for the eradication of PCR inhibitors (Schrader *et al*., 2012). A positive control of standardised mtDNA without the addition of recovery extractions taken from the microcosm effluents containing PCR inhibitors was also analysed.

With the of BSA to both pre-treatment and post-treatment recoveries, the total copy number was significantly increased from



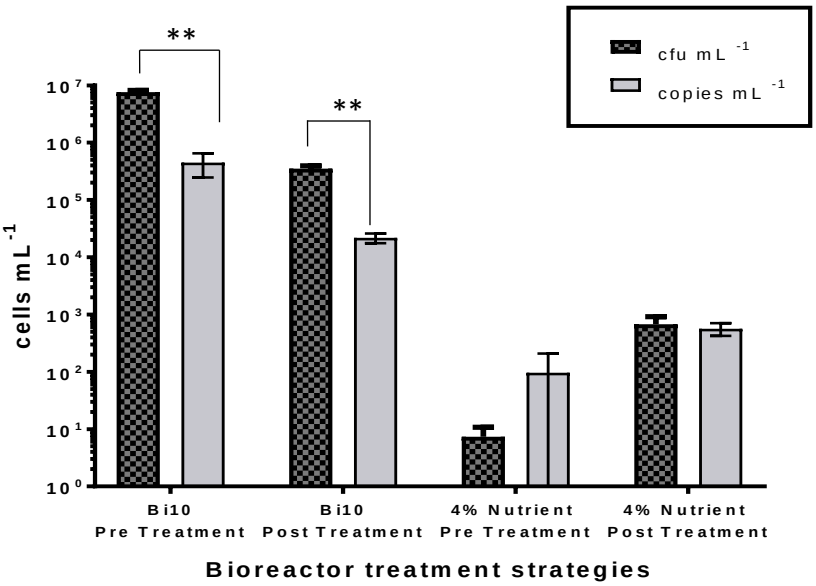


Figure 5.6 Preliminary qPCR results to test for PCR inhibition from heavy oil bioreactors. Studies undertaken to assess for any PCR inhibition on 16S rRNA gene from bioreactor studies. Panel A shows initial qPCR data of four treatment strategies, before and after treatment was applied. Panel B demonstrates discrepancies between 16S rRNA gene copy number and bacterial culturing between treatments of Bi10 and 4% OFM. Values shown as means ± SEM, N=2 (* p <0.05, ** p <0.01, *** p < 0.001).

the untreated (no BSA) samples ($P < 0.05$) for all samples, with the exception of the FW pre treatment, for which there was no recognizable change. A mean increase of 813-1074 gene copies mL^{-1} was seen in the 4% OFM nutrient, Bi10 and PA01 samples after treatment with BSA, in both pre and post MEOR. The greatest reductions in PCR efficiency where identified in the Bi10 pre-treatment and 4% OFM nutrient treatments. However, after BSA addition at 200 ng μL^{-1}, there were no significant differences found in the the copy number determination of mtDNA across all samples. Furthermore, when comparing between BSA treated, and non BSA treated samples, all except for FW pretreatment showed a significantly higher 16S copy number when BSA treatment had been undertaken. The bioreactor samples contained a number of inhibitors that clearly interfeared with the efficiency of the qPCR, giving varied copy numbers. This however was effectively counteracted by using BSA solution (200 ng μL^{-1}) in all PCR reactions. This then provided us with a more accurate relative quantitation of the total copy number of 16S rRNA genes present within the sample, pre and post recoveries.

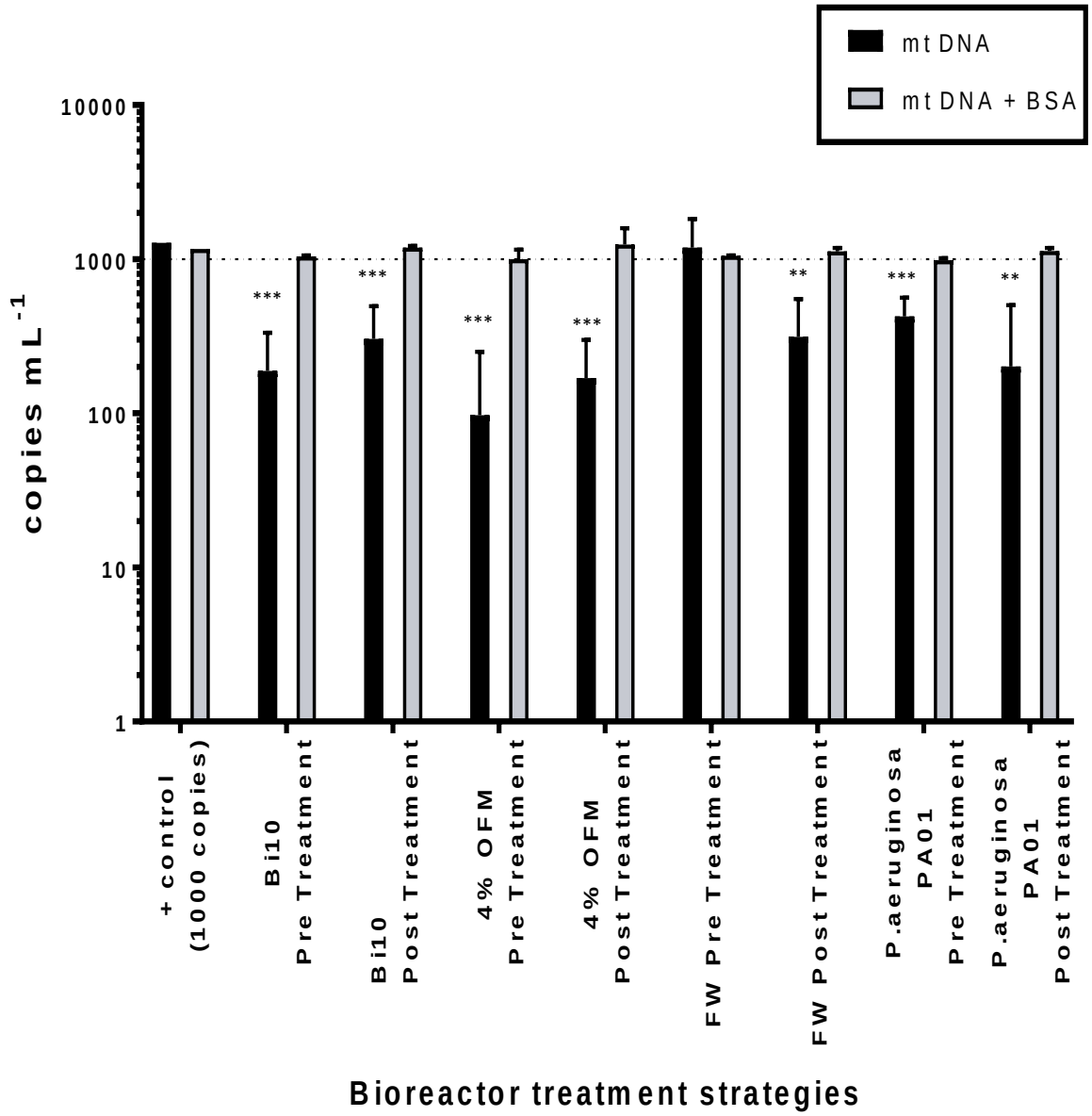


Figure 5.7 Effect of BSA addition on qPCR efficiency of mtDNA. Values shown are mean mtDNA copy number ± SEM with with an unpaired T-test used to identify significant differences between + control and treatment strategies before and after the addition of BSA to the PCR amplification reactions(* p <0.05, ** p <0.01, *** p < 0.001).

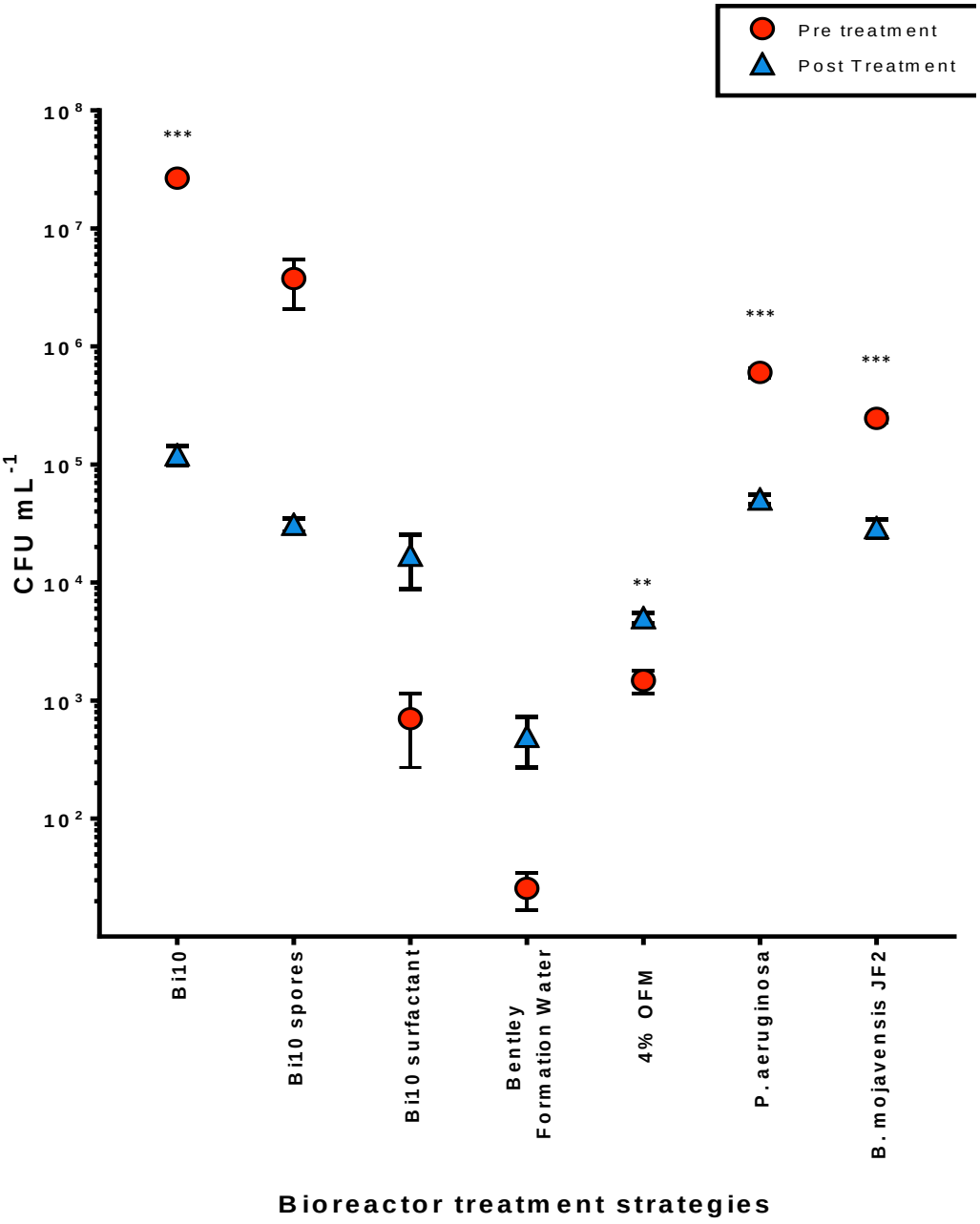


Figure 5.8 Quantification of bacterial numbers in treatment effluent prior to and after the application of MEOR treatment strategies measured by qPCR Values shown as mean copy number ± SEM (16S rRNA gene copy number). Paired T-test used to compare differences before and after application of MEOR treatment (* p <0.05, ** p <0.01, *** p < 0.001)

5.3.4.2 qPCR of bioreactor pre-recovery inoculum and post recovery effluent for accurate determination of 16S rRNA gene copy numbers after PCR inhibitor removal

After determining that BSA facilitated a more accurate determination of gene copy number in our qPCR assays, qPCR with the addition of BSA was performed again to accurately quantify copy number of the 16S rRNA gene. Seven treatment conditions were analysed, both before and after MEOR in the bioreactors: Bi10 cells, Bi10 spores, Bi10 surfactant, FW, Nutrient, *B.mojavensis JF2* and *P.aeruginosa PA01.* qPCR analysis showed that for Bi10 cells ($2.7 \times 10^7 \pm 3.00 \times 10^6$), *P.aeruginosa PA01* ($5.5x\ 10^5 \pm 5.95 \times 10^4$ and *B.mojavensis JF2(* $2.2 \times 10^5 \pm 2.12 \times 10^4$), copies of the 16S rRNA gene were significantly reduced after bioreactor treatment. Conversely, the use of 4% sterile OFM nutrient showed a significant increase in copy number after injection and treatment, exhibiting a 3.5×10^2 (± 609.4) increase after the "shut-in" incubation period. Although a substantial decrease in 16S rRNA gene copy number was also seen in the Bi10 spore treatment (addition of bacterial spores only) after reservoir shut in ($3.7 \times 10^6 \pm 1.25 \times 10^6$ mL^{-1}), an result of low DNA extraction in third repeat bioreactor led to a skewing of the data and the difference not being deemed as significant by a standard paired t-test.

5.3.5 Interfacial tension alteration of pre-recovery inoculum and post recovery effluent for assessment of biosurfactant production

The pre and post MEOR recoveries (Table 5.4) were assessed for interfacial tension alteration using the oil spreading technique. The cell/spore recovery was also assessed using light microscopy. For both of the non-microbial controls, no alteration in oil spreading ability was seen either before or after

MEOR treatment application as neither FW or 4% OFM nutrient treatment was able to reduce the interfacial tension at the oil/water interface. The presence of biosurfactant was identified in all four pre treatment inocula. Both the Bi10 cells and Bi10 spores increased the amount of oil spreading after removal of the inocula from the bioreactor. The same was seen for the *B. mojavensis JF2* strain however, this was not as potent as for the Bi10 strain. However, no alteration in oil spreading potential was observed after MEOR treatment for the *P. aeruginosa PA01* strain. In terms of microscopic analyses, both bacterial cells and spores were present in the pre and post recovery fluids for the Bi10 and *B. licheniformis JF2 treatments.* No viable cells were present in the pre recovery inoculum of the Bi10 spores due to a cell killing treatment. However, post recovery, both bacterial cells and spores were present in the recovered effluent. This is potential evidence of spore germination occurring in the bioreactor, which then may lead to the production of further quantities of biosurfactant.

Table 5.4 Assessment of bacterial existence and metabolite production in pre and post recovery MEOR effluent.

Assay	**Treatment strategy**	**Bi10 cells**	**Bi10 spores**	**FW**	**4% OFM nutrient**	***B. licheniformis JF2***	***P. aeruginosa PA01***
Oil spreading potential Ω	Pre treatment	+	+	-	-	+	+
	Post treatment	+++	+++	-	-	++	+
Bacterial cells †	Pre treatment	+	-	-	-	+	+
	Post treatment	+	+	-	+	+	+
Bacterial spores †	Pre treatment	+	+	-	-	+	-
	Post treatment	+	+	-	-	+	-

Ω Oil spreading diameter potential scored based upon: (-) no clearing zone, (+) clearing zone < 1 cm, (++) clearing zone >1cm but <3 cm, (+++) clearing zone > 3 cm.

† (+) presence of bacterial cells or spores in media or effluent; (–) absence of cells or spores

5.3.6 Application of Bi10 strain for MEOR using non-sterile simulated reservoir

To assess the applicability of the Bi10 strain for MEOR in an oilfield environment, the microcosms must accurately represent downhole conditions. Previous bioreactors have not precisely mimicked oilfield conditions. Although geological and petrochemical parameters were matched as closely as possible, the size and sterile nature of the bioreactors are not replicated in down hole conditions, where multiple microbes will already be present and occupying their established niches (Hibbing, Fuqua, Parsek, & Peterson, 2010). To replicate this competition, a non-sterile bioreactor was trialled. This bioreactor was inoculated with collected formation water from the Bentley field and left to incubate for 7 days. In the non-sterile bioreactor, the AOR efficiency of Bi10 was significantly reduced by 4.51% to 7.34% ± 0.3.

The Bi10 cells were also trialled in a bioreactor model of an increased volume, from 80 mL to ~ 1500 mL empty volume. By increasing the total bioreactor volume by approximately twenty times, it was theorized the system would better model the scaling up process of MEOR moving towards field trial conditions. The increase to a larger bioreactor under non-sterile conditions resulted in no significant difference in the total amount of additional oil recovered ($p > 0.05$). This data indicates that the Bi10 isolate was able to colonize a non-sterile bioreactor to produce an increase in oil recovery and that scaling up of the operation should be possible.

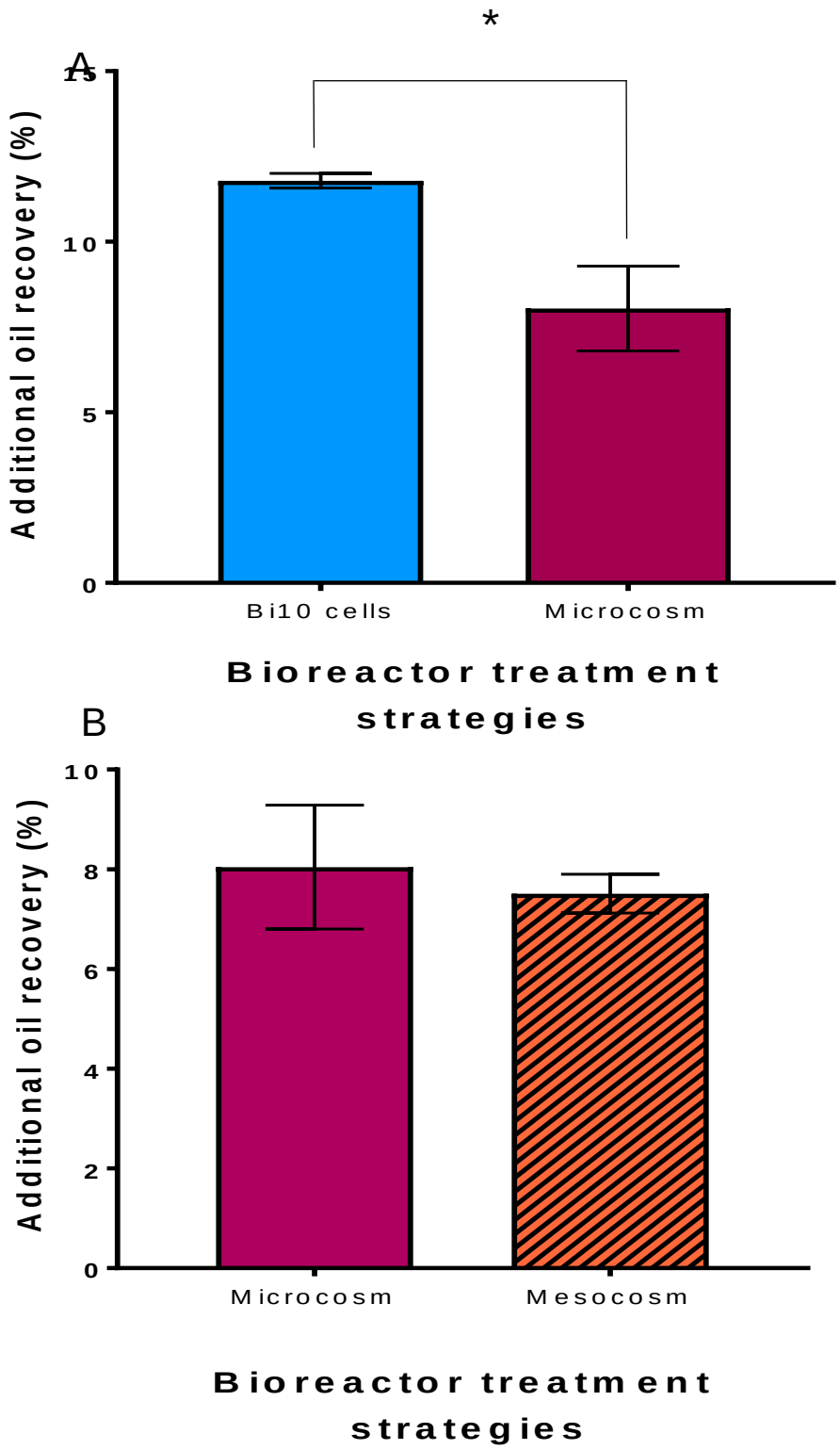


Figure 5.9 Application of MEOR using non-sterile bioreactor systems. Panel A shows the difference between *B. licheniformis* Bi10 cells inoculated into a sterile (left) and non-sterile (right) bioreactors. The scale up of operations in Panel B shows the similarity in AOR of bioreactor testing from a small microcosm (80 mL) to a larger mesocosm (1,500 mL) model system.

5.4 Discussion

A decisive factor in the successful investigation of laboratory simulated MEOR, is the ability to match oilfield parameters to that of the porous models (Gray *et al*., 2008). In a large proportion of laboratory MEOR trials, petrophysical parameters have not been adequately simulated (Table 5.1) |. This has led to a reduced reliability of the obtained results from the laboratory translating to success for field trials, and may even be directly attributable to their failure (Prakash *et al.*, 2012).

In Chapter 3, a number of petrophysical and chemical parameters were identified, *e.g.* FW brine chemistry for water flooding, crude oil type and the oilfield temperatures and pressures. Matching these parameters so closely to the Bentley field itself provides the ability to overcome many of the disadvantages associated with MEOR (Maudgalya *et al*., 2007). However, although the rock porosity and permeability of the sandstone reservoir is known, these parameters must also be accurately recreated. By analysing mixtures of sand of differing grain sizes, it was possible to replicate the reservoir porosity to within 0.5%. Furthermore, after conversations with the reservoir producers, it was determined that MEOR treatments would be applied once this reservoir was categorically depleted, a common practice for EOR treatments. Therefore, after careful analysis of water flooding, it was ascertained that the total flooding volume of 150 mL (post inoculation) would take the microcosms to a level of adequate depletion to mimic this characteristic of the Bentley Oilfield, too.

According to the preliminary results of the study, of the 8 isolates tested for oil recovery enhancement, the two species of *B. licheniformis* (C10 & C36) showed the highest increase in AOR. Furthermore, the strains of *B. subtilis* (C13 & C19) and *P.aeruginosa* (C82) all showed significant increases in AOR. However, use of the *E. coli* and *Acinetobacter sp.* isolates resulted the lowest AORs. These findings are not surprising given that in most studies, performed in both the laboratory and field, *Bacillus* sp. are commonly identified for use in MEOR due to their biosurfactant production and ease of culturing. Members of *Bacillus,* alongside rhamnolipid producing *Pseudomonas* strains were originally identified as the most promising mediators for MEOR (Desai & Banat, 1997). Although there have been some reported cases of *Acinetobacter* use in MEOR (Hui *et al*., 2012; Zou *et al*., 2014), due to its production of rhamnolipid-like compounds, the absence of data showing improved oil recovery and the reduced IFT alteration compared to *Pseudomonas* and *Bacillus spp.* has proven it as an unreliable resource (Hao Dong *et al*., 2016; Zou *et al*., 2014).

Furthermore, even with its robust, biofilm-forming potential, *E. coli* strains have scarcely been reported in the MEOR literature. Only a handful of preliminary studies exist concerning *E.coli* and MEOR, centred around molecular/recombinant technologies to clone various MEOR-related systems, *E.coli* transmembrane proteins for AOR and phage infection (Alvarez Yela *et al*., 2016; Chang & Fu Yen, 1984). Nevertheless, the lack of data for AOR of *E.coli* has pushed the MEOR research down alternate routes. Most MEOR literature focuses on organisms possessing rhamnolipid and lipopeptide

production pathways, as these are the most commonly studied. Knowing the importance of rhamnolipids and lipoprotein production to MEOR, alongside the abundance of literature available on biosurfactants of this nature, this project focused on characterising the *B.licheniformis C10* (or Bi10 as it became known) isolate for its ability to improve oil recovery by MEOR.

With the isolate Bi10 being chosen to undergo further testing, it was important to look at the total bacterial concentration that was to be inoculated into the bioreactor. If many bacteria are introduced into a microcosm or oil well, they we may promote well-bore plugging, or the blocking up of the injector well (Lazar *et al*., 2007). Likewise, the injection of too few bacteria and they will be unable to undertake their MEOR processes or may not become established in the oil well because they are outcompeted by other viable, autochthonous bacteria. By assessing different starting inoculations, we were able to reduce the starting concentration to ~ 5 x 10^7 cfu mL^{-1} microorganisms, whilst still recovering a similar level of oil. Reducing starting cell concentrations to below this value negatively impacted upon AOR, and as such reduces the effectiveness of the MEOR treatment. Allowing cellular growth in fresh media for 24 h in a flask caused either cell death and/or the sporulation of the Bi10 cells due to a number of factors, but primarily the reduced availability of nutrients. Although a reduced cell starting number is shown (Fig 5.3) after 24h incubation, the quantification of dead cells to spores was not calculated after pre-inoculate growth and as such, upon injection to the bioreactor, germination of spores may have taken place. This could be why 24h growth of Bi10

exhibited a similar outcome in terms of AOR to the reduced inocula. However, the effects of sporulation on MEOR will be assessed later on.

A primary concern for the utilization of non-indigenous MEOR treatment strategies is whether the microorganism will be metabolically active in such a diverse environment to where these strains may have been isolated, to drive the oil recovery process. Even if the cells are metabolically active, proliferate in the reservoir, and colonise the entire reservoir, their ability to act on the entrapped residual oil must be demonstrated either in the synthetic laboratory environment or as a field application.

When assessing the effectiveness of the Bi10 strain in comparison to other microorganisms that have shown an MEOR ability (Table 5.1), this study demonstrates that isolate Bi10 was able significantly increased AOR of oils classified as heavy (API <15 º). The recovery of 11.79% ± 0.2 is similar to observations made by other research groups on the effects of MEOR by *B. licheniformis*, conducted on oils with an API > 15º (Gao, 2011; Gudiña *et al*., 2013; Suthar *et al.*, 2009), lighter than oil produced from the Bentley Oilfield. Interestingly, although *B. mojavensis JF2* has facilitated the lowest oil-related IFT reduction ($<0.01 mNm^{-1}$) of all known MEOR bacteria, it did not perform as succesfully (10.768 ± 0.885) as the Bi10 strain. This could be due to the heavy nature of the oil (~12º API), with JF2 having been previously tested only on lighter oils (Mcinerney *et al.*, 1990). The Bi10 cells also produced significantly more oil than its crude cell-free biosurfactant containing culture liquor or the Bi10 spores. This, therefore, suggests that it is not the sheer

presence of surfactant alone that invokes the MEOR response measured in this project, but instead the MEOR response is driven by dual mechanisms between the action of added surfactant and proliferating cells, which are also capable of producing additional surfactants.

The quantitative PCR data and oil spread assays for the bacterial treated model microcosms, demonstrate that not only are microorganisms present in the produced fluids, but they are also metabolically active in the bioreactor. Interestingly, the reduction in IFT by both the Bi10 and JF2 cell injection treatments and presence of spores and cells in the post recovery effluent of these two treatments, show that cell proliferation and production of metabolic by-products has transpired. Yet the substantial reductions in 16S rRNA gene copy number from qPCR results suggest that although cells have progressed through the bioreactor, a large proportion did not make it to the production side after tertiary recovery. As 16S rRNA gene copy number shows no significant correlation to cell viability, it is apparent that had the Bi10 cells flowed through the reactor unhindered, without the occurrence of bioreactor colonization, there should be a similar, if not exponentially greater number of cells present to that of what was injected into the bioreactor (Aellen *et al.*, 2006; Blazewicz *et al.*, 2013).

One hypothesis, is that alongside oil recovery via surfactant production, a selective plugging mediated mechanism could have facilitated the increase in oil recovery due to microbial colonization of Bi10 within the bioreactor microcosm. This strategy utilises microorganisms that produce sizeable

quantities of biomass and EPS in extreme environments, to form biofilms to selectively block the reservoir pore channels. Selective plugging has been previously described in the literature only on a few of occasions and always restricted to strategies primarily focussed on microbial biostimulation (Klueglein *et al*., 2016; SutharI et al., 2009) . However, the plugging mechanism has proven to be more effective when performed by biosurfactants-producing bacterial strains, such as various members of the *Bacillus* (El-Sheshtawy *et al.*, 2016; Suthar *et al*., 2009). Biosurfactants act upon the rock surface like a cleaning agent and thus assist with biofilm adherence. The adherence of the organisms and establishment of a biofilm then forces the flow of formation water into other unexplored channels, allowing the free movement of biosurfactant to facilitate further recovery of oil. The AOR of the Bi10 cells compared to the surfactant alone also provides evidence supporting this hypothesis, as more oil is recovered under the stimulus of actively metabolising cells.

The difference between mean AOR mediated by Bi10 cells and Bi10 spores was similar (< 0.6% difference in AOR). Yet due to the variation in SD between the the Bi10 spore treatment recoveries, there was a significant difference between the Bi10 cells and Bi10 spore MEOR recoveries ($P < 0.05$). This, in hindsight, could possibly be due to the differential numbers of spores and cells being inoculated into the bioreactor, a feature that was harder to control for the spores. If the same number of spores, or future germinating cells, could be inoculated as cells in the Bi10 cell bioreactor, no significant difference in AOR may be seen between the two MEOR treatments. Further to this

hypothesis, the difference in time taken to produce surfactant, due to germination, metabolism and the start of surfactant production, could play a role in AOR difference. Although the total number of 16S rRNA copies decreases for the Bi10 spore treatment microcosm, the presence of spores and cells post recovery suggests that germination of the spores must have taken place. The mechanisms driving MEOR work similarly to that of the Bi10 cells. This is again indicated by the minimal difference between the AOR of Bi10 cells and spores. Spore inoculation for MEOR could therefore prove to be an effective treatment strategy, as spores are easier to maintain, transport, concentrate, store and inject into reservoirs than living cells. However, this cannot be wholly concluded due to the inability of this study to quantify spore counts. As a result, any variation in AOR and bacterial numbers identified could be down to a differential number of spores inoculated in the initial phase of the experiment. An accurate quantification would need to be conducted on the bioreactors at the start and end of the 'lock-in' period to enumerate total spores in comparison to total viable cells to assess if these differences are down to cell viability rather than spore germination.

The primary focus of MEOR biotechnology, once the concept has been proven, is to develop upscaling systems to adapt what has been learned in the laboratory to the field, with the same degree of success. The process of upscaling allows the succesful extrapolation of the lab based microcosm methods to be applied at the oil reservoir level, estimated to be anywhere up to 10^{10} times greater in size (Strappa *et al.*, 2004). Furthermore, it has been identified that certain core reservoir characteristics must be considered for

laboratory data to be practically and successfully upscaled. These key features are: the geometric and kinetic similarities, reservoir characteristics of pressure, porosity and chemistries, alongside oilfield temperature (Gianetto & Islam, 1993).

In this study, the reservoir characteristics and temperature were accurately simulated to mimic conditions found in the Bentley oilfield. By using isolated strains manipulated under their indigenous conditions, growth rates and metabolite production could be accurately controlled for. However, due to bioreactor size and sterility, the geometric and kinetic similarities were not wholly accounted for. To identify if their were as any effects of co-competition between injected cells and a colonised oilfield community, cells were injected into a non-sterile bioreactor. Although the communities in this model would not exactly match that of the Bentley oilfield itself, it was believed that the presence of a pre-existing community derived from Bentley formation water may have some effect on nutrient competition, niche occupation and growth rates. AOR was not as great, suggesting that when in competition for nutrient and potential biofilm production niches, pre-colonised rock surfaces will prevent optimal MEOR effectiveness. However, a significant quantity of oil was still recovered.

To assess the geometric similarity of our microcosm bioreactor, bioreactor up scaling was applied in the laboratory. By increasing the size of the bioreactor model by over 10 fold, no difference in AOR was detected. Therefore we determine that the microcosm data is a reliable testing source for larger scale reservoir simulations for Bentley Oilfield. Nevertheless, although marginally

upscaled, it cannot be assumed that this would still be representitve of a field trial. Further up-scaling, to bioreactors of even larger volmes or field specific computational modeling must also be tested for this to be better supported (Cheng *et al.*, 2014). Yet, in terms of the previous literature available and from the upscaling conducted, we propose that due to the enhancement of oil recovery by this bacterial strain (Bi10) in the microcosm model system , it is important to pursue the goal of an MEOR trial at the Bentley Oilfield.

Chapter 6

Whole genome and comparative genomics analyses of a biosurfactant-producing bacterial isolate from a heavy oil, offshore reservoir

6.1 Introduction

6.1.1 Background

Bacteria are able to synthesize a wide range of extracellular, metabolic products that possess numerous properties and industrial applications, such as uses in agriculture and food technologies, antimicrobial pharmaceuticals and wettability alteration within the oil industry (Galabova, Sotirova, & Karpenko, 2014; Pereira *et al*., 2013; Rincón-Fontán *et al.*, 2017). Of these products, various biosurfactants and lipopeptides are produced by differing bacterial genera, such as *Bacilli* (Simpson *et al.*, 2011; Zhang *et al*, 2016) *and Pseudomonas* (Alvarez Yela *et al*., 2016; Asshifa Md Noh *et al.*, 2012), that have multiple functions. These lipopeptide mediated functions range from mobility to surface tension reduction to facilitating microbial killing amongst other functionalities (Deleu, Paquot, & Nylander, 2008; Kowalewski *et al.*, 2006; Tedesco *et al.*, 2016). However, the bacterial genome is the driving factor in deciding on the actual function of the microorganism and its metabolic product.

With information now available to substantiate the existence of lipopeptide isoforms and differing gene families driving these surfactant production metabolic processes within the same bacterial genera, it is no longer

acceptable to assume the type of lipopeptide solely based on taxonomy and biosurfactant screening assays (Li, Yang, & Mu, 2010). Instead we must look to the process of whole genome sequencing and annotation to gain a better insight into the true individuality of an oilfield isolate and why it behaves in a different manner to its related strains, by identification of gene clusters such as the NRPS operon or *sfp* gene that regulate surfactant production (Martínez-Núñez & López, 2016; Nakano *et al.*, 1992).

6.1.2 DNA sequencing & technologies

Over the last 40 years the concept of sequencing and genomics has revolutionized the biological field. Although the concept of DNA and its structure has existed since the early 1950s (Watson & Crick, 1953) and the first genome, that of a ΦX174 bacteriophage, was sequenced in the 1970's (Sanger *et al*., 1977), it has only been since the advent of next generation sequencing and vast improvements in bioinformatics skill sets and computational power made that the use of sequencing has enabled the new science of comparative genomics to develop. With these major improvements, many feats in DNA analysis have been achieved, such as the completion of the Human Genome Project, the identification of cancer related genes and the understanding of bacterial virulence in human disease (Hood & Rowen, 2013; Sjoblom *et al.*, 2006; Smith *et al.,* 2012).

Long gone are the days of conventional, Sanger chain termination sequencing of short genomic regions. Instead most DNA sequencing is conducted on one of the many next generation sequencing (NGS) platforms. The three most

commercially used platforms are that of the Illumina MiSeq/HiSeq (Illumina, US), the Ion Torrent (ThermoFisher Scientific, US) and the now less commonly used 454 sequencer (Roche, SUI). The advantage of these machines are the improvements in throughput, sequencing mechanisms and also output, with Illumina capable of producing a matched pair of sequences from each end of the DNA fragment (paired-end reads), allowing greater coverage from short length reads (Metzker, 2010; Reuter, Spacek, & Snyder, 2015). Yet with the introduction of exceptionally long-read, single molecule, real-time sequencing from PacBio and the beta tested field capable MinION (Oxford Nanopore, UK), the capabilities and cost reduction of genome sequencing will continue to improve, expanding our sequence databses, knowledge and understanding of cellular, metabolic processes at a genomic level.

6.1.3 Whole Genome Sequencing

In a previous section (Chapter 3), the use of sequencing for the purposes of microbial community determination (relative abundance) and taxonomic analysis was discussed. This proved to be very effective in identifying the microbial consortia present from a hostile environment, in this case an oil reservoir. However, microbial community analysis is not the only use for NGS technologies. For this project, it was paramount to learn more about our oilfield isolate Bi10 by deciphering its genetic disposition. This could have been achieved using gene specific primer sets to amplify a gene loci of interest. However, this can be time consuming due to the multiple primer sets needed to identify each specific genetic trait; and ineffective if the sequences underpinning Bi10's MEOR capacity were unknown or varied between

microbial strains. Furthermore, issues with amplification or primer bias, as discussed in Chapter 1.5.3, may also render our results invalid (see Appendix A). Therefore, whole genome sequencing (WGS) was preferential in this scenario. WGS involves the sequencing of all extracted DNA from an organism and determining the nucleotide order of the whole sequence by computational sequence assembly.

Whole genome sequencing is the primary technique that was used in the Human Genome Project (HGP) (Hood & Rowen, 2013; NHGRI, 2015), however this was not the first organism to have its genome sequenced. In 1995, Fleischmann *et al* successfully sequenced the complete *Haemophilus influenzae* bacterial genome using multiple parallel Sanger sequencing on ABI (Fleischmann *et al*., 1995). This set a precedent for future DNA analysis. Since this date, over 26,000 complete genomes have been sequenced and are available to the public from NCBI, with over 8,000 of these identified as prokaryotic genomes (NCBI Resource Coordinators, 2017). Once the genome has been successfully sequenced, there is much computational work that must go into the analysis before it is uploaded to a database such as the NCBI (NCBI, US). The multiple reads that are returned, particularly from older sequencing technologies, must be carefully assembled and then annotated to produce a completed continuous genome sequence(Ekblom & Wolf, 2014).

6.1.4 Reference mapping vs. *de novo* assembly

To successfully assemble a genome, it is essential to correctly match the returned DNA fragments, or reads, in the correct order. In theory, this should

be far easier with the use of paired-end sequences such as those produced by Illumina sequencers where there is an overlap of sequences, or long sequence reads like those produced by SMRT sequencing (Pacific Biosciences, US). However, this is not always the case and even with the help of computational pipelines, assembly is a complex and time-consuming process.

There are two types of conventional genome assemblies. The first of these is the mapping of sequences to a reference genome. The basic principle of this technique is to align the reads to a known reference sequence that is closely related to the genome to be assembled. From this alignment, a consensus sequence can then be created, where the foundation of bases should match the reference. However, variation can be seen in sequences that do not directly match. Although mapping acquired sequences to a validated reference sequence is an effective method of assembly, it does have its flaws. Firstly, reference sequences do not exist for all known genomes. If the assembled genome does not happen to have a closely related relative, reads will either not align to a reference genome, or align incorrectly, forming gaps or breaks in the sequence contigs/scaffolds. Divergence in sequences between the reference and the assembled genome, which is common, can exacerbate this problem.

Furthermore, areas of sequence repeats within the reference sequence may cause problems. The assembler will be unsure of exactly which location to map the reads to, which normally results in the picking of only the area of

highest similarity, which is not necessarily the correct region (Trapnell & Salzberg, 2009). This may result in an underestimation of sequences, and thus an incorrect sequence alignment which may cause breaks or gaps in the newly assembled sequence.

To overcome these issues, an assembly method that removes the need for a reference sequence was conceived. This "*de novo"* assembly method uses the generation of overlapping reads to create an alignment within itself to form contiguous sequences, or contigs, where a gapless assembly of a number of length 'n's are produced. These contigs can then be aligned to each other to form a scaffold, based on overlap or sequence similarity. This approach nonetheless has its own requirements, such as a deep level of sequencing coverage required across the whole genome and high level computational power for assembly (A. Desai *et al*., 2013; Kisand & Lettieri, 2013). Without a deep level of sequencing there will not be sufficient overlap between contigs and breaks in the genome will form (Ekblom & Wolf, 2014). Therefore, to precisely assemble a gap free genome, it is suggested that a combination of both *de novo* and reference mapped assemblies are used to create the complete genome for annotation

6.1.5 Research hypothesis, aims and objectives

The main aims of this chapter were to successfully assemble sequences of the *B. licheniformis* isolate Bi10 genome into a completed genome. Using comparative genomics, the relationship between Bi10 and other *Bacillus spp*. were investigated. Once the comparisons were undertaken, the next step was to annotate the genome and identify how the Bi10 isolate differs from its other

Bacillus relatives in terms of metabolic pathways that may help improve oilfield survival or drive MEOR, with the focus particularly on surfactant mediated MEOR mechanisms. It was hypothesised that genetic variations within specific genes that are responsible for driving the surfactant production, which in turn will increase the effect of MEOR, can be recognised in the Bi10 bacterial genome.

6.2 Methods

6.2.1 Bacterial culture and DNA extraction

A single colony of Bi10 was picked from an LB agar plate and used to inoculate LB broth. Bi10 was cultured overnight at 37 °C with shaking at 200 rpm. This culture (100 μl) was used to inoculate 50 mL of MSM media which was then cultured for 18 h at 37 °C with shaking at 200 rpm. Total genomic DNA (gDNA) was extracted following a modified protocol from Griffiths *et al.*, 2000. (Section 2.) Following extraction and removal of contaminating RNA, DNA was quantified using the Qubit fluorimeter (Life Technologies) according to manufacturer's instructions

6.2.2 Library preparation for Illumina MiSeq genome sequencing

Libraries for sequencing were prepared by myself using the Nextera XT library preparation kit (Illumina) following manufacturer's instructions. Indexes 1 to 6 were used for each sample, respectively. The clean-up process was carried out using AmPure XP beads (Beckman Coulter) as per manufacturer's instructions. Part of the quality control was done by analysing the resultant DNA on a 1% (w/v) TAE agarose gel to assess the products for appropriate sizes (~ 300 bp). Samples from the 6 indexed libraries were then quantified using a Qubit fluorometer and pooled at the recommended concentrations for the sequencing provider. The completed Nextera XT gDNA libraries were then sequenced on the Illumina MiSeq sequencer platform by the Earlham Institute, Norwich, UK.

6.2.3 Genome assembly

Raw sequence reads were trimmed of sequencing adapters and quality filtered by the Earlham Institute, Norwich,UK. In-house quality control (QC) statistics were carried out upon obtaining the data, to identify and assess poor quality sequence reads. This QC included total read calculations and the analysis of the phred score, N content and sequence overrepresentation (section 6.3.1), using the Galaxy server wrapper FastQC (Afgan *et al*., 2016, available at www.usegalaxy.org). Galaxy is an open source, server based bioinformatics tool that houses numerous tools and pipelines for the analysis of next-generation sequencing data, including FastQC (Andrew, 2010), a tool for a comprehensive analysis of the quality of raw sequencing data. Once libraries had passed this quality control step, assembly of the genome was undertaken. A total of three assemblers, ABySS(Simpson *et al*., 2009), BWA (Li & Durbin, 2009) and SPAdes (Bankevich *et al*., 2012), were used and their outputs compared for both reference based and *de novo* assembly using the filtered forward and reverse fastq reads as input data (Figure 6.1).

The initial assembly was carried out against a reference genome using Burrows-Wheel Aligner (BWA) using the default BWA parameters (Li & Durbin, 2009). The resulting SAM files were evaluated using SAMTools (Li *et al*., 2009). The reference genome sequence used for comparison was from *Bacillus licheniformis*, strain ATCC 14580. For *de novo* assembly, SPAdes (Bankevich *et al*., 2012) and ABySS Simpson *et al*., 2009) assemblers were used. SPAdes was run using the K-mer sizes of 21, 33, 55 and 77. ABySS was run with ABySS-pe command with K-mer sizes for 40, 50, 60, 70 and 80.

For both *de novo* assemblies, scaffolding and alignment of contigs was carried out using SSPACE-BASIC (Baseclear). The resulting fasta files were visualized using Geneious 7 (Kearse *et al*., 2012)

Once assemblies were completed, their quality was assessed. This was achieved using the Quality Assesment Tool for Genome Assemblies (QUAST, Gurevich *et al*, 2013) to determine contig length, sequence gaps, genome coverage and N50 values. Alongside the statistics, to assess which assemblies would be used for annotation, a multiple sequence alignment of the three genome assemblies was carried out using Geneious 7 (Kearse *et al*., 2012) and progressiveMAUVE (Darling *et al.*, 2010). MAUVE, a genome alignment tool, produced a contig rearrangement and allowed us to identify which areas each contiguous sequenced matched on the reference genome. Sequence variations were documented and a consensus sequence produced for further downstream analysis.

6.2.4 Genome annotation

Annotations of the chosen assemblies were carried out using the Rapid Annotations using Subsystems Technology (RAST) pipeline for prokaryotic genome annotation (Aziz *et al*., 2008) using the classic RAST annotation scheme and a concatenated fasta format as the input file. RAST annotations use a high confidence assignment strategy based on the manually curated SEED database to accurately identify coding sequences, gene features and metabolic pathways, before categorizing them into the relevant subsystems. Once a consensus sequence was identified, RAST was used to match and

assign the closest related phylogenetic neighbours to the final annotated genome using gene-candidates from the SEED database, based on similarity score >99%. To create the phylogenetic tree, full length 16S rDNA gene sequences were acquired from RAST and NCBI and input into MEGA5. Further similar sequences were identified via NCBI blast using the megablast algorithm. All sequences were aligned using muscle on MEGA5 and the tree estimated with the neighbor-joining method. To confirm the reliability of the tree, bootstrapping was undertaken.

To maximize annotation and account for any genes missed from the SEED database, Prokka (Seemann, 2014) was also used for secondary annotation on the completed Bi10 genome. Prokka is an automated search tool for open reading frames (ORFs). Once ORFs are identified, Prokka converts the DNA sequences to protein sequences for BLASTp analysis. The sequences are then examined and compared to the reference databases, including Pfam (Finn *et al*., 2016), CDD (Marchler-Bauer *et al*., 2017) and the internal Prokka database.

To identify any regions within the genomes that may contain prophage sequences, each complete genome was run through PHAST (Zhou *et al*, 2011), a search and annotation tool that examines the genome for ORF or protein sequences that have a high similarity to prophage sequences. PHAST is also able to assess the completeness of each prophage and determine whether full or remnant prophage gene integration has occurred.

6.2.5 Comparative genomics

To assess the similarities and differences between the Bi10 genome sequence and the genomes of other similar *Bacillus* strains (*B. licheniformis ATCC14580,* WX02, BL09 & *B. subtilis* 168), Blast Ring Image Generator Version 0.95 (BRIG) was used (Alikhan *et al.*, 2011). By conducting a BLASTn of the 16S rRNA genes it was possible to visualize the dissimilarities between assemblies and also any intra-genera variation across the taxonomic group of *Bacilli*.

6.2.6 Identification of Biosurfactant production gene profiles

The profiling of genes related to biosurfactant production were investigated using the BioSurf database BLAST service (Oliveira *et al*., 2015). The BioSurf database offers a BLAST alignment of genes or whole genomes against the manually curated public database, which collects public and private information based on biosurfactant producing and biodegradative genes from metagenomes, metabolic pathways, biochemical data and bioremediation experiments. To identify any sequences that may not have been previously added to the BioSurf database, a search strategy based on data mining of any returned searches using the terms “Biosurfactant”, “genes”, “lichenysin”, “surfactin”, “*Bacillus*”, “genomics” and “lipopeptide” was conducted on NCBI Database (NCBI Resource Coordinators, 2017). Any returned results were manually searched for in the RAST and Prokka annotations.

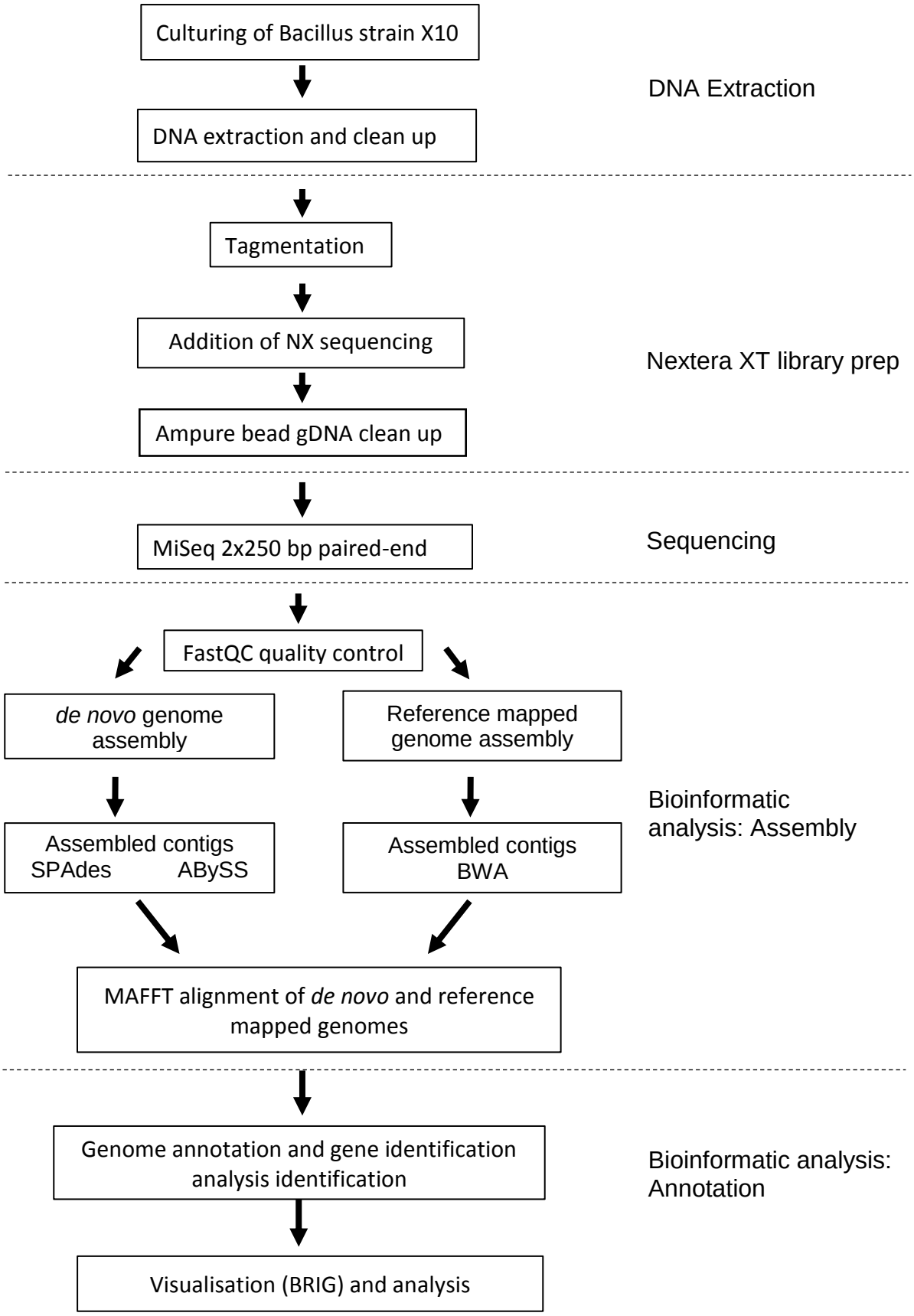


Figure 6.1. Flow diagram illustrating the work flow of sequential steps undertaken for genome assembly and annotation of isolate Bi10.

6.3 Results

6.3.1 Bioinformatic analysis and quality control of sequencing data

The sequencing output from the Illumina MiSeq run consisted of twelve files, each corresponding to 6 indexed samples of either forward or reverse reads. The raw sequencing reads were quality filtered by the Earlham Institute, UK, removing pairs of sequences with illegitimate bases in the reads and adapter sequences from the sequencing indices removed. Filtered reads were downloaded for downstream analysis by myself, following the bioinformatics section of the work flow (Fig. 6.1). A total of 4,341,548 paired-end reads of 350-500 bp in size, were generated from the sequencing of the genomic DNA isolated from the bacterial strain Bi10.

Each indexed file was uploaded to the Galaxy web server and run through the FastQC pipeline. Fast QC produces a detailed statistical analysis for each data file, showing mean quality phred score (error rate), GC%, the presence of ambiguous bases (Ns) and adapter sequences, amongst other statistics (Patel & Jain, 2012). A summary of these statistics are provided for each read file in Table 6.1. For each set of paired reads, the mean Phred score was calculated to be above 37. The number of ambiguous bases and remaining adapter sequences was determined to be below the 1.0% cut off for high quality reads and the GC content was 46% for all files. Furthermore, there were very few overrepresented sequences in the data (<0.8%) and a high proportion of sequences remaining after deduplication of identical reads (> 80%). Quality filtered paired end reads from the QC step were taken forward to the assembly

Table 6.1 FastQC sequencing statistics of Illumina Nextera XT libraries after sequence adapter removal and quality filtering

	Bacillus-genome_Bi10_S1		**Bacillus-genome_Bi10_S2**		**Bacillus-genome_Bi10_S3**		**Bacillus-genome_Bi10_S4**		**Bacillus-genome_Bi10_S5**		**Bacillus-genome_Bi10_S6**	
Read Direction	FW	REV	FW	REV	FW	REV	FW	REV	FW	REV	FW	REV
Total reads	423,122	423,122	464,779	464,779	902,058	902,058	1,017,018	1,017,018	724,054	724,054	810,517	810,517
Mean sequence quality Phred score	38	38	38	38	37	37	38	38	38	38	39	39
Per base N content (%)	< 1.0	< 1.0	< 1.0	< 1.0	< 1.0	<1.0	< 1.0	< 1.0	< 1.0	< 1.0	< 1.0	< 1.0
GC content	46	46	46	46	46	46	46	46	46	46	46	46
Adapter content (%)	< 1.0	< 1.0	< 1.0	< 1.0	< 1.0	< 1.0	< 1.0	< 1.0	< 1.0	< 1.0	< 1.0	< 1.0
Remaining sequences after deduplication (%)	89.8	90.5	89.15	90.03	83.68	85.17	82.77	82.77	85.06	85.75	84.6	85.39
Overrepresented sequences (%)	0.47	0.47	0.47	0.47	0.79	0.79	0.69	0.69	0.56	0.56	0.66	0.66

stage of the workflow. Three different assembly strategies were used as stated in section 6.2.3.

6.3.2 Assembly of metagenomic sequencing data

The initial assembly was undertaken using a reference-based mapping approach. The paired end reads were mapped to a published *B. licheniformis* genome of the strain ATCC14580 (GenBank accession number: CP000002; available at https://www.ncbi.nlm.nih.gov/nuccore/CP000002) using BWA. Of the 4.3 million paired-end reads, 3,842,083 (88.5%) reads were successfully mapped to the reference genome with a mean genome coverage of x 249 (Table 6.2). The GC content was similar to that of the reference genome (46.4%), however 3.5% of the assembly contained base ambiguities in the consensus sequence.

To create *de novo* assemblies, the same paired end reads were input into two differing *de novo* pipelines: ABySS and SPAdes. Although both are de bruijn graph assemblers they do their jobs differently. ABySS is a rapid parallel assembler, designed for short read sequences that works via a two-stage algorithm. The first stage is the building of all possible length k-mer sequences, from which errors are removed and initial contigs are built. To extend over gaps between contigs, paired end reads are then used in the second stage to resolve any ambiguities or gapped sequences between contig pairs (Simpson *et al*., 2009). SPAdes, however, is designed for use on single cell and massively parallel bacterial datasets, such as those provided by PacBio, Illumina and Ion Torrent platforms. SPAdes uses a four-stage assembly

algorithm of multiple sized k-mers. It was specifically designed to overcome common issues with conventional assemblers: the linkage of chimeric sequences and non-uniform sequence coverage (Bankevich *et al*., 2012). SPAdes is a more computationally heavy and slower pipeline than some of its counterparts (Jünemann *et al*., 2014). Both *de novo* assemblers produced scaffolds/contigs of a larger total size than that of the reference assembly and had similar GC content (Table 6.1). The ABySS assembly yielded a genome coverage of x300 with 97.6% of reads successfully mapped into contigs. This was far higher than that of the SPAdes assembly, which only gave a coverage of x93 and was only able to map 37% of reads to contigs. With regards to *de novo* assembly statistics, although the SPAdes assembly had higher N50 and NG50 values, SPAdes produced almost twice as many contigs as the ABySS assembly and also produced an incomplete scaffold when using SSPACE-BASIC.

Alongside these assembly statistics, to evaluate the accuracy and similarity of each assembly, the raw scaffolds/contigs produced by the assembler were passed through the RAST annotation server to identify the total number of gene subsystems recognized within the annotated genome. Each of the scaffold/contig sequences for each assembly were also aligned against each other and the *B.licheniformis* ATCC 14580 strain using the progressive Mauve aligner (Darling *et al*., 2010). The RAST annotation predicted that the SPAdes assembly had the highest number of coding sequences (4952), closely followed by ABySS (4885). The reference assembly possessed far fewer coding sequences (4320). However, by looking closely at each individual

annotation, specifically of the subsystem coverage, an interesting pattern emerges. Although there are more coding sequences present in the *de novo* assemblies, it appears that there are more similarities to CDS in subsystems between the AbySS and BWA reference assembly than there is between the SPAdes and BWA assembly (Figure 6.2). There are notably more RNA's identified in the SPAdes annotation (111) compared to that of the other two assemblies (ABySS: 94, BWA: 93). Furthermore, all subsystems that are involved in metabolism or amino acid production show similar subsystem coverage between BWA and ABySS. Only motility and sulfur metabolism have a comparable level of coverage between the SPAdes and BWA annotations

Alignments of the three assemblies using progressive Mauve shows the homology, or lack of, between certain contigs, after contig reordering had taken place (Figure 6.3). Mauve highlights local collinear blocks (LCB's) which are conserved areas of sequences that have not undergone rearrangement within the genome. These are identified by colour coding. The initial alignment shows a high degree of homology between the BWA (top) and ABySS (bottom) contig alignments. The reordering shows LCB coloured blocks located almost parallel to one another. This is not seen however for the SPAdes alignment, and although similar, LCB blocks appear throughout the reordered genome. The LCB positions in the scaffold of the SPAdes assembly sequence does not correlate to where the blocks are located on the other two assemblies. A consensus sequence of the ABySS assembly aligned to BWA and ATCC was acquired. This assembly was accepted as a true assembly primarily due to its

,

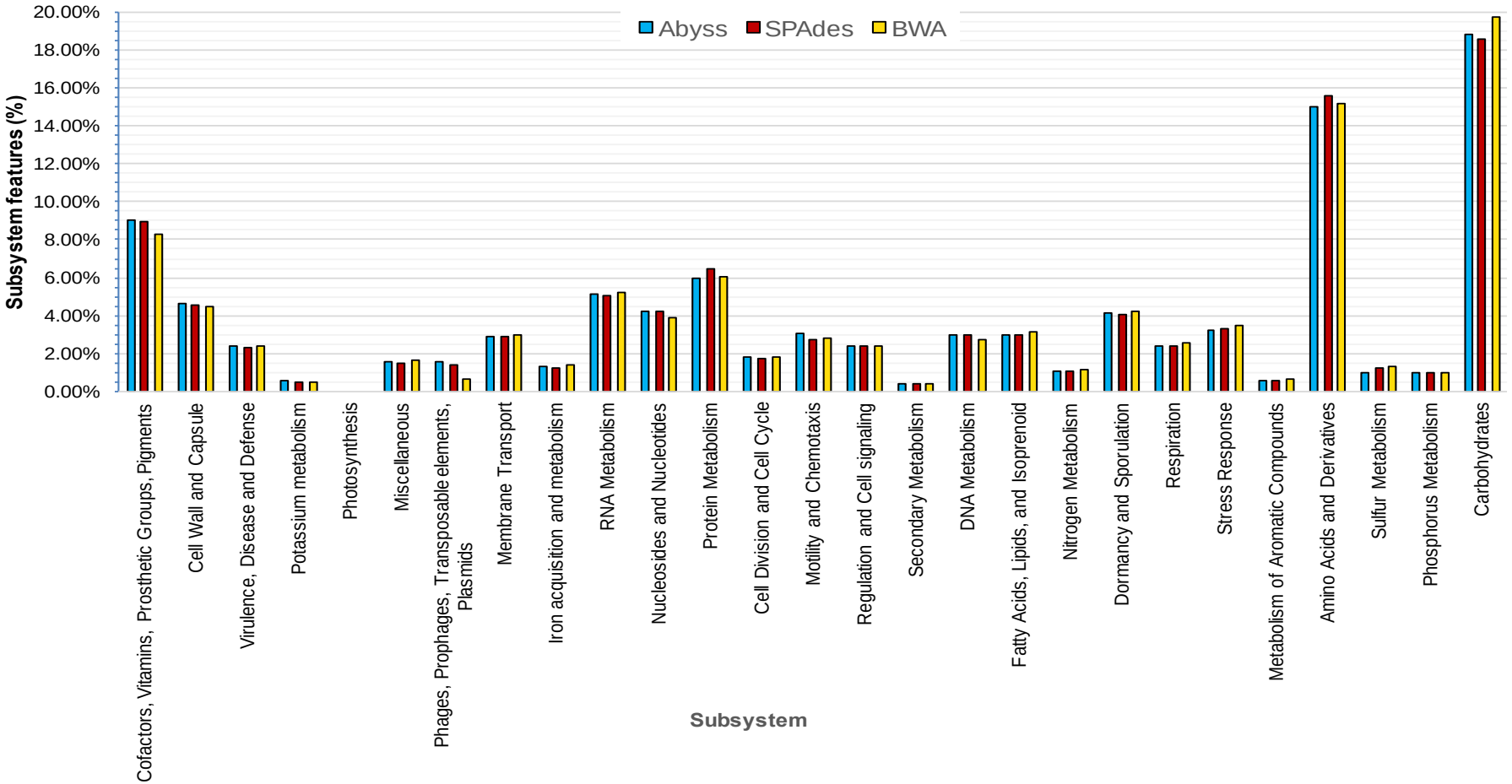


Figure 6.2 Functional annotations of assembled Bi10 genomes. Assemblies were conducted using three different assemblers: ABySS, SPAdes & BWA. All genome assemblies underwent RAST annotation into SEED categories, using the RAST annotation server. SEED categories were used for functionality assignment and are shown as a percentage of total genes assigned to subsystems.

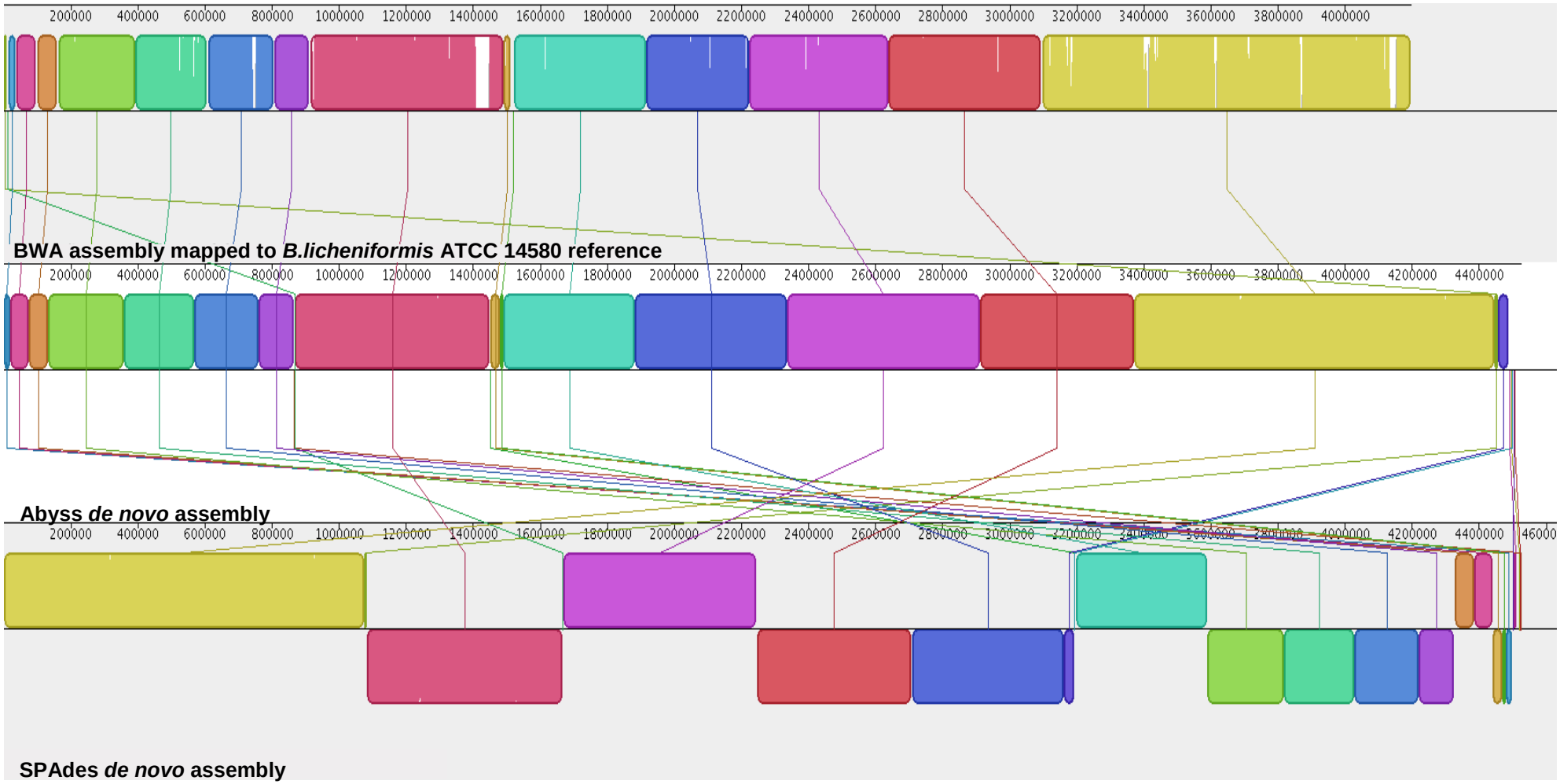


Figure 6.3 ProgressiveMauve alignment of the *B.licheniformis* Bi10 assembled genomes. Contigs resulting from the *de novo* and assemblies of ABySS (middle) and SPAdes (bottom) rearranged using BWA reference based alignment of Bi10 (top) . LCB block rearrangement undertaken using progressiveMauve, with defalt parameters

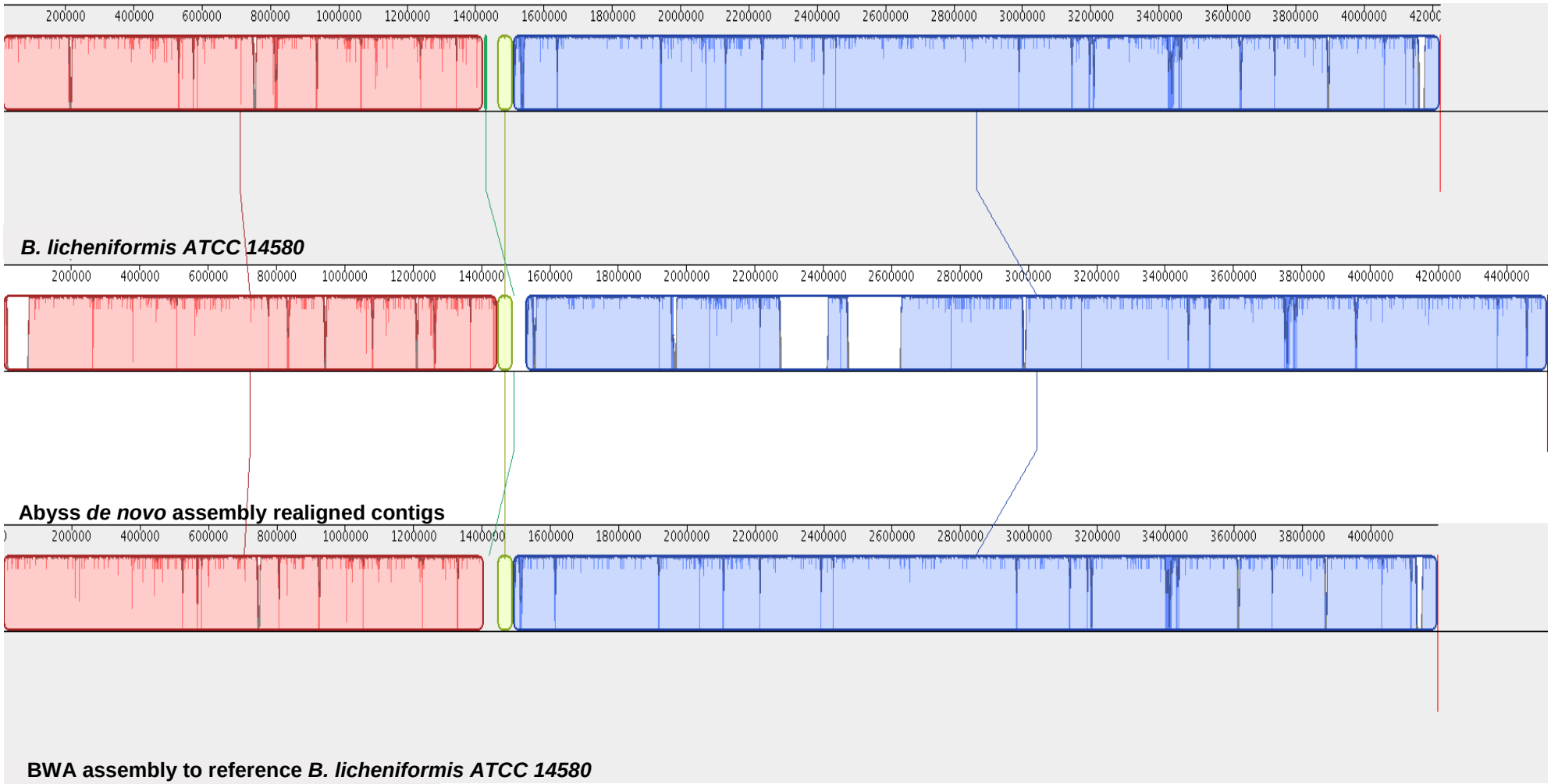


Figure 6.4 Progressive Mauve alignment of the completed *B.licheniformis* Bi10 assembled genome to reference based genome. Contigs resulting from the *de novo* assemblies of ABySS (realigned with BWA assembly) rearranged using ATCC14580 reference genome. Comparison of ATCC 14580 sequence (top) to ABySS assembly (middle) and BWA assembly (bottom). LCB block rearrangement undertaken using progressiveMauve, with defalt parameters

lack of sequence gaps, high coverage and accurate matching Mauve analysis to reference assemblies and the genome of ATCC14580.

When we align the BWA and ABySS reordered consensus assembly to that of the *B.licheniformis* ATCC 14580 strain, the accuracy of the assembliesparticularly the *de novo ABySS* assembly*,* are apparent (Figure 6.4). This un-gapped sequence was then exported into FASTA format and uploaded to both PROKKA and RAST for annotation.

6.3.3 Phylogenomic analysis of isolate Bi10 in relation to other *Bacillus* species

Using PCR and Sanger sequencing it was possible to amplify the 16S rRNA gene and promptly identify the species of bacteria that had been isolated from the oilfield (Chapter 4). However, the nature of using primers that only cover a specific region of the gene makes it difficult to identify distinct differences between species of the same genus. This is important, as by having this phylogenetic information, we are able to pinpoint the exact position of the isolate on an evolutionary tree. This information is of interest, as by knowing which strains of *Bacillus* are most closely related to Bi10, genomic comparative analyses between the oilfield isolate and other sequenced species can be conducted.

Sequence similarity analysis of the annotated 16S rRNA gene region using the NCBI *blastn* program (Available at https://blast.ncbi.nlm.nih.gov/) revealed the most closely related bacterial lineages to strain Bi10 (Table 6.2). Using the

Robust Phylogenetic Analysis tool (Dereeper *et al*., 2008, available at www.phylogeny.fr) an alignment of all *blastn* identified sequences was created and a phylogenetic tree was produced to show the evolutionary distance between Bi10 and other *Bacillus* species (Figure 6.5) based upon the 16S rRNA gene. The sequenced strains with the greatest 16S rRNA gene homology were that of *B.licheniformis ATCC14580, B.licheniformis BL1202* and *B.licheniformis WX02.* 20 other *B.licheniformis* strains were also identified as closely related to Bi10, alongside one *B. paralicheniformis* species.

Of these 21 strains, only four had full genomes available to download from the NCBI database. Fortunately, three of those were the ATCC14580, BL1202 (09) and WX02 strains, Bi10's closest evolutionary relations, as well as the *B.subtilis* 168 strain. It was therefore decided that these strains would be the comparators to drive the comparative genomic analyses upon which the rest of this chapter will focus.

6.3.4 Comparative genome analysis of *B. licheniformis Bi10* to others within the *Bacillus* genus

6.3.4.1 Genome comparisons using Blast Ring Image Generator

The genome of the *B.licheniformis* Bi10 strain sequenced in this study and the publicly available whole genome sequences of *B.licheniformis* ATCC14580, BL-09, WX02 and *B. subtilis* 168 were compared using the Blast Ring Image Generator (BRIG) software. Two comparisons were produced; one using Bi10 as the reference sequence; the second with the sequence of ATCC14580 serving as the reference (Fig 6.6A and 6.6B, respectively). The former

comparison (Fig. 6.6a) clearly shows high similarity between the Bi10 strain and its *B. licheniformis* relatives, especially strains ATCC14580 and WX02.

However, sequence variation between the Bi10 strain and *B. subtilis* 168 is far more apparent, with more gaps identified in the sequence comparison. Despite the genome similarities between the 3 *B. licheniformis* strains, there are four regions of noteworthy disparity (Fig. 6.6A). The first differential region of sequences is situated at the start of the Bi10 genome, located between ~50 and 100 kbp. The second region, situated between ~1,500 – 1,600 kbp of the Bi10 genome. The third fourth regions, both almost 200 kbp in size, are situated adjacently to each other, between 2,300 kbp and 2,700 kbp

When strain ATCC14580 is used as a comparison reference (Fig. 6.6B), there are far fewer regions of dissimilarity between all four *B. licheniformis* strains. Interestingly, again at 1,500 kbp there is a region of dissimilarity between ATCC14580 and the three other strains. However, there appears to be more similarity in these regions to the Bi10 isolate. The only other region of dissimilarity is at 4,200 kbp.

Table 6.2 Nucleotide blast results showing the similarity of top 20 Bacillus strains to Bi10 oilfield isolate. Data shows the sequence coverage (%), match identity of sequence to reference and corresponding E-value. Availability of the full genome sequence on NCBI is also highlighted in red.

Strain	% Identity	% Coverage	E-value	Genome available
B. licheniformis Bi10	100	100	0	
B. licheniformis BL1202 (09)	100	100	0	✓
B. licheniformis SKC B11	99	100	0	
B. licheniformis HRBL15TD17	99	100	0	
B. licheniformis SRC100141	99	100	0	
B. licheniformis SCCB37	99	100	0	
B. licheniformis SCDB14	99	100	0	
B. licheniformis BL14	99	100	0	
B. licheniformis ATCC14580	99	100	0	✓
B. licheniformis WX02	99	100	0	✓
B. licheniformis SB3180	99	99	0	
B. licheniformis SRCM1000027	100	99	0	
B. licheniformis DSM 13	99	100	0	
B. licheniformis SCDB34	99	100	0	
B. subtilis WE2	99	99	0	
B. subtilis JM4	99	99	0	
B. sp. H151	99	100	0	
B. subtilis 168	99	100	0	✓
Bacillus. sp. 1s1	99	100	0	
B. paralicheniformis ATCC9945	99	99	0	
B. paralicheniformis MDJK30	99	99	0	

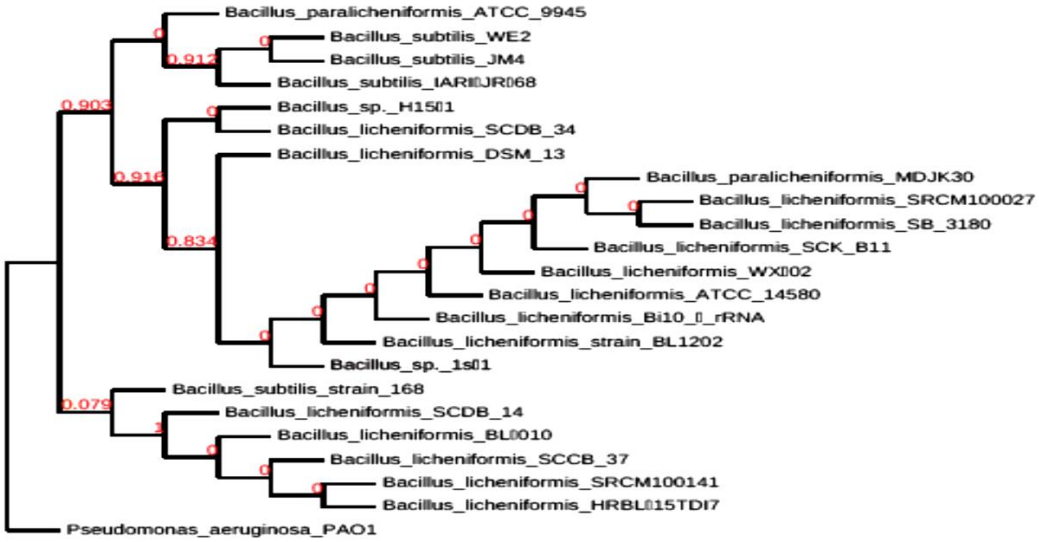


Figure 6.5 Bacterial phylogeny of *Bacillus* species 16S rRNA. Tree shows evolutionary distances between top 20 *Bacillus* species that showed the closest match to *B.licheniformis Bi10* isolate. The tree is rooted against *P.aeruginosa PA01* for comparative purposes.

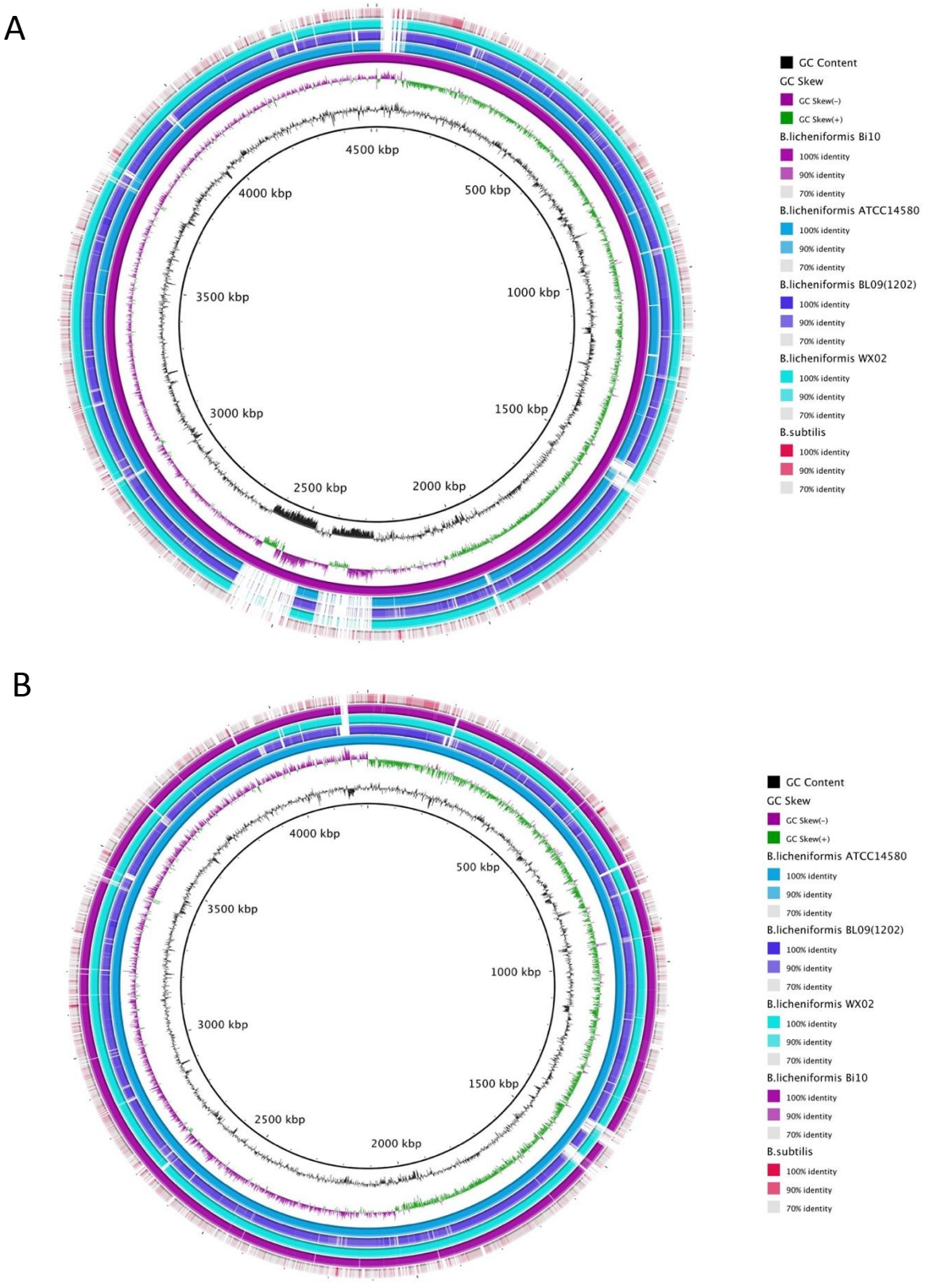


Figure 6.6 BRIG analysis of *Bacillus* genome sequences. BRIG compares whole genome sequences of three *B. licheniformis* and one *B. subtilis* strain from the NCBI database and the *B.licheniformis Bi10* isolate identified in this study. Panel A) Genotype Bi10 was used as the reference genome. Panel B) ATCC14580 was the chosen reference genome. Each individual genome corresponds to the coloured ring shown in the legend.

Genomic analysis using the PHAST tool revealed the presence of four intact prophages, five remnant and one cryptic prophage in the Bi10 isolates genome (Fig. 6.7a). The number of phages found in this whole genome sequence (10) is vastly increased in comparison to the three other *B. licheniformis* strains, all of which contain three complete prophages and fewer than three remnant prophages (Fig. 6.7). All four *Bacillus strains* carried the prophage identified as Bacilli_phi105 (IV). The prophage known as Brevib_jimmer1(X) was also identified in the Bi10, ATCC14580 and WX02 genomes, and Penib_fern was present in both the Bi10 and ATCC14580 genomes (II). The remaining intact prophage, Bacilli_vB_BanS_Tamsa (IX), located at ~2.5mbp-2.6mbp, and all the identified remnant prophage sequences were unique to the Bi10 isolate.

The intact prophage sequences of Bi10 (II, IV, IX & X) each contain 50, 30, 90 and 70 CDS, respectively. Homology to characterized phage genes is limited, with only a proportion are attributed to recognized prophage sequences. The majority of the CDS identified in the phage regions were reported as hypothetical proteins or non-characterized genes by PHAST. However, all essential phage elements, *e.g.* integrase, protease and terminase domains, host-receptor binding sites, portal proteins and other phage-like proteins (PLPs) were identified in the complete prophage regions. (Fig. 6.8). The incomplete or cryptic prophage sequences encompassed predominantly PLPs, attachment sites for site-specific integration and phage portal proteins. Most other important structural and regulatory genes associated with phage

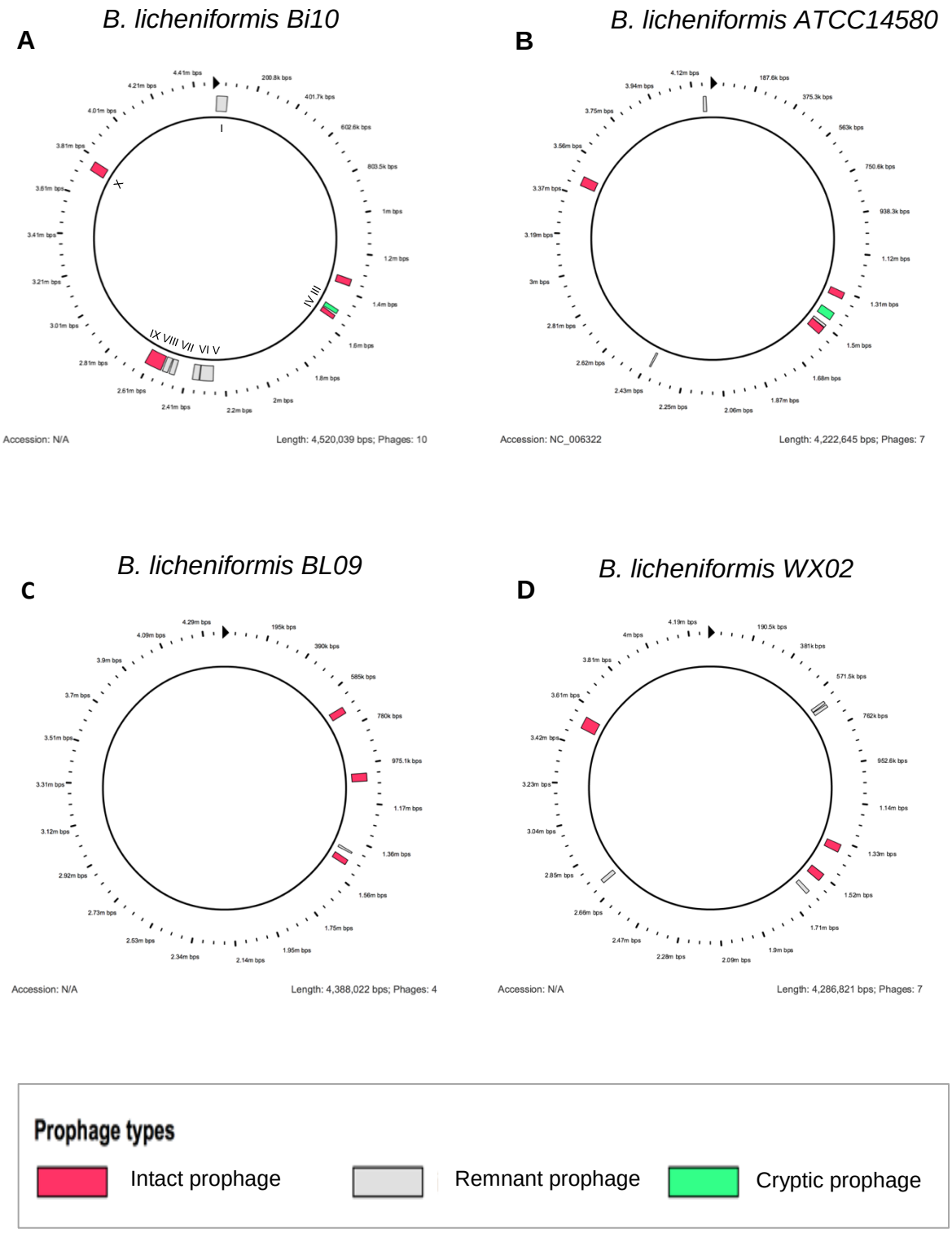


Figure 6.7 Prophage sequences predicted by PHAST for *B.licheniformis* genome sequences. Prophage analysis within the genome sequences of strains of the *bacillus* genus was performed using PHAST. Each panel shows the predicted prophage coding regions for the four individual strains Bi10 (A), ATCC14580 (B), BL09 (C) and WX02 (D). Regions are highlighted based on prophage type as shown in the figure legend.

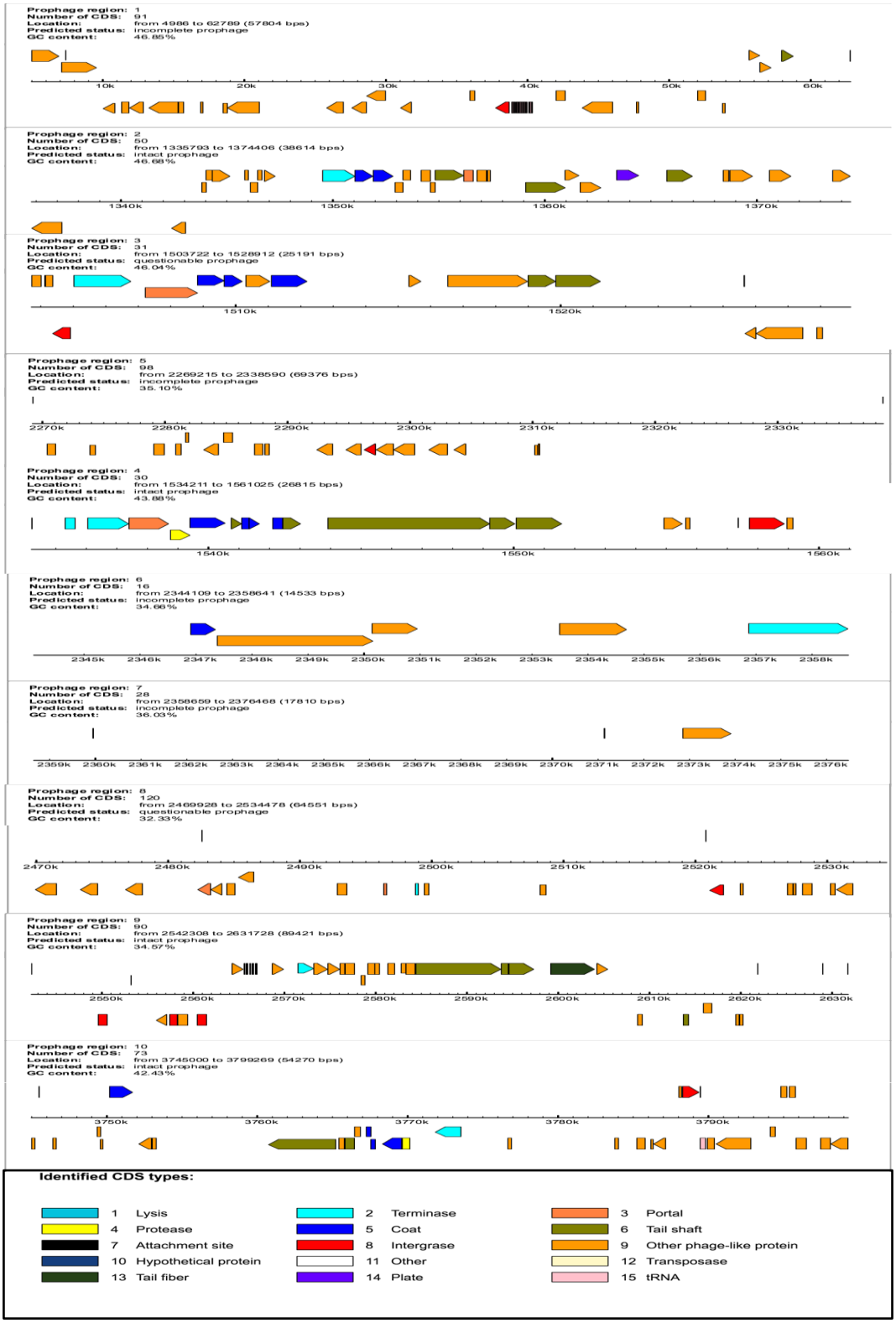


Figure 6.8 Predicted structures of prophage sequence identified in *B. licheniformis* Bi10 genome. Prophage analysis within theBi10 genome was performed using PHAST. Figures II, IV, XI and X show intact prophage sequences. I, V, VI, VII and VII partial or remnant prophages and III relates to a questionable prophage. Each CDS type is identified as a colour according to the figure legend and directionality shown via the arrow of each CDS.

replication were missing from the sequences, or could not be predicted, in this genome.

6.3.5 Genome annotation and gene prediction of *B. licheniformis Bi10*

After initial analysis of contiguous sequences via RAST (Section 6.3.2), it was apparent what the genome would look like upon annotation of the fully assembled sequence. The Bi10 sequence that was taken from the ABySS and BWA assemblies in Section 6.3.2, was then used as the input for gene annotation. This was conducted again using the automated web-based server, RAST, and the local pipeline PROKKA to give two annotations based on differing mechanisms (Seemann, 2014).

Using the RAST annotation server a graphical analysis of the distribution of genes or coding sequences amongst the subsystems was obtained (Fig 6.9). RAST identified 4885 coding sequences from the assembled Bi10 genome, over 3000 of which were deemed to be non-hypothetical genes. By analyzing the data, it is evident that almost half of the coding sequences were fitted into a subsystem, with over 2000 non-hypothetical genes assigned to a subsystem. The most abundant subsystem is that associated with *Carbohydrates*, closely followed by *Amino Acids and Cofactors*, *Vitamins*, *Prosthetic groups and pigments*. There is also evidence that 49 predicted sequences are prophage related genes, an outcome that further supports the data obtained in section 6.3.4.2. The results for the PROKKA annotation were also similar. Out of the 4,848 total annotated gene sequences in the genome, 2832 were determined to be non-hypothetical coding sequences. PROKKA also identified 91 RNA sequences, including 86 tRNA and 4 rRNA sequences.

In comparison, RAST identified 94 RNA sequences, with 86 being assigned to tRNA. Although this information is useful to assess the different genome annotation tools and the total number of genes and their related functions, this information alone does not answer why Bi10 may behave differently to other strains within the genus with regards to MEOR. At this point we used comparative genomics to help find unique attributes carried by Bi10 when compared to the genes carried by the three other *B. licheniformis* (ATCC14580, WX02 & BL09).

By comparing the RAST annotations of the ATCC14580, WX02 and BL09 to the annotation of Bi10, it was possible to assess the commonality between the four genomes and the presence of genes that differentiate between the reference strains and our oilfield isolate. Of the total number of genes annotated, 4102 were found to be shared between all four organisms, appearing to represent the *Bacillus* pan-genome (Figure 6.10). A further 310 genes were present in the Bi10, ATCC14580 and WX02 strains, and 390 found in only Bi10 and ATCC14580. A higher degree of genome variation was seen when comparing Bi10 to WX02 and BL09, with only 18 and 5 sequences shared with Bi10. By focusing specifically on the distribution of genes into the subsystems and comparing the Bi10 isolate to the remaining genomes we can notice these discrepancies more clearly. The three type strains have a relatively similar range of genes in each of the groups, with a few minor exceptions. When assessing the subsystems of Bi10, we can see that some subgroups are substantially different to its other *Bacillus* relatives (Table 6.4). For instance, the number of genes in the *Cofactors, vitamins, prosthetic*

groups & pigments category is greater for Bi10 (291). Predominantly, *Folate and pterine* associated genes accounted for the increase in genes within this subsystem. Similarly, far more prophage related genes are identified in Bi10 (49). Strain WX02, the next most abundant genome for prophage sequences, has only 31 genes in the prophage related subsystem. Although there is some intraspecies variation between genes in subsystems of the four genomes, with exception to those mentioned above, the Bi10 genome does not stand out as being markedly different in any of the subsystem categories to the three comparative *Bacillus* genomes, ATCC14580, WX02 and BL09.

6.3.6 Genes associated with Biosurfactant production

The annotated Bi10 genome sequence was analysed for potential genes of biosurfactant production pathways. Using BioSurfDB alongside the RAST and Prokka annotations, a total of 20 genes potentially involved in surfactant biosynthesis for the Bi10 strain were identified (Table 6.5). Of these 20 genes 15 were identified using the BioSurf database and 5 from the RAST and PROKKA genome annotations. Eleven of the identified genes are likely to be directly involved in the synthesis of a biosurfactant molecule or non-ribosomal peptide (NRP). The remaining 9 genes encode transcriptional regulators, RNA polymerase enzymes and a chaperone protein. As described in the literature, the most essential nonribosomal peptide (NRP) operon for the production of biosurfactant in the Bi10 genome was that of *LchA,* a NRPS gene cluster encoding a protein that directs the synthesis of the *lichenysin lipopeptide*. The four subunits of the *LchA* operon *(LchAB-AD*) combine to synthesize the lichenysin lipopeptide, the potent biosurfactant produced by

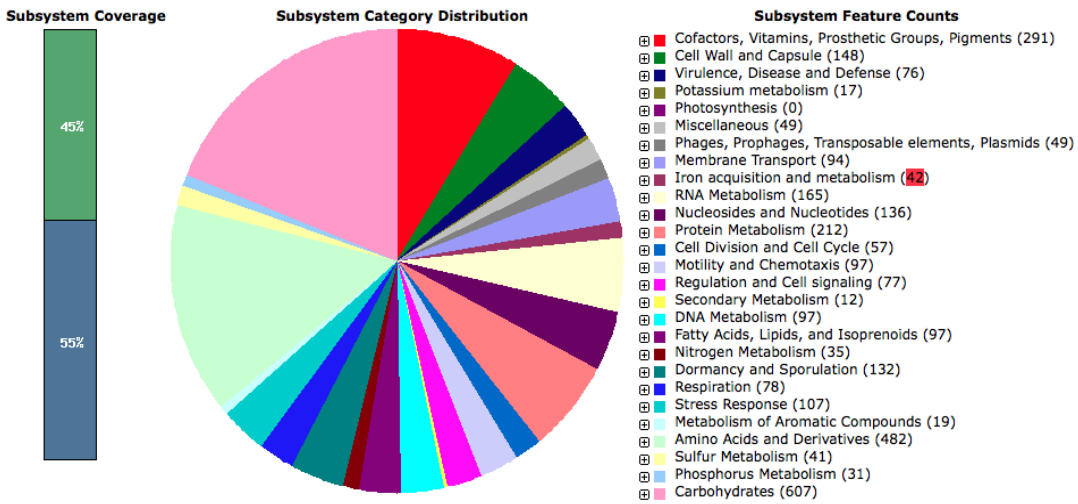


Figure 6.9 Distribution of coding sequences of *B. licheniformis* Bi10 separated into subsystems. Gene prediction was performed using RAST annotation server. The left hand bar chart shows the subsystem coverage from the whole genome annotation Each segment of the pie chart (right) shows the proportion of the categorized whole genome involved in each subsystem, with non-subsytem classified coding regions removed from analysis.

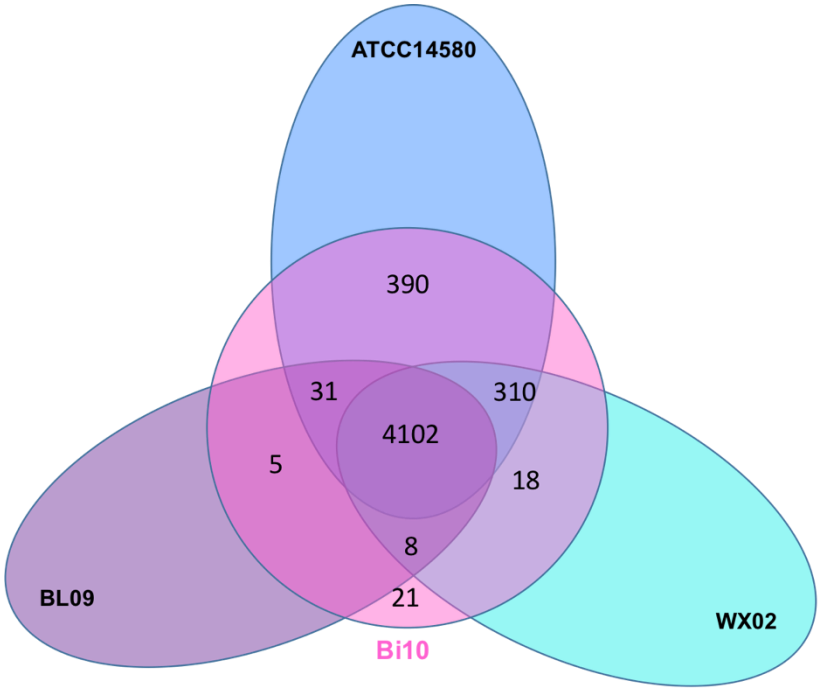


Figure 6.10 Comparison between *B. licheniformis* RAST predicted gene sequences against Bi10 isolate A Venn diagram showing the shared genes of each of the three species in comparison to the Bi10 isolate. Each overlapping segment on the Venn diagram relates to the overlap of gene sequences between the corresponding genomes. RAST annotation conducted on all four *Bacillus* genomes (Bi10, ATCC14580, BL09 & WX02)

Table 6.4 Comparison of total gene number assigned by RAST to each subsystem, for each of the four B. licheniformis genomes

Subsystem	Bi10	ATCC 14580	WX02	BL09
Cofactors, Vitamins, Prosthetic Groups, Pigments	291	269	256	248
Cell Wall and Capsule	148	118	136	146
Virulence, Disease and Defense	76	72	68	74
Potassium metabolism	17	19	20	21
Photosynthesis	0	0	0	0
Miscellaneous	49	66	49	50
Phages, Prophages, Transposable elements, Plasmids	49	6	31	23
Membrane Transport	94	90	92	91
Iron acquisition and metabolism	42	42	41	44
RNA Metabolism	165	176	163	161
Nucleosides and Nucleotides	136	123	118	112
Protein Metabolism	212	231	188	177
Cell Division and Cell Cycle	57	53	55	55
Motility and Chemotaxis	97	67	91	89
Regulation and Cell signaling	77	48	78	80
Secondary Metabolism	12	12	12	11
DNA Metabolism	97	126	88	106
Fatty Acids, Lipids, and Isoprenoid	97	87	99	102
Nitrogen Metabolism	35	27	35	46
Dormancy and Sporulation	132	130	132	131
Respiration	78	88	79	67
Stress Response	107	108	106	111
Metabolism of Aromatic Compounds	19	16	20	20
Amino Acids and Derivatives	482	445	484	519
Sulfur Metabolism	41	40	42	41
Phosphorus Metabolism	31	24	30	30
Carbohydrates	607	683	613	607

B. licheniformis. The four subunits of the *Lch* A gene, *A*-*D*, are located consecutively on the genome and share homology to the *surfactin* and *iturin* gene families of *B. subtilis* strains. *LchA*, *C* and *D* both possess a high similarity of amino acid sequence (> 99%) to that of other *B. licheniformis* strains. Interestingly, the larger subunits of *LchAA* and *LchAB* have a lower similarity, of ~73% and 86%, respectively. This suggests that although they are present in the genome, the sequence carried by Bi10 is different from those produced by other *Bacillus sp.* The remaining biosurfactant synthesis genes all had a similarity of between 70-92% to genes found in species related to the *Bacillus* genera.

Table 6.5 Biosurfactant related genes identified from BioSurf database and genome annotations based on literature search.

Gene	Functional product	Start	End	BioSurf reference	Species	Similarity (%)
Identified from BioSurf DB search						
alnB	Peroxiredoxin	4454956	4455501	GI:45826511	*Acinetobacter radioresistens*	*72.84*
llchAA	Lichenysin synthetase A	414534	423651	GI:45826511	*Bacillus licheniformis.*	*72.84*
bamC	Bacillomycin	417175	417253	GI:1834380	*Bacillus licheniformis*	*92.86*
lchAB	Lichenysin synthetase B	425275	433337	GI:610705801	*Bacillus licheniformis*	*86.25*
lchAC	Lichenysin synthetase C	435021	438869	GI:3029845	*Bacillus licheniformis*	*99.62*
lchAD	Lichenysin synthetase D	438873	439634	GI:3029681	*Bacillus licheniformis*	*99.97*
lpa-14	lpa-14 gene encoding lipopeptide antibiotics iturin A	442245	442494	GI:16053517	*Bacillus licheniformis*	*90.05*
sfpO	Biosurfactant biosynthesis protein (sfpO) gene	442246	442325	GI:432356	*Bacillus subtilis*	*74.51*
ppsA	Plipastatin synthetase subunit	2117102	2118021	GI:524853056	*Bacillus licheniformis*	*83.75*
ppsB	Plipastatin synthetase subunit	2118106	2119132	GI:363747653	*Bacillus subtilis*	*70.73*
ppsC	Plipastatin synthetase subunit	2119344	2120893	GI:363747653	*Bacillus subtilis*	*73.12*
dnaK	Chaperone protein dnaK	2954418	2954878	GI:363747653	*Bacillus subtilis*	*73.24*
arfA	Putative ABC transporter protein	4040243	4040409	GI:60549561	*Pseudomonas putida*	*72.44*
lgrD	Lichenysin synthetase	4043470	4043511	GI:29501264	*Pseudomonas sp*	*75.29*
liaR1	Transcriptional regulatory protein	1019697	1020347	GI:12550509	*Paenibacillus polymyxa*	*88.1*
Identified from RAST/Prokka annotations and from literature search						
sigA	RNA polymerase sigma factor A	2927161	2928282			
comA	Transcriptional regulatory protein	3527458	3528069			
liaR2	Transcriptional regulatory protein	3653767	3654402			
rpoN	RNA polymerase sigma factor 54	3825584	3826897			
dhbF	Dimodular nonribosomal peptide synthetase	4038926	4046083			

6.4 Discussion

The genome sequencing of the *B. licheniformis* Bi10 isolate using the MiSeq NGS sequencing platform produced a complete, single contig, full length, assembled genome. After removal of poor quality and undesirable sequences from the raw datasets, all 12 read files were above the Phred score threshold (37) to be considered for assembly. The Phred scoring system is centered on the logarithmic relationship of the sequence and base calling error rates within that sequence to predict the likelihood of calling a base incorrectly. Therefore, a Phred or Q score of 20 equates to a 99% chance that the base call is accurate (Cliften, 2015) .A Phred cut off of 30 was assigned to our reads and after filtering, all had a mean score of over 37, indicating a > 99.9% likelihood of sequence accuracy. For genome assembly, a number of different assembly methods were trialled, for both reference based and *de novo* sequence assembly, The main aim of conducting *de novo* assembly alongside reference based genome mapping is to produce an accurate replication of the true genome, irrespective of similarity to the reference sequence of choice (Otto, 2015). In this scenario, although a number of complete reference sequences of similar ancestry exist, we cannot be certain that there will be no critical dissimilarities in genome, especially since Bi10 was isolated from an extreme environment.

Comparison of the preliminary annotations of the assemblies showed that the ABySS assembler proved to be the most accurate based upon a combination of sequence gaps, N50 score, GC% content and alignment with another *B. licheniformis* genome (ATCC14580) (Fig 6.3). Not only did ABySS produce

the fewest number of contigs out of the two *de novo* assembly methods, but also produced a gapless sequence upon reordering and alignment of the contigs/scaffold sequences to the reference genome- a feat neither SPAdes or BWA could achieve. The creation of an intermittent or broken assembly can result in the inability to bridge gaps within the sequence, which are caused by indels or low complexity repeated sequences (Higgins *et al.*, 2005; Schatz*et al.*, 2010). The breakage in assembly introduces the need for additional laboratory work involving PCR-based gap closure. The *de novo* assemblies both had a higher total genome length, suggesting that there were some transgenic regions that differ from the reference sequences that were not accounted for in the BWA assembly. This is shown by the substantial decrease in coding sequences found in the BWA RAST analysis. Although the N50 and NG50 values were lower for the ABySS assembly than that of SPAdes, the far superior coverage of the ABySS assembly should compensate for this in terms of assembly quality against contig size. Nevertheless, by aligning the ABySS draft genome against the BWA assembly and the reference genome ATCC 14580, a clear association between the ABySS assembly and the BWA reference assembly and ATCC14580 sequence is evident from the Mauve comparison (Fig. 6.4).

The consensus sequence taken from the ABySS assembly aligned to the BWA assembly and reference genome became our completed genome of *B.licheniformis* Bi10. This sequence was annotated using the RAST and Prokka annotation pipelines (described in 6.2.5). Both pipelines produced a similar output, with Prokka identifying 4848 and RAST 4885 genes, and both

assigned just over half of the identified genes as non-hypothetical proteins. Both pipelines were also able to identify a similar number of non-coding RNA sequences, with both annotations detecting 91 RNA sequences: 86 tRNA, 4 rRNA and 1 tmRNA. These non-coding RNA (ncRNA) sequences do not produce a translated protein like a CDS, but instead produce a functional RNA product that assists with gene regulation. For this reason, ncRNAs are very useful for quality control analysis of bacterial genome assembly and annotation. Due to the evasion of translation and a higher evolutionary incidence, many ncRNAs have a low complexity primary sequence, making it difficult to successfully align using conventional programs (Bussotti, Notredame, & Enright, 2013; Sun *et al.,*, 2012). The number of ncRNAs identified from Bi10 is similar to that of some environmentally isolated *Bacillus* reference strains (Chaudhry *et al.*, 2016; Gkorezis *et al*., 2016), although there is great variation in the number of ncRNA sequences (80-121 ncRNAs) between strains (Lee *et al*., 2017; Ostrov *et al*., 2015; Rey *et al*., 2004; Veith *et al*., 2004).

The existence of repeated elements in the prokaryotic genome and multiple copies of the rRNA gene operons can obstruct accurate assembly in these regions (Wetzel, Kingsford, & Pop, 2011). In the Bi10 assembly, 4 rRNA (two 5S, one 16S and one 23 S) sequences were identified. This is four times lower than the number of rRNAs identified from the ATCC14580 strain, signifying that the assembly of Illumina short read sequences might have underestimated the presence of rRNA operons in the Bi10 genome. To overcome this issue, it has been suggested that the use of single molecule

real-time sequencing from PacBio may be an appropriate option. However, due to financial constraints (PacBio was too expensive) at the time of the project, Illumina MiSeq was the best technology available. Nevertheless, our results do appear to be comparable to results from other WGS approaches for organisms highly related to Bi10, with regards to GC content, size and annotated genes (Gkorezis *et al*., 2016; Lee *et al*., 2017; Rey *et al*., 2004; Wu *et al*., 2013).

Of the paired reads that did correctly assemble, we were able to identify the full length 16S rRNA gene. As described in Chapter 3, the 16S small ribosomal subunit rRNA has desirable properties that make it an excellent choice for phylogenetic characterization. As such, we extracted the 16S rRNA gene sequence from the assembled genome and conducted a BLASTn search to identify Bi10's closest, matching relatives. From the top 20 Blast hits, all of which had a similarity and coverage of > 99%, Bi10 sat within its own clade. However, 8 other strains, 6 of which were *B. licheniformis* were so closely related that the diversity scores were <0.0001. Of these strains, *B. licheniformis ATCC14580, BL1202 (09) and WX02* were the closest ancestral matches, all of which had complete GenBank genome sequences available, according to RAST. It was therefore decided that these three strains would be used for the comparative genomic analyses undertaken in this chapter.

Comparison of the Bi10 genome sequence, at a nucleotide level, against the three *B. licheniformis* reference strain genomes highlighted regions of distinct variation across the four genome sequences. However, after analyzing the

four genomes through PHAST, it is apparent that these ‘gaps’ are caused by the presence of prophage sequences that are exclusive to the Bi10 genome. The integration of prophage sequences accounts for the gaps produced in three regions of the Bi10 genome; 4-62 kbp, 2,260-2,370 kbp and 2,460-2,630 kbp. These prophage sequences are a combination of both complete and incomplete sequences, with some of the regulatory and structural genes missing from the integrated remnant prophage sequences. It is apparent that there is more homology between the Bi10 strain and ATCC1450, than there is between Bi10 and BL09 or WX02. This is further supported by the analysis of the core genome in the Venn diagram (Fig. 6.9), which is no surprise, as phylogenetically, ATCC 14580 was estimated to be the closest relative to that of Bi10. In comparison to the reference strain ATCC 14580, BL09 and WX02 shared more sequence homology with each other than they did with Bi10.

When analyzing the core genes present across all four *B. licheniformis* strains, or the Pan genome, it is apparent that some genes must play a conserved role in cell survival and proliferation, with ~84% of the total genome shared across all four genomes. The majority of these core genes (~80%) were able to be assigned to the subsystem Carbohydrate and Amino Acid Metabolism. A large proportion of genes were also categorized into the Protein Biosynthesis (212), Dormancy & Sporulation (132) and Cofactors subgroups (291). These findings are similar to what has previously been reported for other species with the genera, *Bacillus*, including that of *B. licheniformis* The accessory *B. licheniformis* Bi10 genome (or genes that appear in some but not all *Bacillus* strains) consisted of ~750 genes, all of which were scattered among

subsystem distributions depending on the strains carrying them. However, 21 Bi10 unique genes (or singletons) were identified. All of these genes were associated with prophage integration or classified as hypothetical proteins. We therefore know that the integration of the prophage at these locations is a distinct event that may be a result of environmental pressures, rather than an evolutionary feature of the *Bacillus* genus.

RAST annotation of the four strains into subsystems reveals that the majority of genes that are associated with essential functions and survival in environmental settings, such as sulphur and nitrogen metabolism, cell sporulation, cellular division and proliferation are present in all four genomes. With the exception of a small number of cell division proteins,4 spore coat & spore protease proteins and 2 nitrogen & sulphur related binding proteins, the remaining maintenance genes in these categories make up the core genome. It is the genes that relate to the biosynthesis of vitamins and amino acids, such as biotin and folate, and coenzyme producing genes that are present in a greater abundance in Bi10 than the other three *B. licheniformis* genomes.

The harsh environment that Bi10 was isolated from could explain the nature of its genomic differences. Due to the reduced availability of nutrients in the oilfield environment, vitamin biosynthesis is likely to be an essential source of growth factors and coenzymes to form of fatty acids. Furthermore, it has previously been suggested that the *in situ* biosynthesis of vitamins, like folate, are important to the production of biosurfactants from bacterial isolates, in particular *Bacillus spp.* (Kosaric, 1993; Makkar & Cameotra, 1999; Santos *et*

al., 2016). Additionally, it is the amino acid and fatty acid biosynthesis pathways that have been hypothesised to play a role in regulation of biosurfactant precursor production, by assisting in the formation of the interfacial tension altering hydrophilic and hydrophobic moieties of the biosurfactant, (Siñeriz, 2011). Although this has not been confirmed to be the reason why Bi10 carries additional biosynthesis genes, nor has the role of these extra genes been confirmed in *in situ* surfactant production. However these ideas represent a reasonable hypothesis to explain why the Bi10 strain could possess these extra coding sequences.

The production of Surface active compounds (SACs), or biosurfactants, is a multi-operon, regulated process, with a number of genes linked to amino acid biosynthesis, quorum sensing and non-ribosomal synthase genes (Aguirre-Ramirez *et al.,* 2012; Das *et al.*, 2008). Based on gene prediction from annotation, it was possible to identify 20 distinct biosurfactant related genes from the Bi10 genome (Table 6.5). The largest of these gene families, and potentially most important, was the gene cluster of *LchA*, a large multienzyme synthetase complex consisting of four subunits (lchAA-AD) (Sen, 2010), that encode the lipopeptide *lichenysin*. *Lichenysin* is the primary lipopeptide produced by a *B. licheniformis* subspecies and accounts for a large proportion of industrial-scale biosurfactant production. Furthermore, *lichenysin* produced by the *B. licheniformis* TT42 and *B. mojavensis JF2* strains have been shown to reduce interfacial tension to levels below 10 mN m^{-1}(Joshi *et al*., 2016; Michael J Mcinerney *et al*., 1990; Suthar & Nerurkar, 2016). Although this is a low IFT score, the total amount of additional oil recovered by other known

lichenysin producers in porous microcosm models has been less than that recovered by Bi10 (Chapter 5). Therefore, the identification of this operon could be key to understanding why Bi10 is even more effective at enhancing oil recovery in an oilfield environment. Curiously, although the four subunits are present, it is the similarity between them and their relatives that is striking, particularly for the *LchAA* and *AB* subunits. Yakimov *et al*., (1998), discovered that the *lichenysin* genes show a high degree of sequence conservation to that of other bacterial and fungal peptide synthetase genes. Yet the *LchAA/AB* subunits encoded by Bi10 both have a similarity below 87% to the *LchA* gene subunits identified on the BiosurfDB database. This similarity score is which is considered low for highly conserved genes. Therefore, it can be hypothesised that the distinct evolution that may have occurred in this strain due to its environmental niche and the stresses and selective pressures imposed upon Bi10 in the reservoir could be the principal factors behind the genetic alterations of *LchA*. This, in theory could possibly describe the behavioural changes of this lipopeptide. However this is something we will focus on further in the next chapter. The *pps* operon was also identified from the BioSurf DB search. Although this operon produces another NRPS, the antifungal plipastatin, it is not directly related to IFT reduction or haemolysis, and has previously been exhibited to have negative effects on haemolysis when produced in combination with lipopeptide molecules such as *surfactin* (Coutte *et al*., 2010).

Of the 20 identified biosurfactant related sequences, not all are commonly found in the *Bacillus* genome and so would have been overlooked had it not

been for the additional manual searches of the PROKKA and RAST annotations for other identified surfactant related genes, identified from NCBI database examination. These include the transcriptional regulators *comA* and *liaR2* and the RNA polymerase associated genes *rpoN* and *sigA*. Both *comA and sigA* have been shown to play active roles in stimulating the transcription of the SrfAA gene in *B. subtilis* via promotor recognition, which in turn is regulated by the *codY* gene and members of the rap family (A,F,G & H) in a negative feeback loop to limit lipopeptide production(Ogura *et al.*, 2001; Serror & Sonenshein, 1996). Due to the similar nature of the lipopeptide synthetase operons between *B. licheniformis* and *B. subtilis*, it is suggested that the same regulatory features apply here to the oilfield isolate *B. licheniformis* Bi10.

Our results suggest that the presence of a lipopeptide operon and regulatory genes within the *B. licheniformis* Bi10 genome indicates the ability to produce of some form of lichenysin-related biosurfactant. How similar the lipopeptides produced by Bi10 are to the biosurfactant Lichenysin is the key question that is yet to be identified and will be discussed in the next chapter. However, the genomic variation between Bi10 and other Lichenysin-producing strains could provide a reason as to why this lipopeptide behaves differently to that of other biosurfactants produced by the *lchAA* operon.

Chapter 7

Assessment of the pathways involved in biosurfactant production and the corresponding interfacial tension alteration of produced lipopeptide in a simulated oilfield system

7.1 Introduction

7.1.1 Background

Biosurfactants are biological surface active compounds that have a wide variety of uses in a number of fields: agriculture and medicine (primarily for their antimicrobial properties), cosmetics, food and oilfield technology (J. Patel *et al*., 2015; Rodrigues & Teixeira, 2010; Sachdev & Cameotra, 2013; Siñeriz, 2011). Within oilfield sciences, biosurfactants have been previously utilized for both bioremediation and enhancing crude oil recovery, primarily due their ability to reduce interfacial (IFT) and surface tensions (Daryasafar *et al.*, 2016; Prakash *et al.*, 2012). As discussed in Chapter 1. , there are various types of biosurfactant that have been identified. Of these, *Pseudomonas spp.* produced rhamnolipids and *Bacillus spp.* produced lipopeptides are the most commonly studied biosurfactants in the literature, particularly concerning oilfield technologies. The first notion of biosurfacatant production was elucidated in the late 1940's (Jarvis & Johnson, 1949), however it was not until almost a decade later that the first biosurfactant was isolated and assessed for activity (Arima*et al.*, 1968). Since then, the use of biosurfactants has strengthened, with biosurfactants beginning to replace the standard chemical surfactants, particularly in the oil industry (Marchant & Banat, 2012). Although

their performances are similar in terms of wettability alteration, the lack of toxic components, ease of production and disposal and reduced costs make it a viable alternative to the more traditional chemical surfactants (Gudiña, Fernandes, Rodrigues, Teixeira, & Rodrigues, 2015).

The ability of biosurfactants to reduce IFT is centred around its structural components, containing both hydrophilic (peptide monomers) and a hydrophilic (fatty acids) portion (Cameotra *et al.*, 2010). The existence of both hydrophilic and hydrophobic elements within the biosurfactant molecule, results in the migration of said molecule to the interface or surface with differing degrees of polarity and hydrogen bonding (Desai & Banat, 1997). This then causes a differentiation in forces between the two interfaces, thus altering wettability in the oilfield. The production of these biosurfactants within the bacterial cell is conducted via nonribosomal peptide synthetase (NRPS) driven biosynthesis, which act as the template and synthesizer of the biosurfactant (Bushley & Turgeon, 2010). Unlike with conventional peptide structures, the production of NRPS biosurfactant molecules is undertaken independent of the bacterial ribosome (Mach *et al.*, 1963; Martínez-Núñez & López, 2016).

7.1.2 Structure and diversity of nonribosomal peptide synthetase and their corresponding genes responsible for biosurfactant production

Strains of the genera *Bacillus* and *Psuedomonas,* amongst others, produce a range of different nonribosomal peptides. Within the *Bacillus* genus*,* at least 6 different classifications of surfactant are produced, including those of the surfactant, *iturin* and *fengycin* families to name but a few (Roongsawang *et al*,

2011). These families are then subdivided into further groupings of nonribosomal peptide biosurfactants (e.g. *Lichenysin, Surfactin, Bacillomycin*). Although structurally similar and synthesis driven by the same mechanism, NRPS enzymes, the differences between these lipopeptide biosurfactants are centred around the peptide monomers that make up the hydrophilic portion of the biosurfactant.

The NRPS gene cluster, that codes for the NRPS enzymes driving the lipopeptide production are split into a number of smaller genes or modules. The order of these genes is linear, with each modules gene linked to the gene corresponding to its peptide precursor. Each module is composed of core catalytic domains, primarily that of adenylation, thiolation and condensation of the peptide, however thioesterase domains are also sometimes present at the C-terminus of the NRPS (Felnagle *et al*., 2008). Although they may have the same catalytic domains, a variation within the sequence of the NRPS gene, such as the *LchA* (Madslien *et al*., 2013) or *SrfA* (Jiang *et al*., 2016) genes responsible for the production of Lichenysin and Surfactin, determines the final peptide structure of the protein. Different peptide structures affect the way that the lipopeptide behaves, with variation in amino acid either increasing or decreasing the effectiveness of the lipopeptide By identification of the NRPS gene sequence in the bacterial genome, it may be possible to successfully predict the peptide structure that this coding region will produce (Aleti *et al.*, 2015; De Bruijn *et al*., 2007). Expression of the NRPS genes are controlled by a complicated cascade of other regulators, either up or downstream of the NRPS gene. These all play their own integral part to the production and

inhibition of translation and are variant between strains. However, as shown by the literature, these cascades are found to be highly conserved within genera, or at least for that of *Bacillus spp.,* the most intensively examined biosurfactant regulatory system within the literature (Roongsawang *et al*., 2011)

7.1.3 Aims

With the presence of gene clusters relating to lipopeptide biosurfactant production identified in the previous chapter, the primary aim was to identify the likelihood of the type and structure of the lipopeptide. Once this had been confirmed, successful precipitation and extraction of the biosurfactant was to be undertaken to recover an active fraction that could be used in further experimental work. Following on from this, the secondary aim of the chapter was to assess the alteration in interfacial tension, and thus wettability, within an oilfield system and identify whether the biosurfactant produced by the *B. licheniformis* Bi10 strain is in all likelihood the driving factor behind the increase in additional oil recovery within the simulated Bentley oilfield. The data in this chapter assesses the outcomes of two distinct hypotheses. Firstly, the structure of the produced biosurfactant molecule will be distinctively different to that of the other known surfactants produced by *Bacillus* strains. Secondly, it was hypothesised that the amount of biosurfactant produced would significantly alter the interfacial tension at the oil/water interface, thus driving the MEOR mechanisms.

7.2 Methods

7.2.1 NRPS sequence analysis

From work described in Chapter 6, it was identified that the production of the secondary metabolite surfactant by the oilfield isolate Bi10 was likely to be mediated by a non-ribosomal peptide synthetase (NRPS) operon. The nucleotide and amino acid sequences corresponding to this NRPS operon, *LchA*, including any genes that had been identified as being involved in biosurfactant production (Chapter 6.3.6) were selected for further analyses using Geneious7. The NRPS coding sequences of strain Bi10 were compared to the NRPS codings sequences of the four *Bacillus sp.* previously used for comparative genomics in Chapter 6 (*B. licheniformis* WX02, *B. licheniformis* BL09, *B. licheniformis* ATCC14580, *B. subtilis* 168), using Muscle 3.8.31 multiple alignment tool (Edgar, 2004) and antiSMASH 3.0 (Weber *et al*., 2015). antiSMASH was used for the prediction of the peptide structure produced by the *LchA* operon, using default parameters with PFAM analysis and cluster blast.

7.2.2 Biosurfactant extraction and purification

The biosurfactant was precipitated following the inoculation of 1 L of LB broth (as described in Chapter 2) with the selected bacterial culture. After 24 hour growth, the cell-free supernatant from the culture broth was obtained following centrifugation at 10,000 x g for 20 min at 4 °C. The remaining cells were removed by filtration using a filter with pores of 0.22 µm diameter. The cell free supernatant was subjected to one of three protein recovery strategies: 1) Ammonium sulfate ($(NH_4)_2SO_4$) precipitation, 2) Trichloroacetic acid

precipitation (TCA) and 3) Hydrochloric acid (HCl) precipitation. Both $(NH_4)_2SO_4$ and TCA precipitations were performed by the addition of 60% (w/v) of $(NH_4)_2SO_4$ or TCA to the cell free supernatant, following the protocols of Koontz (2014) and Shah *et al* (2016), respectively. HCl precipitation was conducted in two ways. The first method was performed by adjusting the pH of the cell-free supernatant to 2 using 6 M HCl and incubating the reaction at 4 °C overnight following the protocol of Vater *et al* (2002). The surfactant precipitated by all three precipitation methods was recovered by centrifugation at 10,000 x g for 30 mins to separate the precipitate. The pellet resulting from the HCl treatment was subjected to solvent extraction, using 2 volumes of 100% methanol and the methanol was then evaporated by rotary evaporation. The second HCl extraction protocol involved a slow, progressive decrease of pH over a 10h period to pH 2 in 1.0 pH unit intervals every 2 hours at 4 °C, before being left at pH 2.0 for a further 6 hours The precipitate was recovered by centrifugation at 10,000 x g for 25 min at 4 °C, and extracted two times with methanol, which was evaporated using a rotary evaporator at 70 °C. Finally, the crude biosurfactant was collected and weighed for total surfactant extraction concentration. Oil spread assays were performed on all crude lipopeptide recoveries to assess for activity (Chapter 4.2.3.2).

7.2.3 Experimental setup for Interfacial tension analysis

The interfacial tension values of an oil-water system, mimicking reservoir conditions of temperature, pressure, formation water and crude oil, was assessed by using the Drop Interfacial Tension IFT-10-P (Core Laboratories, USA) and DROPImage Advanced digital image capture software (Rame-hart, US). This equipment was provided by the Petroleum Technology Group,

University of Salford, UK. Figure 7.1 shows the setup of the IFT experimental equipment that was used in this study.

The pressurized bioreactor cell (Fig. 7A) was comprised of a stainless steel cylindrical chamber with a volumetric capacity of ~ 43 cm^3, with two injection ports (top and bottom of cell) and one flow valve, for removal of liquids. At the bottom injection orifice, a stainless steel needle (diameter ~3.03 mm) is used for the generation and release of pendant drops of heavy Bentley crude oil into the liquid. The chamber is equipped with two borosilicate glass windows, with a camera and halogen bulbs placed at opposing sides of the cell, facing inwards.

Using a high-pressure syringe pump, formation water and various inocula were injected into the cell chamber at the appropriate pressure. and left to equilibrate to the working temperature of either 40 °C or 60 °C. Bentley heavy crude oil was pumped into the system using the manual pump and a droplet of oil formed. This droplet is then also left to equilibrate for approximately 10-15 minutes. Once used, the droplet was released and a new droplet formed for further tests. Ten independent IFT measurements were taken for each condition, with the mean and SD calculated.

A total of four conditions were tested: 1) synthetic formation water 2) 4% OFM nutrient 3) Bi10 cell culture 4) Bi10 cell free surfactant supernatant. Synthetic formation water, matching Bentley reservoir geochemistry, was used as the negative control. 4% OFM nutrient was mixed with formation water at a

concentration of 1:1 to assess IFT without bacterial involvement. *B. licheniformis Bi10* as cultured in 4% OFM overnight and either the culture of cell-free culture supernatant were independently mixed at 1:1 with formation water and pumped into the bioreactor cell, following the protocol set out by Ukwungwu *et al* (2017).

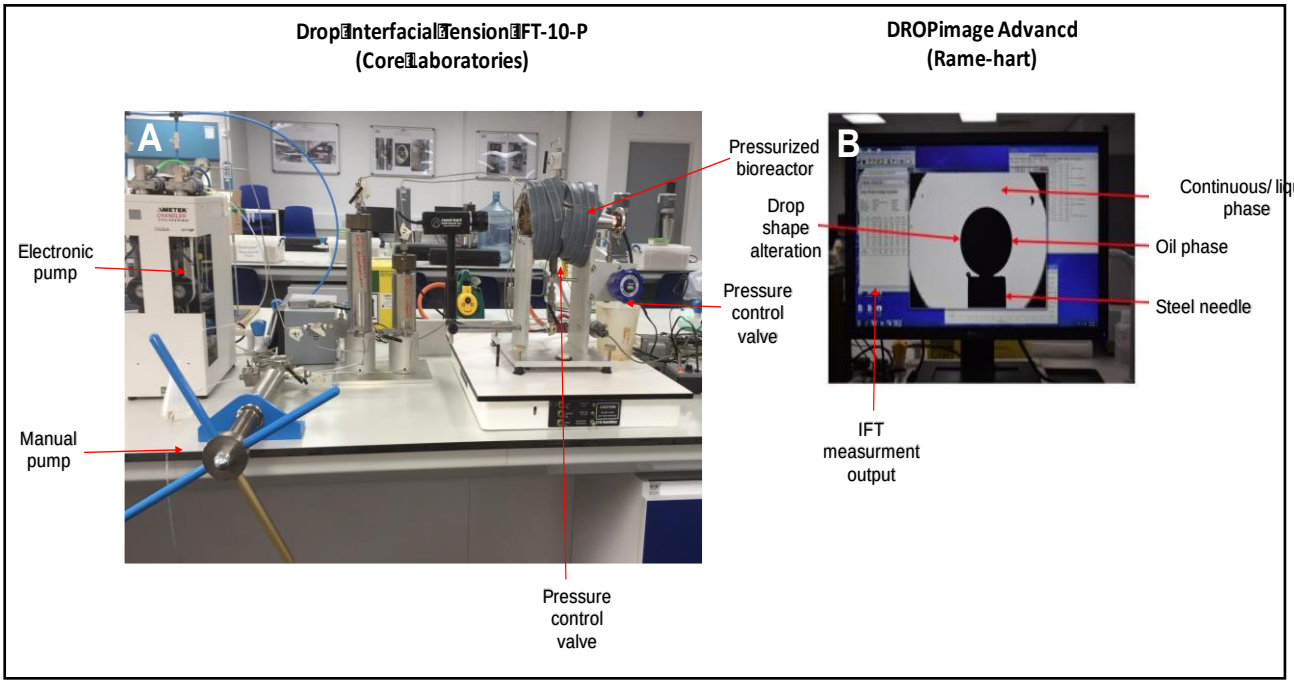


Figure 7. Experimental set up of interfacial tension measurement equipment and visualisation calculation equipment used. Panel A shows the equipment used from the top left, moving clockwise: Pressurised and temperature controlled bioreactor, with two borosilicate glass windows, pressure control valve set to 1,800 psi (rated to 10,000 psi), manual pump for water removal and to drive oil release, electronic pump for filling of bioreactor and surfactant/ media inoculation. Panel B shows the setup of the imageDROP software used to record IFT measurements.

7.3 Results

7.3.1 Comparative analysis of the biosurfactant production pathway of bacteria from the *Bacillus* genus

All of the strains used from the *Bacillus* genus in this study possessed a number of identical, structurally conserved regions in their NRPS operon, represented by *YcxC, YcxD* and *sfp* (Fig 7.2). *YcxA* and *YcxB* are both present in *B. subtilis* 168, but not in any of the *B. licheniformis* strains. Looking specifically at the NRPS gene cluster, additional structurally conserved regions of *LchAC* & *LchAD* were identified in the biosurfactant metabolic pathway in all *B. licheniformis* strains analyzed in this study. The lchAC gene subunit carried by every strain was 3849 bp and coded for a protein containing 1282 amino acids. The lchAD gene was 741 bp and encoded a 246 amino acid protein (Table 7.2). In the *B. licheniformis* WX02 strain, the NRPS *lchAB*-2 gene shared high sequence conservation between WX02 and the Bi10 isolate, with 100% sequence similarity and 100% coverage shown between the genes of these sequences. Furthermore, in both WX02 and Bi10, both *lchAB*-1 and *lchAB*-2 ORFs were identified. The existence of two NRPS AB genes is not apparent in the other two *B.licheniformis* strains (ATCC14580 & BL09) or even in *B. subtilis* 168. The similarity and/or coverage of the NRPS AA & AB genes are substantially reduced in all *Bacillus* strains studied in comparison to the sequence of the *lchAA & AB*-1 in the Bi10 isolate. Both *lchAA* and *lchAB* of Bi10 have a shorter nucleotide sequence length (9669 bp for *lchAA* and 9627 bp *for lchAB*1 *& 2*) and amino acid sequences (3222 a.a. and 2696 a.a., respectively) than the other members of the *Bacillus* genus,

which have sequence lengths of ~10700 – 10900 bp and ~3500 – 3650 a.a. (Table 7.1).

Table 7.1 Amino acid and nucleotide sequence length of NRPS domains AA-AD of Bacillus spp. Strains used in this study. Red cell indicates domain not present in strain

Strain	Sequence	NRPS Domain length				
		AA	***AB1***	***AB2***	***AC***	***AD***
B. licheniformis Bi10	Nucleotide	9669	8091	1536	3849	741
	Amino Acid	3222	2696	511	1282	246
B. licheniformis ATCC 14580	Nucleotide	10743	10767		3849	741
	Amino Acid	3580	3588		1282	246
B. licheniformis WX02	Nucleotide	10758	9315	1632	3849	741
	Amino Acid	3586	3105	544	1282	246
B. licheniformis BL09	Nucleotide	10749	10770		3849	741
	Amino Acid	3580	3588		1282	246
B. subtilis 168	Nucleotide	10764	10762		3828	708
	Amino Acid	3587	3583		1275	242

Through amino acid sequence analysis using antiSMASH 3.0, it was identified that the *LchAA-AD* domains of Bi10 are arranged into a total of 22 modules: 9 domains in *AA*, 7 domains in AB-1, 1 domain in *AB-2*, 4 domains in AC and 1 domain in *AD* (Fig 7.3). Of these, 7 are condensation sites (C), 2 are epimerization domains (E), 6 are peptidyl carrier domains (PCP), 5 are AMP-binding domains (AMP), that are responsible for incorporation of the peptide monomers of the biosurfactant, and 2 thioesterase domains (TE). All of these domains incorporate *gln-leu-val-asp-ile* to make the biosurfactant peptide.

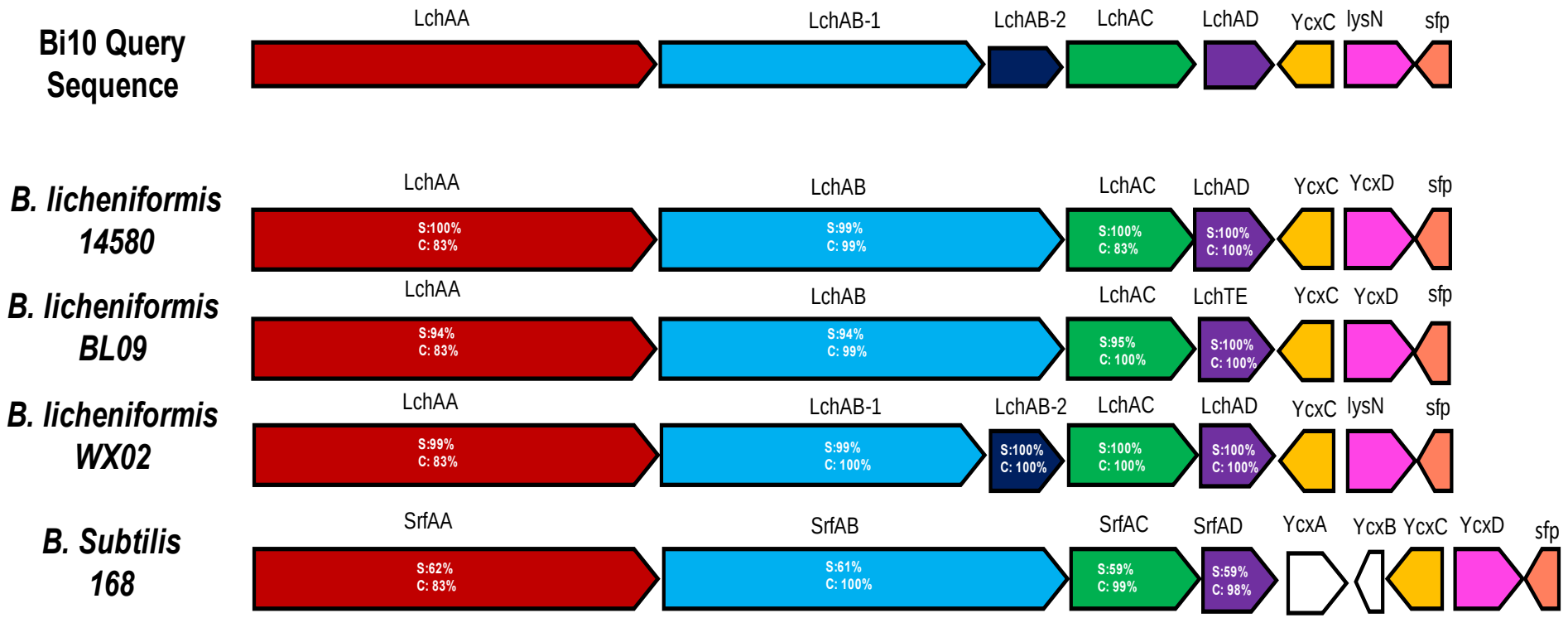


Figure 7.2 Comparative analysis of biosurfactant metabolic pathway in *Bacillus spp.* Arrows with matching colours correspond to orthologous genes in the NRPS operon. Directionality of each gene is indicated by the direction of the arrow. White arrows represent novel genes not present in Bi10 biosurfactant metabolic pathway. Labelling inside NRPS domains A-D correspond to sequence similarity (S) and coverage (C) of genes against Bi10 query sequence.

Strains of the *Bacillus* genus and the Bi10 isolate share these 22 domains, with all responsible for the incorporation of the same amino acids into the peptide structure, with the exception of *B. subtilis,* which incorporates *glutamic acid* instead of *glutamine* at the first AMP binding domain of *SrfAA* and *leucine* instead of isoleucine in *SrfAC*. However, the greatest difference observed between Bi10 and the remaining four *Bacillus* strains is in the lack of two AMP binding sites. These are both found at the ends of the NRPS *AA* and *AB* genes and are responsible for the integration of two further leucine amino acids (leu), that are not present in the Bi10 predicted peptide structure

7.3.2 Biosurfactant recovery.

After 24h growth in the 4% OFM media, recovery of the bacterial lipopeptide was performed. A number of differing methods were tested to attempt to recover a suitable quantity of lipopeptide, whilst keeping the biosurfactant active. The first precipitation protocol followed was that of ammonium sulfate addition and dialysis, a commonly used method for 'salting out' proteins in solution. The ammonium sulfate precipitation produced a small quantity of precipitate (< 0.1 g L^{-1}) after centrifugation at 60 % saturation, however this precipitate possessed no interfacial tension activity via oil drop assay. Solvent based Trichloroacetic Acid (TCA) precipitation was also trialed in the anticipation that pH alteration may pull the biosurfactant out of the suspension. This method produced no pellet or precipitate and also did result in the harvesting of anything capable of oil spread activity.

The final protocol for precipitation of lipopeptide was the use of hydrochloric acid (HCl) precipitation followed by methanol extraction, the preferred recovery method for lipopeptide extractions. The standard HCl precipitation protocol states that cell free supernatant must be reduced to a pH of 2.0 with continuous stirring for ~ 6 h. Although this method produced a sizeable pellet (1.337 g L^{-1}) of precipitate, no activity was observed after methanol extraction and solvent evaporation. Due to this method producing the largest precipitate of all trialled methods, it was therefore decided to alter the protocol to use a progressive decrease in pH over 10 h to a pH of 2.0 to see if activity could be preserved. After a 10 h pH decrease with HCl and overnight incubation at pH 2.0, 1.458 g L^{-1} of precipitate was recovered. Furthermore, this precipitated fraction possessed activity that could drive an oil spread diameter of ~ 1-3 cm.

Table 7.2 Lipopeptide recovery methods and corresponding surfactant activity

Method	**Ammonium sulphate precipitation**	**TCA precipitation**	**HCl precipitation** (pH 2.0)	**HCl Precipitation** (pH 7.0- > 2.0 over 10 h)
Concentration of lipopeptide (g L^{-1})	< 0.1	0.0	1.337	1.458
Lipopeptide activity	-	-	-	+++

N.B. +/- relate to lipopeptide activity as described in **Chapters 2.4 & 4.2.3**

7.3.3 Growth and biosurfactant production profile of Bentley reservoir isolate *B. licheniformis* Bi10.

The profiles of bacterial colony enumeration and corresponding biosurfactant production were assessed against time over a 72 h period (Fig 7.4). After culture set-up and incubation, the Bi10 isolate was able to proliferate in the 4% OFM media, showing a rapid increase (7.3 x 10^6 to 8.4 x 10^6 cfu mL^{-1}) in colony counts in the log phase over the first two hours (Table 7.2). Growth continued at a steadier rate

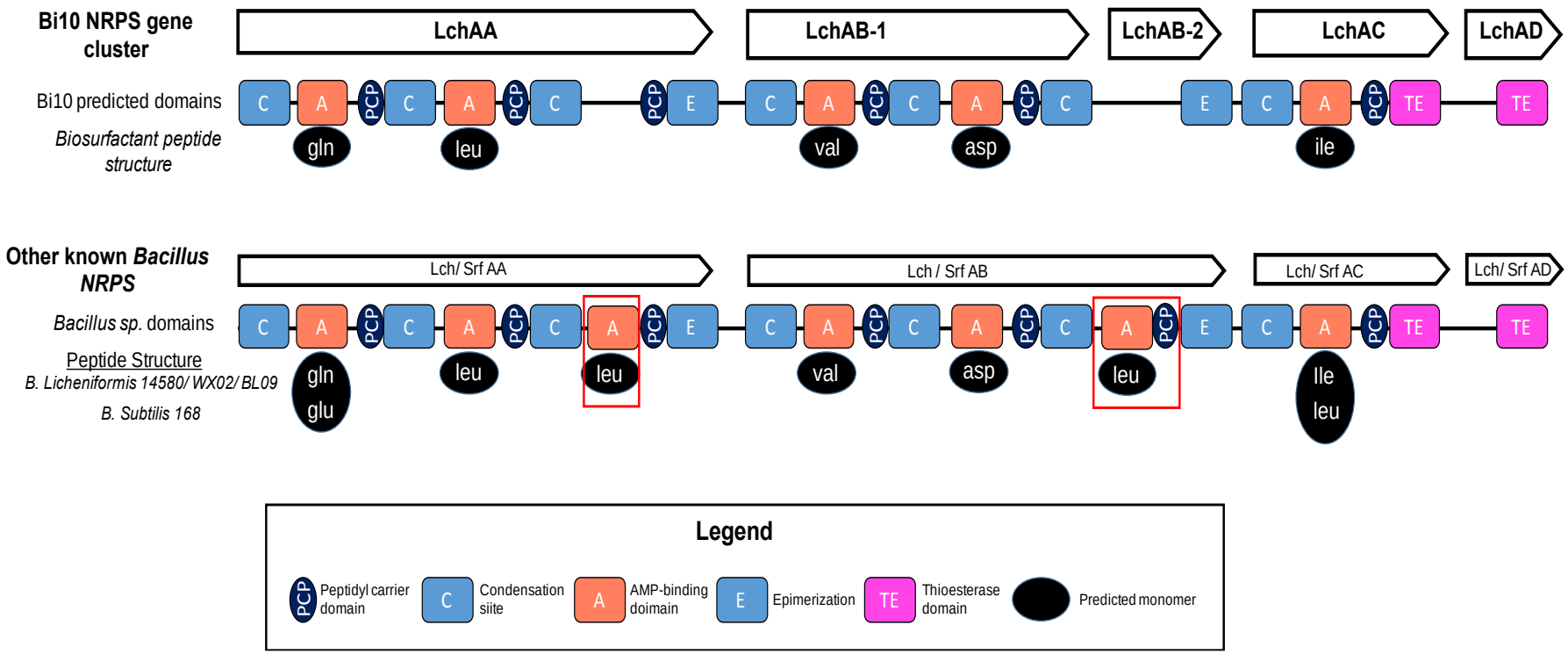


Figure 7.3 Prediction of NRPS gene cluster and peptide structure of *B. licheniformis* Bi10 isolate and other comparative *Bacillus spp.* Upper schematic shows NRPS operon of Bi10 isolate predicted by antiSMASH 3.0. Second schematic shows the NRPS operon of four *Bacillus spp.* (*B. licheniformis ATCC14580/ WX02/ BL09 & B. subtilis 168*). Arrows correspond to orthologous gene clusters present in *Bacillus spp.* Boxes represent peptide domains present in NRPS genes corresponding to the figure legend. Predicted monomers of protein structure are shown below peptide domains (black circle), corresponding to each *Bacillus* strain. Red outline identifies domains that are missing in the Bi10 domain based on NRPS domain prediction.

hereafter until ~ 12-24 h, where Bi10 then enters stationary and death phase of its growth cycle, with colony counts decreasing after 24 h. Conversely, crude biosurfactant production continues to increase over the 72 h time course of this experiment, reaching maximum production of 2.1215 g L^{-1}(±0.028) g L^{-1}. However, the highest rate of crude biosurfactant production occurs between 2-4 h, with production of the lipopeptide increased by 0.4155 g L^{-1} over this two hour period (0.2775 g h^{-1}). The rate of crude biosurfactant production continues on a continuous ascending trajectory between 4-12 h, but this rate is less steep (0.1064 g h^{-1}) than the rate observed between 2-4 h.

Table 7.3 Bacterial growth kinetics and biosurfactant production from Bi10 isolate

Time (h)	**Bacterial enumeration (cfu mL^{-1} ± SD)**		**Interfacial tension (mNm^{-1} ± SD)**	
0	7.3×10^6	±500000	0.0000	±0.000
2	8.7×10^7	±3000000	0.2540	±0.039
4	1.3×10^8	±6000000	0.6695	±0.068
8	2.4×10^8	±8000000	1.0920	±0.082
10	2.6×10^8	±2200000	1.3085	±0.013
12	3.0×10^8	±15000000	1.5205	±0.040
24	1.1×10^6	±1200000	1.8440	±0.048
48	5.6×10^5	±160000	1.9490	±0.042
72	1.1×10^6	±100000	2.1215	±0.028

After 4 h to the end of the experiment (72 h), bacterial cell numbers decrease, due to either cell death or sporulation, before plateauing at approximately 72 h. In contrast biosurfactant production continues to increase over this time period, but at a substantially reduced rate (0.01 g h^{-1}).

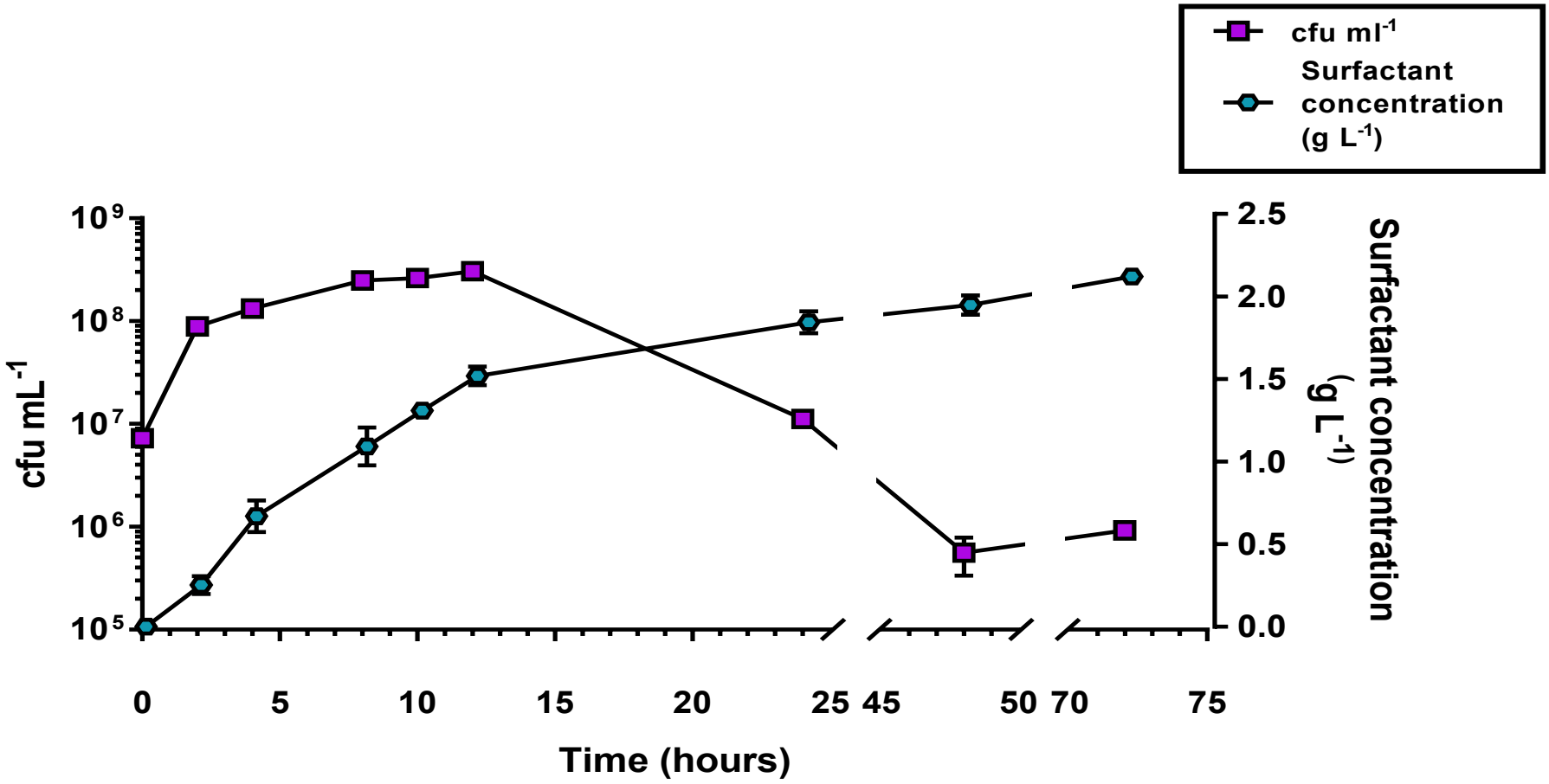


Figure 7.4 Comparison of surfactant concentration against bacterial growth kinetics over a 72 h period. Cell culture removed at appropriate time points and colony counts (left x-axis) and crude surfactant concentration measurements (right x-axis) calculated for each time point. Surfactant mass measurements calculated based on total mass (g L^{-1}) after acid precipitation and methanol extraction. Values are shown as mean cfu mL^{-1} (± SD, *n*=3 replicates) and mean crude surfactant concentration (± SD, *n*=3 replicates).

7.3.4 Interfacial tension measurements of Bi10 and critical micelle concentrations

The ability of the biosurfactant produced by the *B. licheniformis* Bi10 oilfield isolate to reduce interfacial tension (IFT) between crude Bentley oil and synthetic Bentley formation water (Fig 7.5). IFT of both a Bi10 culture and a Bi10 cell free supernatent was calculated based on oil drop analysis and compared to the interfacial tension of sterile synthetic formation water and sterile OFM media at a standardized Bentley oilfield pressure of 1800 psi and two temperatures, 40°C and 60°C. At 40°C, Bi10 cell/ surfactant mixture significantly reduced the IFT (9.48 mNm^{-1}) compared to sterile formation water (55.28 mNm^{-1}) and sterile OFM media (15.61 mNm^{-1})(Fig. 7.5). Bi10 cell free surfactant significantly decreased IFT (1.43mNm^{-1}) compared to the cell/surfactant suspension.

The images showing the contact angle measurments of treated oil droplets at both 40°C and 60°C, clearly show a substantial increase in contact angle in both of the Bi10 cell treated conditions (Fig. 7.6 A & B). At 60°C, a similar pattern emerged with OFM media showing a decrease in IFT, but cell and cell free treatments reducing IFT further. However, the IFT of the cell/ surfactant suspension showed no significant difference (5.61 mNm^{-1}) in comparison to the cell free supernatant (3.05). Furthermore, unlike the decreases seen in all other treatments in IFT between 40°C and 60°C, the cell free supernatant actually increases the IFT at 60°C. This is supported by the contact angle images, which show an increase in contact angle at 40°C compared to °C. This data suggests that the cell free lipopeptide is less effective at higher temperatures.

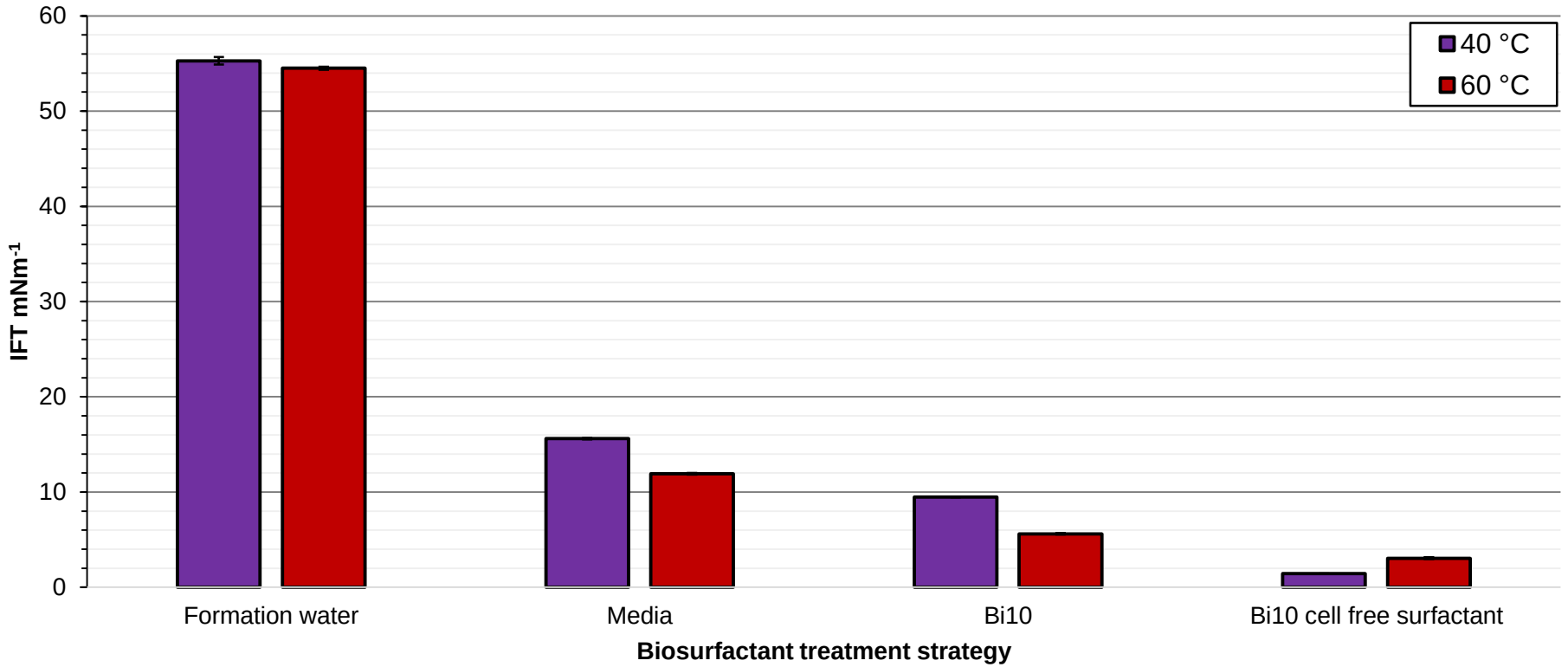


Figure 7.5 Assessment of interfacial tension alteration by biosurfactant produced by *B. licheniformis Bi10* versus apposite controls. Interfacial tension measurements calculated using Drop Interfacial Tension IFT-10-P and imageDROP visualisation software at a steady pressure of ~ 1800 psi and 40 °C & 60 °C. Formation water and 4% OFM media were used as controls to compare the effects of liquid viscosity alteration on IFT when no microorganisms where present. Bi10 cells w/ surfactant and cell free surfactant cultures were also assessed at both temperatures. Values are shown as Mean IFT (± SD, n=10 replicates).

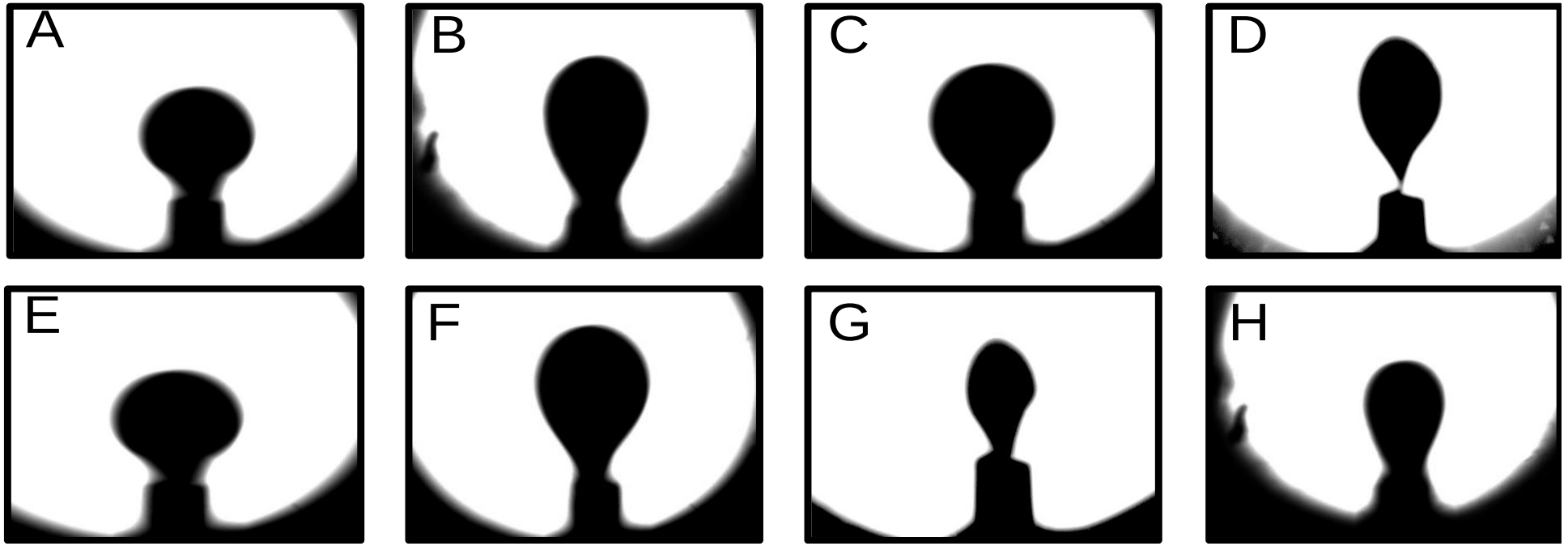


Figure 7.6 Image capture of pendant drop oil measurements over various biological treatments and temperatures. Images were captured from the Drop Interfacial Tension IFT-10-P using imageDROP visualisation software after drops had been left to equilibrate at temperature and pressure. All measurements were taken at a pressure of ~1800 psi and 40 °C & 60 °C. Panel A-D show images of pendant drop analysis at 40 °C Panels E-H show images of pendant drop analysis at 60 °C. Panels A & E show drop analysis corresponding to IFT measurements of sterile formation water. Panels B & F show drop analysis corresponding to IFT measurements of sterile 4% OFM media. Panels C & G show drop analysis corresponding to IFT measurements of overnight culture of Bi10 cells w/ Bi10 surfactant in suspension. Panels D & H show drop analysis corresponding to IFT measurements of cell free Bi10 biosurfactant. Image shown as one representative of n=10 repeats conducted.

After the confirmation of the IFT reduction by the Bi10 isolate, it was essential to identify the critical micelle concentration (CMC) of the surfactant. This analysis was conducted using recovered and extracted crude biosurfactant dissolved in synthetic Bentley FW at differing concentrations, quantifying the IFT for each concentration (Fig. 7.7). At increasing concentrations of crude biosurfactant, IFT is substantially reduced in formation water, until a plateau is reached. At this point, no further increase in biosurfactant concentration is able to decrease IFT further. This is the point at which micelles are formed due to the aggregation of surfactant molecules at the interface of the oil/water. For the Bi10 strain, the CMC was identified to be at a concentration of between 250 – 500 mg L^{-1}.

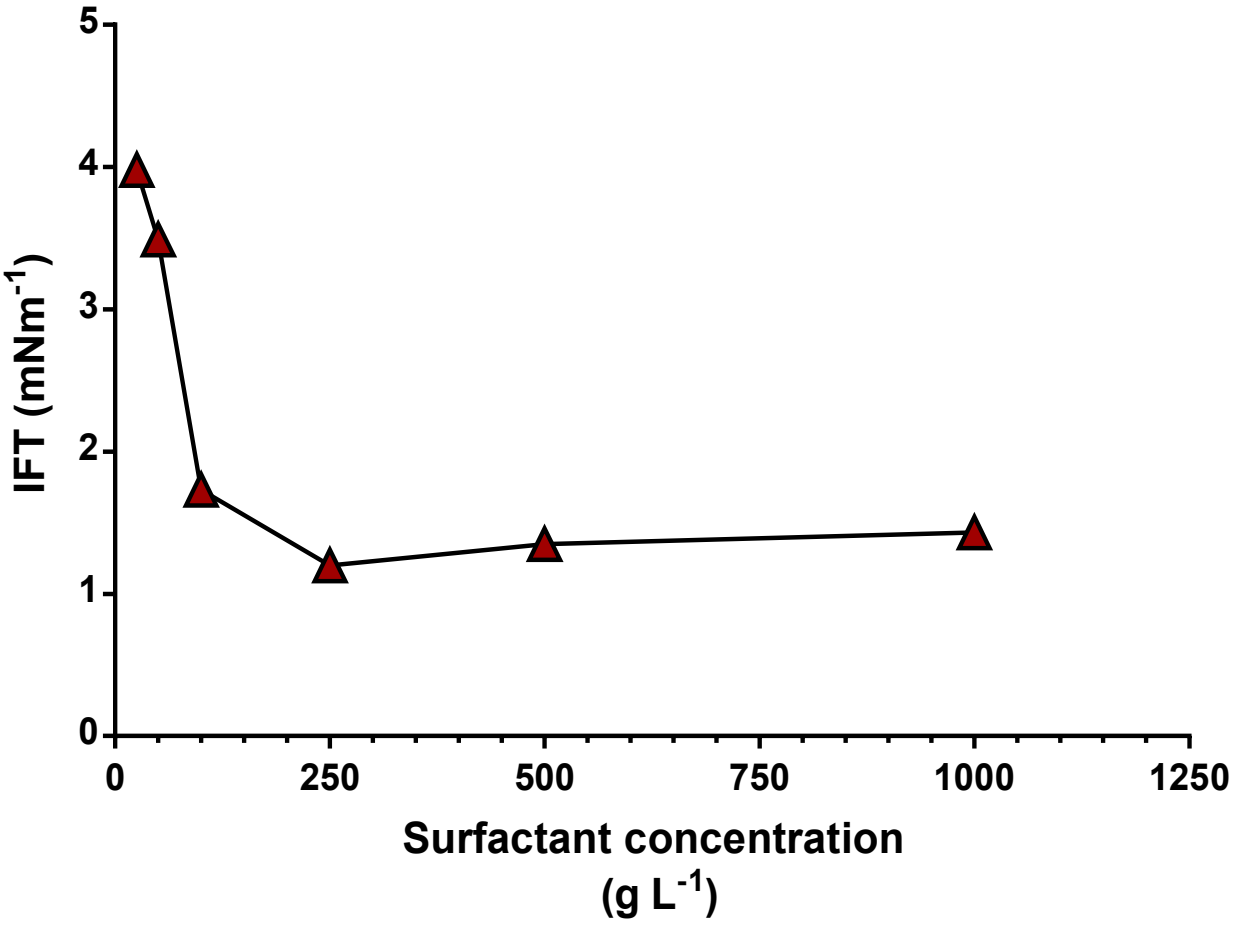


Figure 7.7 Changes in IFT versus biosurfactant concentration. IFT measurements automatically calculated using imageDROP advanced software. Variant concentrations of precipitated crude biosurfactant added to 1L synthetic oilfield formation water as appropriate for biosurfactant concentrations (25 mg L^{-1} – 1 g L^{-1}). Critical micelle concentration of *B. licheniformis Bi10* oilfield isolate calculated from comparison of IFT and surfactant concentration. Values are shown as Mean IFT (± SD, *n*=10 replicates)

7.4 Discussion

The biosurfactant pathway in the *B. licheniformis* strain Bi10 was identified in all likelihood to be controlled by an NRPS domain. This is also the same for all *Bacillus* spp., with the same NRPS domain, LchA identified in both Bi10 and other *B. licheniformis strains* (Domingos *et al*., 2015). The rest of this operon is made up of a number of other genes that are all involved in regulation of synthesized peptide structure, including *sfp*, an gene that is integral in the production of lipopeptides by *Bacillus spp.* (Nakano *et al.*, 1992). This gene, alongside the *ycxC* and *ycxD*, are highly conserved between the Bi10 isolate and other *Bacilus spp.* investigated in this study. However, it is the *lchA* gene cluster where the interesting sequence differentiation is observed. The *lchAC* and *lchAD* subunit genes both share a high level of similarity to other corresponding subunits in *B. licheniformis* species. These similarities include thioesterase domains that are responsible for the addition of amino acids to the peptide. Without conservation of this domain the isolate would be unable to form a cyclic lipopeptide molecule (Domingos *et al*., 2015). The LchAA and LchAB subunits however, have approximately ~ 1000 bp encoding for ~300 AA missing from the ends of the two gene subunits.The genes *lchAA* and *lchAB* both encode an AMP- binding site and associated amino acid sequence. Therefore, unlike most other *surfactin* and *lichenysin* based lipopeptides that comprise of 7 peptides, the Bi10 NRPS protein only has 5 amino monomers. This alteration of the peptide structure could be an interesting feature in determining why the biosurfactant from Bi10 behaves in a fashion different to other known *Bacillus* biosurfactants trialed in MEOR. This

however cannot be confirmed without the chemical characterization of peptide structure.

Choosing an appropriate precipitation method is importance to the identification and characterization of a biosurfactant. Due to the preceding phylogenetic analysis of the Bi10 strain (Chapter 6.3.3) and the investigation of the surfactant pathways, it was assumed that the biosurfactant produced by Bi10 would be a lipopeptide (Figs 7.2 & 7.3). However, it was decided that not only lipopeptides-based precipitation methods, such as HCl based extractions would be used, we also used standard protein precipitation protocols ($(NH_4)_2SO_4$ and TCA). Only the lipopeptides-specific, HCl precipitation method produced a useable quantity of surfactant fraction (>0.1 g L^{-1}). By assessment of these two HCl methodologies, it was identified that a sudden decrease in pH causes inactivation of the surfactant. Conversely, a steady reduction in pH did not cause the loss of activity in Bi10 extracts. This coupled with the previous information, further supports the theory that this biosurfactant is of a lipopeptide nature, potentially of *lichenysin* structure. However, this cannot be confirmed without the procedure of protein specific characterization techniques, such as LCMS, to identify the exact peptide structure of the biosurfactant.

Alongside precipitation of the biosurfactant, it was deemed necessary to assess the effect of contact angle alteration by the biosurfactant produced by the Bi10 isolate on the oil-water interface. Alteration of the oil-water surface tension is directly correlated to surface wettability. This wettability has long

been proposed as a potential mechanism of increased AOR via MEOR (Karimi *et al.*, 2012; Kowalewski *et al.*, 2006). The surface-active compound produced by *B. licheniformis Bi10* was produced in highest quantities during the log phase of bacterial growth, during the first 4 hours after sub culturing (0.2775 g h^{-1}). Hereafter, although the production of biosurfactant continued to increase, the rate of production slowed as cell culture hit the stationary and then death phase of the cell cycle, with a dramatic decrease in enumerated bacterial colonies observed. This is a similar finding to what has been previously seen in the literature, with cell cycle closely correlated to biosurfactant production. Furthermore, Heryani & Putra (2017) also showed that the highest rate of biosurfactant production occurred within the first 4-6 h of bacterial growth in *Bacillus spp.*

The contact angle, determined by obtaining measurements from a drop oil using a pendant drop tensiometer, showed a significant decrease in interfacial tension when compared to sterile nutrient and formation water. This substantial decrease in both the cell and surfactant suspension (9.48 mN m^{-1}) and cell free culture (1.43 mN m^{-1}) showed the alteration of wettability of a hydrophobic surface (oil droplet) to a more water wet state (observed in Fig 7.6 by change in droplet shape). This change in state has been shown to be beneficial in the process of EOR (Strappa *et al*, 2004; Ukwungwu *et al*., 2017). The reduction of IFT of the cell free surfactant to 1.43 mN m^{-1} was not determined to be lower than the IFT previously recorded by *B. mojavensis* JF2 (<0.01 mNm^{-1}) (Mcinerney *et al.*, 1990). However, this IFT is not comparable to that of McInerney *et al* (1990), primarily due to the conditions and equipment

used in the measurement. Rather than a pendant drop, pressurized and heated to oilfield conditions, Mcinerney *et al.* used a spinning DeNouy tensiometer at standard laboratory conditions, using extra light oils (> 25 API°). However, the IFT reduction of 1.43 mNm^{-1} by the Bi10 isolate, was determined to be lower than that of the IFT of other heavy oil biosurfactant producing *Bacillus* strains (reference).

To determine if the differential *lchAA* and *lchAB* genes are the reason for the differences in additional oil recovery and whether the MEOR mechanism is driven by biosurfactant production, it was decided that one approach would be to knock out the *lchA* gene cluster from the Bi10 strain. The *lchA* genes of *B. licheniformis* Bi10 were to be replaced by a kanamycin resistance cassette (Km^{r}) sourced from the pUC4K vector, using a well-established double homologous recombination events (allelic exchange) (Fig 7.8). The loss of the ampicillin resistance marker (Amp^{r}) was to be used as the determinant for the second recombination event.

To create the *lchA* knockout mutant, PCR amplification was used to acquire upstream (*lchAA*) and downstream (*lchAD*) sequences of the respective LchA gene cluster from the Bi10 isolate. Due to the low GC% content of the regions within the *lchA* gene cluster, particularly in the regions surrounding the start and stop codons, a number of primer pairs were trialed to optimize the design for suitable annealing temperatures and performance in PCR amplification. Three primer sets were designed for *lchAA* and *lchAD* subunits (Table 7.4). Of the three primer sets, only primer pairs for the upstream region and for the

downstream region resulted in bands of the correct sizing as determined by TAE agarose gel electrophoresis. Sanger sequencing of the purified amplicons identified them to be the correct fragments. The resulting amplicons were then cloned into the pUC4K cloning vector. The resulting vector could then be introduced into the Bi10 isolate via electroporation to knock out the *srf* gene by homologous recombination, replacing it with the kanamycin resistance cassette. Unfortunately, due to time constraints and the end of the project work, the ligation of both *lchAA*341 & *lchAD*306 and resultant transformation into *B. licheniformis* Bi10 and as such would need to be undertaken after the project has been completed.

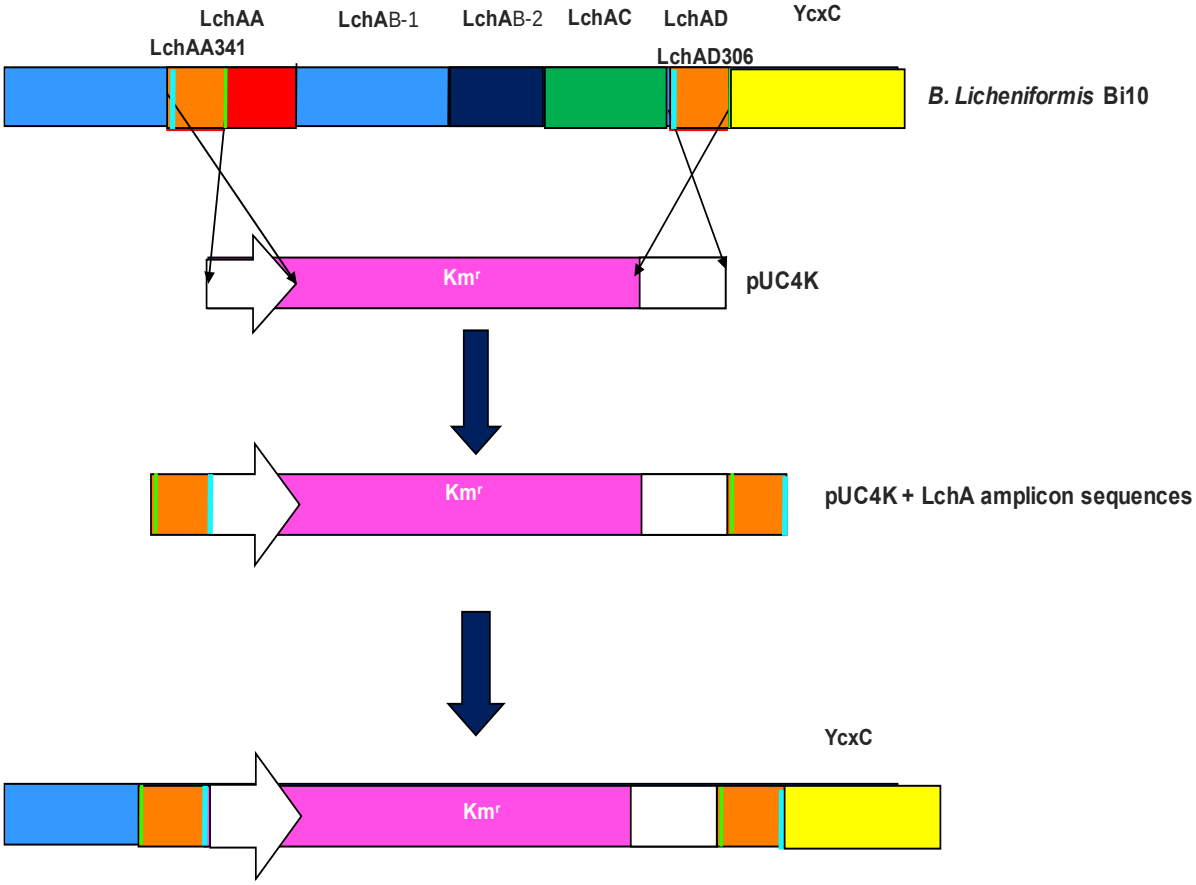


Figure 7.8 Flow diagram for proposed deletion of *LchAA* genes. Methodology for the transformation of *B. licheniformis* Bi10 with pUC4K vector via homologous recombination and replacement of the *LchA* operon. Genes shown via colour coordination and F&R primers identified via green and blue solid lines. Genetic cross over of both amplicons in the vector during transformation will lead to the replacement of the *LchA* gene with the kanamycin resistant cassette.

Table 7.4 Primers for LchA knock-out mutant creation in B. licheniformis Bi10

Primer name	Direction	Sequence	Gene target	Primer Length	GC%	T_m
*LchAA*_331F	Forward	ATACATAACAAGAGATAAAGGAGGA	*LchAA*	25	47	55.1
*LchAA*_272F	Forward	ACATGCTCAAAGGCGAATTTGG	*LchAA*	22	55	60.4
*LchAA*_312F	Forward	ATCGTCATACATAACAAGAG	*LchAA*	20	35	50.1
*LchAA*_R	Reverse	AGAACGCCGTCTGCTCCT	*LchAA*	18	55	61.0
*LchAD324*_F	Forward	GAATACCCAAAGAACTGG	*LchAD*	20	44.4	49.8
*LchAD*_300F	Forward	TTCCTGCCTTCGTTCAGAGC	*LchAD*	20	55	60.3
*LchAD*_R	Reverse	CTGCAGCACACCGATTTTCC	*LchAD*	20	55	61.2

Chapter 8

General Discussion

8.1 Summary of findings

In Chapter 3, the basic geochemical and petrophysical conditions of the Bentley oil reservoir were characterized, alongside the microbial communities present from rock core and produced water fluids from within the reservoir. This study identified a higher abundance of individual taxa than is previously identified in the literature. From the collection of the taxonomic data, two routes of analysis were performed. Firstly, it was possible to identify a correlation between specific microbiota and the petrophysical features of the reservoir. Second, and most importantly to this project, we were able to identify potential biosurfactant producing microorganisms present within the reservoir community that could be used for potential enhancement of oil recovery.

Chapter 4 documents the process of isolation and characterization of potential biosurfactant and biofilm producing microorganisms. Using the collated data of biosurfactant producing organisms identified in Chapter 3, it was possible to enrich, culture and selectively screen strains that we believed may be beneficial for biosurfactant production. Of a total of 309 identified isolates, 8 potential biosurfactant producers and one biofilm producer were characterized. From the characterization and 16S rRNA amplicon sequencing, we identified the following isolates: 4 *Bacillus sp.,* 2 *B. licheniformis* isolates*,*

2 *B. subtilis* isolates), 2 *P. aeruginosa* isolates, 2 *Acinetobacter* spp. and one *E. coli* isolate.

Using these isolates, it was possible to test the additional oil recovery potential of indigenous microorganisms in a simulated oilfield system in Chapter 5. Out of the 8 isolates, one (*B. licheniformis* Bi10) recovered significantly higher quantities of oil. This strain was then utilized for further experiments in the porous bioreactor, to reduce bacterial cell starting concentrations for optimum recovery, comparison of oil recovery in a sterile and non-sterile environment and the addition oil recovery of the Bi10 strain when upscaling the experimental design. Most significantly, the Bi10 strain showed a greater increase in oil recovery than any other tested MEOR related strains in this simulated system, including that of the industry standard *B. mojavensis* JF2. It was also identified that the use of metabolically active cells alongside surfactant showed more success in increasing oil recovery than with the presence of only surfactant. The inoculation of spores into the bioreactor also produced high recoveries of crude oil, and the spores of Bi10 also showed germination within the bioreactors. The most important aspect to this work is the fact that simulated microcosm experiments at oilfield parameters were able to successfully show an increase in additional oil recovery using an indigenous strain for MEOR purposes, in a heavy or unconventional oil reservoir system.

The remaining two chapters, Chapters 6 & 7, explored the structure of both the *B.licheniformis* Bi10 isolate and its synthesized lipopeptide. In Chapter 6, the whole genome of Bi10 was sequenced, assembled and annotated to

identify any genes of keen interest. A number of genes specifically linked to biosurfactant production, such as *LchA*, *sfp* and *comA* were recognized, suggesting that the lipopeptide produced by this *Bacillus* strain could be of *lichenysin* nature. Within the genome, a number of differences in DNA sequence were observed between the Bi10 strain and other known *Bacillus spp.* These dissimilarities coincide with the integration of prophage sequences into the genome, of which there are three regions, coding for the intact prophage Bacilli_vB_BanS_Tamsa and two other remnant phage. Furthermore, a major area of dissimilarity between the genomes is identified at the site of the *LchAA* and *Lch* AB genes, the genes responsible for lipopeptide production. This was further corroborated in Chapter 7, by the identification of two regions, of ~1,000 bp in size missing at the terminus of the genes. The loss of genetic material led to an absence of two peptide monomers in the predicted lipopeptide sequence. However, the lipopeptide prediction analysis did show that in all likelihood, the produced biosurfactant was that of a *lichenysin* surface active compound. This *lichenysin* lipopeptide showed a substantial decrease in interfacial tension at simulated reservoir conditions, to similar levels as what is seen in light oils. Furthermore, a reasonable quantity of biosurfactant was recovered from bacterial cultures of Bi10, whilst keeping activity at levels similar to what was seen when they were analyzed in the pure culture.

8.2 Overall perspective and implications of the thesis

With the decrease in the price of oil in recent years, yet reserves in continual decline, Microbial enhanced oil recovery (MEOR) has become an interesting

concept for increasing oil recovery, due to its low costs of implementation and reduced detrimental effects on the environment (Sen, 2008b). By encouraging the production metabolites of, such as biosurfactants and polymers, cell growth and biomass or other metabolism related processes it is possible to increase the amount of oil that we recover from a reservoir (Banat *et al*, 2000; Patel *et al*., 2015; Youssef *et al*, 2009). However, most of the current work has been conducted in laboratory trials or on lighter oils, of which there is an ever reducing amount. The effects of MEOR on less conventional, heavier oils, which are still present in a great abundance, have not been researched in as considerable detail as its lighter counterparts. Furthermore, conventional MEOR trials in the past, both in the laboratory and field have not been conducted under similar conditions to what is found in the reservoir. Therefore, it is no surprise when the outcome of these tests results in failure to increase oil recovery. The use of microorganisms specifically suited to the environment would be of greater benefit than the preferred 'bugs in a jug' approach that the oilfield industry has lived by in years previous (Adams *et al.*, 2015; Gaytán *et al*., 2015). However the process of identifying, characterizing and utilizing these bacteria and their metabolites is far more difficult in practice. This study aimed to investigate the effect of a metabolite produced by an indigenous microorganism on heavy oil recovery in a porous micromodel bioreactor simulated to specific reservoir conditions. If successful, the mechanisms of oil recovery enhancement would also be assessed alongside the behaviour of the microorganism compared to other known MEOR bacteria.

Much of the previous literature has quashed the theories of MEOR as impractical pipedreams, suggesting that only combined mechanisms of biosurfactant production and microbial plugging could harness the full MEOR potential (Bryant, 1990). Yet it had been suggested that a reduction of IFT to levels that would drive oil recovery is not practically achievable at reservoir condition (Gray *et al*, 2008). This factor, alongside issues with the encouragement of microbial proliferation causing near well plugging in some scenarios had led to a standstill in MEOR research, particularly in heavier oils that are usually a lot more difficult to mobilize.

The outcomes of this study however are on the contrary to much of the previous hypotheses. The results documented here indicate that microbial enhanced oil recovery may actually be a feasible treatment strategy for the increased recovery of oil from unconventional oilfields. Furthermore, the use of a microorganism isolated directly from the reservoir of interest (Bi10) proved to have more success in increasing oil recovery, than those non-native microbes chosen for their MEOR ability. Interestingly, the assessment of mechanisms of action appear to demonstrate that not only does IFT reduction play a significant role, but a secondary mechanism, potentially that of biofilm formation and bacterial colonization of the micromodel, is driving the MEOR effect. The micromodel that showed a substantial increase in AOR was not populated with microorganisms that could compete with the isolate, causing repression of growth and a decrease in metabolic activity. Although this was not precisely indicative of reservoir conditions, the remainder of the experimental conditions go far further than the majority of other previous

laboratory tests have. The fact that this study enabled us to account for a multitude of various petrophysical and geochemical characteristics, ranging from pH & temperature to crude oil density and rock porosity, allowed us to simulate the reservoir conditions closer than many research groups have previously been able to, particularly in the laboratory, with the exception of undertaking large scale field trials. This data, combined with the substantial increase in crude oil recovery of a colonized bioreactor suggests that MEOR may be feasible, at least in the laboratory system.

Appendix A

Comparison of heavy oilfield microbiome identified by the direct sequencing of small subunit rRNA and 16S PCR amplification.

A 1.1 Introduction

Environmental microbial communities consist of a multitude of diverse microorganisms, dependent on a number of factors, such as the distinct physical conditions of temperature, pressure and pH. With this diversity arises a difficulty in establishing the true nature of a microbial community, particularly with culture dependent methods. However, the advent of ribosomal RNA gene sequencing as a phylogenetic marker, combined with PCR has enabled research to specifically focus on the microbial ecology of any environment (Ward, Weller, & Bateson, 1990). It is within the small subunit rRNA, that the 16S (microbial) and 18S (eukaryotic) genes are located which have aided in the correct identification of bacteria, archaea & eukarya over the last 20 years (See Section 1.5.2.3).

However, despite the benefits of utilizing PCR for 16S rRNA gene sequencing, there are some inherent bias' associated with this approach for microbial community characterization. These bias' center around the inaccuracy of primers and databases, primer specificity issues and incorrect template amplification (See Section 1.5.3). Should any one of these issues arise within the processing of a sample, the community structure can be completely differential to what is actually present within the community, by either the under/over representation of microbial consortia or lack of their identification. This leads to major problems downstream, such as inaccurate

treatment strategies for disease or clean up processes, inappropriate culturing techniques and misrepresentation on microbial evolution in an environment.

A number of approaches to overcome the PCR related bias is that of direct sequencing. Previously, total metagenomics and metatranscriptomic approaches have been utilized to bypass the need for amplification, but the vast amounts of data and high proportions of non-SSU rRNA matching to 16S regions have made the comparisons against PCR data difficult to compare (Fan et al., 2012; Lanzén et al., 2011). More recent studies (Li *et al.*, 2014; McDonald *et al.,* 2016), have shown success in the direct gel-purification, reverse transcription and sequencing of rRNA from a variety of environmental and mammalian samples. These studies have identified a differential community diversity, with some samples such as spirochetes in a canine microbiome being underrepresented, in PCR amplified samples compared to that of a direct sequence approach (McDonald *et al.,* 2016). This approach, combined with a new pipeline designed specifically to analyze SSU rRNA sequences (BIONMeta, DEN) has allowed us to deal with large amounts of data in a fast timeframe and identify an accurate community composition, made more meaningful by the removal of amplification biases.

This section aims to build on the recent work by McDonald *et al*., 2016, to demonstrate the abilities of direct SSU rRNA gene sequencing on microbial communities outside of a mammalian host, looking at heavy oilfield environmental samples of both liquid and solid state sample types.

A 1.2 Methods

A 1.2.1 Purification of Small-subunit rRNA

RNA extraction of produced water and drilling mud was undertaken following the protocol mentioned in Chapter 2.8. Total RNA was purified on a 1% low melting point agarose gel containing ethidium bromide via β-Agarase I (NEB, USA), following the manufacturers' protocol. Nucleic acids were briefly visualized under blue light and bands around 1.2- 1.8kb excised from the gel ready for purification using sterile scalpel. SSU rRNA was successively purified via Ammonium Acetate and 100% ethanol precipitation as per (McDonald *et al*., 2016). Following centrifugation at 13,000 rpm for 20 minutes, the RNA pellet was washed in 70% ethanol, resuspended in nuclease-free water and stored at -80°C until reverse transcription preparation was undertaken.

A1.2.2 Reverse-transcription of SSU rRNA into double-stranded cDNA

Up to one microgram of gel extracted and purified SSU rRNA from oilfield samples was reverse-transcribed using the Superscript™ Double-Stranded cDNA Synthesis Kit (Thermofisher Scientific, USA) following the manufacturers' protocol and using random primers (9 mers, NEB, USA). Double-stranded cDNA was stored at -20˚C prior to library preparation for sequencing. Library preparation was undertaken as outlined in general methods section (Sections 2.9 & 2.11).

A 1.3 Results

Oilfield characterization using PCR free analysis of oilfield taxonomy via direct RT-SSU rRNA sequencing

Two sample types, collected in triplicate, from a UK onshore heavy oilfield (matching a similar API to the Bentley field) were collected, produced water samples (PW1F) and samples from collected drilling mud (CX1F). Both sample types underwent a dual DNA/RNA extraction following the protocol according to Griffiths *et al (2000)*. The each sample was then split into two and underwent either PCR amplification or direct small subunit rRNA excision and reverse transcription. Both samples were then sequenced on the Illumina MiSeq platform, producing paired end reads of 250x250 bp.

Analysis of the datasets against the RDP bacterial sequence database shows a distinct picture. For PW1F, members of the phyla *Deferibacteres, Firmicutes* and *Proteobacteria* account for >90% of the total abundance and for CX1F. *Acidobacteria, Actinobacteria, Bacteroidetes, Proteobacteria* account for ~75% of total abundance in both the PCR sequenced and direct SSU-rRNA samples. Although the key taxa make up very similar proportions, it is within their individual groups that there is much variation. For PW1F (figure 3.11) *Deferibacteres* are substantially overrepresented at 15.75% to in total abundance of amplicon sample compared to 7.3% in the SSU-rRNA. The percentage of *Proteobacteria* present increases in the SSU-rRNA, from 67% to 74%. However, using universal primers, the 16S amplicon sample does not detect *Planctomycetes (1.5%)* or *Actinobacteria* (2%). A similar result can be seen for CX1F, with *Actinobacteria, Proteobacteria* and *Gammatimonadetes* being over represented in the 16S amplicon sample. Conversely,

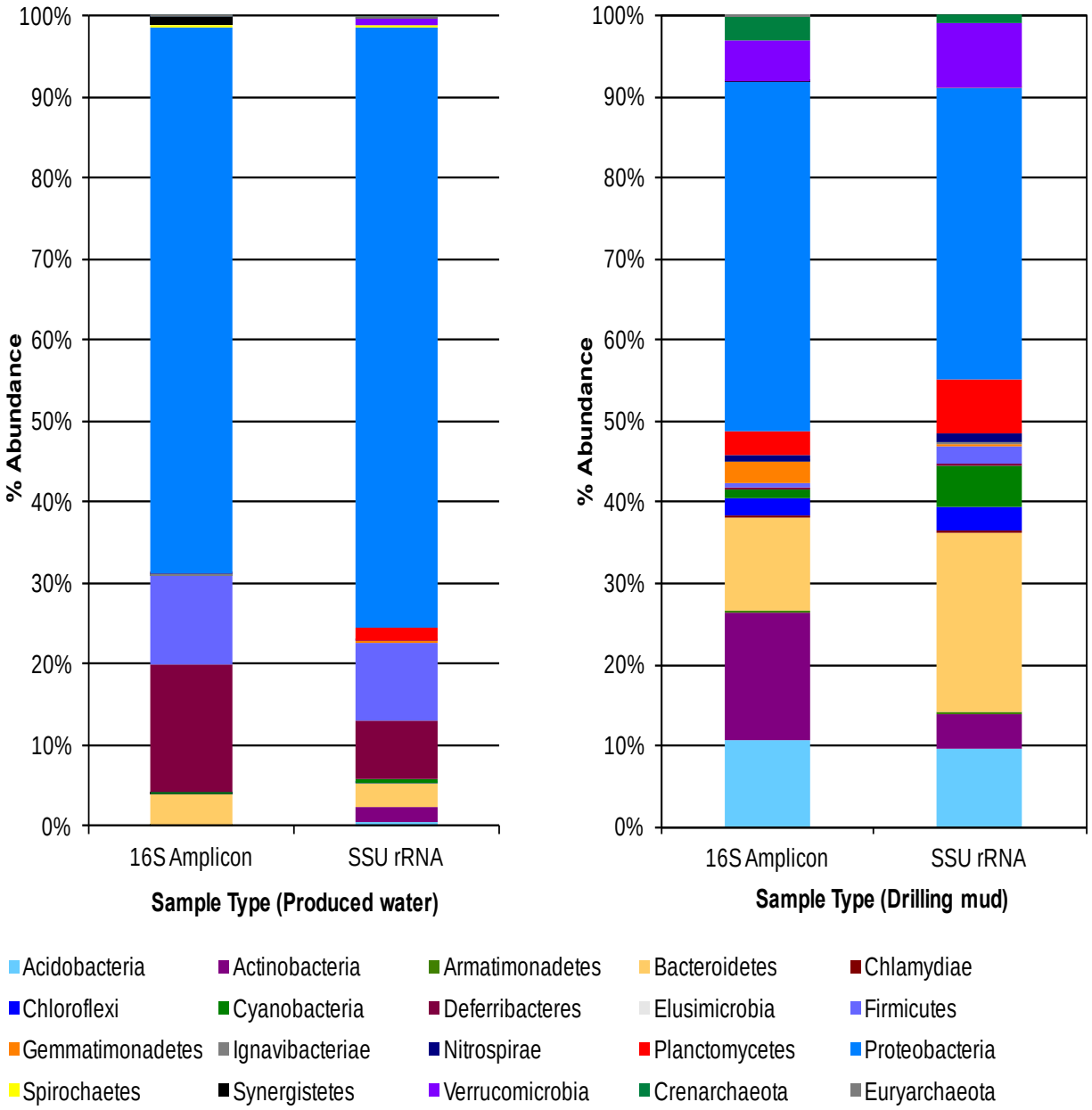


Figure 3.9 Comparison of phylum level characterization of PCR amplicon and RT-SSU rRNA sequence reads derived from. BION-meta was used to classify and compare sequence reads obtained from both amplicon and RT-SSU rRNA sequence data of reservoir production water, PW1F (Left), and drilling mud samples, CX1F (Right), using the RDP database as a reference set. Phylum labelled alphabetically ascending the bar chart

Bacteriodetes, *Cyanobacteria* and *Planctomycetes* are all severely underrepresented.

A 1.4 Discussion

It was anticipated that the PCR/ Direct SSU rRNA comparison would be performed on the reservoir used in this study (Bentley field, 9/03b-7Z). However, due to the reservoir currently being in a state of preproduction and shutdown, it was not practical to collect produced water samples. Even so, the results on this alternate heavy oil reservoir did show some interesting outcomes. The use of the PCR free (Direct SSU) method has previously highlighted some major misinterpretations from the primers of PCR, particularly with Universal 16S primer sets (McDonald *et al*., 2016). Comparing both methods directly for both produced water and drilling mud samples, some bacterial phyla were under/over represented but most critically, some where not identified at all. The data we collected from the alternate reservoir had negative connotations. The drilling mud sample showed the under representation of both *Actinobacteria* and *Bacteriodetes.* For the sample taken from the depths of reservoir itself (PW1F), *Proteobacteria* were the predominant underrepresented phyla in this bacterial phylum contain numerous sulphate reducing bacteria (SRB). Sulphate reducing microorganisms cause souring of the reservoir, which is a reduction in the quality of the oil or gas by the presence of hydrogen sulphide (H_2S). This is not only a physical risk to workers on the production site, but can also cause severe economic issues with pipeline corrosion by formation of iron sulphide

(scale) deposits (Bastin *et al, 1926)* and reducing the quality of the oil. The under representation of these types of bacteria could result in an ineffective treatment strategy of biocide being utilised, which not only could cause insufficient inhibition of detrimental microorganisms , but could also bring about resistance issues within the reservoir microbiota (Davidova *et al, 2001;* Videla & Herrera, 2005*)*. For this reason, it is key to provide as accurate a picture of the reservoir microbiome as technologically possible, an opportunity that this direct sequencing method provides us, without the informatics headache that complete reservoir metagenome sequencing would result in. Furthermore, in non-oilfield related scenarios, this could have detrimental effects, such as the overlooking of disease causing microorganisms in food hygiene or patient diagnostic settings. In the oil industry, although not detrimental to health directly, incorrect diagnosis could have major negative effects such as the identification of MEOR target strains that are not actually in existence.

Appendix B

Identification of known field trials identified from comprehensive literature search of any field trials that assessed the alteration of oil recovery after MEOR treatment

Country	Location	Latitude	Longitude	Year	Site	Oil type	Recovery	Suggested Method	Microorganism
Africa	Alpha	NA	NA	2006	Onshore	Light	15%	Indigenous	Mixed
Africa	Beta	NA	NA	2006	Onshore	Light	25%	Indigenous plugging	Mixed
Argentina	Piedras Coloradas	-40.840492	-65.118137	1997	Onshore	Light	66%	Hydrocarbon degradation	various hydrocarbon degraders
Argentina	Diedama	-45.778036	-67.672011	1994	Onshore	Heavy	0-50%	Exogenous plugging	NA
Australia	Alton, Surat basin	-27.983199	149.315763	1990	Onshore	Light	150%	Exogenous	NA
Brazil	North	-3.713523	-60.952148	2010	Onshore	Light	NA	Indigenous plugging	NA

Canada	Saskatchewan	52.939916	-106.450864	2008	Onshore	Heavy	200%	Indigenous	Mixed
Canada	Saskatchewan	52.939916	-106.450864	1988	Onshore	Heavy	<5 %	Exogenous	*Leuconostoc*
China	Xinjiang	44.069827	84.858398	2015	Onshore	Heavy	307% (short term)	Indigenous surfactant	*Pseudomonas*
China	Dagang, Tianjin	39.043006	117.608991	2007	Onshore	light	13.8-42%	Exogenous/Indigenous	Mixed consortia, SRB, *methanogen*, fermentative
China	Jilin	43.150884	126.44466	1994	Onshore	Light	30%	Indigenous, Gas and Acid	Mixed, *Eubacterium*, *Fusobacterium*, *Bacteriodetes*
China	Chaoyanggou	41.547066	120.625763	2004	Onshore	Light	3%	Exogenous	Unamed
China	Daquing	46.602958	124.570541	2005	Onshore	Light	72%	Hydrocarbon degradation	*Bacillus*, *Brevibacillus*
China	Fuyu	45.013087	126.041336	1996	Onshore	Light	30%	Exogenous	*Enterobacter*, *Bacillus*

China	Qinghai	36.781655	97.198792	1999	Onshore	Light	200%	Exogenous	Mixed hydrocarbon degraders and acid producers
Germany	NA	51.162981	10.463035	1990	Onshore	Light	300%%	Exogenous	*Clostridia*
Hungary	Demjen	47.846616	20.34801	1973	Onshore	Light	60% (short term)	Indigenous, gas and surfactant production	Mixed, SRB, *Pseudomonas*, *Clostridia*
India	Ahmedabad	23.012579	72.501183	2001	Onshore	Light	60%	Indigenous	Multi-bacterial Consortium: *Clostridium Thermoanaerobacterium* sp. and *Thermococcus* sp.
India	Nahorkatiya, Assam	27.286986	95.247552	2000	Onshore	Light	70%	Indigenous	Mixed

Indonesia	Ledok, Java	-6.885785	111.221904	1999	Onshore	Light	12%	Exogenous/Indigenous	Mixed supplemented with B. licheniformis
Malaysia	Bokor	4.202617	113.532715	2000	Offshore	Heavy	2-47%	Indigenous	Mixed
Myanmar	Mann field	20.832502	94.801025	2011	Onshore	Light	10%	EEOR	EEOR
Netherlands	Various	52.677081	6.887827	1958	Onshore	Light	0%	Slime forming	Betacoccus *dextrainicus*
Norway	Norne	65.356532	8.041992	2005	Offshore	Light	NA	NA	NA
Poland	Plawowice	50.172439	20.394573	2011	Onshore	Light	70%	Exogenous	Mixed
Romania	Caldararu	44.450624	24.971066	1992	Onshore	Light	100-200 % over 5 months	Exogenous	*Bacillus*, *Clostridia*, Gram negatives
Romania	Bragadiru	44.356928	25.994511	1990	Onshore	Light	> 50%	Exogenous	*Bacillus*, C*lostridium*, *Arthrobacter*, *Pseudomonas*, Micrococcus
Russia	Vyangpour, Siberia	57.241722	65.714722	1995	Onshore	Light	3%	Indigenous	Mixed

Russia	Siberia	61.01371	99.196656	1995	Onshore	Light	15-30%	Indigenous plugging	*Mixed, Bacillus, Pseudomonas, Actionmyces*
Russia	Romashkino, Tartastan	56.152778	52.489722	1994	Onshore	Light	50%	indigenous	*Clostridia*, *Pseudomonas*
Russia	Romashkinskoye, Urals	60.716568	29.798701	1992	Onshore	Light	15-45%	Indigenous	mixed
Trinidad	Numerous	10.149705	-61.100464	1992	Onshore	Light	0%	Gas producers	numerous facultative anaerobes
UK	Lidsey	41.41684	-83.262291	1991	Onshore	Light	0-10%	Exogenous	Acid fracturing
US	Vassar Vertz, Oklahoma	36.157125	-96.984901	1993	Onshore	Light	< 5%	Indigenous	Mixed
US	Viola, Oklahoma	34.111165	-97.472076	2012	Onshore	Light	10%	Exogenous surfactant	*Bacillus licheniformis* and *subtilis*
US	Nacatoch, Arkansas	32.748916	-93.976776	1954	Onshore	Light	160%	Exogenous plugging	*Clostridia*

US	Delaware-Childers, Oklahoma	36.734439	-95.645795	1990	Onshore	Light	19.20%	Exogenous	*Clostridium*
US	Jack county, Texas	33.256038	-98.221298	1992	Onshore	Light	5%	Indigenous	NA
US	Vasser Vertz, Oklahoma	36.182069	-96.943703	1992	Onshore	Light	300%	Indigenous plugging and gas	Mixed
Venuzeula	Maracaibo	10.527982	-71.801147	1993	Onshore	Light	2%	Indigenous	Mixed

Appendix C

Surfactant & biofilm production analysis of rock core and produced water isolates taken from the Bentley Oilfield, UK.

ID	Sample	Media	Blood Agar	Oil Solubility	Biofilm staining
C1	7400	MB	+	-	0.2739
C2	8650	MB	+	+	0.1551
C3	8650	MB	+	+	0.1335
C4	8650	MB	+++	+	0.2178
C5	8650	MB	+++	+	0.1762
C6	9175	MB	+++	-	0.3928
C7	7400	NB	++	-	0.4549
C8	7400	NB	+	+	0.0717
C9	7400	NB	+	+++	0.2167
C10	9175	0.02	+++	+++	0.4136
C11	9175	NB	++	++	0.5923
C12	9175	NB	++	++	0.596
C13	7400	NB	+++	++	0.3753
C14	7400	NB	++	-	0.5706
C15	8650	NB	++	+++	0.4037
C16	8650	NB	+++	++	0.7207
C17	8650	MB	+	+	0.2021
C18	8650	MB	-	+	0.1128
C19	8650	0.04	+++	+++	0.243
C20	8650	0.04	++	+++	0.4722
C21	8650	MB	+	++	0.2319
C22	7400	0.04	+	+++	0.1602
C23	8650	0.04	++	++	0.4936
C24	8650	MB	+	-	0.3471
C25	8650	MB	++	++	0.3723

ID	Sample	Media	Blood Agar	Oil Solubility	Biofilm staining
C26	7400	MB	+	-	0.267
C27	9175	MB	++	+	0.0918
C28	9175	MB	++	++	0.1881
C29	9175	MB	++	-	0.3742
C30	9175	MB	++	-	0.2397
C31	7400	MB	+	+++	0.5463
C32	8650	MB	+	-	0.2074
C33	9175	MB	+	-	0.3723
C34	7400	MB	+	+	0.3859
C35	7400	MB	+	+	0.7489
C36	7400	MB	+++	+++	0.6671
C37	7400	MB	+	-	1.2427
C38	8650	MB	-	-	0.2774
C39	8250	0.04	+++	-	0.3886
C40	7400	0.04	+++	-	0.3594
C41	9175	0.04	+	-	0.1904
C42	9175	0.04	+	+++	0.1506
C43	8650	MB	+	++	0.151
C44	8250	0.02	+	++	0.3837
C45	8650	0.02	+	+	0.3286
C46	8650	0.02	+	-	0.5463
C47	8650	0.02	-	-	0.6275
C48	8650	0.02	++	+	0.7743
C49	8650	0.02	+	-	0.6825
C50	8650	0.02	+	-	0.441

ID	Sample	Media	Blood Agar	Oil Solubility	Biofilm staining
C51	8250	0.005	+	-	0.2793
C52	8650	MB	+	+++	0.797
C53	7400	0.005	+	++	0.136
C54	7400	0.005	-	-	0.2372
C55	8250	NB	+	++	0.2577
C56	7400	NB	-	-	0.1328
C57	8250	NB	-	-	0.3863
C58	8650	NB	-	-	0.4224
C59	8250	NB	+	-	0.3795
C60	8250	NB	+	++	0.4713
C61	8650	0.02	+	-	0.6895
C62	8650	0.02	++	-	0.611
C63	8650	MB	++	+	0.3234
C64	8250	MB	++	-	0.5957
C65	7400	MB	+	-	0.4474
C66	7400	NB	+	-	0.4462
C67	8650	MB	+++	+	0.3219
C68	8650	MB	+	-	0.4047
C69	8650	MB	+++	-	0.9448
C70	8650	MB	+	+	0.3807
C71	7400	MB	+	-	1.1777
C72	8650	MB	+	-	0.4896
C73	9175	0.04	+	+	0.5481
C74	8650	MB	+	++	0.5437
C75	8650	0.02	+	-	1.4154
C76	8650	MB	+	++	0.4682
C77	9175	NB	++	++	0.4936
C78	8250	NB	-	++	1.0061
C79	8650	0.005	-	-	0.2383
C80	8250	NB	++	++	0.5615
C81	8650	NB	-	++	0.1982
C82	7400	0.04	++	+++	0.7566
C83	8250	0.02	+	++	0.5035

ID	Sample	Media	Blood Agar	Oil Solubility	Biofilm staining
C84	7400	NB	+	-	0.4674
C85	7400	NB	+	-	0.3444
C86	8250	0.04	+++	+	0.7885
C87	8250	0.04	+	++	0.6439
C89	8250	MB	+	+	0.3858
C90	8250	0.04	+	+	0.5404
C91	7400	NB	+	++	0.6502
C92	9175	0.04	+	++	0.9724
C93	9175	NB	+	-	0.2109
C94	8250	NB	+	+	0.4713
C95	9175	NB	+	+	0.2463
C96	8250	MB	++	++	0.9584
C97	9175	MB	++	+	0.2426
C98	9175	MB	++	+++	0.4616
C99	9175	NB	+	++	0.8129
C100	7400	MB	+	++	0.4789
C101	8650	NB	++	++	0.6916
C102	7400	MB	+	++	0.2975
C103	7400	MB	+	+++	0.4224
C104	8650	MB	+	++	0.411
C105	7400	MB	+	-	0.3127
C106	8650	NB	+	++	0.1851
C107	9175	MB	+	+	0.5717
C108	7400	0.02	+++	++	0.2383
C109	9175	MB	+	++	0.3211
C110	8250	NB	+	+++	0.3646
C111	8650	NB	+	+	0.5984
C112	8250	0.02	-	+	0.9421
C113	8650	0.02	-	+	0.9527
C114	9175	MB	+	+++	0.3778
C115	8650	0.02	+	-	0.1873
C116	8650	MB	+	-	0.2419
C117	8650	MB	+	+++	0.1801

ID	Sample	Media	Blood Agar	Oil Solubility	Biofilm staining
C118	9175	MB	+	+++	0.407
C119	9175	MB	-	++	0.279
C120	8650	MB	-	++	0.3375
C121	8650	MB	-	++	0.1952
C122	9175	MB	+	++	0.3191
C123	9175	MB	-	-	0.2273
C124	9175	MB	+	+	0.4029
C125	9175	NB	-	-	0.9763
C126	9175	MB	+	+	0.2334
C127	9175	MB	-	+	0.3127
C128	8250	NB	-	++	0.0806
C129	8650	NB	++	++	0.9835
C130	8650	NB	+	++	0.4367
C131	8650	MB	+	++	0.2193
C132	8250	NB	+	-	0.2857
C133	8650	NB	+	++	0.3175
C134	7400	NB	-	++	0.3858
C135	8650	NB	-	++	0.2593
C136	8650	NB	+	++	0.1852
C137	8250	NB	-	++	0.1947
C138	8250	NB	-	+	0.5824
C139	8650	NB	-	+	0.2015
C140	8650	MB	+	-	0.3059
C141	8650	NB	+	-	0.7119
C142	7400	NB	+	+++	0.9263
C143	8650	NB	-	++	1.228
C144	8650	0.005	-	-	0.6634
C145	8650	0.005	+	+	0.0704
C146	8650	0.005	-	+	0.0562
C147	8250	0.005	-	-	0.1272
C148	7400	0.005	-	++	0.1493
C149	8650	NB	-	+++	0.7154
C150	8250	NB	-	-	0.5472

ID	Sample	Media	Blood Agar	Oil Solubility	Biofilm staining
C151	9175	NB	-	++	1.1255
C152	9175	NB	-	+	0.2189
C153	9175	NB	-	-	0.164
C154	9175	NB	-	-	0.8233
C155	9175	NB	-	+	0.2243
C156	9175	MB	-	-	0.2875
C157	8650	NB	-	++	0.6024
C158	8650	NB	+	+	0.4057
C159	7400	0.04	+	++	0.2167
C160	9175	NB	+	++	0.4828
W1	1	NB	++	+	0.6203
W2	1	NB	++	+	0.7426
W3	1	NB	-	-	0.1428
W4	1	NB	-	-	0.2431
W5	1	NB	++	+	0.1089
W6	1	MB	+	-	0.773
W7	1	MB	-	-	0.0892
W8	1	MB	-	-	0.0791
W9	1	MB	-	-	0.341
W10	1	NB	-	-	0.2301
W11	1	NB	+	-	0.4456
W12	1	NB	+	-	0.3364
W13	1	0.005	-	-	0.1068
W14	1	0.005	-	-	0.1305
W15	1	0.005	-	-	0.0966
W16	1	NB	-	-	0.0432
W17	1	NB	+	-	0
W18	1	MB	+	-	0.1935
W19	1	NB	-	-	0.1348
W20	1	NB	-	-	0.3639
W21	1	NB	-	-	0.2309
W22	1	NB	-	-	0.2801
W23	1	NB	-	-	0.2154

ID	Sample	Media	Blood Agar	Oil Solubility	Biofilm staining
W24	1	NB	-	-	0.2416
W25	1	MB	-	-	0.3251
W26	1	MB	-	-	0.2226
W27	1	NB	-	-	0.0674
W28	1	MB	-	-	0.2228
W29	1	MB	-	-	0.2463
W30	1	MB	-	-	0.3039
W31	1	MB	-	-	0.1473
W32	1	MB	-	-	0.3035
W33	1	MB	-	-	0.5673
W34	1	MB	-	-	0.3953
W35	1	MB	+	-	0.7816
W36	1	NB	+	-	0.8004
W37	2	MB	-	-	0.3782
W38	2	MB	-	-	0.0776
W39	2	NB	-	-	0.2725
W40	2	NB	-	-	0.1608
W41	2	MB	-	-	0.3282
W42	2	MB	-	-	0.2116
W43	2	MB	-	-	0.4091
W44	2	MB	-	-	0.2984
W45	2	MB	-	-	0.1105
W46	2	MB	-	-	0.4009
W47	2	MB	+	-	0.7428
W48	2	MB	-	-	0.2184
W49	2	NB	-	-	0.2735
W50	2	MB	-	-	0.0856
W51	2	MB	-	-	0.0279
W52	2	NB	-	-	0.0583
W53	1	NB	-	-	0.1152
W54	1	0.02	+	-	0.2091
W55	1	NB	-	-	0.0497
W56	1	NB	-	-	0.247

ID	Sample	Media	Blood Agar	Oil Solubility	Biofilm staining
W57	2	NB	-	-	0.0399
W58	2	NB	-	-	0.3186
W59	2	NB	-	-	0.1132
W60	2	NB	-	-	0.0758
W61	2	NB	-	-	0.2052
W62	2	NB	-	-	0.8417
W63	2	MB	-	-	0.0651
W64	2	MB	-	-	0.0577
W65	2	MB	-	-	0.0697
W66	2	MB	-	-	0.2021
W67	2	MB	-	-	0.6094
W68	2	MB	+	-	0.1895
W69	2	MB	-	-	0.3961
W70	2	MB	-	-	0.4944
W71	2	MB	-	-	0.068
W72	2	MB	-	-	0.1368
W73	2	MB	-	-	0.0857
W74	2	0.04	++	-	0.7942
W75	2	0.04	++	++	0.6652
W76	2	MB	-	-	0.1873
W77	2	MB	+	-	0.1398
W78	2	MB	-	-	0.154
W79	2	MB	-	+	0.1384
W80	2	MB	-	-	0.2659
W81	3	0.02	+	-	0.1085
W82	3	NB	-	-	0.271
W83	3	NB	-	-	1.228
W84	3	NB	-	-	0.3684
W85	3	NB	-	-	0.2475
W86	3	NB	-	-	0.2839
W87	3	NB	-	-	0.3386
W88	3	NB	-	+	0.5409
W89	3	NB	-	-	0.134

ID	Sample	Media	Blood Agar	Oil Solubility	Biofilm staining
W90	3	NB	-	-	0.4082
W91	3	NB	-	-	0.361
W92	3	NB	+	-	0.1715
W93	3	0.04	++	-	0.2891
W94	3	0.02	-	-	0.1662
W95	3	0.02	-	-	0.0952
W96	3	0.02	+	-	0.2529
W97	3	0.04	++	+	0.3836
W98	3	MB	+	-	0.1892
W99	3	0.05	-	-	0.173
W100	3	0.05	-	-	0.5047
W101	3	0.05	-	-	0.107
W102	3	0.05	-	-	0.564
W103	3	0.02	+	-	0.0982
W104	3	MB	-	-	0.3481
W105	3	MB	-	-	0.3619
W106	3	MB	-	-	0.2775
W107	3	MB	-	-	0.3615
W108	2	MB	-	-	0.3101
W109	2	MB	-	-	0.1485
W110	2	MB	-	-	0.4688
W111	2	MB	-	-	0.1552
W112	2	MB	-	-	0.2587
W113	2	0.04	-	-	0.1094
W114	2	MB	-	-	0.278
W115	1	MB	+	-	0.6725
W116	1	MB	-	-	0.4731
W117	1	MB	-	-	0.2437
W118	1	MB	-	-	0.2623
W119	1	MB	-	-	0.1679
W120	1	MB	-	-	0.0805
W121	1	MB	-	-	0.1064
W122	1	MB	-	-	0.1811

ID	Sample	Media	Blood Agar	Oil Solubility	Biofilm staining
W123	1	MB	-	-	0.3187
W124	1	MB	+	-	0.6742
W125	1	MB	-	-	0.2183
W126	3	MB	-	-	0.0709
W127	3	MB	+	-	0.4892
W128	3	MB	++	++	0.8924
W129	3	MB	++	+	0.5063
W130	1	MB	-	-	0.445
W131	1	MB	-	-	0.2739
W132	1	MB	-	-	0.284
W133	1	MB	-	-	0.0916
W134	1	MB	-	-	0.698
W135	1	MB	-	-	0.4814
W136	1	MB	-	-	0.222
W137	1	MB	+	+	0.8549
W138	1	MB	+	-	0.3062
W139	1	MB	-	-	0.1068
W140	1	MB	-	-	0.1545
W141	1	MB	-	-	0.1992
W142	1	MB	-	-	0.203
W143	1	MB	-	-	0.3504
W144	1	NB	-	-	0.0623
W145	1	NB	-	-	0.0901
W146	2	NB	-	-	0.1592
W147	2	NB	-	+	0.0984
W148	2	NB	-	-	0.0782
W149	2	0.005	-	-	0.3623

List of references

Abbasian, F., Palanisami, T., Megharaj, M., Naidu, R., Lockington, R., & Ramadass, K. (2016). Microbial Diversity and Hydrocarbon Degrading Gene Capacity of a Crude Oil Field Soil as Determined by Metagenomics Analysis.

Abdelhafiz, Y. A., Manaharan, T., Mohamad, S. Bin, & Merican, A. F. (2017). Whole genome sequencing and functional features of UMX-103: a new Bacillus strain with biosurfactant producing capability. *Genes and Genomics*, *39*(8), 877–886.

Abdolahi, A., Hamzah, E., Ibrahim, Z., & Hashim, S. (2014). Microbially influenced corrosion of steels by Pseudomonas aeruginosa. *Corrosion Reviews*, *32*(3–4), 129–141.

Adams, G. O., Fufeyin, P. T., Okoro, S. E., & Ehinomen, I. (2015). Bioremediation, Biostimulation and Bioaugmention: A Review. *International Journal of Environmental Bioremediation & Biodegradation*, *3*(1), 28–39.

Aellen, S., Que, Y. A., Guignard, B., Haenni, M., & Moreillon, P. (2006). Detection of live and antibiotic-killed bacteria by quantitative real-time PCR of specific fragments of rRNA. *Antimicrobial Agents and Chemotherapy*, *50*(6), 1913–1920.

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., … Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research*, *44*(W1), W3–W10.

Aguirre-Ramirez, M. , Medina, G., Gonzalez-Valdez, A., Grosso-Becerra, V., & Soberon-Chavez, G. (2012). The Pseudomonas aeruginosa rmlBDAC operon, encoding dTDP-L-rhamnose biosynthetic enzymes, is regulated by the quorum-sensing transcriptional regulator RhlR and the alternative sigma factor ?? S. *Microbiology*, *158*(4), 908–916.

Al-Haleem, A A, Abdulah, H H and Saeed, E. A. (2010). Components and treatments of oilfield produced water. *Al-Khwarizmi Engineering Journal*, *6*(1), 24–30.

Al-Sayegh, A., Al-Wahaibi, Y., Al-Bahry, S., Elshafie, A., Al-Bemani, A., & Joshi, S. (2015). Microbial enhanced heavy crude oil recovery through biodegradation using bacterial isolates from an Omani oil field. *Microbial Cell Factories*, *14*(1), 141.

Alasvand Zarasvand, K., & Ravishankar Rai, V. (2016). Identification of the traditional and non-traditional sulfate-reducing bacteria associated with corroded ship hull. *3 Biotech*, *6*(2), 197.

Albokari, M., Mashhour, I., Alshehri, M., Boothman, C., & Al-Enezi, M. (2015). Characterization of microbial communities in heavy crude oil from Saudi Arabia. *Annals of Microbiology*, *65*(1), 95–104.

Aleti, G., Sessitsch, A., & Brader, G. (2015). Genome mining: Prediction of lipopeptides and polyketides from Bacillus and related Firmicutes. *Computational and Structural Biotechnology Journal*. Research Network of Computational and Structural Biotechnology.

Alley, B., Beebe, A., Rodgers, J., & Castle, J. W. (2011). Chemical and physical characterization of produced waters from conventional and unconventional fossil fuel resources. *Chemosphere*, *85*(1), 74–82.

Almeida, P. F., Moreira, R. S., Almeida, R. C. C., Guimarães, A. K., Carvalho, A. S., Quintella, C., Taft, C. A. (2004). Selection and Application of Microorganisms to Improve Oil Recovery. *Engineering in Life Sciences*, *4*(4), 319–325.

Alvarez Yela, A. C., Tibaquirá Martínez, M. A., Rangel Piñeros, G. A., López, V. C., Villamizar, S. H., Núñez Vélez, V. L., González Barrios, A. F. (2016). A comparison

between conventional Pseudomonas aeruginosa rhamnolipids and Escherichia coli transmembrane proteins for oil recovery enhancing. *International Biodeterioration and Biodegradation*, *112*, 59–65.

Amann, R. I., Ludwig, W., Schleifer, K. H., Amann, R. I., & Ludwig, W. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation . Phylogenetic Identification and In Situ Detection of Individual Microbial Cells without Cultivation, *59*(1), 143–169.

Andrew, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Arima, K., Kakinuma, A., & Tamura, G. (1968). Surfactin, a crystalline peptidelipid surfactant produced by Bacillus Subtilis: Isolation, characterization and its inhibition of fibrin clot formation. *Biochemical and Biophysical Research Communications*, *31*(3), 488–494.

Aßhauer, K. P., Wemheuer, B., Daniel, R., & Meinicke, P. (2015). Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data: Fig. 1. *Bioinformatics*, *31*(17), 2882–2884.

Asshifa Md Noh, N., Al-Ashraf Abdullah, A., Nasir Mohamad Ibrahim, M., & Ramli Mohd Yahya, A. (2012). Rhamnolipid produced by Pseudomonas aeruginosa USM-AR2 facilitates crude oil distillation. *The Journal of General and Applied Microbiology*, *58*(2), 153–161.

Augustine, A., Prathap, P., Kamarudheen, N., & Rao, K. V. B. (2017). Isolation, Screening and partial identification of Indigenous Biosurfactant Bacteria from Oil Contaminated Soil. *Research J. Pharm. and Tech*, *10*(6).

Augustinovic, Z., Birketveit, O., Clements, K., Freeman, M., Gopi, S., & Ishoey, T. (2012). Microbes: oilfield enemies or allies. *Oilfield Review*, *24*, 1–14.

Avseth, P., Dvorkin, J., Mavko, G., Rykkje, J. (1998). Diagnosing High-porosity Sands For Reservoir Characterization Using Sonic And Seismic. *Society of Exploration Geophysicists.*

Awan, A. R., Tiegland, R., & Kleppe, J. (2008). A Survey of North Sea Enhanced-Oil-Recovery Projects Initiated During the Years 1975 to 2005. *SPE Res Eval & Eng*, *11*(3), 497–512.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., … Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, *9*(1), 75.

Banat, I. M. (1993). The isolation of a thermophilic biosurfactant producing Bacillus SP. *Biotechnology Letters*, *15*(6), 591–594.

Banat, I. M. (1995). Biosurfactants production and possible uses in microbial enhanced oil recovery and oil pollution remediation: A review. *Bioresource Technology*, *51*(1), 1–12.

Banat, I. M., Makkar, R. S., & Cameotra, S. S. (2000). Potential commercial applications of microbial surfactants. *Applied Microbiology and Biotechnology*, *53*(5), 495–508.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, *19*(5), 455–77.

Bao, M., Pi, Y., Wang, L., Sun, P., Li, Y., & Cao, L. (2014). Lipopeptide biosurfactant production bacteria Acinetobacter sp. D3-2 and its biodegradation of crude oil. *Environmental Science. Processes & Impacts*, *16*(4), 897–903.

Bartholomew, J. W., & Mittwer, T. (1952). The gram stain. *Bacteriological Reviews*, *16*(7), 1–

29.

Bastin, E. S. (1926). The Problem of the Natural Reduction of Sulphates. *AAPG Bulletin*, *10*.

Bastin, E. S., Greer, F. E., Merritt, C. A., & Moulton, G. (1926). THE PRESENCE OF SULPHATE REDUCING BACTERIA IN OIL FIELD WATERS. *Science (New York, N.Y.)*, *63*(1618), 21–4.

Blazewicz, S. J., Barnard, R. L., Daly, R. A., & Firestone, M. K. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *The ISME Journal*, *7*(11), 2061–2068.

Bordoloi, N. K., & Konwar, B. K. (2008). Microbial surfactant-enhanced mineral oil recovery under laboratory conditions. *Colloids and Surfaces B: Biointerfaces*, *63*(1), 73–82.

Bowman, J. S., Rasmussen, S., Blom, N., Deming, J. W., Rysgaard, S., & Sicheritz-Ponten, T. (2012). Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene. *The ISME Journal*, *6*(1), 11–20.

Brennan, B. H., Lucas-Clements, C., Kew, S., & Dempsey, P. F. (2014). De-risking a Heavy Oil Development - A Case Study of the Bentley Field, UKCS Block 9/3b. In *SPE International Heavy Oil Conference and Exhibition* (pp. 1–15).

Brooks, J. P., Edwards, D. J., Harwich, M. D., Rivera, M. C., Fettweis, J. M., Serrano, M. G., … Wiener, M. (2015). The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiology*, *15*(1), 66.

Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Adler, C. J., … Cooper, A. (2013). Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nature Communications*, *4*.

Brown, L.R., Vadie, A.A., Stephens, J. . (2000). Slowing production decline and extending the economic life of an oil field: new MEOR technology. *SPE Paper 59306,*.

Brown, L. R. (2010). Microbial enhanced oil recovery (MEOR). *Current Opinion in Microbiology*, *13*(3), 316–320.

Bryant, R. (1990). *Topical report: Screening criteria for Microbial EOR process*.

Bulat, S. A., Alekhina, I. A., Marie, D., Martins, J., & Petit, J. R. (2011). Searching for life in extreme environments relevant to Jovian's Europa: Lessons from subglacial ice studies at Lake Vostok (East Antarctica). *Advances in Space Research*, *48*(4), 697–701.

Bushley, K. E., & Turgeon, B. G. (2010). Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC Evolutionary Biology*, *10*(1), 26.

Bussotti, G., Notredame, C., & Enright, A. J. (2013). Detecting and comparing non-coding RNAs in the high-throughput era. *International Journal of Molecular Sciences*, *14*(8), 15423–58.

Cai, M., Nie, Y., Chi, C.-Q., Tang, Y.-Q., Li, Y., Wang, X.-B., … Wu, X.-L. (2015). Crude oil as a microbial seed bank with unexpected functional potentials. *Scientific Reports*, *5*, 16057.

Cameotra, S. S., Makkar, R. S., Kaur, J., & Mehta, S. K. (2010). Synthesis of biosurfactants and their advantages to microorganisms and mankind. *Advances in Experimental Medicine and Biology*, *672*, 261–280.

Cao, B., Nagarajan, K., & Loh, K.-C. (2009). Biodegradation of aromatic compounds: current status and opportunities for biomolecular approaches. *Applied Microbiology and Biotechnology*, *85*(2), 207–228.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., … Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, *6*(8), 1621–1624.

Carrillo, P. G., Mardaraz, C., Pitta-Alvarez, S. I., & Giulietti, A. M. (1996). Isolation and selection of biosurfactant-producing bacteria. *World Journal of Microbiology and Biotechnology*, *12*(1), 82–84.

Castorena-Cortés, G., Zapata-Peñasco, I., Roldán-Carrillo, T., Reyes-Avila, J., Mayol-Castillo, M., Román-Vargas, S., & Olguín-Lora, P. (2012). Evaluation of indigenous anaerobic microorganisms from Mexican carbonate reservoirs with potential MEOR application. *Journal of Petroleum Science and Engineering*, *81*, 86–93.

Chai, L., Zhang, F., She, Y., & Banat, I. (2015). Impact of a Microbial-Enhanced Oil Recovery Field Trial on Microbial Communities in a Low-Temperature Heavy Oil Reservoir. *Nature Environment and Pollution Technology*, *14*(3), 455–462.

Chang, P. L., & Fu Yen, T. (1984). Interaction of Escherichia coli B and B/4 and Bacteriophage T4D with Berea Sandstone Rock in Relation to Enhanced Oil Recovery. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, *47*(3), 544–550.

Chaudhry, V., Sharma, S., Bansal, K., & Patil, P. B. (2016). Glimpse into the Genomes of Rice Endophytic Bacteria: Diversity and Distribution of Firmicutes. *Frontiers in Microbiology*, *7*, 2115.

Chen, J., Huang, P. T., Zhang, K. Y., & Ding, F. R. (2012). Isolation of biosurfactant producers, optimization and properties of biosurfactant produced by Acinetobacter sp. from petroleum-contaminated soil. *Journal of Applied Microbiology*, *112*(4), 660–671.

Cheng, M., Lei, G., Gao, J., Xia, T., & Wang, H. (2014). Laboratory Experiment, Production Performance Prediction Model, and Field Application of Multi-slug Microbial Enhanced Oil Recovery. *Energy & Fuels*, *28*(10), 6655–6665.

Cliften, P. (2015). Base Calling, Read Mapping, and Coverage Analysis. In *Clinical Genomics* (pp. 91–107). Elsevier.

Cluff, M. A., Hartsock, A., MacRae, J. D., Carter, K., & Mouser, P. J. (2014). Temporal Changes in Microbial Ecology and Geochemistry in Produced Water from Hydraulically Fractured Marcellus Shale Gas Wells. *Environmental Science & Technology*, *48*(11), 6508–6517.

Coker, J. A. (2016). Extremophiles and biotechnology: current uses and prospects. *F1000Research*, *5*, 396.

Coutte, F., Leclère, V., Béchet, M., Guez, J. S., Lecouturier, D., Chollet-Imbert, M., … Jacques, P. (2010). Effect of pps disruption and constitutive expression of srfA on surfactin productivity, spreading and antagonistic properties of Bacillus subtilis 168 derivatives. *Journal of Applied Microbiology*, *109*(2), 480–491.

Daaou, M., & Bendedouch, D. (2012). Water pH and surfactant addition effects on the stability of an Algerian crude oil emulsion. *Journal of Saudi Chemical Society*, *16*(3), 333–337. https://doi.org/10.1016/j.jscs.2011.05.015

Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE*, *5*(6), e11147

Daryasafar, A., Jamialahmadi, M., Moghaddam, M. B., & Moslemi, B. (2016). Using biosurfactant producing bacteria isolated from an Iranian oil field for application in microbial enhanced oil recovery. *Petroleum Science and Technology*, *34*(8), 739–746.

Das, P., Mukherjee, S., & Sen, R. (2008). Antimicrobial potential of a lipopeptide

biosurfactant derived from a marine Bacillus circulans. *Journal of Applied Microbiology*, *104*(6), 1675–1684.

Das, P., Mukherjee, S., & Sen, R. (2008). Genetic Regulations of the Biosynthesis of Microbial Surfactants: An Overview. *Biotechnology and Genetic Engineering Reviews*, *25*(1), 165–186.

Dastgheib, S. M. M., Amoozegar, M. A., Elahi, E., Asad, S., & Banat, I. M. (2008). Bioemulsifier production by a halothermophilic Bacillus strain with potential applications in microbially enhanced oil recovery. *Biotechnology Letters*, *30*(2), 263–270.

Davidova, I., Hicks, M. S., Fedorak, P. M., & Suflita, J. M. (2001). The influence of nitrate on microbial processes in oil industry production waters. *Journal of Industrial Microbiology & Biotechnology*, *27*(2), 80–6.

Davidson, & S., W. (1988). An MEOR Pilot Test in the Loco Field. *US. DOE Rep.*

De Bruijn, I., De Kock, M. J. D., Yang, M., De Waard, P., Van Beek, T. A., & Raaijmakers, J. M. (2007). Genome-based discovery, structure prediction and functional analysis of cyclic lipopeptide antibiotics in Pseudomonas species. *Molecular Microbiology*, *63*(2), 417–428.

Degre, G. (2014). Chemical EOR for Heavy Oil: The Canadian Experience. In *SPE EOR Conference at Oil and Gas West Asia*. Society of Petroleum Engineers.

Deleu, M., Paquot, M., & Nylander, T. (2008). Effect of Fengycin, a Lipopeptide Produced by Bacillus subtilis, on Model Biomembranes. *Biophysical Journal*, *94*(7), 2667–2679.

Demba Diallo, M., Reinhold-Hurek, B., Hurek, T., JP, M., E, O., GF, S., … K, V. der G. (2008). Evaluation of PCR primers for universal nifH gene targeting and for assessment of transcribed nifH pools in roots of Oryza longistaminata with and without low nitrogen input. *FEMS Microbiology Ecology*, *65*(2), 220–228.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., … Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, *36*(Web Server), W465–W469.

Desai, A., Marwah, V. S., Yadav, A., Jha, V., Dhaygude, K., Bangar, U., … Jere, A. (2013). Identification of Optimum Sequencing Depth Especially for De Novo Genome Assembly of Small Genomes Using Next Generation Sequencing Data. *PLoS ONE*, *8*(4), e60204.

Desai, J. D., & Banat, I. M. (1997). Microbial production of surfactants and their commercial potential. *Microbiology and Molecular Biology Reviews : MMBR*, *61*(1), 47–64.

Dhasayan, A., Selvin, J., & Kiran, S. (2015). Biosurfactant production from marine bacteria associated with sponge Callyspongia diffusa. *3 Biotech*, *5*(4), 443–454.

Domingos, D. F., de Faria, A. F., de Souza Galaverna, R., Eberlin, M. N., Greenfield, P., Zucchi, T. D., … de Oliveira, V. M. (2015). Genomic and chemical insights into biosurfactant production by the mangrove-derived strain Bacillus safensis CCMA-560. *Applied Microbiology and Biotechnology*, *99*(7), 3155–3167.

Dong, H., Xia, W., Dong, H., She, Y., Zhu, P., Liang, K., … Zhang, G. (2016). Rhamnolipids produced by indigenous Acinetobacter junii from petroleum reservoir and its potential in enhanced oil recovery. *Frontiers in Microbiology*, *7*(NOV), 1710.

Dong, H., Zhang, Z. Z., He, Y. L., Luo, Y. J., Xia, W. J., Sun, S. S., … Gao, D. L. (2015). Biostimulation of biogas producing microcosm for enhancing oil recovery in

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Edwards, U., Rogall, T., Blöcker, H., Emde, M., & Böttger, E. C. (1989). Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Research*, *17*(19), 7843–53.

Ehlig-Economides, C. A., Chan, K. S., & Spath, J. B. (1996). Production Enhancement Strategies for Strong Bottom Water Drive Reservoirs. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.

Ehrenberg S N, & Nadeau P H. (2005). Sandstone vs. carbonate petroleum reservoirs: A global perspective on porosity-depth and porosity-permeability relationships. *The American Association of Petroleum Geologists Bulletin*, *89*(4), 435–445.

Ekblom, R., & Wolf, J. B. W. (2014a). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, *7*(9), 1026–1042.

Ekblom, R., & Wolf, J. B. W. (2014b). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, *7*(9), 1026–42.

El-Sheshtawy, H. S., Aiad, I., Osman, M. E., Abo-ELnasr, A. A., & Kobisy, A. S. (2016). Production of biosurfactants by Bacillus licheniformis and Candida albicans for application in microbial enhanced oil recovery. *Egyptian Journal of Petroleum*, *25*(3), 293–298.

Elazzazy, A. M., Abdelmoneim, T. S., & Almaghrabi, O. A. (2015). Isolation and characterization of biosurfactant production under extreme environmental conditions by alkali-halo-thermophilic bacteria from Saudi Arabia. *Saudi Journal of Biological Sciences*, *22*(4), 466–475.

Elyasi, S. (2016). Assessment and evaluation of degree of multilateral well's performance for determination of their role in oil recovery at a fractured reservoir in Iran. *Egyptian Journal of Petroleum*, *25*(1), 1–14.

Endicott, P., Sanchez, J. J., Pichler, I., Brotherton, P., Brooks, J., Egarter-Vigl, E., … Pramstaller, P. (2009, June 19). 1471-2156-10-29-S3.pdf. *BMC Genetics*. BioMed Central.

Energy, R. (2009). *An Independent Evaluation Report on the Bentley Field, UKCS Block 9/3b*.

Eskandari, S., Rashedi, H., Ziaie-Shirkolaee, Y., Mazaheri-Assadi, M., Jamshidi, E., & Bonakdarpour, B. (2009). Evaluation of oil recovery by rhamnolipid produced with isolated strain from Iranian oil wells. *Annals of Microbiology*, *59*(3), 573–577.

Fahrenfeld, N. L., Delos Reyes, H., Eramo, A., Akob, D. M., Mumford, A. C., & Cozzarelli, I. M. (2017). *Shifts in microbial community structure and function in surface waters impacted by unconventional oil and gas wastewater revealed by metagenomics*. *Science of The Total Environment* (Vol. 580).

Fakhru'l-Razi, A., Pendashteh, A., Abdullah, L. C., Biak, D. R. A., Madaeni, S. S., & Abidin, Z. Z. (2009). Review of technologies for oil and gas produced water treatment. *Journal of Hazardous Materials*, *170*(2), 530–551.

Fakruddin, M., & Mannan, K. S. Bin. (2013). Methods for Analyzing Diversity of Microbial Communities in Natural Environments. *Ceylon Journal of Science (Biological Sciences)*, *42*(1), 19–33.

Fan, L., McElroy, K., & Thomas, T. (2012). Reconstruction of ribosomal RNA genes from metagenomic data. *PloS One*, *7*(6), e39948.

Felnagle, E. A., Jackson, E. E., Chan, Y. A., Podevels, A. M., Berti, A. D., McMahon, M. D., & Thomas, M. G. (2008). Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Molecular Pharmaceutics*, *5*(2), 191–211.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., … Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279–D285.

Fleischmann, R., White, O., Adams, M., Clayton, R., Kirkness, E., Kerlavage, A., … Merrick, J. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science*, *269*(5223), 496–512.

Folmsbee, M. J., McInerney, M. J., & Nagle, D. P. (2004). Anaerobic growth of Bacillus mojavensis and Bacillus subtilis requires deoxyribonucleosides or DNA. *Applied and Environmental Microbiology*, *70*(9), 5252–5257.

Frank, Y., Banks, D., Avakian, M., Antsiferov, D., Kadychagov, P., & Karnachuk, O. (2016). Firmicutes is an Important Component of Microbial Communities in Water-Injected and Pristine Oil Reservoirs, Western Siberia, Russia. *Geomicrobiology Journal*, *33*(5), 387–400.

Galabova, D., Sotirova, A., & Karpenko, A. (2014). Role of Microbial Surface-Active Compounds in Environmental Protection. In M. Fanun (Ed.), *The Role of Colloidal Systems in Environmental Protection* (pp. 41–83). Elsevier.

Gao, C. H. (2011). Microbial Enhanced Oil Recovery in Carbonate Reservoir : An Experimental Study. In *SPE EOR Conference* (pp. 1–8). Kuala Lumpur, Malaysia: Society of Petroleum Engineers.

Gao, P. K., Li, G. Q., Tian, H. M., Wang, Y. S., Sun, H. W., & Ma, T. (2015). Differences in microbial community composition between injection and production water samples of water flooding petroleum reservoirs. *Biogeosciences*, *12*, 3403–3414. 5

Gao, P., Li, G., Li, Y., Li, Y., Tian, H., Wang, Y., … Ma, T. (2016). An exogenous surfactant-producing Bacillus subtilis facilitates indigenous microbial enhanced oil recovery. *Frontiers in Microbiology*, *7*(Feb), 1–14.

García-Alcántara, J. A., Maqueda-Gálvez, A. P., Téllez-Jurado, A., Hernández-Martínez, R., & Lizardi-Jiménez, M. A. (2016). Maya Crude-Oil Degradation by a Bacillus licheniformis Consortium Isolated from a Mexican Thermal Source Using a Bubble Column Bioreactor. *Water, Air, & Soil Pollution*, *227*(11), 413.

Garrett, T. R., Bhakoo, M., & Zhang, Z. (2008). Bacterial adhesion and biofilms on surfaces. *Progress in Natural Science*, *18*(9), 1049–1056.

Gaytán, I., Mejía, M. Á., Hernández-Gama, R., Torres, L. G., Escalante, C. A., & Muñoz-Colunga, A. (2015). Effects of indigenous microbial consortia for enhanced oil recovery in a fragmented calcite rocks system. *Journal of Petroleum Science and Engineering*, *128*, 65–72.

Ghojavand, H., Vahabzadeh, F., Mehranian, M., Radmehr, M., Shahraki, K. A., Zolfagharian, F., … Roayaei, E. (2008). Isolation of thermotolerant, halotolerant, facultative biosurfactant-producing bacteria. *Applied Microbiology and Biotechnology*, *80*(6), 1073–1085.

Gianetto, A., & Islam, M. (1993). Mathematical Modelling And Scaling Up Of Microbial Enhanced Oil Recovery. *Journal of Canadian Petroleum Technology*, *32*.

Gibson, T. (1944). 306. A study of Bacillus subtilis and related organisms. *Journal of Dairy Research*, *13*(3), 248–260.

Giovannelli, D., d'Errico, G., Fiorentino, F., Fattorini, D., Regoli, F., Angeletti, L., … Manini, E. (2016). Diversity and Distribution of Prokaryotes within a Shallow-Water Pockmark Field. *Frontiers in Microbiology*, *7*, 941.

Gittel, A., Sørensen, K. B., Skovhus, T. L., Ingvorsen, K., & Schramm, A. (2009). Prokaryotic community structure and sulfate reducer activity in water from high-temperature oil

reservoirs with and without nitrate treatment. *Applied and Environmental Microbiology*, *75*(22), 7086–96.

Gkorezis, P., Van Hamme, J., Bottos, E., Thijs, S., Balseiro-Romero, M., Monterroso, C., … Vangronsveld, J. (2016). Draft Genome Sequence of Bacillus licheniformis Strain GB2, a Hydrocarbon-Degrading and Plant Growth-Promoting Soil Bacterium. *Genome Announcements*, *4*(3), e00608-16.

Gray, M. R., Yeung, A., & Foght, J. M. (2008). Potential Microbial Enhanced Oil Recovery Processes: A Critical Analysis. *SPE Annual Technical Conference and Exhibition*.

Griffiths, R. I., Whiteley, A. S., O'Donnell, A. G., & Bailey, M. J. (2000). Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Applied and Environmental Microbiology*, *66*(12), 5488–91.

Gudiña, E. J., Fernandes, E. C., Rodrigues, A. I., Teixeira, J. A., & Rodrigues, L. R. (2015). Biosurfactant production by Bacillus subtilis using corn steep liquor as culture medium. *Frontiers in Microbiology*, *6*(FEB), 59.

Gudiña, E. J., Pereira, J. F. B., Costa, R., Coutinho, J. A. P., Teixeira, J. A., & Rodrigues, L. R. (2013). Biosurfactant-producing and oil-degrading Bacillus subtilis strains enhance oil recovery in laboratory sand-pack columns. *Journal of Hazardous Materials*, *261*, 106–113.

Gudiña, E. J., Pereira, J. F. B., Rodrigues, L. R., Coutinho, J. A. P., & Teixeira, J. A. (2012). Isolation and study of microorganisms from oil samples for application in Microbial Enhanced Oil Recovery. *International Biodeterioration & Biodegradation*, *68*(0), 56–64.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075.

Hakima, A., & Singleton, I. (2017). Isolation of Indigenous Hydrocarbon Transforming Bacteria from Oil Contaminated Soils in Libya : Selection for Use as Potential Inocula for Soil Bioremediation. *International Journal of Environmental Bioremediation & Biodegradation*, *5*(1), 8–17.

Hallenbeck, P. C. (2012). *Microbial technologies in advanced biofuels production*. Springer.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, *5*(10), 245–249. https://doi.org/10.1016/S1074-5521(98)90108-9

Hanor, J. S. (1994). Origin of saline fluids in sedimentary basins. *Geological Society Special Publications*, *78*.

Heryani, H., & Putra, M. D. (2017). Kinetic study and modeling of biosurfactant production using Bacillus sp. *Electronic Journal of Biotechnology*, *27*, 49–54.

Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews. Microbiology*, *8*(1), 15–25.

Higgins, D. G., Blackshields, G., & Wallace, I. M. (2005). Mind the gaps: Progress in progressive alignment. *PNAS*, *102*(30), 10411–10412.

Hirsch, P., Eckhardt, F., & Palmer, R. (1995). Methods for the study of rock-inhabiting microorganisms—a mini review. *Journal of Microbiological Methods*.

Holloway, B. W., Krishnapillai, V., & Morgan, A. F. (1979). Chromosomal genetics of Pseudomonas. *Microbiological Reviews*, *43*(1), 73–102.

Hong, S., Bunge, J., Leslin, C., Jeon, S., & Epstein, S. S. (2009). Polymerase chain reaction primers miss half of rRNA microbial diversity. *The ISME Journal*, *3*(12), 1365–1373.

Hood, L., & Rowen, L. (2013). The Human Genome Project: big science transforms biology and medicine. *Genome Medicine*, *5*(9), 79.

Hossain, T., & Uddin, K. (2014). OPTIMIZATION OF CULTURE CONDITIONS FOR A BIOFILM FORMING BACTERIUM. *Int J Pharm Bio Sci*, *5*(1), 11–20.

Hosseini-Nasab, S. M., Zitha, P. L. J., Mirhaj, S. A., & Simjoo, M. (2016). A new chemical-enhanced oil recovery method? *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, *507*, 89–95.

Hu, X., Wang, C., & Wang, P. (2015). Optimization and characterization of biosurfactant production from marine Vibrio sp. strain 3B-2. *Frontiers in Microbiology*, *6*, 976.

Hubbert, M. K. (1956). Nuclear Energy and the Fossil Fuels Drilling and Production Practice. *Shell Development Company*, *95*.

Hui, L., Ai, M., Han, S., Guo, Y., Li, N., Shi, R., … Zhang, Y. (2012). Microbial diversity and functionally distinct groups in produced water from the Daqing Oilfield, China. *Petroleum Science*, *9*(4), 469–484.

Ijiri, A., Ohtomo, Y., Morono, Y., Ikehara, M., & Inagaki, F. (2013). Increase in acetate concentrations during sediment sample onboard storage: A caution for pore-water geochemical analyses. *Geochemical Journal*, *47*(5), 567–571.

Jackson, S. C., Alsop, A., Choban, E., Achille, B. D., Fisher, J., Canada, D., Pospisil, G. (2010). SPE 129657 Microbial EOR — Critical Aspects Learned From The Lab. *SPE IOR Symposium*, (April 24-28), SPE 129657.

Janda, J. M., & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of Clinical Microbiology*, *45*(9), 2761–4.

Jarvis, F. G., & Johnson, M. J. (1949). A Glyco-lipide Produced by Pseudomonas Aeruginosa. *Journal of the American Chemical Society*, *71*(12), 4124–4126.

Jennerman, G. E., Knapp, R. M., McInerney, M. J., & Menzie, D. E. (1983). *06/482,308*. US Patent.

Jiang, J., Gao, L., Bie, X., Lu, Z., Liu, H., Zhang, C., … Zhao, H. (2016). Identification of novel surfactin derivatives from NRPS modification of Bacillus subtilis and its antifungal activity against Fusarium moniliforme. *BMC Microbiology*, *16*, 31.

Joshi, S., Bharucha, C., & Desai, A. J. (2008). Production of biosurfactant and antifungal compound by fermented food isolate Bacillus subtilis 20B. *Bioresource Technology*, *99*(11), 4603–4608.

Joshi, S., Bharucha, C., Jha, S., Yadav, S., Nerurkar, A., & Desai, A. J. (2008). Biosurfactant production using molasses and whey under thermophilic conditions. *Bioresource Technology*, *99*(1), 195–199.

Joshi, S. D. (2003). Cost/Benefits of Horizontal Wells. *Society of Petroleum Engineers Inc.*

Joshi, S. J., Al-Wahaibi, Y. M., Al-Bahry, S. N., Elshafie, A. E., Al-Bemani, A. S., Al-Bahri, A., & Al-Mandhari, M. S. (2016). Production, Characterization, and Application of Bacillus licheniformis W16 Biosurfactant in Enhancing Oil Recovery. *Frontiers in Microbiology*, *7*, 1853.

Joshi, S. J., & Desai, A. J. (2013). Bench-Scale Production of Biosurfactants and their Potential in Ex-Situ MEOR Application. *Soil and Sediment Contamination: An International Journal*, *22*(6), 701–715.

Joshi, S. J., Geetha, S. J., & Desai, A. J. (2015). Characterization and Application of Biosurfactant Produced by Bacillus licheniformis R2. *Applied Biochemistry and Biotechnology*, *177*(2), 346–361.

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., … Wong, G. K.-S. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. *Frontiers in Microbiology*, *7*(APR), 459.

Jünemann, S., Prior, K., Albersmeier, A., Albaum, S., Kalinowski, J., Goesmann, A., … Harmsen, D. (2014). GABenchToB: a genome assembly benchmark tuned on bacteria and benchtop sequencers. *PloS One*, *9*(9), e107014.

Karimi, M., Mahmoodi, M., Niazi, A., Al-Wahaibi, Y., & Ayatollahi, S. (2012). Investigating wettability alteration during MEOR process, a micro/macro scale analysis. *Colloids and Surfaces B: Biointerfaces*, *95*(0), 129–136.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649.

Kennedy, J., Flemer, B., Jackson, S. A., Lejon, D. P. H., Morrissey, J. P., O'Gara, F., & Dobson, A. D. W. (2010). Marine metagenomics: New tools for the study and exploitation of marine microbial metabolism. *Marine Drugs*.

Kieft, T. L. (2014). Sampling the Subsurface. *Hydrocarbon and Lipid Microbiology Protocols*, 45–64.

Kip, N., & van Veen, J. A. (2015). The dual role of microbes in corrosion. *The ISME Journal*, *9*(3), 542–51.

Kisand, V., & Lettieri, T. (2013). Genome sequencing of bacteria: sequencing, de novo assembly and rapid analysis using open source tools. *BMC Genomics*, *14*, 211.

Klueglein, N., Kögler, F., Adaktylou, I. J., Wuestner, M. L., Mahler, E., Scholz, J., … Alkan, H. (2016). Understanding Selective Plugging and Biofilm Formation of a Halophilic Bacterial Community for MEOR Application. In *SPE Improved Oil Recovery Conference,*. Tulsa, Oklahoma: Society of Petroleum Engineers.

Koontz, L. (2014). TCA Precipitation. *Methods in Enzymology*, *541*, 3–10.

Kosaric, N. (1993). *Biosurfactants : production, properties, applications*. M. Dekker.

Kostka, J. E., Prakash, O., Overholt, W. A., Green, S. J., Freyer, G., Canion, A., … Huettel, M. (2011). Hydrocarbon-degrading bacteria and the bacterial community response in gulf of Mexico beach sands impacted by the deepwater horizon oil spill. *Applied and Environmental Microbiology*, *77*(22), 7962–74.

Kotlar, H. K., Lewin, A., Johansen, J., Throne-Holst, M., Haverkamp, T., Markussen, S., … Valla, S. (2011). High coverage sequencing of DNA from microorganisms living in an oil reservoir, *Environmental Microbiology Reports*, *3*(6), 674–681.

Kowalewski, E., Rueslåtten, I., Steen, K. H., Bødtker, G., & Torsæter, O. (2006a). Microbial improved oil recovery-bacterial induced wettability and interfacial tension effects on oil production. *Journal of Petroleum Science and Engineering*, *52*(1–4),

Kowalewski, E., Rueslåtten, I., Steen, K. H., Bødtker, G., & Torsæter, O. (2006b). Microbial improved oil recovery-bacterial induced wettability and interfacial tension effects on oil production. *Journal of Petroleum Science and Engineering*, *52*(1–4), 275–286.

Krüger, M., Dopffel, N., Sitte, J., Mahler, E., Mukherjee, S., Herold, A., & Alkan, H. (2016). Sampling for MEOR: Comparison of surface and subsurface sampling and its impact on field applications. *Journal of Petroleum Science and Engineering*, *146*, 1192–1201.

Kumar, G., Kumar, R., & Sharma, A. (2015). Characterization of biosurfactants from indigenous soil bacteria recovered from oil contaminated sites. *Journal of Environmental Biology*, *36*(5), 1101–1104.

Kurata, N., Vella, K., Hamilton, B., Shivji, M., Soloviev, A., Matt, S., … Perrie, W. (2016). Surfactant-associated bacteria in the near-surface layer of the ocean. *Scientific Reports*, *6*(1), 19123.

Lanzén, A., Jørgensen, S. L., Bengtsson, M. M., Jonassen, I., Øvreås, L., & Urich, T. (2011). Exploring the composition and diversity of microbial communities at the Jan Mayen hydrothermal vent field using RNA and DNA. *FEMS Microbiology Ecology*, *77*(3), 577–589.

Latil, M. (1980). *Enhanced oil recovery*. Englewood Cliffs N.J.: Prentice Hall.

Lavalleur, H. J., Colwell, F. S., M, Z., H, W., CJ, N., TG, F., … JC, F. (2013). Microbial characterization of basalt formation waters targeted for geological carbon sequestration. *FEMS Microbiology Ecology*, *85*(1), 62–73.

Lazar, I., Dobrota, S., Stefanescu, M. C., Sandulescu, L., Paduraru, R., & Stefanescu, M. (1992). Recent Field Trials in Romania. In K. Vivirito (Ed.), *Microbial enhancement of oil recovery: Recent advances. Proceedings* (pp. 265–282).

Lazar, I., Petrisor, I. G., & Yen, T. F. (2007). Microbial Enhanced Oil Recovery (MEOR). *Petroleum Science and Technology*, *25*(11), 1353–1366.

Lee, C., Kim, J. Y., Song, H. S., Kim, Y. B., Choi, Y.-E., Yoon, C., … Roh, S. W. (2017). Genomic Analysis of Bacillus licheniformis CBA7126 Isolated from a Human Fecal Sample. *Frontiers in Pharmacology*, *8*, 724.

Lewin, A., Johansen, J., Wentzel, A., Kotlar, H. K., Drablos, F., & Valla, S. (2014). The microbial communities in two apparently physically separated deep subsurface oil reservoirs show extensive DNA sequence similarities. *Environmental Microbiology*, *16*(2), 545–558.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Li, H., Yang, S.-Z., Mu, B.-Z., Rong, Z.-F., & Zhang, J. (2007). Molecular phylogenetic diversity of the microbial community associated with a high-temperature petroleum reservoir at an offshore oilfield. *FEMS Microbiology Ecology*, *60*(1), 74–84.

Li, H., Zhou, E., Ren, Y., Zhang, D., Xu, D., Yang, C., … Yang, K. (2016). Investigation of microbiologically influenced corrosion of high nitrogen nickel-free stainless steel by Pseudomonas aeruginosa. *Corrosion Science*, *111*, 811–821.

Li, X.-R., Lv, Y., Meng, H., Gu, J.-D., & Quan, Z.-X. (2014). Analysis of microbial diversity by pyrosequencing the small-subunit ribosomal RNA without PCR amplification. *Applied Microbiology and Biotechnology*, *98*(8), 3777–3789.

Li, X.-X., Mbadinga, S. M., Liu, J.-F., Zhou, L., Yang, S.-Z., Gu, J.-D., & Mu, B.-Z. (2017). Microbiota and their affiliation with physiochemical characteristics of different subsurface petroleum reservoirs. *International Biodeterioration & Biodegradation*, *120*, 170–185.

Li, Y., Yang, S., & Mu, B. (2010). The surfactin and lichenysin isoforms produced by Bacillus licheniformis HSN 221. *Analytical Letters*, *43*(6), 929–940.

Libes, S. (2009). The Origin of Petroleum in the Marine Environment. In *Introduction to Marine Biogeochemistry*. Elsevier.

Lindgreen, S., Adair, K. L., Gardner, P. P., Dionisio, F., & Rocha, E. P. C. (2016). An evaluation of the accuracy and speed of metagenome analysis tools. *Scientific Reports*, *6*(1), 19233.

Liu, J., Chen, Y., Xu, R., & Jia, Y. (2013). Screening and Evaluation of Biosurfactant-Producing Strains Isolated from Oilfield Wastewater. *Indian Journal of Microbiology*, *53*(2), 168–174.

Liu, Q., Tang, J., Bai, Z., Hecker, M., & Giesy, J. P. (2015). Distribution of petroleum degrading genes and factor analysis of petroleum contaminated soil from the Dagang Oilfield, China. *Scientific Reports*, *5*(1), 11068.

Lortet, M. L. (1892). Researches on the pathogenic microbes f the mud of the Dead Sea. *Palestine Exploration Fund*, 48–50.

Lu, J., Liyanage, P. J., Solairaj, S., Adkins, S., Arachchilage, G. P., Kim, D. H., … Pope, G. A. (2014). New surfactant developments for chemical enhanced oil recovery. *Journal of Petroleum Science and Engineering*, *120*, 94–101.

Mach, B., Reich, E., & Tatum, E. L. (1963). Separation of the Biosynthesis of the Antibiotic Polypeptide Tyrocidine From Protein Biosynthesis*. *Proceedings of the National Academy of Sciences of the United States of America*, *50*(237), 175–181.

Madslien, E. H., Rønning, H. T., Lindbäck, T., Hassel, B., Andersson, M. A., & Granum, P. E. (2013). Lichenysin is produced by most Bacillus licheniformis strains. *Journal of Applied Microbiology*, *115*(4), 1068–1080.

Magot, M. (2005). Indigenous microbial communities in oil fields. In *Petroleum Microbiology* (pp. 21–33). American Society of Microbiology.

Magot, M. (2005). Indigenous Microbial Communities in Oil Fields. In B. Ollivier & M. Magot (Eds.), *Petroleum Microbiology* (pp. 21–32). Washington: ASM Press.

Magot, M., Ollivier, B., & Patel, B. K. C. (2000). Microbiology of petroleum reservoirs. *Antonie van Leeuwenhoek*, *77*(2), 103–116.

Makkar, R. S., & Cameotra, S. S. (1997). Biosurfactant production by a thermophilic Bacillus subtilis strain. *Journal of Industrial Microbiology and Biotechnology*, *18*(1), 37–42.

Makkar, R. S., & Cameotra, S. S. (1999). Biosurfactant production by microorganisms on unconventional carbon sources. *Journal of Surfactants and Detergents*, *2*(2), 237–241.

Marchant, R., & Banat, I. M. (2012). Biosurfactants: a sustainable replacement for chemical surfactants? *Biotechnology Letters*, *34*(9), 1597–1605.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., … Bryant, S. H. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, *45*(D1), D200–D203.

Martínez-Núñez, M. A., & López, V. E. L. y. (2016). Nonribosomal peptides synthetases and their applications in industry. *Sustainable Chemical Processes*, *4*(1), 13.

Maudgalya, S., Knapp, R. M., & McInerney, M. J. (2007). Microbial Enhanced Oil Recovery Technologies: A Review of the Past, Present and Future. In *SPE Production and Operations Symposium*. Oklahoma: Society of Petroleum Engineers.

Maudgalya, S., Mclnerney, M. J., Knapp, R. M., Nagle, D. P., & Folmsbee, M. J. (2005). Tertiary oil recovery with microbial biosurfactant treatment of low-permeability berea sandstone cores. *SPE Production and Operations Symposium 2005: Anticipate the Future, Build on the Present, Celebrate the Past*, 267–273.

McDonald, J. E., Larsen, N., Pennington, A., Connolly, J., Wallis, C., Rooks, D. J., … Allison, H. E. (2016). Characterising the canine oral microbiome by direct sequencing of

reverse-transcribed rRNA molecules. *PLoS ONE*, *11*(6), 1–18.

McInerney, M J, Voordouw, G , Jennerman, G E , Sublette, K. (2007). *Manual of Environmental Microbiology* (3rd ed.). Washington, D.C.: American Society for Microbiology.

Mcinerney, M. J., Javaheri, M., & Nagle, D. P. (1990). Properties of the biosurfactant produced by Bacillus licheniformis strain JF-2. *Journal of Industrial Microbiology*, *5*, 95–102.

McInerney, M. J., Javaheri, M., & Nagle, D. P. (1990). Properties of the biosurfactant produced by Bacillus licheniformis strain JF-2. *Journal of Industrial Microbiology*, *5*(2–3), 95–101. Retrieved from

Mcinerney, M. J., Knapp, R. M., Duncan, K., Simpson, D. R., Youssef, N., Ravi, N., Mcinerney, M. J. (2004). *Development of an in situ biosurfactant production technology for enhanced oil recovery*. Oklahoma.

McInerney, M. J., Knapp, R. M., Duncan, K., Simpson, D. R., Youssef, N., Ravi, N., … Nagle. D. (2005). Development of Microorganisms with Improved Transport and Biosurfactant Activity for Enhanced Oil Recovery: Final Report to the US DoE. *DE-FC-02NT15321 R 02*, *1*(1), 1–178.

Mcinerney, M. J., Knapp, R. M., Duncan, K., Simpson, D. R., Youssef, N., Ravi, N., … Weiland, S. (2007). *Development of an in situ biosurfactant production technology for enhanced oil recovery Final Report*. *DE-FC26-04NT15522*.

Mclandsborough, L. A. (2015). Current Knowledge and Perspectives on Biofilm Formation and Remediation. In A. L. Pometto III & A. Demirci (Eds.), *Biofilms in the Food Environment: Second Edition* (2nd ed., pp. 1–28). John WIley & Sons Ltd.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature Reviews Genetics*, *11*(1), 31–46. Retrieved from

Moberly, J., D'Imperio, S., Parker, A., & Peyton, B. (2016). Microbial community signature in Lake Coeur d'Alene: Association of environmental variables and toxic heavy metal phases. *Applied Geochemistry*, *66*, 174–183.

Mulligan, C. ., Cooper, D. G., & Neufeld, R. J. (1984). Selection of Microbes Producing Biosurfactants in Media without Hydrocarbons. *Journal of Fermentation Technology*, *62*(4), 311–314. Retrieved from http://ci.nii.ac.jp/naid/110002672851/en

Mullis, K. B. (1990). The unusual origin of the polymerase chain reaction. *Scientific American*, *4*, 56–65.

Nakano, M. M., Corbell, N., Besson, J., & Zuber, P. (1992). Isolation and characterization of sfp: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in Bacillus subtilis. *Molecular & General Genetics : MGG*, *232*(2), 313–21.

Nazina, T. N., Shestakova, N. M., Grigor'ian, A. A., Mikhaĭlova, E. M., Turova, T. P., Poltaraus, A. B., … Beliaev, S. S. (2006). [Phylogenetic diversity and activity of anaerobic microorganisms of high-temperature horizons of the Dagang Oilfield (China)]. *Mikrobiologiia*, *75*(1), 70–81.

NCBI Resource Coordinators. (2017). Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *45*(D1), D12–D17.

NHGRI, N. H. G. R. I. (2015). International Consortium Completes Human Genome Project.

Nie, Y., Zhao, J.-Y., Tang, Y.-Q., Guo, P., Yang, Y., Wu, X.-L., & Zhao, F. (2016). Species Divergence vs. Functional Convergence Characterizes Crude Oil Microbial Community Assembly. *Frontiers in Microbiology*, *7*, 1254.

Nmegbu, J. G. (2014). The Effect of Salt Concentration on Microbes during Microbial Enhanced Oil Recovery. *Journal of Engineering Research and Applications Www.ijera.com*, *4*(2).

Nyyssönen, M., Hultman, J., Ahonen, L., Kukkonen, I., Paulin, L., Laine, P., … Auvinen, P. (2014). Taxonomically and functionally diverse microbial communities in deep crystalline rocks of the Fennoscandian shield. *The ISME Journal*, *8*(1), 126–138.

O'Toole, G. A., & Kolter, R. (1998). Initiation of biofilm formation in Pseudomonas fluorescens WCS365 proceeds via multiple, convergent signalling pathways: A genetic analysis. *Molecular Microbiology*, *28*(3), 449–461. https://doi.org/10.1046/j.1365-2958.1998.00797.x

Ogden, B. L. (2000). WATER TECHNOLOGY: UNDERSTANDING, INTERPRETING AND UTILIZING WATER ANALYSIS DATA, 1–12.

Ogura, M., Yamaguchi, H., Yoshida Ki, K., Fujita, Y., & Tanaka, T. (2001). DNA microarray analysis of Bacillus subtilis DegU, ComA and PhoP regulons: an approach to comprehensive analysis of B.subtilis two-component regulatory systems. *Nucleic Acids Research*, *29*(18), 3804–13. Retrieved from

Oliveira, J. S., Araujo, W., Lopes Sales, A. I., Brito Guerra, A. d., Silva Araujo, S. C. d., de Vasconcelos, A. T. R., Freitas, A. T. (2015). BioSurfDB: knowledge and algorithms to support biosurfactants and biodegradation studies.

Ostrov, I., Sela, N., Freed, M., Khateb, N., Kott-Gutkowski, M., Inbar, D., & Shemesh, M. (2015). Draft Genome Sequence of Bacillus licheniformis S127, Isolated from a Sheep Udder Clinical Infection. *Genome Announcements*, *3*(5).

Otto, T. D. (2015). From Sequence Mapping to Genome Assemblies (pp. 19–50). Humana Press, New York, NY.

Owczarzy, R., Tataurov, A. V, Wu, Y., Manthey, J. A., McQuisten, K. A., Almabrazi, H. G., … Peek, A. S. (2008). IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Research*, *36*(Web Server issue), W163-9.

Parshall, J. (2012). Mature Fields Hold Big Expansion Opportunity. *Journal of Petroleum Technology*, *1012-0052*-(October).

Parthipan, P., Preetham, E., Machuca, L. L., Rahman, P. K. S. M., Murugan, K., & Rajasekar, A. (2017). Biosurfactant and degradative enzymes mediated crude oil degradation by bacterium Bacillus subtilis A1. *Frontiers in Microbiology*, *8*(FEB), 193.

Patel, J., Borgohain, S., Kumar, M., Rangarajan, V., Somasundaran, P., & Sen, R. (2015). Recent developments in microbial enhanced oil recovery. *Renewable and Sustainable Energy Reviews*, *52*, 1539–1558.

Patel, R. K., & Jain, M. (2012). NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. *PLoS ONE*, *7*(2), e30619.

Pathak, A. K., Sharma, S., & Shrivastva, P. (2012). Multi-species biofilm of Candida albicans and non-Candida albicans Candida species on acrylic substrate. *Journal of Applied Oral Science*, *20*(1), 70–75.

Patowary, K., Patowary, R., Kalita, M. C., & Deka, S. (2017). Characterization of biosurfactant produced during degradation of hydrocarbons using crude oil as sole source of carbon. *Frontiers in Microbiology*, *8*(FEB), 279.

Pereira, J. F. B., Gudiña, E. J., Costa, R., Vitorino, R., Teixeira, J. A., Coutinho, J. A. P., & Rodrigues, L. R. (2013). Optimization and characterization of biosurfactant production by Bacillus subtilis isolates towards microbial enhanced oil recovery applications. *Fuel*, *111*, 259–268.

Pérez-Armendriz, B., Mauricio-Gutirrez, A., Jimnez Salgado, T., Tapia-Hernndez, A., & Santiesteban Lpez, A. (2013). Emulsification of Hydrocarbons Using Biosurfactant Producing Strains Isolated from Contaminated Soil in Puebla, Mexico. *Biodegradation - Engineering and Technology*, 25–45.

Polz, M. F., & Cavanaugh, C. M. (1998). Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology*, *64*(10), 3724–30.

Prakash, A., Satyanarayana, T., & Johri, B. N. (2012). Biosurfactant-Assisted Bioaugmentation in Bioremediation. In *Microorganisms in environmental management : microbes and environment* (p. 819). Springer.

Pulschen, A. A., Bendia, A. G., Fricker, A. D., Pellizari, V. H., Galante, D., & Rodrigues, F. (2017). Isolation of uncultured bacteria from antarctica using long incubation periods and low nutritional media. *Frontiers in Microbiology*, *8*(JUL), 1–12.

Purwasena, I. A., Sugai, Y., & Sasaki, K. (2009). Estimation of the potential of an oil-viscosity-reducing bacteria, petrotoga sp., isolated from an oilfield for MEOR. *Society of Petroleum Engineers - International Petroleum Technology Conference 2009, IPTC 2009*, *4*(1999), 2811–2818.

Qazi, M., Subhan, M., & Fatima, N. (2013). Role of biosurfactant produced by Fusarium sp. BS-8 in enhanced oil recovery (EOR) through sand pack column. *Int J Biosci Biochem …*, *3*(6).

Rabiei, A., Sharifinik, M., Niazi, A., Hashemi, A., & Ayatollahi, S. (2013). Core flooding tests to investigate the effects of IFT reduction and wettability alteration on oil recovery during MEOR process in an Iranian oil reservoir. *Applied Microbiology and Biotechnology*, *97*(13), 5979–5991.

Rahman, K. S. M., Rahman, T. J., McClean, S., Marchant, R., & Banat, I. M. (2002). Rhamnolipid Biosurfactant Production by Strains of Pseudomonas aeruginosa Using Low-Cost Raw Materials. *Biotechnology Progress*, *18*(6), 1277–1281.

Rajendhran, J., & Gunaskaran, P. (2011). Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. *Microbiological Research*, *166*(2), 99–110.

Rajesh Kanna, A., Gummadi, S. ., & Suresh Kumar, G. (2016). Evaluation of Bio-surfactant on Microbial EOR Using Sand Packed Column. In S. . Gummadi & P. . Vadlani (Eds.), *Biotechnology and Biochemical Engineering* (pp. 121–128). Singapore: Springer.

Ramos-Padrón, E., Bordenave, S., Lin, S., Bhaskar, I. M., Dong, X., Sensen, C. W., … Gieg, L. M. (2011). Carbon and Sulfur Cycling by Microbial Communities in a Gypsum-Treated Oil Sands Tailings Pond. *Environmental Science & Technology*, *45*(2), 439–446.

Rampelotto, P. H. (2010). Resistance of microorganisms to extreme environmental conditions and its contribution to astrobiology. *Sustainability*, *2*(6), 1602–1623.

Ravot, G., Magot, M., Ollivier, B., Patel, B. K. ., Ageron, E., Grimont, P. A. ., … Garcia, J.-L. (1997). Haloanaerobium congolense sp. nov., an anaerobic, moderately halophilic, thiosulfate- and sulfur-reducing bacterium from an African oil field. *FEMS Microbiology Letters*, *147*(1), 81–88.

Ren, H. Y., Zhang, X. J., Song, Z. yong, Rupert, W., Gao, G. J., Guo, S. xue, & Zhao, L. P. (2011). Comparison of microbial community compositions of injection and production well samples in a long-term water-flooded petroleum reservoir. *PLoS ONE*, *6*(8).

Reuter, J. A., Spacek, D. V, & Snyder, M. P. (2015). High-throughput sequencing technologies. *Molecular Cell*, *58*(4), 586–97.

Rey, M. W., Ramaiya, P., Nelson, B. A., Brody-Karpin, S. D., Zaretsky, E. J., Tang, M., … Berka, R. M. (2004). Complete genome sequence of the industrial bacterium Bacillus licheniformis and comparisons with closely related Bacillus species. *Genome Biology*, *5*(10), R77.

Rincón-Fontán, M., Rodríguez-López, L., Vecino, X., Cruz, J. M., & Moldes, A. B. (2017). Influence of micelle formation on the adsorption capacity of a biosurfactant extracted from corn on dyed hair. *RSC Adv.*, *7*(27), 16444–16452.

Rodrigues, L. R., & Teixeira, J. A. (2010). Biomedical and therapeutic applications of biosurfactants. *Advances in Experimental Medicine and Biology*, *672*, 75–87.

Roongsawang, N., Washio, K., & Morikawa, M. (2011). Diversity of nonribosomal peptide synthetases involved in the biosynthesis of lipopeptide biosurfactants. *International Journal of Molecular Sciences*, *12*(1), 141–172.

Rosnes, J. T., Torsvik, T., Lien, T., & Graue, A. (1991). *microbial enhancement of oil recovery—recent advances, Proceedings of the 1990 international conference on microbial enhancement of oil recovery*. *Developments in Petroleum Science* (Vol. 31). Elsevier.

Sachdev, D. P., & Cameotra, S. S. (2013, February 3). Biosurfactants in agriculture. *Applied Microbiology and Biotechnology*. Springer-Verlag.

Safdel, M., Anbaz, M. A., Daryasafar, A., & Jamialahmadi, M. (2017). Microbial enhanced oil recovery, a critical review on worldwide implemented field trials in different countries. *Renewable and Sustainable Energy Reviews*, *74*, 159–172.

Safdel, M., Anbaz, M. A., Daryasafar, A., & Jamialahmadi, M. (2017). Microbial enhanced oil recovery, a critical review on worldwide implemented field trials in different countries. *Renewable and Sustainable Energy Reviews*, *74*(January), 159–172

Saha, P., & Bhaskara Rao, K. V. (2016). Biosurfactants-A Current Perspective on Production and Applications. *Nature Environment and Pollution Technology*, *16*(1), 181–188.

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., … Smith, M. (1977). Nucleotide sequence of bacteriophage φX174 DNA. *Nature*, *265*(5596), 687–695.

Santos, D. K. F., Rufino, R. D., Luna, J. M., Santos, V. A., & Sarubbo, L. A. (2016). Biosurfactants: Multifunctional Biomolecules of the 21st Century. *International Journal of Molecular Sciences*, *17*(3), 401.

Sarma, P. M., Bhattacharya, D., Krishnan, S., & Lal, B. (2004). Assessment of intra-species diversity among strains of *Acinetobacter baumannii* isolated from sites contaminated with petroleum hydrocarbons. *Canadian Journal of Microbiology*, *50*(6), 405–414.

Satpute, S. K., Bhawsar, B. D., Dhakephalkar, P. K., & Chopade, B. A. (2008). Assessment of different screening methods for selecting biosurfactant producing marine bacteria. *Indian Journal of Marine Sciences*, *37*(3), 243–250.

Satpute, S. K., Bhuyan, S. S., Pardesi, K. R., Mujumdar, S. S., Dhakephalkar, P. K., Shete, A. M., & Chopade, B. A. (2010). Molecular genetics of biosurfactant synthesis in microorganisms. *Advances in Experimental Medicine and Biology*, *672*, 14–41.

Schatz, M. C., Delcher, A. L., & Salzberg, S. L. (2010). Assembly of large genomes using second-generation sequencing. *Genome Research*, *20*(9), 1165–73.

Schedler, M., Hiessl, R., Valladares Juárez, A. G., Gust, G., & Müller, R. (2014). Effect of high pressure on hydrocarbon-degrading bacteria. *AMB Express*, *4*, 77.

Schrader, C., Schielke, A., Ellerbroek, L., & Johne, R. (2012). PCR inhibitors – occurrence,

properties and removal.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069.

Sen, R. (2008a). Biotechnology in petroleum recovery: The microbial EOR. *Progress in Energy and Combustion Science*, *34*(6), 714–724.

Sen, R. (2008b). Biotechnology in petroleum recovery: The microbial EOR. *Progress in Energy and Combustion Science*, *34*(6), 714–724. 1

Sen, R. (2010). *Biosurfactants*. Springer Science. Retrieved from https://books.google.co.uk/books?id=tS3wg3E474gC&pg=PA124&lpg=PA124&dq=how+biosurfactants+alter+cell+wall&source=bl&ots=v1MNET_luL&sig=h9qV6hshhgD6Ly47ZaZZaMFh9Ig&hl=en&sa=X&ved=0ahUKEwj05aGQg_3WAhVFJ8AKHUSYANYQ6AEIKjAB#v=onepage&q=how biosurfactants alter cell wall&f=false

Serror, P., & Sonenshein, A. L. (1996). CodY Is Required for Nutritional Repression of Bacillus subtilis Genetic Competence. *JOURNAL OF BACTERIOLOGY*, *178*(20), 5910–5915.

Shah, M. U. H., Sivapragasam, M., Moniruzzaman, M., & Yusup, S. B. (2016). A comparison of Recovery Methods of Rhamnolipids Produced by Pseudomonas Aeruginosa. In *Procedia Engineering* (Vol. 148, pp. 494–500).

Sheehy, A. . (1991a). Microbial physiology and enhancement of oil recovery - recent advances. In E. C. Donaldson (Ed.), *Development in Petroleum Science* (31st ed., pp. 37–44). Amsterdam: Elsevier.

Sheehy, A. J. (1991b). Ch. R-1 Microbial Physiology and Enhanced Oil Recovery. *Developments in Petroleum Science*, *31*, 37–44.

Shelton, J. L., Akob, D. M., McIntosh, J. C., Fierer, N., Spear, J. R., Warwick, P. D., & McCray, J. E. (2016). Environmental Drivers of Differences in Microbial Community Structure in Crude Oil Reservoirs across a Methanogenic Gradient. *Frontiers in Microbiology*, *7*, 1535.

Shimizu, S. (2008). Screening of novel microbial enzymes and their industrial applications. *Journal of Biotechnology*, *136*, S278.

Shukla, P. (2017). Biosurfactant-Aided Bioprocessing : Industrial applications and Environmental Impact. In *Recent advances in applied microbiology* (pp. 50–60).

Siegert, M., Sitte, J., Galushko, A., & Krüger, M. (2014). Starting up microbial enhanced oil recovery. In *Advances in Biochemical Engineering/Biotechnology* (Vol. 142, pp. 1–94).

Simpson, D. R., Natraj, N. R., McInerney, M. J., & Duncan, K. E. (2011). Biosurfactant-producing Bacillus are present in produced brines from Oklahoma oil reservoirs with a wide range of salinities. *Applied Microbiology and Biotechnology*, *91*(4), 1083–1093.

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., & Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Research*, *19*(6), 1117–23.

Siñeriz, F. (2011). Production of Biosurfactants. *Biotechnology -Encyclopedia of Life Support Systems*, *V*(978-1-84826-258–5), 176–202.

Sipos, R., Szekely, A. J., Palatinszky, M., RÃ©vÃ©sz, S., MÃ¡rialigeti, K., & Nikolausz, M. (2007). Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. *FEMS Microbiology Ecology*, *60*(2), 341–350.

Sjoblom, T., Jones, S., Wood, L., William Parsons, D., & Lin, J. (2006). The Consensus Coding Sequences of Human Breast and Colorectal Cancers. *Science*, *314*(5797), 268–274.

Smith, D. L., Rooks, D. J., Fogg, P. C., Darby, A. C., Thomson, N. R., McCarthy, A. J., & Allison, H. E. (2012). Comparative genomics of Shiga toxin encoding bacteriophages. *BMC Genomics*, *13*(1), 311.

Smith, J. M., Castro, H., & Ogram, A. (2007). Structure and function of methanogens along a short-term restoration chronosequence in the Florida Everglades. *Applied and Environmental Microbiology*, *73*, 4135–4141.

Song, Z., Zhao, F., Sun, G., & Zhu, W. (2017). Long-Term Dynamics of Microbial Communities in a High-Permeable Petroleum Reservoir Reveals the Spatiotemporal Relationship between Community and Oil Recovery. *Energy & Fuels*, *31*(10), 10588–10597.

Souayeh, M., Al-Wahaibi, Y., Al-Bahry, S., Elshafie, A., Al-Bemani, A., Joshi, S., … Al-Mandhari, M. (2014). Optimization of a low-concentration Bacillus subtilis strain biosurfactant toward microbial enhanced oil recovery. *Energy and Fuels*, *28*(9), 5606–5611.

Soudmand-asli, A., Ayatollahi, S. S., Mohabatkar, H., Zareie, M., & Shariatpanahi, S. F. (2007). The in situ microbial enhanced oil recovery in fractured porous media. *Journal of Petroleum Science and Engineering*, *58*(1–2), 161–172.

Souza, E. C., Azevedo, P. O. de S. de, Domínguez, J. M., Converti, A., & Oliveira, R. P. de S. (2017). Influence of temperature and pH on the production of biosurfactant, bacteriocin and lactic acid by Lactococcus lactis. *Journal of Food*, *15*(4), 525–530.

Spark, I., Patey, I., Duncan, B., Hamilton, A., Devine, C., & McGovern-Traa, C. (2000). The effects of indigenous and introduced microbes on deeply buried hydrocarbon reservoirs, North Sea. *Clay Minerals*, *35*(1), 5–12.

Stewart, E. J. (2012, August 15). Growing unculturable bacteria. *Journal of Bacteriology*. American Society for Microbiology. https://doi.org/10.1128/JB.00345-12

Strappa, L. A., Lucia, J. P., Maure, M. A., & Llopiz, M. L. (2004). A Novel and Successful MEOR Pilot Project in a Strong Water-Drive Reservoir Vizcacheras Field, Argentina. *Proceedings of SPE/DOE Symposium on Improved Oil Recovery*, 1–24.

Sun, Y., Aljawad, O., Lei, J., & Liu, A. (2012). Genome-scale NCRNA homology search using a Hamming distance-based filtration strategy. *BMC Bioinformatics*, *13 Suppl 3*(Suppl 3), S12.

Suthar, H., Hingurao, K., Desai, A., & Nerurkar, A. (2008). Evaluation of bioemulsifier mediated Microbial Enhanced Oil Recovery using sand pack column. *Journal of Microbiological Methods*, *75*(2), 225–230.

Suthar, H., Hingurao, K., Desai, A., & Nerurkar, A. (2009). Selective plugging strategy based microbial enhanced oil recovery using Bacillus licheniformis TT33. *Journal of Microbiology and Biotechnology*, *19*(June), 1230–1237.

Suthar, H., & Nerurkar, A. (2016). Characterization of Biosurfactant Produced by Bacillus licheniformis TT42 Having Potential for Enhanced Oil Recovery. *Applied Biochemistry and Biotechnology*, *180*(2), 248–260.

Suthar H, Hingurao K, Desai A, N. A. (2009). Selective plugging strategy-based microbial-enhanced oil recovery using Bacillus licheniformis TT33. *J Microbiol Biotechnol.*, *19*(10), 1230–1237.

Taylor, R. G., Walker, D. C., & Mclnnes, R. R. (1993). E.coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic*

Acids Research, *21*(7), 1677–1678.

Tedesco, P., Maida, I., Palma Esposito, F., Tortorella, E., Subko, K., Ezeofor, C. C., … de Pascale, D. (2016). Antimicrobial Activity of Monoramnholipids Produced by Bacterial Strains Isolated from the Ross Sea (Antarctica). *Marine Drugs*, *14*(5).

Tram, G., Korolik, V., & Day, C. J. (2013). MBDS Solvent: An Improved Method for Assessment of Biofilms. *Advances in Microbiology*, *3*(2), 200–204.

Trapnell, C., & Salzberg, S. L. (2009). How to map billions of short reads onto genomes. *Nature Biotechnology*, *27*(5), 455–457.

Tsesmetzis, N., Maguire, M. J., Head, I. M., & Lomans, B. P. (2016). Protocols for investigating the microbial communities of oil and gas reservoirs. In T. McGenity, K. N. Timmis, & B. Nogales Fernandez (Eds.), *Hydrocarbon and Lipid Microbiology Protocols* (p. Tsesmetzis, N., Maguire, M.J., Head, I.M., Lomans,). Springer Berlin Heidelberg.

Udgire, M., Shah, N., & Jadhav, M. (2015). Enrichment , Isolation and Identification of Hydrocarbon Degrading Bacteria. *Int.J.Curr.Microbiol.App.Sci (2015)*, *4*(6), 708–713.

Ukwungwu, S., Abbas, A., Nasr, G., Allison, H., & Goodman, S. (2017). Wettability effects on Bandera Gray Sandstone using biosurfactants, *Journal of Petroleum Technology 6(2)*

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), e115.

Urbieta, M. S., González-Toril, E., Bazán, Á. A., Giaveno, M. A., & Donati, E. (2015). Comparison of the microbial communities of hot springs waters and the microbial biofilms in the acidic geothermal area of Copahue (Neuquén, Argentina). *Extremophiles*, *19*(2), 437–450.

Van De Peer, Y., Chapelle, S., & De Wachter, R. (1996). A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Research*, *24*(17).

Vater, J., Kablitz, B., Wilde, C., Franke, P., Mehta, N., & Cameotra, S. S. (2002). Matrix-assisted laser desorption ionization-time of flight mass spectrometry of lipopeptide biosurfactants in whole cells and culture filtrates of Bacillus subtilis C-1 isolated from petroleum sludge. *Applied and Environmental Microbiology*, *68*(12), 6210–6219.

Veith, B., Herzberg, C., Steckel, S., Feesche, J., Maurer, K. H., Ehrenreich, P., … Gottschalk, G. (2004). The complete genome sequence of Bacillus licheniformis DSM13, an organism with great industrial potential. *Journal of Molecular Microbiology and Biotechnology*, *7*(4), 204–211.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., … Smith, H. O. (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science*, *304*(5667), 66–74.

Vester, J. K., Glaring, M. A., & Stougaard, P. (2015, January). Improved cultivation and metagenomics as new tools for bioprospecting in cold environments. *Extremophiles*. Springer.

Von Wolzogen Kuhr, C A H., van der Vlugt, L. S. (1922). On the Occurrence of Sulfate-reduction in the Deeper Layers of the Earth. *Proceedings of the Koninklijke Nederandse Akademie van Wetenschappen*, *25*, 188–199.

Wagner, M., Roger, A. J., Flax, J. L., Brusseau, G. A., & Stahl, D. A. (1998). Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration.

Journal of Bacteriology, *180*(11), 2975–82.

Wang, F. L., Wang, Z. Y., & Wang, X. D. (2008). MEOR pilot effect and prehension on Block 50 of Chaoyanggou oil field. *PETROLEUM GEOLOGY & OILFIELD DEVELOPMENT IN DAQING.*, *27*(3), 102–105.

Wang, L. Y., Duan, R. Y., Liu, J. F., Yang, S. Z., Yang, J. D., & Mu, B. Z. (2012). Molecular analysis of the microbial community structures in water-flooding petroleum reservoirs with different temperatures. *Biogeosciences*, *9*(11), 4645–4659.

Wang, L. Y., Ke, W. J., Sun, X. B., Liu, J. F., Gu, J. D., & Mu, B. Z. (2014). Comparison of bacterial community in aqueous and oil phases of water-flooded petroleum reservoirs using pyrosequencing and clone library approaches. *Applied Microbiology and Biotechnology*, *98*(9), 4209–4221.

Wang, X., Li, H., Bezemer, T. M., & Hao, Z. (2016). Drivers of bacterial beta diversity in two temperate forests. *Ecological Research*, *31*(1), 57–64.

Ward, D. M., Weller, R., & Bateson, M. M. (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature*, *345*(6270), 63–65.

Watson, J. D., & Crick, F. H. . (1953). Molecular structure of nucleic acids: A structure for Deoxyribose Nucleic Acid. *Nature*, *4356*, 737738.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., … Medema, M. H. (2015). AntiSMASH 3.0-A comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research*, *43*(W1), W237–W243.

Wentzel, A., Ellingsen, T. E., Kotlar, H.-K., Zotchev, S. B., & Throne-Holst, M. (2007). Bacterial metabolism of long-chain n-alkanes. *Applied Microbiology and Biotechnology*, *76*(6), 1209–1221.

Wentzel, A., Lewin, A., Cervantes, F. J., Valla, S., & Kotlar, H. K. (2013). Deep Subsurface Oil Reservoirs as Poly-extreme Habitats for Microbial Life. A Current Review (pp. 439–466). Springer, Dordrecht.

Wetzel, J., Kingsford, C., & Pop, M. (2011). Assessing the benefits of using mate-pairs to resolve repeats in de novo short-read prokaryotic assemblies. *BMC Bioinformatics*, *12*(1), 95.

Wijesekara, S., Seneviratne, M., & Vithanage, M. (2017). Role of Biosurfactants on Microbial Degradation of Oil-Contaminated Soils. In *Agro-Environmental Sustainability* (pp. 165–181). Cham: Springer International Publishing.

Wilson, D. N., & Doudna Cate, J. H. (2012). The structure and function of the eukaryotic ribosome. *Cold Spring Harbor Perspectives in Biology*, *4*(5), a011536.

Wood-McKenszie. (2015). *Bentley Oil Field Asset Report*. Aberdeen.

Woźniak-Karczewska, M., Myszka, K., Sznajdrowska, A., Szulc, A., Zgoła-Grześkowiak, A., Ławniczak, Ł., … Chrzanowski, Ł. (2017). Isolation of rhamnolipids-producing cultures from faeces: Influence of interspecies communication on the yield of rhamnolipid congeners. *New Biotechnology*, *36*, 17–25.

Wu, Q., Peng, S., Yu, Y., Li, Y., & Xu, Y. (2013). Genome Sequence of Bacillus licheniformis CGMCC3963, a Stress-Resistant Strain Isolated in a Chinese Traditional Solid-State Liquor-Making Process. *Genome Announcements*, *1*(1).

Xcite Energy. (2014). Corporate Analysis and Presentation proceedings. *Journal of Petroleum Engineering*, (June).

Xu, Y., & Lu, M. (2011). Microbially enhanced oil recovery at simulated reservoir conditions by use of engineered bacteria. *Journal of Petroleum Science and Engineering*, *78*(2), 233–238.

Yakimov, M. M., Amro, M. M., Bock, M., Boseker, K., Fredrickson, H. L., Kessel, D. G., & Timmis, K. N. (1997). The potential of Bacillus licheniformis strains for in situ enhanced oil recovery. *Journal of Petroleum Science and Engineering*, *18*(1), 147–160.

Yakimov, M. M., Kröger, A., Slepak, T. N., Giuliano, L., Timmis, K. N., & Golyshin, P. N. (1998). A putative lichenysin A synthetase operon in Bacillus licheniformis: initial characterization. *Biochimica et Biophysica Acta*, *1399*(2–3), 141–53.

Youssef, N., Elshahed, M. S., & McInerney, M. J. (2009). *Chapter 6 Microbial Processes in Oil Fields. Culprits, Problems, and Opportunities*. *Advances in Applied Microbiology* (1st ed., Vol. 66). Elesvier Inc.

Youssef, N. H., Duncan, K. E., Nagle, D. P., Savage, K. N., Knapp, R. M., & McInerney, M. J. (2004a). Comparison of methods to detect biosurfactant production by diverse microorganisms. *Journal of Microbiological Methods*, *56*(3), 339–347.

Youssef, N. H., Duncan, K. E., Nagle, D. P., Savage, K. N., Knapp, R. M., & McInerney, M. J. (2004b). Comparison of methods to detect biosurfactant production by diverse microorganisms. *Journal of Microbiological Methods*, *56*(3), 339–347.

Youssef, N., Simpson, D. R., Duncan, K. E., McInerney, M. J., Folmsbee, M., Fincher, T., & Knapp, R. M. (2007). In situ biosurfactant production by Bacillus strains injected into a limestone petroleum reservoir. *Applied and Environmental Microbiology*, *73*(4), 1239–1247. https://doi.org/10.1128/AEM.02264-06

Zhang, F., She, Y.-H., Chai, L.-J., Banat, I. M., Zhang, X.-T., Shu, F.-C., … Hou, D.-J. (2012). Microbial diversity in long-term water-flooded oil reservoirs with different in situ temperatures in China. *Scientific Reports*, *2*, 1–10.

Zhang, F., She, Y. H., Li, H. M., Zhang, X. T., Shu, F. C., Wang, Z. L., … Hou, D. J. (2012). Impact of an indigenous microbial enhanced oil recovery field trial on microbial community structure in a high pour-point oil reservoir. *Applied Microbiology and Biotechnology*, *95*(3), 811–821.

Zhang, J., Xue, Q., Gao, H., Lai, H., & Wang, P. (2016). Production of lipopeptide biosurfactants by Bacillus atrophaeus 5-2a and their potential use in microbial enhanced oil recovery. *Microbial Cell Factories*, *15*(1), 168.

Zhao, F., Zhang, J., Shi, R., Han, S., Ma, F., & Zhang, Y. (2015a). Production of biosurfactant by a Pseudomonas aeruginosa isolate and its applicability to in situ microbial enhanced oil recovery under anoxic conditions. *RSC Adv.*, *5*(45), 36044–36050.

Zhao, F., Zhang, J., Shi, R., Han, S., Ma, F., & Zhang, Y. (2015b). Production of biosurfactant by a Pseudomonas aeruginosa isolate and its applicability to in situ microbial enhanced oil recovery under anoxic conditions. *RSC Adv.*, *5*(45), 36044–36050.

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., & Wishart, D. S. (2011). PHAST: A Fast Phage Search Tool. *Nucleic Acids Research*, *39*(suppl), W347–W352.

Zou, C., Wang, M., Xing, Y., Lan, G., Ge, T., Yan, X., & Gu, T. (2014). Characterization and optimization of biosurfactants produced by Acinetobacter baylyi ZJ2 isolated from crude oil-contaminated soil sample toward microbial enhanced oil recovery applications. *Biochemical Engineering Journal*, *90*, 49–58.