

# **Engineering of an Enzyme Cocktail for Biodegradation of Petroleum Hydrocarbons Based on Known Enzymatic Pathways and Metagenomic Techniques**

BY

Cindy Baburam (216019591)

(Doctor of Philosophy in Biotechnology)



**A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD) in Biotechnology to the Department of Biotechnology, Faculty of Applied and Computer Sciences, Vaal University of Technology.**

**Supervisor: Dr NA Feto (VUT, RSA)**

**Co-Supervisors: Dr T Tsekoa (CSIR, RSA)**

**July 2020**

## **DECLARATION**

I, **Cindy Baburam**, hereby attest that this thesis entitled “**Engineering of an Enzyme Cocktail for Biodegradation of Petroleum Hydrocarbons Based on Known Enzymatic Pathways and Metagenomic Techniques**” is submitted in the fulfilment of the requirements for the degree of Doctor of Philosophy in Biotechnology at Vaal University of Technology and it was composed solely by myself and the work contained herein is my own except explicitly stated otherwise. Being my original work, it has never been submitted for a degree in any other university or any professional qualification.



Cindy Baburam

**5th July 2020**

Date

*“Do not follow where the path may lead. Go instead where there is no path and leave a trail.”*

*Ralph Waldo Emerson*



## ACKNOWLEDGEMENTS

Firstly, and most importantly, I want to thank my Lord and Saviour, Jesus Christ, for providing me with His knowledge, wisdom and understanding during this study and in all aspects of my life. Thank you for keeping me in the hollow of your palm and showing me your never-ending love.

My deep gratitude goes to my supervisor, Dr Naser Feto, who expertly guided me through my study both in the laboratory and the writing of this dissertation. Thank you for sharing with me your valuable knowledge in this field and always making the time to listen to all the adventures in the laboratory. The laughs and long discussions emanating from this study will clearly be etched in my mind. I have learnt immensely from you and wish you the very best in all your future research endeavors. May you continue to grow the OMICS laboratory from strength to strength.

To my co-supervisor, Dr T Tsekoa, based at the CSIR, my sincere thanks for all your assistance.

My appreciation also goes out to my wonderful, supportive and amazing husband, Ramone, who stood by me like a pillar of strength. For days that seemed too long and writing that extended into the late hours of the night, you held my hand and knew the words to say. For this, I will always be entirely grateful and will never forget all the supportive gestures and kind words spoken. Thank you for being my rock when times got tough, my pillar when I needed to lean on and most of all, my best friend when I needed to talk.

To my wonderful and adorable children, Kiara and Riordan, thank you for being my little angels and supporting me through the years. Your little words of encouragement and smiling faces after a long day, became the biggest part of this journey and reminded me to always be thankful for the small things in life, because they are the biggest.

I also want to thank my parents, who set me on the path in life to always strive for the best. Thank you for all you have done for me and the doors of opportunity that you opened so I could be where I am today.

I extend my gratitude to Dr Alfred Mitema for all his suggestions and editing of the dissertation.

To my fellow postgraduate students in the OMICS laboratory at VUT, Grace, Morena, Judith and Jane, I would like to express my heartfelt appreciation for your friendship and the wonderful hours of academic discussions and the jovial environment you provided in the laboratory.

Thank you to the National Research Foundation (NRF) South Africa for the financial support for this study through the Thuthuka NRF grant for doctoral studies and South African Bio-Design Initiatives (SABDI) for laboratory consumables and training support.

My sincere thanks to all members of staff in the Biotechnology department for their support and the Vaal University of Technology for financial support to complete my doctoral studies.

# Engineering of an Enzyme Cocktail for Biodegradation of Petroleum Hydrocarbons Based on Known Enzymatic Pathways and Metagenomic Techniques

## General Abstract

Hydrocarbon pollution is becoming a growing environmental concern in South Africa and globally. This inadvertently supports the need to identify enzymes for their targeted degradation. The search for novel biocatalysts such as monooxygenases, alcohol dehydrogenases and aldehyde dehydrogenases, have relied on conventional culture-based techniques but this allows sourcing of the biomolecules from only 1-10 % of the microbial population leaving the majority of the biomolecules unaccounted for in 90-99 % of the microbial community. The implementation of a metagenomics approach, a culture-independent technique, ensures that more or less than 100 % of the microbial community is assessed. This increases the chance of finding novel enzymes with superior physico-chemical and catalytic traits. Hydrocarbon polluted soils present a rich environment with an adapted microbial diversity. It was thus extrapolated that it could be a potential source of novel monooxygenases, alcohol dehydrogenases (ADH) and aldehyde dehydrogenases (ALDH) involved in hydrocarbon degradation pathways. Therefore, the aim of the study was to extract metagenomic DNA from hydrocarbon contaminated soils and construct a metagenomic fosmid library and screen the library for monooxygenases, alcohol dehydrogenases (ADH) and aldehyde dehydrogenases (ALDH). Accordingly, the fosmid library was constructed from metagenome of hydrocarbon-contaminated soil. Then the library was functionally screened using hexadecane, octadecene and cyclohexane as substrates and fifteen positive clones were selected. The fosmid constructs of the positive clones were sequenced using PacBio next generation sequencing platform. The sequences were *de novo* assembled and analysed using CLC Genomic Workbench. The open reading frames (ORF) of the contigs were identified by blasting the contigs against uniprot database. Accordingly, four novel genes namely *amo-vut1*, *aol-vut3*, *dhy-sc-vut5* and *dhy-g-vut7* that showed close similarity with our target enzymes were further analysed *in silico* and codon-optimized as per *Escherichia coli* codon preference. The codon adjusted sequences were synthesised and cloned into pET30a(+) expression vector. However, it is worth noting that expression of *amo-vut1* was not successful since it was later identified to be a multi-pass member protein, which made it insoluble despite the use of detergent to the effect.

There is a need to meticulously genetically engineer *amo-vut1* to remove the signal and other membrane-bound peptides while maintaining its activity. Yet the other three constructs were successfully transformed and expressed in *E. coli* BL21 (DE3). The enzymes were purified and characterized and cocktail for hydrolysis of hexanol was successfully engineered based on AOL-VUT3, DHY-SC-VUT5 and DHY-G-VUT7. Therefore, novel enzymes were mined from metagenome of fossil-oil contaminated soil and effective hydrocarbon-degrading enzyme cocktails containing their combination were successfully engineered.

**Keywords:** Enzymes, enzyme cocktail, hydrocarbon, pollution, metagenomic, monooxygenase, alcohol dehydrogenases, aldehyde dehydrogenases, bioremediation.



# TABLE OF CONTENTS

<b>GENERAL INTRODUCTION.....</b>	<b>1</b>
<b>Chapter 1: Literature Review.....</b>	<b>4</b>
1.1    Petroleum Hydrocarbon Pollution.....	4
1.2    Some of the Major Global Oil Spills.....	5
1.3    Oil Spills in South Africa.....	6
1.4    The Chemistry of Hydrocarbons.....	7
1.5    Conventional Remediation Strategies for Hydrocarbons.....	9
1.6    Bioremediation of Petroleum Hydrocarbons.....	10
1.6.1    The Use of Microorganisms for Bioremediation Strategies.....	11
1.6.2    Factors affecting biodegradation of hydrocarbons by microorganisms.....	17
1.7    Metagenomics as a Molecular Tool to Mine Genes Encoding for Novel Enzymes ....	21
1.8    Enzymatic Mechanisms Employed by Microorganisms.....	26
1.9    Degradation of Alkanes.....	30
1.9.1    Regulation of alkane degrading pathways.....	32
1.10    Degradation of Alicyclic Hydrocarbons.....	33
1.11    Degradation of Aromatic Compounds.....	34
1.12    Hydroxylases and Hydrocarbon Degradation.....	37
1.13    Alcohol Dehydrogenases and Hydrocarbon Degradation.....	41
1.14    Aldehyde Dehydrogenases and Hydrocarbon Degradation.....	42
1.15    Rationale and/or Motivation.....	42
1.16    Problem Statement.....	43
1.17    Research Aim.....	44
1.18    Objectives.....	44
1.19    References.....	46
<b>Chapter 2: Bacillus species and their invaluable roles in petroleum hydrocarbon bioremediation.....</b>	<b>68</b>
<i>Abstract</i> .....	68

2.1	Introduction .....	71
2.2	The Chemistry of Hydrocarbons in Petroleum Crude Oil.....	73
2.3	Toxicity and Fate of Petroleum Hydrocarbons in the Environment .....	74
2.4	Bioremediation .....	75
2.5	Biostimulation and Bioaugmentation Methods for Bioremediation .....	75
2.6	The Potential of Microorganisms for Hydrocarbon Bioremediation .....	77
2.7	The Mechanisms Employed by <i>Bacillus</i> spp. for Bioremediation of Hydrocarbons.....	80
2.8	Surfactants and Biosurfactant for Bioavailability of Pollutants.....	81
2.9	Bacterial Chemotaxis, Flagellar Motility and Biofilm Formation .....	84
2.10	Uptake and Trans-Membrane Transport of Hydrocarbons .....	85
2.11	Enzymatic Approach for Bioremediation of Hydrocarbons .....	86
2.12	Enzymatic Degradation of Aliphatic Hydrocarbons .....	89
2.13	Enzymatic Degradation of Aromatic Hydrocarbons.....	91
2.14	Enzyme Bioprospecting through Metagenomics .....	93
2.15	Conclusion.....	96
2.16	References .....	97
<b>Chapter 3: Fosmid Library Construction and Functional Screening of Fosmid Library Clones.....</b>		<b>109</b>
	<i>Abstract</i> .....	109
3.1	Introduction .....	112
3.2	Materials and Methods.....	116
3.2.1	Collection of soil samples.....	116
3.2.2	Soil metagenomic DNA extraction.....	117
3.2.3	Metagenomic DNA quality and quantification.....	118
3.2.4	Purification of extracted metagenomic DNA .....	118
3.2.5	Construction of the fosmid metagenomic DNA library.....	120
3.2.6	Storage of fosmid library .....	122
3.2.7	Verification of fosmid library insert size .....	123
3.2.8	Functional screening of the fosmid library clones.....	124
3.2.9	Induction of fosmid clones and extraction of plasmid DNA .....	125
3.2.10	Assessing the quality and quantity of extracted fosmid DNA.....	126

3.3	Results and Discussion.....	127
3.3.1	Metagenomic DNA extraction from hydrocarbon contaminated soils .....	127
3.3.2	Downstream purification of extracted DNA.....	128
3.3.3	Fosmid library construction .....	130
3.3.4	Functional screening of fosmid clones .....	133
3.3.5	Sequencing of candidate fosmid clones.....	138
3.4	Conclusion.....	142
3.5	References .....	143
<b>Chapter 4: Next Generation Sequencing of Metagenomic DNA from Hydrocarbon Contaminated Soils .....</b>		
<b>Abstract .....</b>		<b>147</b>
4.1	Introduction .....	150
4.2	Materials and Methods.....	152
4.2.1	Sample collection of oil contaminated soils .....	152
4.2.2	Metagenomic DNA extraction from soil samples .....	152
4.2.3	Gene sequencing and data analysis.....	153
4.2.4	Assembly and bioinformatics analysis using the JCVI metagenomics Pipeline ..	153
4.2.5	Assembly and bioinformatics analysis using the CLC Genomics Workbench ....	154
4.2.6	Data analysis for microbiome profiling and functional predictions .....	155
4.3	Results and Discussion.....	156
4.3.1	General characterisation of soil community composition .....	156
4.3.2	Predicted functions of bacterial communities.....	166
4.4	Conclusion.....	179
4.5	References .....	180
<b>Chapter 5: Cloning, Expression and Purification of Novel Ammonia Monooxygenase Alpha Subunit (<i>amo-vutI</i>) Mined from Metagenome of Hydrocarbon Contaminated soils .....</b>		
<b>Abstract .....</b>		<b>192</b>
5.1	Introduction .....	195
5.2	Materials and Methods.....	197
5.2.1	Selection of fosmid clones .....	197

5.2.2	Induction of fosmid clones and extraction of fosmid DNA.....	197
5.2.3	Fosmid DNA construct integrity assessment.....	198
5.2.4	Sequence assembly and analysis.....	198
5.2.5	Gene synthesis and cloning.....	199
5.2.6	Transformation of bacterial host cells.....	199
5.2.7	Protein induction and expression.....	201
5.2.8	Protein purification and visualization.....	202
5.2.9	Western blot analysis.....	203
5.3	Results and Discussion.....	204
5.3.1	Fosmid DNA integrity assessment.....	204
5.3.2	Sequence analysis and monooxygenase gene synthesis.....	205
5.3.3	Cloning, expression and protein purification of AMO alpha subunit gene.....	207
5.3.4	Enzymatic activity of AMO alpha subunit.....	214
5.4	Conclusion.....	215
5.5	References.....	216
<b>Chapter 6: Cloning, Expression, Purification and Kinetic Study of Novel Alcohol Dehydrogenase (AOL-VUT3) Mined from Metagenome of Hydrocarbon Contaminated Soils.....</b>		
	<i>Abstract</i> .....	221
6.1	Introduction.....	224
6.2	Materials and Methods.....	226
6.2.1	Selection of fosmid clones.....	226
6.2.2	Induction of fosmid clones and extraction of fosmid DNA.....	226
6.2.3	Fosmid DNA integrity assessment.....	227
6.2.4	<i>De novo</i> sequence assembly and analysis.....	227
6.2.5	Gene synthesis and cloning.....	228
6.2.6	Transformation of bacterial host cells.....	229
6.2.7	Protein induction and expression.....	230
6.2.8	Protein purification and visualization.....	231
6.2.9	Western blot analysis.....	232
6.2.10	Alcohol dehydrogenase (ADH) activity assay.....	232

6.2.11	Kinetic studies (Substrate affinity of ADH) .....	233
6.3	Results and Discussion.....	234
6.3.1	The quality of isolated fosmid DNA.....	234
6.3.2	Sequence Analysis and Histidinol Dehydrogenase Gene Synthesis.....	235
6.3.3	Cloning, expression and protein purification of Histidinol dehydrogenase .....	236
6.3.4	Optimal alcohol dehydrogenase (AOL-VUT3) concentration .....	240
6.3.5	Optimal hexanol substrate concentration and enzyme kinetics .....	242
6.4	Conclusion.....	246
6.5	References .....	247
<b>Chapter 7: Cloning, Expression, Purification and Kinetic Studies of Two Novel Aldehyde Dehydrogenases (DHY-SC-VUT5 and DHY-G-VUT7) Mined from Metagenome of Hydrocarbon Contaminated Soils .....</b>		<b>250</b>
	<i>Abstract</i> .....	250
7.1	Introduction .....	253
7.2	Materials and Methods .....	255
7.2.1	Selection of fosmid clones .....	255
7.2.2	Induction of fosmid clones and extraction of fosmid DNA.....	255
7.2.3	Fosmid DNA construct integrity assessment.....	256
7.2.4	<i>De novo</i> sequence assembly and analysis.....	256
7.2.5	Gene synthesis and cloning.....	257
7.2.6	Transformation of bacterial cells .....	258
7.2.7	Protein induction and expression .....	259
7.2.8	Protein purification and visualization .....	260
7.2.9	Western blot analysis .....	261
7.2.10	Aldehyde dehydrogenase (ALDH) activity assay.....	261
7.2.11	Kinetic studies (Substrate affinity of both aldehyde dehydrogenases).....	262
7.3	Results and Discussion.....	263
7.3.1	The quality of isolated fosmid DNA.....	263
7.3.2	Sequence analysis and ALDH gene synthesis .....	264
7.3.3	Cloning, expression and protein purification.....	266
7.3.4	Optimal aldehyde dehydrogenase (DHY-SC-VUT5 and DHY-G-VUT7) concentration.....	271

7.3.5	Optimal hexanal substrate concentration and enzyme kinetics .....	273
7.4	Conclusion.....	279
7.5	References .....	280
<b>Chapter 8: Engineering of Enzyme Cocktails for Complete Transformation of the Toxic Hexanol into a Useful Product.....</b>		<b>286</b>
	<i>Abstract</i> .....	286
8.1	Introduction .....	289
8.2	Materials and Methods .....	291
8.2.1	Expressed and purified proteins for enzyme cocktails .....	291
8.2.2	Designed enzyme cocktails.....	292
8.2.3	ADH enzyme activity assay.....	293
8.3	Results and Discussion.....	294
8.3.1	Assessing the activity of the enzyme cocktails.....	294
8.4	Conclusion.....	300
8.5	References .....	301
<b>GENERAL CONCLUSION AND FUTURE PROSPECTS.....</b>		<b>304</b>
<b>APPENDIX.....</b>		<b>306</b>

# List of Figures

<b>Figure 1.1:</b> The different types of conventional remediation methods performed for hydrocarbon contaminated soil. ....	10
<b>Figure 1.2:</b> The metagenomics workflow showing sequence-based and functional based screening strategies. ....	23
<b>Figure 1.3:</b> Aerobic pathways for the degradation of alkanes by terminal and subterminal oxidation. ....	31
<b>Figure 1.4:</b> Metabolic pathway for the aerobic degradation of cyclohexane.....	34
<b>Figure 1.5:</b> The two alternative pathways of aerobic degradation of phenol: <i>o</i> -and <i>m</i> -cleavage. ....	35
<b>Figure 1.6:</b> Metabolic pathway of methanotrophs.....	38
<b>Figure 1.7:</b> Degradation of <i>n</i> -alkane by integral-membrane alkane monooxygenase ( <i>alkB</i> )-related alkane hydroxylases.....	40
<b>Figure 2.1:</b> The different classes of hydrocarbons that comprise crude oil.....	73
<b>Figure 2.2:</b> Enzyme engineering to improve enzymatic properties for bioremediation. ....	87
<b>Figure 2.3:</b> Processes involved during the microbial degradation of hydrocarbons as a sole carbon source under aerobic conditions.....	88
<b>Figure 2.4:</b> Metabolic pathway for alkane degradation showing enzymes involved at different stages of degradation for both terminal oxidation and sub-terminal oxidation. ....	90
<b>Figure 2.5:</b> Aerobic degradation pathway of aromatic compounds showing <i>o</i> - and <i>m</i> -cleavage. ....	92
<b>Figure 2.6:</b> Schematic diagram describing the general process of metagenomic strategies for mining novel genes. ....	93
<b>Figure 3.1:</b> The steps involved in the Fosmid Library Construction. ....	113
<b>Figure 3.2:</b> Maps showing the hydrocarbon contaminated soil sampling sites. ....	116
<b>Figure 3.3:</b> Work flow for the developed modified extraction method for the extraction of metagenomic DNA from hydrocarbon contaminated soils. ....	119
<b>Figure 3.4:</b> The vector map of PCC2FOS™ Fosmid vector used for construction of the fosmid metagenomic library. ....	121

<b>Figure 3.5:</b> Illustrative image of a LB agar plate with a numbered grid drawn to show the locations and identities of the candidate fosmid clones selected for functional screening analysis. ....	123
<b>Figure 3.6:</b> Agarose gel electrophoresis of extracted metagenomic DNA from soil samples...	127
<b>Figure 3.7:</b> Analysis of extracted metagenomic DNA from soil.. .....	128
<b>Figure 3.8:</b> Agarose gel electrophoresis to determine the size of the isolated metagenomic DNA from pooled samples.. .....	130
<b>Figure 3.9:</b> LB agar plate showing the growth of colony forming units (CFUs) obtained from the fosmid production library. ....	131
<b>Figure 3.10:</b> Agarose gel electrophoresis showing undigested and digested plasmid DNA obtained from fosmid clones CB1 and CB7. ....	133
<b>Figure 3.11:</b> Functional screening analysis of fosmid library clones grown on Bushnell Haas agar supplemented with hexadecane.....	135
<b>Figure 3.12:</b> Growth of positive controls (S1, S2, R1 and R2) obtained from the culture based technique and the negative control (EPI300-T1 <sup>®</sup> host cells) on LB agar, BH agar with hexadecane, BH agar with cyclohexane and BH agar with octadecene. ....	136
<b>Figure 3.13:</b> Fosmid clones (A1-A8 and B1-B8) grown on Bushnell Haas agar without a hydrocarbon substrate and with a hydrocarbon substrate. ....	137
<b>Figure 3.14:</b> Agarose gel electrophoresis of isolated fosmid DNA from 15 candidate clones obtained from functional screening analysis.....	141
<b>Figure 4.1:</b> The J. Craig Venter Institute (JCVI) metagenomics analysis pipeline used to analyse next generation sequence (NGS) data obtained from metagenomic DNA. ....	154
<b>Figure 4.2:</b> Taxonomic cladogram based on order of abundance showing the composition of soil microbial community based on the metagenomes of four hydrocarbon polluted soil habitats. .	158
<b>Figure 4.3:</b> Taxonomic distribution of the soil microbial communities. Distribution at the phylum level.....	160
<b>Figure 4.4:</b> Taxonomic distribution of the soil microbial communities. Distribution at the genus level.....	163
<b>Figure 4.5:</b> Taxonomic distribution of the soil microbial communities. Cladogram illustrating the distribution of bacterial communities extracted from hydrocarbon contaminated soils.....	165
<b>Figure 4.6:</b> Metabolic pathways using the KOG classes identified in the metagenomic sequences and the total number of reads for each class. ....	167



<b>Figure 4.7:</b> Overall predicted functional pathways of the bacterial communities extracted from hydrocarbon contaminated soils and the relative abundance of each.....	168
<b>Figure 5.1:</b> Map of pET-30a(+) expression vector used for the cloning of target gene sequences.....	199
<b>Figure 5.2:</b> Agarose gel electrophoresis of isolated fosmid DNA from 15 candidate clones obtained from functional screening analysis.....	204
<b>Figure 5.3:</b> Linear map of pET-30a(+)_amo-vut1.....	207
<b>Figure 5.4:</b> Analysis of <i>amo-vut1</i> gene insert.....	208
<b>Figure 5.5:</b> SDS-PAGE analysis of the AMO alpha subunit (AMO-VUT1) band after overexpression in BL21 (DE3) <i>E.coli</i> cells at varying conditions.....	209
<b>Figure 5.6:</b> SDS-PAGE analysis of the AMO alpha subunit (AMO-VUT1) band after overexpression in <i>E.coli</i> Rosetta™2 (DE3) cells at varying conditions.....	210
<b>Figure 5.7:</b> Western blot analysis for AMO alpha subunit (AMO-VUT1) expressed in <i>E. coli</i> cells BL21 (DE3) probed with anti-His antibody.....	211
<b>Figure 5.8:</b> Western blot analysis for AMO alpha subunit (AMO-VUT1) expressed in <i>E. coli</i> Rosetta™ 2(DE3) probed with anti-His antibody.....	212
<b>Figure 6.1:</b> Map of pET-30a(+) expression vector with all the features displayed.....	228
<b>Figure 6.2:</b> Agarose gel electrophoresis of isolated fosmid DNA (3 µl) from 15 candidate clones obtained from functional screening analysis.....	234
<b>Figure 6.3:</b> Linearised map of pET30a(+)_aol-vut3.....	237
<b>Figure 6.4:</b> Analysis for the confirmation of cloning of the <i>aol-vut3</i> gene insert.....	238
<b>Figure 6.5:</b> SDS-PAGE analysis of the AOL-VUT3 protein band after overexpression in <i>E. coli</i> BL21 (DE3) following induction at 15°C for 16 hours with 0.5 mM IPTG and purification using the Ni column.....	239
<b>Figure 6.6:</b> Western blot analysis for AOL-VUT3 expressed in <i>E. coli</i> BL21 (DE3) probed with anti-His antibody.....	240
<b>Figure 6.7</b> Schematic representation showing the enzymatic reaction of the novel alcohol dehydrogenase AOL-VUT3 in the conversion of hexanol to hexanal.....	242
<b>Figure 6.8:</b> The enzymatic activity of novel alcohol dehydrogenase AOL-VUT3 at varying concentration of the substrate, hexanol.....	244
<b>Figure 7.1:</b> Map of pET-30a(+) expression vector with all the features displayed.....	257

<b>Figure 7.2:</b> Agarose gel electrophoresis of isolated fosmid DNA (3 µl) from 15 candidate clones obtained from functional screening analysis. ....	264
<b>Figure 7.3:</b> Linearised maps for cloned novel aldehyde dehydrogenase genes in pET30a(+). 266	
<b>Figure 7.4:</b> Analysis for the confirmation of cloning <i>dhy-sc-vut5</i> gene insert. ....	267
<b>Figure 7.5:</b> Analysis for the confirmation of cloning <i>dhy-g-vut7</i> gene insert.....	268
<b>Figure 7.6:</b> (A) SDS PAGE analysis of novel aldehyde dehydrogenase (DHY-SC-VUT5 ) with the protein band corresponding to a size of 51 kDa after overexpression in BL21 (DE3) <i>E.coli</i> cells following induction at 15°C for 16 hours with 0.5 mM IPTG. (B) Western blot analysis of DHY-SC-VUT5 protein probed with anti-His antibody.....	269
<b>Figure 7.7:</b> (A) SDS PAGE analysis of novel aldehyde dehydrogenase (DHY-G-VUT7) with the protein band corresponding to a size of 67 kDa after overexpression in BL21 (DE3) <i>E.coli</i> cells following induction at 15°C for 16 hours with 0.5 mM IPTG. (B) Western blot analysis of DHY-G-VUT7 protein probed with anti-His antibody.....	270
<b>Figure 7.8</b> Schematic representation of the enzymatic reaction of the two identified novel aldehyde dehydrogenases DHY-SC-VUT5 and DHY-G-VUT7.....	272
<b>Figure 7.9:</b> Comparison of the relative activities of two novel aldehyde dehydrogenase DHY-SC-VUT5 and DHY-G-VUT7 using hexanal as a substrate over a time interval of 3 minutes at 30 °C. ....	274
<b>Figure 7.10:</b> Enzymatic activity of novel aldehyde dehydrogenase DHY-SC-VUT5 at different concentration of hexanal at 30 °C.....	275
<b>Figure 7.11:</b> Enzymatic activity of novel aldehyde dehydrogenase DHY-G-VUT7 at different concentration of hexanal at 30 °C.....	276
<b>Figure 8.1:</b> Diagrammatic representation of the enzymatic activity of two designed cocktails containing an alcohol dehydrogenase and an aldehyde dehydrogenase (AOL-VUT3/DHY-SC-VUT5 and AOL-VUT3/DHY-G-VUT7).....	295
<b>Figure 8.2:</b> Comparison of the relative activity of the enzyme cocktails (AOL-VUT3/DHY-SC-VUT5 and AOL-VUT3/DHY-G-VUT7) on the degradation of hexanol as a substrate.....	296
<b>Figure 8.3:</b> Comparison of the relative activity of the enzyme cocktails (AOL-VUT3/DHY-SC-VUT5 and AOL-VUT3/DHY-G-VUT7) on the degradation of hexanol as a substrate after 20 and 120 sec of incubation at 30 °C....	296

# List of Tables

<b>Table 1.1:</b> Major oil spills in history.....	6
<b>Table 1.2:</b> Classification and general structures of different hydrocarbon classes in crude oil....	8
<b>Table 1.3:</b> Petroleum hydrocarbon-degrading bacteria and their preferred degradation substrates. .....	13
<b>Table 1.4:</b> Some enzymes involved in microbial degradation of petroleum hydrocarbon pollutants. .....	27
<b>Table 1.5:</b> The divergent pathways for aerobic degradation of toluene.....	36
<b>Table 2.1:</b> Microorganisms involved in the degradation of specific classes of petroleum hydrocarbon compounds.....	78
<b>Table 2.2:</b> Enzymes involved in the bioremediation of hydrocarbons and their functions. ....	89
<b>Table 3.1:</b> Names allocated to the different soil samples used in the study.....	117
<b>Table 3.2:</b> The 15 fosmid clones selected following functional screening analysis on hydrocarbon substrates, hexadecane, cyclohexane and octadecene with their respective PacBio sequencing barcode names.....	139
<b>Table 4.1:</b> Number of metagenomic sequences associated with individual KEG pathways in the categories ‘Biodegradation of xenobiotics’ and ‘Carbohydrate metabolism’ and their relative percentage abundance. ....	169
<b>Table 4.2:</b> Classes of enzymes involved in alkane and aromatic hydrocarbon degradation pathways with the respective EC numbers, number of metagenomic reads and open reading frames. .....	171
<b>Table 5.1:</b> Selected fosmid clones based on growth rates on three different hydrocarbon substrates hexadecane, octadecene and cyclohexane. ....	197
<b>Table 6.1:</b> Selected fosmid clones based on growth rates on three different hydrocarbon substrates.....	226
<b>Table 7.1:</b> Selected fosmid clones based on growth rates on three different hydrocarbon substrates.....	255
<b>Table 8.1:</b> Some details of expressed and purified proteins used to design the enzyme cocktails. .....	292

## LIST OF ABBREVIATIONS

<b>Abbreviations</b>	<b>Meaning</b>
%	Percentage
°C	Degrees Celsius
μmol	Micromoles
Arg	Arginine
BLAST	Basic Local Alignment Search Tool
BOD	Biological oxygen demand
BH	Bushnell Haas
bp	Base Pair
CTAB	Cetyltrimethylammonium Bromide
DNA	Deoxyribonucleic acid
<i>E.coli</i>	<i>Escherichia coli</i>
EC	Enzyme Commission
EDTA	Ethylenediamine tetraacetic acid
HCl	Hydrochloric acid
His	Histidine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kb	Kilobase
KOG	EuKaryotic Orthologous Groups
KEGG	Kyoto Encyclopedia of Genes and Genomes
kDa	Kilodaltons
LB	Luria-Bertani
M	Molar
mDNA	Metagenomic DNA
Min	Minutes
mM	Millimolar
MOPS	(3-(N-morpholino)propanesulfonic acid)
NaCl	Sodium chloride
NCBI	National Centre for Biotechnology Information

NAD	Nicotinamide adenine dinucleotide
NGS	Next-Generation Sequencing
ORF	Open Reading Frame
OD	Optical Density
PBS	Phosphate Buffer Solution
PCR	Polymerase Chain Reaction
RPM	Revolutions Per Minute
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide gel electrophoresis
Sec	Seconds
SOC	Super Optimal broth with Catabolite repression
TBE	Trizma-Boric acid- Ethylenediamine tetraacetic acid
TE	Trizma-Hydrochloric acid-Ethylenediamine tetraacetic acid
TEMED	Tetramethylethylenediamine
TGS	Trizma-Glycine Sodium Dodecyl Sulfate
Tris	Trisaminomethane
Tris-HCl	Trizma-Hydrochloric acid

## GENERAL INTRODUCTION

Petroleum is the major source of energy for industrial and everyday human activities (Roy *et al.* 2018). The release of petroleum into both soil and water environments by accidental or human activities has become the main contributor to hydrocarbon pollution. Due to its serious hazard to human health and environmental deterioration, it has been classified as a priority pollutant (Yuniati, 2018). Numerous methods to control oil spillage has been developed, both physicochemical and biological (Adams *et al.* 2015). Physicochemical techniques are expensive due to excavation and transportation of contaminated materials hence green technologies for pollution by biological means termed bioremediation was developed (Taylor *et al.* 2019). Bioremediation is more cost effective and eco-friendly by removing, degrading or transforming contaminants into harmless or less harmful substances. This technology uses either a consortium of microorganisms (Al-Hawash *et al.* 2018, Garrido-Sanz *et al.* 2019) or their enzymes (Peixoto *et al.* 2011).

There is a demand for hydrocarbon degrading enzymes such as monooxygenases, alcohol dehydrogenases (ADH) and aldehyde dehydrogenases (ALDH) originating from microorganisms due to their hydrocarbon degrading nature (Gregson *et al.* 2019, Kim, Dong, *et al.* 2019). However, isolating novel enzymes proves to be a challenge when using the culture-based technique since most microorganisms are unculturable in the laboratory (Berini *et al.* 2017). To solve this issue, metagenomics, a culture-independent method evolved (Stefani *et al.* 2015). Metagenomes within these hydrocarbon contaminated environments are hypothesized to be teeming with novel enzymes for bioremediation applications (Araújo *et al.* 2020). Metagenomics provides direct access to these metagenomes from diversely adapted microorganisms living in extreme conditions (Cowan *et al.* 2015).

This study therefore makes use of metagenomic techniques to identify potential novel monooxygenases, ADHs and ALDHs from a soil hydrocarbon contaminated metagenome to develop enzyme cocktails to target hydrocarbon degradation. The following sections will discuss

in detail petroleum pollution in South Africa and globally, the reasoning behind sampling area selection, the use of metagenomic techniques, the biochemical characteristics of the identified enzymes and their implementation in enzymatic cocktail development.

## CHAPTER 1: Literature Review

### TABLE OF CONTENTS

<b>Chapter 1 .....</b>	<b>4</b>
Literature Review .....	4
1.1 Petroleum Hydrocarbon Pollution.....	4
1.2 Some of the Major Global Oil Spills.....	5
1.3 Oil Spills in South Africa.....	6
1.4 The Chemistry of Hydrocarbons .....	7
1.5 Conventional Remediation Strategies for Hydrocarbons.....	9
1.6 Bioremediation of Petroleum Hydrocarbons.....	10
1.6.1 The Use of Microorganisms for Bioremediation Strategies .....	11
1.6.2 Factors affecting biodegradation of hydrocarbons by microorganisms.....	17
1.6.2.1 Limited physical contact between bacteria and petroleum hydrocarbons.....	17
1.6.2.2 Environmental constraints.....	19
1.6.2.3 Metabolic Restrictions.....	20
1.6.2.4 Time Consuming Factor.....	20
1.7 Metagenomics as a Molecular Tool to Mine Genes Encoding for Novel Enzymes ....	21
1.8 Enzymatic Mechanisms Employed by Microorganisms .....	26
1.9 Degradation of Alkanes.....	30
1.9.1 Regulation of alkane degrading pathways .....	32
1.10 Degradation of Alicyclic Hydrocarbons .....	33
1.11 Degradation of Aromatic Compounds .....	34
1.12 Hydroxylases and Hydrocarbon Degradation .....	37
1.13 Alcohol Dehydrogenases and Hydrocarbon Degradation .....	41
1.14 Aldehyde Dehydrogenases and Hydrocarbon Degradation .....	42
1.15 Rationale and/or Motivation.....	42
1.16 Problem Statement .....	43
1.17 Research Aim .....	44
1.18 Objectives.....	44
1.19 References .....	46



---

# Chapter 1

## Literature Review

---

### 1.1 Petroleum Hydrocarbon Pollution

Petroleum is a natural resource that is sourced from the Earth's crust and is excessively used in everyday life as an energy source for industrial and domestic consumption (Iwuji *et al.* 2016). The major emerging environmental challenges faced as a result of petroleum demand is hydrocarbon pollution, proving to be a major problem in South Africa and globally (Kachienga *et al.* 2018). Hydrocarbon pollution is mainly due to the anthropogenic activities which include industrial and municipal runoffs, effluent release, crude oil spills, refuse from coastal oil refineries, off-shore oil production, shipping activities and accidental spillages of fuels and petroleum products (Siles and Margesin 2018). Moreover, accidental spills and leaks that occur often during exploration, production, refining, transport and storage of petroleum and petroleum products are additional burden of hydrocarbons to soil and water systems (Lim *et al.* 2016). Petroleum contamination globally have reached several millions of tons annually (Antwi-Akomeah *et al.* 2018).

Over the years, petroleum products have been allowed to accumulate in both terrestrial and aquatic environments leading to the growing concern about its harmful carcinogenic, mutagenic and neurotoxic attributes (Egobueze *et al.* 2019). This could pose severe consequences on biotic and abiotic components of the ecosystem that also cause serious effects on humans (Unimke *et al.* 2018). Prolonged exposure can cause detrimental damage to the central nervous systems in humans and animals as well as disrupt the endocrine system resulting in a probability of lung, skin, bladder and liver cancers (Gkorezis *et al.* 2016), and this was confirmed by studies conducted by Kadafa, (2012) and Truskewycz *et al.* (2019) from oil spills in the past 50 years.

## 1.2 Some of the Major Global Oil Spills

Anthropogenic hydrocarbon contamination of soil and water is a global issue (Zampolli *et al.* 2014). Specific petroleum hydrocarbons have been classified by the United States Environmental Protection Agency as priority environmental pollutants due to their harsh impact on animal life, ecosystems and human health (Safdari *et al.* 2018). The Deepwater Horizon oil spill in the United States, Gulf of Mexico in 2010 was classified as a major spill which resulted from the explosion and sinking of the Deepwater Horizon oil rig working on the Macondo exploration well for British Petroleum (BP) (Atlas and Hazen 2011). The US Government estimated the total discharge at 4.9 million barrels (210 million US gal; 780,000 m<sup>3</sup>) and was considered as the largest accidental marine oil spill in the history of the petroleum industry (Lineback, N. & Lineback 2013).

In 2011, an Exxon Mobil pipeline spilled 63,000 gallons of crude oil into the Yellowstone River near Laurel, Montana (Guardian, 2011) and this was regarded as the second largest in US waters after the Deepwater Horizon oil spill. In January 2015, a ruptured oil pipeline leaked up to 50,000 gallons of crude oil into the Yellowstone River in Montana (Lazare, 2015). Days after the spill, officials detected benzene, a cancer-causing agent, in the water supply downstream of the river (CBS-news, 2015). **Table 1.1** shows other major oil spills which have affected land and water ecosystems. By reflecting on the past, one can appreciate the great need for bioremediation methods and preventative measures to be put in place for future incidences.

Table 1.1: Major oil spills in history.\*

<b>Oil spill</b>	<b>Date</b>	<b>Site</b>	<b>Amount spilled (million gallons)</b>
Arabian gulf/Kuwait	January 19, 1991	Persian gulf, Kuwait	382-520
Gulf oil spill	April 22, 2010	Gulf of Mexico	206
Ixtoc 1 oil spill	June 3, 1979	Bay of Campeche, Mexico	140
Atlantic empress oil spill	June 19, 1979	Coast of Trinidad & Tobago	90
Kolva river oil spill	August 6, 1983	Kolva river, Russia	84
Nowruz oil field spill	February 10, 1983	Persian gulf of Iran	80
Castillo de bellven oil spill	August 6, 1983	Saldanha bay, South Africa	79
Amoco cadiz oil spill	March 16, 1978	Portsall, France	69
ABT summer oil spill	May 28, 1991	Coast of Angola	51-81
M/T haven tauker oil spill	April 11, 1991	Italy	45
Odyssey oil spill	November 10, 1988	Coast of Nova scotia, Canada	40.7
The Sea Star oil spill	December 19, 1972	Gulf of Oman	35.3
The Torrey canyon oil spill	March 18, 1967	Scilly Isles, UK	25-36

\*(Moss, 2010)

### 1.3 Oil Spills in South Africa

South Africa has been faced with numerous oil spill incidences off the South African coast as a result of prevailing sea conditions as well as the illegal disposal of oil (Marine or Coastal Oil Spill report 2014). The Coastal region of South Africa was affected by the first major oil spill in 1994, when the MV Apollo Sea sank near Cape Town (Church 2015). Due to the leakage of approximately 2,400 tons of oil from the sunken vessel, it was declared a major environmental

disaster. This resulted in the death of thousands of seabirds including the endangered African Penguins (Church 2015). In May 2013, oil pollution was reported off Bloubergstrand on the South African coastline when a Turkish bulk carrier, Seli One vessel which ran aground in 2009 started to release oil after being battered by rough seas causing one of the tanks to rupture underwater (News24 2013). This posed a major threat to the marine life and the Peninsula's most popular beaches.


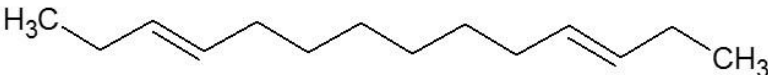
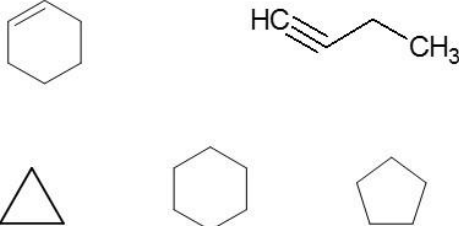
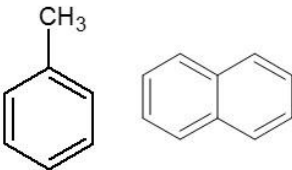
The Engen oil spill in the Richards Bay Port in April 2014 had a great impact on the port's mangrove population (SABC 2014). Six thousand litres of oil leaked from a pipeline belonging to the petroleum giant Engen and, caused major damage to the flora and fauna (SABC 2014). Oil contamination of soil is a major issue faced by environmentalists in South Africa (Makombe and Gwisai 2018). In December 2014, Transnet had an oil spill as a result of a pipeline burst in Hillcrest Durban whereby 2 000 litres of diesel were spilled (Dawood 2014). Recently, up to 400 litres of oil spilt into the Ocean off Port Elizabeth while a vessel was being refueled thus affecting the largest breeding colony of endangered African Penguins in the world (Mail and Guardian, 2019). These are but a few examples of the oil spills that have caused immense damage to the coastlines and the terrestrial areas of South Africa. Such incidences of oil pollution will have detrimental effects in the future unless effective bioremediation strategies are developed and implemented.

#### **1.4 The Chemistry of Hydrocarbons**

Crude petroleum is found naturally beneath the earth's surface as an oily, flammable liquid (Varjani and Upasani 2017a). The crude oil is a mixture of a variety of organic substances, oxygen, nitrogen, sulphur and inorganic metals (Chandra *et al.* 2013). Total petroleum hydrocarbons (TPH) is a term used by the U.S.A Environmental Protection Agency (USEPA) to describe hydrocarbon compounds derived from petroleum sources. These organic compounds are referred to as hydrocarbons because they are mainly comprised of carbon and hydrogen atoms which lack functional groups. They are arranged in different structural configurations which lends itself to their varying physical and chemical characteristics (Gkorezis *et al.* 2016, Adipah 2018a). This

complexity makes the process of degradation highly difficult. The crude oil and refinery compounds responsible for environmental pollution consists of major polluting compounds of different carbon chain lengths (Habib *et al.* 2018). This includes straight *n*-alkanes (aliphatics), branched (isoalkanes), cyclic (naphthenes) and aromatic groups which include mono or polycyclic hydrocarbons (**Table 1.2**) (Elumalai *et al.* 2017).

**Table 1.2:** Classification and general structures of different hydrocarbon classes in crude oil.

Classification of hydrocarbons	General structures
Saturated Alkanes	
Unsaturated alkanes	
Cyclic hydrocarbons	
Aromatic Hydrocarbons	

Petroleum is composed of three hydrocarbon fractions. The most abundant fraction is that of paraffin which contains both linear and branched aliphatic hydrocarbons. Alicyclic hydrocarbon such as naphthenes have one or more saturated rings in their structure. The last fraction includes that of the aromatics which consists of at least one aromatic ring. Hydrocarbons can be composed

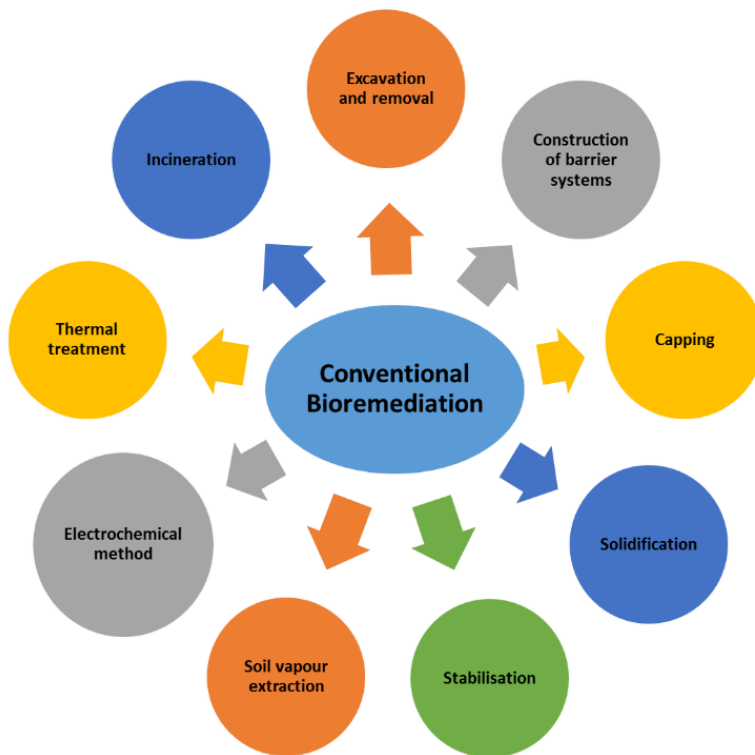
of few carbon containing atoms to greater than 60 atoms which is a determinant of their boiling points (Fuentes *et al.* 2014). These organic contaminants have been classified by the United States National Research Council (NRC) into high and low molecular weight hydrocarbons, oxygenated, aliphatics, halogenated, aromatics, nitro aromatic hydrocarbons and alkanes. These hydrocarbons and alkanes have been listed as priority pollutants which are a biohazard to the environment and natural ecosystems, since they are recalcitrant in the environment due to their non-polar and chemically inert nature (NRC 2000, Elumalai *et al.* 2017).

## 1.5 Conventional Remediation Strategies for Hydrocarbons

The remediation of soil is an action that involves the removal, control, reduction or containment of contaminants so as to ensure that the affected site does not pose any significant risk to human health or the environment (Ossai *et al.* 2020). There has been effort by the scientific community to develop efficient and effective technologies to remediate hydrocarbon polluted soils as a result of the increase in hydrocarbon pollution in developing countries (Adipah 2018a). The conventional methods for petroleum hydrocarbon remediation in soils implemented to date are divided into physical, chemical and biological categories (Venny *et al.* 2012, Khan *et al.* 2018).

The treatment of contaminated soils requires one to understand the nature of soil, characterisation of its organic matrices, sorption/desorption potential and soil microbiome composition (Karlen *et al.* 1997, Sharma *et al.* 2018). There have been a number of physico-chemical treatment options applied to hydrocarbon contaminants in soil and water. These include chemical extraction, incineration and burial in secure landfills which have been the most common conventional methods employed (Unimke *et al.* 2018). Other conventional methods implemented are included in **Figure 1.1**. However, there are multiple downsides to the use of conventional methods as opposed to the biological method. Incineration and underground disposal in secure landfills have proven to be extremely expensive especially when the amounts of pollutants are large. Thus, leading to cleanup delays which leads to the continued leakage of contaminants into soil and water sources. This necessitates the fast removal of hydrocarbon contaminants. Biological methods

which implements bioremediation and biotransformation strategies have arrived on the scene providing an effective and potentially inexpensive way forward.



**Figure 1.1:** The different types of conventional remediation methods performed for hydrocarbon contaminated soil (Unimke et al. 2018).

## 1.6 Bioremediation of Petroleum Hydrocarbons

Many countries are opting for the use of green technologies which has become a more attractive option through the years to aid in the cleanup of petroleum polluted sites using biological means. Bioremediation falls under the umbrella of green technology and can be defined as the use of living organisms such as microorganisms to degrade harmful pollutants into less toxic organic compounds such as  $\text{CO}_2$ ,  $\text{CH}_4$  and  $\text{H}_2\text{O}$  (Varjani and Upasani 2016). This technique has proven to be far more economical, efficient and environmentally sound as compared to the conventional

remediation methods. Hydrocarbon degrading microorganisms available in nature also referred to as oleophilic microbes are able to utilize hydrocarbons which are energy-rich compounds as a sole carbon source (Wu *et al.* 2017).

The method of bioremediation includes bioaugmentation and biostimulation. The bioaugmentation approach involves introducing or inoculating contaminated soil with exogenous microorganisms which have the ability to degrade hydrocarbons. The technology of biostimulation on the other hand involves supplying the contaminated soil with nutrients to facilitate the hydrocarbon degrading capacity of the indigenous microorganisms (Wu *et al.* 2016). Multiple research studies have also reported that the petroleum hydrocarbon degradation efficiency is directly related to the petroleum hydrocarbon degrader population. The rate at which this degradation activity is taking place is still limited and hence more studies in this area are required (Zhang *et al.* 2019). However, this information is of vital importance since it promotes the use of these microorganisms in bioremediation studies.

### **1.6.1 The Use of Microorganisms for Bioremediation Strategies**

The use of microbial bioremediation is a method that is widely used to date for treating petroleum hydrocarbon pollution in both terrestrial and aquatic ecosystems (Varjani and Upasani 2017b). Different microorganisms have the ability to degrade different classes of hydrocarbons ranging from aliphatics to polyaromatics (Lee *et al.* 2019). These microorganisms have the ability to naturally degrade hydrocarbons in the environment to gain energy which they utilize for metabolism and as a carbon source which is an essential component of all cellular constituents. Adaptation to living in hydrocarbon polluted environments is a major benefit for microorganisms which is a result of genetic mutations through the generations preparing them to become hydrocarbon degraders (Fuentes *et al.* 2014). They have evolved mechanisms to activate metabolic intermediates that can be shunted to critical metabolic pathways via aerobic and anaerobic pathways. Hence, this ability to oxidise these substrates allows them to thrive in nutrient limited niches (Kachienga *et al.* 2018).



Aliphatics and aromatic hydrocarbon are the two most predominate classes of hydrocarbon reported in polluted sites (Brzeszcz and Kaszycki 2018). The limited availability of microorganisms in the environment restricts the biodegradability of these oil contaminants. The n-alkanes are considered as the most easily degraded components in a petroleum mixture. It has been reported that biodegradation of n-alkanes with molecular weights up to 44 carbons are possible (Abbasian *et al.* 2015). Hydrocarbons displaying methyl branching generally are found to increase the resistance of hydrocarbons to microbial attack (Das and Chandran 2010). Various reports have concluded that cycloalkanes are particularly resistant to microbial degradation as compared to straight-chain alkanes (Atlas *et al.* 2015). Some petroleum-degrading microorganism that have been reported to degrade specific classes of oil contaminants are enlisted in **Table 1.3**.

**Table 1.3:** Petroleum hydrocarbon-degrading bacteria and their preferred degradation substrates.

<b>Petroleum hydrocarbon compound</b>	<b>Bacterial species</b>	<b>Degradation profile</b>	<b>Reference</b>
Aliphatics	<i>Rhodococcus</i> sp.	<i>n</i> -alkanes (C6-C28)	(Brzeszcz & Kaszycki, 2018)
	<i>Rhodococcus erythropolis</i> BZ4	<i>n</i> -alkanes (C5-C16)	(Brzeszcz & Kaszycki, 2018)
	<i>Rhodococcus rhodochrous</i> TRN7	<i>n</i> -alkanes (C5-C50)	(Brzeszcz & Kaszycki, 2018)
	<i>Pseudomonas</i> sp.	<i>n</i> -alkanes (C9-C17)	(Zheng <i>et al.</i> , 2018)
	<i>Enterobacter</i> sp.	<i>n</i> -alkanes (C9-C30)	(Zheng <i>et al.</i> , 2018)
	<i>Pseudomonas aeruginosa</i>	<i>n</i> -alkanes (C12-C38)	(Muriel-Millán <i>et al.</i> , 2019)
	<i>Aquabacterium</i> sp.	<i>n</i> -alkanes (C12-C35)	(Garrido-Sanz <i>et al.</i> , 2019)
	<i>Geobacillus thermodenitrifican</i>	<i>n</i> -alkanes (C12-C36)	(Elumalai <i>et al.</i> , 2019)
	<i>Alcanivorax</i> sp.	<i>n</i> -alkanes (C6-C12)	(Kadri <i>et al.</i> , 2018)
Aromatics	<i>Rhodococcus aetherivorans</i>	Polyaromatics	(Brzeszcz & Kaszycki, 2018)
	<i>Methylosinus sporium</i>	Polyaromatics	(Park <i>et al.</i> , 2018)
	<i>Neptunomonas</i> sp.	Polyaromatics	(Miller <i>et al.</i> , 2019)
	<i>Pseudomonas aeruginosa</i>	Monoaromatics	(Olowomofe <i>et al.</i> , 2019)
	<i>Alcanivorax</i> sp.	Polyaromatics	(Kadri <i>et al.</i> , 2018)
	<i>Mycobacterium</i> sp.	Polyaromatics	(Suleiman <i>et al.</i> , 2019)
	<i>Achromobacter</i> sp.	Mono/polyaromatics	(Guarino <i>et al.</i> , 2019)

Recent studies have identified 79 bacterial genera capable of degrading hydrocarbons (Tremblay *et al.* 2017). Bacterial species of genera *Achromobacter*, *Acinetobacter brinaer*, *Arthobacter*,

*Azoarcus*, *Brevibacterium*, *Cellulomonas*, *Marinobacter*, *Micrococcus*, *Nocardia*, *Ochrobactrum*, *Pseudomonas*, *Vibrio*, *Stenotrophomaonas*, *Corynebacterium*, *Mycobacterium*, *Burkholderia*, *Staphylococcus*, *Rhodococcus*, *Streptococcus* and *Flavobacterium* have been reported to have hydrocarbon degrading potential (Widdel and Rabus 2001, Mittal and Singh 2009, Chandra *et al.* 2013, Varjani and Upasani 2016, Garrido-Sanz *et al.* 2019). These microorganisms carry out environmental bioremediation using a multitude of processes which include binding, oxidation, volatilization and immobilization. The most commonly employed method is the oxidation of the toxic petroleum hydrocarbons into less harmful products (Malla *et al.* 2018).

Bacterial genera such as *Oleispira*, *Marinobacter*, *Thalassolituus*, *Alcanivorax* and *Cycloclasticus* have been identified from petroleum hydrocarbon polluted sites (Yakimov *et al.* 2007, Ren *et al.* 2018, Liu *et al.* 2019). According to Brooijmans *et al.*, 2009 these indigenous bacteria were present at low or undetectable levels before pollution but in oil polluted soils, their numbers were found to increase and dominate these environments. *Alcanivorax* strains have been found to grow on *n*-alkanes and branched alkanes but are unable to grow on any sugars or amino acids as a carbon source. *Cycloclasticus* strains are capable of growing on aromatic hydrocarbons, naphthalene, phenanthrene and anthracene (Kasai *et al.* 2002) whereas *Oleispira* strains are able to degrade aliphatic hydrocarbon, alkanols and alkanoates as a carbon source (Yakimov *et al.* 2007). This shows the diversification that exists between the different genera with regards to their preferred hydrocarbon source in a polluted environment. Thus, making their co-existence critical to removing the pollutants and introduces one to the multi-domain community that exists in order to understand the complete metabolic potential of indigenous microbial communities (Varjani and Upasani 2017c). Research also carried out by Gregson *et al.* (2018) using *Thalassolituus oleivorans* from oil polluted marine environments, showed the differential protein expression when grown on medium and long chain alkanes. Their study demonstrated the ability of these bacteria to grow on *n*-C<sub>10</sub> up to *n*-C<sub>32</sub> substrates and have enhanced the knowledge in identifying key enzymes expressed during the catabolism of *n*-alkanes.

To further understand the value of microorganisms in hydrocarbon bioremediation, research carried out by Obi *et al.* (2016) using crude oil sludge showed different genera, including

*Stenotrophomonas*, *Pseudomonas*, *Bordetella*, *Brucella*, *Bacillus*, *Achromobacter*, *Mycobacterium* and *Klebsiella* to be present. The study measured the percentage degradation rates of these isolates by measuring the absorbance of the 2,6-dichlorophenol indophenol medium. From these isolates they concluded that *Pseudomonas* was the best degrader with an estimated percentage degradation rate of 73.7% after 7 days of incubation at 28°C. Mittal & Singh (2009) have reported eleven petroleum hydrocarbon-degrading isolates from oil production sites from Lingala oil field and nine isolates from oil-contaminated soil from Haridwar. Varjani & Upasani (2016) conducted a survey study on water surface samples from five different locations in Hooghly-Matla river mouth. Their study concluded that a consortium of three bacteria i.e. *Bacillus*, *P. aeruginosa* and *Micrococcus* strains were responsible for the cleanup of medium and long chain alkanes in diesel contaminated soil.

In existing ecosystems, there are large number of indigenous microbial communities which are capable of degradation of petroleum hydrocarbon (Li *et al.* 2019). The advantage of depending on these indigenous inhabitants rather than adding foreign microorganisms to the soil to degrade the hydrocarbons stems from the fact that the natural populations have developed adequately and are adapted for survival and proliferation in that environment. Most importantly is their ability to utilize hydrocarbons is distributed amongst the diverse microbial population. These populations that exist naturally are interdependent on each other as they work in synergy to metabolise various hydrocarbons that they can individually utilize (Olajire, 2014). Bioremediation of complex mixtures of hydrocarbons therefore necessitates the work of more than a single species since an individual microorganism is capable of metabolising a limited range of substrates. Thus, this application of mixed populations with broad enzymatic activities are required to increase the rate and extent of hydrocarbon degradation (Patowary *et al.* 2016).

Dombrowski *et al.* (2016) reconstructed and confirmed the metabolic pathways of hydrocarbon-degrading bacteria from the Deepwater Horizon oil Spill in the Gulf of Mexico which occurred in 2010. Due to the oil spill, there was a change in the bacterial communities in the water column as well as in the sediments. This shift in community composition was attributed to the existing microbial communities' adaptation to the degradation of hydrocarbons for use in their metabolism

(Tavormina *et al.* 2015). Their findings confirmed that the combined capabilities of the microbial community exceeded those of its individual components i.e. in order for complex hydrocarbon mixtures to be degraded, it requires the non-redundant capabilities of a complex oil-degrading community. This was also confirmed by Wanapaisan *et al.* (2018) in which a bacterial consortium composed of five culturable bacteria have shown synergistic pyrene degradation due to the following aspects: (i) The *Bacillus* strain enhanced the bioavailability of pyrene by producing biosurfactants, (ii) *Mycobacterium* strains initiated the pyrene degradation, and (iii) *Novosphingobium* and *Ochrobactrum* degraded the intermediates of pyrene.

Research conducted by Mukherjee *et al.* (2017) showed the microbial response to hydrocarbon-contaminated environments and found that there was great taxonomic and functional variation in different geographical and spatially isolated oil polluted sites. Their study also showed the relationship that exists between soil microbiome and ecosystem functioning. The work conducted by Hauptmann *et al.* (2017) involving the investigation of microbial metagenomes from the ice sheets of Greenland showed potential microbial genes for the degradation and resistance to contaminants such as polycyclic aromatic hydrocarbons (PAHs). Similar studies were also carried out by Joshi *et al.* (2014) where metagenomes were isolated from petroleum muck. Their studies showed the presence of complex mixture of microorganism inhabiting these contaminated sites and are involved in the breakdown of complex hydrocarbons.

Hydrocarbon metabolizing microorganisms are widely distributed in nature and tools have been developed over time to better understand their microbial abundance and distribution patterns in these environments (Lamichhane *et al.* 2016). The conventional culture-based technique is the most well described, implemented tool which is based on metabolic and physiological characters. The method uses solid and/or liquid media to isolate and cultivate microorganisms in the laboratory growing in hydrocarbon environments. However, this technique does not provide comprehensive information with regards to microbial communities and their diversity (Liu *et al.* 2016). Molecular techniques such as polymerase chain reaction (PCR)-based approaches, restriction fragment length polymorphism (RFLP) and denaturing gradient gel electrophoresis (DGGE) have been used to study specific microorganisms or groups. This has further led to the

study of specific genes to evaluate overall community profiles. This approach is focused mainly on studying non-culturable microorganisms (Varjani and Upasani 2016).

## **1.6.2 Factors affecting biodegradation of hydrocarbons by microorganisms**

Four important factors need to be noted when discussing rates of biodegradation of hydrocarbons by microorganisms. Hydrocarbons found in petroleum and crude oil have low solubilities and are very hydrophobic due to their long hydrocarbon chains (Xu *et al.* 2018, Truskewycz *et al.* 2019). These characteristics limits their biodegradation in the environment (Xu *et al.* 2018). There are a variety of environmental constrains which need to be considered during degradation as these factors cannot be controlled. Metabolic restrictions is also a key player to decreasing biodegradation because these toxic compounds especially in excessive quantities affect microorganism growth (Al-Hawash *et al.* 2018). Lastly, the factor of time that is required by the microorganism to adapt and evolve their metabolism and enzymatic machinery over time (Bleuven and Landry 2016). These factors are discussed in more detail in the following sections.

### **1.6.2.1 Limited physical contact between bacteria and petroleum hydrocarbons**

First step in degradation requires the participation of bacterial membrane-bound oxygenases, direct contact needs to be made between the bacterial cells and the substrates for the introduction of molecular oxygen (Girvan & Munro, 2016; Kim, *et al.*, 2019; Maeng *et al.*, 1996). This limited bioavailability of petroleum hydrocarbons to bacteria therefore compounds the issue of biodegradation (Hua and Wang 2014). Over decades, bacteria have evolved in this respect by putting in countermeasures against petroleum contaminants by improving the adhesion ability of cells by altering their surface components and secreting bioemulsifiers. Bacteria displaying such properties also referred to as bioemulsifier-producing bacteria are highly sort after as environmental remediation agents as they can accelerate the process (Krasowska and Sigler 2014).

Attention has been given mainly to bio-emulsifier producing bacteria possessing two physiological aspects: (i) the ability to solubilize non-polar substrates allowing for bioavailability of substrates, and (ii) improve affinity between cell surfaces and oil-water interfaces (Karlapudi *et al.* 2018) . Ayed *et al.* (2015) reported that biosurfactants produced by *Bacillus amyloliquefaciens* An6 was more efficient than chemically synthesized surfactants since it showed higher solubilisation efficiency towards diesel oil (71.54% at 1 g/L). Recently, Chebbi *et al.* (2017) showed the biosurfactant production by a newly isolated *Pseudomonas* spp. from used motor oil-contaminated soil. Based on Gas chromatography–mass spectrometry (GC-MS) analyses, this biosurfactant was found to degrade 80% of phenanthrene over a period of 30 days at 37 °C.

Patowary *et al.* (2017) isolated a potent bio-emulsifier-producing bacterium, *Pseudomonas aeruginosa* PG1 from hydrocarbon contaminated soil. This bacterium was found to degrade 81.8% of total petroleum hydrocarbons (TPH) after five weeks of cultivation in a minimal media supplemented with crude oil. It should be noted that not all biosurfactants produced by bioemulsifier-producing bacteria can effectively enhance the degradation rate of pollutants. Hua & Wang (2014) observed that degradation by biosurfactants are related to their physico-chemical properties, types of pollutants and the physiological characteristics of the microorganism.

Bacterial surface properties are essential for effective degradation of hydrocarbons which enables adhesion to the substrate (Zhang *et al.* 2015). Ron & Rosenberg, (2014) showed that adherence of hydrophobic pollutants to bacterial cells is related to hydrophobic fimbriae, fibrils, outer-membrane proteins and lipids. An approach that has shown to enhance the bioavailability of petroleum hydrocarbons and it is the application of chemically synthesised surfactants which promotes dissolution leading to the emulsification of petroleum hydrocarbon pollutants (Kleindienst *et al.* 2015, Varjani and Upasani 2017a). Chen *et al.* (2007) observed that the adherence of a *Bacillus* spp. to hydrocarbons used in their research increased by 4% in the presence of rhamnolipids and the degradation of n-hexadecane increased 11.6% as compared to treatment in the absence of rhamnolipids. However, it is important to note that some chemically synthesized surfactants have been reported in the literature to exhibit adverse impacts on oil-degrading bacteria

(Kleindienst *et al.* 2015) due to their toxicity or as a result of competition of the surfactant with hydrocarbons (Liu *et al.* 2016).

### **1.6.2.2 Environmental constraints**

Numerous environmental factors that influence biodegradation reactions are temperature, pH, nutrients, electron acceptors and substrates (Varjani and Upasani 2017b). Researchers have been faced with the problem of mimicking the excellent results obtained for hydrocarbon degradation in the laboratory and in the field-scale tests (Xu *et al.* 2018). For example, the bacterial strains *Acinetobacter* spp. JLS1 and *P. aeruginosa* JLC1 which were isolated from Momoge wetlands in China, showed different sensitivities to temperature fluctuations during the biodegradation of C<sub>16</sub> alkanes. This strongly suggests that temperature plays a pivotal role in biodegradation efficiency since temperature affects bacterial growth and metabolism (Xu *et al.* 2017).

Nutrient deficiency is also another major problem limiting bacterial hydrocarbon degradation since these polluted sites have high concentrations of hydrogen and carbon but bacteria require various other elements in order to sustain their growth (Ron and Rosenberg 2014). The availability of oxygen in the environment is also a limiting step in the reaction of hydrocarbon degradation since the early steps in the catabolism of aliphatic, cyclic and aromatic hydrocarbons by microorganisms involves the oxidation of these substrates by enzymes such as oxygenases and dioxygenases (Van Beilen and Funhoff 2007). Aerobic conditions are therefore vital for microbial degradation since studies have shown that anaerobic degradation occur in negligible rates (Unimke *et al.* 2018).



### **1.6.2.3 Metabolic restrictions**

The ability of microorganisms to biodegrade hydrocarbons lies in the concentration and composition of the hydrocarbons (Al-Hawash *et al.* 2018). Extremely high levels of petroleum hydrocarbons strongly inhibit bacterial growth due to its toxicity which directly decreases the biodegradation efficiency and thus resulting in their death (Ma *et al.* 2018). Varjani (2017) reported that the order of biodegradability of hydrocarbons are listed as follows: linear alkanes > branched alkanes > low molecular weight alkyl aromatics > monoaromatics > cyclic alkanes > polyaromatics > asphaltenes. This is based on the physico-chemical properties of the substrate which affects the availability of the substrate to the bacteria. Studies carried out to date usually focus on the degradation of a single substrate but in fact, the components of petroleum hydrocarbon pollutants are very complex in nature (Wang *et al.* 2016). This presents the problem of reproducing the laboratory results in practical applications in the field. An excellent example to support this idea was shown by Abuhamed *et al.* (2004) in *Pseudomonas putida* F1 which was found to efficiently mineralize benzene, toluene and phenol individually. However, when these substrates were mixed, toluene and benzene enhanced the biodegradation of phenol, but inhibited the biodegradation of benzene and toluene.

### **1.6.2.4 Time consuming factor**

Petroleum oil hydrocarbons are not the substrates intended to be used by microorganisms living in these polluted environments (Galazka *et al.* 2018). They utilize these compounds as an alternate carbon and energy source since there is an absence of their preferred substrate. The function of hydrocarbon-degrading bacteria is dependent on the enzymes involved in degradation i.e. their expression and activity are closely related to the physiological activity of the bacteria (Mukherjee *et al.*, 2017; Song *et al.*, 2017). Sufficient time is required to synthesize these hydrocarbon-degrading enzymes. However, it has been reported that some bacteria are capable of performing this task within several days, or even less than a day under culture conditions (Al-Dhabaan 2019). The limitations of meeting this expected result in practical usage is a major challenge faced by

researchers in the field to date (Chen *et al.* 2017, Zheng *et al.* 2018). This is understood since the combination of various biological and abiotic factors are very complex thus limiting the petroleum hydrocarbon degrading bacteria (Zhao *et al.* 2017, Wang *et al.* 2018).

These factors have contributed to microbial remediation strategies taking a long time as compared to traditional remediation strategies discussed earlier. Furthermore, during emergency pollution incidents, bioremediation will not only remove the contaminants as soon as it occurs but requires time. The time is used to screen for indigenous bacteria and thereafter introduce exogenous bacteria into the area. This requires scientific assessment and approval, hence consuming more time. However, enzymatic applications to the contaminated environments could decrease this time and make bioremediation more feasible (Ivshina *et al.* 2015).

### **1.7 Metagenomics as a Molecular Tool to Mine Genes Encoding for Novel Enzymes**

Development of the culture-based technique has opened up the way for understanding microbial diversity in hydrocarbon contaminated soils (Panigrahi *et al.* 2019). However, the major drawback is that 99% of microbes that inhabit the diverse natural environments are either uncultivable or difficult to culture in the laboratory environment (Malla *et al.* 2018). The use of metagenomics, is a rapidly growing research field targeted at investigating and identifying such unculturable microorganisms (Simon and Daniel 2011). This provides a complete understanding of microbial diversity, their functions, interactions and evolution within highly stressed environments. However, this is not possible with the culture-based technique and hence forced the hand for the development of metagenomics (Malla *et al.* 2018).

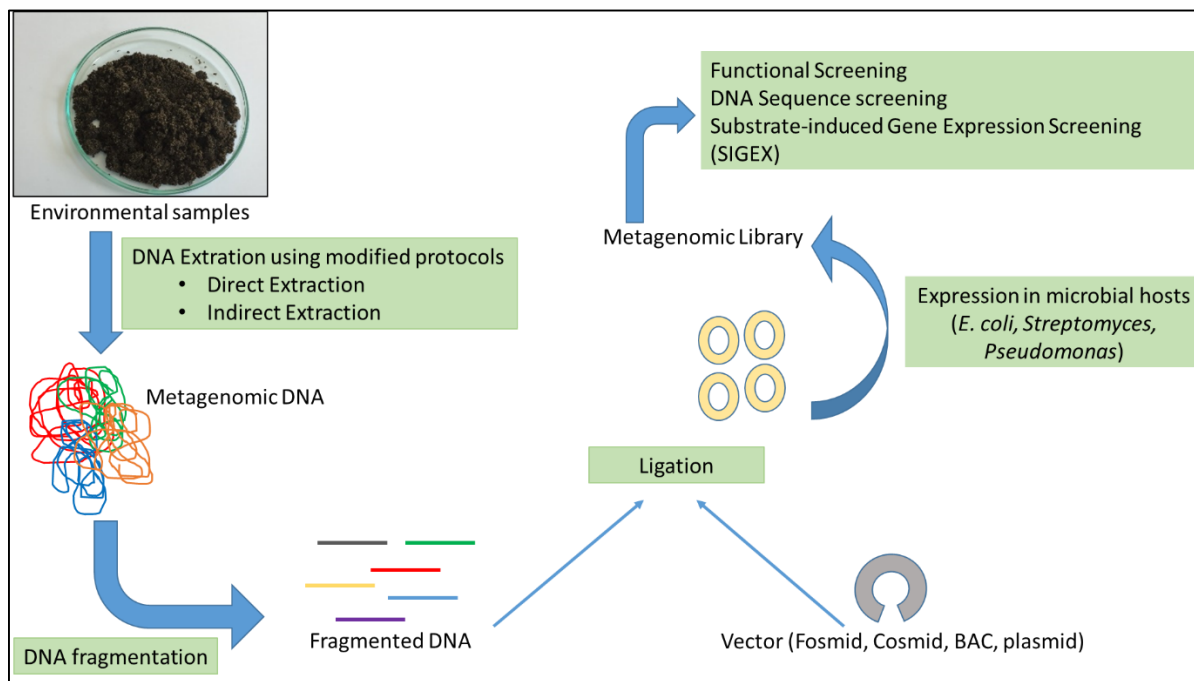
The term metagenomics was first coined by Handelsman *et al.* (1998). Metagenomics combines significant molecular technological developments of the last century thereby enabling scientists to discover and understand the biological potential held by the prokaryotic populations in varying

environments (Steele and Streit 2005). One of these applications is the use of microorganisms or their enzymes for a variety of bioremediation approaches since biocatalysis can provide alternate ways to improve petroleum bioremediation (Cao *et al.* 2009, Peixoto *et al.* 2011, Simon and Daniel 2011, Christova *et al.* 2019). This application is very attractive since novel biocatalysts can be identified timeously compared to traditional molecular methods (Wilson and Piel 2013). Metagenomic bioremediation strategies have been providing more positive results with better degradation ratios as compared to the traditional bioremediation approaches used in the past (Kosaric 2001). This is as a result of metagenomic databases that offer a rich collection of genes for the construction of novel microbial strains as well as isolating targeted gene sequences for targeted use in bioremediation studies (Chandra *et al.* 2013, Silva *et al.* 2019).

Metagenomics is based on analysing DNA from a mixed population of organisms that have adapted over time to growing in stressed environmental conditions (Rodgers and Zhang 2019). Stressed conditions, provide for the targeting of specific enzymes that can function in different conditions such as high temperatures (Silva *et al.* 2019), low temperatures (Liu *et al.* 2019), pH extremes (Merino *et al.* 2019), elevated salinity (Li *et al.* 2019) and high xenobiotic pollution (Phale *et al.* 2019). Such conditions make it possible to isolate and characterise novel target enzymes (Rhee *et al.* 2005, Jin *et al.* 2019).

The metagenomics approach has adopted a culture-independent sequencing-based analysis of DNA (Wilkins *et al.* 2019). It bases the technique on total DNA that has been isolated from environmental samples that are referred to as metagenomes. The metagenome is cloned into a host that can be easily cultivated. This represents the entire genetic complement of a single habitat referred to as a metagenomic library (Handelsman 2005). With the aid of phylotyping, information regarding diversity and function can be investigated at a deeper level. Thus, metagenomics is helping to bridge the gap left by the culture-based technique and with this deeper diversity and functional knowledge, researchers can exploit this information for bioremediation application purposes (Azubuike *et al.* 2016).

There are two different approaches to screening a metagenomic library, the function-based (expression-dependent) screening and the sequenced-based (homology dependent) screening (Lee and Lee 2013, Ngara and Zhang 2018). This is further illustrated in **Figure 1.2**. Individual as well as combined implementation of these methods have been invaluable in discovering the unculturable microorganism diversity from a range of environmental samples (DeCastro *et al.* 2016).



**Figure 1.2:** The metagenomics workflow showing sequence-based and functional based screening strategies.

Sequence-based metagenomic analysis provides microbial diversity irrespective of culturing and depends on the sequence analysis to provide information related to function prediction (Bell and Greer 2015). There is continuous addition of environmental genome sequences that are accessible from various databases. The sequence analysis becomes more informative as more genome databases are compiled from diverse sources. Sequence-based metagenomic analysis can be used for gene identification, genome assemblages, obtaining complete metabolic pathways and comparing organisms from different communities (Malla *et al.* 2018). This method however, produces vast amounts of data as a result of sequencing which places great burdens on computers

and computational scientists. The advent of the bioinformatics tool has provided a way to solve this growing problem and understand the molecular mechanisms involved in degradation pathways (Cardoso *et al.* 2015).

Research carried out by Awasthi *et al.* (2018) showed the implementation of sequence based metagenomics using *in silico* identification and construction of microbial gene clusters which are associated with the biodegradation of xenobiotic compounds. The study focused on the mining of the desired data from the National Center for Biotechnology Information (NCBI) database. The data was analysed through homology approaches using Basic Local Alignment Search Tool (BLAST) searches. The data obtained showed that bacterial populations were the major contributors to xenobiotic degradation specifically *Pseudomonas putida*, *Aspergillus niger* and *Skeletonema coastatum* have the greatest ability to degrade the maximum number of compounds. Their combined traits could be used to find evolutionary relationships between these species. The *in silico* approach was also applied by Kusnezowa & Leichert (2017) to design rational metagenomic libraries for functional studies in the global ocean sampling project (GOS). This is the largest metagenomic project to date initiated by the J. Craig Venter institute (JCVI) in 2007. This study was able to improve the success rate of functional screening approaches and make entire metagenomes available for biochemical characterisation.

However, it cannot be denied that this approach relies solely on information in databases and the available algorithms to infer the functions of newly discovered genes. This in itself leads to a few drawbacks when implementing this approach. Such drawbacks may include that the sequence similarity does not correspond to a functional relationship, novel genes have weak similarity to any genes whose products have been examined biochemically, or if a particular gene is able to carry out numerous functions in the cell (Rodgers and Zhang 2019). Therefore, function-driven screening is the preferred approach for the discovery of genes with novel functions or understanding the sequence diversity of protein families with certain functions (Armstrong *et al.* 2019).

Function-based analysis is a useful method for studying the functional characteristics of genes in order to confirm target genes. The process involves isolating metagenomic DNA from environmental samples in order to identify the functions of the encoded proteins (Lam *et al.* 2015). The metagenomic DNA is fragmented for cloning and expression in a host such as *E.coli* and screened for enzymatic activities using different assay techniques. Usually a metagenomic fosmid library is constructed from the extracted metagenomic DNA and the fosmid clones are functionally screened for potential hydrocarbon degradation activity. The candidate clones are sequenced using Next Generation Sequencing (NGS) or PACBIO sequencing to obtain sequence data that is analysed. The implementation of this method ensures that novel genes are discovered and further provides for the exploration of novel environments such as those polluted with hydrocarbons (Christova *et al.* 2019). Understanding the functional activity of genes allows researchers to construct complete enzymatic pathways which can be used in bioremediation studies with regards to enzymatic degradation of petroleum hydrocarbons in the environment (Vikram *et al.* 2018).

Duarte *et al.* (2017) supported the implementation of a culture- independent function based analysis approach which involved the extraction of metagenomic DNA to prepare a fosmid library. A total of 422 750 fosmid clones were screened for key aromatic ring-cleavage activities using 2,3-dihydroxybiphenyl as a substrate. From their study, nearly two hundred extradiol dioxygenase encoding genes of three different super-families were identified. Another study conducted by Vasconcellos *et al.* (2017) used a functional and genetic characterisation approach to isolate hydrocarbon biodegrader clones from a petroleum metagenomic library. Analysis of sequence data revealed the presence of genes related to hydrocarbon degradation and this was collaborated by the results of hydrocarbon biodegradation production detected following evaluation of the fosmid clones using functional screening.

Recently, a study was carried out by Kachienga *et al.* (2018) in two South African petroleum contaminated water aquifers using a metagenomic profiling approach. This approach assessed the microbial diversity and their adaptation to hydrocarbon degradation in such an environment. The results revealed that protozoa were the most abundant group (62.04 %), followed by fungi (24.49 %) inhabiting these petroleum contaminated aquifers. The overall results demonstrated the ability

of various microorganisms to adapt and survive in such polluted sites. Based on the results from the research studies mentioned and those in the literature, it is conclusive proof that functional based screening approaches hold great potential.

## **1.8 Enzymatic Mechanisms Employed by Microorganisms**

Microorganisms possess evolved enzymatic mechanisms to catabolise petroleum hydrocarbon pollutants to obtain energy or assimilate them into cell biomass (Wu *et al.* 2017). Ability to oxidise these substrates allows them to take advantage of nutrient-limited niches. There exists a wide phylogenetic diversity which is able to degrade pollutants aerobically but researchers through the years have shown that some bacteria such a *Pseudomonas* species and its close relatives have the ability to degrade a wide range of different contaminants (Godini *et al.* 2018). This is therefore beneficial to bioremediation studies and have motivated for their intense investigation.

In order for any pollutant to be completely degraded, a series of steps involving different enzymes are required (Abbasian *et al.* 2015). For example, some of the critical enzyme classes involved in alkane degradation are methane/ammonia monooxygenase, AlkB related alkane hydroxylases, alcohol dehydrogenases and aldehyde dehydrogenases. While for naphthalene, the enzymes include 1,2-dioxygenase, ferredoxin reductases, salicylaldehyde dehydrogenase and lastly benzene dioxygenase, toluene dioxygenase and ethylbenzene dioxygenase work on other petroleum hydrocarbons as shown in **Table 1.4** (Varjani and Upasani 2017b, Bacosa *et al.* 2018).

**Table 1.4:** Some enzymes involved in microbial degradation of petroleum hydrocarbon pollutants.

Enzyme	Hydrocarbon Compounds degraded	Microorganisms	Reference
Soluble/particulate methane/ammonia monooxygenases	C1-C8 alkanes, C1-C5 (halogenated) alkanes, alkenes and cycloalkanes.	<i>Methylococcus, Methylobacter, Pusillimonas, Methylosinus, Pseudomonas, Thalassolituus</i>	Ro <i>et al.</i> (2018), Fisher <i>et al.</i> (2018), Park <i>et al.</i> (2018), Li <i>et al.</i> (2013), Mccarl <i>et al.</i> (2018), Gregson <i>et al.</i> (2018) and Liu <i>et al.</i> (2018)
AlkB related Alkane hydroxylases	C5-C16 alkanes, alkyl benzenes, cycloalkanes, fatty acids	<i>Pseudomonas, Sphingomonas, Dietzia, Gordonia, Mycobacterium, Burkholderia</i>	Hesham <i>et al.</i> (2014), Chen <i>et al.</i> (2017), Silva <i>et al.</i> (2019), Mukherjee <i>et al.</i> (2017) and Liu <i>et al.</i> (2018)
Bacterial P450 oxygenase system (CY153)	C5-C16 alkanes, cycloalkanes, polycyclic Aromatic Hydrocarbon (PAH)-degrading	<i>Acinetobacter, Caulobacter, Mycobacterium</i>	Zafra <i>et al.</i> (2016), Fuentes <i>et al.</i> (2014), Abbasian <i>et al.</i> (2016), Ji <i>et al.</i> (2013) and Lau <i>et al.</i> (2019)
Dioxygenases Alcohol dehydrogenases	C10-C30 alkanes Alcohols (Short and long chain)	<i>Acinetobacter sp. Geobacillus, Bacillus, Rhodococcus, Pseudomonas, Gordonia</i>	Abbasian <i>et al.</i> (2015) Elumalai <i>et al.</i> (2019), Elumalai <i>et al.</i> (2017), Mishra & Singh (2012) and Kim <i>et al.</i> (2019)
Aldehyde dehydrogenases	Aldehydes (Short and long chain)	<i>Alcanivorax, Bacillus, Pseudomonas, Halorientalis, Geobacillus</i>	Gregson <i>et al.</i> (2019), Shin <i>et al.</i> (2019), Zhao <i>et al.</i> (2017) and Liu <i>et al.</i> (2009)



The most common type of bioremediation technique is the implementation of oxidation on toxic organic pollutants, converting them into harmless products (Azubuike *et al.* 2016). Microbial respiration relies on the presence of oxygen, which is the most common electron acceptor under aerobic conditions for the first step in the degradation of organic pollutants ranging from alkanes to arenes. The breakdown of hydrocarbons have always been considered an aerobic process but there have been multiple reports of microorganisms capable of carrying out anaerobic degradation (Widdel and Rabus 2001, Miller *et al.* 2019). Therefore, the metabolic pathways that hydrocarbon degrading heterotrophs can use are either aerobic (i.e. they utilize oxygen as a primary electron acceptor) or anaerobic (i.e. they utilize alternate electron acceptors such as nitrate and sulfate) (Park and Park 2018).

Aerobic and anaerobic pathways are similar in that the reactions of oxidation, reduction, hydroxylation and dehydrogenation are the same when it comes to the pathways degrading hydrocarbons. This was shown by Jaekel *et al.* (2015) and involved anaerobic degradation of cyclohexane by sulphate-reducing bacteria from hydrocarbon contaminated marine sediments. Based on GC-MS analysis, the metabolites detected were 3-cyclohexylpropionate and cyclohexane carboxylate providing evidence that the overall degradation pathway of cyclohexane under anoxic conditions is analogous to *n*-alkanes degradation under aerobic conditions.

The understanding and value of anaerobic degradation is more recent than aerobic degradation therefore there is a lack of information about the genes, the enzymes and regulations involved in anaerobic pathways (Park & Park, 2018). Nitrate, sulphate, ferrous iron and manganese ions have all been reported to be used as terminal electron acceptors under anaerobic conditions (Abbasian *et al.*, 2015; Miller *et al.*, 2019). A sulphate-reducing bacterium was the first anaerobe identified with the ability to utilise *n*-alkanes under anoxic conditions (Aeckersberg *et al.*, 1991). There are two established mechanisms for the degradation of *n*-alkane degradation under anaerobic conditions. The first is the fumarate addition pathway and the second is the carboxylation pathway (Ji *et al.*, 2013a). There have also been a few pure bacterial cultures isolated and identified that are capable of anaerobic degradation of aromatic hydrocarbons (Varjani & Upasani, 2017a). These isolates have been beneficial in elucidating anaerobic degradation pathways and further identifying

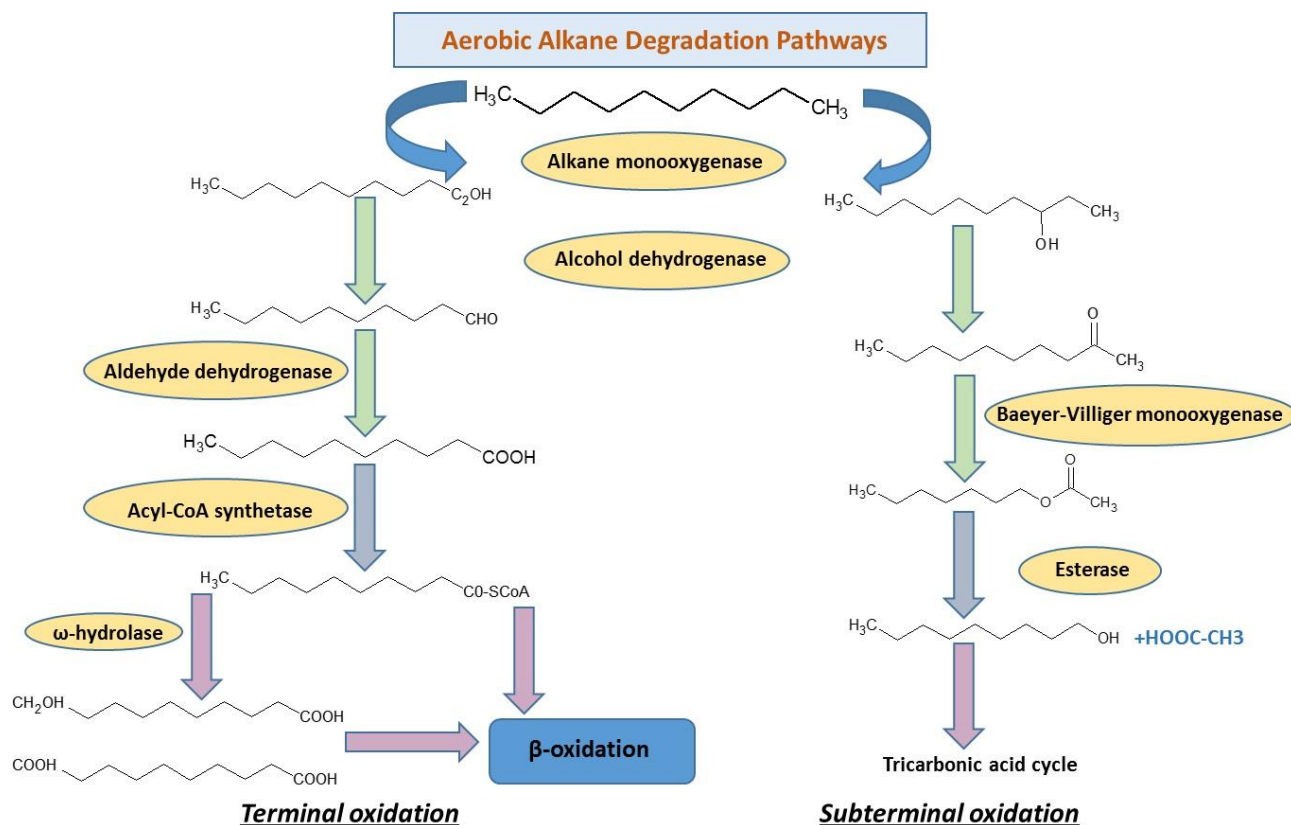
signature metabolites or genes involved with benzene and other polycyclic aromatic hydrocarbons along the way (Rabus et al., 2016).

An interesting study carried out by Atashgahi *et al.* (2018) reported the transcription of genes involved in both aerobic and anaerobic benzene degradation pathways in a benzene-degrading denitrifying continuous culture. The results from this study showed a high gene cluster encoding anaerobic benzene carboxylase with >96 % amino acid similarity to corresponding benzene degradation gene clusters previously described. Concurrently, there were transcription of genes encoding enzymes involved in oxygenase-mediated aerobic benzene degradation. This pointed to the fact that this identified microbial consortium displayed both aerobic and anaerobic benzene degradation pathways. Hence, enabling aerobic microbes to thrive in oxygen-depleted and nitrate containing environments contaminated with aromatic hydrocarbons as well.

The most efficient and rapid degradation of organic pollutants are brought about under aerobic conditions (Olajire and Essien 2014). Aerobic biodegradation of both aromatic and saturated hydrocarbons is a well-studied subject area given much attention over the years unlike that of anaerobic biodegradation. The rate of petroleum hydrocarbon degradation is performed in the following order: *n*-alkanes > branched alkanes > low molecular weight aromatics > cyclic alkanes. The most recalcitrant compounds are resins and asphaltenes (Rojo 2009). The four main pathways involved in petroleum hydrocarbon pollution is terminal oxidation, sub-terminal oxidation, bi-terminal oxidation and the Finnerty pathway (Abbasian *et al.* 2015, Park and Park 2018). The initial breakdown of organic hydrocarbons that occurs extracellularly is an oxidative process. This means that activation and the incorporation of oxygen is the key enzymatic reaction that is catalyzed by oxygenases, dioxygenases and peroxidases. Enzymes such as monooxygenases for example transfer one oxygen atom to the substrate and reduce the other oxygen atom to water (Chandra *et al.* 2013). Dioxygenases on the other hand, incorporate both atoms of molecular oxygen into products of the reaction.

## 1.9 Degradation of Alkanes

Alkanes are a major component in crude oil (constituting more than 50 %) and are readily biodegraded in the environment. Oxidation of alkanes is mainly classified as being terminal or sub-terminal (Olajire and Essien 2014). Alkane degradation starts with the oxidation of the methyl group which results in the formation of an alkanol. The alkanol is dehydrogenated by alcohol dehydrogenase to form an alkylaldehyde that is further dehydrogenated by aldehyde dehydrogenase to the corresponding carboxylic acid that is the intermediate of central intermediary metabolism namely the tricarboxylic acid (TCA) cycle as shown in **Figure 1.3** (Abbasian *et al.* 2015). This pathway was first demonstrated in *Pseudomonas putida* GPo1 and found to degrade alkanes of medium chain length. With the great advances in research over 250 *alkB* gene homologs have been reported in other hydrocarbonoclastic bacteria (Wang and Shao 2013).



**Figure 1.3:** Aerobic pathways for the degradation of alkanes by terminal and subterminal oxidation (Rojo 2009).

The most common degradation pathway for alkanes involves mono-terminal oxidation forming a primary alcohol which is further oxidised to an aldehyde and fatty acid (Das and Chandran 2010). Other oxidation pathways are di-terminal and sub-terminal oxidation. A di-terminal pathway, involves oxidation of both ends of the alkane molecule through  $\omega$  hydroxylation to obtain fatty acids which are further converted into a dicarboxylic acid and processed by  $\beta$ -oxidation (Abbasian *et al.* 2015). A sub-terminal oxidation on the other hand yields a secondary alcohol. The first step in this pathway is the oxidation of alkanes by monooxygenases to secondary alcohols, then to ketones and fatty acids. The ester produced is then hydrolysed to produce an alcohol and a fatty acid (McDonald *et al.* 2006).

Recently, research carried out by Gregson *et al.* (2018) on *Thalassolituus oleivorans* MIL-1 first isolated from seawater and sediment in Milazzo Harbor in Italy showed the expression of an alkane 1-monooxygenase (TOL\_1175) involved in terminal oxidation during growth on medium-chain alkanes ( $n$ -C<sub>14</sub>). This protein had a 99% sequence identity to an alkane 1-monooxygenase isolated from *T. oleivorans* and 65% identity to an alkane 1-monooxygenase isolated from *Oleispira antarctica* encoded for by the *alkB2* gene. Findings also revealed that an alkane monooxygenase (TOL\_2658) was also upregulated during growth on long chain alkanes ( $n$ -C<sub>28</sub>) and catalysing sub-terminal oxidation. An alcohol dehydrogenase was expressed 2.5-fold greater during growth of  $n$ -C<sub>28</sub> compared to  $n$ -C<sub>14</sub> together with a flavin-binding family monooxygenase identified as a Baeyer-Villiger monooxygenase (BVMO) class enzyme. The BVMOs catalyse the conversion of a ketone to an ester via the insertion of an oxygen in a ketone next to the carbonyl atom. To further support the use of sub-terminal oxidation during growth on long-chain alkanes, an esterase was found 17-fold upregulated during growth on the  $n$ -C<sub>28</sub> as compared to  $n$ -C<sub>14</sub>. Esterases are required to hydrolyse the subsequent ester of BVMO sub-terminal oxidation to produce alcohol and a fatty acid (Gregson *et al.*, 2018).

### 1.9.1 Regulation of alkane degrading pathways

The expression of the bacterial genes involved in alkane degradation pathways are tightly regulated. Genes involved in this pathway are down-regulated by complex global regulatory controls that ensure the genes are expressed only under the correct physiological conditions or in the absence of the preferred substrate (Rojo, 2009). For up-regulation of genes alkane-responsive regulators ensure that alkane degradation genes are induced only in the presence of the substrate (Wang & Shao, 2013).

For each specific microorganism there are several sets of alkane degradation systems specific to that microorganism for a respective alkane substrate and/or a physiological condition which becomes activated when the conditions are suitable (Rojo 2009). The mechanism of alkane hydroxylase regulation has been well studied in the OCT plasmid of *Pseudomonas putida* GPo1 (Dinamarca *et al.* 2003). These bacteria contain the gene clusters, alkBFGHJKL and alkST. When the inducer (n-alkanes) are present, the LuxR family activator AlkS induces the expression of both these gene clusters. However, in the absence of the inducer, AlkS will bind to its own promoter in order to repress its expression. This regulates the expression of alkane hydroxylases (van Beilen *et al.* 2001).

An investigation carried out on 137 environmental metagenomes by Nie *et al.* (2014) revealed that *Actinobacteria*-related *alkB* genes are significantly abundant in hydrocarbon polluted environments, suggesting the importance of Gram-positive bacteria in n-alkane degradation. A putative TetR family regulator (TFR) gene located downstream of *alkB* was found in the genomes of many of these Gram-positive bacteria. This finding suggests there might be a link between TFRs and *alkB* genes in Gram-positive bacteria involved in alkane degradation. TFRs compose the third most common transcriptional regulator family in bacteria that regulate a multitude of cellular activities (Cuthbertson and Nodwell 2013, Perrone *et al.* 2017).

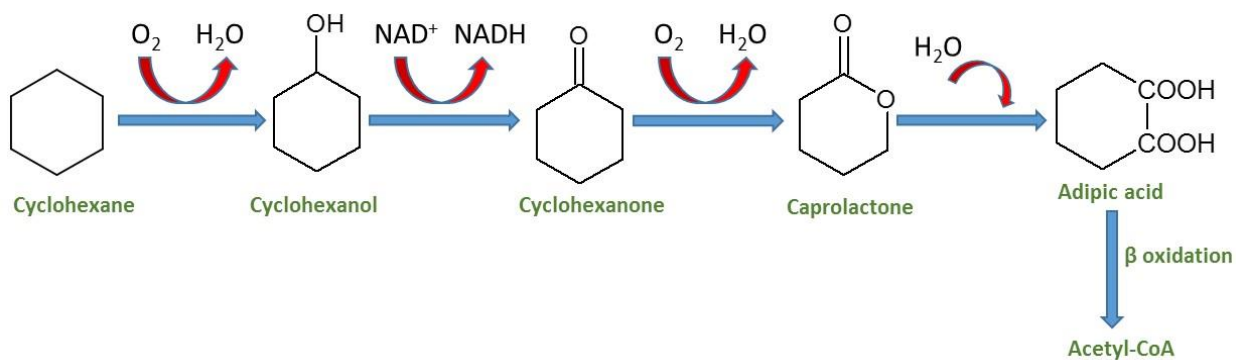
Studies by Liang *et al.* (2015) showed for the first time, the regulation mode of *n*-alkane metabolism in the Gram-positive bacterium *Dietzia* spp. DQ12-45-1b. The results concluded that TFR repressed the transcription of the *alkB* gene by cooperatively binding to its promoter. Their study went further to suggest that AlkX could sense the concentrations of *n*-alkane degradation metabolites and a model proposed suggested that AlkX controls *alkW1* expression in a metabolite-dependent manner. Hence, when *n*-alkanes are completely degraded to fatty acids, the repression is halted via an auto-regulatory feedback mechanism. However, more research is still required to confirm whether TFRs can regulate *alkB* expression (Liang 2017).

### 1.10 Degradation of Alicyclic Hydrocarbons

The minor components (20%-40%) found in mineral oil are cycloalkanes which have been found to be relatively recalcitrant to microbial degradation mechanisms as compared to *n*-alkanes (Fritsche and Hofrichter 2008). This can be as a result of its higher toxicity as compared to *n*-alkanes. Cyclopentane, cyclohexane and their alkylated derivatives are the most abundant (Jaekel *et al.* 2015). *n*-alkanes are degraded based on the initial attack on the terminal methyl group but the absence of an exposed terminal methyl group in cycloalkanes complicates this process (Olajire and Essien 2014).

Cyclohexane is used as a model compound in many studies due its stable chemical structure. Only a relatively low number of aerobic bacterial strains are able to use cycloalkanes as a sole carbon source and very few have been isolated to date. These include *Actinobacteria* and *Proteobacteria* (Rouvière and Chen 2003). Work carried out by Lee & Cho (2008) have reported for the first time the ability of *Rhodococcus* spp. EC1 to degrade cyclohexanes and the rates are within the range previously reported for other cyclohexane-degrading bacteria such as *Xanthobacter* spp., *Pseudomonas* spp., and *B. petroleovorans*. Degradation of cyclohexane was also demonstrated by *Bacillus lentus* strain LP32 (Opere *et al.* 2013). The pathway for the degradation of cycloalkanes under aerobic conditions is the activation by a cyclohexane monooxygenase forming cyclohexanol, this is further oxidized to cyclohexanone, caprolactone and adipate as shown in

**Figure 1.4.** A monooxygenase introduces an oxygen into the cyclic ketone that further cleaves the cyclic ring (Cheng *et al.* 2002).

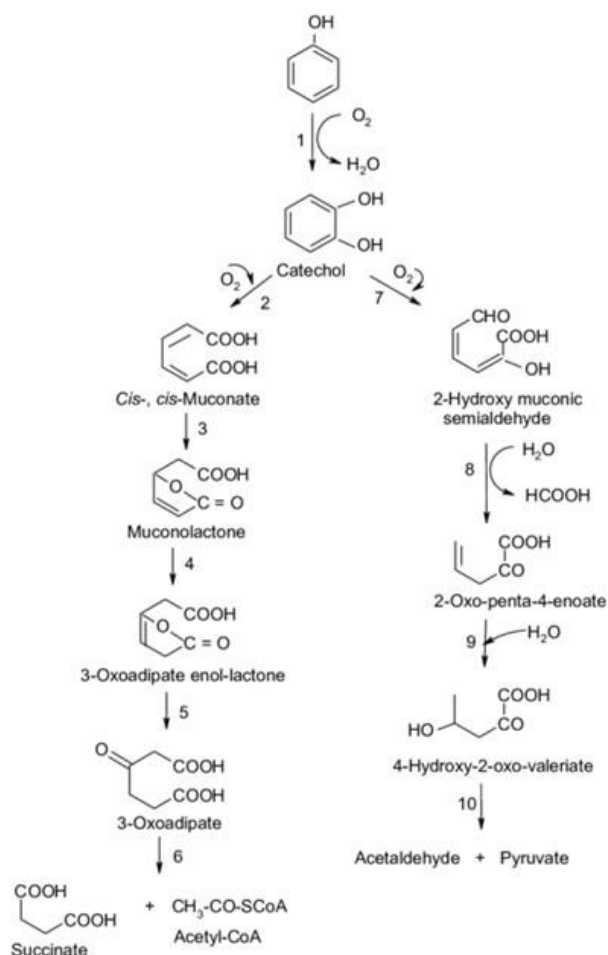


**Figure 1.4:** Metabolic pathway for the aerobic degradation of cyclohexane (Olajire and Essien 2014).

### 1.11 Degradation of Aromatic Compounds

Aromatic hydrocarbons are of great public concern due to their prolonged persistence in the environment and their harmful effects on human health (Abdel-Shafy and Mansour 2016, Thompson and Darwish 2019). Polycyclic aromatic hydrocarbons (PAHs) are especially of concern since they are known to be carcinogenic (Chebbi *et al.* 2017b). The PAHs are made up of four or more benzene rings which are strongly absorbed by soil and relatively insoluble in water. Hence, their degradation in contaminated soils is governed mainly by indigenous microbial degradation (Wu, Chen, *et al.* 2013). This process is slow, facilitating the need to identify PAH-degrading bacteria and understand the metabolic pathways involved for their complete degradation (Olowomofe *et al.* 2019).

The first step in aerobic degradation of PAHs is the hydroxylation of an aromatic ring via a monooxygenase or dioxygenase, with the formation of a *cis*-dihydrodiol. This gets rearomatised to a diol intermediate by the action of a dehydrogenase (Ghosal *et al.* 2016). The monooxygenases cleave the oxygen-oxygen bond between O<sub>2</sub> and inserts one oxygen atom into the aromatic ring and the other is reduced to water (Ladino-Orjuela *et al.* 2015). Under aerobic conditions, these diol intermediates may be cleaved by intradiol or extradiol ring-cleaving dioxygenase through either an ortho-cleavage or meta-cleavage pathway (**Figure 1.5**). The intermediates from this pathway, such as catechols are converted to TCA cycle intermediates (Mallick *et al.* 2011).



**Figure 1.5:** The two alternative pathways of aerobic degradation of phenol: *o*- and *m*-cleavage (Olajire and Essien 2014).



There are many catabolic pathways that have been elucidated for aromatic hydrocarbon degradation (Kadri *et al.* 2017). A good example is toluene, which can be degraded by bacteria in five different pathways as summarized in

**Table 1.5.** The key enzymes involved in degrading these compounds are only induced and synthesised in appreciable amounts when the substrates are present (Olajire and Essien 2014).

Microorganism	Pathways elucidated	References
TOL Plasmid	<pre> graph LR     A[Toluene] --&gt; B[Benzyl alcohol]     B --&gt; C[Benzaldehyde]     C --&gt; D[Benzoate]     D --&gt; E[TCA Cycle Intermediates]           </pre>	Harayama <i>et al.</i> (1999)
<i>P. putida</i> F1	<pre> graph LR     A[Toluene] --&gt; B[Addition of two hydroxyl groups]     B --&gt; C[cis-toluene dihydrodiol]     C --&gt; D[3-methylcatechol]           </pre>	Panke <i>et al.</i> (1998)
<i>P. mendocina</i> KR1	<pre> graph LR     A[Toluene] --&gt; B[Toluene 4-monoxygenase]     B --&gt; C[p-cresol]     C --&gt; D[Oxidation of methyl side chain]     D --&gt; E[p-hydroxybenzoate]           </pre>	Harayama <i>et al.</i> (1999) and Suyama <i>et al.</i> (1996)
<i>P. pickettii</i> PKO1	<pre> graph LR     A[Toluene] --&gt; B[Toluene 4-monoxygenase]     B --&gt; C[m-cresol]     C --&gt; D[Oxidation by monoxygenase]     D --&gt; E[3-methylcatechol]           </pre>	Harayama <i>et al.</i> (1999) and Suyama <i>et al.</i> (1996)
<i>Bukholderia cepacia</i> G4	<pre> graph LR     A[Toluene] --&gt; B[Toluene 4-monoxygenase]     B --&gt; C[o-cresol]     C --&gt; D[Oxidation by monoxygenase]     D --&gt; E[3-methylcatechol]           </pre>	Harayama <i>et al.</i> (1999)

**Table 1.5:** The divergent pathways for aerobic degradation of toluene.

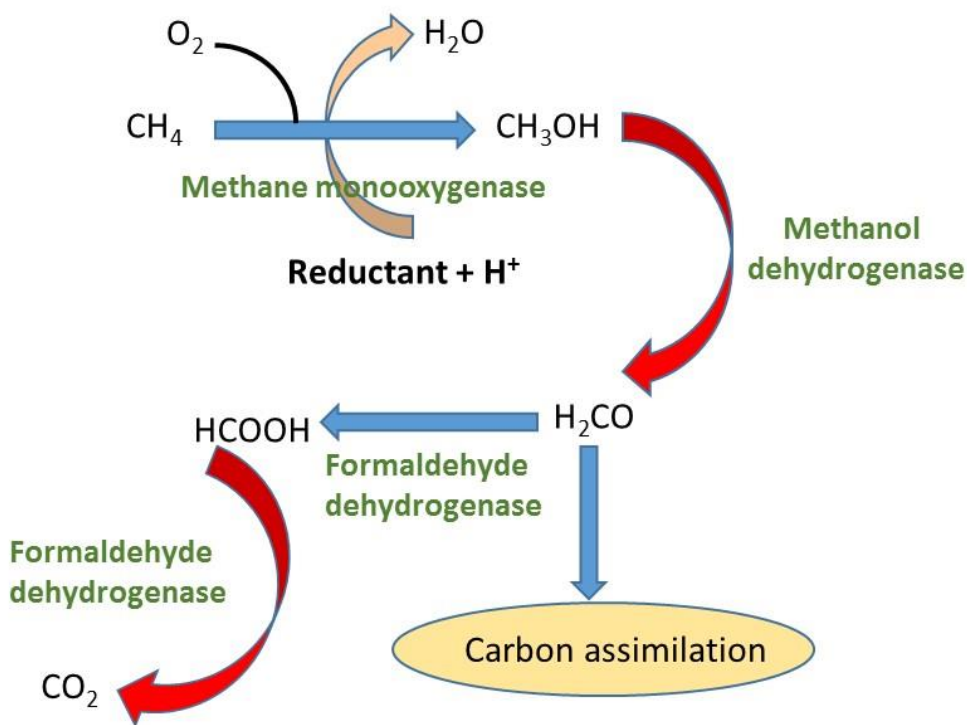
Microorganism	Pathways elucidated	References
TOL Plasmid	<pre> graph LR     Toluene --&gt; Benzyl_alcohol[Benzyl alcohol]     Benzyl_alcohol --&gt; Benzaldehyde     Benzaldehyde --&gt; Benzoate     Benzoate --&gt; TCA_Cycle[TCA Cycle Intermediates]             </pre>	Harayama <i>et al.</i> (1999)
<i>P. putida</i> F1	<pre> graph LR     Toluene --&gt; Hydroxyl[Addition of two hydroxyl groups]     Hydroxyl --&gt; cis_toluene[cis-toluene dihydrodiol]     cis_toluene --&gt; 3_methylcatechol[3-methylcatechol]             </pre>	Panke <i>et al.</i> (1998)
<i>P. mendocina</i> KR1	<pre> graph LR     Toluene --&gt; Toluene_4_monooxygenase[Toluene 4-monoxygenase]     Toluene_4_monooxygenase --&gt; p_cresol[p-cresol]     p_cresol --&gt; Oxidation[Oxidation of methyl side chain]     Oxidation --&gt; p_hydroxybenzoate[p-hydroxybenzoate]             </pre>	Harayama <i>et al.</i> (1999) and Suyama <i>et al.</i> (1996)
<i>P. pickettii</i> PKO1	<pre> graph LR     Toluene --&gt; Toluene_4_monooxygenase[Toluene 4-monoxygenase]     Toluene_4_monooxygenase --&gt; m_cresol[m-cresol]     m_cresol --&gt; Oxidation[Oxidation by monoxygenase]     Oxidation --&gt; 3_methylcatechol[3-methylcatechol]             </pre>	Harayama <i>et al.</i> (1999) and Suyama <i>et al.</i> (1996)
<i>Bukholderia cepacia</i> G4	<pre> graph LR     Toluene --&gt; Toluene_4_monooxygenase[Toluene 4-monoxygenase]     Toluene_4_monooxygenase --&gt; o_cresol[o-cresol]     o_cresol --&gt; Oxidation[Oxidation by monoxygenase]     Oxidation --&gt; 3_methylcatechol[3-methylcatechol]             </pre>	Harayama <i>et al.</i> (1999)

## 1.12 Hydroxylases and Hydrocarbon Degradation

The rate-limiting enzymes reported for aerobic alkane degradation are classified as alkane hydroxylases (AHs) (Liang *et al.* 2015). Microbial degradation requires an oxygen-dependent monooxygenase that converts alkanes to alkylalcohols, which are further oxidised to fatty acids and catabolised via the  $\beta$ -oxidation pathway. To date, four aerobic alkane degradation pathways have been identified through the years of study on alkane aerobic degradation (Wang and Shao 2013). However, the number of alkane hydroxylases isolated, characterised and analysed remains

limited (Ji *et al.* 2013a). Studies on these limited hydroxylases have shown that they belong to different enzyme families and classified based on their substrate range and degradation characteristics. These include methane monooxygenases, AlkB family hydroxylases and the cytochrome P450 alkane hydroxylases.

Methane monooxygenases (MMO) are involved in the first step of the catabolism of methane to form methanol (Park *et al.* 2019). Methanol is oxidised by methanol dehydrogenase to form formaldehyde which is converted to formate and carbon dioxide by formaldehyde dehydrogenase to provide energy for the cells (**Figure 1.6**) (Lieberman and Rosenzweig 2004).



**Figure 1.6:** Metabolic pathway of methanotrophs (Lieberman and Rosenzweig 2004).

Ammonia monooxygenase (AMO) is the only other enzyme besides the MMO-containing methanotrophic bacteria that can activate the C-H bond in methane. Methanol is an easily transported liquid fuel that can be easily obtained from a cheap and abundant natural gas (Shih *et*

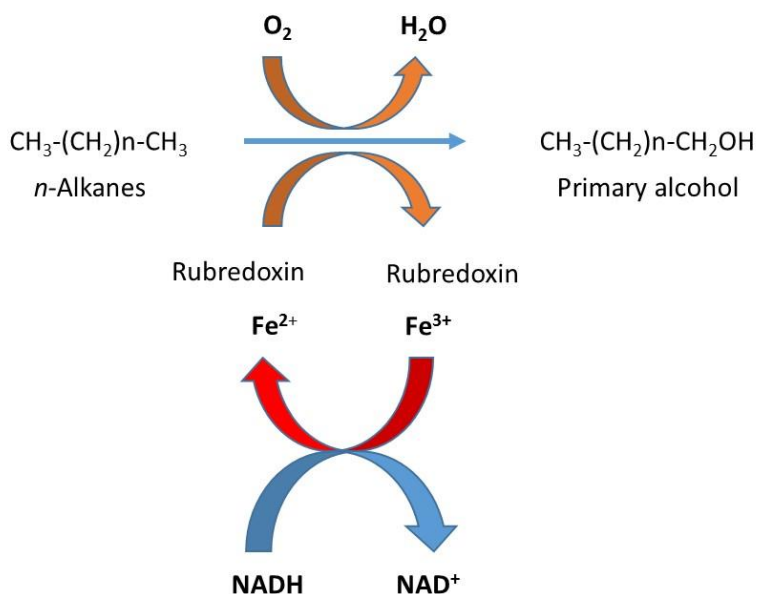
*al.* 2018). However, industries are faced with the problem of converting methane to methanol under ambient temperature and pressure due to its inert nature (104 kcal mol<sup>-1</sup> C-H bond) (Park *et al.* 2019). The identification of more MMOs and AMOs can be valuable in solving this problem. Another important property of MMOs is its ability to oxidise a variety of alkanes and alkenes in addition to methane.

There are two types of MMO, which includes a membrane-bound particulate MMO (pMMO) and a cytoplasmic, water-soluble MMO (sMMO) (Ross and Rosenzweig 2017). Most methanotrophs such as *Methylococcus capsulatus* contain both pMMO and sMMO and expression of specific ones is mostly dependent on the concentration of copper ions in the medium. Under low copper concentration, the cells express only sMMO while a higher copper concentration leads to the over expression of pMMO (Guggenheim *et al.* 2019). Furthermore, pMMO has a relatively narrow substrate specificity (short *n*-alkanes i.e. fewer than 5 carbons) while sMMO oxidises a broad range of substrates (saturated alkanes, alkenes, aromatics and chlorinated aromatics (Sirajuddin and Rosenzweig 2015, Lock *et al.* 2017).

The most predominate methane oxidation catalyst in nature are pMMOs but it has proven to be difficult to isolate due its multi-pass transmembrane protein nature (Ro *et al.* 2019). As a result, these proteins are difficult to purify for biochemical and biophysical characterisation due to its instability in detergents used to solubilise the enzyme (Wang *et al.*, 2017). Hence, researchers have opted to study sMMO that consist of three components: a hydroxylase (MMOH), a reductase (MMOR) and a regulatory protein (MMOB). The biochemistry, structure and mechanisms of sMMO are well understood to date which is in contrast to that of pMMO where further studies are required to obtain comprehensive information (Ross and Rosenzweig 2017).

The integral-membrane alkane monooxygenase (*alkB*)-related alkane hydroxylases are the most commonly found AHs occurring in both Gram-negative and Gram-positive bacteria with great diversity in sequences (Nie *et al.* 2014). The most widely characterised alkane degradation system is that of the *Alk* system identified from the *Pseudomonas putida* GPoI that oxidises middle and

long chain *n*-alkanes to alkanols. This system is comprised of three components namely, alkane hydroxylase (*alkB*), rubredoxin (AlkG), and rubredoxin reductase (AlkT) (van Beilen *et al.* 1994). The *alkB* is a non heme-iron integral membrane protein that is involved with the hydroxylation reaction. The reduction step takes place when electrons are transferred from the reduced NADH via its cofactor FAD to rubredoxin reductase, then subsequently to soluble rubredoxin, membrane bound hydroxylase and finally into the alkane as shown in **Figure 1.7** (Fuentes *et al.* 2014).



**Figure 1.7:** Degradation of *n*-alkane by integral-membrane alkane monooxygenase (*alkB*)-related alkane hydroxylases.

A diverse range of *alkB* genes have been detected and characterized from different microorganisms and environments (Mukherjee, Bhagowati, *et al.* 2017, Garrido-Sanz *et al.* 2019, Gregson *et al.* 2019, Kim, Dong, *et al.* 2019). Such studies have concluded that these monooxygenases are responsible for the initial degradation of medium and long chain *n*-alkanes under aerobic conditions.

The cytochrome P450 alkane hydroxylases are terminal monooxygenases that can be detected in all domains of life. Due to their ubiquitous nature, they have a diverse substrate range that includes fatty acids, steroids, organic solvents, ethanol, alkylaryl hydrocarbon products, carcinogens and pesticides (Ji *et al.* 2013a). However, they are commonly found in alkane-degrading bacteria that lack *alkB* but research has shown their co-existence with *alkB* thus expanding the *n*-alkane range of the host strain (Nie *et al.* 2014). These enzymes are categorised into two classes and are determined by the differences in the components of the P450 monooxygenase systems. These classes are the class I P450 (isolated from the mitochondria and bacteria) and the class II P450s. A large number of P450 enzymes are known (>4000) but those able to degrade *n*-alkanes are very few (Ji *et al.* 2013b).

### 1.13 Alcohol Dehydrogenases and Hydrocarbon Degradation

Alcohol dehydrogenases (ADHs) are widely present in all organisms and are involved in catalysing the reversible oxidation of alcohols to their corresponding aldehydes or ketones (Kim *et al.*, 2019). ADHs have a broad substrate range and play a pivotal role in a variety of physiological processes (Liu *et al.* 2009). With regards to its role in alkane biodegradation, ADH catalyses the second reaction step of the terminal oxidation pathways which involves the oxidation of an alkyl alcohol to an alkyl aldehyde (Wentzel *et al.* 2007). NAD(P)-dependent ADHs are divided into three major groups based on their metal content and molecular properties. These include medium-chain zinc-containing ADHs, short-chain ADHs and long-chain or iron-activated ADHs (Wu *et al.*, 2013).

A wide selection of ADHs have been isolated from alkane degrading bacteria (Elumalai *et al.* 2017). Some include *Bacillus* and *Geobacillus* (Elumalai *et al.* 2019), *Thermococcus kodakarensis* KOD1 (Wu *et al.*, 2013), *Alcanivorax borkumensis* SK2 (Gregson *et al.* 2019), *Gordonia iterans* (Kim, Dong, *et al.* 2019), *Geobacillus thermodenitrificans* NG80-2 (Liu *et al.* 2009), *Pseudomonas synxantha* LSH-7 (Meng *et al.* 2017). Such characterised ADHs are potentially useful in crude oil and hydrocarbon bioremediation as well as other bioconversion processes.

These enzymes work in synergy with other hydrocarbon related enzymes to degrade hydrocarbons in the environment until they are completely bio-transformed into less toxic compounds (Elumalai *et al.* 2017).

#### **1.14 Aldehyde Dehydrogenases and Hydrocarbon Degradation**

Aldehyde dehydrogenases (ALDH) are a group of diverse enzymes that catalyse the oxidation of aldehydes to their corresponding carboxylic acids (Sierra-García *et al.* 2014). These enzymes show a preference to either NAD or NADP<sup>+</sup> as a coenzyme (Kwok and Weiner 2005). Their roles in cellular metabolism are pivotal to the functioning of various processes and these have been extensively studied from bacteria to humans (Yakushi *et al.* 2018, Vassalli 2019). Hence, ALDHs have shown to be ubiquitous in all life forms by functioning as a detoxification enzyme to cope with toxic aldehydes generated from numerous cellular metabolic pathways (Li *et al.* 2010). With regards to its role in hydrocarbon degradation, they have been shown to be involved in the degradation of alkanes in which alkyl intermediates are generated (Okibe *et al.* 1999).

A number of reported ALDHs have been characterised from bacterial origin and oxidise short to medium chain alkyl aldehydes (Yakushi *et al.* 2018, Gregson *et al.* 2019, Vassalli 2019). Some of these bacteria include *Acetobacter pasteurianus* SKU1108 (Yakushi *et al.* 2018), *Thalassolituus oleivorans* MIL-1 (Gregson *et al.*, 2018), *Alcanivorax borkumensis* SK2<sup>T</sup> (Gregson *et al.* 2019), *Acinetobacter* spp. strain ADP1 (Hillen 1998), *Pseudomonas aeruginosa* strain ASP-53 (Mukherjee *et al.*, 2017), *Geobacillus thermoparaffinivorans* IR2 and *Bacillus licheniformis* MN6 (Elumalai *et al.* 2017). Studies have shown that a consortium of such bacteria are able to completely degrade various hydrocarbons in the environment since the array of enzymes work in synergy (Truskewycz *et al.* 2019). In theory, ALDHs should be able to work in synergy with monooxygenases and alcohol dehydrogenases since the alkyl aldehydes are a substrate for ALDHs.

### 1.15 Rationale and/or Motivation

Anthropogenic practices such as industrial activities, petroleum and derivatives and incomplete combustion of fossil fuels have caused an increase of petroleum hydrocarbons in the environment (Santos *et al.* 2011, Habib *et al.* 2018). Removal of these hydrocarbons is considered one of the main concerns with regards to bioremediation activities. The biotechnological tools consist of the use of whole microorganisms or isolated enzymes to degrade petroleum hydrocarbons into non- or less-toxic compounds (Van Hamme *et al.* 2003, Santos *et al.* 2011). However, there are certain drawbacks and concerns of the bioremediation strategy that is solely based on introduction of microbes whether genetically modified or not. This includes the time required for the microorganism to grow and produce the biodegrading enzyme, low levels of fitness and growth as compared to the indigenous population, the possibility of altering a natural ecosystem, thermal conditions and the lack of monitoring or control of released microorganisms in the long term (Wilson and Piel 2013). The use of enzymatic bioremediation methods that do not involve microorganisms especially that of an enzyme cocktail, has great potential. Enzymatic bioremediation can provide real benefits to the environment by avoiding the conditions that are required by whole-cell applications especially in extreme conditions, does not generate toxic by-products and offers increased rates of biodegradation. Some isolated attempts based on cultural and classical molecular biology techniques have been made in terms of developing or identifying enzymes that can play a role in degrading certain hydrocarbons (Rhee *et al.* 2005, Zampolli *et al.* 2014). A larger number of novel genes encoding for biotechnologically important enzymes were identified based on metagenomics than classical molecular biology approaches (Glogauer *et al.* 2011, Iqbal *et al.* 2014, Bouhajja *et al.* 2017, Fisher *et al.* 2018, Garrido-Sanz *et al.* 2019, Ji *et al.* 2019, Muriel-Millán *et al.* 2019). Besides, there is a need to develop a green technology that could be employed in a targeted progressive deconstruction of hydrocarbon complexes. As far as literature reviewed is concerned, no cocktail of enzymes targeting specific hydrocarbons has so far been developed based on metagenomic techniques.



## **1.16 Problem Statement**

The most important classes of organic pollutants in the environment are mineral oil constituents and halogenated products of petro-chemicals. Hydrocarbons are problematic environmental pollutants due to bioaccumulation in both soil and water environments and difficulty in biodegrading them, thereby posing detrimental health effects on living organisms (Habib *et al.* 2018, Liu *et al.* 2018, Ossai *et al.* 2020). There is a broad range of petrol hydrocarbon mixtures; these include crude oil, engine oil, diesel fuels and other fuel oil materials (Macaulay 2014). The major reason for the increase in these petrol hydrocarbon mixtures is the excessive production and consumption which is directly related to the impact of industrialised factories on the environment (Chandra *et al.* 2013, Pathak *et al.* 2015). The continued demand for crude oil as a fuel source in the absence of other solid and sustainable alternatives further increases the problem of oil contamination which has a significant ecological impact. It is also reported that most petroleum compounds have carcinogenic properties (Cao *et al.* 2009, Ghosal *et al.* 2016, Haleyur *et al.* 2016). As a result of this toxicity, several hydrocarbons are classified by the Agency of Environmental Protection (USA), the World Health Organisation and the European Union as top priority pollutants (Haleyur *et al.* 2016, Adipah 2018b). Therefore there is a growing demand to develop an environmentally friendly green technology that could be employed in a targeted progressive deconstruction of hydrocarbon complexes in oil-contaminated water and soil bodies.

## **1.17 Research Aim**

The aim of the study was to identify genes encoding for hydrocarbon degrading enzymes from oil contaminated soil using known enzymatic pathways and metagenomic techniques and engineer a cocktail of enzymes for biodegradation of petroleum hydrocarbons.

## 1.18 Objectives

1. To isolate total metagenomic DNA from hydrocarbon contaminated soil samples using a modified CTAB extraction method.
2. To construct a Fosmid library using metagenomic DNA from soil using the CopyControl™ HTP Fosmid Library Production Kit.
3. To conduct function driven analysis using specific hydrocarbon substrates.
4. To sequence fosmid constructs of the positive clones using PacBio Next Generation Sequencing platform and analyse using different bioinformatics tools.
5. To clone target genes into pET-30a(+) expression vector and transform into chemically competent *Escherichia coli*.
6. To purify the expressed proteins using Fast Protein Liquid Chromatography (FPLC).
7. To conduct kinetic studies of the purified candidate enzymes.
8. To engineer enzyme cocktails based on the candidate enzymes that act in synergy and evaluate the cocktails against hexanol as a substrate.

## 1.19 References

- Abbasian, F., Lockington, R., Mallavarapu, M., and Naidu, R., 2015. A Comprehensive Review of Aliphatic Hydrocarbon Biodegradation by Bacteria. *Applied Biochemistry and Biotechnology*, 176 (3), 670–699.
- Abdel-Shafy, H.I. and Mansour, M.S.M., 2016. A review on polycyclic aromatic hydrocarbons: Source, environmental impact, effect on human health and remediation. *Egyptian Journal of Petroleum*, 25 (1), 107–123.
- Abuhamed, T., Bayraktar, E., Mehmetoğlu, T., and Mehmetoğlu, Ü., 2004. Kinetics model for growth of *Pseudomonas putida* F1 during benzene, toluene and phenol biodegradation. *Process Biochemistry*, 39 (8), 983–988.
- Adams, G.O., Fufeyin, P.T., Okoro, S.E., and Ehinomen, I., 2015. Bioremediation, Biostimulation and Bioaugmentation: A Review. *International Journal of Environmental Bioremediation & Biodegradation*, 3 (1), 28–39.
- Adipah, S., 2018a. Remediation of Petroleum Hydrocarbons Contaminated Soil by Fenton's Oxidation. *Journal of Environmental Science and Public Health*, 02 (04), 168–178.
- Adipah, S., 2018b. Introduction of Petroleum Hydrocarbons Contaminants and its Human Effects. *Journal of Environmental Science and Public Health*, 03 (01), 1–9.
- Al-Dhabaan, F.A., 2019. Morphological, biochemical and molecular identification of petroleum hydrocarbons biodegradation bacteria isolated from oil polluted soil in Dhahran, Saud Arabia. *Saudi Journal of Biological Sciences*, 26 (6), 1247–1252.
- Al-Hawash, A.B., Dragh, M.A., Li, S., Alhujaily, A., Abbood, H.A., Zhang, X., and Ma, F., 2018. Principles of microbial degradation of petroleum hydrocarbons in the environment. *Egyptian Journal of Aquatic Research*, 44 (2), 71–76.
- Antwi-Akomeah, S., Fei-Baffoe, B., Belford, E.J.D.J.D., and Borigu, M., 2018. Hydrocarbon contaminated water remediation using a locally constructed multi-stage bioreactor incorporated with media filtration. *Global Journal of Environmental Science and Management*, 4 (4), 413–426.

- Araújo, S.C. da S., Silva-Portela, R.C.B., de Lima, D.C., da Fonsêca, M.M.B., Araújo, W.J., da Silva, U.B., Napp, A.P., Pereira, E., Vainstein, M.H., and Agnez-Lima, L.F., 2020. MBSBP1: a biosurfactant protein derived from a metagenomic library with activity in oil degradation. *Scientific Reports*, 10 (1), 1–13.
- Armstrong, Z., Liu, F., Kheirandish, S., Chen, H., Mewis, K., Duo, T., and Morgan-lang, C., 2019. as a Resource for Biocatalyst Development, 4 (4), 1–19.
- Atashgahi, S., Hornung, B., Van Der Waals, M.J., Da Rocha, U.N., Hugenholtz, F., Nijssse, B., Molenaar, D., Van Spanning, R., Stams, A.J.M., Gerritse, J., and Smidt, H., 2018. A benzene-degrading nitrate-reducing microbial consortium displays aerobic and anaerobic benzene degradation pathways. *Scientific Reports*, 8 (1), 1–12.
- Atlas, R.M. and Hazen, T.C., 2011. Atlas 2011 Oil spill history. *Environmental Science and Technology*, (Table 1), 6709–6715.
- Atlas, R.M., Stoeckel, D.M., Faith, S.A., Minard-Smith, A., Thorn, J.R., and Benotti, M.J., 2015. Oil Biodegradation and Oil-Degrading Microbial Populations in Marsh Sediments Impacted by Oil from the Deepwater Horizon Well Blowout. *Environmental Science and Technology*, 49 (14), 8356–8366.
- Awasthi, G., Kumari, A., Pant, A.B., and Srivastava, P., 2018. In silico identification and construction of microbial gene clusters associated with biodegradation of xenobiotic compounds. *Microbial Pathogenesis*, 114 (December 2017), 340–343.
- Ayed, H., Jemil, N., Maalej, H., Bayouhd, A., Hmidet, N., and Nasri, M., 2015. Enhancement of solubilization and biodegradation of diesel oil by biosurfactant from *Bacillus amyloliquefaciens* An6. *International Biodeterioration and Biodegradation*, 99 (April), 8–14.
- Azubuike, C.C., Chikere, C.B., and Okpokwasili, G.C., 2016. Bioremediation techniques—classification based on site of application: principles, advantages, limitations and prospects. *World Journal of Microbiology and Biotechnology*, 32 (11), 1–18.
- Bacosa, H.P., Erdner, D.L., Rosenheim, B.E., Shetty, P., Seitz, K.W., Baker, B.J., and Liu, Z., 2018. Hydrocarbon degradation and response of seafloor sediment bacterial community in

- the northern Gulf of Mexico to light Louisiana sweet crude oil. *ISME Journal*, 12 (10), 2532–2543.
- Van Beilen, J.B. and Funhoff, E.G., 2007. Alkane hydroxylases involved in microbial alkane degradation. *Applied Microbiology and Biotechnology*, 74 (1), 13–21.
- van Beilen, J.B., Panke, S., Lucchini, S., Franchini, A.G., Rothlis-berger, M., and Witholt, B., 2001. Analysis of *Pseudomonas putida* alkane degradation gene clusters and flanking insertion sequences: Evolution and regulation of the alk-genes. *Microbiology (United Kingdom)*, 147 (May), 1621–1630.
- van Beilen, J.B., Wubbolts, M.G., and Witholt, B., 1994. Genetics of alkane oxidation by *Pseudomonas oleovorans*. *Biodegradation*, 5 (3–4), 161–174.
- Bell, T. and Greer, C.W., 2015. Encyclopedia of Metagenomics, (December 2017).
- Berini, F., Casciello, C., Marcone, G.L., and Marinelli, F., 2017. Metagenomics: Novel enzymes from non-culturable microbes. *FEMS Microbiology Letters*, 364 (21), 1–19.
- Bleuven, C. and Landry, C.R., 2016. Molecular and cellular bases of adaptation to a changing environment in microorganisms. *Proceedings of the Royal Society B: Biological Sciences*, 283 (1841).
- Bouhajja, E., McGuire, M., Liles, M.R., Bataille, G., Agathos, S.N., and George, I.F., 2017. Identification of novel toluene monooxygenase genes in a hydrocarbon-polluted sediment using sequence- and function-based screening of metagenomic libraries. *Applied Microbiology and Biotechnology*, 101 (2), 797–808.
- Brooijmans, R.J.W., Pastink, M.I., and Siezen, R.J., 2009. Hydrocarbon-degrading bacteria: The oil-spill clean-up crew. *Microbial Biotechnology*, 2 (6), 587–594.
- Brzeszcz, J. and Kaszycki, P., 2018. Aerobic bacteria degrading both n-alkanes and aromatic hydrocarbons: an undervalued strategy for metabolic diversity and flexibility. *Biodegradation*, 29 (4), 359–407.
- Cao, B., Nagarajan, K., and Loh, K.C., 2009. Biodegradation of aromatic compounds: Current status and opportunities for biomolecular approaches. *Applied Microbiology and*

*Biotechnology*, 85 (2), 207–228.

Cardoso, J., Anderson, M., Herrgard, M., and Sonnenschein, N., 2015. Analysis of genetic variation and potential applications in genome-scale metabolic modeling. *Frontiers in Bioengineering and Biotechnology*, 3 (February), 1–12.

Cbsnews, 2015. Yellowstone River spill: Oil detected in water supplies. [online]. Available from: <http://www.cbsnews.com/news/yellowstone-river-spill-oil-detected-in-water-supplies/> [Accessed 28 Jan 2019].

Chandra, S., Sharma, R., Singh, K., and Sharma, A., 2013. Application of bioremediation technology in the environment contaminated with petroleum hydrocarbon. *Annals of Microbiology*, 63 (2), 417–431.

Chebbi, A., Hentati, D., Zaghden, H., Baccar, N., Rezgui, F., Chalbi, M., Sayadi, S., and Chamkha, M., 2017a. Polycyclic aromatic hydrocarbon degradation and biosurfactant production by a newly isolated *Pseudomonas* sp. strain from used motor oil-contaminated soil. *International Biodeterioration and Biodegradation*, 122, 128–140.

Chebbi, A., Hentati, D., Zaghden, H., Baccar, N., Rezgui, F., Chalbi, M., Sayadi, S., and Chamkha, M., 2017b. Polycyclic aromatic hydrocarbon degradation and biosurfactant production by a newly isolated *Pseudomonas* sp. strain from used motor oil-contaminated soil. *International Biodeterioration and Biodegradation*, 122.

Chen, W., Li, J., Sun, X., Min, J., and Hu, X., 2017. High efficiency degradation of alkanes and crude oil by a salt-tolerant bacterium *Dietzia* species CN-3. *International Biodeterioration and Biodegradation*, 118, 110–118.

Chen, Y., Wang, H., Wang, R., and Yun, Y., 2007. Effects of rhamnolipid on the biodegradation of n-hexadecane by microorganism and the cell surface hydrophobicity. *Environmental Science*, 28, 2117–2122.

Cheng, Q., Thomas, S.M., and Rouvière, P., 2002. Biological conversion of cyclic alkanes and cyclic alcohols into dicarboxylic acids: biochemical and molecular basis. *Applied Microbiology and Biotechnology*, 58 (6), 704–711.

Christova, N., Kabaivanova, L., Nacheva, L., Petrov, P., and Stoineva, I., 2019. Biodegradation of

- crude oil hydrocarbons by a newly isolated biosurfactant producing strain. *Biotechnology and Biotechnological Equipment*, 33 (1), 863–872.
- Church, P., 2015. Five of the most devastating pollution disasters in South Africa's history [online]. Available from: <https://www.thesouthafrican.com/the-worst-pollution-disasters-in-south-africa/> [Accessed 28 Jan 2019].
- Cowan, D.A., Ramond, J.B., Makhalanyane, T.P., and De Maayer, P., 2015. Metagenomics of extreme environments. *Current Opinion in Microbiology*, 25, 97–102.
- Cuthbertson, L. and Nodwell, J.R., 2013. The TetR Family of Regulators. *Microbiology and Molecular Biology Reviews*, 77 (3), 440–475.
- Das, N. and Chandran, P., 2010. Microbial Degradation of Petroleum Hydrocarbon Contaminants: An Overview. *Biotechnology Research International*, 2011 (March), 1–13.
- DeCastro, M.E., Rodríguez-Belmonte, E., and González-Siso, M.I., 2016. Metagenomics of thermophiles with a focus on discovery of novel thermozyms. *Frontiers in Microbiology*, 7 (SEP), 1–21.
- Dinamarca, M.A., Aranda-Olmedo, I., Puyet, A., and Rojo, F., 2003. Expression of the *Pseudomonas putida* OCT plasmid alkane degradation pathway is modulated by two different global control signals: Evidence from continuous cultures. *Journal of Bacteriology*, 185 (16), 4772–4778.
- Dombrowski, N., Donaho, J.A., Gutierrez, T., Seitz, K.W., Teske, A.P., and Baker, B.J., 2016. Reconstructing metabolic pathways of hydrocarbon-degrading bacteria from the Deepwater Horizon oil spill. *Nature Microbiology*, 1 (May), 1–8.
- Duarte, M., Nielsen, A., Camarinha-Silva, A., Vilchez-Vargas, R., Bruls, T., Wos-Oxley, M.L., Jauregui, R., and Pieper, D.H., 2017. Functional soil metagenomics: elucidation of polycyclic aromatic hydrocarbon degradation potential following 12 years of in situ bioremediation. *Environmental Microbiology*.
- Egobueze, F.E., Ayotamuno, J.M., Iwegbue, C.M.A., Eze, C., and Okparanma, R.N., 2019. Effects of organic amendment on some soil physicochemical characteristics and vegetative properties of *Zea mays* in wetland soils of the Niger Delta impacted with crude oil. *International Journal*

*of Recycling of Organic Waste in Agriculture*, 8 (s1), 423–435.

- Elumalai, P., Parthipan, P., Karthikeyan, O.P., and Rajasekar, A., 2017. Enzyme-mediated biodegradation of long-chain n-alkanes (C32 and C40) by thermophilic bacteria. *3 Biotech*, 7 (2), 1–10.
- Elumalai, P., Parthipan, P., Narenkumar, J., Anandakumar, B., Madhavan, J., Oh, B.T., and Rajasekar, A., 2019. Role of thermophilic bacteria (*Bacillus* and *Geobacillus*) on crude oil degradation and biocorrosion in oil reservoir environment. *3 Biotech*, 9 (3), 0.
- Fisher, O.S., Kenney, G.E., Ross, M.O., Ro, S.Y., Lemma, B.E., Batelu, S., Thomas, P.M., Sosnowski, V.C., DeHart, C.J., Kelleher, N.L., Stemmler, T.L., Hoffman, B.M., and Rosenzweig, A.C., 2018. Characterization of a long overlooked copper protein from methane- and ammonia-oxidizing bacteria. *Nature Communications*, 9 (1), 1–12.
- Fritsche, W. and Hofrichter, M., 2008. Aerobic Degradation by Microorganisms. *Biotechnology: Second, Completely Revised Edition*, 11–12, 144–167.
- Fuentes, S., Méndez, V., Aguila, P., and Seeger, M., 2014. Bioremediation of petroleum hydrocarbons: Catabolic genes, microbial communities, and applications. *Applied Microbiology and Biotechnology*, 98 (11), 4781–4794.
- Galazka, A., Grzadziel, J., Galazka, R., Ukalska-Jaruga, A., Strzelecka, J., and Smreczak, B., 2018. Genetic and functional diversity of bacterial microbiome in soils with long term impacts of petroleum hydrocarbons. *Frontiers in Microbiology*, 9 (AUG), 1–17.
- Garrido-Sanz, D., Redondo-Nieto, M., Guirado, M., Pindado Jiménez, O., Millán, R., Martín, M., and Rivilla, R., 2019. Metagenomic Insights into the Bacterial Functions of a Diesel-Degrading Consortium for the Rhizoremediation of Diesel-Polluted Soil. *Genes*, 10 (6), 456.
- Ghosal, D., Ghosh, S., Dutta, T.K., and Ahn, Y., 2016. Current state of knowledge in microbial degradation of polycyclic aromatic hydrocarbons (PAHs): A review. *Frontiers in Microbiology*, 7 (AUG).
- Girvan, H.M. and Munro, A.W., 2016. Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology. *Current Opinion in Chemical Biology*, 31, 136–145.



- Gkorezis, P., Daghighi, M., Franzetti, A., Van Hamme, J.D., Sillen, W., and Vangronsveld, J., 2016. The interaction between plants and bacteria in the remediation of petroleum hydrocarbons: An environmental perspective. *Frontiers in Microbiology*, 7 (NOV).
- Glogauer, A., Martini, V.P., Faoro, H., Couto, G.H., Müller-Santos, M., Monteiro, R.A., Mitchell, D.A., de Souza, E.M., Pedrosa, F.O., and Krieger, N., 2011. Identification and characterization of a new true lipase isolated through metagenomic approach. *Microbial cell factories*, 10 (1), 54.
- Godini, K., Samarghandi, M.R., Zafari, D., Rahmani, A.R., Afkhami, A., and Arabestani, M.R., 2018. Isolation and identification of new strains of crude oil degrading bacteria from Kharg Island, Iran. *Petroleum Science and Technology*, 36 (12), 869–874.
- Gregson, Benjamin H. McKew, B.A., Metodiev, M. V., Metodieva, G., and Golyshin, P.N., 2018. Differential Protein Expression During Growth on Medium Versus Long-Chain Alkanes in the Obligate Marine Hydrocarbon-Degrading Bacterium *Thalassolituus oleivorans* MIL-1. *Frontiers in Microbiology*, 9 (December), 1–14.
- Gregson, B.H., Metodieva, G., Metodiev, M. V., and McKew, B.A., 2019. Differential protein expression during growth on linear versus branched alkanes in the obligate marine hydrocarbon-degrading bacterium *Alcanivorax borkumensis* SK2T. *Environmental Microbiology*, 21 (7), 2347–2359.
- Guggenheim, C., Brand, A., Bürgmann, H., Sigg, L., and Wehrli, B., 2019. Aerobic methane oxidation under copper scarcity in a stratified lake. *Scientific Reports*, 9 (1), 1–11.
- Habib, S., Ahmad, S.A., Lutfi, W., Johari, W., Yunus, M., Shukor, A., and Yasid, N.A., 2018. SCIENCE & TECHNOLOGY Bioremediation of Petroleum Hydrocarbon in Antarctica by Microbial Species: An Overview. *Pertanika J. Sci. & Technol*, 26 (1), 1–20.
- Haleyur, N., Shahsavari, E., Mansur, A.A., Koshlaf, E., Morrison, P.D., Osborn, A.M., and Ball, A.S., 2016. Comparison of rapid solvent extraction systems for the GC-MS/MS characterization of polycyclic aromatic hydrocarbons in aged, contaminated soil. *MethodsX*, 3, 364–370.
- Van Hamme, J.D., Singh, A., and Ward, O.P., 2003. Recent Advances in Petroleum Microbiology

- Recent Advances in Petroleum Microbiology. *Microbiology and Molecular Biology Reviews*, 67 (4), 503–549.
- Handelsman, J., 2005. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*, 69 (1), 195–195.
- Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M., 1998. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chemistry and Biology*, 5 (10).
- Hauptmann A, Pachebat, J.A., Ii, M.C.K., Klein, A.L., Sicheritz-pontén, T., Cameron, K.A., and Bælum, J., 2017. Contamination of the Arctic reflected in microbial metagenomes from the Greenland ice sheet. *Environmental Research Letters*, 12, 074019.
- Hillen, W., 1998. Expression of Alkane Hydroxylase from *Acinetobacter* sp . Strain ADP1 Is Induced by a Broad Range of n -Alkanes and Requires the Transcriptional Activator AlkR, 180 (22), 5822–5827.
- Hua, F. and Wang, H.Q., 2014. Uptake and trans-membrane transport of petroleum hydrocarbons by microorganisms. *Biotechnology and Biotechnological Equipment*, 28 (2), 165–175.
- Iqbal, H.A., Craig, J.W., and Brady, S.F., 2014. Antibacterial enzymes from the functional screening of metagenomic libraries hosted in *Ralstonia metallidurans*. *FEMS Microbiology Letters*, 354 (1), 19–26.
- Ivshina, I.B., Kuyukina, M.S., Krivoruchko, A. V., Elkin, A.A., Makarov, S.O., Cunningham, C.J., Peshkur, T.A., Atlas, R.M., and Philp, J.C., 2015. Oil spill problems and sustainable response strategies through new technologies. *Environmental Sciences: Processes and Impacts*, 17 (7), 1201–1219.
- Iwuji, C.C., Okeke, O.C., Ezenwoke, B.C., Amadi, C.C., and Nwachukwu, H., 2016. Earth Resources Exploitation and Sustainable Development: Geological and Engineering Perspectives. *Engineering*, 08 (01), 21–33.
- Jaekel, U., Zedelius, J., Wilkes, H., and Musat, F., 2015. Anaerobic degradation of cyclohexane by sulfate-reducing bacteria from hydrocarbon-contaminated marine sediments. *Frontiers in Microbiology*, 6 (FEB), 1–11.

- Ji, L., Fu, X., Wang, M., Xu, C., Chen, G., Song, F., Guo, S., and Zhang, Q., 2019. Enzyme cocktail containing NADH regeneration system for efficient bioremediation of oil sludge contamination. *Chemosphere*, 233, 132–139.
- Ji, Y., Mao, G., Wang, Y., and Bartlam, M., 2013a. Structural insights into diversity and n-alkane biodegradation mechanisms of alkane hydroxylases. *Frontiers in Microbiology*, 4 (September 2014).
- Ji, Y., Mao, G., Wang, Y., and Bartlam, M., 2013b. Structural insights into diversity and n-alkane biodegradation mechanisms of alkane hydroxylases. *Frontiers in Microbiology*, 4 (March), 1–13.
- Jin, M., Gai, Y., Guo, X., Hou, Y., and Zeng, R., 2019. Properties and applications of extremozymes from deep-sea extremophilic microorganisms: A mini review. *Marine Drugs*, 17 (12).
- Joshi, M.N., Dhebar, S. V, Bhargava, P., Pandit, a S., Patel, R.P., Saxena, a K., and Bagatharia, S.B., 2014. Metagenomic Approach for Understanding Microbial Population from Petroleum Muck. *Genome Announcements*, 2 (3), e00533-14.
- Kachienga, L., Jitendra, K., and Momba, M., 2018. Metagenomic profiling for assessing microbial diversity and microbial adaptation to degradation of hydrocarbons in two South African petroleum-contaminated water aquifers /631/80 /631/326 /704/172 /13 /13/89 /82 /82/58 /82/75 /82/80 /82/83 /82/16 article. *Scientific Reports*, 8 (1), 1–6.
- Kadafa, A., 2012. Oil Exploration and Spillage in the Niger Delta of Nigeria. *Civil and Environmental Research*, 2 (3), 38–51.
- Kadri, T., Rouissi, T., Kaur Brar, S., Cledon, M., Sarma, S., and Verma, M., 2017. Biodegradation of polycyclic aromatic hydrocarbons (PAHs) by fungal enzymes: A review. *Journal of Environmental Sciences (China)*, 51, 52–74.
- Karlapudi, A.P., Venkateswarulu, T.C., Tammineedi, J., Kanumuri, L., Ravuru, B.K., Dirisala, V. ramu, and Kodali, V.P., 2018. Role of biosurfactants in bioremediation of oil pollution-a review. *Petroleum*, 4 (3), 241–249.
- Karlen, D.L., Mausbach, M., J., ., Doran, J.W., Cline, R.G., Harris, R.F., and Schuman, G.E., 1997.

- Soil Quality: A Concept, Definition, and Framework for Evaluation. *Soil Science Society of America Journal*, 61 (1), 4.
- Kasai, Y., Kishira, H., and Harayama, S., 2002. Bacteria Belonging to the Genus. *Applied and environmental microbiology*, 68 (11), 5625–5633.
- Khan, N.T., Jameel, N., and Khan, M.J., 2018. A Brief Overview of Contaminated Soil Remediation Methods. *BioTechnology: An Indian Journal Research* /, 14 (4), 168.
- Kim, G., Azmi, L., Jang, S., Jung, T., Hebert, H., Roe, A.J., Byron, O., and Song, J.J., 2019. Aldehyde-alcohol dehydrogenase forms a high-order spiroosome architecture critical for its activity. *Nature Communications*, 10 (1).
- Kim, H.S., Dong, K., Kim, J., and Lee, S.S., 2019. Characteristics of crude oil-degrading bacteria *Gordonia iterans* isolated from marine coastal in Taeae sediment. *MicrobiologyOpen*, 8 (6), 1–10.
- Kleindienst, S., Paul, J.H., and Joye, S.B., 2015. Using dispersants after oil spills: Impacts on the composition and activity of microbial communities. *Nature Reviews Microbiology*, 13 (6), 388–396.
- Kosaric, N., 2001. Biosurfactants and their application for soil bioremediation. *Food Technology and Biotechnology*, 39 (295–304), no-no.
- Krasowska, A. and Sigler, K., 2014. How microorganisms use hydrophobicity and what does this mean for human needs? *Frontiers in Cellular and Infection Microbiology*, 4 (August), 1–7.
- Kusnezowa, A. and Leichert, L.I., 2017. In silico approach to designing rational metagenomic libraries for functional studies. *BMC Bioinformatics*, 18 (1), 1–11.
- Kwok, K.H. and Weiner, H., 2005. Isolation and characterization of an aldehyde dehydrogenase encoded by the aldB gene of *Escherichia coli*. *Journal of Bacteriology*, 187 (3), 1067–1073.
- Ladino-Orjuela, G., Gomes, E., and Silva, R. da, 2015. Reviews of Environmental Contamination and Toxicology. *Reviews of Environmental Contamination and Toxicology*, 236 (November 2017), 1–297.
- Lam, K.N., Cheng, J., Engel, K., Neufeld, J.D., and Charles, T.C., 2015. Current and future

- resources for functional metagenomics. *Frontiers in Microbiology*, 6 (OCT), 1–8.
- Lamichhane, S., Krishna, K.C.B., and Sarukkalige, R., 2016. Chemosphere Polycyclic aromatic hydrocarbons ( PAHs ) removal by sorption : A review. *Chemosphere*, 148, 336–353.
- Lee, E.H. and Cho, K.S., 2008. Characterization of cyclohexane and hexane degradation by *Rhodococcus* sp. EC1. *Chemosphere*, 71 (9), 1738–1744.
- Lee, M.H. and Lee, S.-W., 2013. Bioprospecting Potential of the Soil Metagenome: Novel Enzymes and Bioactivities. *Genomics & Informatics*, 11 (3), 114.
- Lee, Y., Lee, Y., and Jeon, C.O., 2019. Biodegradation of naphthalene, BTEX, and aliphatic hydrocarbons by *Paraburkholderia aromaticivorans* BN5 isolated from petroleum-contaminated soil. *Scientific Reports*, 9 (1), 24–30.
- Li, H., Li, X., Yu, T., Wang, F., and Qu, C., 2019. Study on Extreme Microbial Degradation of Petroleum Hydrocarbons. *IOP Conference Series: Materials Science and Engineering*, 484 (1).
- Li, X., Li, Y., Wei, D., Li, P., Wang, L., and Feng, L., 2010. Characterization of a broad-range aldehyde dehydrogenase involved in alkane degradation in *Geobacillus thermodenitrificans* NG80-2. *Microbiological Research*, 165 (8), 706–712.
- Liang, J.-L., 2017. Crystal Structure of TetR Family Biodegradation of n -Alkanes, 83 (21), 1–15.
- Liang, J.-L., Wang, M., Nie, Y., Wu, X.-L., Maser, E., Xiong, G., and Wang, Y.-P., 2015. Regulation of alkane degradation pathway by a TetR family repressor via an autoregulation positive feedback mechanism in a Gram-positive *Dietzia* bacterium. *Molecular Microbiology*, 99 (2), 338–359.
- Lieberman, R.L. and Rosenzweig, A.C., 2004. Biological methane oxidation: Regulation, biochemistry, and active site structure of particulate methane monooxygenase. *Critical Reviews in Biochemistry and Molecular Biology*, 39 (3), 147–164.
- Lim, M.W., Lau, E. Von, and Poh, P.E., 2016. A comprehensive guide of remediation technologies for oil contaminated soil — Present works and future directions. *Marine Pollution Bulletin*.
- Lineback, N. & Lineback, M., 2013. World's Largest Oil Spills [online]. Available from:

<http://voices.nationalgeographic.com/2013/10/12/geography-in-the-news-oil-spills/>  
[Accessed 28 Jan 2019].

- Liu, B., Liu, J., Ju, M., Li, X., and Yu, Q., 2016. Purification and characterization of biosurfactant produced by *Bacillus licheniformis* Y-1 and its application in remediation of petroleum contaminated soil. *Marine Pollution Bulletin*, 107 (1), 46–51.
- Liu, J., Zheng, Y., Lin, H., Wang, X., Li, M., Liu, Y., Yu, M., Zhao, M., Pedentchouk, N., Lea-Smith, D.J., Todd, J.D., Magill, C.R., Zhang, W.J., Zhou, S., Song, D., Zhong, H., Xin, Y., Yu, M., Tian, J., and Zhang, X.H., 2019. Proliferation of hydrocarbon-degrading microbes at the bottom of the Mariana Trench. *Microbiome*, 7 (1), 1–13.
- Liu, X., Dong, Y., Zhang, J., Zhang, A., Wang, L., and Feng, L., 2009. Two novel metal-independent long-chain alkyl alcohol dehydrogenases from *Geobacillus thermodenitrificans* NG80-2. *Microbiology*, 155 (6), 2078–2085.
- Liu, Y., Ding, A., Sun, Y., Xia, X., and Zhang, D., 2018. Impacts of n-alkane concentration on soil bacterial community structure and alkane monooxygenase genes abundance during bioremediation processes. *Frontiers of Environmental Science and Engineering*, 12 (5), 1–13.
- Lock, M., Nichol, T., Murrell, J.C., and Smith, T.J., 2017. Mutagenesis and expression of methane monooxygenase to alter regioselectivity with aromatic substrates. *FEMS microbiology letters*, 364 (13), 1–6.
- M D Yuniati, 2018. Bioremediation of petroleum-contaminated soil : A Review Bioremediation of petroleum-contaminated soil : A Review. *IOP Conference Series: Earth and Environmental Science*, 118 (1).
- Ma, Y., Li, X., Mao, H., Wang, B., and Wang, P., 2018. Remediation of hydrocarbon–heavy metal co-contaminated soil by electrokinetics combined with biostimulation. *Chemical Engineering Journal*, 353 (April), 410–418.
- Maeng, J.H.O., Sakai, Y., Tani, Y., and Kato, N., 1996. Isolation and characterization of a novel oxygenase that catalyzes the first step of n-alkane oxidation in *Acinetobacter* sp. strain M-1. *Journal of Bacteriology*, 178 (13), 3695–3700.

- Makombe, N. and Gwisai, R.D., 2018. Soil Remediation Practices for Hydrocarbon and Heavy Metal Reclamation in Mining Polluted Soils. *Scientific World Journal*, 2018, 6–13.
- Malla, M.A., Dubey, A., Yadav, S., Kumar, A., Hashem, A., and Abd-Allah, E.F., 2018. Understanding and designing the strategies for the microbe-mediated remediation of environmental contaminants using omics approaches. *Frontiers in Microbiology*, 9 (JUN).
- Mallick, S., Chakraborty, J., and Dutta, T.K., 2011. Role of oxygenases in guiding diverse metabolic pathways in the bacterial degradation of low-molecular-weight polycyclic aromatic hydrocarbons: A review. *Critical Reviews in Microbiology*, 37 (1), 64–90.
- Marine or Coastal Oil Spill report, 2014. Marine or Coastal Oil Spill [online]. Available from: <https://www.capetown.gov.za/en/DRM/Pages/MarineorCoastalOilSpill.aspx> [Accessed 28 Jan 2018].
- McDonald, I.R., Murrell, J.C., Wendlandt, K.D., Miguez, C.B., Rogge, G., Groleau, D., and Bourque, D., 2006. Diversity of soluble methane monooxygenase-containing methanotrophs isolated from polluted environments. *FEMS Microbiology Letters*, 255 (2), 225–232.
- Meng, L., Li, H., Bao, M., and Sun, P., 2017. Metabolic pathway for a new strain *Pseudomonas synxantha* LSH-7<sup>T</sup>: From chemotaxis to uptake of n-hexadecane. *Scientific Reports*, 7 (January), 1–13.
- Merino, N., Aronson, H.S., Bojanova, D.P., Feyhl-Buska, J., Wong, M.L., Zhang, S., and Giovannelli, D., 2019. Living at the extremes: Extremophiles and the limits of life in a planetary context. *Frontiers in Microbiology*, 10 (MAR).
- Miller, J.I., Techtmann, S., Fortney, J., Mahmoudi, N., Joyner, D., Liu, J., Olesen, S., Alm, E., Fernandez, A., Gardinali, P., GaraJayeva, N., Askerov, F.S., and Hazen, T.C., 2019. Oil hydrocarbon degradation by caspian sea microbial communities. *Frontiers in Microbiology*, 10 (MAY), 1–15.
- Mittal, A. and Singh, P., 2009. Isolation of hydrocarbon degrading bacteria from soils contaminated with crude oil spills. *Indian Journal of Experimental Biology*, 47 (9), 760–765.
- Mukherjee, A., Chettri, B., Langpoklakpam, J.S., Basak, P., Prasad, A., Mukherjee, A.K., Bhattacharyya, M., Singh, A.K., and Chattopadhyay, D., 2017. Bioinformatic Approaches

- Including Predictive Metagenomic Profiling Reveal Characteristics of Bacterial Response to Petroleum Hydrocarbon Contamination in Diverse Environments. *Scientific Reports*, 7 (1), 1–22.
- Mukherjee, A.K., Bhagowati, P., Biswa, B.B., Chanda, A., and Kalita, B., 2017. A comparative intracellular proteomic profiling of *Pseudomonas aeruginosa* strain ASP-53 grown on pyrene or glucose as sole source of carbon and identification of some key enzymes of pyrene biodegradation pathway. *Journal of Proteomics*, 167, 25–35.
- Muriel-Millán, L.F., Rodríguez-Mejía, J.L., Godoy-Lozano, E.E., Rivera-Gómez, N., Gutierrez-Rios, R.M., Morales-Guzmán, D., Trejo-Hernández, M.R., Estradas-Romero, A., and Pardo-López, L., 2019. Functional and Genomic Characterization of a *Pseudomonas aeruginosa* Strain Isolated From the Southwestern Gulf of Mexico Reveals an Enhanced Adaptation for Long-Chain Alkane Degradation. *Frontiers in Marine Science*, 6 (September), 1–15.
- Ngara, T.R. and Zhang, H., 2018. Recent Advances in Function-based Metagenomic Screening. *Genomics, Proteomics and Bioinformatics*, 16 (6), 405–415.
- Nie, Y., Chi, C.Q., Fang, H., Liang, J.L., Lu, S.L., Lai, G.L., Tang, Y.Q., and Wu, X.L., 2014. Diverse alkane hydroxylase genes in microorganisms and environments. *Scientific Reports*, 4, 1–11.
- NRC, 2000. *Natural attenuation for groundwater remediation, committee on intrinsic remediation, water science and technology board, board on radioactive waste management. National Research Council*, 292.
- Obi, L.U., Atagana, H.I., and Adeleke, R.A., 2016. Isolation and characterisation of crude oil sludge degrading bacteria. *SpringerPlus*, 5 (1).
- Okibe, N., Amada, K., Hirano, S.I., Haruki, M., Imanaka, T., Morikawa, M., and Kanaya, S., 1999. Gene cloning and characterization of aldehyde dehydrogenase from a petroleum-degrading bacterium, strain HD-1. *Journal of Bioscience and Bioengineering*, 88 (1), 7–11.
- Olajire, A. and Essien, J., 2014. Aerobic Degradation of Petroleum Components by Microbial Consortia. *Journal of Petroleum & Environmental Biotechnology*, 05 (05).
- Olowomofe, T.O., Oluyeye, J.O., Aderiye, B.I., and Oluwole, O.A., 2019. Degradation of poly



- aromatic fractions of crude oil and detection of catabolic genes in hydrocarbon-degrading bacteria isolated from Agbabu bitumen sediments in Ondo State, 5 (September), 308–323.
- Opere, B.O., Obayori, O.S., and Raji, a a, 2013. Degradation of Cyclohexane and Cyclohexanone by *Bacillus lentus* Strain LP32. *African Journal of Biotechnology*, 12 (47), 6632–6635.
- Ossai, I.C., Ahmed, A., Hassan, A., and Hamid, F.S., 2020. Remediation of soil and water contaminated with petroleum hydrocarbon: A review. *Environmental Technology and Innovation*, 17, 100526.
- Panigrahi, S., Velraj, P., and Subba Rao, T., 2019. Functional Microbial Diversity in Contaminated Environment and Application in Bioremediation. *Microbial Diversity in the Genomic Era*, (January), 359–385.
- Park, C. and Park, W., 2018. Survival and energy producing strategies of Alkane degraders under extreme conditions and their biotechnological potential. *Frontiers in Microbiology*, 9 (MAY), 1–15.
- Park, M.B., Park, E.D., and Ahn, W.-S., 2019. Recent Progress in Direct Conversion of Methane to Methanol Over Copper-Exchanged Zeolites. *Frontiers in Chemistry*, 7 (July), 1–7.
- Pathak, C., Education, M., Chandrakant, H., Ahmedabad, M., and Corporation, M., 2015. Petroleum Industries : Environmental Pollution Effects , Management and Treatment Methods E NVIRONMENTAL P OLLUTION E FFECTS , (October).
- Patowary, K., Patowary, R., Kalita, M.C., and Deka, S., 2016. Development of an efficient bacterial consortium for the potential remediation of hydrocarbons from contaminated sites. *Frontiers in Microbiology*, 7 (JUL), 1–14.
- Patowary, K., Patowary, R., Kalita, M.C., and Deka, S., 2017. Characterization of biosurfactant produced during degradation of hydrocarbons using crude oil as sole source of carbon. *Frontiers in Microbiology*, 8 (FEB), 1–14.
- Peixoto, R.S., Vermelho, A.B., and Rosado, A.S., 2011. Petroleum-degrading enzymes: Bioremediation and new prospects. *Enzyme Research*, 2011 (1).
- Perrone, F., De Siena, B., Muscariello, L., Kendall, S.L., Waddell, S.J., and Sacco, M., 2017. A

- Novel TetR-like transcriptional regulator is induced in acid-nitrosative stress and controls expression of an efflux pump in mycobacteria. *Frontiers in Microbiology*, 8 (OCT), 1–10.
- Phale, P.S., Sharma, A., and Gautam, K., 2019. *Microbial degradation of xenobiotics like aromatic pollutants from the terrestrial environments*. Pharmaceuticals and Personal Care Products: Waste Management and Treatment Technology. Elsevier Inc.
- Ren, M., Zhang, Z., Wang, X., Zhou, Z., Chen, D., Zeng, H., Zhao, S., Chen, L., Hu, Y., Zhang, C., Liang, Y., She, Q., Zhang, Y., and Peng, N., 2018. Diversity and contributions to nitrogen cycling and carbon fixation of soil salinity shaped microbial communities in Tarim Basin. *Frontiers in Microbiology*, 9 (MAR), 1–14.
- Rhee, J.K., Ahn, D.G., Kim, Y.G., and Oh, J.W., 2005. New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. *Applied and Environmental Microbiology*, 71 (2), 817–825.
- Ro, S.Y., Schachner, L.F., Koo, C.W., Purohit, R., Remis, J.P., Kenney, G.E., Liauw, B.W., Thomas, P.M., Patrie, S.M., Kelleher, N.L., and Rosenzweig, A.C., 2019. Native top-down mass spectrometry provides insights into the copper centers of membrane-bound methane monooxygenase. *Nature Communications*, 10 (1), 1–12.
- Rodgers, T. and Zhang, H., 2019. Recent Advances in Function-based Metagenomic Screening. *Genomics, Proteomics & Bioinformatics*, 16 (6), 405–415.
- Rojo, F., 2009. Degradation of alkanes by bacteria: Minireview. *Environmental Microbiology*, 11 (10), 2477–2490.
- Ron, E.Z. and Rosenberg, E., 2014. Enhanced bioremediation of oil spills in the sea. *Current Opinion in Biotechnology*, 27, 191–194.
- Ross, M.O. and Rosenzweig, A.C., 2017. A tale of two methane monooxygenases. *Journal of Biological Inorganic Chemistry*, 22 (2–3), 307–319.
- Rouvière, P.E. and Chen, M.W., 2003. Isolation of *Brachymonas petroleovorans* CHX, a novel cyclohexane-degrading  $\beta$ -proteobacterium. *FEMS Microbiology Letters*, 227 (1), 101–106.
- Roy, A., Dutta, A., Pal, S., Gupta, A., Sarkar, J., Chatterjee, A., Saha, A., Sarkar, P., Sar, P., and

- Kazy, S.K., 2018. Biostimulation and bioaugmentation of native microbial community accelerated bioremediation of oil refinery sludge. *Bioresource Technology*, 253.
- Safdari, M.S., Kariminia, H.R., Rahmati, M., Fazlollahi, F., Polasko, A., Mahendra, S., Wilding, W.V., and Fletcher, T.H., 2018. Development of bioreactors for comparative study of natural attenuation, biostimulation, and bioaugmentation of petroleum-hydrocarbon contaminated soil. *Journal of Hazardous Materials*, 342, 270–278.
- Santos, H.F., Carmo, F.L., Paes, J.E.S., Rosado, A.S., and Peixoto, R.S., 2011. Bioremediation of mangroves impacted by petroleum. *Water, Air, and Soil Pollution*, 216 (1–4), 329–350.
- Sharma, S., Tiwari, S., Hasan, A., Saxena, V., and Pandey, L.M., 2018. Recent advances in conventional and contemporary methods for remediation of heavy metal-contaminated soils. *3 Biotech*, 8 (4), 1–18.
- Shih, C.F., Zhang, T., Li, J., and Bai, C., 2018. Powering the Future with Liquid Sunshine. *Joule*, 2 (10), 1925–1949.
- Sierra-García, I.N., Alvarez, J.C., De Vasconcellos, S.P., De Souza, A.P., Dos Santos Neto, E.V., and De Oliveira, V.M., 2014. New hydrocarbon degradation pathways in the microbial metagenome from brazilian petroleum reservoirs. *PLoS ONE*, 9 (2).
- Siles, J.A. and Margesin, R., 2018. Insights into microbial communities mediating the bioremediation of hydrocarbon-contaminated soil from an Alpine former military site. *Applied Microbiology and Biotechnology*, 102 (10), 4409–4421.
- Silva, N.M., De Oliveira, A.M.S.A., Pegorin, S., Giusti, C.E., Ferrari, V.B., Barbosa, D., Martins, L.F., Morais, C., Setubal, J.C., Vasconcellos, S.P., Da Silva, A.M., De Oliveira, J.C.F., Pascon, R.C., and Viana-Niero, C., 2019. Characterization of novel hydrocarbon-degrading *Gordonia paraffinivorans* and *Gordonia sihwensis* strains isolated from composting. *PLoS ONE*, 14 (4), 1–16.
- Simon, C. and Daniel, R., 2011. Metagenomic analyses: Past and future trends. *Applied and Environmental Microbiology*, 77 (4), 1153–1161.
- Sirajuddin, S. and Rosenzweig, A.C., 2015. Enzymatic oxidation of methane. *Biochemistry*, 54 (14), 2283–2294.

- Song, M., Yang, Y., Jiang, L., Hong, Q., Zhang, D., Shen, Z., Yin, H., and Luo, C., 2017. Characterisation of the phenanthrene degradation-related genes and degrading ability of a newly isolated copper-tolerant bacterium. *Environmental Pollution*, 220, 1059–1067.
- Steele, H.L. and Streit, W.R., 2005. Metagenomics: Advances in ecology and biotechnology. *FEMS Microbiology Letters*, 247 (2), 105–111.
- Stefani, F.O.P., Bell, T.H., Marchand, C., De La Providencia, I.E., El Yassimi, A., St-Arnaud, M., and Hijri, M., 2015. Culture-dependent and -independent methods capture different microbial community fractions in hydrocarbon-contaminated soils. *PLoS ONE*, 10 (6), 1–16.
- Tavormina, P.L., Shiller, A.M., Villareal, T.A., Hunter, K.S., Nigro, L.M., Joye, S.B., Amon, R.M.W., Orphan, V.J., Asper, V.L., Montoya, J.P., Diercks, A.-R., Crespo-Medina, M., Meile, C.D., Bracco, A., Battles, J.J., Chanton, J.P., Wood, A.M., and Joung, D.-J., 2015. Erratum: The rise and fall of methanotrophy following a deepwater oil-well blowout. *Nature Geoscience*, 8 (6), 490–490.
- Taylor, M.J., Alabdrabalameer, H.A., and Skoulou, V., 2019. Choosing physical, physicochemical and chemical methods of pre-treating lignocellulosic wastes to repurpose into solid fuels. *Sustainability (Switzerland)*, 11 (13).
- Thompson, L.A. and Darwish, W.S., 2019. Environmental Chemical Contaminants in Food: Review of a Global Problem. *Journal of Toxicology*, 2019.
- Tremblay, J., Yergeau, E., Fortin, N., Cobanli, S., Elias, M., King, T.L., Lee, K., and Greer, C.W., 2017. Chemical dispersants enhance the activity of oil- and gas condensate-degrading marine bacteria. *ISME Journal*, 11 (12), 2793–2808.
- Truskewycz, A., Gundry, T.D., Khudur, L.S., Kolobaric, A., Taha, M., Aburto-Medina, A., Ball, A.S., and Shahsavari, E., 2019. Petroleum hydrocarbon contamination in terrestrial ecosystems—fate and microbial responses. *Molecules*, 24 (18), 1–20.
- Unimke, A.A., Mmuoegbulam, O.A., and Anika, O.C., 2018. Microbial Degradation of Petroleum Hydrocarbons: Realities, Challenges and Prospects. *Biotechnology Journal International*, 22 (2), 1–10.
- Varjani, S.J., 2017. Microbial degradation of petroleum hydrocarbons. *Bioresource Technology*,

223, 277–286.

- Varjani, S.J. and Upasani, V.N., 2016. Biodegradation of petroleum hydrocarbons by oleophilic strain of *Pseudomonas aeruginosa* NCIM 5514. *Bioresource Technology*, 222.
- Varjani, S.J. and Upasani, V.N., 2017a. A new look on factors affecting microbial degradation of petroleum hydrocarbon pollutants. *International Biodeterioration and Biodegradation*, 120, 71–83.
- Varjani, S.J. and Upasani, V.N., 2017b. Critical review on biosurfactant analysis, purification and characterization using rhamnolipid as a model biosurfactant. *Bioresource Technology*, 232, 389–397.
- Varjani, S.J. and Upasani, V.N., 2017c. A new look on factors affecting microbial degradation of petroleum hydrocarbon pollutants. *International Biodeterioration and Biodegradation*.
- Vasconcellos, S.P., Sierra-Garcia, I.N., Dellagnezze, B.M., Vicentini, R., Midgley, D., Silva, C.C., Santos Neto, E. V., Volk, H., Hendry, P., and Oliveira, V.M., 2017. Functional and genetic characterization of hydrocarbon biodegrader and exopolymer-producing clones from a petroleum reservoir metagenomic library. *Environmental Technology*, 3330 (August), 1–12.
- Vassalli, G., 2019. Aldehyde dehydrogenases: Not just markers, but functional regulators of stem cells. *Stem Cells International*, 2019.
- Venny, Gan, S., and Ng, H.K., 2012. Inorganic chelated modified-Fenton treatment of polycyclic aromatic hydrocarbon (PAH)-contaminated soils. *Chemical Engineering Journal*, 180, 1–8.
- Vikram, S., Singh, V., Singh, D., Kumar, S., and Tripathi, M., 2018. Metagenomic Approach towards Bioprospection of Novel Biomolecule(s) and Environmental Bioremediation. *Annual Research & Review in Biology*, 22 (2), 1–12.
- Wanapaisan, P., Laothamteep, N., Vejarano, F., Chakraborty, J., Shintani, M., Muangchinda, C., Morita, T., Suzuki-Minakuchi, C., Inoue, K., Nojiri, H., and Pinyakong, O., 2018. Synergistic degradation of pyrene by five culturable bacteria in a mangrove sediment-derived bacterial consortium. *Journal of Hazardous Materials*, 342, 561–570.
- Wang, H., Wang, B., Dong, W., and Hu, X., 2016. Co-acclimation of bacterial communities under

- stresses of hydrocarbons with different structures. *Scientific Reports*, 6 (1), 34588.
- Wang, V.C.C., Maji, S., Chen, P.P.Y., Lee, H.K., Yu, S.S.F., and Chan, S.I., 2017. Alkane Oxidation: Methane Monooxygenases, Related Enzymes, and Their Biomimetics. *Chemical Reviews*, 117 (13), 8574–8621.
- Wang, W. and Shao, Z., 2013. Enzymes and genes involved in aerobic alkane degradation. *Frontiers in Microbiology*, 4 (MAY), 1–7.
- Wang, Y., Liang, J., Wang, J., and Gao, S., 2018. Combining stable carbon isotope analysis and petroleum-fingerprinting to evaluate petroleum contamination in the Yanchang oilfield located on loess plateau in China. *Environmental Science and Pollution Research*, 25 (3), 2830–2841.
- Wentzel, A., Ellingsen, T.E., Kotlar, H.K., Zotchev, S.B., and Throne-Holst, M., 2007. Bacterial metabolism of long-chain n-alkanes. *Applied Microbiology and Biotechnology*, 76 (6), 1209–1221.
- Widdel, F. and Rabus, R., 2001. Anaerobic biodegradation of saturated and aromatic hydrocarbons. *Current Opinion in Biotechnology*, 12 (3), 259–276.
- Wilkins, L.G.E., Ettinger, C.L., Jospin, G., and Eisen, J.A., 2019. Metagenome-assembled genomes provide new insight into the microbial diversity of two thermal pools in Kamchatka, Russia. *Scientific Reports*, 9 (1), 1–15.
- Wilson, M.C. and Piel, J., 2013. Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chemistry and Biology*, 20 (5), 636–647.
- Wu, M., Chen, L., Tian, Y., Ding, Y., and Dick, W.A., 2013. Degradation of polycyclic aromatic hydrocarbons by microbial consortia enriched from three soils using two different culture media. *Environmental Pollution*, 178, 152–158.
- Wu, M., Dick, W.A., Li, W., Wang, X., Yang, Q., Wang, T., Xu, L., Zhang, M., and Chen, L., 2016. Bioaugmentation and biostimulation of hydrocarbon degradation and the microbial community in a petroleum-contaminated soil. *International Biodeterioration and Biodegradation*, 107, 158–164.

- Wu, M., Li, W., Dick, W.A., Ye, X., Chen, K., Kost, D., and Chen, L., 2017. Bioremediation of hydrocarbon degradation in a petroleum-contaminated soil and microbial population and activity determination. *Chemosphere*, 169.
- Wu, X., Zhang, C., Orita, I., Imanaka, T., Fukui, T., and Xing, X.H., 2013. Thermostable alcohol dehydrogenase from *Thermococcus kodakarensis* KOD1 for enantioselective bioconversion of aromatic secondary alcohols. *Applied and Environmental Microbiology*, 79 (7), 2209–2217.
- Xu, X., Tian, S., Jiang, P., Li, H., Wang, W., Gao, X., Qi, Q., Liu, W., Li, F., and Yu, H., 2018. Petroleum Hydrocarbon-Degrading Bacteria for the Remediation of Oil Pollution Under Aerobic Conditions: A Perspective Analysis. *Frontiers in Microbiology*, 9 (December), 1–11.
- Xu, X., Zhai, Z., Li, H., Wang, Q., Han, X., and Yu, H., 2017. Synergetic effect of biophotocatalytic hybrid system: g-C<sub>3</sub>N<sub>4</sub> and *Acinetobacter* sp. JLS1 for enhanced degradation of C<sub>16</sub> alkane. *Chemical Engineering Journal*, 323 (June), 520–529.
- Yakimov, M.M., Timmis, K.N., and Golyshin, P.N., 2007. Obligate oil-degrading marine bacteria. *Current Opinion in Biotechnology*, 18 (3), 257–266.
- Yakushi, T., Fukunari, S., Kodama, T., Matsutani, M., Nina, S., Kataoka, N., Theeragool, G., and Matsushita, K., 2018. Role of a membrane-bound aldehyde dehydrogenase complex AldFGH in acetic acid fermentation with *Acetobacter pasteurianus* SKU1108. *Applied Microbiology and Biotechnology*, 102 (10), 4549–4561.
- Zampolli, J., Collina, E., Lasagni, M., and Di Gennaro, P., 2014. Biodegradation of variable-chain-length n-alkanes in *Rhodococcus opacus* R7 and the involvement of an alkane hydroxylase system in the metabolism. *AMB Express*, 4 (1), 1–9.
- Zhang, B., Zhang, L., and Zhang, X., 2019. Bioremediation of petroleum hydrocarbon-contaminated soil by petroleum-degrading bacteria immobilized on biochar. *RSC Advances*, 9 (60), 35304–35311.
- Zhang, X., Zhang, Q., Yan, T., Jiang, Z., Zhang, X., and Zuo, Y.Y., 2015. Quantitatively predicting bacterial adhesion using surface free energy determined with a spectrophotometric method.

*Environmental Science and Technology*, 49 (10), 6164–6171.

Zhao, D., Kumar, S., Zhou, J., Wang, R., Li, M., and Xiang, H., 2017. Isolation and complete genome sequence of *Halorientalis hydrocarbonoclasticus* sp. nov., a hydrocarbon-degrading haloarchaeon. *Extremophiles*, 21 (6), 1081–1090.

Zheng, J., Feng, J.-Q., Mu, B.-Z., Gu, J.-D., Mbadanga, S.M., and Zhou, L., 2018. Characterization of bacterial composition and diversity in a long-term petroleum contaminated soil and isolation of high-efficiency alkane-degrading strains using an improved medium. *World Journal of Microbiology and Biotechnology*, 34 (2), 1–11.



---

# Chapter 2

## ***Bacillus* species and their invaluable roles in petroleum hydrocarbon bioremediation**

---

A review article submitted for publication

Cindy Baburam, Alfred Mitema, Naser Aliye Feto\*

OMICS Research Group, Department of Biotechnology, Vaal University of Technology, Vanderbijlpark, South Africa.

\*Correspondence: naserf@vut.ac.za/anaser22@yahoo.com

### *Abstract*

Exploration, production activities as well as improper waste disposal practices have led to major contamination of both aquatic and terrestrial ecosystems. Oil spills in particular have proven to be a global problem with oil spills in the Gulf of Mexico, the coasts of Cape Town (South Africa) and Montara (CL, USA) being some of the landmarks. Efficient biodegradation of petroleum hydrocarbon (H-C) necessitates the presence of a stable biosurfactant and a microbe that can use the petroleum H-C as a sole carbon source. In light of this fact, *Bacillus* spp. have been reported to both produce stable biosurfactant and tolerate high crude oil concentration and degrade aliphatics, monoaromatics and polyaromatics, which are the major constituents of crude petroleum. Such peculiarity makes the species the major source of potential candidates for H-C degradation. However, most of the oil-spill bioremediation products are presented as a magic cocktail of microorganisms, even though *Bacillus* spp. is the central genus if not the only one. Hence, this review aims to highlight the central and significant roles played by *Bacillus* spp. in oil-spill bioremediation and the mechanisms therein. Most importantly, highlighting the central roles played by *Bacillus* spp. and the mechanisms involved will contribute to targeted development of

microbial and/or enzyme cocktail products and metabolic engineering to design next-generation oil-eaters in the future. Therefore, details of the major metabolic pathways, enzymes and biosurfactants deployed by *Bacillus* spp. during H-C biodegradation are addressed at possible depth.

**Keywords:** *Bacillus* species, biosurfactant, enzyme cocktail, hydrocarbon biodegradation, oil-spill bioremediation, petroleum.

<b>TABLE OF CONTENTS</b>	<b>PAGES</b>
2.1 Introduction .....	71
2.2 The Chemistry of Hydrocarbons in Petroleum Crude Oil.....	73
2.3 Toxicity and Fate of Petroleum Hydrocarbons in the Environment .....	74
2.4 Bioremediation .....	75
2.5 Biostimulation and Bioaugmentation Methods for Bioremediation .....	75
2.6 The Potential of Microorganisms for Hydrocarbon Bioremediation .....	77
2.7 The Mechanisms Employed by <i>Bacillus</i> spp. for Bioremediation of Hydrocarbons.....	80
2.8 Surfactants and Biosurfactant for Bioavailability of Pollutants.....	81
2.9 Bacterial Chemotaxis, Flagellar Motility and Biofilm Formation .....	84
2.10 Uptake and Trans-Membrane Transport of Hydrocarbons .....	85
2.11 Enzymatic Approach for Bioremediation of Hydrocarbons .....	86
2.12 Enzymatic Degradation of Aliphatic Hydrocarbons .....	89
2.13 Enzymatic Degradation of Aromatic Hydrocarbons .....	91
2.14 Enzyme Bioprospecting through Metagenomics .....	93
2.15 Conclusion.....	96
2.16 References .....	97

## 2.1 Introduction

Our planet suffers greatly from various pollution problems, which include water, soil and air pollutions. Petroleum hydrocarbons are amongst the most toxic compounds released into oceans and spilled onto soil, thereby making them some of the major contributing factors to water and soil pollutions. Oil is introduced into the environment through natural seepage and human activities, which include pipeline and tanker leaks and spills. For instance, the British Petroleum (BP) /Deepwater Horizon discharge in 2010 stands out as the largest marine open water hydrocarbon discharge to date. The landmark incident caused a discharge of five million barrels of oil and at least 250, 000 metric tonnes of natural gas into the Gulf of Mexico (Joye *et al.* 2016).

The most important classes of organic pollutants in the environment include the mineral oil constituents and the halogenated products of petrochemicals (Chandra *et al.* 2013). Most petroleum compounds are reported to have carcinogenic properties (Souza *et al.* 2014). These compounds have a wide distribution and high toxicity hence hydrocarbons are considered as the principle organic markers of the anthropogenic activity in the ecosystems (Koshlaf and S Ball 2017). Petroleum is a heterogeneous mixture of hydrocarbons (organic compounds containing carbon and hydrogen) that includes aliphatic (n-alkanes), alicyclic and aromatic hydrocarbons (polycyclic aromatic hydrocarbons), which are insoluble in water (Joye *et al.* 2016). There is, therefore, an increasing need to develop a green technology to remove or transform hydrocarbons to a non-toxic or less toxic version of the compound through application of hydrocarbon-degrading microbial or enzyme cocktails.

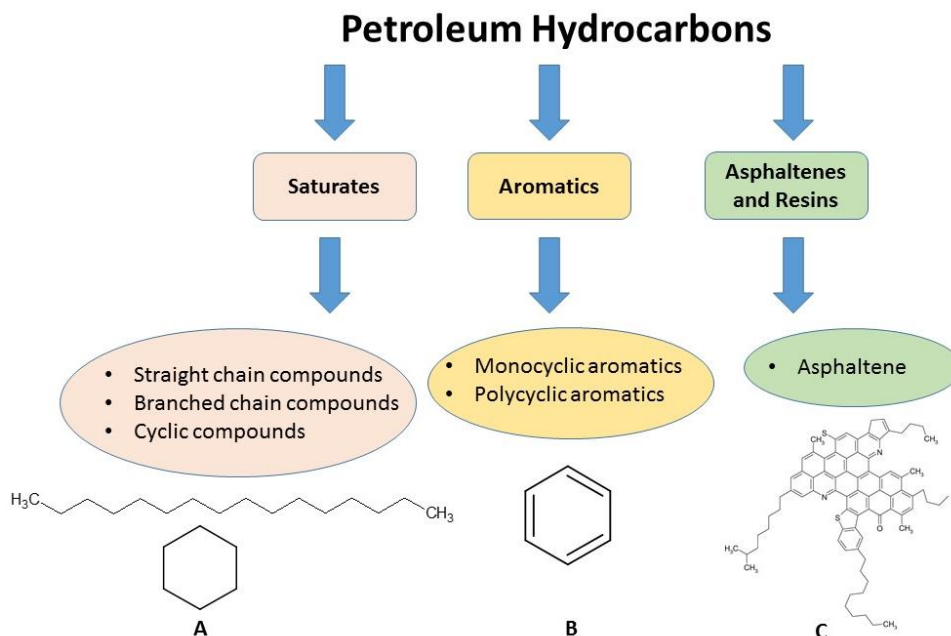
Bioremediation is a less invasive and less expensive process as compared to the classical decontamination methods (Zafra *et al.* 2016). Some microbes have been documented to inhabit and proliferate in hydrocarbon-contaminated environments, thus, using hydrocarbon as sole carbon and energy sources (Chebbi *et al.* 2017). The characterisation of certain metabolic pathways indicates that high enzymatic capacity allows the microbial communities to degrade complex hydrocarbons and this therefore promotes the importance of enzymes in the bioremediation process (Peixoto *et al.* 2011). Bioremediation can be carried out either through

bioattenuation, biostimulation and/or bioaugmentation (Adams *et al.* 2015). The classical bioremediation approach that has been in use is bioaugmentation, which is application of a consortium of *in vitro*-cultivated inocula into the polluted site. However, there are multiple factors that lessen the effectiveness of this technique. This includes biotic factors such as predation and niche-exclusion, which could in turn affect the activity of the introduced microorganisms and result in the decline of the introduced populations. Moreover, bioaugmentation poses a potential risk with regards to the addition of microorganisms, in some cases a consortium that contains genetically modified ones, into the environment, which could be difficult to monitor and contain in the natural environment. Besides, relatively the time required for the microorganism to grow and be able to produce the enzyme required to biodegrade the hydrocarbons, has negative implications on the rate of degradation (Adams *et al.* 2015).

There are a multitude of extreme environments on earth which are colonized by microorganisms called extremophiles which can thrive under diverse harsh conditions (Mirete *et al.* 2016). Particularly those microorganism with the evolved ability to grow in petroleum hydrocarbon contaminated soils and water as highlighted in this chapter. Many technologies already developed through the years for remediation, rehabilitation and restoration of contaminated environments exploit the potential of microbial biological systems to use toxic compounds as substrates for growth and convert them into harmless by-products. Hence, the vast amount of research conducted to date concentrates on the capabilities of a single or a consortia of microbes exhibiting robust and effective hydrocarbon degradative properties (Mukherjee *et al.* 2017). The *Bacillus* spp. has been put in the spot light over the years due to its increased capabilities for hydrocarbon degradation as compared to other microorganism studied to date (Parthipan *et al.* 2017). This provides promising potential for independent as well as co-culture application for bioremediation as reviewed extensively in this review article.

## 2.2 The Chemistry of Hydrocarbons in Petroleum Crude Oil

Hydrocarbons are compounds that consist exclusively of carbon and hydrogen and lack functional groups. The lack of functional groups make hydrocarbons exhibit apolar properties, low chemical reactivity at room temperature and low water solubility in long chain hydrocarbons. Hydrocarbons can be classified into the following groups: (a) aliphatic groups which include straight-chain (n-alkanes), branched-chain and cyclic compounds and (b) aromatic groups which include mono or polycyclic hydrocarbons and many important compounds which contain aliphatic hydrocarbon chains (e.g. alkylbenzenes) (Koshlaf and Ball 2017). Petroleum crude oil is a complex mixture of low and high molecular weight hydrocarbons like alkanes, aromatics, asphaltenes and resins associated with other organic compounds containing sulphur, nitrogen and oxygen (**Figure 2.1**). Among these hydrocarbons mentioned, aromatic hydrocarbons constitute a major fraction (Arun *et al.* 2011). The structural complexity of the various hydrocarbon constituents from petroleum makes the process of degradation highly difficult.



**Figure 2.1:** The different classes of hydrocarbons that comprise crude oil are (a) saturated hydrocarbon; (b) aromatics and (c) asphaltenes and resins.

### 2.3 Toxicity and Fate of Petroleum Hydrocarbons in the Environment

It is of utmost importance to expand our knowledge about the fate of hydrocarbons in the environment in order to have some level of pollution control. Due to its persistence and toxicity they cause permanent damage to the ecosystem (Atlas *et al.* 2015). The toxicity of the petroleum hydrocarbons depends on the solubility and bioavailability of the hydrocarbons. The water soluble fractions of the aromatics and polyaromatics are considered to be the most toxic (mutagenic, teratogenic and carcinogenic), though it is generally reported that most petroleum hydrocarbons have carcinogenic properties (Pandey *et al.* 2016). As a result of the reported toxicity, several hydrocarbons are classified by the Agency of Environmental Protection, the World Health Organisation and the European Union as top priority pollutants (Koshlaf and S Ball 2017). Anthropogenic hydrocarbon contamination of soil and water is a global issue throughout the industrialised world (Umeh *et al.* 2017). Anthropogenic activity in different ecosystems are based on hydrocarbons as the principle organic markers due to their wide distribution and their high toxicity (Maikudi Usman *et al.* 2016).

Soil acts as a repository for many hydrocarbons; this poses a great concern due to the adverse effects on human health and the high levels of accumulation in the environment over long periods of time (Mukherjee *et al.* 2017). Throughout history there have been major global oil spills which have had catastrophic implications on the environment (Zampolli *et al.* 2014, Joye *et al.* 2016). The Deepwater Horizon oil spill also referred to as the Gulf of Mexico oil spill in 2010 was classified as a major oil spill. The US Government estimated the total discharge at 4.9 million barrels (210 million US gal; 780,000 m<sup>3</sup>) (On Scene Coordinator Report on Deepwater Horizon Oil Spill 2011). In addition, in 2011, an Exxon Mobil pipeline spilled 63,000 gallons of crude oil into the Yellowstone River near Laurel, Montana. Following this, in January 2015 a ruptured oil pipeline leaked up to 50,000 gallons of crude oil into the Yellowstone River in Montana. Days after the spill, officials detected benzene, a cancer-causing agent, in the water supply downstream of the river (CBS news 2015). These are but a few of the major oil spills which have affected land

and water ecosystems. Therefore, there is a great need for bioremediation methods and preventative measures to be put in place to bioremediate and prevent future incidences.

## **2.4 Bioremediation**

Technologies commonly used for soil remediation of petroleum hydrocarbons include mechanical burying, evaporation, dispersion and washing. However, these measures for remediation are very costly, time consuming and not effective in the long run. Bioremediation is the preferred method since it is cost effective, efficient and eco-friendly as compared to other treatments (Varjani 2017). Bioremediation is applied after the use of physical and chemical methods and natural attenuation methods have not been successful (Roy *et al.* 2018). Bioremediation is the preferred method since it is cost effective, efficient and eco-friendly as compared to other treatments (Varjani 2017). Various species of archaea, bacteria, algae, and fungi have been reported to play significant role in bioremediation oil-spills in different environments (Sharma, Dangi, *et al.* 2018). The details of different bioremediation approaches in general and bioremediation of oil-spills in particular are discussed in the following sections.

## **2.5 Biostimulation and Bioaugmentation Methods for Bioremediation**

There are two main types of bioremediation methods, these include biostimulation and bioaugmentation (Wu *et al.* 2016). Many reports have shown that both methods have greatly enhanced the biodegradation of hydrocarbons in oil-polluted soils. The effects of both methods are case-specific and results could be inconsistent. Hence, the use of bioaugmentation and biostimulation needs to be carefully planned and studied for each type of contaminant and environmental conditions (Roy *et al.* 2018).



Biostimulation involves the modification of natural conditions of a site to increase the rates of biological degradation by stimulating indigenous microbial community endowed with catabolic pathways for hydrocarbon-degradation. The environmental conditions can be modified through the addition of surfactants, nutrients, water and chemical species acting as donors or acceptors of electrons (Zrafi-Nouira and Saidane-Mosbahi 2012). Perfumo *et al.* (2007) added nitrogen, phosphate, inorganic potassium and surfactant on the bioremediation of a ground polluted with hexadecanes. Following treatment of the ground, there was a 10% increase in the degradation. Bragg *et al.* (1994) used fertilisation of the marine environment to stimulate the natural biomass in areas of Alaska following the accident of the Exxon Valdez. Recently Wu *et al.* (2016) compared the efficiency of bioaugmentation and biostimulation approaches for hydrocarbon biodegradation. Accordingly, after six weeks of treatment they reported that 60 % and 34 % of the total petroleum hydrocarbons biodegradation, respectively were achieved.

In some cases, the indigenous microbial populations of polluted ecosystems may not contain the metabolic tools to carry out complete degradation of the pollutants (Wu *et al.* 2013). In such cases bioaugmentation can be used to overcome discrepancy by incorporating microorganisms (*in vitro*-cultivated inoculum) with adapted metabolism for hydrocarbon degradation in contaminated sites (Zrafi-Nouira and Saidane-Mosbahi 2012). For instance, Da Silva *et al.* (2005) showed an improvement of anaerobic degradation of a mixture composed of benzene, toluene, ethylbenzene, xylene and ethanol after the addition of methanogen consortia.

It is important to emphasize that the success of the implementation of bioaugmentation for example depends on competitiveness of the inoculated strains in different environments (Adams *et al.* 2015). Genetically modified organisms can also prove to be beneficial in the bioremediation process but there are some limitations to their applications such as problems with government legislation and monitoring and long-term assessment is therefore very difficult. In either case of introduction of genetically modified or wild-type strains, the potential impacts of introducing degrading microorganisms in the presence of indigenous microbes must be evaluated and monitored (Lim *et al.* 2016).

Recent studies by Roy *et al.* (2018) demonstrated an effective strategy for the bioremediation of petroleum refinery sludge using a combined biostimulation-bioaugmentation approach. This nutrient amendment to the sludge resulted in a native microbial shift that promoted hydrocarbon degrading populations within the sludge microbiome. It resulted in 57-75 % total petroleum hydrocarbon reduction in the sludge. Therefore, nutrient(s) induced community dynamics and metabolic interplay might be involved in accelerating bioremediation applications.

## **2.6 The Potential of Microorganisms for Hydrocarbon Bioremediation**

Microorganisms have the amazing ability to metabolise many organic contaminants, using them as an energy source or converting them to non-toxic products (carbon dioxide, water and biomass). Many petroleum-degrading strains are being characterised specifically for their metabolic pathways involved in biodegradation of petroleum. This natural biodegradation by microorganisms prevents the dispersion of petroleum hydrocarbons in the soil and water and minimises the pollution levels to some degree (Koshlaf and Ball 2017). Numerous microorganisms, such as bacteria, cyanobacteria, green algae and fungi are capable of either degrading or producing hydrocarbons; this completely depends on the metabolic pathways involved and the different environmental conditions (e.g. aerobic, anaerobic, varied pHs and salinities (Varjani 2017). The use of microbes and their enzymes for the removal of pollutants has proven to be effective, safe and a less expensive method. Native microbes growing in a polluted environment are expected to be more robust in degradation than non-native species since they produce certain degrading metabolites, allowing them to out-perform non-native species (Parthipan *et al.* 2017). Alternatively, the method of co-cultivation of native microbes with efficient pollutant degrading microbes is believed to be a valuable strategy to increasing contaminant removal in a short period of time (Patowary *et al.* 2016).

A large number of bacterial species have been classified as hydrocarbonoclastic bacteria and play key roles in the removal of hydrocarbons from polluted environments (Parthipan *et al.* 2017, Sharma, Dangi, *et al.* 2018). Other examples of some hydrocarbonoclastic alkane degrading

bacteria found in literature include *Bacillus* spp. (Freitas de Oliveira *et al.* 2013, Borah and Yadav 2014a, Liu *et al.* 2016, Parthipan *et al.* 2017), *Pseudomonas* spp. (Varjani and Upasani 2016) , *Streptomyces* spp. (Ferradji *et al.* 2014), *Sphingomonas* spp. (Ibrahim *et al.* 2013, Noparat *et al.* 2014) and *Staphylococcus* spp. (Shekhar *et al.* 2015). Some microorganism have the ability to degrade aliphatics, monoaromatics or polyaromatics while others are involved in the degradation of resins (Varjani 2017). Some examples of such microorganisms are listed in **Table 2.1**.

**Table 2.1:** Microorganisms involved in the degradation of specific classes of petroleum hydrocarbon compounds.

<b>Petroleum hydrocarbon</b>	<b>Microorganism responsible for biodegradation of the H-C</b>	<b>References</b>
Aliphatics	<i>Acinetobacter</i> spp.	(Mittal and Singh 2009)
	<i>Alcanivorax</i> spp.	(Brooijmans <i>et al.</i> 2009, Kadri <i>et al.</i> 2017)
	<i>Bacillus</i> spp.	(Ahmed <i>et al.</i> 2010, Darsa and Thatheyus 2014, Liu <i>et al.</i> 2016, Parthipan <i>et al.</i> 2017)
	<i>Pseudomonas</i> spp.	(Das and Mukherjee 2007, Ahmed <i>et al.</i> 2010, Sajna <i>et al.</i> 2015, Patowary <i>et al.</i> 2017)
	<i>Rhodococcus</i> spp.	(Abbasian <i>et al.</i> 2015)
Monoaromatics	<i>Sphingobacterium</i> spp.	(Noparat <i>et al.</i> 2014)
	<i>Alcanivorax</i> spp.	(Kadri <i>et al.</i> 2018)
	<i>Acinetobacter</i> spp.	(Batista <i>et al.</i> 2006)
Polyaromatics	<i>Bacillus</i> spp.	(Opere <i>et al.</i> 2013)
	<i>Sphingobacterium</i> spp.	(Noparat <i>et al.</i> 2014)
	<i>Alcanivorax</i> spp.	(Kadri <i>et al.</i> 2018)
	<i>Bacillus</i> spp.	(Mittal and Singh 2009, Borah and Yadav 2014a, 2014b)
	<i>Geobacillus</i> spp.	(Arun <i>et al.</i> 2011)
	<i>Pseudomonas</i> spp.	(Chettri <i>et al.</i> 2016, Thomas <i>et al.</i> 2016)
	<i>Vibrio</i> spp.	(Widdel and Rabus 2001)

Work carried out by Brooijmans *et al.* (2009) showed that there has been an evolution of hydrocarbonoclastic bacterial genera specifically *Oleispira*, *Marinobacter*, *Thalassolituus*, *Alcanivorax* and *Cycloclasticus* from petroleum contaminated sites. These bacteria were found to be present at low almost undetectable levels before pollution but after contamination of the area, the population started to grow and dominate. A vast amount of research has been done on one particular marine bacterium, *Alcanivorax borkumensis*, which can assimilate linear and branched alkanes but not aromatic hydrocarbons as a carbon source (Schneiker *et al.* 2006). *Alcanivorax* spp. are present in non-polluted sea water in low numbers but these strains become predominant after a crude oil spill. It is therefore believed to play an important role in natural bioremediation of oil spills worldwide (Kadri *et al.* 2018).

Although many bacteria are able to metabolise organic pollutants, a single bacterium is not capable of degrading all or even most of the compounds in polluted soils. The main reason for this stems from the fact that a single bacterium does not possess the enzymatic capability to complete such a task. Therefore, mixed microbial communities are most effective in degrading complex mixtures of hydrocarbons present in contaminated areas since the genetic information from all the microorganisms is present (Dombrowski *et al.* 2016a). This process occurs gradually by sequential metabolism of its compounds. The genes involved in this degradation process can be located either on chromosomal or plasmid DNA. Therefore, it is crucial to assess biodegradation as a multi-domain community for one to completely appreciate the metabolic potential of the indigenous microbial community (Varjani 2017).

The importance of efficient bacterial consortiums for potential hydrocarbon remediation was reinforced by Patowary *et al.* (2016). Fourteen different bacterial consortia were designed involving both biosurfactant producing and non-producing isolates. The results obtained from this study showed that a consortium comprising two *Bacillus* stains namely, *Bacillus pumilus* KS2 and *Bacillus cereus* R2 displayed the best result in the desired degradation of crude oil. This consortium showed degradation of up to 84.15% of TPH after 5 weeks of incubation using gravimetric analysis, further cementing the central and dominant role played by *Bacillus* spp. Fourier transform infrared (FTIR) and Gas chromatography-mass spectrometer (GC-MS) analyses

correlated with the gravimetric results revealed that the consortium had removed a wide range of petroleum hydrocarbons including both aliphatic and aromatic hydrocarbons. Furthermore, analysis carried out by Dombrowski et al. (Dombrowski *et al.* 2016a) through reconstruction of metabolic pathways of hydrocarbon-degrading bacteria from the Deepwater Horizon oil spill, further supports the complexity of oil-degrading community and the co-ordination of its metabolic pathways to ensure complete degradation of petroleum hydrocarbons in the environment.

The enzymatic mechanisms are also crucial to the success of microbial degradation of hydrocarbons. Different microbial electron acceptors such as oxygen, nitrate, manganese, iron and sulphate are all involved in the biotransformation of aliphatic and aromatic hydrocarbons. For example, the breakdown of n-alkanes are catalysed by alkane activating enzymes, which are monooxygenases that lower chemical reactivity of the hydrocarbon by generating reactive oxygen species. The best alkane-degrading pathway identified is encoded by the OCT plasmid of *P. putida* GPo1 (van Beilen *et al.* 2001). The first enzyme in this pathway is an integral-membrane non-haem diiron monooxygenase (AlkB) that hydroxylates alkanes at the terminal position. This produces a primary alcohol which is further oxidised by alcohol and aldehyde dehydrogenases to form the corresponding aldehyde. The product is finally converted to fatty acid via oxidation which is channeled into the  $\beta$ -oxidation pathway in the form of acetyl-CoA.

## **2.7 The Mechanisms Employed by *Bacillus* spp. for Bioremediation of Hydrocarbons**

A number of bacterial species belonging to *Bacillus* have been identified as petroleum hydrocarbon degraders and have specifically been involved in naphthalene and pyrene degradation (Annweiler *et al.* 2000). Thamer et al. (Thamer *et al.* 2013) conducted studies showing that *Bacillus thuringiensis* has great capacity for the biodegradation of crude oil. This bacteria exhibited the ability to break down hydrocarbon compounds by 80% and total biomass reaches 5g/l, while the amount of emulsion reaches 2.3 g/l. Numerous studies have supported the implementation of *Bacillus* spp. in bioremediation activities. Sakthi Priya *et al.* (2015) showed that *Bacillus subtilis* isolated from a polymer dump site in India was used for degradation of crude oil. It was noted that

the crude oil degradation and viscosity reduction was observed to be 80% and 60% respectively within a period of 10 days. It was concluded that high microbial adherence, surface tension reduction, emulsification activity, quantity of biosurfactant produced and stability provides a true indication that *Bacillus subtilis* is a potential microorganism for oil spill treatment. These results compared with other studies conducted (Borah and Yadav 2014a, Darsa and Thatheyus 2014, Jabeen *et al.* 2015, Vinothini, C., Sudhhakar, S. and Ravikumar 2015, Parthipan *et al.* 2017, Tao *et al.* 2017).

Work done by Bujang and Ibrahim, Noor Azlina (2013) demonstrated a very good biodegradation capability of oily wastewater by *Bacillus cereus* from three different automotive workshops. Oily waste water has a complex composition of hydrocarbons and it was found that the degradation was up to 91% after 5 days of incubation. This finding was further supported by studies conducted (Borah and Yadav 2014b, 2017) on *Bacillus cereus* DRDU1 which was found to efficiently degrade 96% of kerosene. The results obtained for this biosurfactants emulsification index for kerosene, crude oil and used engine oil was in a good range thus making it an attractive future application for MEOR process. Other species of *Bacillus* recently found to have hydrocarbon degradation potential include *Bacillus thermoleovorans* (Annweiler *et al.* 2000) *Bacillus amyloliquefaciens* An6 (Ayed *et al.* 2015), *Bacillus licheniformis* Y-1 (Liu *et al.* 2016), *Bacillus pumilus* (Patowary *et al.* 2016) and *Bacillus methylotrophicus* USTBa (Chandankere *et al.* 2014).

## **2.8 Surfactants and Biosurfactant for Bioavailability of Pollutants**

Surfactants are surface active substances which consists of hydrophilic (polar) and hydrophobic (nonpolar) portions on its molecule (Mulligan 2005). The polar part of the biosurfactant can be an amino acid, a carbohydrate, and/or a phosphate group. The long chain fatty acid makes up the nonpolar portion. Synthetic surfactants are widely used to treat oil spills by dispersing it and accelerating its mineralisation (Fernandes *et al.* 2016). The mechanism is the ability to reduce surface and interface tensions between liquids, solids and gases, allowing hydrocarbons to readily

emulsify in water by forming aggregates called micelles (Maikudi Usman *et al.* 2016). Knowing the number of bacteria involved with biosurfactant production will enhance hydrocarbon degradation of hydrocarbons in two ways, the first is a direct method which favors the solubility of hydrophobic compounds and secondly by indirectly increasing the availability of hydrophobic compounds to the native microorganisms (Noparat *et al.* 2014). This bioavailability of petroleum crude pollutants to microbes is crucial for the process of bioremediation (Varjani and Upasani 2017a).

The demand for surfactants for many years has been met by those derived from petroleum which proved to be toxic to the environment and non-biodegradable. This has therefore led to using biosurfactants which serve the same purpose but are produced extracellularly or as part of the cell membrane by bacteria, fungi and yeast and utilise different substrates like simple sugars, oils and hydrocarbons from contaminated environments. Biosurfactants have advantages over its chemical counterparts with regards to its effective response to extreme environmental conditions, biodegradability, lower toxicity and can be produced from cheap organic sources, which facilitates commercialisation (DÃ-az De Rienzo *et al.* 2016). Biosurfactants are produced extracellularly by a few microorganisms including bacteria, fungi and yeast. *Bacillus* spp. (Bezza and Chirwa 2015), *Streptomyces* spp. (Ferradji *et al.* 2014), *Sphingobacterium* spp. (Noparat *et al.* 2014) are but a few examples of biosurfactant producing bacteria. The addition of biosurfactant producing bacteria such as those mentioned above, can resolve unavailability of hydrocarbons which is hindered by the pollutants lack of solubility (Varjani and Upasani 2016, 2017b).

The quality, quantity and type of biosurfactant produced are influenced by the concentration of carbon, nitrogen, iron and phosphorous irons sources. Environmental factors such as growth conditions (pH, agitation, temperature and oxygen accessibility) are also a major contributor and are also valuable in assessing biosurfactant production (Maikudi Usman *et al.* 2016). This was clearly demonstrated by Khan *et al.* (Ali Khan *et al.* 2017) with regards to the role of nutrients in bacterial biosurfactant production by four bacterial strains (*Pseudomonas poae* BA1, *Acinetobacter bouveti* BP18, *Bacillus thuringiensis* BG3 and *Stenotrophomonas rhizophila*

BG32). Further, pH and temperature were shown to influence biosurfactant production by *Pseudomonas* spp. as concluded by Müller *et al.* (Müller *et al.* 2011). Kinetics of nutrient enhanced crude oil degradation was also carried out by Chettri *et al.* (Chettri *et al.* 2016) for *Bacillus* spp. AKS2 isolated from a refinery in Guwahati and further supports work done by Khan *et al.* (Ali Khan *et al.* 2017). The results showed a seven fold increase in biodegradation due to nutrient enhancement.

Many recent reports have shown the application of biosurfactant producing microbes (Bezza and Chirwa 2015, DÃ-az De Rienzo *et al.* 2016, Joy *et al.* 2017). The remediation of petroleum contaminated soils and water in the presence of microorganisms can be enhanced by the production of biosurfactants (Ferradji *et al.* 2014). It is therefore important to characterise biosurfactant producing bacteria. A vast amount of research have concluded that the *Bacillus* spp. are potential biosurfactant producers (Chandankere *et al.* 2013, Ferradji *et al.* 2014, DÃ-az De Rienzo *et al.* 2016, Maikudi Usman *et al.* 2016, Parthipan *et al.* 2017), biodegrading microbes and have been extensively used in microbial enhanced oil recovery (MEOR) (Gudiña *et al.* 2013, Youssef *et al.* 2013, Al-Wahaibi *et al.* 2014), bioremediation purposes (Bezza and Chirwa 2015, Liu *et al.* 2016) and biodegradation (Bujang and Ibrahim, Noor Azlina 2013, Thamer *et al.* 2013, Darsa and Thatheyus 2014).

There are pathways for biosurfactants synthesis that are well described by many studies for the *Bacillus* genera for the production of surfactin linked to the following genes (spf, srfAA, srfAB, srfAC and srfAD among others). In particular, two novel surfactin molecules isolated from cell-free cultures of *Bacillus subtilis* HSO121 were presented by Liu *et al.* (2016) . Surfactins have the ability to form sphere-like micelles, this amphipathic and surface properties play role in the minor polar and major hydrophobic domains of a surfactin molecule (Liu *et al.* 2015). Currently, a vast amount of research has been focused on the discovery of a novel hydrocarbon degrading biosurfactant that would enhance biodegradation in hydrocarbon polluted environments. Ayed *et al.* (2015) had successfully produced a biosurfactant from a bacterial strain *Bacillus amyloliquefaciens* An6. They reported that this biosurfactant displayed good stability over a wide



range of temperatures, pH and salinity. The solubility of diesel was enhanced by up to 10% as compared to SDS or Tween 80 respectively.

Comparative studies have also been carried out between *B. thailandensis* E264 and *P. aeruginosa* ATCC 9027 with regards to rhamnolipid biosurfactant production (DÃ-az De Rienzo *et al.* 2016). Furthermore, Chandankere *et al.* (2014) carried out work to show the properties and characterise the biosurfactant in crude oil biodegradation by *Bacillus methylotrophicus* USTBa. The biosurfactant produced during the course of hydrocarbon degradation in this study was monitored by surface tension and cell hydrophobicity measurements. The results showed the ability of this biosurfactant to decrease the surface tension of water from 72 to 28 mN/m, with a critical micelle concentration of 35 mg/L. It exhibited 90% emulsification activity on the crude oil. The combination of these results made this biosurfactant an appropriate candidate for bioremediation of crude oil contaminants.

## **2.9 Bacterial Chemotaxis, Flagellar Motility and Biofilm Formation**

The exposure of bacteria to pollutants such as hydrocarbon contaminants induces chemoattraction or chemorepellent reactions. Recent research suggests that the capacity to degrade harmful compounds have co-evolved in bacteria by reacting chemotactically to the pollutant therefore increasing their bioavailability which directly enhances biodegradation rates (Krell *et al.* 2013). Biosurfactant synthesis can be related to chemotaxis, flagellar motility and biofilm production (Chrzanowski *et al.* 2012). Structures used by microorganisms for locomotion such as flagella and pili have important roles in cell motility and chemotaxis. This helps the bacteria to move to relatively better niches as well as to attach to surfaces so as to start the biofilm formation through the complex cell-to-cell signaling. In hydrocarbon polluted niches, the process of alkane degradation will take place in the oil-water interfaces. This makes more carbon available for growth, hence increasing cell density leading to the biosynthesis of biosurfactants which can disperse hydrophobic compounds and promote their bioavailability (Nie *et al.* 2012).

Research recently carried out by Vasconcellos *et al.* (2017) in which 16 genes associated with chemotaxis and flagellar motility in prokaryotes were found in fosmid clones from a constructed fosmid library using metagenomic DNA from a petroleum reservoir in Northeast Brazil. Thus, showing the link between hydrocarbon degradation and chemotaxis. Anthracene degradation studies carried out by Das *et al.* (2017) using *Bacillus Cereus* strain JMG-01 isolated from hydrocarbon contaminated soils showed the presence of anthracene and the process of degradation lead to the modification of the cell surface morphology and the formation of an exopolymeric matrix. The filamentous growth of the biomass in the form of biofilm reveals the chemotaxis behaviour of this *Bacillus* strain in enhanced anthracene degradation.

Further research by Dombrowski *et al.* (2016a) showed the presence of major genes for motility and for use of scarce nutrients which could suggest that microorganisms are well adapted for chemotactic motility towards a hydrocarbon substrate. These genes might be of great importance for both the survival and growth of these oil-degrading organisms found in the Gulf of Mexico.

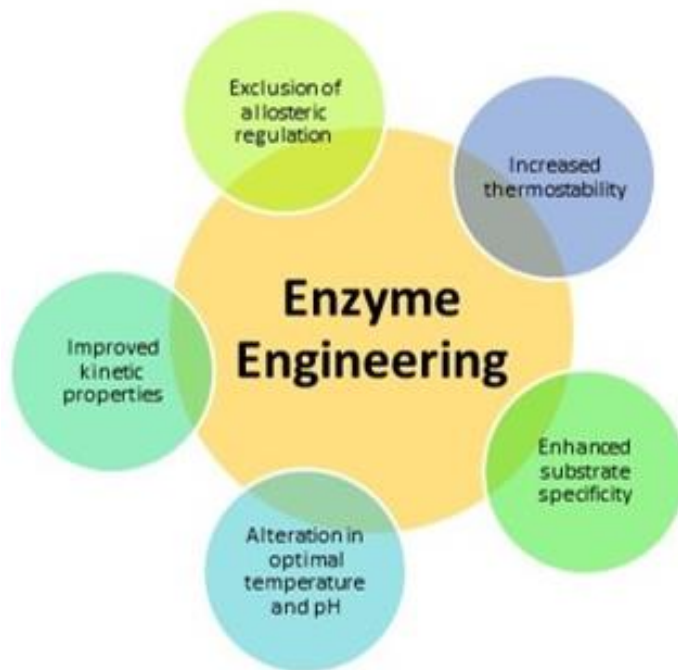
## **2.10 Uptake and Trans-Membrane Transport of Hydrocarbons**

The uptake of hydrocarbons have been the subject of many studies through the years. In particular, biosurfactants discussed previously has received a lot of the lime light. However, the transport mechanisms by which hydrocarbons cross the cell membrane have received little attention. The membrane of bacterial cells is hydrophobic. This causes an issue with biodegradation due to the decreased availability of hydrocarbons for uptake by the cells. Most substrates need to undergo cellular attachment to become accessible. Bacteria need to have access to the substrate hence they need to be either dissolved in the aqueous phase or the bacteria have to directly adhere to the hydrocarbon. Hydrocarbons that are adsorbed on the surface of the cell are transported across the membrane into the interior of the microorganism. These hydrocarbons will be degraded in the presence of enzymes which is a quick response (Hua and Wang 2014).

Following the physical interaction between bacterial cells and hydrocarbons, the next step is trans-membrane transport. Microbial cells have resources to physically access soluble, emulsified hydrocarbons and large oil droplets and transport these substrates across cell membranes and form inclusions before the hydrocarbons are metabolised (Alvarez *et al.* 1997, Bouchez-Naïtali and Vandecasteele 2008). Studies have been conducted on trans-membrane transport of phenanthrene, naphthalene and *n*-hexadecane (Kim *et al.* 2002, Kallimanis *et al.* 2007). For Gram-negative bacteria, outer membrane proteins have been reported to be involved in the transport of hydrocarbons across the cell membrane. Majority of the research has been carried out on *E. coli* with respect to long chain fatty acid transporter, FadL (Wiener and Horanyi 2011). This mechanism of trans-membrane transport is still poorly understood in and needs to be further investigated to close the knowledge gap. However, it can be concluded that it is a huge player in the success for the ability of microorganisms to degrade petroleum hydrocarbons.

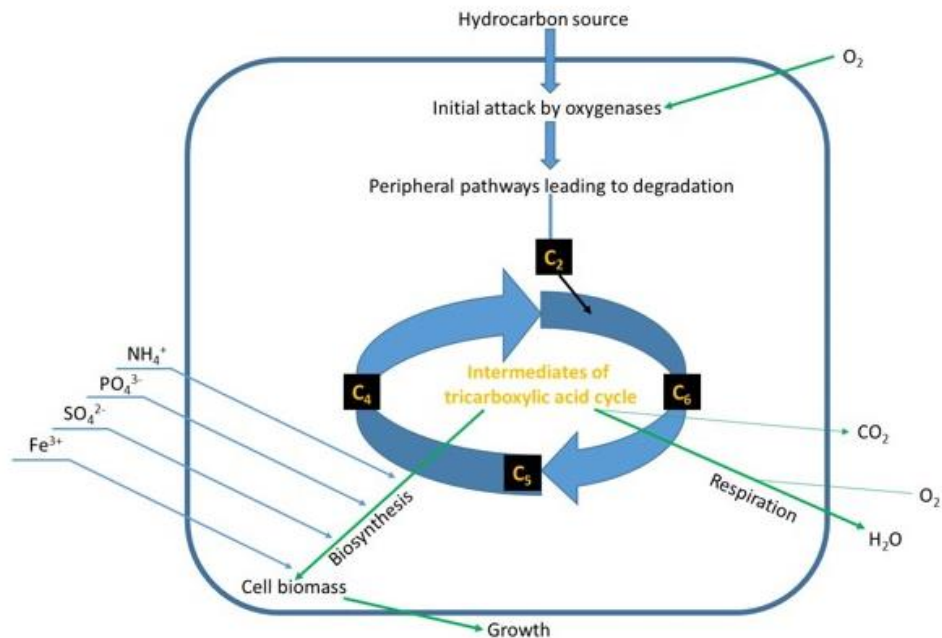
### **2.11 Enzymatic Approach for Bioremediation of Hydrocarbons**

Enzymes are complex biological molecules with the sole purpose of catalyzing a number of biochemical reactions involved in the degradation of hydrocarbons (Kalogerakis *et al.* 2017). Enzyme engineering has also been employed to improve the catalytic activity of isolated enzymes in different environments (**Figure 1.1****Figure 2.2**). This is done by changing or modifying the basic amino acid structure of the enzymes (Sharma, Dangi, *et al.* 2018).



**Figure 2.2:** Enzyme engineering to improve enzymatic properties for bioremediation.

Biodegradation of hydrocarbons including both aliphatic and aromatic compounds may occur under aerobic or anaerobic condition (Koshlaf and Ball 2017). When considering aerobic conditions, oxygenase enzymes introduce oxygen atoms into hydrocarbons (monooxygenases introduce one oxygen atom to a substrate while dioxygenases introduces two (Parthipan *et al.* 2017). The most rapid and complete degradation of the majority of pollutants is brought about under aerobic conditions. **Figure 2.3** shows the processes involved during the microbial degradation of hydrocarbons as a sole carbon source under aerobic conditions. Anaerobic degradation is catalysed by anaerobic bacteria; these include sulphate-reducing bacteria that use terminal electron acceptors (Van Hamme *et al.* 2003). The catabolism of hydrocarbons under aerobic conditions is faster than under anaerobic conditions due to the availability of oxygen as an electron acceptor (Cao *et al.* 2009).



**Figure 2.3:** Processes involved during the microbial degradation of hydrocarbons as a sole carbon source under aerobic conditions modified from (Fritsche and Hofrichter 2008).

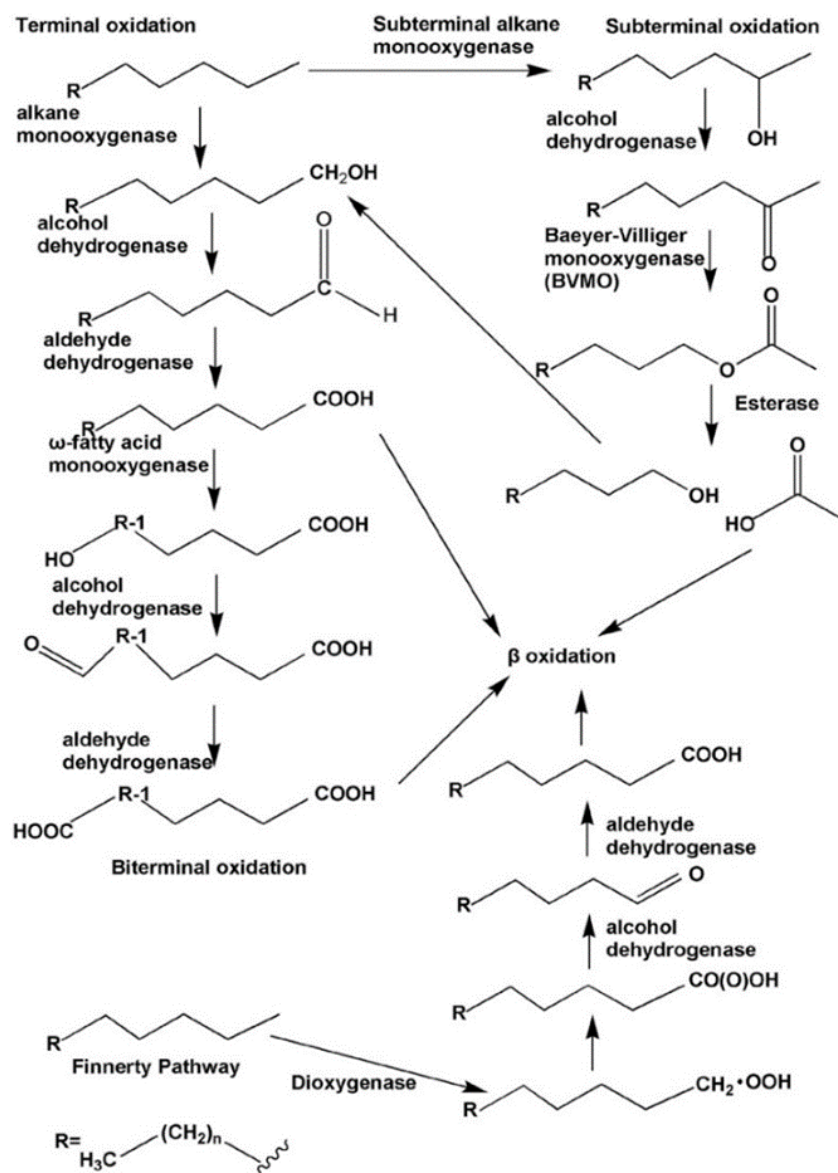
There are several enzymes involved in the hydrocarbon bioremediation process which have been identified. These include oxidoreductases (laccases, oxygenases and peroxidases) and hydrolases that are depicted in **Table 2.2** with their catalytic function (Kadri *et al.* 2017).

**Table 2.2:** Enzymes involved in the bioremediation of hydrocarbons and their functions.

Enzyme family	Examples	Functions	Reference
Oxidoreductases	Oxygenases	Catalyses the cleavage of the ring in aromatic compounds by adding one or two molecules of oxygen.	(Thomas <i>et al.</i> 2016, Sharma, Tiwari, <i>et al.</i> 2018)
	Laccases	Catalyses the oxidation of phenolic and aromatic compounds.	(Shekher <i>et al.</i> 2011)
	Peroxidases	Catalyses reduction reactions in the presence of hydrogen peroxide and generate reactive free radicals after oxidation of organic compounds.	(Sharma, Dangi, <i>et al.</i> 2018)
Hydrolases	Haloalkane	Catalyses the biodegradation of halogenated aliphatic compounds.	(Nagata <i>et al.</i> 2015)
	Dehalogenase		

## 2.12 Enzymatic Degradation of Aliphatic Hydrocarbons

When saturated aliphatic hydrocarbons are oxidised, the final product is acetyl-CoA which is catabolised in the citric acid cycle as well as the production of electrons in the electron transport chain. This chain of events is repeated to further degrade the aliphatic hydrocarbons which are then oxidised to carbon dioxide (Peixoto *et al.* 2011). The *n*-Alkanes are the main constituents of mineral oil contamination (Varjani 2017). Alkane metabolism is not a straightforward method since they are very hydrophobic, and less soluble in water (Rojo 2009). This poses a great challenge with regards to their uptake and hence facilitates their accumulation in the cytoplasmic membrane. Such accumulation could be detrimental as it might alter the membranes' fluidity. Despite such problems, many microorganisms have acquired the ability to degrade alkanes and use them as a carbon source (Wentzel *et al.* 2007). Four pathways have been identified for the initial attack on *n*-alkanes as shown in **Figure 2.4**.



**Figure 2.4:** Metabolic pathway for alkane degradation showing enzymes involved at different stages of degradation for both terminal oxidation and sub-terminal oxidation (Ji *et al.* 2013).

First is the terminal oxidation pathway, which has been well studied in *Geobacillus thermodenitrificans* NG80-2 (Li *et al.* 2008). In this pathway the alkanes are first attacked at their terminal methyl group to yield corresponding primary alcohols. These alcohols are further oxidised by alcohol dehydrogenases and aldehyde dehydrogenases to yield fatty acids which enter the β-oxidation pathway. The second pathway is called the biterminal oxidation, in which the

termini of the *n*-alkane undergo oxidation to the corresponding fatty acid without rupturing of the carbon chain. The fatty acid produced undergoes  $\omega$ -hydroxylation at the terminal methyl group. This yields a  $\omega$ -hydroxy fatty acid that is further converted to a dicarboxylic acid which enters  $\beta$ -oxidation (Coon 2005, Li *et al.* 2008). Subterminal oxidation has also been identified in a number of microorganisms and takes place when alkanes are oxidised at the subterminal position to form a primary alcohol and a secondary alcohol with the same chain length as the substrate.

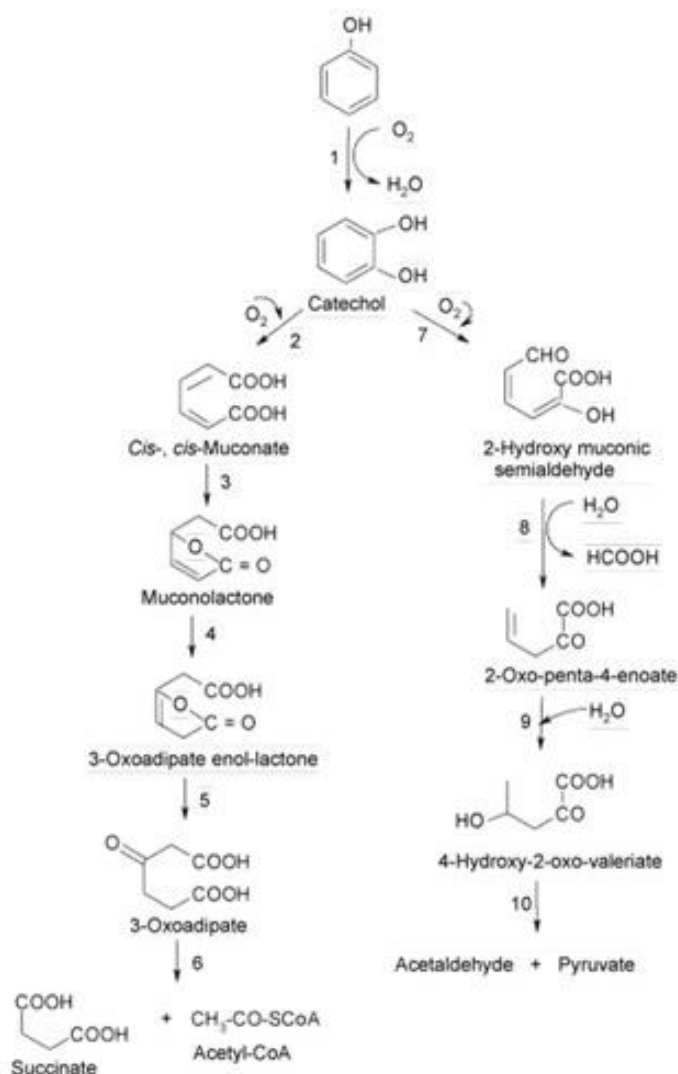
Microorganisms degrading short chain alkanes (C<sub>2</sub>-C<sub>4</sub>) have enzymes related to methane monooxygenase (Wang, Maji, *et al.* 2017). Those that degrade medium chain alkanes (C<sub>5</sub>-C<sub>11</sub>) or long chain alkanes (>C<sub>12</sub>), contain integral membrane non-heme iron monooxygenases related to a well characterised *Pseudomonas putida* GPo1 AlkB alkane hydroxylase (Rojo 2009). Several strains that assimilate alkanes greater than 18 carbon atoms contain alkane hydroxylases that are apparently unrelated to the former ones and have only recently been characterised (Wang, Nie, *et al.* 2017) Alkane hydroxylases are alkane-degrading enzymes that are distributed among many different species of bacteria, yeast, fungi and algae (Van Beilen and Funhoff 2007). The most extensively studied alkane degradation pathway is that for *Pseudomonas putida* GPo1, encoded by the OCT plasmid (van Beilen *et al.* 1994) and *Gordonia* sp. TF6 (FUJII *et al.* 2004). This pathway describes the conversion of an alkane into an alcohol using membrane monooxygenases, soluble rubredoxin and rubredoxin reductase (Van Hamme *et al.* 2003).

### **2.13 Enzymatic Degradation of Aromatic Hydrocarbons**

Cyclic alkanes represent minor components of mineral oil and have been found to be relatively resistant to microbial attack. Only a few species are capable of using cyclohexane as sole carbon source. Alkyl side chains of cycloalkanes facilitate their enzymatic degradation. With regards to aromatic hydrocarbons e.g. benzene, toluene, ethylbenzene, xylenes and naphthalene, they belong to the large volume of petrochemicals which are used as fuels and industrial solvents. Many microorganisms have evolved catabolic pathways to degrade aromatic compounds. The



compounds can be enzymatically converted to natural intermediates of the degradation: catechol and protocatechuate (Fritsche and Hofrichter 2008) (**Figure 2.5**).



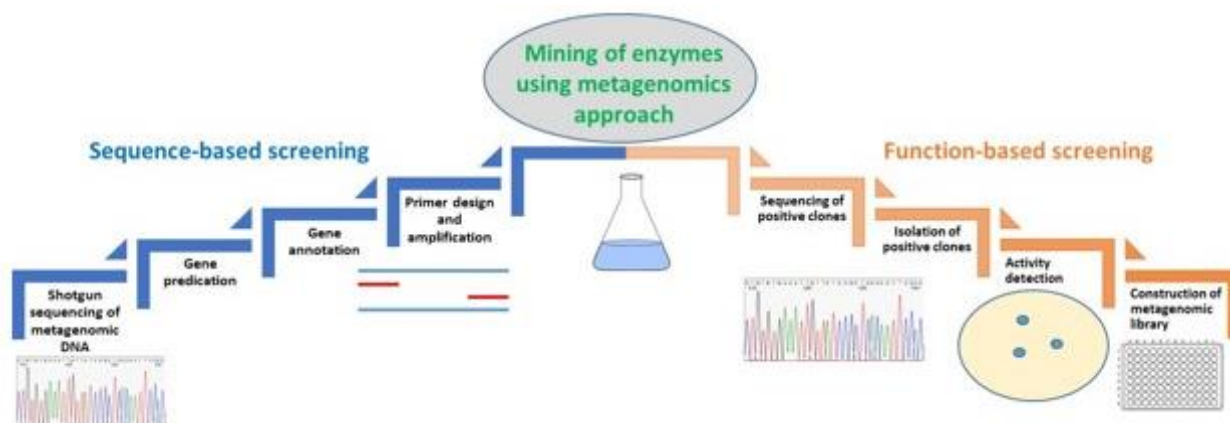
**Figure 2.5:** Aerobic degradation pathway of aromatic compounds showing *o*- and *m*-cleavage (Rojo 2009).

Benzene and related compounds are much more thermodynamically stable than aliphatics. They are very problematic since they are highly water soluble and toxic (Bello-Akinosho *et al.* 2016). Only a few reports on bacteria capable of degrading benzene have been reported but great progress has been made in this area through the years (Smith 1990, Vogt *et al.* 2011, Ren *et al.* 2015). The

class of enzymes involved in the degradation of aerobic aromatic hydrocarbons is catechol dioxygenase which are bacterial iron-containing enzymes. These enzymes are involved in aromatic ring cleavage since they are able to catalyse the addition of molecular oxygen atoms (Van Hamme *et al.* 2003).

## 2.14 Enzyme Bioprospecting through Metagenomics

Over the past few years, the field of metagenomics has proven to be a very valuable tool for accessing the biosynthetic machinery of uncultured microbiota, that constitute for at least 80-99% of the total population, as opposed to the conventional culture-based method (Lam *et al.* 2015). Novel biocatalysts and small molecule biosynthetic genes can be identified in a very short span of time as compared to the conventional molecular biology methods (Duarte *et al.* 2017). The two complimentary parallel approaches of metagenomic strategies for mining novel enzyme are as illustrated in **Figure 2.6**.



**Figure 2.6:** Schematic diagram describing the general process of metagenomic strategies for mining novel genes.

The sequence-based approach allows for the identification of candidate genes and as a result of the development of Next Generation Sequencing (NGS) technology and improved bioinformatics, has significantly advanced this methodology. The major drawback of this technology is that it does

not allow direct conclusions about the biochemical nature and functionality of the encoded proteins and furthermore it is limited to the identification of homologs of already known motifs) and reveals only partial genes which make expression and biochemical analysis of the gene products difficult (Rabausch *et al.* 2013).

The function-based approach on the other hand results in the detection of complete and functional genes from a metagenome clones library further allowing identification of novel enzymes. Metagenomic libraries are constructed using high quality metagenomic DNA, extracted using a modified DNA extraction method depending on the sample type. The DNA is sheared using mechanical or enzymatic techniques which is ligated into a linearised cloning vector such as plasmids, fosmids, cosmids or bacterial artificial chromosomes (BAC) analysed (Mirete *et al.* 2016). The resulting metagenomic library is expressed in a microbial host, mainly *Escherichia coli* and screened for desired traits on respective substrates.

Accordingly, there have been a great number of successes over the past few years with regards to the identification of various novel enzymes involved in hydrocarbon degradation using metagenomic approaches. Recently works carried out by Duarte *et al.* (2017) involving a metagenomic function-based screening approach was used to assess the microbial catabolome for polycyclic aromatic hydrocarbons from soils subjected to 12 years of *in situ* bioremediation. A total of 422 750 fosmid clones were screened for aromatic ring-cleavage activities using 2,3 dihydroxybiphenyl as a substrate. The positive fosmid clones that were sequenced showed that nearly two hundred extradiol dioxygenase encoding genes of three different superfamilies were identified. Other genes including oxygenases were also identified and provided detailed information on enzymes activating aromatic compounds, thus providing insights into the complex microbial enzyme network. This was also shown by Dombrowski *et al.* (2016a) working on reconstructing metabolic pathways of hydrocarbon-degrading bacteria from the Deepwater Horizon oil spill, where a broad gene set for degrading hydrocarbons was identified and found to work as a coordinated network for polycyclic aromatic degradation.

Moreover, in an ongoing experiment, Baburam and Feto (unpublished data), also used a function-driven metagenomic approach to identify diverse and potentially novel hydrocarbon biodegraders from petroleum contaminated soils. A fosmid library was constructed using metagenomic DNA from contaminated soil samples. Hexadecane, octadecene and cyclohexane were used to screen the library for hydrocarbon-degrading fosmid clones. The growth of clones over a period of 5 days and the size of the fosmid clones on the different substrates (hexadecane, octadecene and cyclohexane) was an indication of positive clones with enzymatic degrading capabilities for a specific hydrocarbon. Unpublished data from this research study showed a large number of clones with the ability to grow on multiple substrates. The ability of the positive fosmid clones to use the difficult-to-degrade substrate as a sole carbon source indicates the presence of the enzymatic machinery and the necessary catabolic pathways to degrade the candidate hydrocarbons. Therefore, the preliminary data shows the potential of metagenomics approach to mine potent enzymes that could be part of an enzyme cocktail with a potential for an efficient bioremediation of oil-spills. Most importantly, unravelling the role of an enzyme cocktail as an efficient oil-eater could lay solid ground for future development of next-generation oil-eaters through metabolic engineering.

## 2.15 Conclusion

Natural oil seepages and anthropogenic oil discharges continue to plague the water body including its flora and fauna. Some microorganisms have been found to be efficient in degrading hydrocarbons as they are capable of surviving in hydrocarbon contaminated waters and soils across the world. Developing a deeper understanding of the regulation and capacity for microbial hydrocarbon remediation in various environments is critical. While much has been learned over the past few decades, there is still more that needs to be understood. From the review it is apparent that data and findings observed by various researchers in the field, the central roles of the *Bacillus* spp. cannot be denied and its future uses in the environment be it natural or modified will be of great value for bioremediation advances in the future. Research further points to the benefit of mining enzymes especially novel enzymes with hydrocarbon degrading characteristics from *Bacillus* spp. As a matter of fact the majority of the biomolecules discovered to date are developed using the culture-based approach, which accounts only for 1-10% of the microbial community. However, with the onset of the metagenomics approach, which literally deal with 100% of the population, scientist have been unravelling a number of novel biocatalysts with outstanding physiochemical as well as catalytic activities. Such approach will definitely contribute to discover novel biocatalysts with a potential to turn the tide against the oil-spills across the globe. Nonetheless, the field is yet to develop, and thus, there are a number of limitations that should be addressed in the future research endeavor, from handling of huge data generated through Next Generation Sequencing (NGS) of metagenome, NGS-sequence analysis, accurate annotation, robust heterologous expression to lack of a modular high throughput (HTP) screening techniques.

## 2.16 References

- Abbasian, F., Lockington, R., Mallavarapu, M., and Naidu, R., 2015. A Comprehensive Review of Aliphatic Hydrocarbon Biodegradation by Bacteria. *Applied Biochemistry and Biotechnology*, 176 (3), 670–699.
- Adams, G.O., Fufeyin, P.T., Okoro, S.E., and Ehinomen, I., 2015. Bioremediation, Biostimulation and Bioaugmentation: A Review. *International Journal of Environmental Bioremediation & Biodegradation*, 3 (1), 28–39.
- Ahmed, A.M., Naif, A.H., and Salem, A.D., 2010. Hexadecane degradation by bacterial strains isolated from contaminated soils. *African Journal of Biotechnology*, 9 (44), 7487–7494.
- Al-Wahaibi, Y., Joshi, S., Al-Bahry, S., Elshafie, A., Al-Bemani, A., and Shibulal, B., 2014. Biosurfactant production by *Bacillus subtilis* B30 and its application in enhancing oil recovery. *Colloids and Surfaces B: Biointerfaces*, 114, 324–333.
- Ali Khan, A.H., Tanveer, S., Alia, S., Anees, M., Sultan, A., Iqbal, M., and Yousaf, S., 2017. Role of nutrients in bacterial biosurfactant production and effect of biosurfactant production on petroleum hydrocarbon biodegradation. *Ecological Engineering*, 104, 158–164.
- Alvarez, H.M., Pucci, O.H., and Steinbüchel, A., 1997. Lipid storage compounds in marine bacteria. *Applied Microbiology and Biotechnology*, 47 (2), 132–139.
- Annweiler, E., Richnow, H.H., Antranikian, G., Hebenbrock, S., Garms, C., Franke, S., Francke, W., and Michaelis, W., 2000. Naphthalene degradation and incorporation of naphthalene-derived carbon into biomass by the thermophile *Bacillus thermoleovorans*. *Applied and Environmental Microbiology*, 66 (2), 518–523.
- Arun, K., Ashok, M., Rajesh, S., Vidyapith, B., Sciences, L., and Pradesh, H., 2011. Crude oil PAH constitution , degradation pathway and associated bioremediation microflora : an overview. *International Journal of Environmental Sciences*, 1 (7), 1420–1439.
- Atlas, R.M., Stoeckel, D.M., Faith, S.A., Minard-Smith, A., Thorn, J.R., and Benotti, M.J., 2015. Oil Biodegradation and Oil-Degrading Microbial Populations in Marsh Sediments Impacted by Oil from the Deepwater Horizon Well Blowout. *Environmental Science and Technology*, 49 (14), 8356–8366.

- Ayed, H., Jemil, N., Maalej, H., Bayoudh, A., Hmidet, N., and Nasri, M., 2015. Enhancement of solubilization and biodegradation of diesel oil by biosurfactant from *Bacillus amyloliquefaciens* An6. *International Biodeterioration and Biodegradation*, 99 (April), 8–14.
- Batista, S.B., Mounteer, A.H., Amorim, F.R., and Tótola, M.R., 2006. Isolation and characterization of biosurfactant/bioemulsifier-producing bacteria from petroleum contaminated sites. *Bioresource Technology*, 97 (6), 868–875.
- Van Beilen, J.B. and Funhoff, E.G., 2007. Alkane hydroxylases involved in microbial alkane degradation. *Applied Microbiology and Biotechnology*, 74 (1), 13–21.
- van Beilen, J.B., Panke, S., Lucchini, S., Franchini, A.G., Rothlis-berger, M., and Witholt, B., 2001. Analysis of *Pseudomonas putida* alkane degradation gene clusters and flanking insertion sequences: Evolution and regulation of the alk-genes. *Microbiology (United Kingdom)*, 147 (May), 1621–1630.
- van Beilen, J.B., Wubbolts, M.G., and Witholt, B., 1994. Genetics of alkane oxidation by *Pseudomonas oleovorans*. *Biodegradation*, 5 (3–4), 161–174.
- Bello-Akinosho, M., Makofane, R., Adeleke, R., Thantsha, M., Pillay, M., and Chirima, G.J., 2016. Potential of Polycyclic Aromatic Hydrocarbon-Degrading Bacterial Isolates to Contribute to Soil Fertility. *BioMed Research International*, 2016.
- Bezza, F.A. and Chirwa, E.M.N., 2015. Production and applications of lipopeptide biosurfactant for bioremediation and oil recovery by *Bacillus subtilis* CN2. *Biochemical Engineering Journal*, 101, 168–178.
- Borah, D. and Yadav, R.N.S., 2014a. Biodegradation of complex hydrocarbon by a novel *Bacillus cereus* strain. *Journal of Environmental Science and Technology*.
- Borah, D. and Yadav, R.N.S., 2014b. Biodegradation of Diesel, Crude Oil, Kerosene and Used Engine Oil by a Newly Isolated *Bacillus cereus* Strain DRDU1 from an Automobile Engine in Liquid Culture. *Arabian Journal for Science and Engineering*, 39 (7), 5337–5345.
- Borah, D. and Yadav, R.N.S., 2017. Bioremediation of petroleum based contaminants with biosurfactant produced by a newly isolated petroleum oil degrading bacterial strain. *Egyptian Journal of Petroleum*, 26 (1).

- Bouchez-Naïtali, M. and Vandecasteele, J.P., 2008. Biosurfactants, an help in the biodegradation of hexadecane? the case of Rhodococcus and Pseudomonas strains. *World Journal of Microbiology and Biotechnology*, 24 (9), 1901–1907.
- Brooijmans, R.J.W., Pastink, M.I., and Siezen, R.J., 2009. Hydrocarbon-degrading bacteria: The oil-spill clean-up crew. *Microbial Biotechnology*, 2 (6), 587–594.
- Bujang, M. and Ibrahim, Noor Azlina, A. a/l E.R., 2013. Biodegradation of oily wastewater by pure culture of Bacillus cereus. *ARPJ Journal of Agricultural and Biological Science*, 8 (2), 108–115.
- Cao, B., Nagarajan, K., and Loh, K.C., 2009. Biodegradation of aromatic compounds: Current status and opportunities for biomolecular approaches. *Applied Microbiology and Biotechnology*, 85 (2), 207–228.
- Chandankere, R., Yao, J., Cai, M., Masakorala, K., Jain, A.K., and Choi, M.M.F., 2014. Properties and characterization of biosurfactant in crude oil biodegradation by bacterium Bacillus methylotrophicus USTBa. *Fuel*, 122, 140–148.
- Chandankere, R., Yao, J., Choi, M.M.F., Masakorala, K., and Chan, Y., 2013. An efficient biosurfactant-producing and crude-oil emulsifying bacterium Bacillus methylotrophicus USTBa isolated from petroleum reservoir. *Biochemical Engineering Journal*, 74, 46–53.
- Chandra, S., Sharma, R., Singh, K., and Sharma, A., 2013. Application of bioremediation technology in the environment contaminated with petroleum hydrocarbon. *Annals of Microbiology*, 63 (2), 417–431.
- Chebbi, A., Hentati, D., Zaghden, H., Baccar, N., Rezgui, F., Chalbi, M., Sayadi, S., and Chamkha, M., 2017. Polycyclic aromatic hydrocarbon degradation and biosurfactant production by a newly isolated Pseudomonas sp. strain from used motor oil-contaminated soil. *International Biodeterioration and Biodegradation*, 122, 128–140.
- Chettri, B., Mukherjee, A., Langpoklakpam, J.S., Chattopadhyay, D., and Singh, A.K., 2016. Kinetics of nutrient enhanced crude oil degradation by Pseudomonas aeruginosa AKS1 and Bacillus sp. AKS2 isolated from Guwahati refinery, India. *Environmental Pollution*, 216, 548–558.
- Chrzanowski, Ł., Ławniczak, Ł., and Czaczyk, K., 2012. Why do microorganisms produce



- rhamnolipids? *World Journal of Microbiology and Biotechnology*, 28 (2), 401–419.
- Coon, M.J., 2005. Omega Oxygenases: Nonheme-iron enzymes and P450 cytochromes. *Biochemical and Biophysical Research Communications*, 338 (1), 378–385.
- DÃ-az De Rienzo, M.A., Kamalanathan, I.D., and Martin, P.J., 2016. Comparative study of the production of rhamnolipid biosurfactants by *B. thailandensis* E264 and *P. aeruginosa* ATCC 9027 using foam fractionation. *Process Biochemistry*, 51 (7), 820–827.
- Darsa, K. V. and Thatheyus, A.J., 2014. Biodegradation of Petroleum Compound Using *Pseudomonas aeruginosa*. *OALib*, 01 (05), 1–9.
- Das, K. and Mukherjee, A.K., 2007. Crude petroleum-oil biodegradation efficiency of *Bacillus subtilis* and *Pseudomonas aeruginosa* strains isolated from a petroleum-oil contaminated soil from North-East India. *Bioresource Technology*, 98 (7), 1339–1345.
- Das, M., Bhattacharya, A., Banu, S., and Kotoky, J., 2017. Enhanced Biodegradation of Anthracene by *Bacillus Cereus* Strain JMG-01 Isolated from Hydrocarbon Contaminated Soils. *Soil and Sediment Contamination*, 26 (5), 510–525.
- Dombrowski, N., Donaho, J.A., Gutierrez, T., Seitz, K.W., Teske, A.P., and Baker, B.J., 2016a. Reconstructing metabolic pathways of hydrocarbon-degrading bacteria from the Deepwater Horizon oil spill. *Nature Microbiology*, 1 (May), 1–8.
- Dombrowski, N., Donaho, J.A., Gutierrez, T., Seitz, K.W., Teske, A.P., and Baker, B.J., 2016b. Reconstructing metabolic pathways of hydrocarbon-degrading bacteria from the Deepwater Horizon oil spill. *Nature Microbiology*, 1 (7).
- Duarte, M., Nielsen, A., Camarinha-Silva, A., Vilchez-Vargas, R., Bruls, T., Wos-Oxley, M.L., Jauregui, R., and Pieper, D.H., 2017. Functional soil metagenomics: elucidation of polycyclic aromatic hydrocarbon degradation potential following 12 years of in situ bioremediation. *Environmental Microbiology*.
- Fernandes, P.L., Rodrigues, E.M., Paiva, F.R., Ayupe, B.A.L., McInerney, M.J., and Tótola, M.R., 2016. Biosurfactant, solvents and polymer production by *Bacillus subtilis* RI4914 and their application for enhanced oil recovery. *Fuel*, 180, 551–557.
- Ferradji, F.Z., Mnif, S., Badis, A., Rebbani, S., Fodil, D., Eddouaouda, K., and Sayadi, S., 2014.

- Naphthalene and crude oil degradation by biosurfactant producing *Streptomyces* spp. isolated from Mitidja plain soil (North of Algeria). *International Biodeterioration and Biodegradation*, 86, 300–308.
- Freitas de Oliveira, D.W., Lima França, Í.W., Nogueira Félix, A.K., Lima Martins, J.J., Aparecida Giro, M.E., Melo, V.M.M., and Gonçalves, L.R.B., 2013. Kinetic study of biosurfactant production by *Bacillus subtilis* LAMI005 grown in clarified cashew apple juice. *Colloids and Surfaces B: Biointerfaces*, 101, 34–43.
- Fritsche, W. and Hofrichter, M., 2008. Aerobic Degradation by Microorganisms. *Biotechnology: Second, Completely Revised Edition*, 11–12, 144–167.
- FUJII, T., NARIKAWA, T., TAKEDA, K., and KATO, J., 2004. Biotransformation of Various Alkanes Using the *Escherichia coli* Expressing an Alkane Hydroxylase System from *Gordonia* sp. TF6. *Bioscience, Biotechnology, and Biochemistry*, 68 (10), 2171–2177.
- Gudiña, E.J., Pereira, J.F.B., Costa, R., Coutinho, J.A.P., Teixeira, J.A., and 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.
- Van Hamme, J.D., Singh, A., and Ward, O.P., 2003. Recent Advances in Petroleum Microbiology. *Recent Advances in Petroleum Microbiology. Microbiology and Molecular Biology Reviews*, 67 (4), 503–549.
- Hua, F. and Wang, H.Q., 2014. Uptake and trans-membrane transport of petroleum hydrocarbons by microorganisms. *Biotechnology and Biotechnological Equipment*, 28 (2), 165–175.
- Ibrahim, M.L., Ijah, U.J.J., Manga, S.B., Bilbis, L.S., and Umar, S., 2013. Production and partial characterization of biosurfactant produced by crude oil degrading bacteria. *International Biodeterioration and Biodegradation*, 81, 28–34.
- Jabeen, S., Khan, M.A.S., Hassan, Q.M., Ahmed, M.S., Nishat, N., and Zain, H., 2015. Comparative Study of Bioremediation of Crude Oil by *Bacillus Subtilis* and Organic Substances. *Journal of Sciences J. Chem. Bio. Phy. Sci. Sec. B*, 55 (33), 2621–2633.
- Ji, Y., Mao, G., Wang, Y., and Bartlam, M., 2013. Structural insights into diversity and n-alkane biodegradation mechanisms of alkane hydroxylases. *Frontiers in Microbiology*, 4 (September 2014).

- Joy, S., Rahman, P.K.S.M., and Sharma, S., 2017. Biosurfactant production and concomitant hydrocarbon degradation potentials of bacteria isolated from extreme and hydrocarbon contaminated environments. *Chemical Engineering Journal*, 317, 232–241.
- Joye, S., Kleindienst, S., Gilbert, J., Handley, K., Weisenhorn, P., Overholt, W., and Kostka, J., 2016. Responses of Microbial Communities to Hydrocarbon Exposures. *Oceanography*, 29 (3).
- Kadri, T., Magdouli, S., Rouissi, T., and Brar, S.K., 2018. Ex-situ biodegradation of petroleum hydrocarbons using *Alcanivorax borkumensis* enzymes. *Biochemical Engineering Journal*, 132, 279–287.
- Kadri, T., Rouissi, T., Kaur Brar, S., Cledon, M., Sarma, S., and Verma, M., 2017. Biodegradation of polycyclic aromatic hydrocarbons (PAHs) by fungal enzymes: A review. *Journal of Environmental Sciences (China)*, 51, 52–74.
- Kallimanis, A., Frillingos, S., Drainas, C., and Koukkou, A.I., 2007. Taxonomic identification, phenanthrene uptake activity, and membrane lipid alterations of the PAH degrading *Arthrobacter* sp. strain Sphe3. *Applied Microbiology and Biotechnology*, 76 (3), 709–717.
- Kalogerakis, N., Fava, F., and Corvini, P.F.X., 2017. Bioremediation advances. *New Biotechnology*, 38, 41–42.
- Kim, I.S., Foght, J.M., and Gray, M.R., 2002. Selective transport and accumulation of alkanes by *Rhodococcus erythropolis* S+14He. *Biotechnology and Bioengineering*, 80 (6), 650–659.
- Koshlaf, E. and S Ball, A., 2017. Soil bioremediation approaches for petroleum hydrocarbon polluted environments. *AIMS Microbiology*, 3 (1), 25–49.
- Krell, T., Lacal, J., Reyes-Darias, J.A., Jimenez-Sanchez, C., Sungthong, R., and Ortega-Calvo, J.J., 2013. Bioavailability of pollutants and chemotaxis. *Current Opinion in Biotechnology*, 24 (3), 451–456.
- Lam, K.N., Cheng, J., Engel, K., Neufeld, J.D., and Charles, T.C., 2015. Current and future resources for functional metagenomics. *Frontiers in Microbiology*, 6 (OCT), 1–8.
- Li, L., Liu, X., Yang, W., Xu, F., Wang, W., Feng, L., Bartlam, M., Wang, L., and Rao, Z., 2008. Crystal Structure of Long-Chain Alkane Monooxygenase (LadA) in Complex with

- Coenzyme FMN: Unveiling the Long-Chain Alkane Hydroxylase. *Journal of Molecular Biology*, 376 (2), 453–465.
- Lim, M.W., Lau, E. Von, and Poh, P.E., 2016. A comprehensive guide of remediation technologies for oil contaminated soil — Present works and future directions. *Marine Pollution Bulletin*, 109 (1), 14–45.
- Liu, B., Liu, J., Ju, M., Li, X., and Yu, Q., 2016. Purification and characterization of biosurfactant produced by *Bacillus licheniformis* Y-1 and its application in remediation of petroleum contaminated soil. *Marine Pollution Bulletin*, 107 (1), 46–51.
- Liu, J.F., Mbadanga, S.M., Yang, S.Z., Gu, J.D., and Mu, B.Z., 2015. Chemical structure, property and potential applications of biosurfactants produced by *Bacillus subtilis* in petroleum recovery and spill mitigation. *International Journal of Molecular Sciences*, 16 (3), 4814–4837.
- Maikudi Usman, M., Dadrasnia, A., Tzin Lim, K., Fahim Mahmud, A., and Ismail, S., 2016. Application of biosurfactants in environmental biotechnology; remediation of oil and heavy metal. *AIMS Bioengineering*, 3 (3), 289–304.
- Mirete, S., Morgante, V., and González-Pastor, J.E., 2016. Functional metagenomics of extreme environments. *Current Opinion in Biotechnology*, 38, 143–149.
- Mittal, A. and Singh, P., 2009. Isolation of hydrocarbon degrading bacteria from soils contaminated with crude oil spills. *Indian Journal of Experimental Biology*, 47 (9), 760–765.
- Mukherjee, A., Chettri, B., Langpoklakpam, J.S., Basak, P., Prasad, A., Mukherjee, A.K., Bhattacharyya, M., Singh, A.K., and Chattopadhyay, D., 2017. Bioinformatic Approaches Including Predictive Metagenomic Profiling Reveal Characteristics of Bacterial Response to Petroleum Hydrocarbon Contamination in Diverse Environments. *Scientific Reports*, 7 (1), 1–22.
- Müller, M.M., Hörmann, B., Kugel, M., Syldatk, C., and Hausmann, R., 2011. Evaluation of rhamnolipid production capacity of *Pseudomonas aeruginosa* PAO1 in comparison to the rhamnolipid over-producer strains DSM 7108 and DSM 2874. *Applied Microbiology and Biotechnology*, 89 (3), 585–592.
- Mulligan, C.N., 2005. Environmental applications for biosurfactants. *Environmental Pollution*,

133 (2), 183–198.

- Nagata, Y., Ohtsubo, Y., and Tsuda, M., 2015. Properties and biotechnological applications of natural and engineered haloalkane dehalogenases. *Applied Microbiology and Biotechnology*, 99 (23), 9865–9881.
- Nie, Y., Tang, Y.Q., Li, Y., Chi, C.Q., Cai, M., and Wu, X.L., 2012. The genome sequence of polymorphum gilvum SL003B-26A1 T reveals its genetic basis for crude oil degradation and adaptation to the saline soil. *PLoS ONE*, 7 (2).
- Noparat, P., Maneerat, S., and Saimmai, A., 2014. Application of biosurfactant from *Sphingobacterium spiritivorum* AS43 in the biodegradation of used lubricating oil. *Applied Biochemistry and Biotechnology*, 172 (8), 3949–3963.
- Opere, B.O., Obayori, O.S., and Raji, a a, 2013. Degradation of Cyclohexane and Cyclohexanone by *Bacillus lentus* Strain LP32. *African Journal of Biotechnology*, 12 (47), 6632–6635.
- Pandey, P., Pathak, H., and Dave, S., 2016. Microbial Ecology of Hydrocarbon Degradation in the Soil : A Review. *Research Journal of Environmental Toxicology*, 10 (1), 1–15.
- Parthipan, P., Preetham, E., Machuca, L.L., Rahman, P.K.S.M., Murugan, K., and Rajasekar, A., 2017. Biosurfactant and Degradative Enzymes Mediated Crude Oil Degradation by Bacterium *Bacillus subtilis* A1. *Frontiers in Microbiology*, 8 (February), 1–14.
- Patowary, K., Patowary, R., Kalita, M.C., and Deka, S., 2016. Development of an efficient bacterial consortium for the potential remediation of hydrocarbons from contaminated sites. *Frontiers in Microbiology*, 7 (JUL), 1–14.
- Patowary, K., Patowary, R., Kalita, M.C., and Deka, S., 2017. Characterization of biosurfactant produced during degradation of hydrocarbons using crude oil as sole source of carbon. *Frontiers in Microbiology*, 8 (FEB), 1–14.
- Peixoto, R.S., Vermelho, A.B., and Rosado, A.S., 2011. Petroleum-degrading enzymes: Bioremediation and new prospects. *Enzyme Research*, 2011 (1).
- Perfumo, A., Banat, I.M., Marchant, R., and Vezzulli, L., 2007. Thermally enhanced approaches for bioremediation of hydrocarbon-contaminated soils. *Chemosphere*, 66 (1), 179–184.
- R. Bragg, J., Prince, R., James Harner, E., and M. Atlas, R., 1994. *Effectiveness of bioremediation*

for the Exxon Valdez oil spill. *Nature*.

- Rabausch, U., Juergensen, J., Ilmberger, N., Böhnke, S., Fischer, S., Schubach, B., Schulte, M., and Streit, W. V., 2013. Functional screening of metagenome and genome libraries for detection of novel flavonoid-modifying enzymes. *Applied and Environmental Microbiology*, 79 (15), 4551–4563.
- Ren, Y., Yang, J., and Chen, S., 2015. The fate of a nitrobenzene-degrading bacterium in pharmaceutical wastewater treatment sludge. *Chemosphere*, 141, 13–18.
- Rojo, F., 2009. Degradation of alkanes by bacteria: Minireview. *Environmental Microbiology*, 11 (10), 2477–2490.
- Roy, A., Dutta, A., Pal, S., Gupta, A., Sarkar, J., Chatterjee, A., Saha, A., Sarkar, P., Sar, P., and Kazy, S.K., 2018. Biostimulation and bioaugmentation of native microbial community accelerated bioremediation of oil refinery sludge. *Bioresource Technology*, 253 (November 2017), 22–32.
- Sajna, K.V., Sukumaran, R.K., Gottumukkala, L.D., and Pandey, A., 2015. Crude oil biodegradation aided by biosurfactants from *Pseudozyma* sp. NII 08165 or its culture broth. *Bioresource Technology*, 191, 133–139.
- Sakthi Priya, N., Doble, M., and Sangwai, J.S., 2015. Bioremediation of costal and marine pollution due to crude oil using a microorganism *Bacillus subtilis*. *Procedia Engineering*, 116 (1), 213–220.
- Schneiker, S., Dos Santos, V.A.P.M., Bartels, D., Bekel, T., Brecht, M., Buhrmester, J., Chernikova, T.N., Denaro, R., Ferrer, M., Gertler, C., Goesmann, A., Golyshina, O. V., Kaminski, F., Khachane, A.N., Lang, S., Linke, B., McHardy, A.C., Meyer, F., Nechitaylo, T., Pühler, A., Regenhardt, D., Rupp, O., Sabirova, J.S., Selbitschka, W., Yakimov, M.M., Timmis, K.N., Vorhölter, F.J., Weidner, S., Kaiser, O., and Golyshin, P.N., 2006. Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium *Alcanivorax borkumensis*. *Nature Biotechnology*, 24 (8), 997–1004.
- Sharma, B., Dangi, A.K., and Shukla, P., 2018. Contemporary enzyme based technologies for bioremediation : A review. *Journal of Environmental Management*, 210, 10–22.
- Sharma, S., Tiwari, S., Hasan, A., Saxena, V., and Pandey, L.M., 2018. Recent advances in

- conventional and contemporary methods for remediation of heavy metal-contaminated soils. *3 Biotech*, 8 (4), 1–18.
- Shekhar, S.K., Godheja, J., and Modi, D.R., 2015. Original Research Article Hydrocarbon Bioremediation Efficiency by five Indigenous Bacterial Strains isolated from Contaminated Soils, *3 (9)*, 892–905.
- Shekher, R., Sehgal, S., Kamthania, M., and Kumar, A., 2011. Laccase : Microbial Sources , Production , Purification , and Potential Biotechnological Applications, 2011.
- Da Silva, M.L.B., Ruiz-Aguilar, G.M.L., and Alvarez, P.J.J., 2005. Enhanced anaerobic biodegradation of BTEX-ethanol mixtures in aquifer columns amended with sulfate, chelated ferric iron or nitrate. *Biodegradation*, 16 (2), 105–114.
- Smith, M.R., 1990. The biodegradation of aromatic hydrocarbons by bacteria. *Biodegradation*, 1 (2–3), 191–206.
- Souza, E.C., Vessoni-Penna, T.C., and De Souza Oliveira, R.P., 2014. Biosurfactant-enhanced hydrocarbon bioremediation: An overview. *International Biodeterioration and Biodegradation*, 89, 88–94.
- Tao, K., Liu, X., Chen, X., Hu, X., Cao, L., and Yuan, X., 2017. Biodegradation of crude oil by a defined co-culture of indigenous bacterial consortium and exogenous *Bacillus subtilis*. *Bioresource Technology*, 224, 327–332.
- Thamer, M., Al-Kubaisi, A.R., Zahraw, Z., Abdullah, H.A., Hindy, I., and Khadium, A.A., 2013. Biodegradation of Kirkuk light crude oil by *Bacillus thuringiensis*, Northern of Iraq. *Natural Science*, 5 (7), 865–873.
- Thomas, F., Lorgeoux, C., Faure, P., Billet, D., and Cébron, A., 2016. Isolation and substrate screening of polycyclic aromatic hydrocarbon degrading bacteria from soil with long history of contamination. *International Biodeterioration and Biodegradation*, 107, 1–9.
- Umeh, A.C., Duan, L., Naidu, R., and Semple, K.T., 2017. Residual hydrophobic organic contaminants in soil: Are they a barrier to risk-based approaches for managing contaminated land? *Environment International*, 98, 18–34.
- Varjani, S.J., 2017. Microbial degradation of petroleum hydrocarbons. *Bioresource Technology*,

223, 277–286.

- Varjani, S.J. and Upasani, V.N., 2016. Biodegradation of petroleum hydrocarbons by oleophilic strain of *Pseudomonas aeruginosa* NCIM 5514. *Bioresource Technology*, 222, 195–201.
- Varjani, S.J. and Upasani, V.N., 2017a. Critical review on biosurfactant analysis, purification and characterization using rhamnolipid as a model biosurfactant. *Bioresource Technology*, 232, 389–397.
- Varjani, S.J. and Upasani, V.N., 2017b. A new look on factors affecting microbial degradation of petroleum hydrocarbon pollutants. *International Biodeterioration and Biodegradation*.
- Vasconcellos, S.P., Sierra-Garcia, I.N., Dellagnezze, B.M., Vicentini, R., Midgley, D., Silva, C.C., Santos Neto, E. V., Volk, H., Hendry, P., and Oliveira, V.M., 2017. Functional and genetic characterization of hydrocarbon biodegrader and exopolymer-producing clones from a petroleum reservoir metagenomic library. *Environmental Technology*, 3330 (August), 1–12.
- Vinothini, C., Sudhhakar, S. and Ravikumar, R., 2015. Biodegradation of petroleum and crude oil by *Pseudomonas putida* and *Bacillus cereus*. *International Journal of Current Microbiology and Applied Sciences*, 4 (1), 318–329.
- Vogt, C., Kleinsteuber, S., and Richnow, H.H., 2011. Anaerobic benzene degradation by bacteria. *Microbial Biotechnology*, 4 (6), 710–724.
- Wang, V.C.C., Maji, S., Chen, P.P.Y., Lee, H.K., Yu, S.S.F., and Chan, S.I., 2017. Alkane Oxidation: Methane Monooxygenases, Related Enzymes, and Their Biomimetics. *Chemical Reviews*, 117 (13), 8574–8621.
- Wang, Y., Nie, M., Wan, Y., Tian, X., Nie, H., Zi, J., and Ma, X., 2017. Functional characterization of two alkane hydroxylases in a versatile *Pseudomonas aeruginosa* strain NY3. *Annals of Microbiology*, 67 (7), 459–468.
- Wentzel, A., Ellingsen, T.E., Kotlar, H.K., Zotchev, S.B., and Throne-Holst, M., 2007. Bacterial metabolism of long-chain n-alkanes. *Applied Microbiology and Biotechnology*, 76 (6), 1209–1221.
- Widdel, F. and Rabus, R., 2001. Anaerobic biodegradation of saturated and aromatic hydrocarbons. *Current Opinion in Biotechnology*, 12 (3), 259–276.



- Wiener, M.C. and Horanyi, P.S., 2011. How hydrophobic molecules traverse the outer membranes of Gram-negative bacteria. *Proceedings of the National Academy of Sciences*, 108 (27), 10929–10930.
- Wu, M., Chen, L., Tian, Y., Ding, Y., and Dick, W.A., 2013. Degradation of polycyclic aromatic hydrocarbons by microbial consortia enriched from three soils using two different culture media. *Environmental Pollution*, 178, 152–158.
- Wu, M., Dick, W.A., Li, W., Wang, X., Yang, Q., Wang, T., Xu, L., Zhang, M., and Chen, L., 2016. Bioaugmentation and biostimulation of hydrocarbon degradation and the microbial community in a petroleum-contaminated soil. *International Biodeterioration and Biodegradation*, 107, 158–164.
- Youssef, N., Simpson, D.R., McInerney, M.J., and Duncan, K.E., 2013. In-situ lipopeptide biosurfactant production by *Bacillus* strains correlates with improved oil recovery in two oil wells approaching their economic limit of production. *International Biodeterioration and Biodegradation*, 81, 127–132.
- Zafra, G., Taylor, T.D., Absalón, A.E., and Cortés-Espinosa, D. V., 2016. Comparative metagenomic analysis of PAH degradation in soil by a mixed microbial consortium. *Journal of Hazardous Materials*, 318 (55), 702–710.
- Zampolli, J., Collina, E., Lasagni, M., and Di Gennaro, P., 2014. Biodegradation of variable-chain-length n-alkanes in *Rhodococcus opacus* R7 and the involvement of an alkane hydroxylase system in the metabolism. *AMB Express*, 4 (1), 1–9.
- Zrafi-Nouira, I. and Saidane-Mosbahi, D., 2012. Crude Oil Metagenomics for Better Bioremediation of Contaminated Environments. *'Introduction to Enhanced Oil Recovery (EOR) Processes and Bioremediation of Oil-Contaminated Sites'*, 31.

---

# Chapter 3

## Fosmid Library Construction and Functional Screening of Fosmid Clone Libraries

---

### *Abstract*

The functional metagenomics approach is a powerful tool for studying microbial diversity as well as holistic gene functions in hydrocarbon contaminated soil. Accordingly, hydrocarbon contaminated soil samples were collected from a petrol depot on the R57 in Vanderbijlpark and the Sasol truck stop located in Sasolburg, South Africa. In order to successfully construct fosmid libraries from metagenome one has to obtain high quality DNA free of any contaminants like humic acid, traces of alcohol and salt that could interfere with downstream applications. Extracting high quality metagenomic DNA of the said standard from hydrocarbon contaminated soil was initially not possible using the conventional CTAB method. Hence, we developed an optimized version of the CTAB method. Therefore, in this chapter we discuss details of the modified method for extraction of high molecular weight and quality metagenomic DNA from hydrocarbon contaminated soils. Besides, functional screening of the fosmid libraries against target substrate is also discussed. Accordingly, optimized method yielded high quality metagenomic DNA that was successfully used to construct large insert fosmid libraries. Eighty fosmid clones were functionally screened using an agar screening method with Bushnell Haas media supplemented with three different hydrocarbon substrates (hexadecane, octadecene and cyclohexane) as the sole carbon sources. The fosmid clones were allowed to grow for five days at 37 °C. The screening process yielded fifteen candidate positive clones displaying hydrocarbon degrading potential based on the growth rate and colony size. We believe, the optimized method, which also won a best poster presentation award at an international conference in Melbourne (Australia), could be used as a

blueprint to extract high molecular weight and quality metagenomic DNA from hydrocarbon contaminated soil. In addition, the constructed fosmid library could further be screened for biomolecules of interest for developing enzyme cocktails for potential biodegradation of plastics, which are mostly of hydrocarbon sources.

**Keywords:** Metagenomics, microorganism, hydrocarbon, PacBio sequencing, bioremediation, metagenomic DNA.

<b>TABLE OF CONTENTS</b>	<b>PAGES</b>
3.1 Introduction .....	112
3.2 Materials and Methods .....	116
3.2.1 Collection of soil samples .....	116
3.2.2 Soil metagenomic DNA extraction .....	117
3.2.3 Metagenomic DNA quality and quantification .....	118
3.2.4 Purification of extracted metagenomic DNA .....	118
3.2.5 Construction of the fosmid metagenomic DNA library .....	120
3.2.6 Storage of fosmid library .....	122
3.2.7 Verification of fosmid library insert size .....	123
3.2.8 Functional screening of the fosmid library clones .....	124
3.2.9 Induction of fosmid clones and extraction of plasmid DNA .....	125
3.2.10 Assessing the quality and quantity of extracted fosmid DNA .....	126
3.3 Results and Discussion .....	127
3.3.1 Metagenomic DNA extraction from hydrocarbon contaminated soils .....	127
3.3.2 Downstream purification of extracted DNA .....	128
3.3.3 Fosmid library construction .....	130
3.3.4 Functional screening of fosmid clones .....	133
3.3.5 Sequencing of candidate fosmid clones .....	138
3.4 Conclusion .....	142
3.5 References .....	143

### 3.1 Introduction

The continuous use of petroleum as a principle source of energy and its products for industrial applications have led to the increase in hydrocarbon soil pollution globally (Xu *et al.* 2018). The factors contributing to this ever-growing environmental issue lies in accidental spillages both on land and in marine environments (Joshi *et al.* 2014). Therefore, the need for bioremediation has become a high priority strategy to detoxify and breakdown these hydrocarbon pollutants in the soil (Lu *et al.* 2019).

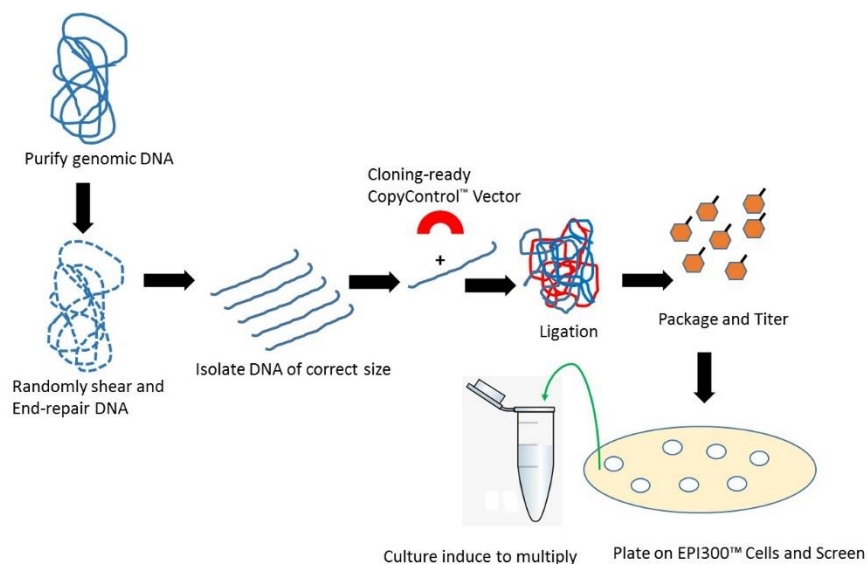
Soil is known to have a diverse naturally occurring microbial inhabitants with a variety of species of microorganisms (Verma *et al.* 2017). Those species found in contaminated soils work in consortium to break down different hydrocarbon classes, thus leading to complete detoxification of an environment over time (Zafra *et al.* 2016, Olowomofe *et al.* 2019). Soil with high hydrocarbon pollution need to be given more attention in order for these microbial communities to be exhausted in terms of understanding their function and adaptation of their evolved enzymatic pathways for hydrocarbon degradation. Metagenomic studies of such polluted sites have shown the vital role it plays in understanding and manipulating such environments (Duarte *et al.* 2017).

The metagenomics approach is a culture-dependent approach which is the alternative method to conventional microbial screening. It has proven to be invaluable to the scientific community since it allows access to whole microbial communities and to further determine the actively expressed genes of these organisms or a mixture of these without the need to culture them in the laboratory (Ezekoye *et al.*, 2018). This is important since normally, around 1% of the total soil microbial species are culturable using conventional plate culture techniques (Verma *et al.* 2017).

The value of the metagenomics approach lies in the discovery and exploitation of novel biocatalysts and other valuable biomolecules using either function or sequence based screening from unculturable microbial communities found in diverse ecosystems (Solomon *et al.* 2016). However, in order for this approach to be implemented, the availability of high quality metagenomic DNA in appreciable quantities are a pre-requisite (Solomon *et al.* 2016). This is a challenging process since the shearing of DNA during the extraction process and the co-extraction

of inhibitory compounds such as humic acids and other phenolic compounds reduce the quality of the isolated nucleic acids thus making it unsuitable for the construction of large insert metagenomic libraries (Nesme *et al.*, 2016; Singh *et al.*, 2014). This challenge has been faced by many researchers in the field based on other developed methods that suffered with regards to either poor yields or quality of metagenomic DNA. Therefore, this study focused on standardizing a metagenomic DNA extraction protocol for isolation of high-molecular weight and good quality metagenomic DNA from hydrocarbon contaminated soil samples.

The construction of a metagenomic library generated from DNA directly isolated from the environment is crucial for identifying novel biocatalysts since it provides access to the entire gene content of an environment (Ferrer *et al.* 2016). The following steps (**Figure 3.1**) are required to construct these libraries, (i) extraction of high molecular weight quality metagenomic DNA, (ii) end-repair of the isolated DNA, (iii) ligation into an appropriate vector (which could be a plasmid, cosmid or fosmid) and (iv) transformation of an appropriate host strain which is usually *E.coli* for heterologous expression (Simon and Daniel 2011). For most of the functional metagenomic studies, the genomic libraries created made use of PCC2FOS<sup>TM</sup> vector (fosmid) from Epicentre ([www.epibio.com](http://www.epibio.com)) because the fosmid vector is used for insert DNA of 25-40 kb.



**Figure 3.1:** The steps involved in the Fosmid Library Construction (Simon and Daniel 2011).

The rich genetic resource from the constructed fosmid libraries can be investigated in two ways i.e. the sequence or function driven screening. Function driven screening approach has fast become a preferred method when it comes to discovering genes with novel functions or investigating sequence diversity of protein families with certain functions (Christova *et al.* 2019). This method has been successful in identifying a variety of novel genes encoding novel enzymes (Vasconcellos *et al.* 2017, Park and Park 2018). The screening approach involves a variety of formats but screening colonies on a plate with a specific media is least labor intensive since a large number of clones can be screened using a few plates especially when it involves selecting for growth that has resulted from a specific function (Martinez *et al.* 2010).

Such screenings involve enzyme based activity, pigmentation, antibacterial and antifungal screens, antibiotic resistant screens and screens for genes that allow for the utilization of a specific substrate (Brahmi *et al.* 2019). The current study focused on a function driven screening technique that is based on the ability of fosmid library clones expressed in a heterologous host capable of using a hydrocarbon substrate as a sole carbon source that the host normally would not be able to use. This ensures that the clones selected will have the potential to degrade hydrocarbons. The (BH) Bushnell Haas media was used in this study to best assess efficient degradation of hydrocarbons. BH media is widely used as a selective media in hydrocarbon degradation studies (Borah and Yadav 2014). Numerous studies have also shown its value in identifying fosmid clones with HC degrading potential

Second-generation sequencing (SGS) technologies has been a great improvement over Sanger sequencing with its limitations of short read lengths making it poorly suited for the assembly and the determination of complex genomic regions (Ambardar *et al.* 2016). Single-molecule real-time (SMRT) sequencing, developed by Pacific BioSciences (PacBio), offers an alternative approach to overcome many of these limitations (Rhoads and Au 2015). This sequencing platform commonly referred to as PacBio sequencing, offers much longer reads and faster runs than SGS methods (Song *et al.* 2019). Such a platform can be utilised to study large genomes due to the advanced computational techniques and improvements and the long reads allows for identification of complete gene sequences (Rhoads and Au 2015).

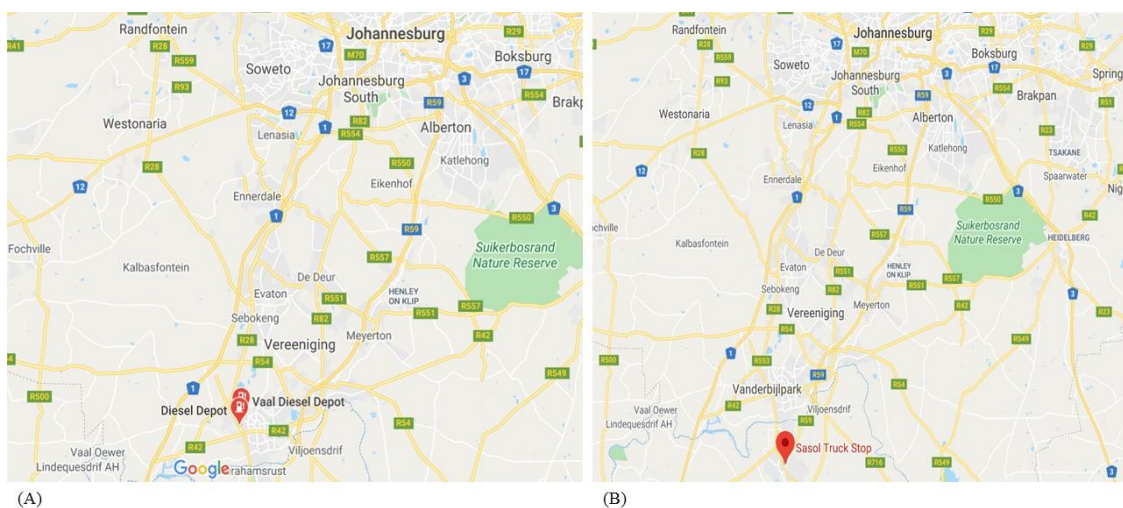
Therefore, the aim of this study was to create a high quality metagenomic library that would subsequently be screened for hydrocarbon degrading activities. The positive clones will be selected on their ability to survive on different hydrocarbon substrates and the insert DNA from selected clones will be sequenced using a PacBio sequencing platform.



## 3.2 Materials and Methods

### 3.2.1 Collection of soil samples

Different petrol, diesel and used oil contaminated soil samples were collected based on the sampling method by Singh *et al.* (2014) from the Sasol Truck stop in Sasolburg in the Freestate ( $26^{\circ}48'38.5''\text{S } 27^{\circ}50'33.5''\text{E}$ ) and the R57 Diesel Depot in Vanderbijlpark, Golden Highway ( $26^{\circ}42'04.3''\text{S } 27^{\circ}48'15.8''\text{E}$ ) as shown in **Figure 3.2**.



**Figure 3.2:** Maps showing the hydrocarbon contaminated soil sampling sites from two locations, Sasol Truck stop in Sasolburg, Freestate (A) and R57 Diesel Depot in Vanderbijlpark, Golden Highway (B).

Both locations have a regular overflow of petrol and diesel on the soil surface due to the filling of trucks and cars on a daily basis. Approximately 1kg of the soil samples were taken from a depth of 10cm using sterilised coring equipment. Soil samples were placed in sterilized polyethylene bags and placed on ice for transport to the OMICS laboratory at the Vaal University of Technology and stored at  $4^{\circ}\text{C}$  for further analysis. Samples were labeled according to their specific sampling location on site and the given names were used throughout the study (**Table 3.1**).

**Table 3.1:** Names allocated to the different soil samples used in the study.

Full sample name	Abbreviated sample name	Specific Location	GPS co-ordinates
Truck stop	TS	Sasol Truck stop in Sasolburg in the Freestate	26°48'38.5"S 27°50'33.5"E
Natref truck parking	N	Sasol Truck stop in Sasolburg in the Freestate	26°48'38.5"S 27°50'33.5"E
Tank spill	T	R57 Diesel Depot in Vanderbijlpark, Golden Highway	26°42'04.3"S 27°48'15.8"E
Ground spill	G	R57 Diesel Depot in Vanderbijlpark, Golden Highway	26°42'04.3"S 27°48'15.8"E

### 3.2.2 Soil metagenomic DNA extraction

DNA extraction from the soil samples was carried out using a modified CTAB extraction method according to Singh *et al.* (2014). Five grams of soil material from the four sites were weighed into sterile 50 mL centrifuge tubes. The soil was mixed with 13.5 ml of DNA extraction buffer (100 mM Tris-HCl pH 8.0, 100 mM EDTA pH 8.0, 100 mM sodium phosphate pH 8.0, 1.5 M NaCl, 1 % (w/v) CTAB, mercaptoethanol and 100 µl proteinase K (10 mg/mL). The tubes were incubated at 37°C for 30 minutes with constant shaking, thereafter 1.5 mL of 20% (w/v) SDS was added. The samples were incubated at 65°C for two hours with inversion of the tubes at 20 minute intervals. Samples were then centrifuged at 6,000g for 10 minutes at room temperature.

The supernatant was poured through double layers of sterile mesh gauze placed over the mouth of 50 mL centrifuge tubes to remove the fatty oil later floating on the surface. This ensured minimal contamination from the insoluble layer formed at the surface. An equal volume of

chloroform:isoamyl alcohol (24:1) was added to the supernatant and the aqueous phase recovered by centrifugation at 15,000 g for 10 minutes. This step was repeated twice followed by the addition of RNase A and incubation at 37°C for 30 minutes. The DNA was precipitated with 1/10 volume sodium acetate and 2-2.5 volumes ice cold 100 % (v/v) ethanol. Tubes were inverted to ensure complete mixing and allowed to precipitate for 2 hours at -20°C. The tubes were centrifuged at maximum speed for 20 minutes and the recovered DNA pellet was washed with 7 % (v/v) ethanol, air-dried and dissolved in 5 mL TE buffer (pH 8.0) (10mM Tris-HCl and 1mM ethylenediaminetetraacetic acid (EDTA)).

### **3.2.3 Metagenomic DNA quality and quantification**

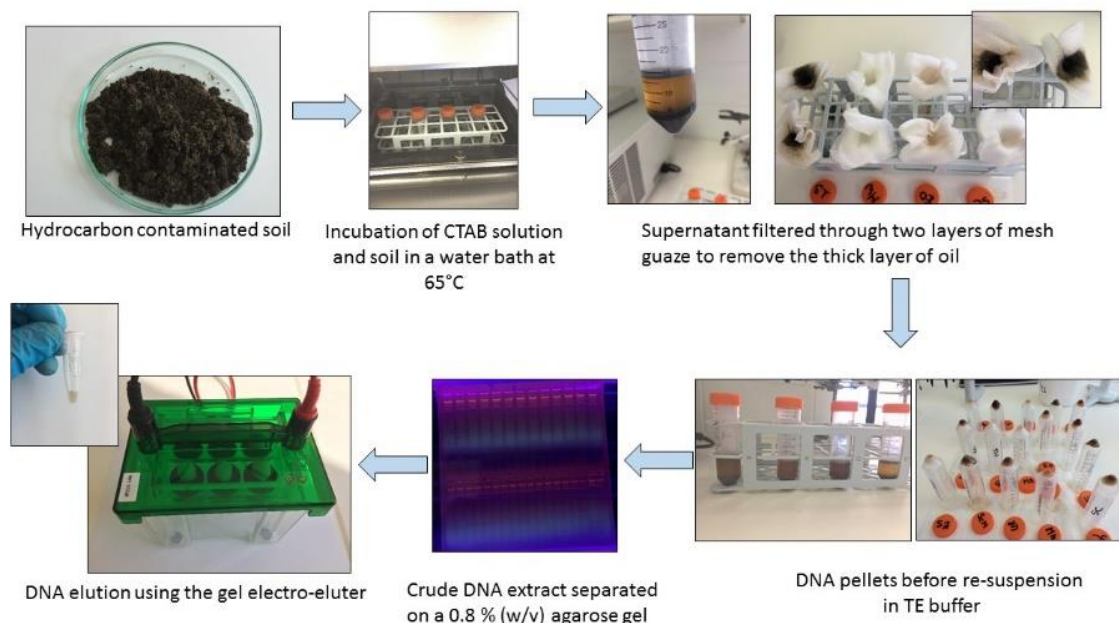
The extracted metagenomic DNA was separated on a 0.8% (w/v) agarose gel stained with ethidium bromide and run at 120 V for 45 minutes in 1X TBE buffer pH 8.0 (0.089M Tris Base, 0.089M Boric Acid, and 0.002M EDTA). A KAPA™ Biosystems Universal ladder was used to confirm the high molecular weight band that was expected. The ethidium bromide stained gel image was visualised and captured using the Gel Doc™ System (Bio-Rad, South Africa). High quality DNA was confirmed by sharp high molecular weight bands with no smears. The Nanodrop (Thermo Fisher Scientific) was used to quantify and check the quality of the extracted DNA from each location.

### **3.2.4 Purification of extracted metagenomic DNA**

High quality metagenomic DNA was required for downstream applications but the high levels of humic acid contamination of the crude DNA posed a challenge. Further purification was therefore required and the extracted DNA from each sample were separated on a 0.8% (w/v) agarose gel stained with ethidium bromide at 120 V for 45 minutes. The bands were excised from the gel placed on a UV light box using a sharp scalpel and the DNA eluted in DNA elution buffer (40 mM

Tris, 20mM glacial acetic acid, 1 mM EDTA pH 8.0 and 0.1% SDS) using the gel electro-eluter (Bio-Rad, South Africa) set at a current of 40 mA for 1 hour. Multiple bands were excised and pooled to ensure high DNA concentrations were obtained. The membrane caps were therefore used multiple times. The DNA was precipitated with 100 % (v/v) ethanol and 1/10 volume 3M ammonium acetate pH 8.0 overnight at 4 °C. The precipitated DNA was centrifuged at 4000 rpm at room temperature for 30 minutes. The pelleted DNA was washed with 70 % (v/v) ethanol and further centrifuged at 4000 rpm for 30 minutes. The DNA pellet was resuspended in 50 µl TE buffer and all volumes combined.

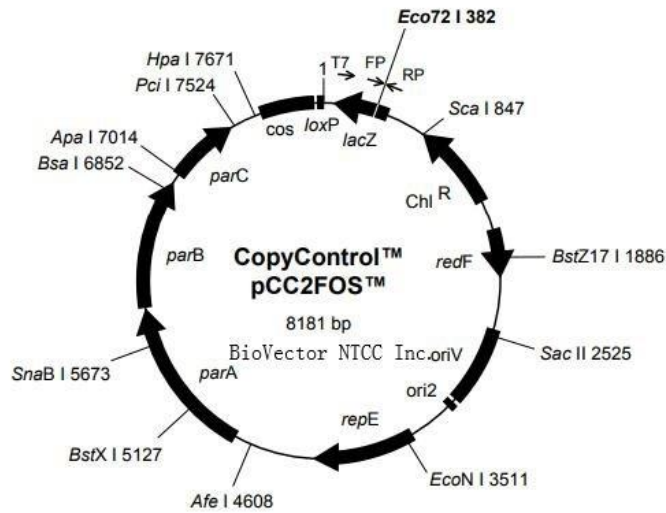
The eluted DNA from the four soil samples (N, T, G and TS) were pooled and separated on a 0.8% (w/v) agarose gel stained with ethidium bromide at 120 V for 45 minutes. A KAPA™ Biosystems Universal ladder was used to confirm the high molecular weight band that was expected. The eluted DNA was further concentrated using the Amicon® DNA concentrator filters (Merck, USA). The complete modified DNA extraction and purification protocol developed is depicted in **Figure 3.3**.



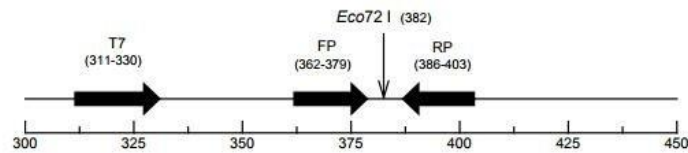
**Figure 3.3:** Work flow for the developed modified extraction method for the extraction of metagenomic DNA from hydrocarbon contaminated soils.

### 3.2.5 Construction of the fosmid metagenomic DNA library

The fosmid library was constructed using the CopyControl™ HTP Fosmid production kit (Epicentre, USA) with pCC2FOS™ vector (**Figure 3.4**) using the microbial consortium obtained from the pooled DNA as per manufacturer's instructions. The shearing of DNA into approximately 40 kb fragments was not necessary since the size was confirmed to be in this region by electrophoretic separation of the DNA and the fosmid control DNA (provided by the manufacturer) on an agarose gel. Approximately 20 µg of the insert DNA was end-repaired to generate blunt-ended 5'-phosphorylated DNA. The reaction was carried out using sterile water, 8 µl 10X End-repair buffer, 8µl 2.5 mM dNTP mix, 8µl ATP, 20µg insert DNA and 4µl End-repair enzyme mix in a total reaction volume of 80 µl. The ligation reaction of the insert DNA and the CopyControl pCC2FOS Vector was carried out in a reaction volume of 10 µl which consisted of sterile water, 1 µl 10X Fast-link ligation buffer, 1 µl 10mM ATP, 1µl CopyControl pCC2FOS vector (0.5 µg/µl), concentrated insert DNA (0.25 µg of 40 kb DNA) and 1 µl Fast-link DNA ligase. The ligation reaction was carried out at room temperature for 4 hours and the enzyme deactivated at 70 °C for 10 minutes.



Note: Not all restriction enzymes that cut only once are indicated above. See Appendix F for complete restriction information. Primers are not drawn to scale.



**Figure 3.4:** The vector map of PCC2FOS™ Fosmid vector used for construction of the fosmid metagenomic library (CopyControl™ Fosmid Library Production Kit with pCC1FOS™ Vector manual, Epicentre, USA).

The packaging of the CopyControl Fosmid clones were carried out using MaxPlax Lambda Packaging extracts (Epicentre, USA), EPI300-T1® plating strain. Fifty milliliters of Luria-Bertani (LB) broth with 10 mM MgSO<sub>4</sub> and 0.2 % (w/v) maltose was inoculated with 0.5 ml of EPI300-T1® overnight culture. The flask was shaken at 37 °C to an A<sub>600</sub> of 0.8-1.0 (± 2 hours). For every ligation reaction, a volume of 25 µl of the thawed MaxPlax Lambda Packaging extracts was used with 10 µl of the ligation reaction. The solution was mixed well by pipetting and the reaction was allowed to proceed at 30 °C for 2 hours. The remaining 25 µl of MaxPlax Lambda Packaging extracts was added to the reaction mixture and this was allowed to proceed for a further 2 hours. At the end of the second incubation, Phage dilution buffer (PDB) was added to a final volume of

1 ml in each tube and mixed gently. Twenty five microliters of chloroform was added to each tube and stored at 4 °C.

The titering of the packaged CopyControl Fosmid clones were carried out to determine the number of plates and dilutions required to obtain an extensive library. Titering was carried out by making serial dilutions of the 1 ml of packaged phage particles into PDB in sterile microcentrifuge tubes. Serial dilutions of 1:10<sup>1</sup>, 1:10<sup>2</sup> and 1:10<sup>3</sup> were prepared using the concentrated packaged phage particles. Ten microlitres from each dilution was added to 10 µl of the undiluted phage and mixed with 100 µl of the prepared EPI300-T1<sup>®</sup> host cells. This was incubated for one hour at 37 °C. The infected EPI300-T1<sup>®</sup> cells were spread on LB agar plates supplemented with chloramphenicol (12.5 µg/ml) and incubated at 37°C overnight to select for the CopyControl Fosmid clones. The colonies on the plates for each dilution was counted and the titer of the packaged phage particles were calculated using the following equation:

$$\textit{Titer} = \frac{(\textit{number of colonies})(\textit{dilution factor})(1000 \mu\textit{l/ml})}{\textit{volume of phage plated} (\mu\textit{l})}$$

The infection of cells with the remaining phage was carried out and these were plated on LB agar plates supplemented with chloramphenicol (12.5 µg/ml) and incubated at 37 °C overnight. This was carried out to ensure the fosmid library constructed would have an extensive coverage.

### **3.2.6 Storage of fosmid library**

The plates were washed with 2 ml of LB broth supplemented with chloramphenicol (12.5 µg/ml) in order to resuspend all the bacterial cells and then transferred to sterile cryotubes. Sterile glycerol was added to the resuspended liquid to a final concentration of 20 % (v/v). The solution was mixed and stored in aliquots of 1 ml at -80 °C until required. Each aliquot constitutes a library with the calculated coverage.

### 3.2.7 Verification of fosmid library insert size

The clones from the prepared fosmid library were plated onto LB agar plates supplemented with chloramphenicol (12.5 µg/ml) and grown overnight at 37 °C. A total of 80 fosmid clones were randomly selected for downstream functional screening analysis and plated onto an LB plate marked with a grid and the respective clones were named according to the grid numbers as shown in **Figure 3.5**. This was carried out by using a toothpick to select individual clones and spot onto a grid location. These clones were grown overnight at 37 °C.



**Figure 3.5:** Illustrative image of a LB agar plate with a numbered grid drawn to show the locations and identities of the candidate fosmid clones selected for functional screening analysis.

Two fosmid clones (CB1 and CB7) were selected for quality control purposes to confirm the ligation of the insert DNA in the pCC2FOS™ vector. CB1 and CB2 fosmid clones were picked from the LB agar plate grid and grown overnight at 37 °C in 50 ml of LB broth supplemented with chloramphenicol (12.5 µg/ml) and 100 µl 500X CopyControl Fosmid autoinduction solution with shaking (225-250 rpm). The use of the autoinduction ensures that the fosmid clones can be induced



to high copy numbers for high yields of DNA for downstream applications. The plasmid DNA was isolated from the clones using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) according to manufacturer's instructions. The plasmid DNA was eluted with TE buffer in a final volume of 40  $\mu$ l. The EPI300-T1<sup>®</sup> host cells were used as the negative control since it does not contain any plasmid DNA.

Five microliters of the isolated plasmid DNA was separated on a 0.8 % (w/v) agarose gel stained with ethidium bromide at 120 V for 45 minutes to determine the quality and quantity of the isolated plasmid. A KAPA<sup>™</sup> Biosystems Universal ladder was used to confirm the estimated size of the plasmid DNA. The quantity and quality of the plasmid DNA was obtained using the Nanodrop (Thermo Fisher Scientific). Restriction digests of the plasmid DNA isolated from fosmid clones CB1 and CB7 and pCC2FOS<sup>™</sup> vector (negative control) were carried out using the restriction enzyme *Xba*I (New England BioLabs<sup>®</sup>) to confirm the presence of the insert DNA and the plasmid backbone. The restriction digest was setup using 16.8  $\mu$ l sterile water, 2  $\mu$ l RE 10X Buffer D, 0.2  $\mu$ l BSA, 0.5  $\mu$ l plasmid DNA (200 ng) and 0.5  $\mu$ l *Xba*I (New England BioLabs<sup>®</sup>) in a total reaction volume of 20  $\mu$ l. The restriction digests were separated on a 0.8 % (w/v) agarose gel stained with ethidium bromide at 120 V for 45 minutes. A KAPA<sup>™</sup> Biosystems Universal ladder was used to confirm the expected sizes of the restricted DNA fragments. The fosmid control vector was used to compare with DNA fragment sizes obtained from the digests of the plasmid DNA from the candidate clones.

### **3.2.8 Functional screening of the fosmid library clones**

The 80 candidate fosmid clones obtained from the library were screened for hydrocarbon degrading potential using a modified method described by (Joy *et al.* 2017). Master plates of the fosmid clones were prepared on LB agar supplemented with chloramphenicol (12.5  $\mu$ g/ml) and gridded according to their reference name. These plates were incubated overnight at 37 °C. Fosmid clones were transferred from the master plates using sterile toothpicks onto gridded Bushnell Haas (BH) agar plates (composition  $g\ l^{-1}$ : MgSO<sub>4</sub> 0.2, CaCl<sub>2</sub> 0.02, KH<sub>2</sub>PO<sub>4</sub> 1.0, K<sub>2</sub>HPO<sub>4</sub>

1.0, NH<sub>4</sub>NO<sub>3</sub> 1.0, FeCl<sub>3</sub> 0.05, agar-agar 20.0, pH 7.0 at 25°C) supplemented with chloramphenicol (12.5 µg/ml), with and without 1 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) and 1 % (v/v) filter sterilised n-hexadecane (Sigma-Aldrich, USA), 1- octadecene (Sigma-Aldrich, USA) and cyclohexane (Sigma-Aldrich, USA) respectively as the sole carbon and energy source. The experiments were carried out in duplicates. The BH agar plates were incubated at 37 °C for up to 5 days to ensure ample time for colony growth. Candidate fosmid clones were selected based on their growth rate and the size of the colony after five days of incubation at 37 °C. Five fosmid clones were selected for each of the hydrocarbon sources and sent to Inqaba Biotec™ for PacBio SMRT sequencing.

The negative control used in this experiment was the EPI300-T1® host cells. The positive controls (S1, S2, R1 and R2 ) were obtained from a culture-based method from a previous study using the soil samples from the different locations to isolate and characterize hydrocarbon degrading bacteria growing in a M9 minimal salts media (Sigma-Aldrich) supplemented with 5 % (v/v) benzene (Sigma-Aldrich, USA), 5 % (v/v) hexadecane (Sigma-Aldrich, USA), 5 % (v/v) octadecene (Sigma-Aldrich, USA) and 5 % (v/v) cyclohexane (Sigma-Aldrich) as the sole carbon source.

### **3.2.9 Induction of fosmid clones and extraction of plasmid DNA**

An overnight culture of each fosmid clone was grown in 10 ml of Luria-Bertani (LB) broth supplemented with chloramphenicol (12.5 µg/ml). This stock culture was used to inoculate 25 ml LB broth supplemented with chloramphenicol (12.5 µg/ml). This was autoinduced by adding 50 µl 500X CopyControl™ Fosmid Autoinduction Solution as specified by the manufacturer. The autoinduced culture was grown overnight at 37 °C with shaking at 170 rpm in a shaking incubator. EPI 300™ bacterial host cells grown in 25 ml LB broth served as the negative control. The bacterial cells were collected from 10 ml of overnight culture by centrifugation at 4000 rpm for 5 minutes. The supernatant was removed and the pellet resuspended in 250 µl suspension buffer and the plasmid DNA was extracted using the GeneJET Plasmid Miniprep Kit (Thermoscientific)

according to the manufactures instructions. Isolated plasmid DNA was eluted in 50 µl Elution buffer and stored at -80 °C.

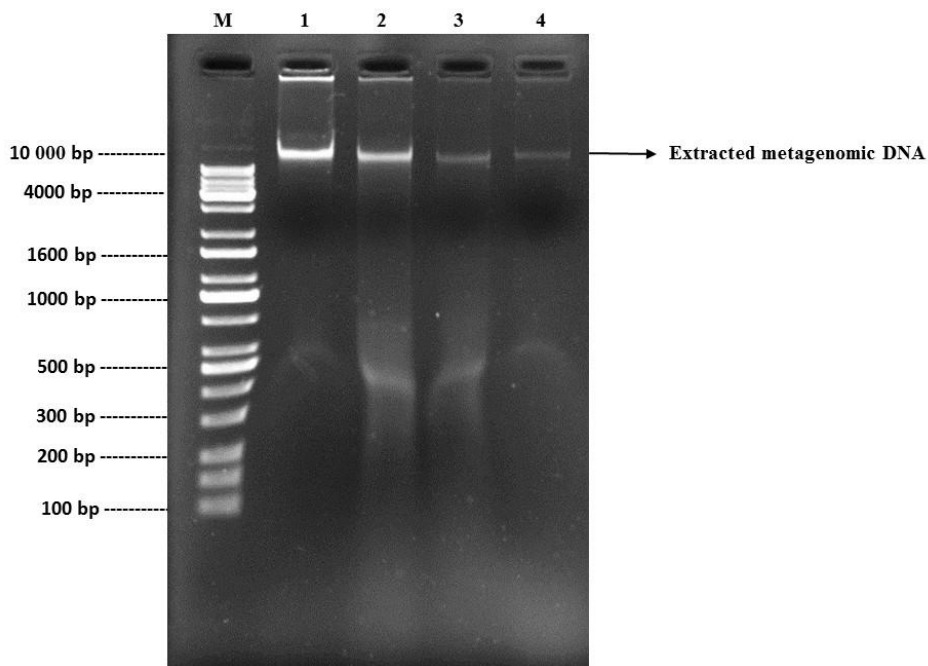
### **3.2.10 Assessing the quality and quantity of extracted fosmid DNA**

The extracted fosmid DNA for the 15 candidate fosmid clones were visualised on a 0.8 % (w/v) agarose gel stained with ethidium bromide and run at 100V. The pCC2FOS™ vector from the CopyControl™ HTP Fosmid Library Production Kit was also separated on the agarose gel and compared to fosmid clones containing the insert DNA. The plasmid DNA for the 15 fosmid clones were sent for PACBIO SMRT® sequencing at Inqaba Biotec™ in Pretoria, South Africa.

### 3.3 Results and Discussion

#### 3.3.1 Metagenomic DNA extraction from hydrocarbon contaminated soils

Metagenomic DNA was isolated from four hydrocarbon contaminated soil samples using a modified CTAB method by Singh *et al.*, 2014. A total DNA concentration of  $\pm 4 \mu\text{g/g}$  was obtained from the starting material from each soil sample used in this study. Based on the results shown in **Figure 3.6**, the size of the extracted metagenomic DNA was found to be  $\geq 11 \text{ kb}$  and corresponds to results obtained in other research (Fatima *et al.* 2014, Gastauer *et al.* 2019).

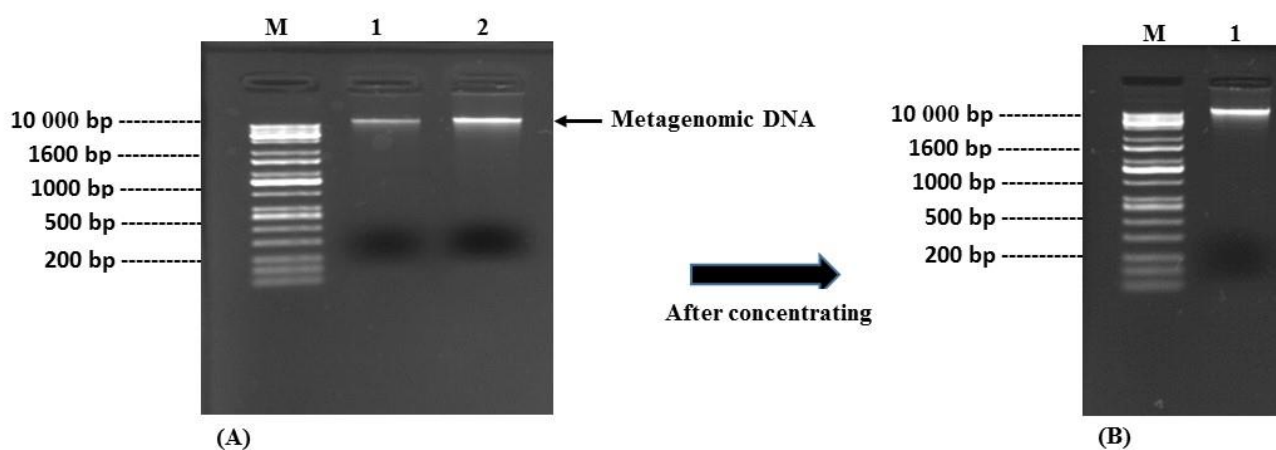


**Figure 3.6:** 0.8 % (w/v) Agarose gel electrophoresis of extracted metagenomic DNA from soil samples. M: 10 kb DNA ladder, Lane 1: Truck stop (Sasolburg), Lane 2: Natref Truck parking (Sasolburg), Lane 3: Tank spill (Vanderbijlpark) and Lane 4: Ground spill (Vanderbijlpark).

The agarose gel showed minimal smearing of the DNA in the lanes and sharp bands were observed. The  $A_{260/230}$  and  $A_{260/A280}$  ratios were in the region of 0.60 – 0.85 and 1.50 – 1.70 respectively for the different soil samples. This confirmed the presence of contaminants co-extracted with the metagenomic DNA such as humic acid, phenolic compounds, salts and protein contamination was also evident. Such findings have also been shown by other researchers working with hydrocarbon contaminated soil samples (Lee and Hallam 2009, Solomon *et al.* 2016, Tanveer *et al.* 2016). This therefore, necessitated further downstream purification steps to remove the residual contaminants.

### 3.3.2 Downstream purification of extracted DNA

The extracted metagenomic DNA from the four locations were pooled, separated on an agarose gel and extracted using agarose gel elution. The method ensured that the pooled metagenomic DNA would be removed of contaminants such as humic acid and phenolic compounds that would hinder the downstream fosmid library construction. This was confirmed using agarose gel electrophoresis in **Figure 3.7(A)** showing sharp high molecular weight bands with no smearing in the lanes.



**Figure 3.7:** Analysis of extracted metagenomic DNA from soil. (A) 0.8 % (w/v) Agarose gel electrophoresis of extracted pooled and purified metagenomic DNA. M: 10 kb DNA ladder, Lane 1: 2 µl metagenomic DNA and Lane 2: 4 µl metagenomic DNA. (B) 0.8 % (w/v) Agarose gel

electrophoresis of concentrated, pooled and purified metagenomic DNA for library construction. M: 10 kb DNA ladder and Lane 1: 2 µl of concentrated pooled DNA using the Amicon® DNA concentrator filters (Merck, USA).

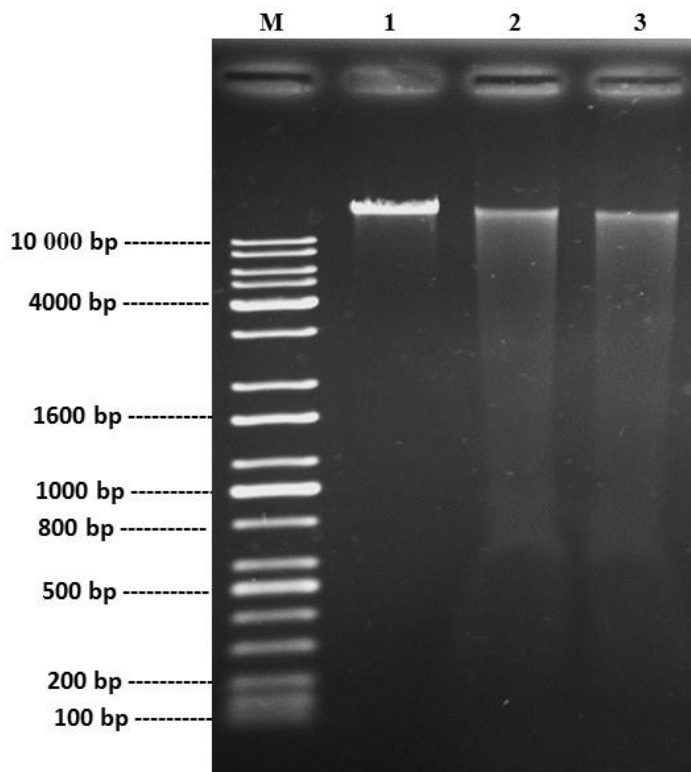
This was further confirmed using the Nanodrop, the  $A_{260/230}$  and  $A_{260/A280}$  ratios were in the region of 1.20 – 1.60 and 1.70 – 1.85 respectively. These values indicate higher quality DNA with reduced contaminants as originally observed for the crude extractions. The concentration of DNA was found to be 50 ng/µl in a total eluted volume of 500 µl. The DNA was further concentrated using the Amicon DNA concentrator kit (Merck, USA) and separated on an agarose gel as shown in **Figure 3.7(B)**.

The isolation of metagenomic DNA requires to be of the highest quality as this is crucial for the application of the metagenomic approach in this study which involved preparing a fosmid library for functional screening. The method described by Singh *et al.* (2014) for the isolation of metagenomic DNA from soil was not sufficient to remove co-extracted contaminants such as humic acid and other phenolic compounds from the extracted DNA. Therefore, further purification steps involving filtration and agarose gel extractions were implemented to ensure that the quality of DNA was up to standard for the preparation of the fosmid library. The isolated DNA was therefore shown to be free of contaminants that could interfere with enzymatic reactions involved in constructing metagenomic libraries (Hassan *et al.* 2018). The metagenomic DNA was therefore suitable for use in constructing large insert metagenomic libraries.

This result is conclusive in demonstrating the value of the modified extraction method developed in this study to obtain high quality DNA in appreciable amounts. Many methods have been used in the past but posed problems of high contamination from co-extracted compounds (Singh *et al.*, 2014; Verma *et al.*, 2017). Posing a challenge to many researchers in the field since it is the primary step in all meaningful metagenomic studies aiming to characterise and harness the benefits of microbial communities. This method can therefore be recommended, verified and adopted for metagenomic DNA extraction based on the current findings.

### 3.3.3 Fosmid library construction

In order to construct the metagenomic fosmid library, it was essential to determine the size of the metagenomic DNA extracted from the soil samples. The fosmid control DNA (40 kb) supplied by the manufacturer and the extracted metagenomic DNA was separated on an agarose gel to compare the sizes as shown in **Figure 3.8**. The size of the metagenomic DNA was found to be within 35-40 kb in range when compared to the fosmid control DNA of 40 kb. Since the size of the DNA was within the range described for use in the CopyControl™ HTP Fosmid production kit, the shearing of DNA was not necessary. The DNA was therefore of a good size range to use in the library production.



**Figure 3.8:** 0.8 % (w/v) Agarose gel electrophoresis to determine the size of the isolated metagenomic DNA from pooled samples. M: 10 kb DNA ladder, Lane 1: Fosmid control DNA (40 kb), Lane 2 and 3: 2  $\mu$ l of purified metagenomic DNA.

The fosmid library was constructed using the CopyControl™ HTP Fosmid production kit with pCC2FOS™ vector. This resulted in a library size of approximately 500 000 colony forming units (cfu) as calculated using the titer package phage equation described in the materials and methods section of this chapter (3.2.5) This is indicative of a rich genomic library ensuring for increased probability of identifying genes of interest. This is in keeping with similar work carried out by (Silva *et al.* 2013) in which a total of 13 200 clones were obtained from a constructed fosmid library obtained from a bioreactor sludge of petroleum refinery wastewater treatment system. The results yielded 400 positive clones displaying phenol degradation following functional screening analysis. The number of colony forming units on the plate was found to increase by increasing the time of infection of bacteria with phage for 24 hours and thereafter concentrating the phage. The colonies obtained were found to be of a good size, number and high growth rate for different dilution factors as shown in **Figure 3.9**.

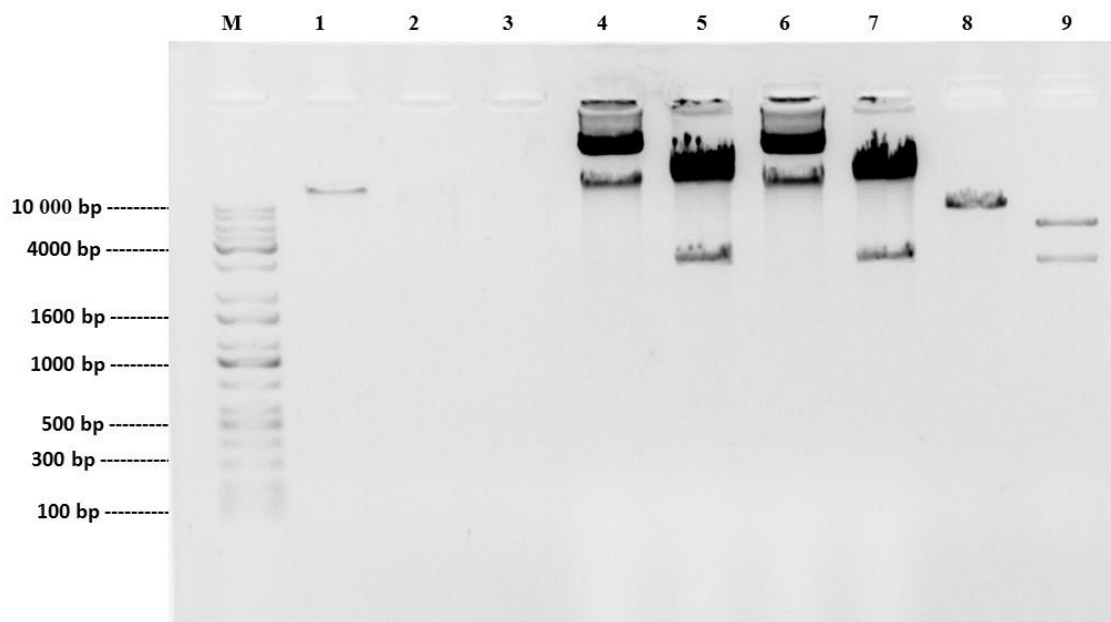


**Figure 3.9:** LB agar plate showing the growth of colony forming units (CFUs) obtained from the fosmid production library, dilution factor of  $10^{-5}$ .



Large-insert libraries with fosmid vectors are essential for cloning large gene clusters required to express more complex metabolic pathways. Especially those seen with alkane degradation pathways involving three enzymes working in synergy to accomplish complete breakdown. The large number of fosmid clones obtained showed the high coverage of the library which is important for finding genes of interest. Function based screening was adopted in this study to screen for fosmid clones able to degrade hydrocarbons using an agar-based screening method in which a minimal media (BH media) was supplemented with hydrocarbon substrates. A total of 15 fosmid clones were identified using this method. This screening method has proven to be very successful in identifying hydrocarbon degrading enzymes such as dioxygenases involved in polycyclic aromatic degradation (Duarte *et al.* 2017), toluene monooxygenases involved in toluene oxidation (Bouhajja *et al.* 2017) and catechol 1,2-dioxygenase and catechol 2,3-dioxygenase involved in phenol degradation (Pessoa *et al.* 2013).

Two randomly selected fosmid clones were selected from the LB plate and identified as CB1 and CB7. The fosmid DNA was isolated and restriction digestion performed with *Xba*I in order to confirm the quality of the fosmid DNA and the size of the insert DNA as shown in **Figure 3.10**.



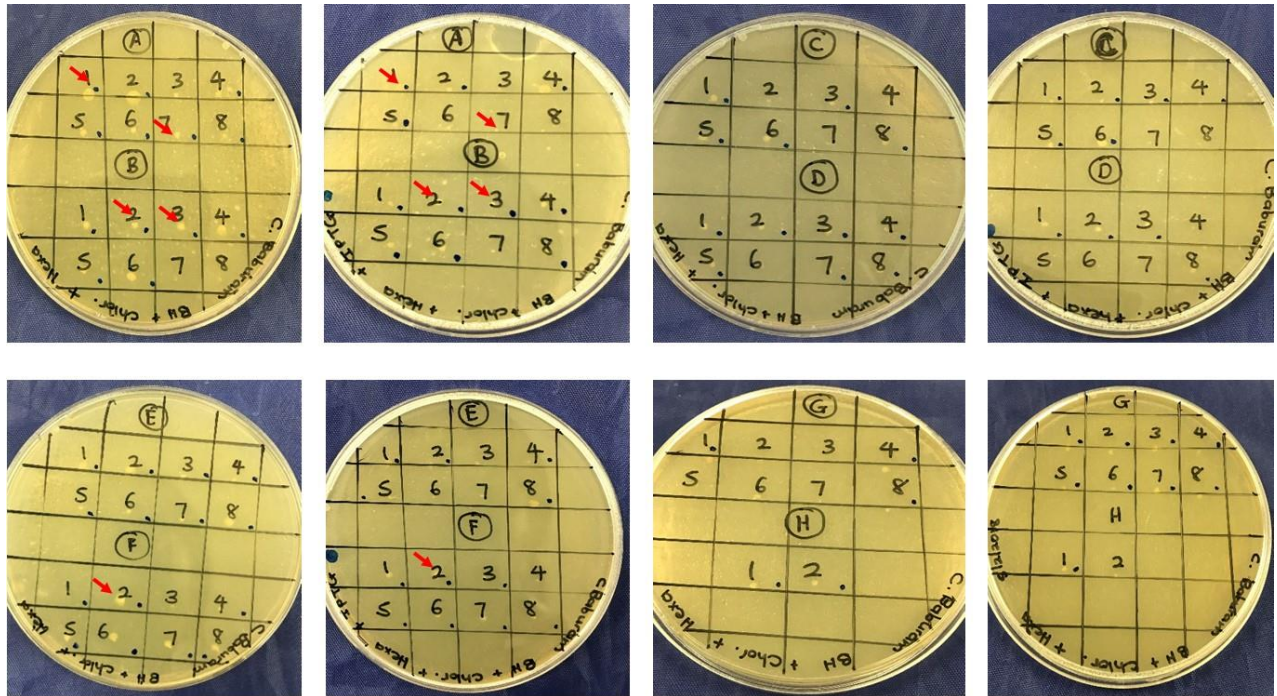
**Figure 3.10:** 0.8 % (w/v) Agarose gel electrophoresis showing undigested and digested plasmid DNA obtained from fosmid clones CB1 and CB7, EPI300-T1<sup>®</sup> host cells and pCC2FOS CopyControl vector for verification purposes. M: 10 kb DNA ladder, Lane 1: Fosmid control DNA (40 kb), Lane 2: EPI300-T1<sup>®</sup> host cell plasmid DNA, Lane 3: EPI300-T1<sup>®</sup> host cell plasmid DNA digested with *Xba I*, Lane 4: 20 µl Plasmid DNA isolated from CB1, Lane 5: 20 µl Plasmid DNA from CB1 digested with *Xba I*, Lane 6: 20 µl Plasmid DNA isolated from CB7, Lane 7: 20 µl Plasmid DNA from CB7 digested with *Xba I*, Lane 8: 250 ng undigested pCC2FOS CopyControl vector and Lane 9: 250 ng pCC2FOS CopyControl vector digested with *Xba I*.

The DNA quality and concentration were found to be high as observed by readings obtained from the Nanodrop. *Xba I* is expected to cleave at two locations within the pCC2FOS CopyControl vector (413 bp and 3234 bp). The expected sizes were observed on the agarose gel hence confirming the vector contained the insert DNA. This further confirmed that the ligation reaction was efficient and the clones growing on the plate contained the vector and the insert, thus allowing for further investigation using functional screening. The EPI300-T1<sup>®</sup> host cells were used as a negative control to show that no plasmid DNA was present. Hence, the plasmid DNA observed from the fosmid clones would only be from the library preparation. From the results, no plasmid DNA was obtained from the EPI300-T1<sup>®</sup> host cells. The pCC2FOS CopyControl vector was restricted with *Xba I* and showed the expected banding pattern with fragments of 3234 bp and 2821 bp based on the vector map. The band of 2821 was observed for the plasmid DNA isolated from the fosmid clones and restricted with *Xba I*. This confirmed that the vector with insert were contained within the host cells.

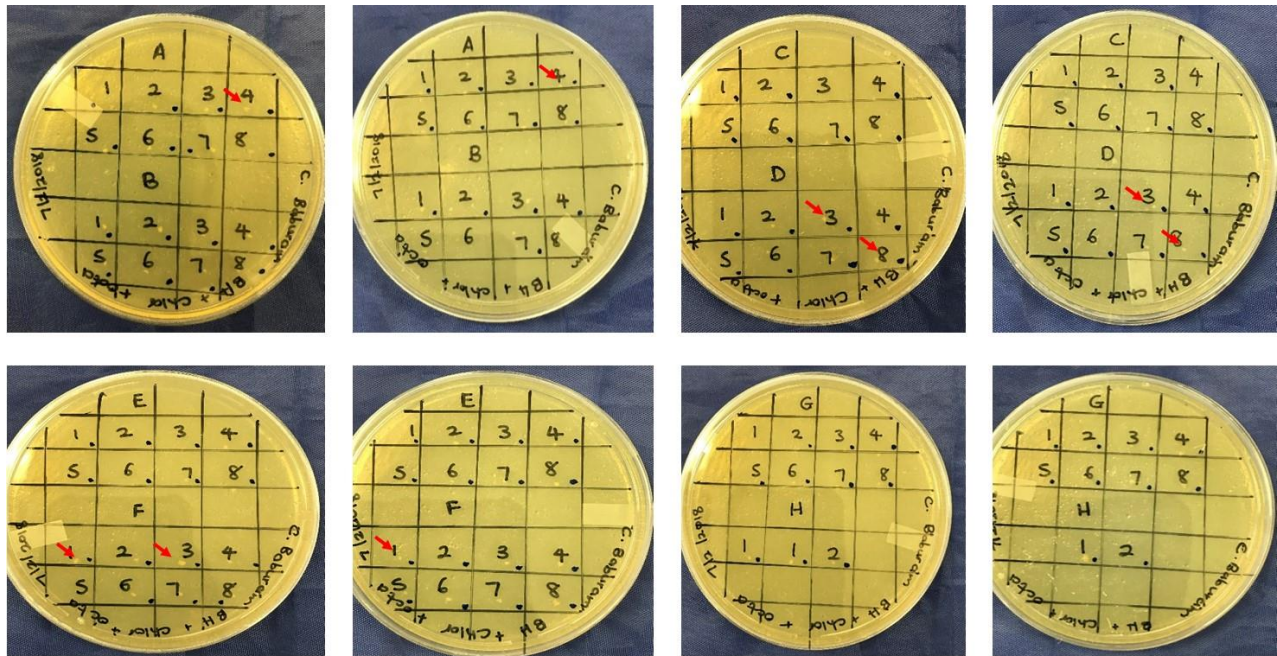
### 3.3.4 Functional screening of fosmid clones

Eighty candidate fosmid clones from the current study were randomly chosen from the prepared library and screened for hydrocarbon degrading potential using a modified method described by (Joy *et al.* 2017). The fosmid clones grown on a minimal media of Bushnell Haas was

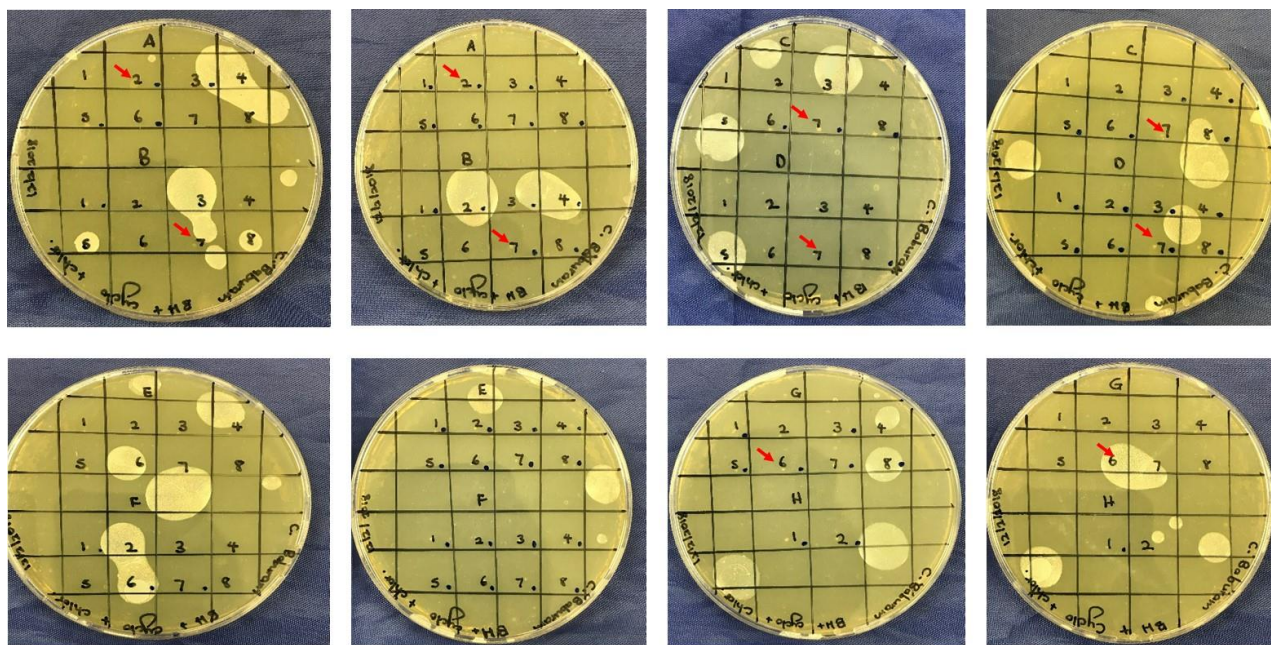
supplemented with hexadecane, octadecene and cyclohexane respectively. Growth of the fosmid clones were recorded after 5 days as shown in **Figure 3.11(A), (B) and (C)**.



(A)



(B)



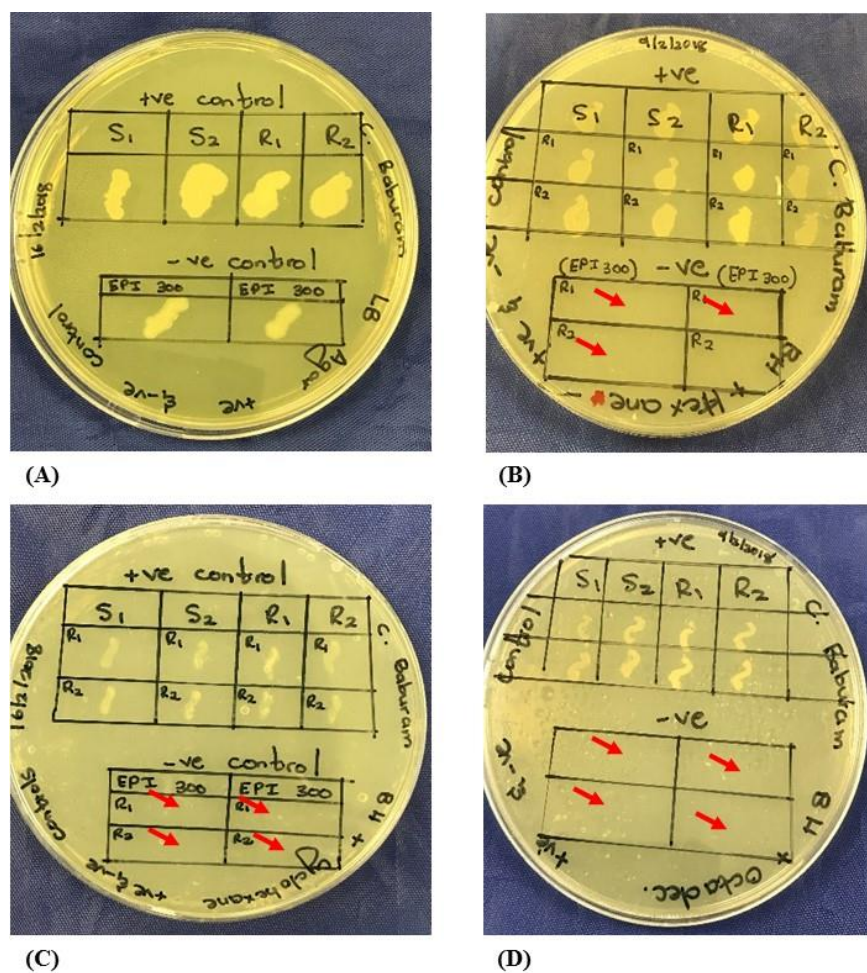
(C)

**Figure 3.11:** Functional screening analysis of fosmid library clones grown on Bushnell Haas agar supplemented with hexadecane (A), octadecene (B) and cyclohexane (C) as a sole carbon source in duplicates. Arrows denote candidate fosmid clones that displayed high growth rates with large colony sizes after incubation for 5 days at 37 °C.

The results obtained showed that some clones grew at a rate faster than others as denoted by the red arrows in **Figure 3.11** while some showed no growth. This was as a result of hydrocarbons being a sole carbon source, hence only clones with hydrocarbon degrading potential would be able to grow as they contain the genes involved in the degradation pathway. This result is a promising premise that the genes in these pathways have been captured by the constructed library. They are invaluable to the continuation of this study and will be further characterised downstream to determine their identities and functions.

Similar results have also been observed by other researchers who have implemented this functional screening approach for identifying hydrocarbon degrading fosmid clones (Bouhajja *et al.* 2017,

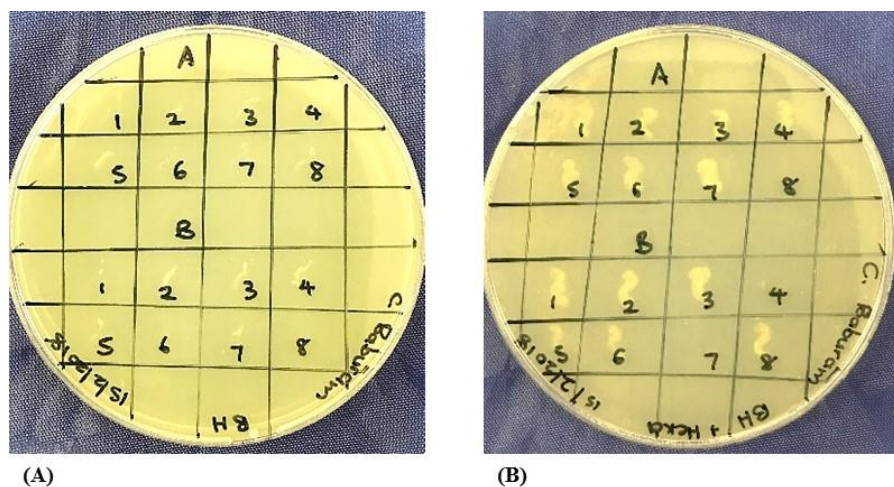
Duarte *et al.* 2017, Verma *et al.* 2017, Ngara and Zhang 2018). Hence, supporting the value of using metagenomic DNA to construct fosmid libraries to ultimately identify potential candidate genes involved in hydrocarbon degradation. The target clones will be sequenced with confidence knowing that identifying genes of interest are of a high probability due to the hydrocarbon degrading potential observed from functional screening. The results for the growth of positive and negative controls on Bushnell Haas media with hydrocarbon substrates and on LB agar are shown in **Figure 3.12**.



**Figure 3.12:** Growth of positive controls (S1, S2, R1 and R2) obtained from the culture based technique and the negative control (EPI300-T1<sup>®</sup> host cells) on LB agar (A), BH agar with hexadecane (B), BH agar with cyclohexane (C) and BH agar with octadecene (D). The red arrows

denote the negative controls showing no growth on the BH media supplemented with the different substrates.

The positive controls (S1, S2, R1 and R2) and the negative control (EPI300-T1<sup>®</sup> host cell) were both found to grow on the LB agar plates within 24 hours. With regards to growth on BH media supplemented with various hydrocarbons, the negative controls showed no growth as denoted by the red arrows in **Figure 3.12**. This result was expected since the host cells cannot survive on a minimal media containing only hydrocarbons. The positive controls obtained from the culture based technique on the other hand showed accelerated growth on all three substrates (hexadecane, octadecene and cyclohexane). Confirming their ability to thrive on a hydrocarbon source due to their evolution to survive over time in a toxic environment and adapt enzymatic mechanisms to degrade and utilize hydrocarbons as a sole carbon source. These results validated the use of BH media in the study and eliminated false positives by confirming that clones able to grow on this media was only made possible by the supplemented hydrocarbon source. The growth of the candidate fosmid clones on Bushnell Haas media with and without hydrocarbon substrates were investigated and the results are shown in **Figure 3.13**.



**Figure 3.13:** Fosmid clones (A1-A8 and B1-B8) grown on Bushnell Haas agar without a hydrocarbon substrate (A) and with a hydrocarbon substrate (hexadecane) (B).

Fosmid clones grown on BH media showed slow and for some no growth whereas those grown on Bushnell Haas supplemented with a hydrocarbon substrate showed accelerated growth. BH media is a minimal media which contains all the nutrients except a carbon source that is necessary for the growth of bacteria. Only those fosmid clones that are able to degrade hydrocarbons will grow on this media. This selective method will ensure that only fosmid clones containing specific genes involved in pathways of hydrocarbon degradation will be selected for further analysis in identifying these genes. The result could be due to some fosmid clones being capable of hydrolyzing the hydrocarbon leading to accelerated growth as compared to those clones with no substrate in the media. This confirms that the experiment was suitable for selecting those clones that can utilize the hydrocarbons. Substantiating that the growth of the clones are only as a result of hydrocarbon availability in the media. Hence, the BH media can be assumed to be carbon free in its composition as stipulated by the manufacturer and supported by other researchers in their use for identifying fuel degrading microorganisms (Bushnell and Haas 1941, Panda *et al.* 2013, Ali Khan *et al.* 2017, Kim *et al.* 2019).

### **3.3.5 Sequencing of candidate fosmid clones**

Fifteen candidate fosmid clones were chosen for downstream PacBio sequence analysis in the current study. The clones were selected based on their colony size and vigor on different hydrocarbon substrates after a growth period of 5 days. The clones selected were pFos\_A1, pFos\_A2, pFos\_A4, pFos\_A7, pFos\_B2, pFos\_B3, pFos\_B7, pFos\_C7, pFos\_D3, pFos\_D7, pFos\_D8, pFos\_F1, pFos\_F2, pFos\_F3 and pFos\_G6 (**Table 3.2**).

**Table 3.2:** The 15 fosmid clones selected following functional screening analysis on hydrocarbon substrates, hexadecane, cyclohexane and octadecene with their respective PacBio sequencing barcode names.

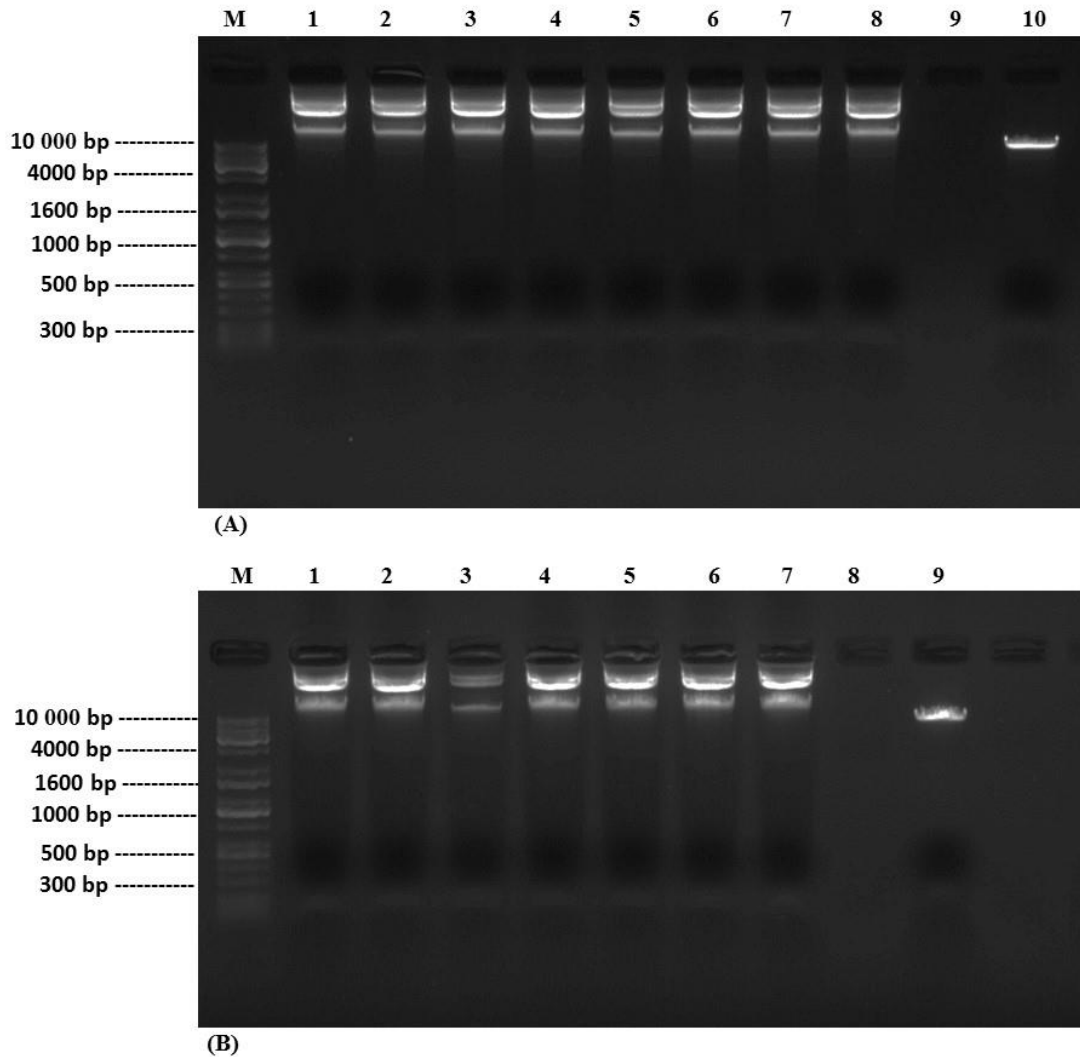
<b>Name of Fosmid clone</b>	<b>Substrate type</b>	<b>Barcode Name</b>
pFos_A1	Hexadecane	lbc1
pFos_A2	Cyclohexane	lbc2
pFos_A4	Octadecene	lbc3
pFos_A7	Hexadecane	lbc4
pFos_B2	Hexadecane	lbc5
pFos_B3	Hexadecane	lbc6
pFos_B7	Cyclohexane	lbc7
pFos_C7	Cyclohexane	lbc8
pFos_D3	Octadecene	lbc10
pFos_D7	Cyclohexane	lbc11
pFos_D8	Octadecene	lbc12
pFos_F1	Octadecene	lbc13
pFos_F2	Hexadecane	lbc14
pFos_F3	Octadecene	lbc15
pFos_G6	Cyclohexane	lbc16

PacBio sequencing was successfully completed and generated long sequence reads of high quality that were further analysed and discussed in Chapter 4. Such a platform was chosen since it provides long-read sequences with high consensus accuracy and uniform coverage (Wei and Zhang 2018). It can therefore be concluded that the DNA extraction method developed in this



study produced high quality DNA that can be successfully used to construct large insert DNA libraries as well as be sequenced using PacBio sequencing platforms. Many studies have shown the challenges experienced when extracting high quality DNA from hydrocarbon contaminated soils (Lee and Hallam 2009, Fatima *et al.* 2014, Tanveer *et al.* 2016). This study has therefore developed a potential standardised DNA extraction method that can be applied to all hydrocarbon contaminated soil samples displaying high co-extracted contaminants.

Fosmid DNA was isolated from these 15 candidate clones and separated on a 0.8 % (w/v) agarose gel to determine the quality and quantity of the DNA as shown in **Figure 3.14(A)** and (B). The results obtained showed good quality DNA with sharp bands and no smearing on the agarose gel. This was sufficient to proceed with PacBio sequencing.



**Figure 3.14:** 0.8 % (w/v) Agarose gel electrophoresis of isolated fosmid DNA from 15 candidate clones obtained from functional screening analysis. (A) M: KAPA™ Biosystems Universal ladder (10 kb), Lane 1: pFos\_A1, Lane 2: pFos\_A2, Lane 3: pFos\_A4, Lane 4: pFos\_A7, Lane 5: pFos\_B2, Lane 6: pFos\_B3, Lane 7: pFos\_B7, Lane 8: pFos\_C7 and Lane 10: Fosmid control DNA (40 kb). (B) M: KAPA™ Biosystems Universal ladder (10 kb), Lane 1: pFos\_D3, Lane 2: pFos\_D7, Lane 3: pFos\_D8, Lane 4: pFos\_F1, Lane 5: pFos\_F2, Lane 6: pFos\_F3, Lane 7: pFos\_G6 and Lane 9: Fosmid control DNA (40 kb).

### **3.4 Conclusion**

The modified metagenomic DNA extraction method used in this study is valuable for extracting high quality DNA and for the successful construction of metagenomic libraries. The functional screening method was also successful in identifying candidate fosmid clones based on their ability to survive on the hydrocarbon substrates like hexadecane, octadecene and cyclohexane. The rich metagenomic library constructed from hydrocarbon contaminated soil proved to be a valuable source of clones harbouring hydrocarbon-degrading genes. Therefore, we believe, the optimized method, which won a best poster presentation award at an international conference in Melbourne (Australia), could be used as a blueprint to extract high molecular weight and quality metagenomic DNA from hydrocarbon contaminated soil. In addition, the constructed fosmid libraries could further be screened for any biomolecules of interest including candidates that could be part of enzymes cocktail for potential biodegradation of plastics, which are mostly of hydrocarbon origin.

### 3.5 References

- Ali Khan, A.H., Tanveer, S., Alia, S., Anees, M., Sultan, A., Iqbal, M., and Yousaf, S., 2017. Role of nutrients in bacterial biosurfactant production and effect of biosurfactant production on petroleum hydrocarbon biodegradation. *Ecological Engineering*, 104, 158–164.
- Ambardar, S., Gupta, R., Trakroo, D., Lal, R., and Vakhlu, J., 2016. High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian Journal of Microbiology*, 56 (4), 394–404.
- Borah, D. and Yadav, R.N.S., 2014. BH MEDIUM.pdf. *Biotechnology*.
- Bouhajja, E., McGuire, M., Liles, M.R., Bataille, G., Agathos, S.N., and George, I.F., 2017. Identification of novel toluene monooxygenase genes in a hydrocarbon-polluted sediment using sequence- and function-based screening of metagenomic libraries. *Applied Microbiology and Biotechnology*, 101 (2), 797–808.
- Brahmi, A., Castonguay, A., and Déziel, É., 2019. Novel ‘Bacteriospray’ Method Facilitates the Functional Screening of Metagenomic Libraries for Antimicrobial Activity. *Methods and Protocols*, 2 (1), 4.
- Bushnell, L.D. and Haas, H.F., 1941. The Utilization of Certain Hydrocarbons by Microorganisms. *Journal of bacteriology*, 41 (5), 653–673.
- CC, E., CB, C., and GC, O., 2018. Field Metagenomics of Bacterial Community Involved in Bioremediation of Crude Oil-Polluted Soil. *Journal of Bioremediation & Biodegradation*, 09 (05).
- Christova, N., Kabaivanova, L., Nacheva, L., Petrov, P., and Stoineva, I., 2019. Biodegradation of crude oil hydrocarbons by a newly isolated biosurfactant producing strain. *Biotechnology and Biotechnological Equipment*, 33 (1), 863–872.
- Duarte, M., Nielsen, A., Camarinha-Silva, A., Vilchez-Vargas, R., Bruls, T., Wos-Oxley, M.L., Jauregui, R., and Pieper, D.H., 2017. Functional soil metagenomics: elucidation of polycyclic aromatic hydrocarbon degradation potential following 12 years of in situ bioremediation. *Environmental Microbiology*.
- Fatima, F., Pathak, N., and Rastogi Verma, S., 2014. An Improved Method for Soil DNA Extraction to Study the Microbial Assortment within Rhizospheric Region. *Molecular Biology*

*International*, 2014, 1–6.

Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W.R., Golyshina, O. V., and Golyshin, P.N., 2016. Estimating the success of enzyme bioprospecting through metagenomics: Current status and future trends. *Microbial Biotechnology*, 9 (1), 22–34.

Gastauer, M., Vera, M.P.O., de Souza, K.P., Pires, E.S., Alves, R., Caldeira, C.F., Ramos, S.J., and Oliveira, G., 2019. Data descriptor: A metagenomic survey of soil microbial communities along a rehabilitation chronosequence after iron ore mining. *Scientific Data*, 6 (February), 1–10.

Hassan, M., Essam, T., and Megahed, S., 2018. Illumina sequencing and assessment of new cost-efficient protocol for metagenomic-DNA extraction from environmental water samples. *Brazilian Journal of Microbiology*, 49, 1–8.

Joshi, M.N., Dhebar, S. V, Bhargava, P., Pandit, a S., Patel, R.P., Saxena, a K., and Bagatharia, S.B., 2014. Metagenomic Approach for Understanding Microbial Population from Petroleum Muck. *Genome Announcements*, 2 (3), e00533-14.

Joy, S., Rahman, P.K.S.M., and Sharma, S., 2017. Biosurfactant production and concomitant hydrocarbon degradation potentials of bacteria isolated from extreme and hydrocarbon contaminated environments. *Chemical Engineering Journal*, 317, 232–241.

Kim, H.S., Dong, K., Kim, J., and Lee, S.S., 2019. Characteristics of crude oil-degrading bacteria *Gordonia iterans* isolated from marine coastal in Taean sediment. *MicrobiologyOpen*, 8 (6), 1–10.

Lee, S. and Hallam, S.J., 2009. Extraction of High Molecular Weight Genomic DNA from Soils and Sediments. *Journal of Visualized Experiments*, (33), 2–5.

Lu, C., Hong, Y., Liu, J., Gao, Y., Ma, Z., Yang, B., Ling, W., and Waigi, M.G., 2019. A PAH-degrading bacterial community enriched with contaminated agricultural soil and its utility for microbial bioremediation. *Environmental Pollution*, 251, 773–782.

Martinez, A., Tyson, G.W., and Delong, E.F., 2010. Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses, 12, 222–238.

Nesme, J., Achouak, W., Agathos, S.N., Bailey, M., Baldrian, P., Brunel, D., Frosteg??rd, ??sa,

Heulin, T., Jansson, J.K., Jurkevitch, E., Kruus, K.L., Kowalchuk, G.A., Lagares, A., Lappin-Scott, H.M., Lemanceau, P., Le Paslier, D., Mandic-Mulec, I., Murrell, J.C., Myrold, D.D., Nalin, R., Nannipieri, P., Neufeld, J.D., O’Gara, F., Parnell, J.J., Pöhler, A., Pylro, V., Ramos, J.L., Roesch, L.F.W., Schlöter, M., Schleper, C., Sczyrba, A., Sessitsch, A., Sjöling, S., Sørensen, J., Sørensen, S.J., Tebbe, C.C., Topp, E., Tsiamis, G., Van Elsas, J.D., Van Keulen, G., Widmer, F., Wagner, M., Zhang, T., Zhang, X., Zhao, L., Zhu, Y.G., Vogel, T.M., and Simonet, P., 2016. Back to the future of soil metagenomics. *Frontiers in Microbiology*, 7 (FEB), 1–5.

Ngara, T.R. and Zhang, H., 2018. Recent Advances in Function-based Metagenomic Screening. *Genomics, Proteomics and Bioinformatics*, 16 (6), 405–415.

Olowomofe, T.O., Oluyeye, J.O., Aderiye, B.I., and Oluwole, O.A., 2019. Degradation of poly aromatic fractions of crude oil and detection of catabolic genes in hydrocarbon-degrading bacteria isolated from Agbabu bitumen sediments in Ondo State, 5 (September), 308–323.

Panda, S.K., Kar, R.N., and Panda, C.R., 2013. Isolation and identification of petroleum hydrocarbon degrading microorganisms from oil contaminated environment. *International Journal of Environmental Sciences*, 3 (5), 1314–1321.

Park, C. and Park, W., 2018. Survival and energy producing strategies of Alkane degraders under extreme conditions and their biotechnological potential. *Frontiers in Microbiology*, 9 (MAY), 1–15.

Pessoa, T.B.A., de Souza, S.S., Cerqueira, A.F., Rezende, R.P., Pirovani, C.P., and Dias, J.C.T., 2013. Construction and validation of metagenomic DNA libraries from landfarm soil microorganisms. *Genetics and Molecular Research*, 12 (2), 2148–2155.

Rhoads, A. and Au, K.F., 2015. PacBio Sequencing and Its Applications. *Genomics, Proteomics and Bioinformatics*, 13 (5), 278–289.

S, R., T, D., V, V., and R, S., 2014. Comparative Studies on the Extraction of Metagenomic DNA from Various Soil and Sediment Samples of Jammu and Kashmir Region in Prospect for Novel Biocatalysts. *IOSR Journal of Environmental Science, Toxicology and Food Technology*, 8 (4), 46–56.

Silva, C.C., Hayden, H., Sawbridge, T., Mele, P., De Paula, S.O., Silva, L.C.F., Vidigal, P.M.P.,

Vicentini, R., Sousa, M.P., Torres, A.P.R., Santiago, V.M.J., and Oliveira, V.M., 2013. Identification of Genes and Pathways Related to Phenol Degradation in Metagenomic Libraries from Petroleum Refinery Wastewater. *PLoS ONE*, 8 (4), 1–11.

Simon, C. and Daniel, R., 2011. Metagenomic analyses: Past and future trends. *Applied and Environmental Microbiology*, 77 (4), 1153–1161.

Solomon, S., Kachiprath, B., Jayanath, G., Sajeevan, T.P., Bright Singh, I.S., and Philip, R., 2016. High-quality metagenomic DNA from marine sediment samples for genomic studies through a preprocessing approach. *3 Biotech*, 6 (2), 1–5.

Song, W., Thomas, T., and Edwards, R.J., 2019. Complete genome sequences of pooled genomic DNA from 10 marine bacteria using PacBio long-read sequencing. *Marine Genomics*, 48 (May), 100687.

Tanveer, A., Yadav, S., and Yadav, D., 2016. Comparative assessment of methods for metagenomic DNA isolation from soils of different crop growing fields. *3 Biotech*, 6 (2), 1–5.

Vasconcellos, S.P., Sierra-Garcia, I.N., Dellagnezze, B.M., Vicentini, R., Midgley, D., Silva, C.C., Santos Neto, E. V., Volk, H., Hendry, P., and Oliveira, V.M., 2017. Functional and genetic characterization of hydrocarbon biodegrader and exopolymer-producing clones from a petroleum reservoir metagenomic library. *Environmental Technology*, 3330 (August), 1–12.

Verma, S.K., Singh, H., and Sharma, P.C., 2017. An improved method suitable for isolation of high-quality metagenomic DNA from diverse soils. *3 Biotech*, 7 (3), 1–7.

Wei, Z.G. and Zhang, S.W., 2018. NPBSS: A new PacBio sequencing simulator for generating the continuous long reads with an empirical model. *BMC Bioinformatics*, 19 (1), 1–9.

Xu, X., Tian, S., Jiang, P., Li, H., Wang, W., Gao, X., Qi, Q., Liu, W., Li, F., and Yu, H., 2018. Petroleum Hydrocarbon-Degrading Bacteria for the Remediation of Oil Pollution Under Aerobic Conditions: A Perspective Analysis. *Frontiers in Microbiology*, 9 (December), 1–11.

Zafra, G., Taylor, T.D., Absalón, A.E., and Cortés-Espinosa, D. V., 2016. Comparative metagenomic analysis of PAH degradation in soil by a mixed microbial consortium. *Journal of Hazardous Materials*, 318 (55), 702–710.

---

# Chapter 4

## Next Generation Sequencing of Metagenomic DNA from Hydrocarbon Contaminated Soils

---

### *Abstract*

Soils contaminated with petroleum hydrocarbons are rich sources of microbial consortia with an abundance of enzymes suitable for degrading a range of hydrocarbons. In this study we discuss *in silico* analyses of MiSeq data obtained from next generation sequencing of metagenomic DNA extracted from hydrocarbon-contaminated soil. We profiled microbial communities and predicted the functional pathways in the metagenomic pool. Accordingly, *Proteobacteria* and *Actinobacteria* were found to be the predominant phyla in the metagenomic DNA and detected at roughly 84 % and 8 % phylum abundance, respectively based on both SILVA.123 and Greengene databases. Thus, there is a clear increase in *Proteobacteria* and *Actinobacteria* growth rates following hydrocarbon contamination of soil environments, which negatively affected the microbial diversity. This is in agreement with findings observed by other researchers in the field. It can be assumed that the abundant phyla in the soils have evolved enzymatic mechanisms to survive in this harsh environment. Several functional genes that play pivotal role in the metabolism of aliphatic and aromatic hydrocarbons were also identified. These include methane/ammonia monooxygenases and alkane hydroxylases which are crucial for the initial step of alkane degradation by the introduction of oxygen. The same goes for aromatic hydrocarbons in which dioxygenases were identified. A range of enzyme classes known to be involved in specific hydrocarbon pathways were also identified and including aldehyde dehydrogenases, alcohol dehydrogenases, decarboxylases, hydrolases, dehalogenases, reductases, transferases, synthases



and hydrolases. Based on this rich collection of enzymes from the metagenome, they piece together like a puzzle to show pathways that can lead to near-complete degradation of hydrocarbons. The outcomes of this sequence-based approach can be used to develop a bacterial consortia for application in bioremediation studies. Furthermore, the identification of a number of hydrocarbon-degrading enzymes during *in silico* analysis further confirms the richness of the fosmid library that was constructed from the same pool.

**Keywords:** Hydrocarbons, metagenomic DNA, alkanes, aromatics, hydroxylases, dehydrogenases, microbial consortia, enzymes, bioremediation.

<b>TABLE OF CONTENTS</b>	<b>PAGES</b>
4.1 Introduction .....	150
4.2 Materials and Methods .....	152
4.2.1 Sample collection of oil contaminated soils .....	152
4.2.2 Metagenomic DNA extraction from soil samples .....	152
4.2.3 Gene sequencing and data analysis.....	153
4.2.4 Assembly and bioinformatics analysis using the JCVI metagenomics Pipeline ..	153
4.2.5 Assembly and bioinformatics analysis using the CLC Genomics Workbench ....	154
4.2.6 Data analysis for microbiome profiling and functional predictions .....	155
4.3 Results and Discussion.....	156
4.3.1 General characterisation of soil community composition .....	156
4.3.2 Predicted functions of bacterial communities.....	166
4.3.2.1 Aliphatic hydrocarbon degradation.....	171
4.3.2.2 Aromatic hydrocarbon degradation.....	178
4.4 Conclusion.....	179
4.5 References .....	180

## 4.1 Introduction

The most widespread environmental pollutants on earth are those of oil hydrocarbons. These include *n*-alkanes, cycloalkanes and polycyclic aromatic hydrocarbon (PAHs) (Yuniati, 2018). These have been classified as having serious effects on both ecological and public health (Spini *et al.* 2018). Through the years, petroleum oil contamination of ecosystems have increased as a result of crude oil drilling, transportation, refining and related activities. The public awareness has increased and as a result, driven a need to find effective and affordable technologies for the treatment of oil industrial wastes (Varjani and Upasani 2017).

This has opened up an avenue for bioremediation which uses microorganisms to decompose pollutants from contaminated soils. It has been found to be clean, cost-effective and environmentally friendly (Vikram *et al.* 2018). Most importantly, it provides a complete degradation of pollutants unlike other conventional methods which leave behind toxic secondary by-products (Godini *et al.* 2018). Soil environments with hydrocarbon contaminants are considered to have the most diverse natural microbial communities that can constitute a huge reservoir of enzymes for bioremediation of oil contaminants (Xu *et al.* 2014, Kumar *et al.* 2019, Pannekens *et al.* 2019). Traditional approaches have focused on isolating and cultivating culturable microorganisms from these soil environments but these only account for less than 1% of the soil microbial populations (Xu *et al.* 2014, Stefani *et al.* 2015).

During the past years, metagenomics has emerged as a powerful techniques to tap into rich genetic resources of unculturable bacteria by skipping the culturing of microorganisms and isolating genomic DNA directly from environmental samples (Handelsman 2005, Rodgers and Zhang 2019). In this study the sequence-based metagenomics approach was implemented and used to characterize not only the microbial diversity or bacterial abundance but also the functional metabolic potential in a complex hydrocarbon contaminated environment.

The value of understanding the microbial diversity and abundance of certain bacteria in these soil environments acts as a guide to characterizing microorganisms with the ability to produce hydrocarbon degrading enzymes (Barone *et al.* 2014, Lam *et al.* 2015, Abbasian *et al.* 2016). In addition, it is clear from the literature that bacterial consortia are essential for the complete degradation of hydrocarbon contaminants (Ganesh Kumar *et al.* 2014, Elumalai *et al.* 2017, Wanapaisan *et al.* 2018). This is made possible by evolved enzymatic mechanisms orchestrated by cocktails of enzymes working in synergy (Ji *et al.* 2019). Hence, using metagenomic DNA isolated from these bacterial consortiums will be invaluable for constructing large insert libraries. This increases the confidence level of locating target gene sequences from the libraries following functional screening analysis. This approach has proven to be successful by many researchers involved in such bioremediation studies (DeCastro *et al.* 2016, Sarkar *et al.* 2016a, Kusnezowa and Leichert 2017, Roy, Dutta, *et al.* 2018). The present study therefore aimed to better capture the bacterial communities and their overall enzymatic contributions by implementing next generation sequencing (NGS) and *in silico* analysis of metagenomic DNA isolated from hydrocarbon contaminated soil samples.

## **4.2 Materials and Methods**

### **4.2.1 Sample collection of oil contaminated soils**

Soil samples were collected from petrol, diesel and used oil contaminated sites located at the Truck stop in Sasolburg in the Freestate (26°48'38.5"S 27°50'33.5"E) and the R57 Diesel Depot in Vanderbijlpark, Golden Highway (26°42'04.3"S 27°48'15.8"E). Sampling was carried out at a depth of 10 cm using sterile equipment coring device. Samples were kept in sterile polyethylene bags and stored on ice for transfer to the Omics laboratory at the Vaal University of Technology. All samples were stored at 4°C until further analysis. Samples were labelled according to their specific sampling location as described in Chapter 3 section 3.1.1.

### **4.2.2 Metagenomic DNA extraction from soil samples**

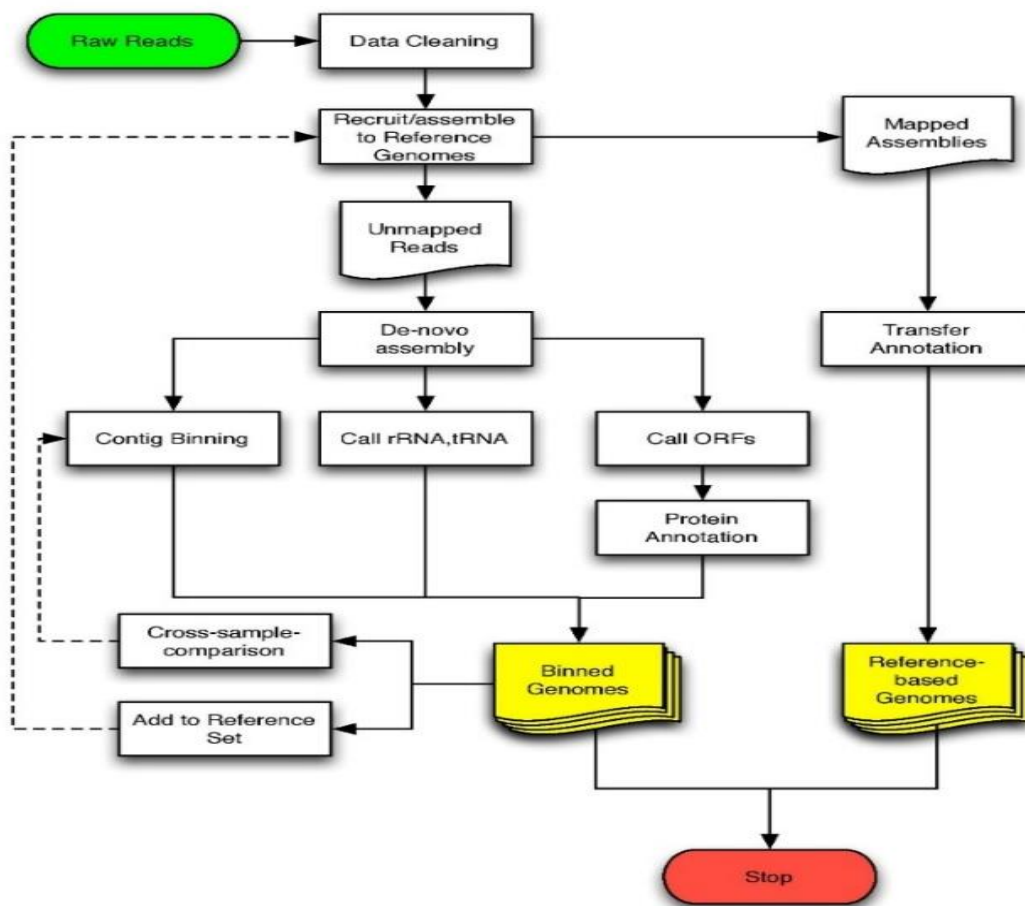
DNA extraction from soil samples (TS and N) from the Sasol Truck stop in Sasolburg in the Freestate and soil samples (T and G) from the R57 diesel depot in Vanderbijlpark was carried out using a modified CTAB extraction method according to (Singh *et al.*, 2014). The method developed for the extraction and purification of high quality pooled metagenomic DNA from the hydrocarbon contaminated soils has been described in more detail in Chapter 3, section 3.1.4. Extracting high quality metagenomic DNA in appreciable amounts is essential for the downstream application of next generation sequencing (NGS). Extracted metagenomic DNA preparations were quantified and quality checked using a Nanodrop™ 2000/c spectrophotometer (Thermo Fischer Scientific™). Samples were also electrophoretically separated on a 0.8 % (w/v) agarose gel to check for high molecular weight DNA with no smears indicative of intact DNA. Extracting high quality metagenomic DNA is essential for the downstream application of next generation sequencing (NGS).

### 4.2.3 Gene sequencing and data analysis

The microbial diversity, community composition and gene classes within the pooled metagenomic DNA extracted from the four soil samples TS, N, T and G were studied using the next generation sequencing (NGS) based approach. This was carried out at the Inqaba Biotec<sup>TM</sup> facility in Pretoria, South Africa. The metagenomic DNA was fragmented using an enzyme-based approach according to the NEBNext Ultra II kit workflow. The resulting fragments were purified (size selected), end-repaired and an Illumina specific adapter sequence ligated to all fragments. Following quantification, the samples were individually indexed, and a second size selection step was performed using the AMPure XP beads. Quality control was carried out on a DNA chip (Agilent 2100 Bioanalyzer) and sequenced on Illumina's MiSeq platform, using a MiSeq v3 (600 cycle) kit according to the manufacturer's instruction.

### 4.2.4 Assembly and bioinformatics analysis using the JCVI metagenomics Pipeline

The raw MiSeq data was sent from the company in FASTQ format and quality control was carried out by the company to ensure that all adapters and indices were trimmed. This was important in ensuring that the assembler does not join regions that are not biologically relevant which can result in the assembler taking a long time and resulting in misleading results. The NGS data was analysed using the J. Craig Venter Institute (JCVI) prokaryotic metagenomics pipeline (**Figure 4.1**) at the JCVI in Maryland, Washington, USA.



**Figure 4.1:** The J. Craig Venter Institute (JCVI) metagenomics analysis pipeline used to analyse next generation sequence (NGS) data obtained from metagenomic DNA extracted from hydrocarbon contaminated soil samples (Tanenbaum *et al.* 2010).

#### 4.2.5 Assembly and bioinformatics analysis using the CLC Genomics Workbench

To confirm the results obtained from the JCVI pipeline and make a clear and comparative analysis, the CLC Genomics Workbench 9.5 (CLC Bio, Qiagen) was used to analyse the microbial diversity and compositions of the sequenced data. Metagenome *de novo* assembly was carried out on the contigs (clean reads) using the CLC Genomics Workbench (Qiagen). The parameters for the

minimum threshold contig length was set at 200-bp to filter the short contigs after assembly. The assembler was run in longer contigs mode which produces significantly longer reads.

#### **4.2.6 Data analysis for microbiome profiling and functional predictions**

Microbiome profiling based on a database of taxonomic marker genes was carried out using the computational tools, MetaPhlAn 2 (ref MetaPhlAn2 for enhanced metagenomic) and the output of the MetaPhlAn 2 analyses was plotted using GraPhlAn based on the metagenomic reads (Asnicar *et al.* 2015). The top 15 (relative abundance) operational taxonomic units (OTUs) were chosen and displayed in a bar graph at the summed phylum and genus level abundance. Krona (<http://krona.sourceforge.net>), a web based metagenomics visualization software tool was also used to display the relative abundance of taxa across multiple levels of hierarchy (Ondov *et al.* 2011). SILVA.123 and Greengene databases were used for the detection of operational taxonomic units (OTUs) which showed a percentage similarity of 97 %.

Functional annotation which is the assignment of the most probable biological role for a given peptide was also carried out with the use of BLASTP against the most recent version of JCVI's PANDA data resource. This is an internal collection of non-redundant protein and nucleotide data sets from a variety of public databases (e.g. Uniprot, NCBI GenBank). The analysis makes use of BLOSUM62 substitution matrix and an E value cutoff of  $1e-5$ . The ten most significant alignments were stored for each peptide and the defined BLAST hits were delineated and ordered from high confidence to low confidence. The output format was in a tab delimited format. The OTU abundance table (obtained from Greengene database) was normalized to predict the functions of bacterial communities and categorise them into KEGG pathways using PICRUSt (Langille *et al.* 2013).



## 4.3 Results and Discussion

### 4.3.1 General characterisation of soil community composition

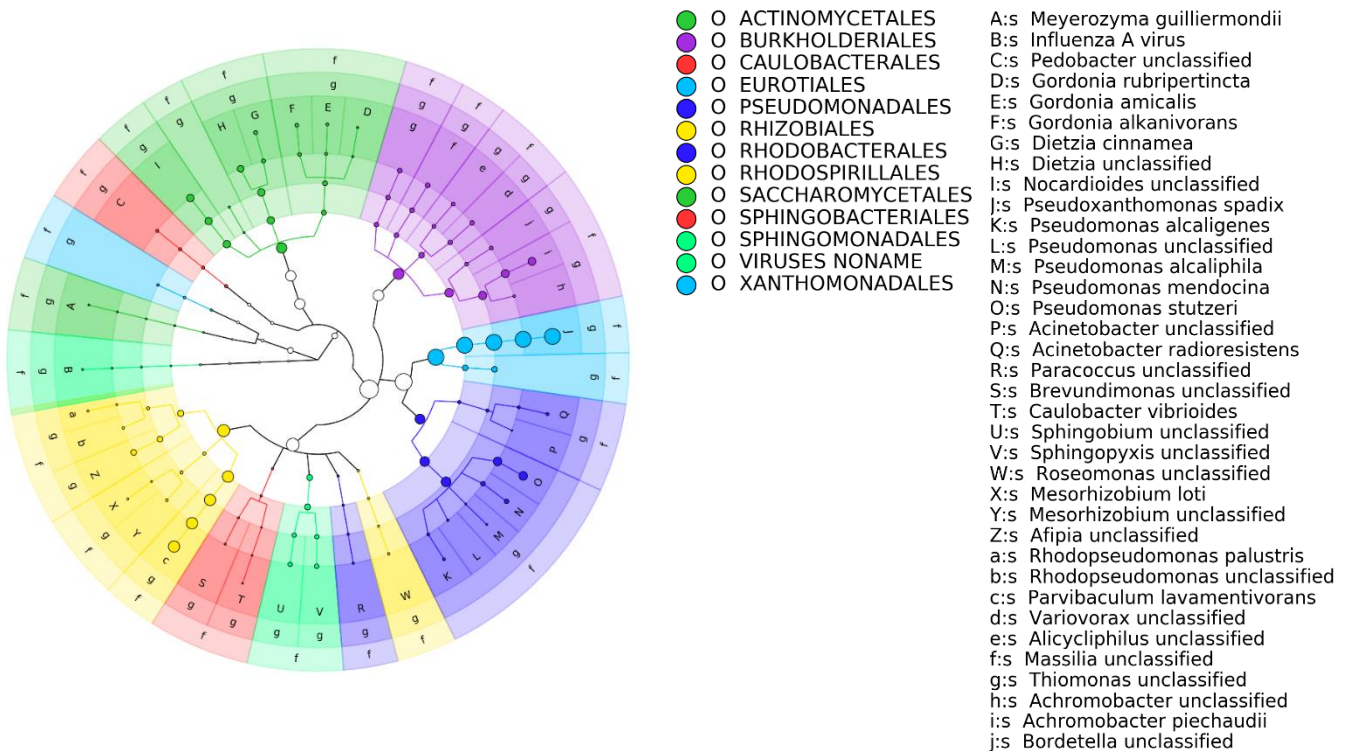
A comprehensive characterisation of taxonomic compositions in the soil microbiota from four hydrocarbon contaminated soil samples was carried out in this study. In order to achieve this, the next generation sequencing (NGS) method was used to capture the structure and abundance of the microorganisms inhabiting the soils. High quality metagenomic DNA was isolated from four soil samples in hydrocarbon-rich environments. Next generation sequencing was carried out on the pooled metagenomic DNA and bioinformatics analysis of the data was carried out using the JCVI metagenomics pipeline (Tanenbaum *et al.* 2010).

This pipeline encompasses both structural and functional annotations. This pipeline is designed to process data sets which are in the range of tens of millions of sequence reads (Tanenbaum *et al.* 2010). The analysis carried out with the data set from this study, involved gene finding or structural annotation. Structural annotation results in the most probable proteins present in the nucleotide sequence data and further identifies the best possible open reading frames from the metagenomics shotgun sequencing reads. It is important to remember that the putative proteins identified can be fragments of the full-length protein hence the beginning and end of each read were treated as putative start and stop sites.

The first step involved identifying open reading frames (ORFs) with a minimum size of 180 nucleotides based on the expected size of typical bacterial genes. Thereafter, the use of MetaGene Annotator a gene predictor tool and comparison of ORFs obtained with those from the MetaGene Annotator produced the longest possible putative proteins from the sequence data. The data was analysed using the JCVI metagenomics pipeline. This pipeline is unlike MG-RAST (Metagenomic Rapid Annotations using Subsystems Technology - open-source web application) because it does not confine itself to BLASTX based peptide identification which uses a defined data set. This

results in a larger yield of putative proteins for the same number of input reads. Therefore, the JCVI metagenomics pipeline was implemented in this study.

Based on the assessment by MetaPhlAn which explored the profiling of microbial community compositions by analyzing the metagenomic reads using unique clade-specific marker genes (Truong *et al.* 2015), a taxonomic circular cladogram (**Figure 4.2**) was generated. This helped in determining the major types of microbial species and their relationship with each other in these type of soil samples. This showed the abundance of the following taxonomic orders in decreasing order, *Actinomycetales*, *Burkholderiales*, *Caulobacterales*, *Eurotiales*, *Pseudomonadales*, *Rhizobiales*, *Rhodobacterales*, *Rhodospirillales*, *Saccharomycetales*, *Sphingobacterales*, *Sphingomonadales* and *Xanthomonadales*. This indicates that the soil samples contain an abundant consortium of microorganisms and the microbial community existing in these polluted soils are diverse.



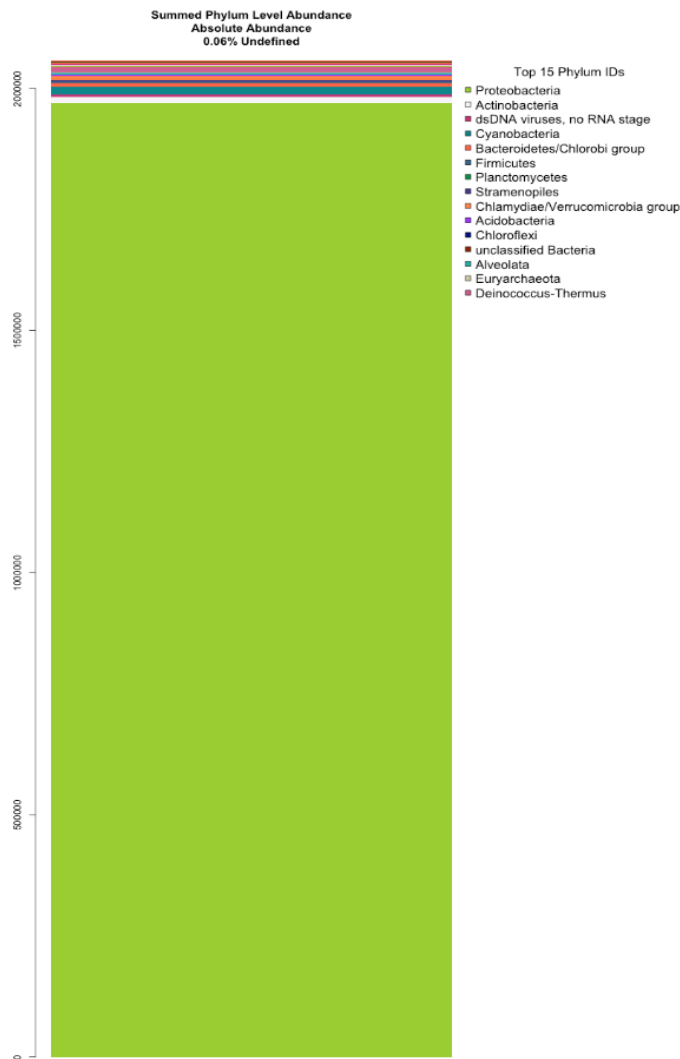
**Figure 4.2:** Taxonomic cladogram based on order of abundance showing the composition of soil microbial community based on the metagenomes of four hydrocarbon polluted soil habitats and their relation to each other using the computational tools, MetaPhlAn 2 and GraPhlAn.

Based on the cladogram, the predominant taxonomic orders from these soil samples were found to be *Actinomycetales* which are classified as gram positive bacteria and have been documented in literature as being involved in complex polysaccharide (Yeager *et al.* 2017) and hydrocarbon degradation in soil (Abbasian *et al.* 2016). Work carried out by Yeager *et al.*, 2017, examined the lignocellulolytic and chitinolytic potential of 112 *Actinomycetales* strains from semi-arid grasslands in Colorado in Utah. It was found that *Actinomycetales* sequences comprised up to 20 % of the bacterial 16s rRNA gene libraries from these soils and were highly involved in the degradation of lignocellulose and other complex biopolymers in soils.

With regards to the role of *Actinomycetales* in hydrocarbon degradation, research carried out by Abbasian *et al.* (2016) showed that *Actinomycetales* are dominant microorganisms in soils contaminated with crude oil. The study further showed that several functional genes (12.35 %) belonging to *Actinomycetales* play a role in the metabolism of aliphatic and aromatic hydrocarbons. Another recent study carried out by Miller *et al.* (2019) in the Caspian Sea which receives substantial input of hydrocarbons also showed high abundance levels of *Actinomycetales*. This therefore justifies the abundant nature of this taxonomic order found in soil samples used in this study which were also obtained from hydrocarbon polluted soils. Furthermore, it can be suggested that these microorganisms are well adapted for survival in this environment and have the enzymatic capabilities required to degrade both aromatic and aliphatic compounds.

A taxonomic distribution of the soil microbial communities based on phylum level abundance was also carried out using the computation analysis tool MetaPhlAn 2. This generated the top fifteen phylum (**Figure 4.3**) identities found in the soil samples. The bacterial abundance was at a much higher proportion than Archaea, Eukaryota and Viroids. Knowing these abundant phylum and

genus further shows the diversity of microorganism capable of surviving in hydrocarbon contaminated soil environments. They have become highly adapted to this extreme and toxic environment and have colonized this ecological niche. A host of research through the years shows that the abundant phylum and genus found in this study also predominate in other similar locations and thrive over time using hydrocarbons as a sole carbon source (Deshpande *et al.* 2018, Spini *et al.* 2018). Furthermore, it can be suggested that a consortium of microorganisms are required for complete hydrocarbon degradation which is a syntrophic process, whereby different members of the microbial community perform different steps in an overall metabolic process and cannot be carried out by a single member (Pannekens *et al.* 2019).



**Figure 4.3:** Taxonomic distribution of the soil microbial communities. Distribution at the phylum level.

With regards to the bacteria domain, *Proteobacteria*, *Actinobacteria* and *Cyanobacteria* were found to be predominant in the investigated soils. *Proteobacteria* and *Actinobacteria* were detected at roughly 84 % and 8 % phylum abundance respectively based on both SILVA.123 and Greengene databases (**Figure 4.3**). This finding can be collaborated with similar work done by Ren *et al.* (2018), Xu *et al.* (2014) and Pedron *et al.* (2019). Furthermore, work carried out by Pedron *et al.* (2019) using 16S profiling showed that *Proteobacteria* also dominated the water microbiome in all their sampling sites. Specifically, for the summed phylum level abundance, *Proteobacteria* was of the highest abundance in the microbial community as compared to any other phylum.

Similarly work done on desert soils from the Tarim Basin in northwestern China, concluded that *Proteobacteria* were the most abundant based on metagenomic data (Ren *et al.* 2018). The same finding was also observed by Xu *et al.* (2014), who used 33 publicly available metagenomes from diverse soil sites (i.e. grassland, forest soil, desert, Artic soil, and mangrove sediment). This clearly indicates that as suggested in the literature that *Proteobacteria* species have been reported in different complex ecosystems (Sun *et al.* 2014, Sarkar *et al.* 2016b, Lu *et al.* 2019) and have been proven to be ubiquitous and are able to survive in a wide range of environments including hydrocarbon contaminated ones clearly observed in this study as well.

It is also worth noting that *Firmicutes* species found to a lesser extent in this study have also been observed in a number of diverse environments around the world especially in those contaminated with petroleum (Prince and Mcgenity 2010). In the *Firmicutes* phylum, orders such as *Bacillales*, *Lactobacillales* and *Clostridiales* contain hydrocarbon-degrading genera and hence inferred to be able to degrade recalcitrant components. Usually they work in consortia with other bacteria due

to their ability to degrade hydrocarbons. Furthermore, their spore-forming nature allows them to survive in adverse nutritional depleted and environmental conditions (Guerra *et al.* 2018).

*Actinobacteria* ranking second on the abundance list also shows high resilience to toxic pollutants such as small volatile to long chain hydrocarbons (Brzeszcz and Kaszycki 2018). Recent studies have described or characterised new actinobacterial strains applicable for bioremediation, e.g., *Gordonia polyisoprenivorans* capable of cleavage of poly(cis-1,4-isoprene) (Anderl *et al.* 2018), *Gordonia* sp. capable of biodegradation of phthalic esters (Fan *et al.* 2018), *Gordonia rubripertincta* CWB2 capable of biodegradation of styrene (Heine, Zimmerling, *et al.* 2018), or *Gordonia phthalatica* sp. capable of biodegradation of various phthalic esters (Jin *et al.* 2017).

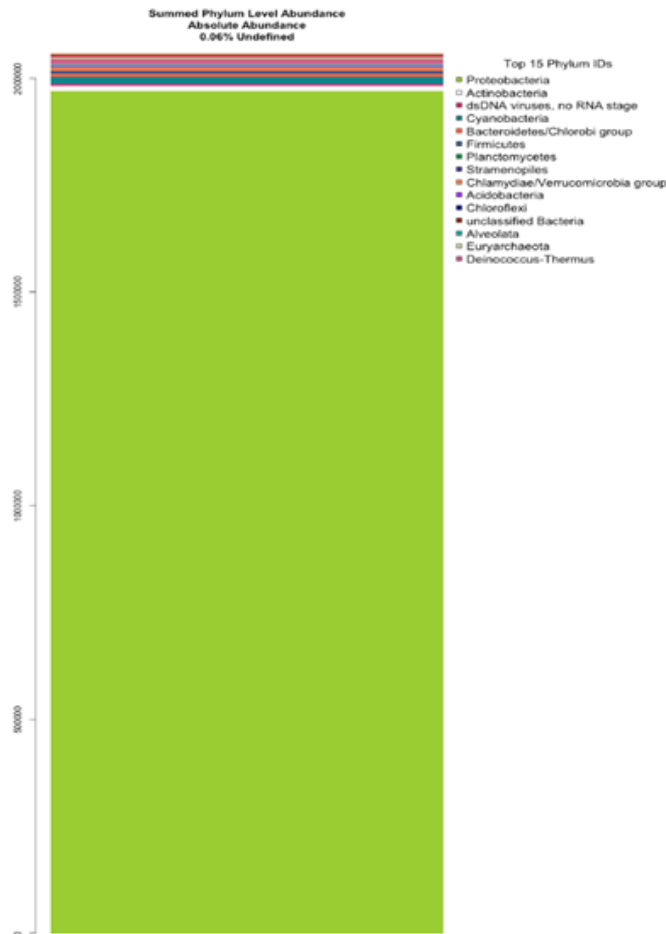
Many researchers have ventured even further to correlate the elevation in the abundance of *Proteobacteria* with other phyla as petroleum hydrocarbon concentrations increase (Yang *et al.* 2014, Garrido-Sanz *et al.* 2019). This stems from the fact that according to literature, there is a strong relation between pollutant loads and specific microbial communities (Spini *et al.* 2018). Moreover, *Actinobacteria*, *Acidobacteria*, *Cyanobacteria* and *Firmicutes* together with *Proteobacteria* have all been reported as the dominant phyla in several oil contaminated soil sites around the world (Joshi *et al.* 2014, Guerra *et al.* 2018).

Research carried out by Godini *et al.* (2018) for example, involved the isolation and identification of crude oil degrading bacteria from Kharg Island in Iran. The research identified strains belonging to *Proteobacteria* and *Firmicutes* which are infamous for producing catabolic enzymes effective in petroleum bioremediation for adaptation to oil sludge or highly polluted soils. Tiralerdpanich *et al.* (2018) also carried out work to identify the microbial consortium in the biodegradation of diesel, hexadecane and phenanthrene in mangrove sediments collected in Bangkhuntien in Bangkok. The 16S rRNA gene amplicon sequence analysis revealed that the major bacterial assemblages were *Gammaproteobacteria*, *Deltaproteobacteria* and *Alphaproteobacteria*. Again, studies by Yang *et al.*, 2014 on bacterial communities along the China-Russia crude oil pipeline

route showed the dominance of *Proteobacteria*, *Actinobacteria* and *Firmicutes*. The list of literature confirming the trend observed in this study is extensive (Abbasian *et al.*, 2016; Deshpande *et al.*, 2018; Pedron *et al.*, 2019; Roy *et al.*, 2018; Xu *et al.*, 2014). It can be said that after contamination, certain microbial communities become dominated by those taxa capable of metabolizing or tolerating complex hydrocarbons as mentioned above.

Interestingly, this study represented viruses as the third most abundant source of DNA from the soil samples (**Figure 4.3**). This is not uncommon since they have been commonly identified in oil reservoirs and other hydrocarbon rich environments (Cai *et al.* 2015, Abbasian *et al.* 2016, Crispim *et al.* 2018). Even though viruses are not the focus of this study, it is important to acknowledge that viruses have a major impact on microbial communities and their ecology. The bioremediation processes in oil-contaminated sites have been proposed using a phage-driven microbial loop (Lu *et al.* 2012). This is accomplished by them lysing their bacterial hosts thus causing the persistent release and turnover of nutritional biomass enabling bacterial hydrocarbon degradation (Rosenberg *et al.* 2010, Wang *et al.* 2016). However, it remains that there are more questions than answers with regards to their ecological importance and the extent to which they shape and control microbial communities. Therefore, more research is still required in this area to fill the knowledge gap.

A taxonomic distribution of the soil microbial communities based on genus level abundance was also carried out using the computational analysis tool MetaPhlAn 2. This generated the top fifteen genus identities found in the soil samples as shown in **Figure 4.4**. The findings indicates that the OTUs with the highest relative abundance with hydrocarbon degrading capabilities belong to *Parvibaculum* (Schleheck *et al.* 2004), *Pseudoxanthomonas* (Patel *et al.* 2012, Duarte *et al.* 2017), *Burkholderia* (Lee *et al.* 2016) and *Acinetobacter* (Xu *et al.* 2017). These genera fall under the wider class of *Gammaproteobacteria*. Importantly, it has been reported that this class also includes the genera *Alcanivorax*, *Marinobacter* and *Acinetobacter* and according to literature, are acceptably known as oil degraders in the scientific community (Kostka *et al.* 2011, Dombrowski *et al.* 2016, Bouhajja *et al.* 2017, Kadri *et al.* 2018).



**Figure 4.4:** Taxonomic distribution of the soil microbial communities. Distribution at the genus level.

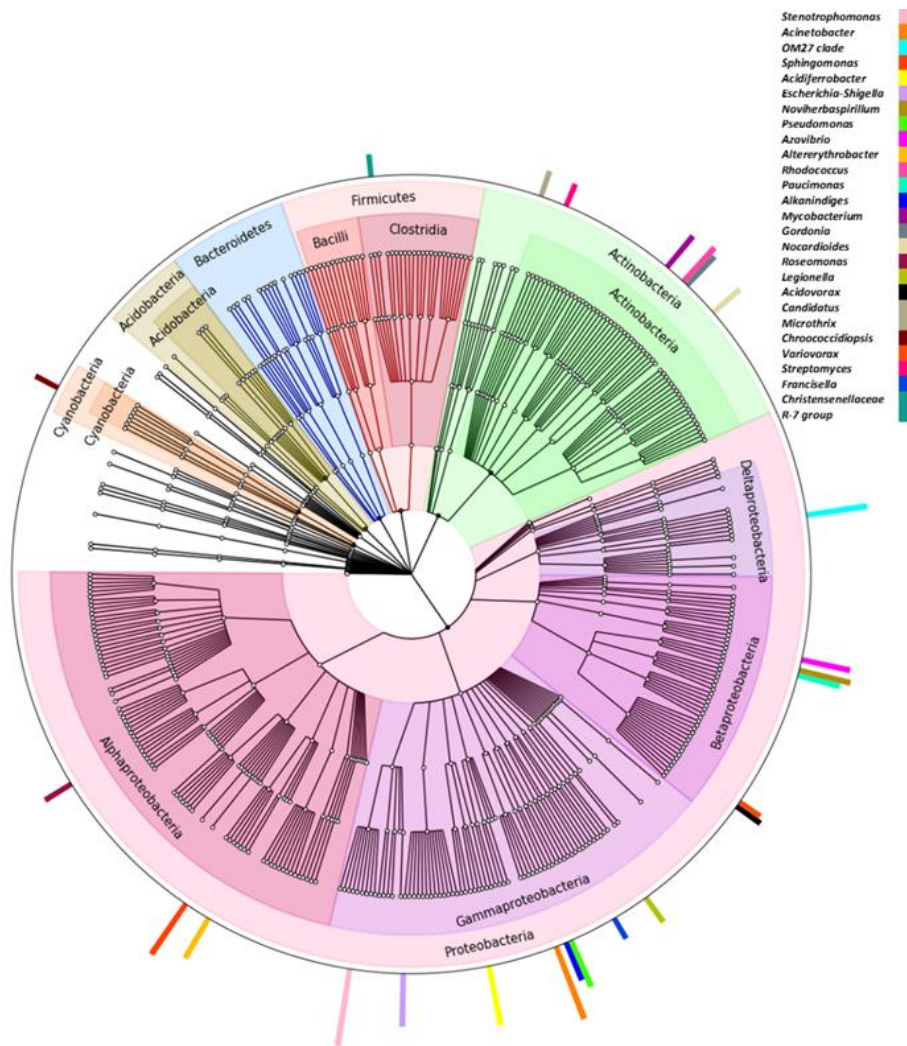
A closer analysis of *Parvibaculum* was warranted due its abundant nature in this study and its involvement with aliphatic-hydrocarbon degradation (Schleheck *et al.* 2011, Rosario-Passapera *et al.* 2012, Park and Park 2018) and poly aromatic hydrocarbon degradation (Schleheck *et al.* 2011, Deshpande *et al.* 2018) . Polymerase chain reaction (PCR) based studies examining the Atlantic Ocean surface seawater and deep-sea hydrothermal vents revealed that only *Parvibaculum* strains possess an alkane-oxidising cytochrome P450 (CYP)-like protein to degrade alkanes (Park and Park 2018). This finding was further echoed very recently by Garrido-Sanz *et al.* (2019) using metagenomic insights to understand the bacterial functions of diesel-degrading consortium from



diesel-polluted soil in San Fernando, Spain. Their findings showed that *Parvibaculum* contains CYP153 (family of cytochrome P450) which display hydroxylating activity towards alkanes specifically oxidation at terminal and bi-terminal positions. Interestingly, this oxidising cytochrome-like protein is only specific to this genus. This is valuable information and sheds light on how this genus possesses adaptive mechanisms allowing for its abundance in the H-C soil samples investigated in this study.

Other genera showing a considerably high level of abundance and worth mentioning is *Pseudoxanthomonas* and *Acinetobacter*. Both have been well categorised in the literature as being xenobiotic degraders (specifically benzene, toluene, ethyl-benzene and *o*-, *m*-, *p*- xylene) and colonizing various hydrocarbon contaminated environments (Patel *et al.* 2012, Bello-Akinosho *et al.* 2016, Mohapatra *et al.* 2018, Spini *et al.* 2018). The genus *Pseudoxanthomonas* was first isolated by Finkmann *et al.* (2000) from hydrocarbon contaminated biofilters used for waste gas treatment. There are a few older reports of *Pseudoxanthomonas* genera from hydrocarbon contaminated sediments such as *P. kalamensis* JA40 and *P. spadix* (Thierry *et al.* 2004, Young *et al.* 2007, Patel *et al.* 2012). More recently however, Mohapatra *et al.* (2018), showed that *Pseudoxanthomonas arseniciresistens* sp. isolated from arsenic contaminated ground water in India has the ability to utilize hydrocarbons such as BTEX, chrysene and phenanthrene. While on the other hand, Tiralerdpanich *et al.* (2018), focused on the isolation of *Acinetobacter* from a mangrove sediment and discovered their ability to degrade diesel, hexadecane and phenanthrene. These are all conclusive evidence that such genera are well documented to be hydrocarbon degraders hence resulting in the microbial shift observed in this study.

Sequencing analysis in this study has identified several bacteria that have previously been shown to be hydrocarbon degraders and have been compiled together into a cladogram illustrating the distribution of bacterial communities (**Figure 4.5**).



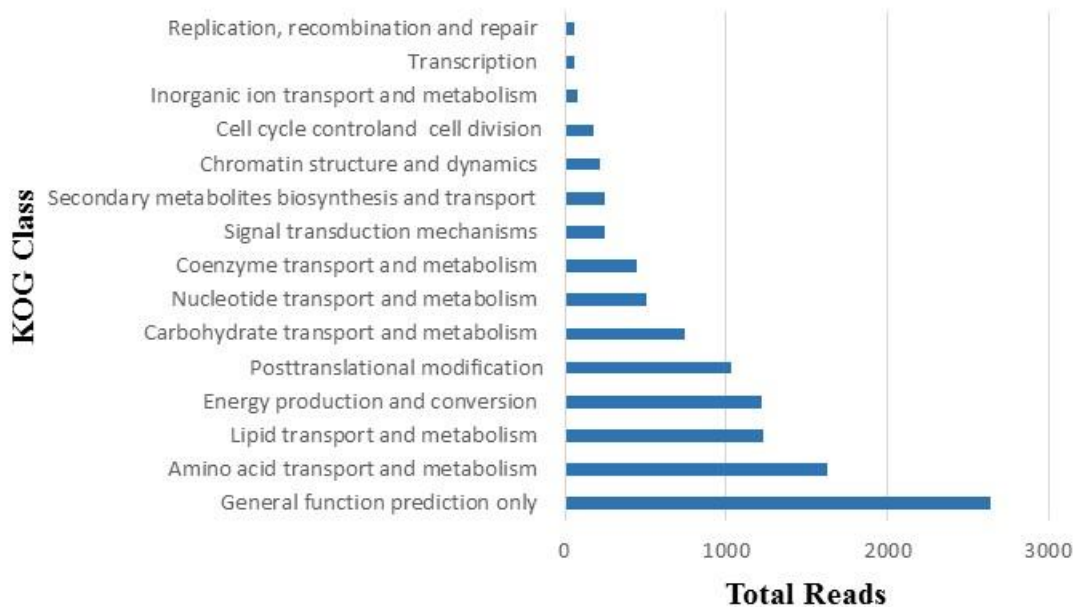
**Figure 4.5:** Taxonomic distribution of the soil microbial communities. Cladogram illustrating the distribution of bacterial communities extracted from hydrocarbon contaminated soils. The outer rings display the most dominant genera in each phylum and the height of each bar represents the abundance of operational taxonomic units (OTUs) found in each genus.

Understanding the cosmopolitan nature of these organisms and their known degradation capabilities could be used as part of the toolbox for remediation practices and predict the biodegradation potential of these bacteria working in consortium. Having a consortium of bacteria to efficiently remove hydrocarbons from an environment has proven to be successful by numerous

studies (Patowary *et al.* 2016, Elumalai *et al.* 2017, Tao *et al.* 2017, Wanapaisan *et al.* 2018) as in the case of the consortium of bacteria observed in this study. This confirms that the metagenomic DNA used to construct the fosmid library in Chapter 3 would comprise of a rich diversity of microbial DNA from hydrocarbon degraders.

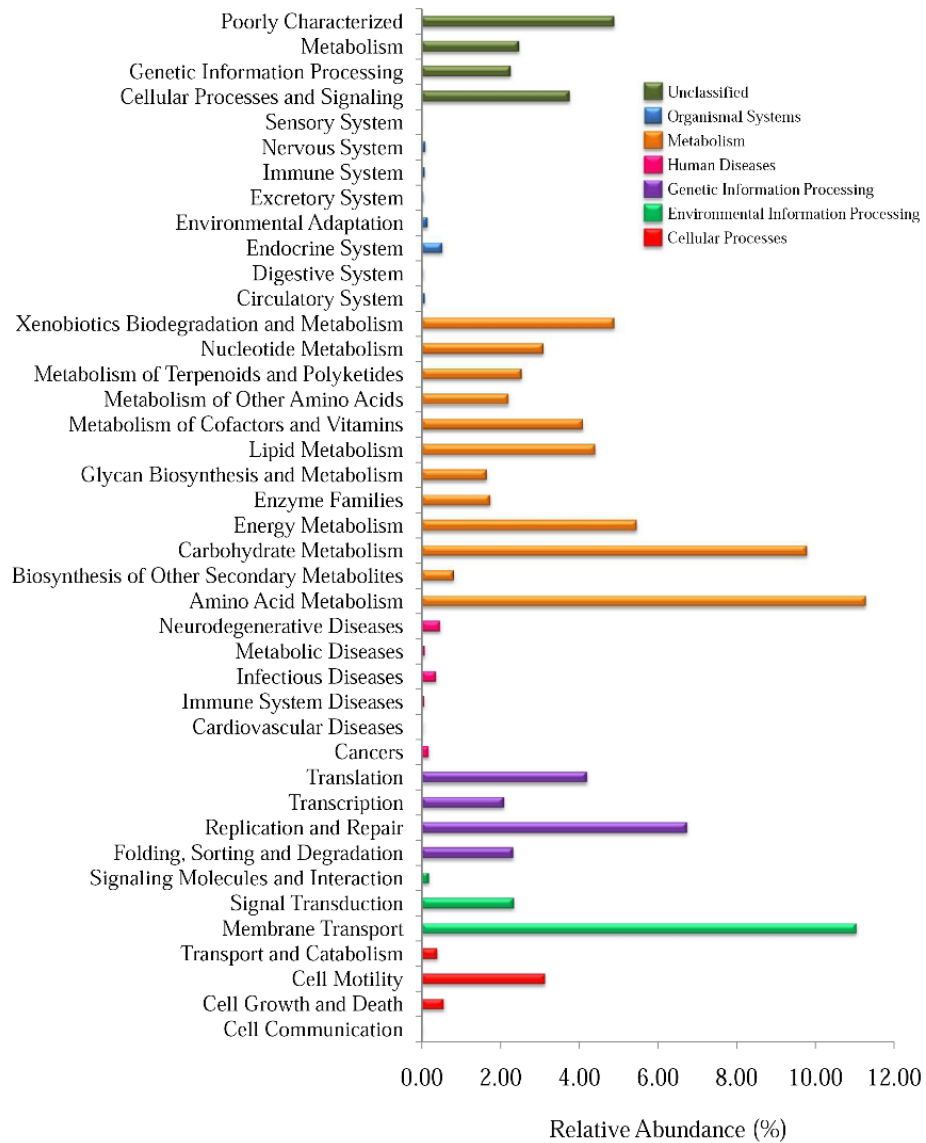
### 4.3.2 Predicted functions of bacterial communities

Functional annotation which is the assignment of the most probable biological role for a given peptide was carried out with the use of BLASTP against the most recent version of JCVI's PANDA data resource. The functional proteins based on the KOG database and the total reads of each in the hydrocarbon soil samples are shown in **Figure 4.6**. The major KOG classes depicted by a high abundance of reads were found to be transport and metabolism of amino acids (1627 reads), lipids (1228 reads), carbohydrates (745 reads), nucleotides (505 reads) co-enzymes (441 reads) and energy production and conversion (1217 reads). This data analysis has the potential to show that the metagenome capacity of this hydrocarbon contaminated soil samples was wide enough to carry out most known metabolic pathways that are essential to the overall growth and activity of the microorganisms inhabiting the soil. The majority of the genes related to these metabolic pathways are involved in routine microbial activity such as metabolism and transport of biological molecules and intracellular clustering processes. It can be inferred that these genes would mainly belong to bacteria, especially *Proteobacteria* and *Actinobacteria* which were found to be the most abundant in this study at 84 % and 8 % respectively.



**Figure 4.6:** Metabolic pathways using the KOG classes identified in the metagenomic sequences and the total number of reads for each class.

A more in-depth analysis of the putative metabolic properties of the microbial communities in the hydrocarbon contaminated soil samples were predicted using the PICRUSt analysis (**Figure 4.7**). The results further supports the information shown in **Figure 4.6** above in which metabolism and transport of biological molecules such as lipids, nucleotides, carbohydrates, co-enzymes and amino acids are in the highest relative abundance. However, it is important to note the abundance of genes involved in xenobiotic biodegradation and metabolism (6 %) which is clearly as a result of adaptation of microbial populations to high hydrocarbon concentrations and the degradation of such compounds in order to be used as a carbon source. These microbial populations are specifically bacteria and include *Proteobacteria* and *Actinobacteria* based on their high abundance levels. Research by Abbasian *et al.* (2016) on microbial diversity and hydrocarbon degrading gene capacity in crude oil field soils also showed similar results.



**Figure 4.7:** Overall predicted functional pathways of the bacterial communities extracted from hydrocarbon contaminated soils and the relative abundance of each. The functional categories were obtained based on the KEGG database (Level 2).

This finding can be further supported by the number of metagenomic sequences associated with the KEGG pathways in the categories of biodegradation of xenobiotics and carbohydrate metabolism shown in **Table 4.1**.

**Table 4.1:** Number of metagenomic sequences associated with individual KEG pathways in the categories ‘Biodegradation of xenobiotics’ and ‘Carbohydrate metabolism’ and their relative percentage abundance.

KEGG category	Reads obtained from metagenomic sequences	Percentage abundance (%)
Carbon metabolism	67929	22.1
Pyruvate metabolism	27821	9.1
Glycolysis / Gluconeogenesis	20923	6.8
Carbon fixation pathways in prokaryotes	20337	6.6
Propanoate metabolism	19332	6.3
Starch and sucrose metabolism	17646	5.7
Methane metabolism	15659	5.1
Butanoate metabolism	13336	4.3
Benzoate degradation	10775	3.5
Carbon fixation in photosynthetic organisms	10143	3.3
Degradation of aromatic compounds	9568	3.1
Pentose phosphate pathway	8862	2.9
Galactose metabolism	4950	1.6
Chloroalkane and chloroalkene degradation	4925	1.6
Aminobenzoate degradation	4711	1.5
Metabolism of xenobiotics by cytochrome P450	4337	1.4
Limonene and pinene degradation	4279	1.4
Glucagon signaling pathway	4148	1.4
Caprolactam degradation	3984	1.3
Fructose and mannose metabolism	3631	1.2
Atrazine degradation	3626	1.2
Drug metabolism - other enzymes	3404	1.1
Central carbon metabolism in cancer	3182	1.0
Synthesis and degradation of ketone bodies	2749	0.9
Naphthalene degradation	2653	0.9
Pentose and glucuronate interconversions	2443	0.8
Xylene degradation	2360	0.8
Inositol phosphate metabolism	2135	0.7
Fluorobenzoate degradation	1801	0.6
Nitrotoluene degradation	1163	0.4
Styrene degradation	1092	0.4
Dioxin degradation	1022	0.3
Chlorocyclohexane and chlorobenzene degradation	816	0.3
Glycosaminoglycan degradation	585	0.2
Toluene degradation	486	0.2
Ethylbenzene degradation	343	0.1

Notes: KEGG, Kioto Encyclopedia of Genes and Genomes.

A number of KEGG categories related to hydrocarbon metabolism and degradation have been highlighted and include the following hydrocarbons in decreasing percentage abundance: propanoate, methane, butanoate, benzoate, aromatic compounds, chloroalkane, chloroalkene, xenobiotics, limonene, pinene, naphthalene, xylene, fluorobenzoate, nitrotoluene, styrene, chlorocyclohexane, chlorobenzene, toluene and ethylbenzene. The highest percentage abundance of hydrocarbon metabolism was seen for propanoate (6.3 %). This is expected since propanoate can serve as a single carbon source for many bacteria especially *Proteobacteria* (Suvorova *et al.* 2012). Its metabolism was also shown to be strongly connected to the malonate metabolic pathway and central metabolism which includes the tricarboxylic acid (TCA) cycle where propanoate is converted to pyruvate via the methylcitrate pathway (Suvorova *et al.* 2012). However, attention was placed on methane/alkane metabolism and aromatic degradation which showed to have relatively high abundances of 5.1 % and 3.1 % respectively and is discussed in more detail in the subsequent sections.

An interesting KEGG category involving metabolism of xenobiotics by cytochrome P450 showed a percentage abundance of 1.4 % with a total of 4337 reads from the metagenomic sequences. The cytochrome P450 is classified as a super enzyme and constitutes a family of heme-thiolate monooxygenases that play a major role in the microbial degradation of oil, chlorinated hydrocarbons, fuel additives and other complex hydrocarbons under aerobic conditions (Van Beilen and Funhoff 2007, Oliveira *et al.* 2017, Al-Hawash *et al.* 2018, Unimke *et al.* 2018). There were a number of reads for chlorocyclohexane and chlorobenzene (816), fluorobenzoate (1801), chloroalkanes and chloroalkenes (4925) degradation which are classified as chlorinated hydrocarbons and with the availability of cytochrome P450 protein, it can be inferred that they are involved in their break down in the soil. Cytochrome P450 can only be found in a few bacterial communities and *Parvibaculum* is one of them (Park and Park 2018). Not surprising since it is the most abundant genus identified in the soil samples in this study. Hence, one can extrapolate that the high number of reads for chloroalkane and chloroalkene degradation could be attributed to the abundance of *Parvibaculum* sp. in the soil samples.

### 4.3.2.1 Aliphatic hydrocarbon degradation

Several enzymes related to aliphatic metabolism such as monooxygenases i.e. alkanal monooxygenase (EC:1.14.14.3), cyclohexanone monooxygenase (EC:1.14.13.22), limonene 1,2-monooxygenase (EC:1.14.13.107), alkane-1-monooxygenase (EC:1.14.15.3), methane/ammonia monooxygenase and alkanesulfonate monooxygenases were present in the metagenome (**Table 4.2**). The monooxygenase enzyme is crucial for the initial oxidation reaction of alkanes by bacterial communities in the soil (Dubbels *et al.* 2007, Heine, Tischler, *et al.* 2018). There are also a number of alcohol dehydrogenases and aldehyde dehydrogenases identified in the metagenome and these play a role in the downstream break down of alkanes following the oxidation by monooxygenases (Elumalai *et al.*, 2017; Gregson *et al.*, 2018; Li *et al.*, 2010; Liu *et al.*, 2009; Meng *et al.*, 2017). This suggests strong evidence for the involvement of bacterial monooxygenases and dehydrogenases working in synergy to achieve complete in aliphatic hydrocarbon degradation. This therefore supports the high abundance of reads from the sequenced DNA related to methane/alkane metabolism shown in **Table 4.1** above.

**Table 4.2:** Classes of enzymes involved in alkane and aromatic hydrocarbon degradation pathways with the respective EC numbers, number of metagenomic reads and open reading frames.

KO value	Class of Enzymes with ECs	Number of reads	ORFs
<b>Monooxygenases</b>			
K00494	alkanal monooxygenase alpha chain [EC:1.14.14.3]	1027	8
K03379	cyclohexanone monooxygenase [EC:1.14.13.22]	811	5
K00483	4-hydroxyphenylacetate 3-monooxygenase	429	1



	[EC:1.14.14.9]		
K14733	limonene 1,2-monooxygenase	337	2
	[EC:1.14.13.107]		
K00496	alkane 1-monooxygenase	313	3
	[EC:1.14.15.3]		
K10946	methane/ammonia monooxygenase subunit C	80	1
K04091	alkanesulfonate monooxygenase	55	1
	[EC:1.14.14.5]		
K00459	nitronate monooxygenase	648	3
	[EC:1.13.12.16]		
K06134	ubiquinone biosynthesis monooxygenase Coq7 [EC:1.14.13.-]	318	3
K00486	kynurenine 3-monooxygenase	75	1
	[EC:1.14.13.9]		

### Dehydrogenases

K00128	aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	1122	10
K04117	cyclohexanecarboxyl-CoA dehydrogenase [EC:1.3.99.-]	852	8
K01491	methylenetetrahydrofolate dehydrogenase (NADP+) / methenyltetrahydrofolate cyclohydrolase [EC:1.5.1.5 3.5.4.9]	452	3
K00100	butanol dehydrogenase [EC:1.1.1.-]	398	3
K07535	2-hydroxycyclohexanecarboxyl-CoA dehydrogenase [EC:1.1.1.- ]	237	1
K09019	3-hydroxypropanoate dehydrogenase [EC:1.1.1.-]	154	1
K00055	aryl-alcohol dehydrogenase [EC:1.1.1.90]	93	1
K00013	histidinol dehydrogenase [EC:1.1.1.23]	931	8
K03520	carbon-monoxide dehydrogenase large subunit [EC:1.2.99.2]	697	2

K03885	NADH dehydrogenase [EC:1.6.99.3]	336	2
K13774	citronellol/citronellal dehydrogenase	325	5
K00020	3-hydroxyisobutyrate dehydrogenase [EC:1.1.1.31]	664	4
K00252	glutaryl-CoA dehydrogenase [EC:1.3.8.6]	477	2
K00248	butyryl-CoA dehydrogenase [EC:1.3.8.1]	473	5
K00151	5-carboxymethyl-2-hydroxymuconic-semialdehyde dehydrogenase [EC:1.2.1.60]	417	2
K01782	3-hydroxyacyl-CoA dehydrogenase / enoyl-CoA hydratase / 3-hydroxybutyryl-CoA epimerase [EC:1.1.1.35 4.2.1.17 5.1.2.3]	386	2
K00074	3-hydroxybutyryl-CoA dehydrogenase [EC:1.1.1.157]	325	1
K03473	erythronate-4-phosphate dehydrogenase [EC:1.1.1.290]	264	2
K13953	alcohol dehydrogenase, propanol-preferring [EC:1.1.1.1]	274	3
K00114	alcohol dehydrogenase (cytochrome c) [EC:1.1.2.8]	1151	5
K00001	alcohol dehydrogenase [EC:1.1.1.1]	597	6
K00282	glycine dehydrogenase subunit 1 [EC:1.4.4.2]	591	3
K04073	acetaldehyde dehydrogenase [EC:1.2.1.10]	246	1
K15238	2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase 2 [EC:1.3.1.-]	65	1
K00123	formate dehydrogenase major subunit	258	4
K01491	methylenetetrahydrofolate dehydrogenase (NADP+) / methenyltetrahydrofolate cyclohydrolase [EC:1.5.1.5 3.5.4.9]	452	3
K10217	2-hydroxymuconate-6-semialdehyde dehydrogenase [EC:1.2.1.32 1.2.1.85]	92	1

### Dioxygenases

K03119	taurine dioxygenase [EC:1.14.11.17]	1341	6
K05549	benzoate/toluate 1,2-dioxygenase subunit alpha [EC:1.14.12.10 1.14.12.-]	1212	10

K00529	3-phenylpropionate/trans-cinnamate dioxygenase ferredoxin reductase component [EC:1.18.1.3]	665	3
K14583	1,2-dihydroxynaphthalene dioxygenase [EC:1.13.11.56]	515	1
K05708	3-phenylpropionate/trans-cinnamate dioxygenase subunit alpha [EC:1.14.12.19]	485	2
K08689	biphenyl 2,3-dioxygenase subunit alpha [EC:1.14.12.18]	215	3
K05550	benzoate/toluate 1,2-dioxygenase subunit beta [EC:1.14.12.10 1.14.12.-]	196	2
K00457	4-hydroxyphenylpyruvate dioxygenase [EC:1.13.11.27]	112	1
K08967	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase [EC:1.13.11.53 1.13.11.54]	110	1
K10619	p-cumate 2,3-dioxygenase subunit alpha [EC:1.14.12.-]	62	1
K05713	2,3-dihydroxyphenylpropionate 1,2-dioxygenase [EC:1.13.11.16]	53	1
K00446	catechol 2,3-dioxygenase [EC:1.13.11.2]	56	1

### Hydroxylases

K03185	2-octaprenyl-6-methoxyphenol hydroxylase [EC:1.14.13.-]	602	3
K05712	3-(3-hydroxy-phenyl)propionate hydroxylase [EC:1.14.13.127]	410	1

### Dehalogenases

K01563	haloalkane dehalogenase [EC:3.8.1.5]	302	1
--------	--------------------------------------	-----	---

### Transferases

K02548	1,4-dihydroxy-2-naphthoate octaprenyltransferase [EC:2.5.1.74 2.5.1.-]	58	1
K00568	2-polyprenyl-6-hydroxyphenyl methylase / 3-demethylubiquinone-9 3-methyltransferase [EC:2.1.1.222 2.1.1.64]	360	3
K03179	4-hydroxybenzoate polyprenyltransferase [EC:2.5.1.39]	480	4

### Reductases

K00297	methylenetetrahydrofolate reductase (NADPH) [EC:1.5.1.20]	531	4
K03527	4-hydroxy-3-methylbut-2-enyl diphosphate reductase [EC:1.17.1.2]	528	5
<b>Synthases</b>			
K00574	cyclopropane-fatty-acyl-phospholipid synthase [EC:2.1.1.79]	910	4
K03526	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase [EC:1.17.7.1 1.17.7.3]	594	4
<b>Synthetase</b>			
K01665	para-aminobenzoate synthetase component I [EC:2.6.1.85]	63	1
<b>Hydratases</b>			
K02509	2-oxo-hept-3-ene-1,7-dioate hydratase [EC:4.2.1.-]	445	2
<b>Aldolases</b>			
K02510	4-hydroxy-2-oxoheptanedioate aldolase [EC:4.1.2.52]	444	2
<b>Decarboxylase</b>			
K03182	4-hydroxy-3-polyprenylbenzoate decarboxylase [EC:4.1.1.98]	336	1
K01607	4-carboxymuconolactone decarboxylase [EC:4.1.1.44]	83	1
K01617	2-oxo-3-hexenedioate decarboxylase [EC:4.1.1.77]	62	1
<b>Hydrolase</b>			
K05714	2-hydroxy-6-oxonona-2,4-dienedioate hydrolase [EC:3.7.1.14]	293	1
<b>Ligases</b>			
K04116	cyclohexanecarboxylate-CoA ligase [EC:6.2.1.-]	500	6
K01906	6-carboxyhexanoate--CoA ligase [EC:6.2.1.14]	63	1
<b>Others</b>			
K01061	carboxymethylenebutenolidase [EC:3.1.1.45]	393	2
K08195	MFS transporter, AAHS family, 4-hydroxybenzoate transporter	85	1

Notes: Ec, Enzyme commission number.

The involvement of monooxygenase genes are the key and rate-limiting enzymes for the aerobic breakdown of hydrocarbons and is a well understood mechanism based on many research outputs through the years (Liang *et al.* 2015, Bouhajja *et al.* 2017, Liu *et al.* 2018). This reaction involves introducing oxygen atom(s) into alkane substrates to yield a primary alcohol which is further oxidised by alcohol dehydrogenase and aldehyde dehydrogenase. Research has shown that the *alkB* like genes (alkane monooxygenases) are widely distributed in *Proteobacteria* and *Actinobacteria* (Fuentes *et al.* 2014). Thus, correlating the abundance of this genera with the high number of reads associated with alkane monooxygenases in this study. The monooxygenases identified and their respective number of reads from the metagenomic DNA include, cyclohexanone monooxygenase (811 reads), alkanal monooxygenase (1027 reads), 4-hydroxyphenylacetate 3-monooxygenase (429 reads), limonene 1,2-monooxygenase (337 reads), alkane 1-monooxygenase (313 reads), methane/ammonia monooxygenase subunit C (80 reads) and alkanesulfonate monooxygenase (55 reads). Hence, based on the variety and number of sequence reads identified for monooxygenase involved in alkane degradation, one can conclude that bacterial species co-existing in this soil environment are well equipped and evolved to degrade alkanes.

Two enzymes that need to be highlighted are methane/ammonia monooxygenase subunit C monooxygenase and alkane-1-monooxygenase due to their value in alkane degradation and the volumes of research conducted on them from different bacterial species. Methane monooxygenase (MMO) is involved in the catabolism of methane to form methanol, which is further oxidised to formaldehyde by methanol dehydrogenase and then to formate by formate dehydrogenase which is used by the cells as an energy source (Ji *et al.* 2013). A number of alcohol dehydrogenases have been identified from the metagenomic sequences and particularly a formate dehydrogenase (EC:1.2.1.2). This is an indication that the enzymatic pathway for the catabolism of methane was carried out by the bacterial consortia inhabiting the soil samples which could represent methanotrophic bacteria.

With respect to alkane-1-monooxygenase, they are classified as an important hydroxylases responsible for aerobic alkane degradation in bioremediation of oil-polluted environments next to CYP153 (Nie *et al.* 2014). These are integral-membrane non-heme di-iron monooxygenases that hydroxylates alkanes to alcohols at a terminal position (Van Beilen and Funhoff 2007, Mukherjee *et al.* 2017). There are many bacterial strains that possess this enzyme and have been found to degrade intermediate C<sub>5</sub>-C<sub>11</sub> and long C<sub>12</sub> *n*-alkanes (Rojo 2009, Hesham *et al.* 2014, Sun *et al.* 2014, Chen *et al.* 2017). According to Nie *et al.* (2014), hundreds of *alkB* and CYP153 genes most of which were novel showing low sequence identities to known genes were retrieved by mining metagenome databases. These were predominantly found to be from *Proteobacteria* and *Actinobacteria*. Although alkane hydroxylases have been extensively studied since they were first identified in 1977 (Benson *et al.* 1977), hydroxylases are complex enzymes and purification of all the components of a soluble three-component diiron with physiologically relevant activity has proven to be very difficult (Wang *et al.*, 2017).

To further confirm the aerobic breakdown of alkanes occurring in soil samples by native organisms, a number of metagenomic sequence reads were identified for alcohol and aldehyde dehydrogenase which are involved in the downstream breakdown of alkanes following the oxidation by hydroxylases. From the analysed metagenomic data the read abundance for alcohol dehydrogenase includes butanol dehydrogenase [EC:1.1.1.-] (398 reads), alcohol dehydrogenase (cytochrome c) [EC:1.1.2.8] (1151 reads) alcohol dehydrogenase [EC:1.1.1.1] (597 reads), propanol dehydrogenase [EC:1.1.1.1] (274 reads) and 2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase 2 [EC:1.3.1.-] (65 reads). The read abundance for aldehyde dehydrogenase includes aldehyde dehydrogenase (NAD<sup>+</sup>) [EC:1.2.1.3] (1122 reads), NADH dehydrogenase [EC:1.6.99.3] (336 reads) and acetaldehyde dehydrogenase [EC:1.2.1.10] (246 reads).

#### 4.3.2.2 Aromatic hydrocarbon degradation

The analysis of the metagenomic data revealed the presence of several enzymes involved in both aerobic and anaerobic degradation of aromatic hydrocarbons such as benzoate, toluene, phenol and polycyclic aromatic, naphthalene. For aerobic metabolism such monooxygenases these included dioxygenases such as 4-hydroxyphenylacetate 3-monooxygenase [EC:1.14.14.9], 2-octaprenyl-6-methoxyphenol hydroxylase [EC:1.14.13.-], 3-(3-hydroxy-phenyl)propionate hydroxylase [EC:1.14.13.127], benzoate/toluene 1,2-dioxygenase subunit alpha [EC:1.14.12.10 1.14.12.-], benzoate/toluene 1,2-dioxygenase subunit beta [EC:1.14.12.10 1.14.12.-], -phenylpropionate/trans-cinnamate dioxygenase ferredoxin reductase component [EC:1.18.1.3], 3-phenylpropionate/trans-cinnamate dioxygenase subunit alpha [EC:1.14.12.19], 1,2-dihydroxynaphthalene dioxygenase [EC:1.13.11.56], biphenyl 2,3-dioxygenase subunit alpha [EC:1.14.12.18], 2,3-dihydroxyphenylpropionate 1,2-dioxygenase [EC:1.13.11.16] and catechol 2,3-dioxygenase [EC:1.13.11.2].

Apart from the monooxygenases, contig analysis also showed the presence of several other enzymes in this metagenome such as aryl-alcohol dehydrogenase [EC:1.1.1.90]. These are a group of enzymes with broad specificity towards primary alcohols with an aromatic or cyclohex-1-ene ring, but with low or no activity towards short-chain aliphatic alcohols (Yang *et al.* 2018). Other enzymes include 5-carboxymethyl-2-hydroxymuconic-semialdehyde dehydrogenase [EC:1.2.1.60], para-aminobenzoate synthetase component I [EC:2.6.1.85], glutaryl-CoA dehydrogenase [EC:1.3.8.6], 2-hydroxy-6-oxonona-2,4-dienedioate hydrolase [EC:3.7.1.14], 4-carboxymuconolactone decarboxylase [EC:4.1.1.44], 2-hydroxymuconate-6-semialdehyde dehydrogenase [EC:1.2.1.32 1.2.1.85] and 2-oxo-3-hexenedioate decarboxylase [EC:4.1.1.77] indicating the presence of near-complete degradation pathways for aromatics. The occurrence of aerobic degradation of aromatic compounds can be supported by the presence of the genera *Pseudoxanthomonas* (Bello-Akinosho *et al.* 2016, Lu *et al.* 2019) *Pseudomonas* (Suyama *et al.* 1996, Chebbi *et al.* 2017a, 2017b, Wang *et al.* 2017, Deshpande *et al.* 2018) and *Burkholderia* (Silva *et al.* 2013, Lee *et al.* 2016) identified in this study and shown to be participants in aromatic hydrocarbon degradation.

#### 4.4 Conclusion

It can be concluded that the application of next generation sequencing (NGS) methods can be a powerful tool in metagenomic studies. Thus, opening new doors of understanding by expanding our mind to the future uses of bacterial consortia as well as enzyme cocktails that can be developed to completely break down harmful pollutants in the soil environment. Lastly, a number of monooxygenases targeting different substrates were identified and based on literature it can be concluded that isolating and purifying these proteins proves to be a challenge since many fall in the class of particulate monooxygenases. We are able to identify a range of enzyme classes known to be involved in specific hydrocarbon pathways using *in silico* analyses of the metagenomic pool including aldehyde dehydrogenases, alcohol dehydrogenases, decarboxylases, hydrolases, dehalogenases, reductases, transferases, synthases and hydrolases. Therefore, the identification of a number of hydrocarbon-degrading enzymes during *in silico* analysis further confirms the richness of the fosmid library that was constructed from the same pool as discussed in the Chapter 3.



## 4.5 References

- Abbasian, F., Palanisami, T., Megharaj, M., Naidu, R., Lockington, R., and Ramadass, K., 2016. Microbial diversity and hydrocarbon degrading gene capacity of a crude oil field soil as determined by metagenomics analysis. *Biotechnology progress*, 32 (3), 638–648.
- Al-Hawash, A.B., Dragh, M.A., Li, S., Alhujaily, A., Abbood, H.A., Zhang, X., and Ma, F., 2018. Principles of microbial degradation of petroleum hydrocarbons in the environment. *Egyptian Journal of Aquatic Research*, 44 (2), 71–76.
- Andler, R., Hiessl, S., Yücel, O., Tesch, M., and Steinbüchel, A., 2018. Cleavage of poly(cis-1,4-isoprene) rubber as solid substrate by cultures of *Gordonia polyisoprenivorans*. *New Biotechnology*, 44 (March), 6–12.
- Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C., and Segata, N., 2015. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ*, 2015 (6), 1–17.
- Barone, R., De Santi, C., Palma Esposito, F., Tedesco, P., Galati, F., Visone, M., Di Scala, A., and De Pascale, D., 2014. Marine metagenomics, a valuable tool for enzymes and bioactive compounds discovery. *Frontiers in Marine Science*, 1 (38), 1–6.
- Van Beilen, J.B. and Funhoff, E.G., 2007. Alkane hydroxylases involved in microbial alkane degradation. *Applied Microbiology and Biotechnology*, 74 (1), 13–21.
- Bello-Akinosho, M., Makofane, R., Adeleke, R., Thantsha, M., Pillay, M., and Chirima, G.J., 2016. Potential of Polycyclic Aromatic Hydrocarbon-Degrading Bacterial Isolates to Contribute to Soil Fertility. *BioMed Research International*, 2016.
- Benson, S., Fennewald, M., Shapiro, J., and Huettner, C., 1977. Fractionation of inducible alkane hydroxylase activity in *Pseudomonas putida* and characterization of hydroxylase-negative plasmid mutations. *Journal of Bacteriology*, 132 (2), 614–621.
- Bouhajja, E., McGuire, M., Liles, M.R., Bataille, G., Agathos, S.N., and George, I.F., 2017. Identification of novel toluene monooxygenase genes in a hydrocarbon-polluted sediment using sequence- and function-based screening of metagenomic libraries. *Applied*

*Microbiology and Biotechnology*, 101 (2), 797–808.

- Brzeszcz, J. and Kaszycki, P., 2018. Aerobic bacteria degrading both n-alkanes and aromatic hydrocarbons: an undervalued strategy for metabolic diversity and flexibility. *Biodegradation*, 29 (4), 359–407.
- Cai, M., Nie, Y., Chi, C.Q., Tang, Y.Q., Li, Y., Wang, X.B., Liu, Z.S., Yang, Y., Zhou, J., and Wu, X.L., 2015. Crude oil as a microbial seed bank with unexpected functional potentials. *Scientific Reports*, 5, 1–12.
- Chebbi, A., Hentati, D., Zaghden, H., Baccar, N., Rezgui, F., Chalbi, M., Sayadi, S., and Chamkha, M., 2017a. Polycyclic aromatic hydrocarbon degradation and biosurfactant production by a newly isolated *Pseudomonas* sp. strain from used motor oil-contaminated soil. *International Biodeterioration and Biodegradation*, 122, 128–140.
- Chebbi, A., Hentati, D., Zaghden, H., Baccar, N., Rezgui, F., Chalbi, M., Sayadi, S., and Chamkha, M., 2017b. Polycyclic aromatic hydrocarbon degradation and biosurfactant production by a newly isolated *Pseudomonas* sp. strain from used motor oil-contaminated soil. *International Biodeterioration and Biodegradation*, 122.
- Chen, W., Li, J., Sun, X., Min, J., and Hu, X., 2017. High efficiency degradation of alkanes and crude oil by a salt-tolerant bacterium *Dietzia* species CN-3. *International Biodeterioration and Biodegradation*, 118, 110–118.
- Crispim, J.S., Dias, R.S., Vidigal, P.M.P., De Sousa, M.P., Da Silva, C.C., Santana, M.F., and De Paula, S.O., 2018. Screening and characterization of prophages in *Desulfovibrio* genomes. *Scientific Reports*, 8 (1), 1–10.
- DeCastro, M.E., Rodríguez-Belmonte, E., and González-Siso, M.I., 2016. Metagenomics of thermophiles with a focus on discovery of novel thermozyms. *Frontiers in Microbiology*, 7 (SEP), 1–21.
- Deshpande, R.S., Sundaravadivelu, D., Techtmann, S., Conmy, R.N., Santo Domingo, J.W., and Campo, P., 2018. Microbial degradation of Cold Lake Blend and Western Canadian select dilbits by freshwater enrichments. *Journal of Hazardous Materials*, 352 (March), 111–120.

- Dombrowski, N., Donaho, J.A., Gutierrez, T., Seitz, K.W., Teske, A.P., and Baker, B.J., 2016. Reconstructing metabolic pathways of hydrocarbon-degrading bacteria from the Deepwater Horizon oil spill. *Nature Microbiology*, 1 (May), 1–8.
- Duarte, M., Nielsen, A., Camarinha-Silva, A., Vilchez-Vargas, R., Bruls, T., Wos-Oxley, M.L., Jauregui, R., and Pieper, D.H., 2017. Functional soil metagenomics: elucidation of polycyclic aromatic hydrocarbon degradation potential following 12 years of in situ bioremediation. *Environmental Microbiology*.
- Dubbels, B.L., Sayavedra-Soto, L.A., and Arp, D.J., 2007. Butane monooxygenase of ‘*Pseudomonas butanovora*’: Purification and biochemical characterization of a terminal-alkane hydroxylating diiron monooxygenase. *Microbiology*, 153 (6), 1808–1816.
- Elumalai, P., Parthipan, P., Karthikeyan, O.P., and Rajasekar, A., 2017. Enzyme-mediated biodegradation of long-chain n-alkanes (C32 and C40) by thermophilic bacteria. *3 Biotech*, 7 (2), 1–10.
- Fan, S., Wang, J., Li, K., Yang, T., Jia, Y., Zhao, B., and Yan, Y., 2018. Complete genome sequence of *Gordonia* sp. YC-JH1, a bacterium efficiently degrading a wide range of phthalic acid esters. *Journal of Biotechnology*, 279 (May), 55–60.
- Finkmann, W., Altendorf, K., Stackebrandt, E., and Lipski, A., 2000. Characterization of N<sub>2</sub>O-producing Xanthomonas-like isolates from biofilters as *Stenotrophomonas nitritireducens* sp. nov., *Luteimonas mephitis* gen. nov., sp. nov. and *Pseudoxanthomonas broegbernensis* gen. nov., sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 50 (1), 273–282.
- Fuentes, S., Méndez, V., Aguila, P., and Seeger, M., 2014. Bioremediation of petroleum hydrocarbons: Catabolic genes, microbial communities, and applications. *Applied Microbiology and Biotechnology*, 98 (11), 4781–4794.
- Ganesh Kumar, A., Vijayakumar, L., Joshi, G., Magesh Peter, D., Dharani, G., and Kirubakaran, R., 2014. Biodegradation of complex hydrocarbons in spent engine oil by novel bacterial consortium isolated from deep sea sediment. *Bioresource Technology*, 170, 556–564.

- Garrido-Sanz, D., Redondo-Nieto, M., Guirado, M., Pindado Jiménez, O., Millán, R., Martín, M., and Rivilla, R., 2019. Metagenomic Insights into the Bacterial Functions of a Diesel-Degrading Consortium for the Rhizoremediation of Diesel-Polluted Soil. *Genes*, 10 (6), 456.
- Godini, K., Samarghandi, M.R., Zafari, D., Rahmani, A.R., Afkhami, A., and Arabestani, M.R., 2018. Isolation and identification of new strains of crude oil degrading bacteria from Kharg Island, Iran. *Petroleum Science and Technology*, 36 (12), 869–874.
- Gregson, Benjamin H. McKew, B.A., Metodiev, M. V., Metodieva, G., and Golyshin, P.N., 2018. Differential Protein Expression During Growth on Medium Versus Long-Chain Alkanes in the Obligate Marine Hydrocarbon-Degrading Bacterium *Thalassolituus oleivorans* MIL-1. *Frontiers in Microbiology*, 9 (December), 1–14.
- Guerra, A.B., Oliveira, J.S., Silva-Portela, R.C.B., Araújo, W., Carlos, A.C., Vasconcelos, A.T.R., Freitas, A.T., Domingos, Y.S., de Farias, M.F., Fernandes, G.J.T., and Agnez-Lima, L.F., 2018. Metagenome enrichment approach used for selection of oil-degrading bacteria consortia for drill cutting residue bioremediation. *Environmental Pollution*, 235, 869–880.
- Handelsman, J., 2005. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*, 69 (1), 195–195.
- Heine, T., Tischler, D., van Berkel, W., van Pée, K.-H., and Gassner, G., 2018. Two-Component FAD-Dependent Monooxygenases: Current Knowledge and Biotechnological Opportunities. *Biology*, 7 (3), 42.
- Heine, T., Zimmerling, J., Ballmann, A., Kleeberg, S.B., Busche, T., Winkler, A., Poetsch, A., Scholtissek, A., Schmidt, G., and Tischler, D., 2018. crossm Degradation in *Gordonia rubripertincta* CWB2, 1–16.
- Hesham, A.E.L., Mawad, A.M.M., Mostafa, Y.M., and Shoreit, A., 2014. Biodegradation ability and catabolic genes of petroleum-degrading *Sphingomonas koreensis* strain ASU-06 isolated from Egyptian oily soil. *BioMed Research International*, 2014.
- Ji, L., Fu, X., Wang, M., Xu, C., Chen, G., Song, F., Guo, S., and Zhang, Q., 2019. Enzyme cocktail containing NADH regeneration system for efficient bioremediation of oil sludge

- contamination. *Chemosphere*, 233, 132–139.
- Ji, Y., Mao, G., Wang, Y., and Bartlam, M., 2013. Structural insights into diversity and n-alkane biodegradation mechanisms of alkane hydroxylases. *Frontiers in Microbiology*, 4 (March), 1–13.
- Jin, D., Kong, X., Jia, M., Yu, X., Wang, X., Zhuang, X., Deng, Y., and Bai, Z., 2017. *Gordonia phthalatica* sp. nov., a di-n-butyl phthalate-degrading bacterium isolated from activated sludge. *International Journal of Systematic and Evolutionary Microbiology*, 67 (12), 5128–5133.
- Joshi, M.N., Dhebar, S. V., Dhebar, S. V., Bhargava, P., Pandit, A., Patel, R.P., Saxena, A., and Bagatharia, S.B., 2014. Metagenomics of petroleum muck: Revealing microbial diversity and depicting microbial syntrophy. *Archives of Microbiology*, 196 (8), 531–544.
- Kadri, T., Magdoui, S., Rouissi, T., and Brar, S.K., 2018. Ex-situ biodegradation of petroleum hydrocarbons using *Alcanivorax borkumensis* enzymes. *Biochemical Engineering Journal*, 132, 279–287.
- Kostka, J.E., Prakash, O., Overholt, W.A., Green, S.J., Freyer, G., Canion, A., Delgado, J., Norton, N., Hazen, T.C., and 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), 7962–7974.
- Kumar, S., Suyal, D.C., Yadav, A., Shouche, Y., and Goel, R., 2019. Microbial diversity and soil physiochemical characteristic of higher altitude. *PLoS ONE*, 14 (3), 1–15.
- Kusnezowa, A. and Leichert, L.I., 2017. In silico approach to designing rational metagenomic libraries for functional studies. *BMC Bioinformatics*, 18 (1), 1–11.
- Lam, K.N., Cheng, J., Engel, K., Neufeld, J.D., and Charles, T.C., 2015. Current and future resources for functional metagenomics. *Frontiers in Microbiology*, 6 (OCT), 1–8.
- Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepille, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G., and Huttenhower, C., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene

- sequences. *Nature Biotechnology*, 31 (9), 814–821.
- Lee, S.Y., Kim, G.H., Yun, S.H., Choi, C.W., Yi, Y.S., Kim, J., Chung, Y.H., Park, E.C., and Kim, S. II, 2016. Correction: Proteogenomic characterization of monocyclic aromatic hydrocarbon degradation pathways in the aniline-degrading bacterium *Burkholderia* sp. K24. *PLoS ONE*, 11 (6), 1–14.
- Li, X., Li, Y., Wei, D., Li, P., Wang, L., and Feng, L., 2010. Characterization of a broad-range aldehyde dehydrogenase involved in alkane degradation in *Geobacillus thermodenitrificans* NG80-2. *Microbiological Research*, 165 (8), 706–712.
- Liang, J.-L., Wang, M., Nie, Y., Wu, X.-L., Maser, E., Xiong, G., and Wang, Y.-P., 2015. Regulation of alkane degradation pathway by a TetR family repressor via an autoregulation positive feedback mechanism in a Gram-positive *Dietzia* bacterium. *Molecular Microbiology*, 99 (2), 338–359.
- Liu, X., Dong, Y., Zhang, J., Zhang, A., Wang, L., and Feng, L., 2009. Two novel metal-independent long-chain alkyl alcohol dehydrogenases from *Geobacillus thermodenitrificans* NG80-2. *Microbiology*, 155 (6), 2078–2085.
- Liu, Y., Ding, A., Sun, Y., Xia, X., and Zhang, D., 2018. Impacts of n-alkane concentration on soil bacterial community structure and alkane monooxygenase genes abundance during bioremediation processes. *Frontiers of Environmental Science and Engineering*, 12 (5), 1–13.
- Lu, C., Hong, Y., Liu, J., Gao, Y., Ma, Z., Yang, B., Ling, W., and Waigi, M.G., 2019. A PAH-degrading bacterial community enriched with contaminated agricultural soil and its utility for microbial bioremediation. *Environmental Pollution*, 251, 773–782.
- Lu, Z., Deng, Y., Van Nostrand, J.D., He, Z., Voordeckers, J., Zhou, A., Lee, Y.J., Mason, O.U., Dubinsky, E.A., Chavarria, K.L., Tom, L.M., Fortney, J.L., Lamendella, R., Jansson, J.K., D’Haeseleer, P., Hazen, T.C., and Zhou, J., 2012. Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. *ISME Journal*, 6 (2), 451–460.
- M D Yuniati, 2018. Bioremediation of petroleum-contaminated soil : A Review Bioremediation

- of petroleum-contaminated soil: A Review. *IOP Conference Series: Earth and Environmental Science*, 118 (1).
- Meng, L., Li, H., Bao, M., and Sun, P., 2017. Metabolic pathway for a new strain *Pseudomonas synxantha* LSH-7<sup>T</sup>: From chemotaxis to uptake of n-hexadecane. *Scientific Reports*, 7 (January), 1–13.
- Miller, J.I., Techtmann, S., Fortney, J., Mahmoudi, N., Joyner, D., Liu, J., Olesen, S., Alm, E., Fernandez, A., Gardinali, P., GaraJayeva, N., Askerov, F.S., and Hazen, T.C., 2019. Oil hydrocarbon degradation by caspian sea microbial communities. *Frontiers in Microbiology*, 10 (MAY), 1–15.
- Mohapatra, B., Sar, P., Kazy, S.K., Maiti, M.K., and Satyanarayana, T., 2018. Taxonomy and physiology of *pseudoxanthomonas arseniciresistens* sp. nov., an arsenate and nitrate-reducing novel gammaproteobacterium from arsenic contaminated groundwater, India. *PLoS ONE*, 13 (3), 1–18.
- Mukherjee, A.K., Bhagowati, P., Biswa, B.B., Chanda, A., and Kalita, B., 2017. A comparative intracellular proteomic profiling of *Pseudomonas aeruginosa* strain ASP-53 grown on pyrene or glucose as sole source of carbon and identification of some key enzymes of pyrene biodegradation pathway. *Journal of Proteomics*, 167, 25–35.
- Nie, Y., Chi, C.Q., Fang, H., Liang, J.L., Lu, S.L., Lai, G.L., Tang, Y.Q., and Wu, X.L., 2014. Diverse alkane hydroxylase genes in microorganisms and environments. *Scientific Reports*, 4, 1–11.
- Oliveira, J.S., Araújo, W.J., Figueiredo, R.M., Silva-Portela, R.C.B., De Brito Guerra, A., Da Silva Araújo, S.C., Minnicelli, C., Carlos, A.C., De Vasconcelos, A.T.R., Freitas, A.T., and Agnez-Lima, L.F., 2017. Biogeographical distribution analysis of hydrocarbon degrading and biosurfactant producing genes suggests that near-equatorial biomes have higher abundance of genes with potential for bioremediation. *BMC Microbiology*, 17 (1).
- Ondov, B.D., Bergman, N.H., and Phillippy, A.M., 2011. Interactive metagenomic visualization in a Web browser Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, 12 (1), 385.

- Pannekens, M., Kroll, L., Müller, H., Mbow, F.T., and Meckenstock, R.U., 2019. Oil reservoirs, an exceptional habitat for microorganisms. *New Biotechnology*, 49 (November 2018), 1–9.
- Park, C. and Park, W., 2018. Survival and energy producing strategies of Alkane degraders under extreme conditions and their biotechnological potential. *Frontiers in Microbiology*, 9 (MAY), 1–15.
- Patel, V., Cheturvedula, S., and Madamwar, D., 2012. Phenanthrene degradation by *Pseudoxanthomonas* sp. DMVP2 isolated from hydrocarbon contaminated sediment of Amlakhadi canal, Gujarat, India. *Journal of Hazardous Materials*, 201–202, 43–51.
- Patowary, K., Patowary, R., Kalita, M.C., and Deka, S., 2016. Development of an efficient bacterial consortium for the potential remediation of hydrocarbons from contaminated sites. *Frontiers in Microbiology*, 7 (JUL), 1–14.
- Pedron, R., Esposito, A., Bianconi, I., Pasolli, E., Tett, A., Asnicar, F., Cristofolini, M., Segata, N., and Jousson, O., 2019. Genomic and metagenomic insights into the microbial community of a thermal spring. *Microbiome*, 7 (1), 1–13.
- Prince, R. and Mcgenity, T.J., 2010. Handbook of Hydrocarbon and Lipid Microbiology. *Handbook of Hydrocarbon and Lipid Microbiology*, (January).
- Ren, M., Zhang, Z., Wang, X., Zhou, Z., Chen, D., Zeng, H., Zhao, S., Chen, L., Hu, Y., Zhang, C., Liang, Y., She, Q., Zhang, Y., and Peng, N., 2018. Diversity and contributions to nitrogen cycling and carbon fixation of soil salinity shaped microbial communities in Tarim Basin. *Frontiers in Microbiology*, 9 (MAR), 1–14.
- Rodgers, T. and Zhang, H., 2019. Recent Advances in Function-based Metagenomic Screening. *Genomics, Proteomics & Bioinformatics*, 16 (6), 405–415.
- Rojo, F., 2009. Degradation of alkanes by bacteria: Minireview. *Environmental Microbiology*, 11 (10), 2477–2490.
- Rosario-Passapera, R., Keddis, R., Wong, R., Lutz, R.A., Starovoytov, V., and Vetriani, C., 2012. *Parvibaculum hydrocarboniclasticum* sp. nov., a mesophilic, alkane-oxidizing alphaproteobacterium isolated from a deep-sea hydrothermal vent on the East Pacific Rise.



- International Journal of Systematic and Evolutionary Microbiology*, 62 (12), 2921–2926.
- Rosenberg, E., Bittan-Banin, G., Sharon, G., Shon, A., Hershko, G., Levy, I., and Ron, E.Z., 2010. The phage-driven microbial loop in petroleum bioremediation. *Microbial Biotechnology*, 3 (4), 467–472.
- Roy, A., Dutta, A., Pal, S., Gupta, A., Sarkar, J., Chatterjee, A., Saha, A., Sarkar, P., Sar, P., and Kazy, S.K., 2018. Biostimulation and bioaugmentation of native microbial community accelerated bioremediation of oil refinery sludge. *Bioresource Technology*, 253.
- Roy, A., Sar, P., Sarkar, J., Dutta, A., Sarkar, P., Gupta, A., Mohapatra, B., Pal, S., and Kazy, S.K., 2018. Petroleum hydrocarbon rich oil refinery sludge of North-East India harbours anaerobic, fermentative, sulfate-reducing, syntrophic and methanogenic microbial populations. *BMC Microbiology*, 18 (1), 1–22.
- S, R., T, D., V, V., and R, S., 2014. Comparative Studies on the Extraction of Metagenomic DNA from Various Soil and Sediment Samples of Jammu and Kashmir Region in Prospect for Novel Biocatalysts. *IOSR Journal of Environmental Science, Toxicology and Food Technology*, 8 (4), 46–56.
- Sarkar, J., Kazy, S.K., Gupta, A., Dutta, A., Mohapatra, B., Roy, A., Bera, P., Mitra, A., and Sar, P., 2016a. Biostimulation of indigenous microbial community for bioremediation of petroleum refinery sludge. *Frontiers in Microbiology*, 7 (SEP), 1–20.
- Sarkar, J., Kazy, S.K., Gupta, A., Dutta, A., Mohapatra, B., Roy, A., Bera, P., Mitra, A., and Sar, P., 2016b. Biostimulation of indigenous microbial community for bioremediation of petroleum refinery sludge. *Frontiers in Microbiology*, 7 (SEP).
- Schleheck, D., Tindall, B.J., Rosselló-Mora, R., and Cook, A.M., 2004. Parvibaculum lavamentivorans gen. nov., sp. nov., a novel heterotroph that initiates catabolism of linear alkylbenzenesulfonate. *International Journal of Systematic and Evolutionary Microbiology*, 54 (5), 1489–1497.
- Schleheck, D., Weiss, M., Pitluck, S., Bruce, D., Land, M.L., Han, S., Saunders, E., Tapia, R., Detter, C., Brettin, T., Han, J., Woyke, T., Goodwin, L., Pennacchio, L., Nolan, M., Cook,

- A.M., Kjelleberg, S., and Thomas, T., 2011. Complete genome sequence of *Parvibaculum lavamentivorans* type strain (DS-1 1). *Standards in Genomic Sciences*, 5 (3), 298–310.
- Silva, C.C., Hayden, H., Sawbridge, T., Mele, P., De Paula, S.O., Silva, L.C.F., Vidigal, P.M.P., Vicentini, R., Sousa, M.P., Torres, A.P.R., Santiago, V.M.J., and Oliveira, V.M., 2013. Identification of Genes and Pathways Related to Phenol Degradation in Metagenomic Libraries from Petroleum Refinery Wastewater. *PLoS ONE*, 8 (4), 1–11.
- Spini, G., Spina, F., Poli, A., Blieux, A.-L., Regnier, T., Gramellini, C., Varese, G.C., and Puglisi, E., 2018. Molecular and Microbiological Insights on the Enrichment Procedures for the Isolation of Petroleum Degrading Bacteria and Fungi. *Frontiers in Microbiology*, 9 (October).
- Stefani, F.O.P., Bell, T.H., Marchand, C., De La Providencia, I.E., El Yassimi, A., St-Arnaud, M., and Hijri, M., 2015. Culture-dependent and -independent methods capture different microbial community fractions in hydrocarbon-contaminated soils. *PLoS ONE*, 10 (6), 1–16.
- Sun, J.Q., Xu, L., Zhang, Z., Li, Y., Tang, Y.Q., and Wu, X.L., 2014. Diverse bacteria isolated from microtherm oil-production water. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 105 (2), 401–411.
- Suvorova, I.A., Ravcheev, D.A., and Gelfand, M.S., 2012. Regulation and evolution of malonate and propionate catabolism in proteobacteria. *Journal of Bacteriology*, 194 (12), 3234–3240.
- Suyama, A., Iwakiri, R., Kimura, N., Nishi, A., Nakamura, K., and Furukawa, K., 1996. Engineering hybrid pseudomonads capable of utilizing a wide range of aromatic hydrocarbons and of efficient degradation of trichloroethylene. *Journal of Bacteriology*, 178 (14), 4039–4046.
- Tanenbaum, D.M., Goll, J., Murphy, S., Kumar, P., Zafar, N., Thiagarajan, M., Madupu, R., Davidsen, T., Kagan, L., Kravitz, S., Rusch, D.B., and Yooseph, S., 2010. The JCVI standard operating procedure for annotating prokaryotic metagenomic shotgun sequencing data. *Standards in Genomic Sciences*, 2 (2), 229–237.
- Tao, K., Liu, X., Chen, X., Hu, X., Cao, L., and Yuan, X., 2017. Biodegradation of crude oil by a defined co-culture of indigenous bacterial consortium and exogenous *Bacillus subtilis*.

*Bioresource Technology*, 224, 327–332.

Thierry, S., Macarie, H., Iizuka, T., Geißdörfer, W., Assih, E.A., Spanevello, M., Verhe, F., Thomas, P., Fudou, R., Monroy, O., Labat, M., and Ouattara, A.S., 2004. *Pseudoxanthomonas mexicana* sp. nov. and *Pseudoxanthomonas japonensis* sp. nov., isolated from diverse environments, and emended descriptions of the genus *Pseudoxanthomonas* Finkmann et al. 2000 and of its type species. *International Journal of Systematic and Evolutionary Microbiology*, 54 (6), 2245–2255.

Tiralerdpanich, P., Sonthiphand, P., Luepromchai, E., Pinyakong, O., and Pokethitiyook, P., 2018. Potential microbial consortium involved in the biodegradation of diesel, hexadecane and phenanthrene in mangrove sediment explored by metagenomics analysis. *Marine Pollution Bulletin*, 133 (June), 595–605.

Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., and Segata, N., 2015. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nature Methods*, 12 (10), 902–903.

Unimke, A.A., Mmuoegbulam, O.A., and Anika, O.C., 2018. Microbial Degradation of Petroleum Hydrocarbons: Realities, Challenges and Prospects. *Biotechnology Journal International*, 22 (2), 1–10.

Varjani, S.J. and Upasani, V.N., 2017. A new look on factors affecting microbial degradation of petroleum hydrocarbon pollutants. *International Biodeterioration and Biodegradation*.

Vikram, S., Singh, V., Singh, D., Kumar, S., and Tripathi, M., 2018. Metagenomic Approach towards Bioprospection of Novel Biomolecule(s) and Environmental Bioremediation. *Annual Research & Review in Biology*, 22 (2), 1–12.

Wanapaisan, P., Laothamteep, N., Vejarano, F., Chakraborty, J., Shintani, M., Muangchinda, C., Morita, T., Suzuki-Minakuchi, C., Inoue, K., Nojiri, H., and Pinyakong, O., 2018. Synergistic degradation of pyrene by five culturable bacteria in a mangrove sediment-derived bacterial consortium. *Journal of Hazardous Materials*, 342, 561–570.

Wang, H., Wang, B., Dong, W., and Hu, X., 2016. Co-acclimation of bacterial communities under

- stresses of hydrocarbons with different structures. *Scientific Reports*, 6 (1), 34588.
- Wang, Y., Nie, M., Wan, Y., Tian, X., Nie, H., Zi, J., and Ma, X., 2017. Functional characterization of two alkane hydroxylases in a versatile *Pseudomonas aeruginosa* strain NY3. *Annals of Microbiology*, 67 (7), 459–468.
- Xu, X., Tian, S., Jiang, P., Li, H., Wang, W., Gao, X., Qi, Q., Liu, W., Li, F., and Yu, H., 2018. Petroleum Hydrocarbon-Degrading Bacteria for the Remediation of Oil Pollution Under Aerobic Conditions: A Perspective Analysis. *Frontiers in Microbiology*, 9 (December), 1–11.
- Xu, X., Zhai, Z., Li, H., Wang, Q., Han, X., and Yu, H., 2017. Synergetic effect of biophotocatalytic hybrid system: g-C<sub>3</sub>N<sub>4</sub> and *Acinetobacter* sp. JLS1 for enhanced degradation of C<sub>16</sub> alkane. *Chemical Engineering Journal*, 323 (June), 520–529.
- Xu, Z., Hansen, M.A., Hansen, L.H., Jacquiod, S., and Sørensen, S.J., 2014. Bioinformatic approaches reveal metagenomic characterization of soil microbial community. *PLoS ONE*, 9 (4).
- Yang, D.D., de Billerbeck, G.M., Zhang, J.J., Rosenzweig, F., and Francois, J.M., 2018. Deciphering the origin, evolution, and physiological function of the subtelomeric arylalcohol dehydrogenase gene family in the yeast *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology*, 84 (1), 1–16.
- Yang, S., Wen, X., Zhao, L., Shi, Y., and Jin, H., 2014. Crude oil treatment leads to shift of bacterial communities in soils from the deep active layer and upper permafrost along the China-Russia Crude Oil Pipeline route. *PLoS ONE*, 9 (5), 12–14.
- Yeager, C.M., Gallegos-Graves, L.V., Dunbar, J., Hesse, C.N., Daligault, H., and Kuske, C.R., 2017. Polysaccharide degradation capability of Actinomycetales soil isolates from a semiarid grassland of the Colorado Plateau. *Applied and Environmental Microbiology*, 83 (6), 1–19.
- Young, C.C., Ho, M.J., Arun, A.B., Chen, W.M., Lai, W.A., Shen, F.T., Rekha, P.D., and Yassin, A.F., 2007. *Pseudoxanthomonas spadix* sp. nov., isolated from oil-contaminated soil. *International Journal of Systematic and Evolutionary Microbiology*, 57 (8), 1823–1827.

---

# Chapter 5

## Cloning, Expression and Purification of a Novel Ammonia Monooxygenase Alpha Subunit (AMO-VUT1) Mined from Metagenome of Hydrocarbon Contaminated Soils

---

### *Abstract*

Hydrocarbon contaminated soils are rich sources of methanotrophic bacterial community that are inherently dependent on ammonia monooxygenases (AMO) for the initial break down of alkanes as a sole carbon source. Fifteen fosmid clones with hydrocarbon degrading potential were obtained from metagenomic DNA isolated from hydrocarbon contaminated soils. The samples were screened to identify monooxygenase genes in comparison to available sequences on the Uniprot database (<https://www.uniprot.org/>). An open reading frame (ORF) of 960 bp encoding a polypeptide of 319 amino acids with 36 kDa predicted molecular mass sharing a sequence similarity of only 21.3 % to AMO  $\alpha$ -subunit (amoA1) (EC:1.14.99.39), which shows the novelty of the gene. The gene sequence was then codon optimized as per *Escherichia coli* codon preference and synthesized. The synthesised gene was successfully cloned into a pET-30a(+) expression vector and transformed into chemically competent *E. coli* Rosetta™ (DE3) cells. The protein expression was induced using 1 mM Isopropyl- $\beta$ -D-thiogalactoside (IPTG) and incubated

at 15 °C for 16 hours. The expressed protein was purified and run on SDS-PAGE and the correct band size of 36 kDa was observed and further confirmed using Western Blotting Analysis using anti-His antibody. However, getting soluble AMOA-VUT1 was not achieved since it was later found to be an integral multi-pass membrane-bound protein, which made it insoluble despite the use of detergent to the effect. The solubility challenge was also reported in a number of other published studies. Therefore, it is pertinent for future studies to focus on improving the solubility of ammonia monooxygenase through meticulously genetically engineering the enzyme to remove the signal and other membrane-bound peptides without compromising its activity. Research should also be conducted to independently clone, express and purify all three subunits and reconstitute the purified subunits and evaluate if they could catalyze the oxygen attack of hydrocarbons.

**Keywords:** Hydrocarbons, methanotrophs, microorganisms, ammonia monooxygenase, multi-pass membrane protein, oxidation, bioremediation.

<b>TABLE OF CONTENTS</b>	<b>PAGES</b>
5.1 Introduction .....	195
5.2 Materials and Methods .....	197
5.2.1 Selection of fosmid clones .....	197
5.2.2 Induction of fosmid clones and extraction of fosmid DNA.....	197
5.2.3 Fosmid DNA construct integrity assessment .....	198
5.2.4 Sequence assembly and analysis.....	198
5.2.5 Gene synthesis and cloning.....	199
5.2.6 Transformation of bacterial host cells.....	199
5.2.7 Protein induction and expression .....	201
5.2.8 Protein purification and visualization .....	202
5.2.9 Western blot analysis .....	203
5.3 Results and Discussion.....	204
5.3.1 Fosmid DNA integrity assessment.....	204
5.3.2 Sequence analysis and monooxygenase gene synthesis .....	205
5.3.3 Cloning, expression and protein purification of AMO alpha subunit gene .....	207
5.3.4 Enzymatic activity of AMO alpha subunit .....	214
5.4 Conclusion.....	215
5.5 References .....	216

## 5.1 Introduction

The value of producing hydrocarbon degrading enzymes at an industrial level for bioremediation strategies have placed great focus on recombinant protein production in microbial cell factories (Hausjell *et al.* 2018). The gene is cloned, and the protein mass produced in a compatible host system such as bacteria, yeast, insects or plants. *Escherichia coli* always stands out as the most attractive option due to its genetic manipulations and fast growing nature on inexpensive media leading to high yields (Rosano and Ceccarelli 2014, Jia and Jeon 2016).

BL21 and BL21 (DE3) derived strains of *E. coli* are such examples making it the most used strains for expression studies (Rosano and Ceccarelli 2014, Jia and Jeon 2016, Castiñeiras *et al.* 2018). For this study to take place purified soluble active recombinant proteins are required hence it is important to have a means to detect expression and purification, attain maximum solubility and be able to easily purify it from *E. coli*. Therefore, small peptide tags are fused to the gene such as poly-His and recovered by immobilisation metal ion affinity chromatography (Bornhorst and Falke 2011, Paraskevopoulou and Falcone 2018).

Recombinant protein production does not come without its many issues and the most frustrating of these is low protein levels that cannot be detected even by sensitive techniques such as Western blotting. The problem lies in the harmful effect that the heterologous protein exerts on the host cells (Freudl 2018). Furthermore, inclusion body formation can also result when a foreign gene is introduced into *E. coli* (Fakruddin *et al.* 2013, Rosano and Ceccarelli 2014). When there are high levels of expression, hydrophobic stretches in the polypeptide regions are available for interaction with other similar regions. Contributing to protein instability and protein aggregation resulting in inclusion bodies (Slouka *et al.* 2019).

The common feature in diverse metabolic pathways is that the initial oxidation is always mediated by a monooxygenase (MO) (Wang and Shao 2013). MO are a valuable enzyme family for



application in bioremediation (Mccarl *et al.* 2018). MO studies from bacteria have been found to be diverse and are classified into six types based on their cofactor requirements and/or subcellular locations (Daubner *et al.* 1997, Leahy *et al.* 2003, Balasubramanian *et al.* 2010, Bui and Steiner 2016, Girvan and Munro 2016, Fürst *et al.* 2019).

Ammonia monooxygenase (AMO) are closely related to methane monooxygenase (MMO) occurring in ammonia oxidisers and methanotrophs respectively as first shown by (Holmes *et al.* 1995). Their results show that particulate methane monooxygenase (pMMO) and ammonia monooxygenase (AMO) are evolutionarily related despite their different physiological roles in bacteria and belong to a novel family of membrane-bound monooxygenases (Khadka *et al.* 2018). A derived model shows they consist of three subunits and metal centers of copper and iron (Semrau *et al.* 2018). Research has been hindered through the years due to decreasing enzymatic activity by 10 to 100 fold upon membrane isolation and protein purification (Sirajuddin and Rosenzweig 2015, Fisher *et al.* 2018, Hausjell *et al.* 2018). This can be attributed to the absence of unidentified protein components that are involved in loading, assembly and stabilisation of the active sites and/or delivery of protons and electrons (Fisher *et al.* 2018). Thus, a purified AMO sample with enzymatic activity has never been obtained (Fisher *et al.* 2018).

Therefore, the objective of the current study was to isolate the monooxygenase alpha subunit (amoA1) from metagenomic DNA isolated from hydrocarbon contaminated soils. It is envisioned that the enzyme can be added to the toolbox of bioremediation studies by increasing the number of purified novel monooxygenases found to date.

## 5.2 Materials and Methods

### 5.2.1 Selection of fosmid clones

A total of 15 candidate fosmid clones obtained from a large insert metagenomic library described in more detail in Chapter 3, section 3.1.5, were selected based on functional screening analysis using three hydrocarbon substrates (hexadecane, octadecene and cyclohexane). The functional screening of the fosmid library was performed as described previously in Chapter 3, section 3.1.8. Briefly, five clones were chosen based on their hydrocarbon degrading potential when grown on a minimal media (Bushnell Hass) supplemented with one hydrocarbon substrate (**Table 5.1**). This was assessed based on colony growth rate and size over a time period.

**Table 5.1:** Selected fosmid clones based on growth rates on three different hydrocarbon substrates hexadecane, octadecene and cyclohexane.

Hydrocarbon substrates	Selected clones
Hexadecane	pFos_A1, pFos_A7, pFos_B2, pFos_B3 and pFos_F2
Octadecene	pFos_D3, pFos_D8, pFos_F1, pFos_F3 and pFos_A4
Cyclohexane	pFos_A2, pFos_C7, pFos_B7, pFos_D7 and pFos_G3

### 5.2.2 Induction of fosmid clones and extraction of fosmid DNA

An overnight culture of each fosmid clone was grown in 10 ml of Luria-Bertani (LB) broth supplemented with chloramphenicol (12.5 µg/ml) to make a stock culture and 1 ml was used to inoculate 25 ml LB broth supplemented with chloramphenicol (12.5 µg/ml). This was autoinduced by adding 50 µl 500X CopyControl™ Fosmid Autoinduction Solution as previously described by the manufacturer (Epicentre, USA). The autoinduced culture was grown overnight at 37 °C in a shaking incubator at 170 rpm. EPI 300™ bacterial host cells grown in 25 ml LB broth was used

as the negative control. The bacterial cells were harvested from 10 mls of overnight culture by centrifugation at 4000 rpm for 5 minutes. The supernatant was discarded and the pellet resuspended in 250 µl suspension buffer (Epicentre, USA). The fosmid DNA was thereafter, extracted using the GeneJET Plasmid Miniprep Kit (Thermoscientific) according to the manufacturer's instructions. Isolated plasmid DNA was eluted in 50 µl Elution buffer supplied by the manufacturer and stored at -80 °C.

### **5.2.3 Fosmid DNA construct integrity assessment**

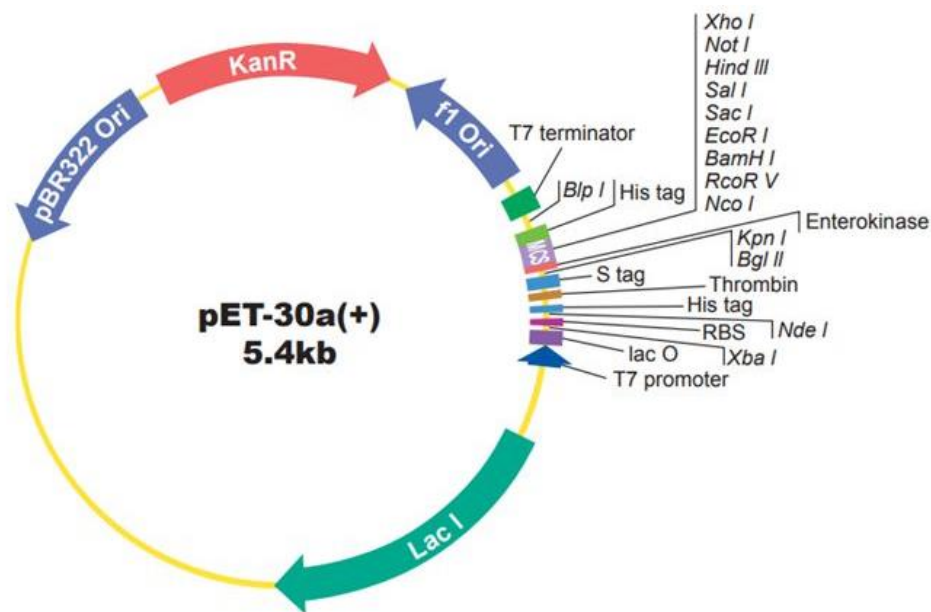
The extracted fosmid DNA for the 15 positive fosmid clones were visualised on a 0.8 % (w/v) agarose gel stained with ethidium bromide and run at 100V for 30 minutes. The pCC2FOS™ vector supplied with the CopyControl™ HTP Fosmid Library Production Kit (Epicentre, USA) was run on the agarose gel electrophoresis as a control. The fosmid constructs were sequenced using PACBIO SMRT® next generation sequencing at Inqaba Biotec™ (Pretoria, South Africa).

### **5.2.4 Sequence assembly and analysis**

*Denova* assembly was carried out on the PACBIO sequence reads using the CLC Genomics Workbench 11.0.1 (CLC Bio, Qiagen). The quality of the data was assessed and the contig sequence length was set at 200 bp. Filtering of reads were carried out based on default parameters set in the software. The software program was used to extract the longest contig assembly. The open reading frames (ORFs) along with the translated protein were determined from the assembled contig sequences for each of the fifteen fosmid clones using ORF finder on the NCBI database (<https://www.ncbi.nlm.nih.gov/orffinder/>). A protein Basic Local Alignment Search Tool (BLAST) called Uniprot-Swiss prot (<https://www.uniprot.org/>) was carried out on the individual ORF sequences in order to find regions of local similarity between known sequences to identify members of gene families involved in hydrocarbon degradation pathways. These ORFs were flagged for further analysis.

## 5.2.5 Gene synthesis and cloning

The ORF sequence designated *amo-vut1* was codon optimised for codon compatibility in *E.coli* to increase protein expression. His tags were added to the sequence to facilitate for downstream protein purification and detection using the Akta Start Protein Purification System (Merck KGaA, Germany). The optimised gene sequence for ammonia/methane monooxygenase corresponding to a specific enzyme class found in the alkane degradation pathway was selected and synthesized by GenScript® USA Inc. This DNA sequence was cloned into a pET-30a(+) expression vector system (Novagen, USA) (Figure 5.1) with a T7 *lac* promoter according to the pET System manual (Novagen, USA). The DNA sequence was cloned at cloning sites *Nde I* and *Hind III*.



**Figure 5.1:** Map of pET-30a(+) expression vector used for the cloning of target gene sequences.

## 5.2.6 Transformation of bacterial host cells

The preparation of chemically competent *E. coli* BL21 (DE3) pLysS cells was carried out using the Rubidium chloride method. An overnight culture of 0.5 ml was used to inoculate 500 ml LB

broth. This was incubated at 37 °C with aeration until the optical density (OD<sub>600 nm</sub>) reached 0.48 using the NanoDrop™ One spectrophotometer (Thermo Fisher Scientific™). These were decanted into 20 ml tubes and incubated on ice for 15 minutes followed by centrifugation at 5000xg at 4 °C for 5 minutes until all cells were pelleted. The supernatant was discarded and 0.4 volumes of TfbI solution (30 mM Potassium acetate, 100 mM Rubidium chloride, 10 mM Calcium chloride, 50 mM Manganese chloride and 15% (v/v) glycerol) was used to resuspend the pellet and thereafter kept on ice for 15 minutes. The cells were pelleted at 5000xg for 5 minutes at 4 °C and resuspended in 0.04 volumes of TfbII solution pH 6.5 [10 mM MOPS 3-(N-morpholino) propane sulfonic acid], 75 mM Calcium chloride, 10 mM Rubidium chloride and 15% (v/v) glycerol). This solution was kept on ice for 15 minutes and thereafter 0.5 ml aliquots were placed in 1 ml cryotubes and stored at -80 °C.

The transformation of the chemically competent *E. coli* BL21 (DE3) pLysS cells with pET-30a(+) expression vector ligated with the modified *amo-vut1* was carried out according to the protocol for E.cloni® 10G chemically competent cells (Lucigen®). This involved thawing 50 µl of chemically competent *E. coli* BL21 (DE3) pLysS cells on ice and transformation was carried out using 4 µl plasmid vector with insert. The mixture was incubated on ice for 30 minutes followed by incubation for 45 seconds at 42 °C in a water bath. The heated solution was placed immediately on ice for two minutes and 960 µl of Recovery medium (E.cloni® 10G, Lucigen). Transformation was also carried out using Rosetta™2 (DE3) *E. coli* cells to check for enhanced expression levels. The culture was allowed to shake at 250 rpm at 37 °C for 1 hour and thereafter spread plated onto LB agar plates supplemented with kanamycin (50 µg/ml). This was allowed to grow overnight at 37 °C. Transformed host cells with plasmid DNA without insert were used as the negative control. The growth of colonies on the plate indicated the transformed bacteria and were selected for further analysis.

A single transformed colony with plasmid and insert resulting in the designated name pET30a(+)\_amo-vut1 was picked and grown in 100 ml LB broth supplemented with kanamycin (50 µg/ml) and grown overnight at 37 °C under aeration. Glycerol stocks of these were prepared

and stored at -80 °C. In order to carry out quality control and confirm the successful transformation of the *E.coli* host cells, 1.5 ml of the overnight culture was used to isolate plasmid DNA from the pelleted cells using the ZR miniprep plasmid isolation kit (Zymo Research) according to the manufacturer's instructions. The isolated plasmid DNA was visualised on a 0.8 % (w/v) agarose gel stained with ethidium bromide to determine the quality and quantity of the isolated plasmid DNA. Restriction analysis was carried out on the plasmid DNA using the enzymes *Xba I* and *Hind III* in a double digest. Digestion was allowed to proceed for 3 hours at 37 °C and the restriction mixture was separated on a 0.8 % (w/v) agarose gel stained with ethidium bromide. This was carried out to determine the successful transformation of the *E. coli* host cells and confirm the size of the insert.

### **5.2.7 Protein induction and expression**

The expression of His-tagged proteins was carried out using *E. coli* BL21 (DE3) pLysS cells as the host strain. An overnight seed culture prepared from the glycerol stock for pET30a(+)\_amo-vut1 was used to inoculate 200 ml LB broth supplemented with kanamycin (50 µg/ml) with aeration at 37 °C. The cells were grown until they reached an optical density (OD<sub>600 nm</sub>) of 0.5-0.6 using the NanoDrop™ One spectrophotometer (Thermo Fisher Scientific™) and followed by induction with Isopropyl-β-D-thiogalactoside (IPTG) at a final concentration of 1 mM. The expression was monitored over various time periods for induction at 25 °C with shaking at 250 rpm. A corresponding culture without IPTG was used as the uninduced control. In order to obtain the highest expression levels for each clone, parameters such as growth temperatures (16 °C, 25 °C and 37 °C), the IPTG inducer concentration (0.3 – 1 mM final concentration), varied densities of optical density (OD<sub>600 nm</sub>) for induction (0.3 - 0.9) and the growth period (3 hours, 8 hours and overnight) were assessed.

Following the various parameters tested for induction, the bacterial cells were pelleted by centrifugation at 5000 x g for 10 minutes and protein extraction was carried out using the B-PER®

Bacterial Protein Extraction Reagent (Thermo Scientific™) according to the manufacturer's instructions. A test expression protein sample and a solubility protein sample were collected for each expression analysis for induced and uninduced experiments. The uninduced samples represented the negative controls and Bovine serum albumin (Sigma Aldrich) at concentrations of 1 µg and 2 µg concentrations were used as a negative control.

### **5.2.8 Protein purification and visualization**

The expressed protein profile was visualised using Sodium dodecyl sulfate– polyacrylamide gel electrophoresis (SDS–PAGE) (15 % resolving gel and 5% stacking gel) as described by the Laemmli procedure (Laemmli 1970). Protein samples were prepared by adding Laemmli 2x concentrate sample buffer (Sigma-Aldrich) to the protein supernatant and this was heat denatured at 95 °C in a water bath for 5 minutes. The proteins were stained with Coomassie blue stain solution (40 % ethanol, 0.125 % Coomassie® blue, distilled water and 10 % acetic acid) and destained with Coomassie blue destain solution (5 % ethanol and 7.5 % acetic acid). Crude extracts were centrifugation at 18,000g for 30 minutes to remove any particulate matter and applied to a HisTrap™ FF column (GE Healthcare Life Sciences) according to the manufacturer's instruction using the Akta Start Protein Purification System (Merck KGaA, Germany). The unbound proteins were washed out with a wash buffer (50 mM KH<sub>2</sub>PO<sub>4</sub>–K<sub>2</sub>HPO<sub>4</sub>, 300 mM NaCl, 40 mM imidazole, pH 8.0). His-tagged fusion proteins were eluted with an elution buffer (50 mM KH<sub>2</sub>PO<sub>4</sub>–K<sub>2</sub>HPO<sub>4</sub>, 300 mM NaCl, 300 mM imidazole, pH 8.0). Fractionation was set to 1.5 ml volumes. This was further dialyzed against 50 mM KH<sub>2</sub>PO<sub>4</sub>–K<sub>2</sub>HPO<sub>4</sub> (pH 8.0) containing 20 % glycerol. Protein concentrations were determined by the method of Bradford (1976) using bovine serum albumin (BSA, Sigma Aldrich) as a standard.

### 5.2.9 Western blot analysis

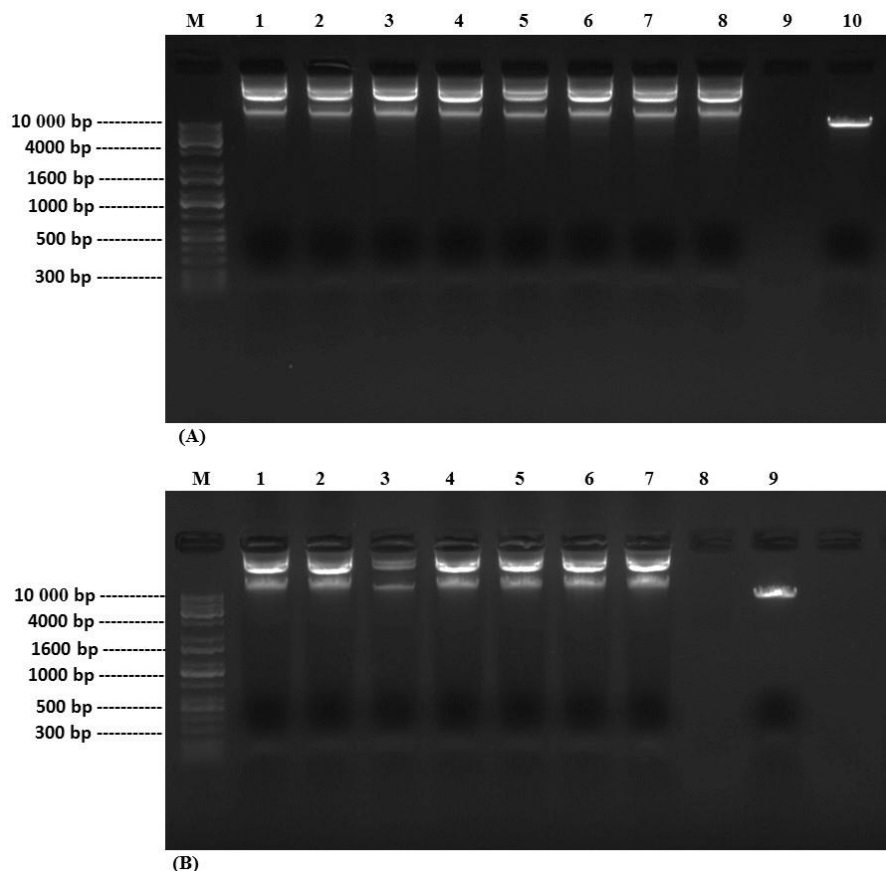
The purified protein AMO-VUT1 was visualised using Sodium dodecyl sulfate– polyacrylamide gel electrophoresis (SDS–PAGE) (15 % resolving gel and 5% stacking gel) as described by the Laemmli procedure (Laemmli 1970). The iBind™ Western automated system (iBlot 2™ and iBind™) was used to carry out Western blotting according to the manufacturer’s instructions included in the iBlot 2 and iBind Western Starter Kit. Following the blotting, the membranes were blocked with a 1X iBind™ solution. Mouse-anti-His mAb (ThermoFisher Scientific, USA) was used as the primary antibodies for the antibody binding step. Proteins were detected using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Scientific, USA) that is an enhanced chemiluminescent (ECL) horseradish peroxidase (HRP) substrate according to manufacturer’s instructions.



## 5.3 Results and Discussion

### 5.3.1 Fosmid DNA integrity assessment

Fosmid DNA isolated from the selected fosmid clones based on functional screening were found to be of high quality with no smearing of bands (**Figure 5.2**). This was important for downstream PACBIO SMRT<sup>®</sup> sequencing that requires high quality DNA to ensure successful long sequence reads to be obtained. Selection was based on their ability to utilize one of the three hydrocarbon substrates hexadecane, cyclohexane and octadecene. The isolated fosmid DNA was separated using agarose gel electrophoresis and visualised using ethidium bromide.



**Figure 5.2:** Agarose gel 0.8 % (w/v) electrophoresis of isolated fosmid DNA from 15 candidate clones obtained from functional screening analysis. (A) M: KAPA<sup>™</sup> Biosystems Universal ladder (10 kb), Lane 1: pFos\_A1, Lane 2: pFos\_A2, Lane 3: pFos\_A4, Lane 4: pFos\_A7, Lane 5:

pFos\_B2, Lane 6: pFos\_B3, Lane 7: pFos\_B7, Lane 8: pFos\_C7 and Lane 10: Fosmid control DNA (40 kb). (B) M: KAPA<sup>TM</sup> Biosystems Universal ladder (10 kb), Lane 1: pFos\_D3, Lane 2: pFos\_D7, Lane 3: pFos\_D8, Lane 4: pFos\_F1, Lane 5: pFos\_F2, Lane 6: pFos\_F3, Lane 7: pFos\_G6 and Lane 9: Fosmid control DNA (40 kb).

### 5.3.2 Sequence analysis and monooxygenase gene synthesis

PACBIO SMRT<sup>®</sup> sequencing platform was chosen for this study since the platform grants long-read sequences with high consensus accuracy and uniform coverage. *De novo* assembly of the sequences was carried out using the CLC Genomics Workbench 11.0.1 (CLC Bio, Qiagen). Data for assembly shown in Appendix B1. The longest assembled contigs for each clone obtained from the software program was used to determine ORF and consequently the regions of local similarity between known sequences to identify monooxygenase genes involved in the initial step of hydrocarbon degradation. An ORF of 960 bp, encoding a polypeptide of 319 amino acids with 36 kDa predicted molecular mass was found to share a sequence similarity of 21.3 % to ammonia monooxygenase alpha subunit (*amoA1*) (EC:1.14.99.39). The complete *amo-vut1* nucleotide sequence was submitted to GenBank and assigned an accession number MT606178. According to the Uniprot-Swissprot database (Bateman *et al.* 2017) this identified monooxygenase subunit requires copper as a cofactor. Thus, the enzyme can be classified under the family of monooxygenases related to copper-containing multi-pass membrane proteins.

The low percentage similarity (21.3 %) for the sequence further indicates the novelty of our monooxygenase *amo-vut1*. A number of researchers employed metagenomics techniques to mine novel monooxygenases (Sierra-García *et al.* 2014, Girvan and Munro 2016, Bouhajja *et al.* 2017, Heine *et al.* 2018). Bouhajja *et al.* (2017), identified three novel toluene monooxygenase genes in hydrocarbon sediments from a tar oil-contaminated site in Flingern, Germany.

To further support the low percentage similarity observed, it is worth mentioning that the sampling locations chosen for this study as mentioned previously (Chapter 3, section 3.1.1) has been under investigated. Metagenomic DNA isolated from such soil samples hold the potential for isolating highly adapted microorganisms, or those which have been poorly characterised due to their unculturable nature or are totally unclassified at this point. By focusing on their ability to degrade hydrocarbons and utilise them as a sole carbon source, novel monooxygenase genes involved in the initial step of hydrocarbon degradation can potentially be targeted and further characterised.

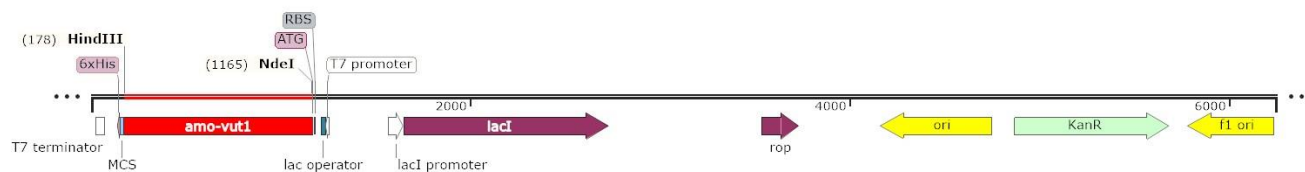
The ammonia monooxygenase AMO1 was also shown to have been previously isolated from *Nitrosomomas europaea* (strain ATCC 19718 / CIP 103999 / KCTC 2705 / NBRC 14298) according to the Uniprot-Swissprot database (Bateman *et al.* 2017). *Nitrosomomas europaea* falls under the taxonomic lineage of *Proteobacteria*. We have reported earlier in Chapter 4 that *Proteobacteria* was the most abundant bacterial domain (84%). Ren and colleagues also reported that *Proteobacteria* are abundant in different complex ecosystems (Ren *et al.* 2018). They are ubiquitous and able to survive in a wide range of environments including hydrocarbon contaminated areas. Researchers (Shen *et al.* 2018, Siles and Margesin 2018, Auti *et al.* 2019, Garrido-Sanz *et al.* 2019) have shown the same correlation with respect to the elevation in abundance of *Proteobacteria* and other phyla with increasing petroleum hydrocarbon concentrations in the environment and their manipulation for valuable application in bioremediation studies.

It is important at this point to highlight that only the ammonia monooxygenase alpha subunit was identified from the ORFs and not all the subunits which are essential to ensuring the proper functional activity of the enzyme which involves catalysing the oxidation of ammonia to hydroxylamine or alkanes to alkanols. No other ORFs identified from the 15 fosmid clone sequences were found to show sequence similarity to the other subunits or other soluble monooxygenases in the Uniprot-swissprot database. Hence, posing a challenge of assessing the enzymatic activity of this enzyme. The characteristics of this class of enzymes with multiple

subunits and each with their varying metal content has complicated the efforts to isolate a complete working monooxygenase complex in this study.

### 5.3.3 Cloning, expression and protein purification of AMO alpha subunit gene

The ligated pET-30a(+) expression vector with *amo-vut1* gene insert was transformed into chemically competent *E. coli* BL21 (DE3) pLysS cells and was designated pET30a(+)\_amo-vut1 (**Figure 5.3**). Complete modified sequences for DNA and protein shown in Appendix B4 and B5. Numerous studies have shown the potential of using the pET expression vector for cloning and expression studies making it the most appropriate system (Hu *et al.* 2013, Yue *et al.* 2017). This is attributed to easier cloning, detection and purification of target proteins (Choi and Geletu 2018).

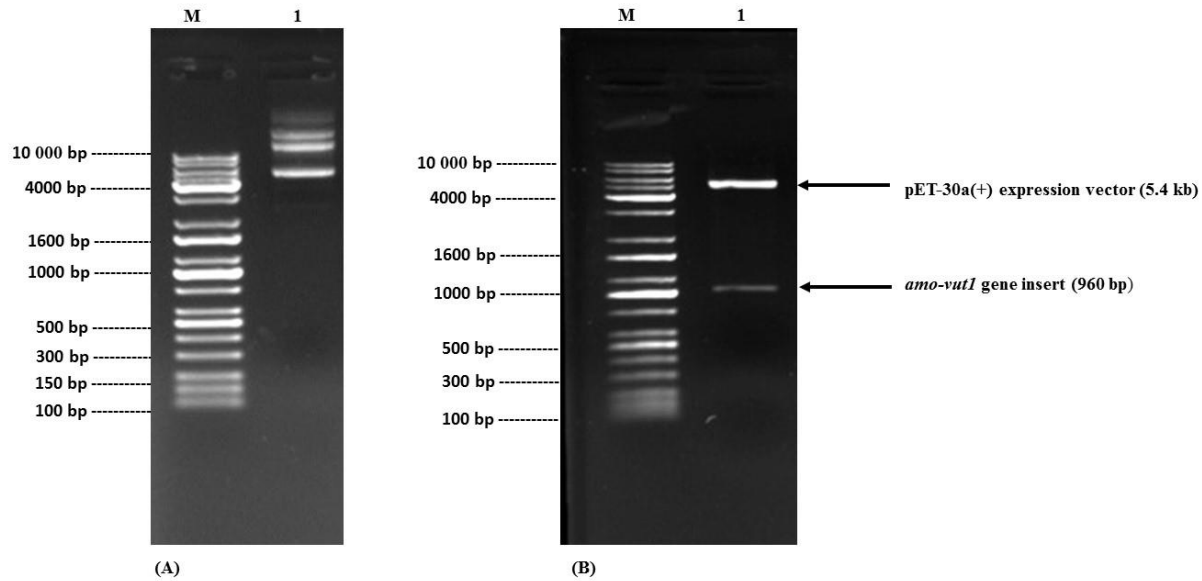


**Figure 5.3:** Linear map of pET30a(+)\_amo-vut1.

The plasmid DNA was extracted and separated on an agarose gel stained with ethidium bromide (

**Figure 5.4(A)**) thereafter, restriction digest performed using XbaI and HindIII (

**Figure 5.4(B)**) to confirm the ligation product. The extracted plasmid DNA was found to be of high quality and the restriction digest showed the successful cloning of the insert DNA with the expected band size of 960 bp.

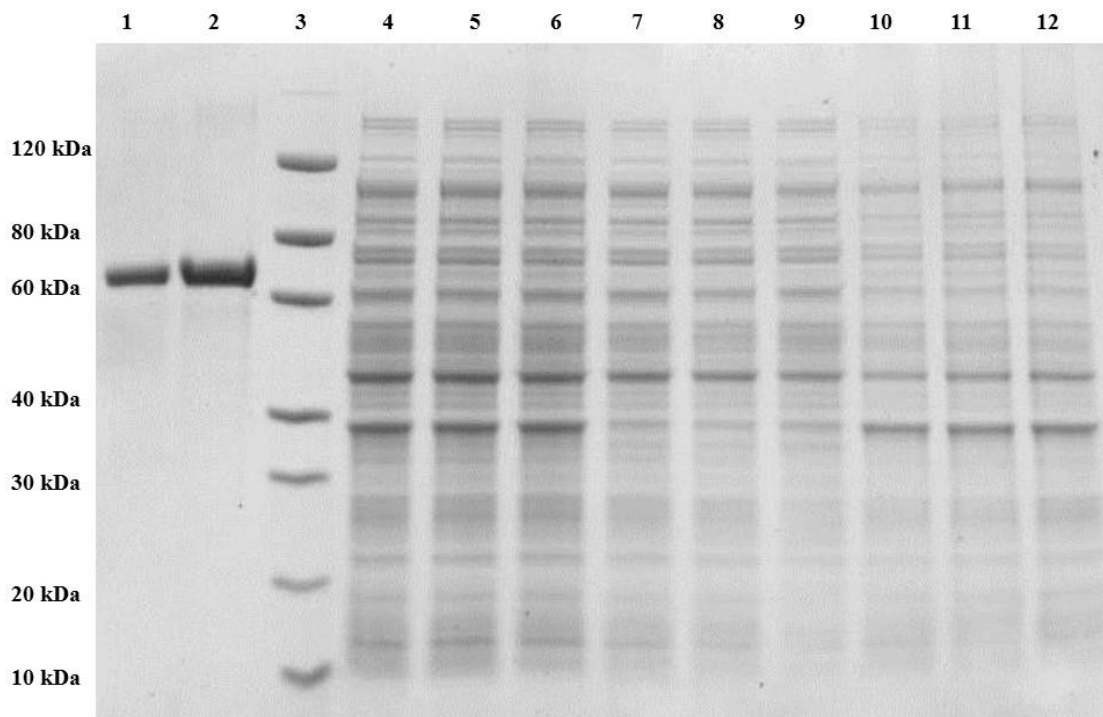


**Figure 5.4:** Analysis of *amo-vut1* gene insert. (A) Image of agarose gel/EtBr electrophoresis of extracted plasmid DNA. M: 10 kb DNA ladder. Lane 1: Extracted plasmid DNA from transformed *E. coli* BL21 (DE3) pLysS competent cells. (B) M. 10 kb Ladder. Lane 1: Image of restriction digests with *XbaI* and *HindIII* showing the successful cloning of the AMO alpha subunit gene insert at 960 bp.

The soluble and insoluble fractions of the cell extracts AMO alpha subunit protein were analysed on SDS-PAGE to determine the expression level of the protein following varying conditions of induction. The estimated molecular mass of the protein subunit was calculated from the amino acid sequence in order to compare with the standard protein marker found to be 36 kDa. Varying experimental conditions for induction including time and temperature were setup to maximize expression levels as well as the use of both BL21 (DE3) *E. coli* and Rosetta™2 (DE3) *E. coli* cell lines for transformation.

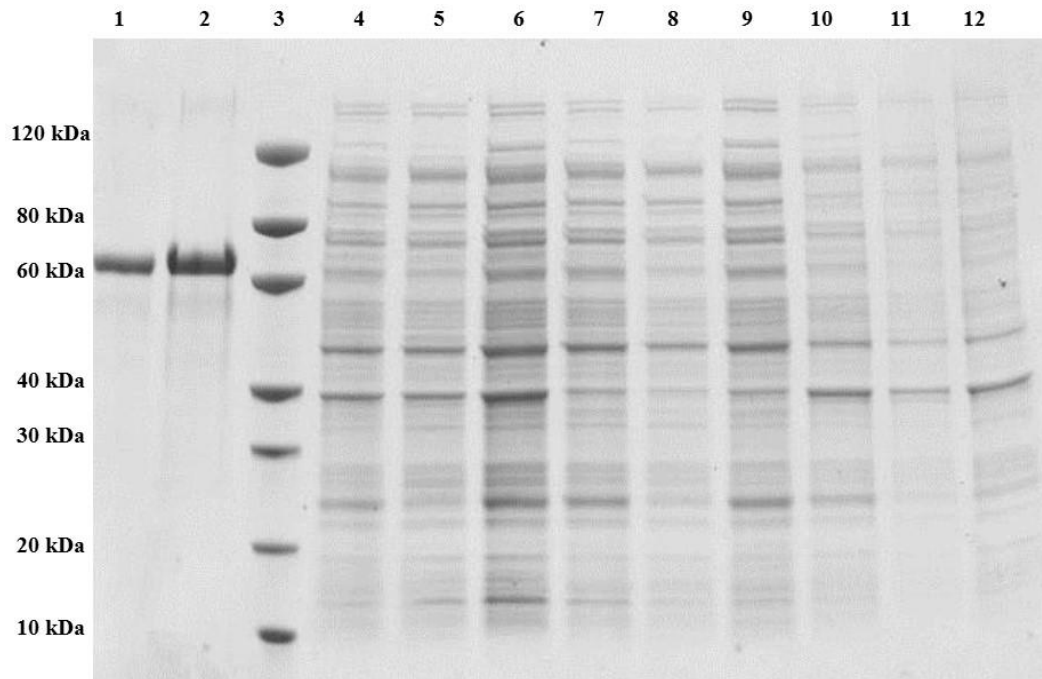
SDS-PAGE gel images for induced and non-induced AMO-VUT1 in BL21 (DE3) and *E. coli* cells Rosetta™2 (DE3) *E. coli* with 0.5 mM IPTG are shown in **Figure 5.5** and **Figure 5.6** respectively.

The overexpressed protein of expected band size 36 kDa was not clearly visible for the induced protein for the cell lysate, supernatant of cell lysate and pellet of cell lysate since the level of expression obtained was very low. The optimal induction conditions were observed at 16 hours at 15 °C and 4 hours at 37 °C but the expression levels were still found to be too low for both expression in *E. coli* BL21 (DE3) and *E. coli* Rosetta™2 (DE3) host cells. Such results does not seem to be uncommon when reviewing the literature which shows the difficulty in expressing particulate monooxygenases in *E. coli* host cells (Furuya *et al.* 2013) as compared to soluble monooxygenases such as alkene monooxygenases in a Gram-negative host (Mccarl *et al.* 2018).



**Figure 5.5:** SDS-PAGE analysis of the AMO alpha subunit (AMO-VUT1) band after overexpression in BL21 (DE3) *E. coli* cells at varying conditions. Lane 1: BSA (1 µg), Lane 2: BSA (2 µg), Lane 3: Protein marker, Lane 4: Cell lysate without induction, Lane 5: Cell lysate with induction for 16 h at 15 °C, Lane 6: Cell lysate with induction for 4 h at 37 °C, Lane 7: Supernatant of cell lysate without induction, Lane 8: Supernatant of cell lysate with induction for 16 h at 15 °C, Lane 9: Supernatant of cell lysate with induction for 4 h at 37 °C, Lane 10: Pellet of

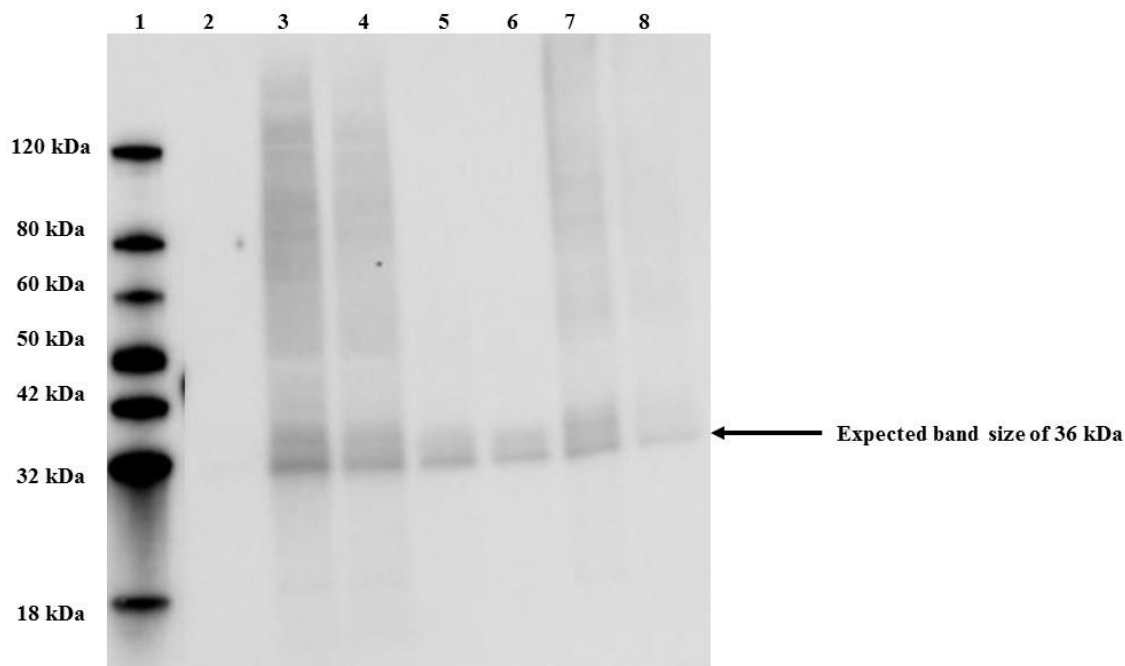
cell lysate without induction, Lane 11: Pellet of cell lysate with induction for 16 h at 15 °C and Lane 12: Pellet of cell lysate with induction for 4 h at 37 °C.



**Figure 5.6:** SDS-PAGE analysis of the AMO alpha subunit (AMO-VUT1) band after overexpression in *E.coli* Rosetta™2 (DE3) cells at varying conditions. Lane 1: BSA (1 µg), Lane 2: BSA (2 µg), Lane 3: Protein marker, Lane 4: Cell lysate without induction, Lane 5: Cell lysate with induction for 16 h at 15 °C, Lane 6: Cell lysate with induction for 4 h at 37 °C, Lane 7: Supernatant of cell lysate without induction, Lane 8: Supernatant of cell lysate with induction for 16 h at 15 °C, Lane 9: Supernatant of cell lysate with induction for 4 h at 37 °C, Lane 10: Pellet of cell lysate without induction, Lane 11: Pellet of cell lysate with induction for 16 h at 15 °C and Lane 12: Pellet of cell lysate with induction for 4 h at 37 °C.

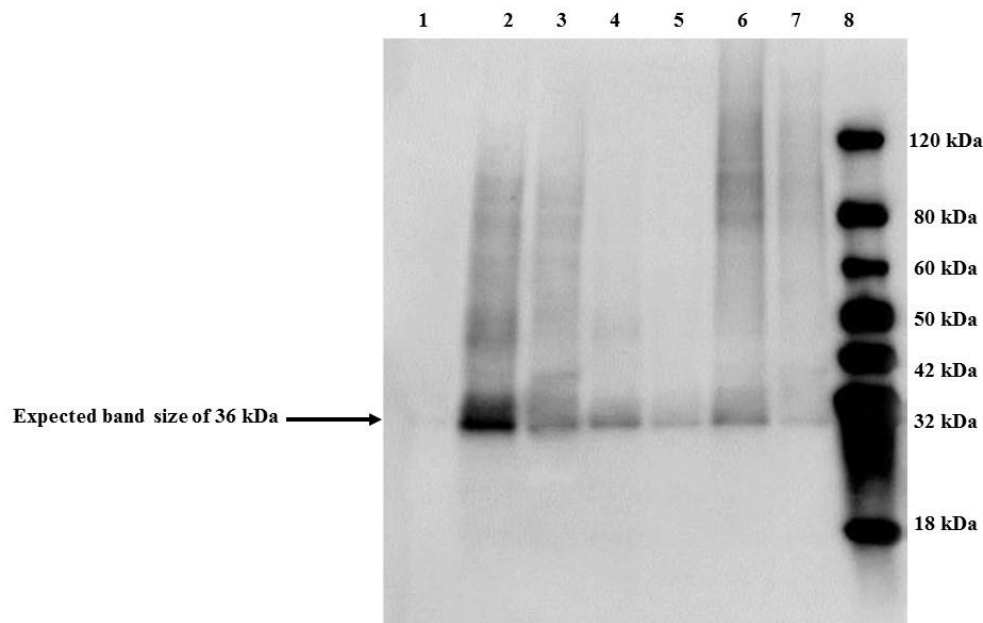
Bands observed on the gel for the uninduced experiment were similarly detected for the induced expression hence, Western blot analysis was performed to confirm the expression results. This

was done using the anti-His antibody to probe the Western blots using chemiluminescence detection methods for both transformation cell lines for His-tagged proteins of interest as shown in **Figure 5.7** and **Figure 5.8**.



**Figure 5.7:** Western blot analysis for AMO alpha subunit (AMO-VUT1) expressed in *E. coli* cells BL21 (DE3) probed with anti-His antibody. Lane 1: Western blot marker, Lane 2: Cell lysate without induction, Lane 3: Cell lysate with induction for 16 h at 15 °C, Lane 4: Cell lysate with induction for 4 h at 37 °C, Lane 5: Supernatant of cell lysate with induction for 16 h at 15 °C, Lane 6: Supernatant of cell lysate with induction for 4 h at 37 °C, Lane 7: Pellet of cell lysate with induction for 16 h at 15 °C and Lane 8: Pellet of cell lysate with induction for 4 h at 37 °C.





**Figure 5.8:** Western blot analysis for AMO alpha subunit (AMO-VUT1) expressed in *E. coli* Rosetta™ 2(DE3) probed with anti-His antibody. Lane 1: Cell lysate without induction, Lane 2: Cell lysate with induction for 16 h at 15 °C, Lane 3: Cell lysate with induction for 4 h at 37 °C, Lane 4: Supernatant of cell lysate with induction for 16 h at 15 °C, Lane 5: Supernatant of cell lysate with induction for 4 h at 37 °C, Lane 6: Pellet of cell lysate with induction for 16 h at 15 °C, Lane 7: Pellet of cell lysate with induction for 4 h at 37 °C and Lane 8: Western blot marker.

The results obtained showed that the AMO-VUT1 with 6x His-tags was present on the SDS-PAGE with an expected band size of 36 kDa (**Figure 5.7**), hence expression of the protein was taking place but at very low levels. The cell lysate without induction showed no bands (**Figure 5.7**) therefore no His-tagged proteins were present as was expected. Hence, it can be confirmed that the expressed His-tagged proteins observed was only as a result of the induction carried out.

Expression levels were found to be higher for cell lysate with induction for 16 h at 15 °C and cell lysate with induction for 4 h at 37 °C in comparison to the cell lysate supernatant and pellet.

Additionally, expression in Rosetta™ 2(DE3) *E. coli* cells showed better expression than BL21 (DE3) *E. coli* cells. Low expression levels in the current study can be supported by the numerous studies indicating the difficulty of expression and purification of the particulate monooxygenase enzyme complex. The complexity of this protein arises from its structure having multiple subunits as well as its complex multi-pass membrane bound nature that makes its solubilisation a challenge (Ro *et al.* 2018).

To further explain these results, it is important for one to understand the complex nature of monooxygenases and their difficulty to be purified as detailed in the literature through the years (Sirajuddin and Rosenzweig 2015). This is most particular for those that fall under the class of particulate monooxygenases. Particulate monooxygenases as is the case with the identified sequence in this study are integral membrane bound proteins and unlike soluble monooxygenases found in the cytoplasm, adds to the challenge of solubilizing and purifying active particulate monooxygenases. Furthermore, these enzymes are copper-dependent hence the availability of this element is essential in specific concentrations for enzyme activity. As a result, many studies have been hindered by the loss of enzymatic activity upon its removal from the native membrane and its solubilisation with detergents. In many reports, activity was even found to be completely abolished. These challenges have been highlighted by a number of researchers and recently pertinent work carried out by Ro *et al.* (2018) studied whether loss of enzyme activity was only largely due to the assumption that solubilisation and purification of monooxygenases lead to the removal of catalytically essential copper ions or whether the removal of particulate monooxygenases from the membrane itself has deleterious effects on its activity on the whole.

Challenges of isolating most of the subunits of monooxygenase with its varying copper content from the membrane using detergents seems to be a difficult task due to loss of enzyme activity. However, there seems to be a way forward according to work done by Ro *et al.* (2018), which involved the implementation of bicelle reconstitution of purified protein from *Methylomicrobium* (Mm.) *alcaliphilum* 20Z (20Z-pMMO) and *Methylococcus capsulatus* (Bath) (Bath-pMMO) into bicelles which recovers the oxidative activity of this enzyme without the addition of exogenous copper ions or substantial alteration in the copper sites. This could prove to be a way forward if

monooxygenases are to be further characterised, studied and implemented in enzyme cocktails targeted at a synergistic hydrocarbon degradation approach.

#### **5.3.4 Enzymatic activity of AMO alpha subunit**

In the current study, only the novel AMO alpha subunit of the monooxygenase gene with a molecular mass of 36 kDa was expressed and purified. pMMO from four methanotrophs has been characterized using a crystallography method. In all instances the oligomerization state was a ~300 kDa  $\alpha_3\beta_3\gamma_3$  trimer composed of PmoA (24 kDa) and PmoC (22 kDa), almost entirely transmembrane domains, and PmoB (42 kDa), a primarily periplasmic subunit comprising two cupredoxin-like  $\beta$ -barrels joined by two trans-membrane helices (Sirajuddin *et al.* 2014). Thus, showing the complexity of the protein class.

These proteins require all the subunits in order to carry out its function hence, enzymatic activity testing could not be assessed. Hence, its enzymatic activity based on parameters such as optimal enzyme concentration and substrate concentration could not be assessed in this study. Furthermore, based on the low expression levels visualised on the SDS-PAGE gels and the presence of expressed His-tagged proteins in the cell pellet lysate, this protein can be classified as a particulate membrane protein. Particulate membrane proteins pose the problem of being difficult to solubilise and purify and the use of detergents in solubilisation, further leads to their inactivation downstream since they become denatured. Hence, this expressed novel alpha subunit alone could not be used in the developed enzyme cocktail with other downstream enzymes in the pathway to show complete alkane degradation. This is due to the requirement of all four subunits to attain complete functionality of the enzyme.

## 5.4 Conclusion

Ammonia monooxygenase gene *amo-vut1* showed low sequence identity (21.3 %) against monooxygenase genes in the Uniprot database, implying the novelty of the gene. The gene was successfully cloned and expressed. The expressed protein was purified and run on SDS-PAGE and the correct band size of 36 kDa was observed and further confirmed using Western Blotting Analysis using anti-His antibody. However, getting soluble AMOA-VUT1 was not achieved since it was later found to be an integral multi-pass membrane-bound protein, which made it insoluble despite the use of detergent to the effect. The solubility challenge was also reported in a number of other published studies. Therefore, it is pertinent for future studies to focus on improving the solubility of ammonia monooxygenase through meticulously genetically the enzyme to remove the signal and other membrane-bound peptides without compromising its activity. Research should also be conducted to independently clone, express and purify all three subunits and reconstitute the purified subunits and evaluate if they could catalize the oxygen attack of hydrocarbons.

## 5.5 References

- Auti, A.M., Narwade, N.P., Deshpande, N.M., and Dhotre, D.P., 2019. Microbiome and imputed metagenome study of crude and refined petroleum-oil-contaminated soils: Potential for hydrocarbon degradation and plant-growth promotion. *Journal of biosciences*, 44 (5), 0–16.
- Balasubramanian, R., Smith, S.M., Rawat, S., Yatsunyk, L.A., Stemmler, T.L., and Rosenzweig, A.C., 2010. Oxidation of methane by a biological dicopper centre. *Nature*, 465 (7294), 115–119.
- Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-AJee, H., Cowley, A., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Castro, L.G., Figueira, L., Garmiri, P., Georghiou, G., Gonzalez, D., Hatton-Ellis, E., Li, W., Liu, W., Lopez, R., Luo, J., Lussi, Y., MacDougall, A., Nightingale, A., Palka, B., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Rosanoff, S., Saidi, R., Sawford, T., Shypitsyna, A., Speretta, E., Turner, E., Tyagi, N., Volynkin, V., Wardell, T., Warner, K., Watkins, X., Zaru, R., Zellner, H., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Aimo, L., ArgoudPuy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-Casas, C., De Castro, E., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Nospikel, N., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A.L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Ross, K., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S., and Zhang, J., 2017. UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, 45 (D1), D158–D169.
- Bornhorst, B.J.A. and Falke, J.J., 2011. WITHDRAWN: Reprint of: Purification of Proteins Using Polyhistidine Affinity Tags. *Protein Expression and Purification*, (February 2000).

- Bouhajja, E., McGuire, M., Liles, M.R., Bataille, G., Agathos, S.N., and George, I.F., 2017. Identification of novel toluene monooxygenase genes in a hydrocarbon-polluted sediment using sequence- and function-based screening of metagenomic libraries. *Applied Microbiology and Biotechnology*, 101 (2), 797–808.
- Bradford, M.M., 1976. A Rapid and Sensitive Method for the Quantitation Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Analytical Biochemistry*, 72, 248–254.
- Bui, S. and Steiner, R.A., 2016. New insight into cofactor-free oxygenation from combined experimental and computational approaches. *Current Opinion in Structural Biology*, 41, 109–118.
- Castiñeiras, T.S., Williams, S.G., Hitchcock, A.G., and Smith, D.C., 2018. E. coli strain engineering for the production of advanced biopharmaceutical products. *FEMS Microbiology Letters*, 365 (15), 1–10.
- Choi, T.J. and Geletu, T.T., 2018. High level expression and purification of recombinant flounder growth hormone in E. coli. *Journal of Genetic Engineering and Biotechnology*, 16 (2), 347–355.
- Daubner, S.C., Hillas, P.J., and Fitzpatrick, P.F., 1997. Characterization of chimeric pterin-dependent hydroxylases: Contributions of the regulatory domains of tyrosine and phenylalanine hydroxylase to substrate specificity. *Biochemistry*, 36 (39), 11574–11582.
- Fakruddin, M., Mohammad Mazumdar, R., Bin Mannan, K.S., Chowdhury, A., and Hossain, M.N., 2013. Critical Factors Affecting the Success of Cloning, Expression, and Mass Production of Enzymes by Recombinant E. coli . *ISRN Biotechnology*, 2013 (3), 1–7.
- Fisher, O.S., Kenney, G.E., Ross, M.O., Ro, S.Y., Lemma, B.E., Batelu, S., Thomas, P.M., Sosnowski, V.C., DeHart, C.J., Kelleher, N.L., Stemmler, T.L., Hoffman, B.M., and Rosenzweig, A.C., 2018. Characterization of a long overlooked copper protein from methane- and ammonia-oxidizing bacteria. *Nature Communications*, 9 (1), 1–12.
- Freudl, R., 2018. Signal peptides for recombinant protein secretion in bacterial expression

- systems. *Microbial Cell Factories*, 17 (1), 1–10.
- Fürst, M.J., Fiorentini, F., and Fraaije, M.W., 2019. Beyond active site residues: overall structural dynamics control catalysis in flavin-containing and heme-containing monooxygenases. *Current Opinion in Structural Biology*, 59, 29–37.
- Furuya, T., Hayashi, M., and Kino, K., 2013. Reconstitution of active mycobacterial binuclear iron monooxygenase complex in escherichia coli. *Applied and Environmental Microbiology*, 79 (19), 6033–6039.
- Garrido-Sanz, D., Redondo-Nieto, M., Guirado, M., Pindado Jiménez, O., Millán, R., Martín, M., and Rivilla, R., 2019. Metagenomic Insights into the Bacterial Functions of a Diesel-Degrading Consortium for the Rhizoremediation of Diesel-Polluted Soil. *Genes*, 10 (6), 456.
- Girvan, H.M. and Munro, A.W., 2016. Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology. *Current Opinion in Chemical Biology*, 31, 136–145.
- Hausjell, J., Halbwirth, H., and Spadiut, O., 2018. Recombinant production of eukaryotic cytochrome P450s in microbial cell factories. *Bioscience Reports*, 38 (2), 1–13.
- Heine, T., Tischler, D., van Berkel, W., van Pée, K.-H., and Gassner, G., 2018. Two-Component FAD-Dependent Monooxygenases: Current Knowledge and Biotechnological Opportunities. *Biology*, 7 (3), 42.
- Holmes, A.J., Costello, A., Lidstrom, M.E., and Murrell, J.C., 1995. Evidence that participate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. *FEMS Microbiology Letters*, 132 (3), 203–208.
- Hu, W., Liu, N., Tian, Y., and Zhang, L., 2013. Molecular cloning, expression, purification, and functional characterization of dammarenediol synthase from panax ginseng. *BioMed Research International*, 2013 (3).
- Jia, B. and Jeon, C.O., 2016. High-throughput recombinant protein expression in Escherichia coli: Current status and future perspectives. *Open Biology*, 6 (8).

- Khadka, R., Clothier, L., Wang, L., Lim, C.K., Klotz, M.G., and Dunfield, P.F., 2018. Evolutionary History of Copper Membrane Monooxygenases. *Frontiers in Microbiology*, 9 (October), 1–13.
- Laemmli, U.K., 1970. 227680a0. *Nature*, 227, 680–685.
- Leahy, J.G., Batchelor, P.J., and Morcomb, S.M., 2003. Evolution of the soluble diiron monooxygenases. *FEMS Microbiology Reviews*, 27 (4), 449–479.
- Mccarl, V., Somerville, M. V, Ly, M., Henry, R., Liew, E.F., Wilson, N.L., Holmes, A.J., and Coleman, N. V, 2018. crossm Bacterial Hosts, 84 (15), 1–16.
- Paraskevopoulou, V. and Falcone, F., 2018. Polyionic Tags as Enhancers of Protein Solubility in Recombinant Protein Expression. *Microorganisms*, 6 (2), 47.
- Ren, M., Zhang, Z., Wang, X., Zhou, Z., Chen, D., Zeng, H., Zhao, S., Chen, L., Hu, Y., Zhang, C., Liang, Y., She, Q., Zhang, Y., and Peng, N., 2018. Diversity and contributions to nitrogen cycling and carbon fixation of soil salinity shaped microbial communities in Tarim Basin. *Frontiers in Microbiology*, 9 (MAR), 1–14.
- Ro, S.Y., Ross, M.O., Deng, Y.W., Batelu, S., Lawton, T.J., Hurley, J.D., Stemmler, T.L., Hoffman, B.M., and Rosenzweig, A.C., 2018. From micelles to bicelles: Effect of the membrane on particulate methane monooxygenase activity. *Journal of Biological Chemistry*, 293 (27), 10457–10465.
- Rosano, G.L. and Ceccarelli, E.A., 2014. Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*, 5 (APR), 1–17.
- Semrau, J.D., Dispirito, A.A., Gu, W., and Yoon, S., 2018. crossm, 84 (6), 7–14.
- Shen, Y., Ji, Y., Li, C., Luo, P., Wang, W., Zhang, Y., and Nover, D., 2018. Effects of phytoremediation treatment on bacterial community structure and diversity in different petroleum-contaminated soils. *International Journal of Environmental Research and Public Health*, 15 (10).
- Sierra-García, I.N., Alvarez, J.C., De Vasconcellos, S.P., De Souza, A.P., Dos Santos Neto, E.V.,



- and De Oliveira, V.M., 2014. New hydrocarbon degradation pathways in the microbial metagenome from brazilian petroleum reservoirs. *PLoS ONE*, 9 (2).
- Siles, J.A. and Margesin, R., 2018. Insights into microbial communities mediating the bioremediation of hydrocarbon-contaminated soil from an Alpine former military site. *Applied Microbiology and Biotechnology*, 102 (10), 4409–4421.
- Sirajuddin, S., Barupala, D., Helling, S., Marcus, K., Stemmler, T.L., and Rosenzweig, A.C., 2014. Effects of zinc on particulate methane monooxygenase activity and structure. *Journal of Biological Chemistry*, 289 (31), 21782–21794.
- Sirajuddin, S. and Rosenzweig, A.C., 2015. Enzymatic oxidation of methane. *Biochemistry*, 54 (14), 2283–2294.
- Slouka, C., Kopp, J., Spadiut, O., and Herwig, C., 2019. Perspectives of inclusion bodies for bio-based products: curse or blessing? *Applied Microbiology and Biotechnology*, 103 (3), 1143–1153.
- Wang, W. and Shao, Z., 2013. Enzymes and genes involved in aerobic alkane degradation. *Frontiers in Microbiology*, 4 (MAY), 1–7.
- Yue, Q., Yang, Y., Zhao, J., Zhang, L., Xu, L., Chu, X., Liu, X., Tian, J., and Wu, N., 2017. Identification of bacterial laccase cueO mutation from the metagenome of chemical plant sludge. *Bioresources and Bioprocessing*, 4 (1).

---

# Chapter 6

## Cloning, Expression, Purification and Kinetic Study of a Novel Alcohol Dehydrogenase (AOL-VUT3) Mined from Metagenome of Hydrocarbon Contaminated Soils

---

Article submitted for publication

### *Abstract*

Microorganisms from hydrocarbon contaminated soils have adapted to harsh environments by utilizing specific enzymatic pathways to catabolise different classes of hydrocarbon. Over the years enzymes have become highly sought-after biomolecules for environmental, pharmaceutical and industrial and other applications. Alcohol dehydrogenase lies at the heart of aerobic hydrocarbon degradation since it is involved in the second step of catalysing the oxidation of alcohols to aldehydes. Fifteen candidate fosmid clones with hydrocarbon degrading potential were screened to identify alcohol dehydrogenase genes by comparison to sequences available in the Uniprot database. An open reading frame of 759 bp encoding for alcohol dehydrogenase (AOL-VUT3) with 24.7 % sequence similarity to histidinol dehydrogenase (HDH) from *Mycobacterium bovis* (strain ATCC BAA-935 / AF2122/97) (EC:1.1.1.23) was found and the low sequence similarity shows the novelty of the *aol-vut3*. The gene sequence was codon-optimized as per *Escherichia coli* codon preference and synthesized. The synthesised gene was successfully cloned into a pET-30a(+) expression vector and transformed into chemically competent *Escherichia coli*

Rosetta™ (DE3) cells. The protein expression was induced using 1 mM Isopropyl-β-D-thiogalactoside (IPTG) and incubated at 15 °C for 16 hours. The protein was expressed, purified and its kinetic parameters studied. The enzyme successfully converted hexanol to a less complex hexanal. Accordingly, kinetics study showed a low  $K_m$  value of 2.875 mM with a  $V_{max}$  of 0.008614  $\mu\text{mol}\cdot\text{min}^{-1}$ , showing the high affinity of the enzyme to its substrate (hexanol) thus, making it a valuable candidate for degradation of alkanol. Since, the discovered enzyme is novel it is important to conduct crystallographic study of such indispensable protein and solve its structure.

**Keywords:** Hydrocarbons, hexanol, hexanal, alkanols, enzyme kinetics, alcohol dehydrogenase, fosmid clones, reduction, bioremediation.

<b>TABLE OF CONTENTS</b>	<b>PAGES</b>
6.1 Introduction .....	224
6.2 Materials and Methods .....	226
6.2.1 Selection of fosmid clones .....	226
6.2.2 Induction of fosmid clones and extraction of fosmid DNA.....	226
6.2.3 Fosmid DNA integrity assessment.....	227
6.2.4 <i>De novo</i> sequence assembly and analysis .....	227
6.2.5 Gene synthesis and cloning.....	228
6.2.6 Transformation of bacterial host cells.....	229
6.2.7 Protein induction and expression .....	230
6.2.8 Protein purification and visualization .....	231
6.2.9 Western blot analysis .....	232
6.2.10 ADH activity assay .....	232
6.2.11 Kinetic studies (Substrate affinity of ADH) .....	233
6.3 Results and Discussion.....	234
6.3.1 The quality of isolated fosmid DNA.....	234
6.3.2 Sequence Analysis and Histidinol Dehydrogenase Gene Synthesis.....	235
6.3.3 Cloning, expression and protein purification of Histidinol dehydrogenase .....	236
6.3.4 Optimal alcohol dehydrogenase (AOL-VUT3) concentration .....	240
6.3.5 Optimal hexanol substrate concentration and enzyme kinetics .....	242
6.4 Conclusion.....	246
6.5 References .....	247

## 6.1 Introduction

Soil is a complex environment teeming with microbial diversity for gene mining and a vast resource for natural product discovery such as hydrocarbon degrading enzymes for industrial application (Choi and Geletu 2018, Hausjell *et al.* 2018, Gurusinghe *et al.* 2019). Targeting polluted soil sites is important for identifying novel alcohol dehydrogenases (ADHs) that are involved in aerobic pathways for petroleum compound breakdown by the indigenous microbiota. Metagenomic DNA isolated from these soils allow for access to genomes from microorganisms that are troublesome when subjected to culturing techniques (Chaudhary *et al.* 2019). The candidate gene is cloned, and the protein is mass produced in a compatible host system such as bacteria, yeasts, insects or plants. *Escherichia coli* is the most employed expression host due to its rapid growth on inexpensive media, expression, ease of culturing and high productivity (Rosano and Ceccarelli 2014, Jia and Jeon 2016).

The most widely used strain for expression studies are BL21 and BL21(DE3) derived strains of *E. coli* (Rosano and Ceccarelli 2014) . They have been well documented and have undergone numerous modifications through the years (Castiñeiras *et al.* 2018). The expression system used in this study was the pET-30a(+) vector, documented as the most powerful system developed for cloning and expression of recombinant proteins in *E. coli* (Hu *et al.* 2013; Choi and Geletu 2018). The high efficiency of this plasmid lies in the activity of the T7 RNA polymerase which when fully induced drives almost all of the cell's resources towards target gene expression.

ADHs are highly diverse enzymes catalysing the reversible oxidation of primary or secondary alcohols to aldehydes or ketones, and the reactions are coupled to the reduction/oxidation of a pyridine. Due to their different functions, these biocatalysts are of great interest for industrial applications (Shinde *et al.* 2018). The main function of ADHs is to catalyse the reversible oxidation of alcohols and are wide spread in all organisms (Liu *et al.* 2009). They display a broad substrate range and play a variety of physiological roles. With regards to alkane degradation,

ADH catalyses the second reaction step of the terminal oxidation pathway and have been isolated from many alkane degrading bacteria (Wentzel *et al.* 2007, Elumalai *et al.* 2017).

The enzymatic activity of purified ADH has been assessed on a specific enzymatic assay developed by Li *et al.* (2010). This assay utilizes ADH enzymes to oxidize alcohols into aldehydes. In this process the co-factor NAD (nicotinamide adenine dinucleotide) is converted into NADH. NAD does not absorb light at 340 nm whereas, NADH does. The change in absorbance caused by the formation of NADH due to the oxidation of the substrate is measured spectrophotometrically. In this assay one unit of enzyme is defined as the amount of the enzyme that catalyses the formation of 1.0  $\mu\text{mol}$  of NADH per minute.

The objective of this study was to develop an expression plasmid for an identified ADH gene (*aol-vut3*) obtained from a fosmid clone, and to subsequently purify and characterize the protein. Enzyme activity of the purified protein was assessed using an ADH assay by determining NADH produced or oxidised when catalysing an alkanol. The optimal enzyme and substrate concentrations were determined followed by enzyme kinetics studies. Identification and characterization of a novel ADH with hydrocarbon degrading potential holds value for future bioremediation studies.

## 6.2 Materials and Methods

### 6.2.1 Selection of fosmid clones

A total of 15 candidate fosmid clones from the prepared metagenomic library described in Chapter 3, section 3.1.5, were selected based on functional screening analysis using three hydrocarbon substrates hexadecane, octadecene and cyclohexane. The functional screening of the fosmid library has been described previously (Chapter 3, section 3.1.8). Five clones were chosen for each of the substrates based on their growth rate and colony size on minimal media (Bushnell Haas media) supplemented with a hydrocarbon source (**Table 6.1**).

**Table 6.1:** Selected fosmid clones based on growth rates on three different hydrocarbon substrates.

Hydrocarbon substrates	Selected clones
Hexadecane	pFos_A1, pFos_A7, pFos_B2, pFos_B3 and pFos_F2
Octadecene	pFos_D3, pFos_D8, pFos_F1, pFos_F3 and pFos_A4
Cyclohexane	pFos_A2, pFos_C7, pFos_B7, pFos_D7 and pFos_G3

### 6.2.2 Induction of fosmid clones and extraction of fosmid DNA

An overnight culture of each fosmid clone was grown in 10 ml of Luria-Bertani (LB) broth supplemented with chloramphenicol (12.5 µg/ml) to make a stock culture and 1 ml was used to inoculate 25 ml LB broth supplemented with chloramphenicol (12.5 µg/ml). This was autoinduced by adding 50 µl 500X CopyControl™ Fosmid Autoinduction Solution as previously described by the manufacturer (Epicentre, USA). The autoinduced culture was grown overnight at 37 °C in a

shaking incubator at 170 rpm. EPI 300™ bacterial host cells grown in 25 ml LB broth was used as the negative control. The bacterial cells were harvested from 10 mls of overnight culture by centrifugation at 4000 rpm for 5 minutes. The supernatant was discarded and the pellet resuspended in 250 µl suspension buffer (Epicentre, USA). The fosmid DNA was thereafter, extracted using the GeneJET Plasmid Miniprep Kit (Thermoscientific) according to the manufacturer's instructions. Isolated fosmid DNA was eluted in 50 µl Elution buffer supplied by the manufacturer and stored at -80 °C.

### **6.2.3 Fosmid DNA integrity assessment**

The extracted fosmid DNA for the 15 candidate fosmid clones were visualised on a 0.8 % (w/v) agarose gel stained with ethidium bromide and run at 100V for 30 minutes. The pCC2FOS™ vector supplied with the CopyControl™ HTP Fosmid Library Production Kit (Epicentre, USA) was separated on the agarose gel as a control. The fosmid constructs were sent for PACBIO SMRT® next generation sequencing at Inqaba Biotec™ (Pretoria, South Africa).

### **6.2.4 *De novo* sequence assembly and analysis**

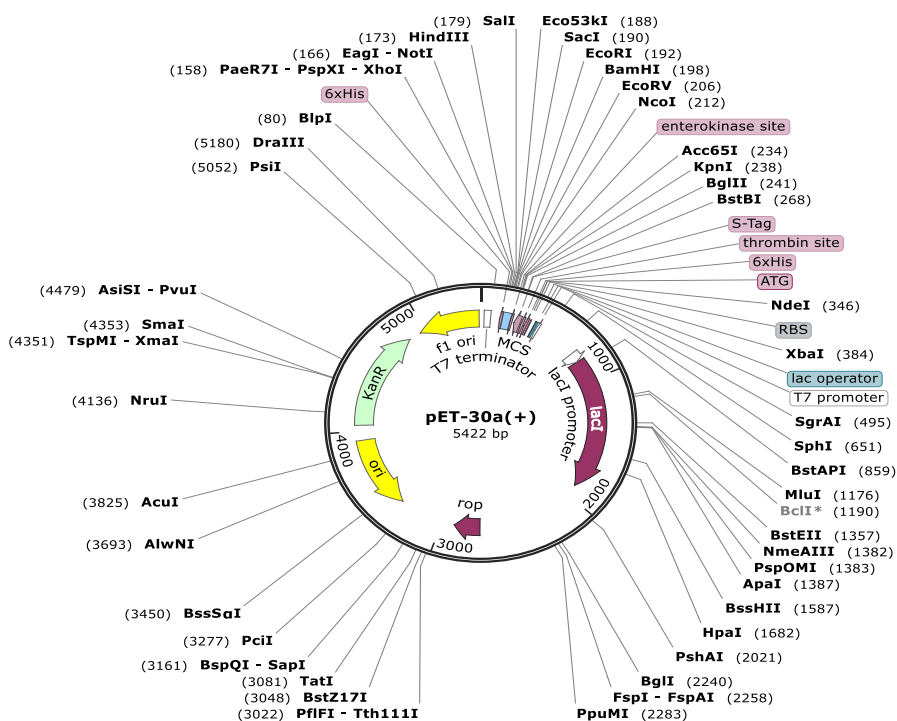
*De novo* sequence assembly was carried out on the PACBIO sequences obtained from Inqaba Biotec™ using the CLC Genomics Workbench 11.0.1 (CLC Bio, Qiagen). The quality of data was checked, the contig sequence length was set at 200 bp and filtering was carried out based on default parameters in the software. The software program was used to extract the longest contig assembly. The open reading frames (ORFs) along with the translated protein were determined from the assembled contig sequences for each of the fifteen fosmid clones using ORF finder on the NCBI database (<https://www.ncbi.nlm.nih.gov/orffinder/>). A protein Basic Local Alignment Search Tool (BLAST) called Uniprot-Swiss prot (<https://www.uniprot.org/>) was carried out on the individual ORF sequences in order to find regions of local similarity between known sequences to



identify members of gene families involved in hydrocarbon degradation pathways. These ORFs were flagged for further analysis.

### 6.2.5 Gene synthesis and cloning

The ORF sequence designated *aol-vut3* was codon optimised for codon compatibility in *E. coli* to increase protein expression and His tags were added to the sequence to facilitate for downstream protein purification and detection using the ÄKTA™ Start protein purification system (GE Healthcare Lifesciences). The optimised gene sequence was synthesized by GenScript® USA Inc. The synthesised DNA sequence was cloned into a pET-30a(+) expression vector system (Novagen, USA) (**Figure 6.1**) with a T7 *lac* promoter according to the pET System manual (Novagen, USA). The DNA sequence was cloned at cloning sites *NdeI* and *Hind III*.



**Figure 6.1:** Map of pET-30a(+) expression vector with all the features displayed.

### 6.2.6 Transformation of bacterial host cells

The preparation of chemically competent *E. coli* BL21 (DE3) pLysS cells was carried out using the Rubidium chloride method. An overnight culture of 0.5 ml was used to inoculate 500 ml LB broth. This was incubated at 37 °C with aeration until the optical density (OD<sub>600 nm</sub>) of 0.48 using the NanoDrop™ One spectrophotometer (Thermo Fisher Scientific™) was reached. These were decanted into 20 ml tubes and incubated on ice for 15 minutes followed by centrifugation at 5000xg at 4 °C for 5 minutes until all cells were pelleted. The supernatant was discarded and 0.4 volumes of TfbI solution (30 mM Potassium acetate, 100 mM Rubidium Chloride, 10 mM Calcium chloride, 50 mM Manganese chloride and 15% (v/v) glycerol) was used to resuspend the pellet and thereafter kept on ice for 15 minutes. The cells were pelleted at 5000xg for 5 minutes at 4 °C and resuspended in 0.04 volumes of TfbII solution pH 6.5 (10 mM MOPS (3-(N-morpholino)propanesulfonic acid), 75 mM Calcium chloride, 10 mM Rubidium chloride and 15% (v/v) glycerol). This solution was kept on ice for 15 minutes and thereafter 0.5 ml aliquots were placed in 1 ml cryotubes tubes and stored at -80 °C.

The transformation of the chemically competent *E. coli* BL21 (DE3) pLysS cells with pET-30a(+) expression vector ligated with the sequence of interest (*aol-vut3*) was carried out according to the protocol for E.cloni® 10G chemically competent cells (Lucigen®). This involved thawing 50 µl of chemically competent *E. coli* BL21 (DE3) pLysS cells on ice and transformation was carried out using 4 µl plasmid vector with insert. The mixture was incubated on ice for 30 minutes followed by incubation for 45 seconds at 42 °C in a water bath. The heated solution was placed immediately on ice for two minutes and 960 µl of Recovery medium (E. cloni® 10G, Lucigen). Transformation was also carried out using Rosetta™2 (DE3) *E. coli* cells to check for enhanced expression levels. The culture was allowed to shake at 250 rpm at 37 °C for 1 hour and thereafter spread plated onto LB agar plates supplemented with kanamycin (50 µg/ml). This was allowed to grow overnight at 37 °C. A negative control was also setup and used plasmid DNA without insert to transform the host cells. The growth of colonies on the plate indicated the transformed bacteria which were selected for further analysis.

A single transformed colony with plasmid and insert resulting in the designated name pET30(+)\_aol-vut3 was picked and grown in 100 ml LB supplemented with kanamycin (50 µg/ml) overnight at 37 °C with aeration. Glycerol stocks of these were prepared and stored at -80 °C. In order to carry out quality control and confirm the successful transformation of the *E. coli* host cells, 1.5 ml of the overnight culture was used to isolate plasmid DNA from the pelleted cells using the ZR miniprep plasmid isolation kit (Zymo Research) according to the manufacturer's instructions. The isolated plasmid DNA was visualised on a 0.8 % (w/v) agarose gel stained with ethidium bromide to determine the quality of the plasmid DNA. Restriction analysis was also carried out on the plasmid DNA using a double digestion involving the enzymes *XbaI* and *HindIII*. Digestion was allowed to proceed for 3 hours at 37 °C and the restriction mixture was separated on a 0.8 % (w/v) agarose gel stained with ethidium bromide. This was carried out to determine the size of the inserts for each clone and the successful transformation of the *E. coli* host cells.

### **6.2.7 Protein induction and expression**

The expression of His-tagged proteins was carried out using *E. coli* BL21 (DE3) pLysS cells as the host strain. An overnight seed culture for pET30a(+)\_aol-vut3 was prepared from a glycerol stock and used to inoculate 200 ml LB broth supplemented with kanamycin (50 µg/ml) with aeration at 37 °C. The cells were grown until they reached an optical density (OD<sub>600 nm</sub>) of 0.5-0.6 using the NanoDrop™ One spectrophotometer (Thermo Fisher Scientific™) and followed by induction with Isopropyl-β-D-thiogalactoside (IPTG) at a final concentration of 1mM. The expression was monitored over various time periods at 25 °C with shaking at 250 rpm. A corresponding culture without IPTG was used as the uninduced control. In order to obtain the highest expression levels for each clone, parameters such as growth temperatures (16 °C, 25 °C and 37 °C), the inducer IPTG concentration (0.3 – 1 mM) final concentration, varied densities of optical density (OD<sub>600 nm</sub>) for induction (0.3- 0.9) and the growth period (3 hours, 8 hours and overnight) were assessed.

Following the various parameters for induction, the bacterial cells were pelleted by centrifugation at 5000 x g for 10 minutes and protein extraction was carried out using the B-PER<sup>®</sup> Bacterial Protein Extraction Reagent (Thermo Scientific<sup>™</sup>) according to the manufacturer's instructions. A test expression protein sample and a solubility protein sample were collected for each expression analysis for induced and uninduced experiments. The uninduced samples represented the negative controls and BSA at 1 µg and 2 µg concentrations were used as a negative control.

### **6.2.8 Protein purification and visualization**

The expressed protein profile for AOL-VUT3 was visualised using Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) (15 % resolving gel and 5% stacking gel) as described by the Laemmli procedure (Laemmli 1970). Protein samples were prepared by adding Laemmli 2x concentrate sample buffer (Sigma-Aldrich) to the protein supernatant and this was heat denatured at 95 °C in a water bath for 5 minutes. The proteins were stained with Coomassie blue stain solution (40 % ethanol, 0.125 % Coomassie<sup>®</sup> blue, distilled water and 10 % acetic acid) and destained with Coomassie blue destain solution (5 % ethanol and 7.5 % acetic acid). Crude extracts were centrifugation at 18,000g for 30 minutes to remove any particulate matter and applied to a HisTrap<sup>™</sup> FF column (GE Healthcare Life Sciences) according to the manufacturer's instruction using the Akta Start Protein Purification System (Merck KGaA, Germany). The unbound proteins were washed out with a wash buffer (50 mM KH<sub>2</sub>PO<sub>4</sub>–K<sub>2</sub>HPO<sub>4</sub>, 300 mM NaCl, 40 mM imidazole, pH 8.0). His-tagged fusion proteins were eluted with an elution buffer (50 mM KH<sub>2</sub>PO<sub>4</sub>–K<sub>2</sub>HPO<sub>4</sub>, 300 mM NaCl, 300 mM imidazole, pH 8.0). Fractionation was set to 1.5 ml volumes. This was further dialyzed against 50 mM KH<sub>2</sub>PO<sub>4</sub>–K<sub>2</sub>HPO<sub>4</sub> (pH 8.0) containing 20 % glycerol. Protein concentrations were determined by the method of Bradford (1976) using bovine serum albumin (BSA, Sigma Aldrich) as a standard.

### 6.2.9 Western blot analysis

The purified protein AOL-VUT3 was visualised using Sodium dodecyl sulfate– polyacrylamide gel electrophoresis (SDS–PAGE) as described by the Laemmli procedure (Laemmli 1970). The iBind™ Western automated system (iBlot 2™ and iBind™) was used to carry out Western blotting according to the manufacturer’s instructions included in the iBlot 2 and iBind Western Starter Kit. Following the blotting, the membranes were blocked with a 1X iBind™ solution. Mouse-anti-His mAb (ThermoFisher Scientific, USA) was used as the primary antibodies for the antibody binding step. Proteins were detected using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (ThermoFisher Scientific, USA) that is an enhanced chemiluminescent (ECL) horseradish peroxidase (HRP) substrate according to manufacturer’s instructions.

### 6.2.10 Alcohol dehydrogenase (ADH) activity assay

The activity of the purified protein (AOL-VUT3) was carried out according to the protocol by Liu *et al.* (2009) for assessing ADH activity. The standard reaction mixture for the oxidation reaction (200 µl) contained hexanol (Merck, USA) as the substrate, 50 mM Nicotinamide adenine dinucleotide (NAD) in 50 mM Tris-HCl (pH 8.0) and an appropriate amount of enzyme. Varying concentrations of hexanol substrate (25 mM, 50 mM, 100 mM, 150 mM and 200 mM) and enzyme concentrations (0.001 mg/ml, 0.003 mg/ml, 0.007 mg/ml and 0.014 mg/ml) were tested respectively in order to determine the optimal concentrations for each. ADH activity was assayed spectrophotometrically at 340 nm in a 96 well plate by determining the Nicotinamide adenine dinucleotide phosphate (NADH) produced or oxidised, using the BioTek® Epoch 2 microplate reader (Agilent Technologies, USA). The enzyme was added to the reaction mixture to start the reaction. This was mixed thoroughly and immediately placed in the microplate reader. The absorbance values were recorded at twenty second intervals for 3 minutes at 30 °C. The enzymatic activity of the purified protein was measured by the increase in  $A_{340}$  due to the formation of NADH.

### 6.2.11 Kinetic studies (Substrate affinity of ADH)

After identifying the optimum enzyme concentration as previously described in section 6.1.10, a range of hexanol concentrations (25mM, 50 mM, 100 mM, 150 mM and 200 mM) were prepared to study the effect of hexanol concentration on the activity of the enzyme. This was carried out to determine the substrate affinity of the recombinant ADH and carry out kinetic studies to obtain the Michaelis-Menten constant ( $K_m$ ) and  $V_{max}$  using GraphPad Prism for Michaelis Menten kinetics. The Beer's Law equation was used to determine the initial velocity for each substrate concentration at a specific time interval. The wavelength specific absorption co-efficient for NADH ( $\epsilon$ ) is 6220 mol.L<sup>-1</sup>cm<sup>-1</sup> and a path length of 0.58 cm was used for a 96 well plate with 200  $\mu$ l volume.

$$A = \epsilon l c$$

Where:

$\epsilon$  = Substance and wavelength specific absorption coefficient,

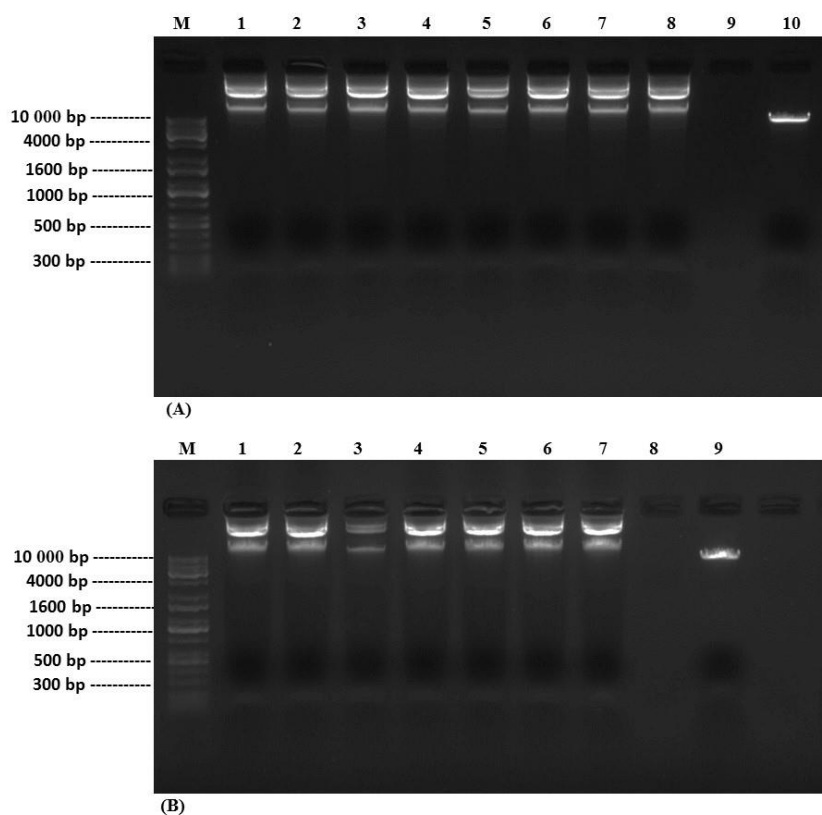
$l$  = Length of the light that travels through the sample

$c$  = Sample concentration.

## 6.3 Results and Discussion

### 6.3.1 The quality of isolated fosmid DNA

Plasmid DNA was isolated from 15 fosmid clones selected on the basis of a screening procedure described in detail in Chapter 3. This was based on their ability to utilize one of the three hydrocarbon substrates i.e. hexadecane, cyclohexane and octadecene. The isolated plasmid DNA was separated using agarose gel electrophoresis and was found to be of high quality and quantity (**Figure 6.2**). This was important for downstream PACBIO SMRT® sequencing that requires high quality DNA to ensure successful long sequence reads.



**Figure 6.2:** 0.8 % (w/v) agarose gel electrophoresis of isolated fosmid DNA (3  $\mu$ l) from 15 candidate clones obtained from functional screening analysis. (A) M: KAPA™ Biosystems Universal ladder (10 kb), Lane 1: pFos\_A1, Lane 2: pFos\_A2, Lane 3: pFos\_A4, Lane 4: pFos\_A7, Lane 5: pFos\_B2, Lane 6: pFos\_B3, Lane 7: pFos\_B7, Lane 8: pFos\_C7 and Lane 10: Fosmid control DNA (40 kb). (B) M: KAPA™ Biosystems Universal ladder (10 kb), Lane 1: pFos\_D3,

Lane 2: pFos\_D7, Lane 3: pFos\_D8, Lane 4: pFos\_F1, Lane 5: pFos\_F2, Lane 6: pFos\_F3, Lane 7: pFos\_G6 and Lane 9: Fosmid control DNA (40 kb).

### 6.3.2 Sequence analysis and Alcohol Dehydrogenase Gene Synthesis

*De novo* assembly was carried out on the PACBIO sequences for the 15 fosmid clones using the CLC Genomics Workbench 11.0.1 (CLC Bio, Qiagen). Data for assembly shown in Appendix B1. The longest assembled contigs for each clone obtained from this software program was used to obtain open reading frames and consequently find regions of local similarity between known sequences to identify ADH genes involved in the initial step of hydrocarbon degradation. An ORF of 759 bp in length encoding a polypeptide of 252 amino acids with 25 kDa predicted molecular mass was found to share a sequence similarity of 24.7 % to histidinol dehydrogenase (*hisD* gene) from *Mycobacterium bovis* (strain ATCC BAA-935 / AF2122/97) (EC:1.1.1.23) with an unknown subcellular location. The complete alcohol dehydrogenase *aol-vut3* gene sequence was submitted to GenBank and assigned an accession number MT606179. According to the database, this identified ADH (EC: 1.1.1.23) is NAD dependent and requires  $Zn^{2+}$  as a cofactor i.e. it binds one zinc ion per subunit. The size of the protein is in accordance with that of other ADH isolated in other studies (Wu *et al.* 2013).

The low percentage similarity to this sequence on the database was expected since it is well documented that the use of metagenomic DNA is a valuable tool for tapping into the source of non-culturable bacteria. This inadvertently leads to an increased possibility of discovering novel gene sequences (Sierra-García *et al.* 2014). Similar understanding was supported by researchers in the field of metagenomics and implemented as a tool to identifying novel ADH sequences. Liu *et al.* (2009), conducted a research involving the characterization of two novel metal-dependent long-chain alkyl ADH from *Geobacillus thermodenitrificans* NG80-2.



The low percentage sequence similarity of *aol-vut3* to the ADH observed in the current study is supported by the previous findings (Chapter 3, section 3.1.1). The selection of sampling sites with specific stress factors is key to metagenomics studies (Rocca *et al.* 2019). Thus, holding the potential for isolating highly adapted microorganisms which have been under characterised or are unknown for their ability to degrade and utilize hydrocarbons as a sole carbon source. Consequently, identifying novel ADH genes involved in catalysing the NAD-dependent reversible oxidation of alcohols to aldehydes.

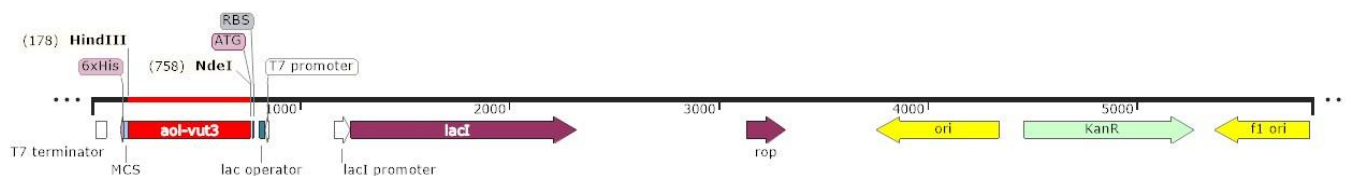
The ADH gene sequence showed similarity to *Mycobacterium bovis* isolate (strain ATCC BAA-935 / AF2122/97) based on Uniprot database search. *M. bovis* is one of the major bacteria known to cause tuberculosis in human and animals (Gelalcha *et al.* 2019). *M. bovis* is known to originate from the taxonomic lineage *Actinobacteria* and have been shown to display 8 % phylum abundance in the soil samples according to microbiome profiling studies conducted previously (Chapter 4, section 4.2.1). *Actinobacteria* have been reported of possessing high resilience to toxic pollutants, described and characterized as a new actinobacterial strain potential for bioremediation of hydrocarbons (Brzeszcz and Kaszycki 2018).

The unknown subcellular location of the ADH based on information obtained from the Uniprot database proved to be challenging in determining the levels of expression due to minimal information in protein solubility. However, following protein purification using the Ni column, the protein was purified from the supernatant of the cell lysate indicating the subcellular location of this protein was in the cytoplasm and the protein was therefore soluble.

### **6.3.3 Cloning, expression and protein purification of Alcohol Dehydrogenase**

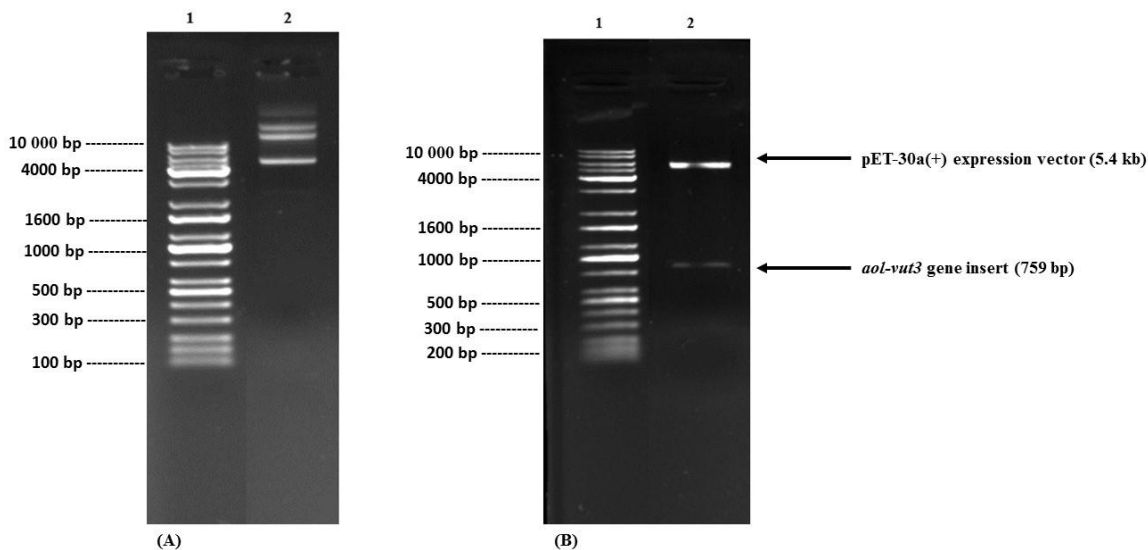
Obtaining high levels of recombinant proteins production was crucial for the current study. The modified genes were cloned into a pET30a(+) expression vector. Complete modified sequences for DNA and protein shown in Appendix B4 and B5. The pET vector system is a powerful system

developed for the cloning and expression of recombinant proteins in *E. coli*. The cloning strategy involved linearizing the pET30a(+) vector by restriction digestion at the *HindIII* and *XbaI* restriction sites. The gene of interest (*aol-vut3*) was inserted and ligated into pET30a(+) expression vector and designated pET30a(+)\_*aol-vut3* (**Figure 6.3**).



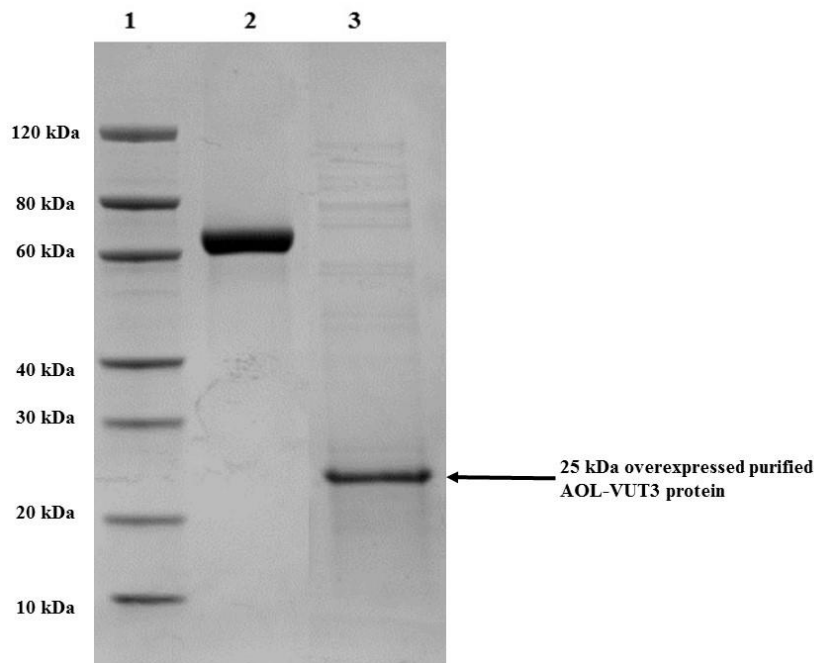
**Figure 6.3:** Linearised map of pET30a(+)\_*aol-vut3*.

The plasmid constructs pET30a(+)\_*aol-vut3* were transformed into *E. coli* BL21 DE3 pLysS competent cells containing T7 RNA polymerase, for expression studies. The plasmid DNA was extracted and assessed on an agarose/EtBr gel (**Figure 6.4(A)**) followed by restriction digest with *XbaI* and *HindIII* (**Figure 6.4(B)**) to confirm the ligation reaction. The extracted plasmid DNA was found to be of high quality and the restriction digest exhibited successful cloning of the insert DNA with the expected band size of 759 bp.



**Figure 6.4:** Analysis for the confirmation of cloning of the *aol-vut3* gene insert. (A) Agarose gel electrophoresis of extracted plasmid DNA. Lane 1: KAPA™ Biosystems Universal ladder (10 kb). Lane 2: Extracted plasmid DNA from transformed *E. coli* BL21 (DE3) pLysS competent cells. (B) Lane 1: KAPA™ Biosystems Universal ladder (10 kb) and Lane 2: Restricted plasmid DNA with *XbaI* and *HindIII* showing the successful cloning of the alcohol dehydrogenase gene insert (759 bp).

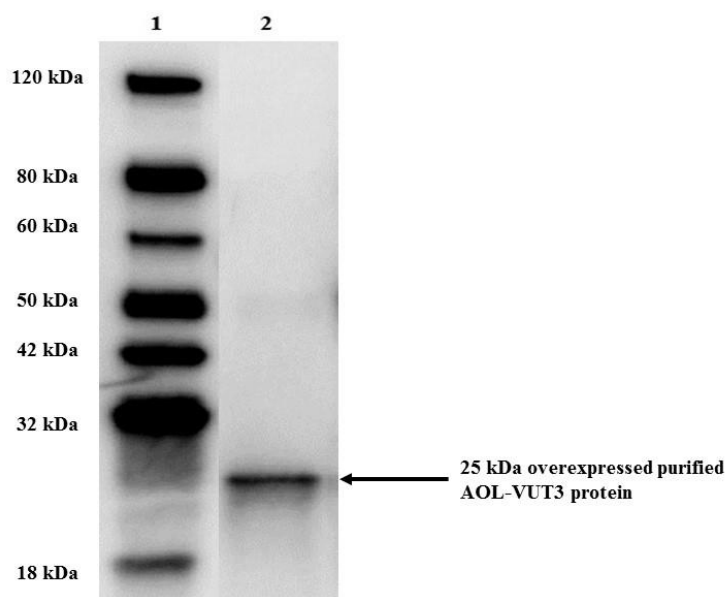
The soluble and insoluble fractions of the cell extracts AOL-VUT3 protein was analysed on SDS-PAGE to determine the expression level of the protein following varying conditions of induction. The estimated molecular mass of the protein was calculated from the amino acid sequence in comparison to the standard protein marker and found to be 25 kDa. Varying experimental conditions for induction time and temperature were setup to maximize expression levels. The SDS PAGE gel image for induced AOL-VUT3 protein in BL21 (DE3) following purification using the affinity to Ni ions using the Akta Start Protein Purification System (Merck KGaA, Germany) is shown in **Figure 6.5**.



**Figure 6.5:** SDS-PAGE analysis of the AOL-VUT3 protein band after overexpression in *E. coli* BL21 (DE3) following induction at 15°C for 16 hours with 0.5 mM IPTG and purification using the Ni column. Lane 1: Protein Marker, GenScript, Lane 2: BSA (2 µg) and Lane 3: Overexpressed purified AOL-VUT3 protein (2 µg).

The overexpressed purified protein migrated as a band of ~25 kDa for the induced protein from the cell lysate thus, correlated to the calculated molecular mass. The results demonstrated that the ADH proteins were expressed in BL21 (DE3) cells appropriately. Additionally, the results showed that the 6× His-tag that was added to the C-terminal facilitated the purification process. This was confirmed by the presence of a single band corresponding to the expected protein size (**Figure 6.5**). The optimal induction condition was observed using 0.5 mM IPTG at 15 °C for an incubation period of 16 hours. These parameters revealed to be optimum for ensuring the overexpression of these recombinant proteins due to their solubility and lack of storage as inclusion bodies. The concentration obtained for the protein using Bradford protein assay exhibited 0.57 mg/ml.

Western blot analysis was carried out to confirm the purity, molecular weight and expression results of the purified protein using a primary anti-His antibody (Mouse-anti-His mAb) **Figure 6.6**. The results obtained showed the binding of the anti-His antibody to the histidinol ADH protein with 6x His-tags at an expected band size of 25 kDa. Thus, confirming the successful expression of the target protein.



**Figure 6.6:** Western blot analysis for AOL-VUT3 expressed in *E. coli* BL21 (DE3) probed with anti-His antibody. Lane 1: Western blot marker, Lane 2: Purified AOL-VUT3 protein of 25kDa using nickel ion affinity chromatography.

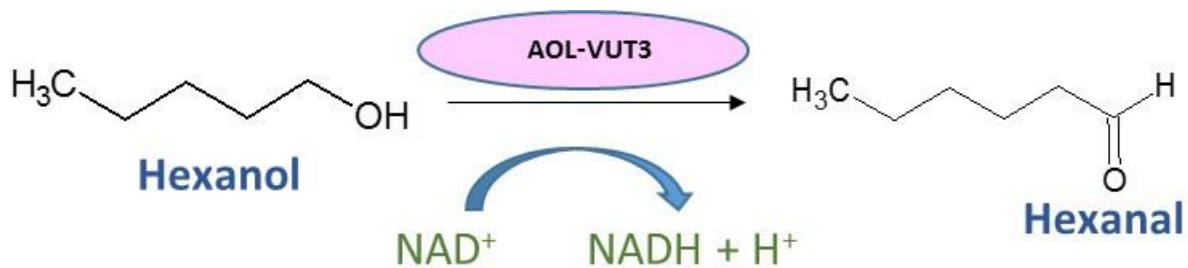
#### 6.3.4 Optimal alcohol dehydrogenase (AOL-VUT3) concentration

Varying concentrations of the purified novel AOL-VUT3 enzyme (0.001 mg/ml, 0.003 mg/ml, 0.007 mg/ml and 0.014 mg/ml) were tested in the ADH assay as stipulated by Liu *et al.* (2009).

This was carried out to determine the optimal concentration of enzyme to be used for the kinetics study and the cocktail of enzymes designed in Chapter 8, section 8.1.2. The effect of varying enzyme concentrations in this assay is depicted by change in absorbance value at a wavelength of 340 nm. This is indicative of the amount of NADH formed by the reduction of the coenzyme NAD<sup>+</sup> using different enzyme concentrations. Absorbance data and graphs shown in Appendix C. Moreover, NAD<sup>+</sup> was used as a coenzyme in this study since most ADH produced from microorganisms exhibit NAD<sup>+</sup>-dependent ADH activity during growth on different alcohols (Jendrossek *et al.* 1990).

A similar trend for each of the enzyme concentrations were exhibited and peaks emerged after 20 seconds leading to plateau formation. Such a trend agrees with the general literature with regards to the functioning of enzymes. Enzyme molecules follow these steps which involves binding to its substrate, catalysing the reaction and lastly releasing the product. The following steps each require a specific amount of time and the more enzyme molecules that are available, the more product can be produced. Therefore, the more enzyme there is available, the more quickly the substrate can be converted into product. Once all the enzyme has been used up, the reaction will be complete. It can be clearly seen that the enzyme catalyses the reaction rapidly until 20 seconds when it can be assumed that the enzyme has been exhausted and all possible product has been made.

The maximum absorbance (0.060 a.u) was obtained at a concentration of 0.003 mg/ml at 20 seconds and generally no further increase for the duration of the reaction. Thus, signaling the completion of the enzymatic reaction since either enzyme or substrate was the limiting reagent. Moreover, maximum NADH production took place within the first 20 seconds, an indication of a very fast rate of reaction. The findings were in agreement with the previous studies (Wu *et al.* 2013) which showed the reduction of NAD<sup>+</sup> to NADH by an ADH exhibits UV absorption at 340 nm and an increase in absorbance value, is indicative of the amount of NADH produced upon catalyzation. Thus, confirming the conversion of hexanol into hexanal. Therefore, it can be concluded that the class of novel dehydrogenase enzyme identified in this study as AOL-VUT3 is indeed an ADH with the ability to convert hexanol to hexanal as schematically shown in **Figure 6.7**.



**Figure 6.7** Schematic representation showing the enzymatic reaction of the novel alcohol dehydrogenase AOL-VUT3 in the conversion of hexanol to hexanal.

From the results it can be deduced that an increase in enzyme concentration has a direct effect on the amount of NAD<sup>+</sup> converted to NADH as depicted by the increase in the absorbance values. Thus, the amount of hexanal increases due to the conversion of hexanol. Statistical analysis using two-way analysis of variance (ANOVA) confirmed the concentration of AOL-VUT3 enzyme accounting for 67 % of the total variance which was considered to be of significant effect ( $p = 0.0021$ ). The highest absorbance was obtained at 0.003 mg/ml with subsequent decline at concentrations of 0.007 mg/ml and 0.0014 mg/ml respectively. Thus, the optimal concentration of enzyme for both the kinetics study and the enzyme cocktail preparation was subsequently identified to be 0.003 mg/ml for successive use.

### 6.3.5 Optimal hexanol substrate concentration and enzyme kinetics

Varying substrate concentrations of hexanol (25 mM, 50 mM, 100 mM, 150 mM and 200 mM) were tested in the ADH assay. The current research deployed and determined the optimal concentration of substrate for the kinetics study and the cocktail of enzymes for subsequent use. The effects of varying substrate concentrations were depicted by change in absorbance value at a wavelength of 340 nm. Thus, an indicative of the amount of NADH formed by the reduction of the coenzyme NAD<sup>+</sup>. Absorbance data and graphs for the assay shown in Appendix C. NAD<sup>+</sup>

was used as a coenzyme in the current study since most ADH produced from microorganisms exhibit NAD<sup>+</sup>-dependent ADH activity during growth on different alcohols and this was also observed by Jendrossek *et al.* (1990).

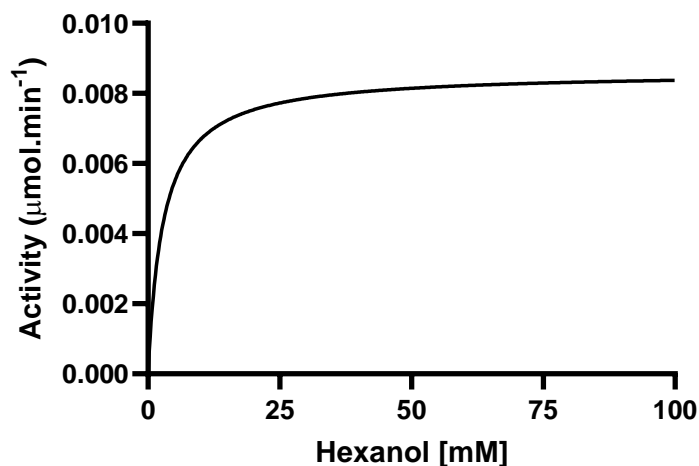
A consistent trend was observed in which the absorbance values increased between 20 – 60 seconds followed by plateau formation. The maximum absorbance value of 0.107 a.u was detected for 150 mM hexanol, thus, an indication of maximum NAD<sup>+</sup> reduction to NADH at that concentration. From the results it can be concluded that an increase in substrate concentration has an effect on the amount of NAD<sup>+</sup> converted to NADH as shown by the increase in the absorbance values. A 2-way ANOVA analysis confirmed the concentration of hexanol accounting for 9.174 % of the total significant variance ( $p = 0.0006$ ). Further confirming that AOL-VUT3 can be classified as an alcohol dehydrogenase due to the production of hexanal.

From the results it is evidently obvious to conclude that an increase in the substrate concentration has an effect on the amount of NAD<sup>+</sup> converted to NADH. There is a gradual increase in absorbance from 50 mM to 150 mM since more hexanol is available for conversion to hexanal. However, at concentration 200 mM, the absorbance value starts to decline. This can be explained by the fact that as the concentration of substrate increases, the enzyme becomes saturated with substrate and adding more substrate will not significantly affect the rate of the reaction.

The absorbance values obtained at the 60 sec interval for the various hexanol substrate concentrations were used to construct the Michaelis-Menten graph for the AOL-VUT3 protein using GraphPad Prism (**Figure 6.8**). The graph describes the rate of the enzymatic reaction in relation to the reaction rate (rate of formation of product) to the concentration of a substrate. It can be concluded that the enzymatic reaction followed a typical first order of the Michaelis-Menten function when hexanol was used as a substrate in the presence of NAD<sup>+</sup>. Similar results were also obtained by Zhu *et al.* (2017) but using a novel glutamate dehydrogenase from *Geotrichum candidum* with high affinity for hexanol. From the graph it can be depicted that the initial velocity



of the reaction increases as the hexanol concentration increases but a peak of 0.01  $\mu\text{mol}/\text{min}$  was observed at 150 mM substrate concentration and thereafter decreases.



**Figure 6.8:** The enzymatic activity of novel alcohol dehydrogenase AOL-VUT3 at varying concentration of the substrate, hexanol. The activity of the enzyme was measured as the amount of  $\text{NAD}^+$  reduced to NADH by the action of the enzyme on its substrate at incubation temperature of 30 °C for 60 second. Absorbance values were recorded spectrophotometrically at 340 nm wavelength. The kinetic data was fitted into Michaelis-Menten model using GraphPad Prism Version 8.4.1 (460).

The Michaelis-Menten constant ( $K_m$ ) value obtained from the Michaelis-Menten graph represents the concentration of substrates used by the enzyme when the reaction reaches half of maximal velocity ( $V_{\text{max}}$ ). The  $V_{\text{max}}$  and  $K_m$  values were determined to be 0.008614  $\mu\text{mol}\cdot\text{min}^{-1}$  and 2.875 mM, respectively (**Figure 6.8**). The relationship between rate of reaction and concentration of substrate depends on the affinity of the enzyme for its substrate. Since a small  $K_m$  value indicates that the enzyme has a high affinity for the substrate, it can be concluded that AOL-VUT3 displays a high affinity for hexanol as a substrate. By using 2.875 mM of hexanol, the AOL-VUT3 protein reached half of  $V_{\text{max}}$  and thus, acts at a more or less constant rate, regardless of variations in the

concentration of substrate within the physiological range. The obtained novel enzyme exhibited a property that distinguished it from previously reported hexanol dehydrogenases (Veeranagouda *et al.* 2008, Wu *et al.* 2013, Tavakoli and Hamzah 2017). The reports showed higher values for  $K_m$  based on kinetics studies for hexanol as a substrate than depicted in the current study. Accordingly, this enzyme can be implemented in future research involving the conversion of other alkanols into alkanals due to its high affinity for hexanol.

It is important to understand the harmful and long term effects of hexanol pollution in soils on plants, animals and humans (Zhu *et al.* 2017). Hexanol is present in the soil due to hydrocarbon pollution by pesticide application to crops and crude oil spillages. Hexanol is also known to build up over time as a secondary pollutant following the breakdown of hexane, a toxic compound, by naturally occurring bacteria in the soil under the control of monooxygenases under aerobic conditions as shown by (Huang *et al.* 2018). Therefore, the discovery of this novel alcohol dehydrogenase (AOL-VUT3) is invaluable in transforming this pollutant into hexanal which can feed into the degradation pathway and speed up the process of converting hexanol into hexanal, which is a slow naturally occurring process.

A few ADH targeted at hydrocarbon degradation have been isolated. Wu *et al.* (2013) isolated a novel thermostable alcohol dehydrogenase (ADH) from *Thermococcus kodakarensis* KOD1 for enantioselective bioconversion of aromatic secondary alcohols. This ADH was found to prefer secondary alcohols and accepted various ketones and aldehyde as substrates. Additionally, a NAD-dependent secondary alcohol dehydrogenase was purified by Ludwig *et al.* (1995) from the alkane-degrading bacterium, *Rhodococcus erythropolis* ATCC 4277. The ADH was active against a broad range of substrates, particularly long-chain secondary aliphatic alcohols. Optimal oxidation activity was observed with linear 2-alcohols containing between 6 and 11 carbon atoms including secondary alcohols as long as 2-tetradecanol at a rate of 25 %. Recently, Tavakoli and Hamzah (2017), expressed, isolated and characterised a benzyl Alcohol Dehydrogenase (BADH) from *Rhodococcus ruber* UKMP-5M, a member of medium chain alcohol dehydrogenases.

## 6.4 Conclusion

The enzyme was successfully purified using Akta Start Protein Purification System (Merck KGaA, Germany) and Western blot analysis confirmed the purity and molecular weight of the protein with computed size of 25 kDa. The alcohol dehydrogenase assay showing the activity of the purified novel AOL-VUT3 enzyme by varying both enzyme and substrate concentration were found to be optimal at 0.003 mg.ml<sup>-1</sup> and 150 mM, respectively. The 2-way ANOVA showed that varying concentrations had a significant effect on the amount of NAD<sup>+</sup> reduced to NADH as a result of action of alcohol dehydrogenase AOL-VUT3 on hexanol and subsequent conversion into hexanal. Accordingly, the enzyme successfully converted hexanol to a less complex hexanal, which is an intermediate product in the complete degradation or transformation of hexanol. Accordingly, kinetics study showed a low K<sub>m</sub> value of 2.875 mM with a V<sub>max</sub> of 0.008614 μmol.min<sup>-1</sup>, showing the high affinity of the enzyme to its substrate (hexanol) thus, making it a valuable candidate for degradation of alkanol. Since, the discovered enzyme is novel it is important to conduct crystallographic study of such indispensable protein and solve its structure.

## 6.5 References

- Bradford, M.M., 1976. A Rapid and Sensitive Method for the Quantitation Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Analytical Biochemistry*, 72, 248–254.
- Brzeszcz, J. and Kaszycki, P., 2018. Aerobic bacteria degrading both n-alkanes and aromatic hydrocarbons: an undervalued strategy for metabolic diversity and flexibility. *Biodegradation*, 29 (4), 359–407.
- Castiñeiras, T.S., Williams, S.G., Hitchcock, A.G., and Smith, D.C., 2018. E. coli strain engineering for the production of advanced biopharmaceutical products. *FEMS Microbiology Letters*, 365 (15), 1–10.
- Chaudhary, D.K., Khulan, A., and Kim, J., 2019. Development of a novel cultivation technique for uncultured soil bacteria. *Scientific Reports*, 9 (1), 1–11.
- Choi, T.J. and Geletu, T.T., 2018. High level expression and purification of recombinant flounder growth hormone in E. coli. *Journal of Genetic Engineering and Biotechnology*, 16 (2), 347–355.
- Elumalai, P., Parthipan, P., Karthikeyan, O.P., and Rajasekar, A., 2017. Enzyme-mediated biodegradation of long-chain n-alkanes (C32 and C40) by thermophilic bacteria. *3 Biotech*, 7 (2), 1–10.
- Gelalcha, B.D., Zewude, A., and Ameni, G., 2019. Tuberculosis Caused by Mycobacterium bovis in a Sheep Flock Colocated with a Tuberculous Dairy Cattle Herd in Central Ethiopia. *Journal of Veterinary Medicine*, 2019, 1–6.
- Gurusinghe, S., Brooks, T.L., Barrow, R.A., Zhu, X., and Weston, L.A., 2019. Technologies for the Selection, Culture and Metabolic. *Molecules*, 24, 1–20.
- Hausjell, J., Halbwirth, H., and Spadiut, O., 2018. Recombinant production of eukaryotic cytochrome P450s in microbial cell factories. *Bioscience Reports*, 38 (2), 1–13.
- Hu, W., Liu, N., Tian, Y., and Zhang, L., 2013. Molecular cloning, expression, purification, and

- functional characterization of dammarenediol synthase from panax ginseng. *BioMed Research International*, 2013 (3).
- Huang, Y., Xiao, L., Li, F., Xiao, M., Lin, D., Long, X., and Wu, Z., 2018. Microbial degradation of pesticide residues and an emphasis on the degradation of cypermethrin and 3-phenoxy benzoic acid: A review. *Molecules*, 23 (9).
- Jendrossek, D., Kruger, N., and Steinbuchel, A., 1990. Characterization of alcohol dehydrogenase genes of derepressible wild-type *Alcaligenes eutrophus* H16 and constitutive mutants. *Journal of Bacteriology*, 172 (9), 4844–4851.
- Jia, B. and Jeon, C.O., 2016. High-throughput recombinant protein expression in *Escherichia coli*: Current status and future perspectives. *Open Biology*, 6 (8).
- Laemmli, U.K., 1970. 227680a0. *Nature*, 227, 680–685.
- Li, X., Li, Y., Wei, D., Li, P., Wang, L., and Feng, L., 2010. Characterization of a broad-range aldehyde dehydrogenase involved in alkane degradation in *Geobacillus thermodenitrificans* NG80-2. *Microbiological Research*, 165 (8), 706–712.
- Liu, X., Dong, Y., Zhang, J., Zhang, A., Wang, L., and Feng, L., 2009. Two novel metal-independent long-chain alkyl alcohol dehydrogenases from *Geobacillus thermodenitrificans* NG80-2. *Microbiology*, 155 (6), 2078–2085.
- Ludwig, B., Akundi, A., and Kendall, K., 1995. A long-chain secondary alcohol dehydrogenase from *Rhodococcus erythropolis* ATCC 4277. *Applied and Environmental Microbiology*, 61 (10), 3729–3733.
- Rocca, J.D., Simonin, M., Blaszcak, J.R., Ernakovich, J.G., Gibbons, S.M., Midani, F.S., and Washburne, A.D., 2019. The Microbiome Stress Project: Toward a global meta-analysis of environmental stressors and their effects on microbial communities. *Frontiers in Microbiology*, 10 (JAN).
- Rosano, G.L. and Ceccarelli, E.A., 2014. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology*, 5 (APR), 1–17.

- Shinde, P., Musameh, M., Gao, Y., Robinson, A.J., and Kyrtziz, I. (Louis), 2018. Immobilization and stabilization of alcohol dehydrogenase on polyvinyl alcohol fibre. *Biotechnology Reports*, 19, e00260.
- Sierra-García, I.N., Alvarez, J.C., De Vasconcellos, S.P., De Souza, A.P., Dos Santos Neto, E.V., and De Oliveira, V.M., 2014. New hydrocarbon degradation pathways in the microbial metagenome from brazilian petroleum reservoirs. *PLoS ONE*, 9 (2).
- Tavakoli, A. and Hamzah, A., 2017. Biochemical characterization of recombinant benzyl alcohol dehydrogenase from *Rhodococcus ruber* UKMP-5M, 3 (4), 613–622.
- Veeranagouda, Y., Benndorf, D., Heipieper, H.J., and Karegoudar, T.B., 2008. Purification and characterization of NAD-dependent n-butanol dehydrogenase from solvent-tolerant n-butanol-degrading *Enterobacter* sp. VKGH12. *Journal of Microbiology and Biotechnology*, 18 (4), 663–669.
- Wentzel, A., Ellingsen, T.E., Kotlar, H.K., Zotchev, S.B., and Throne-Holst, M., 2007. Bacterial metabolism of long-chain n-alkanes. *Applied Microbiology and Biotechnology*, 76 (6), 1209–1221.
- Wu, X., Zhang, C., Orita, I., Imanaka, T., Fukui, T., and Xing, X.H., 2013. Thermostable alcohol dehydrogenase from *Thermococcus kodakarensis* KOD1 for enantioselective bioconversion of aromatic secondary alcohols. *Applied and Environmental Microbiology*, 79 (7), 2209–2217.
- Zhu, J., Lu, K., Xu, X., Wang, X., and Shi, J., 2017. Purification and characterization of a novel glutamate dehydrogenase from *Geotrichum candidum* with higher alcohol and amino acid activity. *AMB Express*, 7 (1).

---

# Chapter 7

## Cloning, Expression, Purification and Kinetic Studies of Two Novel Aldehyde Dehydrogenases (DHY- SC-VUT5 and DHY-G-VUT7) Mined from Metagenome of Hydrocarbon Contaminated Soils

---

Article submitted for publication

### *Abstract*

Dehydrogenases are vital for aerobic hydrocarbon degradation and is involved in the last step of catalysing the oxidation of aldehydes to carboxylic acids. Fifteen fosmid clones with hydrocarbon degrading potential were functionally screened to identify dehydrogenase enzymes. Accordingly, the fosmid insert of the positive clones were sequenced using PacBio next generation sequencing platform and *de novo* assembled using CLC Genomic Work Bench. The 1233 bp long open reading frame (ORF) for DHY-SC-VUT5 was found to share a protein sequence similarity of 96.7 % to short-chain dehydrogenase from *E. coli*. The 1470 bp long ORF for DHY-G-VUT7 was found to share a nucleotide sequence similarity of 23.9 % to glycine dehydrogenase (decarboxylating) (EC 1.4.4.2) from *Caulobacter vibrioides* (strain NA1000 / CB15N) (*Caulobacter crescentus*). The *in silico* analyses and blast against UNIPROT protein database with the stated similarity show that the two dehydrogenases are novel. Both dehydrogenase gene sequences were codon-optimized as per *Escherichia coli* codon preference and synthesized. The

synthesised genes were successfully cloned into a pET-30a(+) expression vector and transformed into chemically competent *E. coli* BL21 (DE3) cells. The protein expression was induced using 0.5 mM Isopropyl- $\beta$ -D-thiogalactoside (IPTG) and incubated at 15 °C for 16 hours. The highest relative activity was observed at substrate concentrations of 150 mM and 50 mM for DHY-SC-VUT5 and DHY-G-VUT7, respectively. The  $K_m$  values were found to be 13.77 mM with a  $V_{max}$  of 0.009135  $\mu$ mol/min and 2.832 mM with a  $V_{max}$  of 0.005886  $\mu$ mol/min for DHY-SC-VUT5 and DHY-G-VUT7, respectively. The DHY-G-VUT7 had a higher substrate affinity as displayed by its lower  $K_m$  value than DHY-SC-VUT5 dehydrogenase thus, a potent and efficient enzyme for alkyl aldehyde conversion to carboxylic acid. The dehydrogenases could be used to make up enzyme cocktails for biodegradation of alkanes. Moreover, since the discovered enzymes are novel it is important to conduct crystallographic studies of such invaluable enzymes to solve their structures and explore the down stream applications.

**Keywords:** Microorganisms, metagenomics, aldehyde dehydrogenase, hydrocarbons, alkanal, fosmid clones, expression host, bioremediation.



<b>TABLE OF CONTENTS</b>	<b>PAGES</b>
7.1 Introduction .....	253
7.2 Materials and Methods .....	255
7.2.1 Selection of fosmid clones .....	255
7.2.2 Induction of fosmid clones and extraction of fosmid DNA.....	255
7.2.3 Fosmid DNA construct integrity assessment .....	256
7.2.4 <i>De novo</i> sequence assembly and analysis .....	256
7.2.5 Gene synthesis and cloning.....	257
7.2.6 Transformation of bacterial cells .....	258
7.2.7 Protein induction and expression .....	259
7.2.8 Protein purification and visualization .....	260
7.2.9 Western blot analysis .....	261
7.2.10 ALDH activity assay.....	261
7.2.11 Kinetic studies of novel DHY-SC-VUT5 and DHY-G-VUT7 dehydrogenases ..	262
7.3 Results and Discussion.....	263
7.3.1 The quality of isolated fosmid DNA.....	263
7.3.2 Sequence analysis and ALDH gene synthesis .....	264
7.3.3 Cloning, expression and protein purification.....	266
7.3.4 Optimal aldehyde dehydrogenase (DHY-SC-VUT5 and DHY-G-VUT7) concentration.....	271
7.3.5 Optimal hexanal substrate concentration and enzyme kinetics .....	273
7.4 Conclusion.....	279
7.5 References .....	280

## 7.1 Introduction

Anthropogenic disturbances such as hydrocarbon contamination in soil environments can affect the microbial community structure and biodiversity since it is toxic to microbial cells (Jung *et al.* 2016, Liu *et al.* 2018). Some microorganisms have the ability to use hydrocarbons as a sole carbon source thus such soils represent a complex environment teeming with microbial diversity which is preferred for gene mining and natural product discovery (Gurusinghe *et al.* 2019). Polluted soil sites can be targeted with confidence in identifying novel aldehyde dehydrogenases (ALDHs) involved in aerobic pathways for petroleum compound breakdown (Hausjell *et al.* 2018). Furthermore, metagenomic DNA from these soils reveal the potential of unculturable microorganisms that are otherwise troublesome to culturing techniques as a source of novel enzymes (Berini *et al.* 2017). Numerous studies have shown the success of employing this technique for example novel lipases (Yan *et al.* 2017), alkane hydroxylases (Silva *et al.* 2019), monooxygenases (Li *et al.* 2013, Bouhajja *et al.* 2017), alcohol dehydrogenases (Liu *et al.* 2009) to list but a few.

ALDHs are a diverse group of related enzymes that catalyze the oxidization of aldehydes to carboxylic acids, using NAD<sup>+</sup> or NADP<sup>+</sup> as the coenzyme. ALDHs are ubiquitous in all life forms and are classified as detoxification enzymes involved with coping with toxic aldehydes generated from various cellular metabolisms. However, more specifically related to this project, aldehyde dehydrogenases are involved in numerous bioconversion processes for example the degradation of alkanes in which alkyl aldehydes are generated as intermediates and represents the third step of the alkane degradation pathway. Due to their specific function, these biocatalysts are of great interest for industrial applications (Shinde *et al.* 2018).

To date there are only a few reports of ALDHs of bacterial origin involved in oxidation of short and medium chain alkyl aldehydes (C<sub>1</sub> – C<sub>14</sub>) such as those from *Acinetobacter* sp. strain M-1 (Ald1, C<sub>4</sub>–C<sub>14</sub>) (Ishige *et al.* 2000), *Geobacillus thermodenitrificans* NG80-2 (C<sub>1</sub> – C<sub>20</sub>) (Li *et al.*

2010), *Rhodococcus erythropolis* (ThcA, C<sub>1</sub>–C<sub>8</sub>), *Acetobacter pasteurianus* SKU1108 (Yakushi *et al.* 2018) and *Alcanivorax borkumensis* SK2T (Gregson *et al.* 2019).

This study set out to develop two expression plasmids for two identified novel aldehyde dehydrogenase genes obtained from fosmid clones and subsequent purification and characterization of the proteins. Induction parameters such as temperature, time and IPTG concentration were tested to ensure maximum expression of the candidate protein. Enzyme activity of the purified proteins were assessed using an aldehyde dehydrogenase assay by Li *et al.* (2010). It utilizes aldehyde dehydrogenase enzymes to oxidize aldehydes into carboxylic acids. In this process the co-factor NAD (nicotinamide adenine dinucleotide) is converted into NADH and absorbs light at 340 nm. The optimal enzyme and substrate concentrations were determined using the assay followed by enzyme kinetics studies to determine the affinity of the enzymes to convert aldehydes into a less toxic compound, carboxylic acids. Thus, the identification and characterisation of novel aldehyde dehydrogenases will be valuable in targeting bioremediation of alkyl aldehyde contaminated soil environments.

## 7.2 Materials and Methods

### 7.2.1 Selection of fosmid clones

A total of 15 candidate fosmid clones from the prepared metagenomic library were selected based on functional screening analysis using three hydrocarbon substrates (hexadecane, octadecene and cyclohexane). The functional screening of the fosmid library has been explained in more detail in Chapter 3. Five clones were chosen for each of the substrates based on their growth rate and colony size on minimal media (Bushnell Haas media) supplemented with a hydrocarbon source (**Table 7.1**).

**Table 7.1:** Selected fosmid clones based on growth rates on three different hydrocarbon substrates.

Hydrocarbon substrates	Selected clones
Hexadecane	pFos_A1, pFos_A7, pFos_B2, pFos_B3 and pFos_F2
Octadecene	pFos_D3, pFos_D8, pFos_F1, pFos_F3 and pFos_A4
Cyclohexane	pFos_A2, pFos_C7, pFos_B7, pFos_D7 and pFos_G3

### 7.2.2 Induction of fosmid clones and extraction of fosmid DNA

An overnight culture of each fosmid clone was grown in 10 ml of Luria-Bertani (LB) broth supplemented with chloramphenicol (12.5 µg/ml) to make a stock culture and 1 ml was used to inoculate 25 ml LB broth supplemented with chloramphenicol (12.5 µg/ml). This was autoinduced by adding 50 µl 500X CopyControl™ Fosmid Autoinduction Solution as previously described by the manufacturer (Epicentre, USA). The autoinduced culture was grown overnight at 37 °C in a

shaking incubator at 170 rpm. EPI 300™ bacterial host cells grown in 25 ml LB broth was used as the negative control. The bacterial cells were harvested from 10 mls of overnight culture by centrifugation at 4000 rpm for 5 minutes. The supernatant was discarded and the pellet resuspended in 250 µl suspension buffer (Epicentre, USA). The fosmid DNA was thereafter, extracted using the GeneJET Plasmid Miniprep Kit (Thermoscientific) according to the manufacturer's instructions. Isolated fosmid DNA was eluted in 50 µl Elution buffer supplied by the manufacturer and stored at -80 °C.

### **7.2.3 Fosmid DNA construct integrity assessment**

The extracted fosmid constructs from 15 positive fosmid clones were visualised on a 0.8 % (w/v) agarose gel stained with ethidium bromide and run at 100V for 30 minutes. The pCC2FOS™ vector supplied with the CopyControl™ HTP Fosmid Library Production Kit (Epicentre, USA) was run on the agarose gel electrophoresis as a control. The fosmid constructs were then sequenced using PACBIO SMRT® next generation sequencing platform at Inqaba Biotec™ (Pretoria, South Africa).

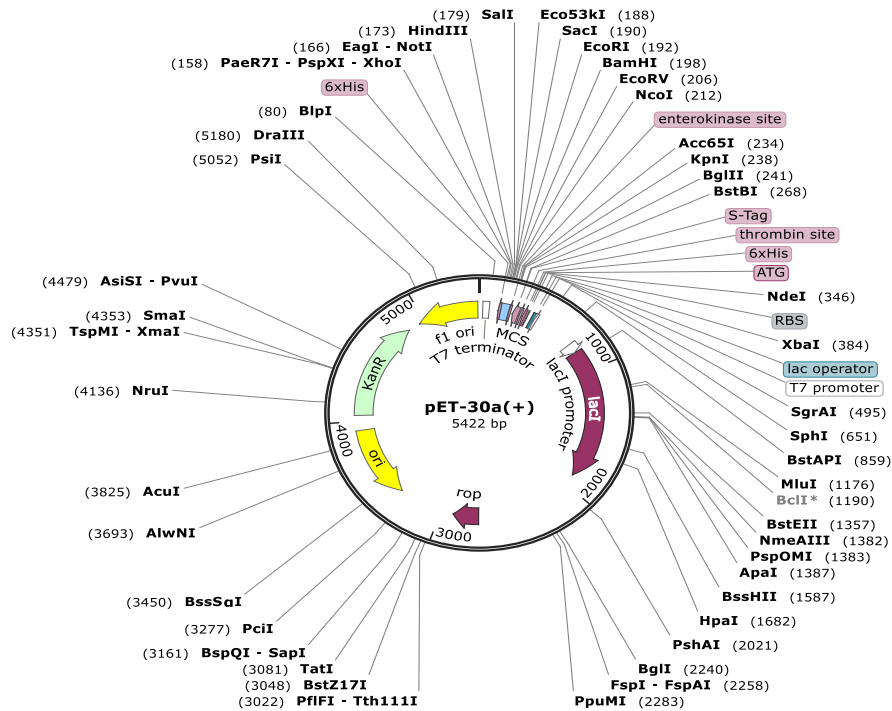
### **7.2.4 *De novo* sequence assembly and analysis**

*De novo* assembly was carried out on the PACBIO sequences obtained from Inqaba Biotec™ using the CLC Genomics Workbench 11.0.1 (CLC Bio, Qiagen). The quality of data was checked, the contig sequence length was set at 200 bp and filtering was carried out based on default parameters in the software. The software program was used to extract the longest contig assembly. The open reading frames along with the translated protein were determined from the assembled contig sequences for each of the fifteen fosmid clones using ORF finder on the NCBI database (<https://www.ncbi.nlm.nih.gov/orffinder/>). A protein Basic Local Alignment Search Tool (BLAST) called Uniprot-Swiss prot (<https://www.uniprot.org/>) was carried out on the individual ORF sequences in order to find regions of local similarity between known sequences to identify

members of gene families involved in hydrocarbon degradation pathways. These ORFs were flagged for further analysis.

### 7.2.5 Gene synthesis and cloning

The two identified ORF sequences designated *dhy-sc-vut5* and *dhy-g-vut7* were codon optimised for codon compatibility in *E.coli* to increase protein expression and His tags were added to the sequence to facilitate for downstream protein purification and detection using the Akta Start Protein Purification System (Merck KGaA, Germany). The optimised gene sequences were synthesized by GenScript® USA Inc. The synthesised DNA sequences was cloned into a pET-30a(+) expression vector system (Novagen, USA) (**Figure 7.1**) with a T7 *lac* promoter according to the pET System manual (Novagen, USA). The DNA sequence was cloned at cloning sites *NdeI* and *Hind III*.



**Figure 7.1:** Map of pET-30a(+) expression vector with all the features displayed.

### 7.2.6 Transformation of bacterial cells

The preparation of chemically competent *E. coli* BL21 (DE3) pLysS cells was carried out using the Rubidium chloride method. An overnight culture of 0.5 ml was used to inoculate 500 ml LB broth. This was incubated at 37 °C with aeration until the optical density (OD<sub>600 nm</sub>) of 0.48 using the NanoDrop™ One spectrophotometer (Thermo Fisher Scientific™) was reached. These were decanted into 20 ml tubes and incubated on ice for 15 minutes followed by centrifugation at 5000xg at 4 °C for 5 minutes until all cells were pelleted. The supernatant was discarded and 0.4 volumes of TfbI solution (30 mM Potassium acetate, 100 mM Rubidium Chloride, 10 mM Calcium chloride, 50 mM Manganese chloride and 15 % (v/v) glycerol) was used to resuspend the pellet and thereafter kept on ice for 15 minutes. The cells were pelleted at 5000xg for 5 minutes at 4 °C and resuspended in 0.04 volumes of TfbII solution pH 6.5 (10 mM MOPS (3-(N-morpholino)propanesulfonic acid), 75 mM Calcium chloride, 10 mM Rubidium chloride and 15% (v/v) glycerol). This solution was kept on ice for 15 minutes and thereafter 0.5 ml aliquots were placed in 1 ml cryotubes tubes and stored at -80 °C.

The transformation of the chemically competent *E. coli* BL21 (DE3) pLysS cells with pET-30a(+) expression vector ligated with the gene sequences of interest (*dhy-sc-vut5* and *dhy-g-vut7*) was carried out according to the protocol by for E.cloni® 10G chemically competent cells (Lucigen®). This involved thawing 50 µl of chemically competent *E. coli* BL21 (DE3) pLysS cells on ice and transformation was carried out using 4 µl plasmid vector with insert. The mixture was incubated on ice for 30 minutes followed by incubation for 45 seconds at 42 °C in a water bath. The heated solution was placed immediately on ice for two minutes and 960 µl of Recovery medium (E.cloni® 10G, Lucigen). Transformation was also carried out using Rosetta™2 (DE3) *E.coli* cells to check for enhanced expression levels. The culture was allowed to shake at 250 rpm at 37 °C for 1 hour and thereafter spread plated onto LB agar plates supplemented with kanamycin (50 µg/ml). This was allowed to grow overnight at 37 °C. A negative control was also setup and used plasmid DNA without insert to transform the host cells. The growth of colonies on the plate indicated the transformed bacteria which were selected for further analysis.

A single transformed colony for each plasmid and insert resulting in pET30a(+)\_dhy-g-vut7 and pET30a(+)\_dhy-sc-vut5 were picked and grown in 100 ml LB supplemented with kanamycin (50 µg/ml) overnight at 37 °C with aeration. Glycerol stocks of these were prepared and stored at -80 °C. In order to carry out quality control and confirm the successful transformation of the *E. coli* host cells, 1.5 ml of the overnight culture was used to isolate plasmid DNA from the pelleted cells using the ZR miniprep plasmid isolation kit (Zymo Research) according to the manufacturer's instructions. The isolated plasmid DNA was visualised on a 0.8 % (w/v) agarose gel stained with ethidium bromide to determine the quality of the plasmid DNA. Restriction analysis was also carried out on the plasmid DNA using a double digestion involving the enzymes *XbaI* and *HindIII*. Digestion was allowed to proceed for 3 hours at 37 °C and the restriction mixture was separated on a 0.8 % (w/v) agarose gel stained with ethidium bromide. This was carried out to determine the size of the inserts for each clone and the successful transformation of the *E. coli* host cells.

### **7.2.7 Protein induction and expression**

The expression of novel His-tagged proteins (DHY-SC-VUT5 and DHY-G-VUT7) was carried out using *E. coli* BL21 (DE3) pLysS cells as the host strain. An overnight seed culture for each of the clones pET30a(+)\_dhy-g-vut7 and pET30a(+)\_dhy-sc-vut5 were prepared from glycerol stocks and used to inoculate 200 ml LB broth supplemented with kanamycin (50 µg/ml) with aeration at 37 °C. The cells were grown until they reached an optical density (OD<sub>600</sub> nm) of 0.5-0.6 using the NanoDrop™ One spectrophotometer (Thermo Fisher Scientific™) and followed by induction with Isopropyl-β-D-thiogalactoside (IPTG) at a final concentration of 1mM. The expression was monitored over various time periods at 25 °C with shaking at 250 rpm. A corresponding culture without IPTG was used as the uninduced control. In order to obtain the highest expression levels for each clone, parameters such as growth temperatures (16 °C, 25 °C and 37 °C), the inducer IPTG concentration (0.3 – 1 mM) final concentration, varied densities of optical density (OD<sub>600</sub> nm) for induction (0.3- 0.9) and the growth period (3 hours, 8 hours and overnight) were assessed.



Following the various parameters for induction, the bacterial cells were pelleted by centrifugation at 5000 x g for 10 minutes and protein extraction was carried out using the B-PER<sup>®</sup> Bacterial Protein Extraction Reagent (Thermo Scientific<sup>™</sup>) according to the manufacturer's instructions. A test expression protein sample and a solubility protein sample were collected for each expression analysis for induced and uninduced experiments. The uninduced samples represented the negative controls and BSA at 1 µg and 2 µg concentrations were used as a negative control.

## 7.2.8 Protein purification and visualization

The expressed protein profile for DHY-SC-VUT5 (pET30a(+)\_dhy-sc-vut5) and DHY-G-VUT7 (pET30a(+)\_dhy-g-vut7) were visualised using Sodium dodecyl sulfate– polyacrylamide gel electrophoresis (SDS–PAGE) (15 % resolving gel and 5% stacking gel) as described by the Laemmli procedure (Laemmli 1970). Protein samples were prepared by adding Laemmli 2x concentrate sample buffer (Sigma-Aldrich) to the protein supernatant and this was heat denatured at 95 °C in a water bath for 5 minutes. The proteins were stained with Coomassie blue stain solution (40 % ethanol, 0.125 % Coomassie<sup>®</sup> blue, distilled water and 10 % acetic acid) and destained with Coomassie blue destain solution (5 % ethanol and 7.5 % acetic acid). Crude extracts were centrifugation at 18,000g for 30 minutes to remove any particulate matter and applied to a HisTrap<sup>™</sup> FF column (GE Healthcare Life Sciences) according to the manufacturer's instruction using the Akta Start Protein Purification System (Merck KGaA, Germany). The unbound proteins were washed out with a wash buffer (50 mM KH<sub>2</sub>PO<sub>4</sub>–K<sub>2</sub>HPO<sub>4</sub>, 300 mM NaCl, 40 mM imidazole, pH 8.0). His-tagged fusion proteins were eluted with an elution buffer (50 mM KH<sub>2</sub>PO<sub>4</sub>–K<sub>2</sub>HPO<sub>4</sub>, 300 mM NaCl, 300 mM imidazole, pH 8.0). Fractionation was set to 1.5 ml volumes. This was further dialyzed against 50 mM KH<sub>2</sub>PO<sub>4</sub>–K<sub>2</sub>HPO<sub>4</sub> (pH 8.0) containing 20 % glycerol. Protein concentrations were determined by the method of Bradford (1976) using bovine serum albumin (BSA, Sigma Aldrich) as a standard.

### 7.2.9 Western blot analysis

The purified proteins (DHY-SC-VUT5 and DHY-G-VUT7) were visualised using Sodium dodecyl sulfate– polyacrylamide gel electrophoresis (SDS–PAGE) (15 % resolving gel and 5% stacking gel) as described by the Laemmli procedure (Laemmli 1970). The iBind™ Western automated system (iBlot 2™ and iBind™) was used to carry out Western blotting according to the manufacturer's instructions included in the iBlot 2 and iBind Western Starter Kit. Following the blotting, the membranes were blocked with a 1X iBind™ solution. Mouse-anti-His mAb (ThermoFisher Scientific, USA) was used as the primary antibodies for the antibody binding step. Proteins were detected using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Scientific) that is an enhanced chemiluminescent (ECL) horseradish peroxidase (HRP) substrate according to manufacturer's instructions.

### 7.2.10 Aldehyde dehydrogenase (ALDH) activity assay

The activity of the purified proteins (DHY-SC-VUT5 and DHY-G-VUT7) were carried out according to the protocol by Li *et al.* (2010) for assessing ALDH activity. The standard reaction mixture for the oxidation reaction (200  $\mu$ l) contained hexanal (Merck, USA) as the substrate, 50 mM Nicotinamide adenine dinucleotide (NAD) in 50 mM Tris-HCl (pH 8.0) and an appropriate amount of enzyme. Varying concentrations of hexanal substrate (25 mM, 50 mM, 100 mM, 150 mM and 200 mM) and enzyme concentrations (0.001 mg/ml, 0.003 mg/ml, 0.007 mg/ml and 0.014 mg/ml) were tested respectively in order to determine the optimal concentrations for each. ALDH activity was assayed spectrophotometrically at 340 nm in a 96 well plate by determining the Nicotinamide adenine dinucleotide phosphate (NADH) produced or oxidised, using the BioTek® Epoch 2 microplate reader (Agilent Technologies, USA). The enzymes were added to the reaction mixture to start the reaction. This was mixed thoroughly and immediately placed in the microplate reader. The absorbance values were recorded at twenty second intervals for 3 minutes at 30 °C. The enzymatic activity of the purified proteins was measured by the increase in  $A_{340}$  due to the formation of NADH.

### 7.2.11 Kinetic studies of novel DHY-SC-VUT5 and DHY-G-VUT7 dehydrogenases

After identifying the optimum enzyme concentration as previously described in section 7.1.10, a range of hexanal concentrations (25 mM, 50 mM, 100 mM, 150 mM and 200 mM) were prepared to study the effect of hexanal concentration on the activity of the enzyme, to determine the substrate affinity of the recombinant alcohol dehydrogenase and carry out kinetic studies to obtain the Michaelis-Menten constant ( $K_m$ ) and  $V_{max}$  using GraphPad Prism for Michaelis-Menten kinetics. The equation of Beer's Law shown below was used to determine the initial velocity for each substrate concentration at a specific time interval. The wavelength specific absorption coefficient for NADH ( $\epsilon$ ) is  $6220 \text{ mol.L}^{-1}\text{cm}^{-1}$  and a path length of 0.58 cm was used for a 96 well plate with 200  $\mu\text{l}$  volume.

$$A = \epsilon l c$$

**Where:**

$\epsilon$  = Substance and wavelength specific absorption coefficient,

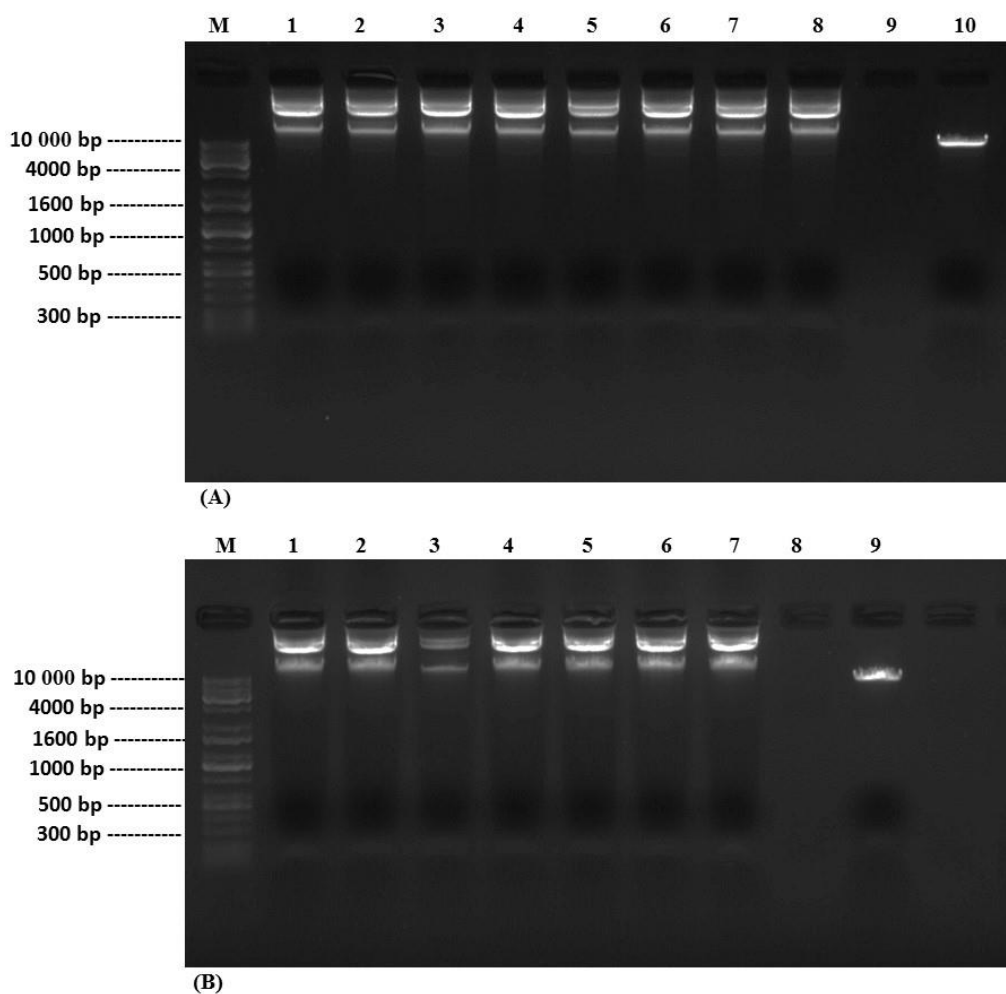
$l$  = Length of the light that travels through the sample

$c$  = Sample concentration.

## 7.3 Results and Discussion

### 7.3.1 The quality of isolated fosmid DNA

Fosmid DNA was isolated from 15 fosmid clones selected on the basis of a screening procedure described in detail in Chapter 3. This was based on their ability to utilize one of the three hydrocarbon substrates i.e. hexadecane, cyclohexane and octadecene. The isolated fosmid DNA was separated using agarose gel electrophoresis and was found to be of high quality and quantity (**Figure 7.2**). This was important for downstream PACBIO SMRT® sequencing that requires high quality DNA to ensure successful long sequence reads.



**Figure 7.2:** 0.8 % (w/v) agarose gel electrophoresis of isolated fosmid DNA (3  $\mu$ l) from 15 candidate clones obtained from functional screening analysis. (A) M: KAPA™ Biosystems Universal ladder (10 kb), Lane 1: pFos\_A1, Lane 2: pFos\_A2, Lane 3: pFos\_A4, Lane 4: pFos\_A7, Lane 5: pFos\_B2, Lane 6: pFos\_B3, Lane 7: pFos\_B7, Lane 8: pFos\_C7 and Lane 10: Fosmid control DNA (40 kb). (B) M: KAPA™ Biosystems Universal ladder (10 kb), Lane 1: pFos\_D3, Lane 2: pFos\_D7, Lane 3: pFos\_D8, Lane 4: pFos\_F1, Lane 5: pFos\_F2, Lane 6: pFos\_F3, Lane 7: pFos\_G6 and Lane 9: Fosmid control DNA (40 kb).

### 7.3.2 Sequence analysis and ALDH gene synthesis

*De novo* assembly was carried out on the PACBIO sequences for the 15 fosmid clones using the CLC Genomics Workbench 11.0.1 (CLC Bio, Qiagen). Data for assembly shown in Appendix B1. The longest assembled contigs for each clone obtained from this software program was used to obtain open reading frames and consequently find regions of local similarity between known sequences to identify aldehyde dehydrogenase genes involved in the third step of aerobic hydrocarbon degradation Uniprot-Swiss prot (<https://www.uniprot.org/>) database. An ORF of 1233 bp in length encoding a polypeptide of 406 amino acids with 50.6 kDa predicted molecular mass was found to share a sequence similarity of 96.7 % to short-chain dehydrogenase from *E. coli*. The percentage similarity shows that this sequence is novel. It has the potential to catalyse the oxidation of aldehydes generated by ADH to carboxylic acids using NAD<sup>+</sup> or NADP<sup>+</sup> as the coenzymes in the last step of aerobic hydrocarbon degradation. The complete *dhy-sc-vut5* nucleotide sequence was submitted to GenBank and assigned an accession number MT606180.

The second ORF of 1470 bp encoding a polypeptide of 637 amino acids with a predicted molecular mass of 67.6 kDa was found to share a sequence similarity of 23.9 % to glycine dehydrogenase (EC 1.4.4.2) from *Caulobacter vibrioides* (strain NA1000 / CB15N) (*Caulobacter crescentus*). The complete *dhy-g-vut7* nucleotide sequence was submitted to GenBank and assigned an accession number MT606181. The low sequence similarity is attributed to its novelty and its ability to catalyse the oxidation of aldehydes to less toxic carboxylic acids. Moreover, the ORF

lengths identified are similar to those seen in the literature for other isolated aldehyde dehydrogenases (Ishige *et al.* 2000, Chen *et al.* 2014). Similarly, the low percentage similarity can be supported further by the investigations conducted on the hydrocarbon contaminated sampling locations discussed previously (Chapter 3, section 3.1.1). The selection of sampling sites with specific stress factors is key for all metagenomics study. This, therefore, holds the potential for isolating highly adapted microorganisms which have been poorly characterised or are totally unknown till this point and home in on their ability to degrade hydrocarbons and utilise it as a sole carbon source. Thus, identifying novel dehydrogenase genes involved in catalysing the NAD-dependent reversible oxidation of aldehydes to carboxylic acids.

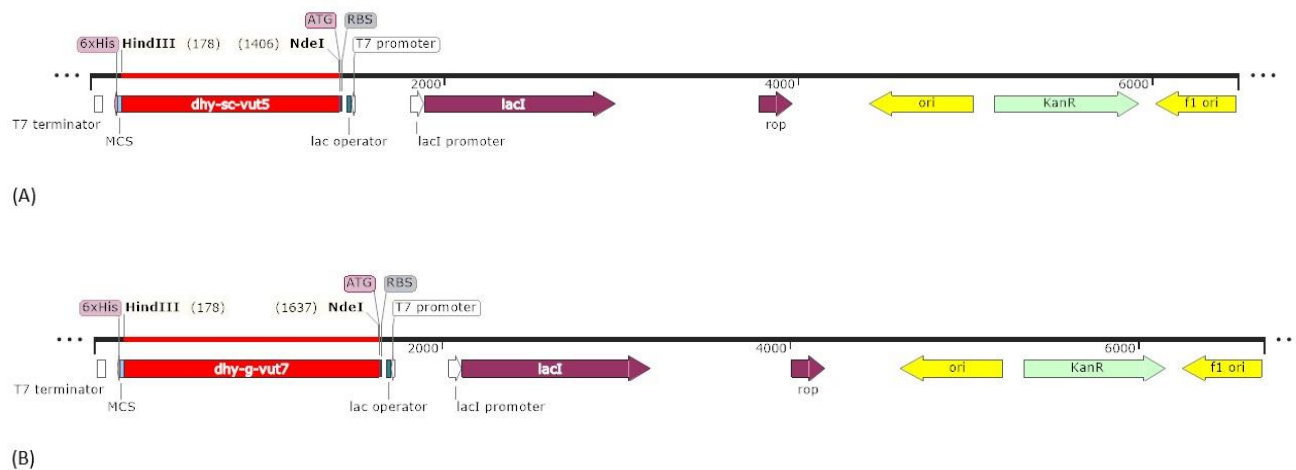
The dehydrogenase gene sequence showed similarity to that isolated from *Caulobacter vibrioides* (strain NA1000 / CB15N) (*Caulobacter crescentus*) according to the Uniprot database search. *Caulobacter vibrioides* is from the taxonomic lineage *Proteobacteria* which was shown to display an 84 % phylum abundance in the soil samples according to microbiome profiling studies carried out in Chapter 4, section 4.2.1. Work carried out by (Yang *et al.* 2014) which involved identifying potential degraders within bacterial profiles, showed that oil samples were dominated by the bacterial taxa of *Caulobacteraceae* (esp. *Caulobacter*). Their study showed that they were either tolerant to or capable of degrading crude oil compounds especially, PAHs. The findings have been supported by previous studies (Lu *et al.* 2019, Pedron *et al.* 2019).

It is important to note that the novel dehydrogenase genes identified here have the potential to convert the harmful compound hexanal into a less toxic compound, hexanoic acid. Moreover, the fosmid clones were selected based on functional screening on hydrocarbon substrates thus, the increased probability of the dehydrogenase involvement in the hydrocarbon degradation pathway specifically alkyl aldehydes. There are a number of studies that have implemented the metagenomics approach similar to the one in this study (Uchiyama and Miyazaki 2009, Niranjani *et al.* 2016, Garrido-Sanz *et al.* 2019, Pedron *et al.* 2019). Some industrially relevant enzymes discovered by metagenomics include cellulases (Gutiérrez-Lucas 2014), amylases (DeCastro *et al.* 2016), chitinases (Berini *et al.* 2019), lipases (Hårdeman and Sjöling 2007), carboxylesterase

(Rashamuse *et al.* 2009), alcohol dehydrogenase (Itoh *et al.* 2014), aldehyde dehydrogenase (Chen *et al.* 2014) and proteases (Iqbal *et al.* 2014).

### 7.3.3 Cloning, expression and protein purification

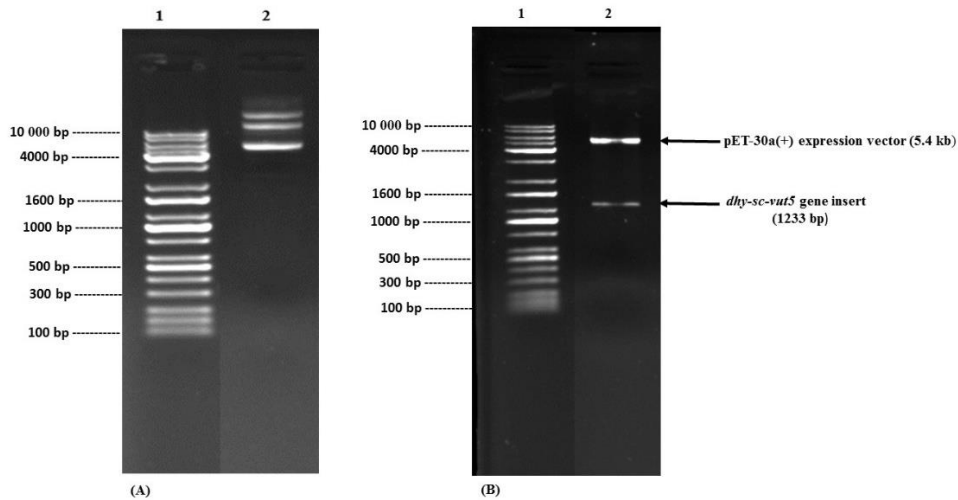
The expression of high levels of recombinant proteins is essential for this study hence the two modified gene sequences were cloned into a pET30a (+) expression vector. Complete modified sequences for DNA and protein shown in Appendix B4 and B5. The pET vector system is a powerful system developed for the cloning and expression of recombinant proteins in *E. coli*. The cloning strategy involved linearizing the pET30a(+) vector by restriction digestion at the *HindIII* and *XbaI* restriction sites. The genes of interest (*dhy-sc-vut5* and *dhy-g-vut7*) were then inserted and ligated. The pET-30a(+) expression vector ligated with the synthesized genes (pET30a(+)\_*dhy-sc-vut5* and pET30a(+)\_*dhy-g-vut7*) were transformed into *E. coli* BL21 (DE3) pLysS competent cells for expression studies (**Figure 7.3(A)** and **Figure 7.3(B)**).



**Figure 7.3:** Linearised maps for cloned novel aldehyde dehydrogenase genes in pET-30a(+).

(A) Linearise map showing features for pET30a(+)\_*dhy-sc-vut5*. (B) Linearised map showing features for pET30a(+)\_*dhy-g-vut7*.

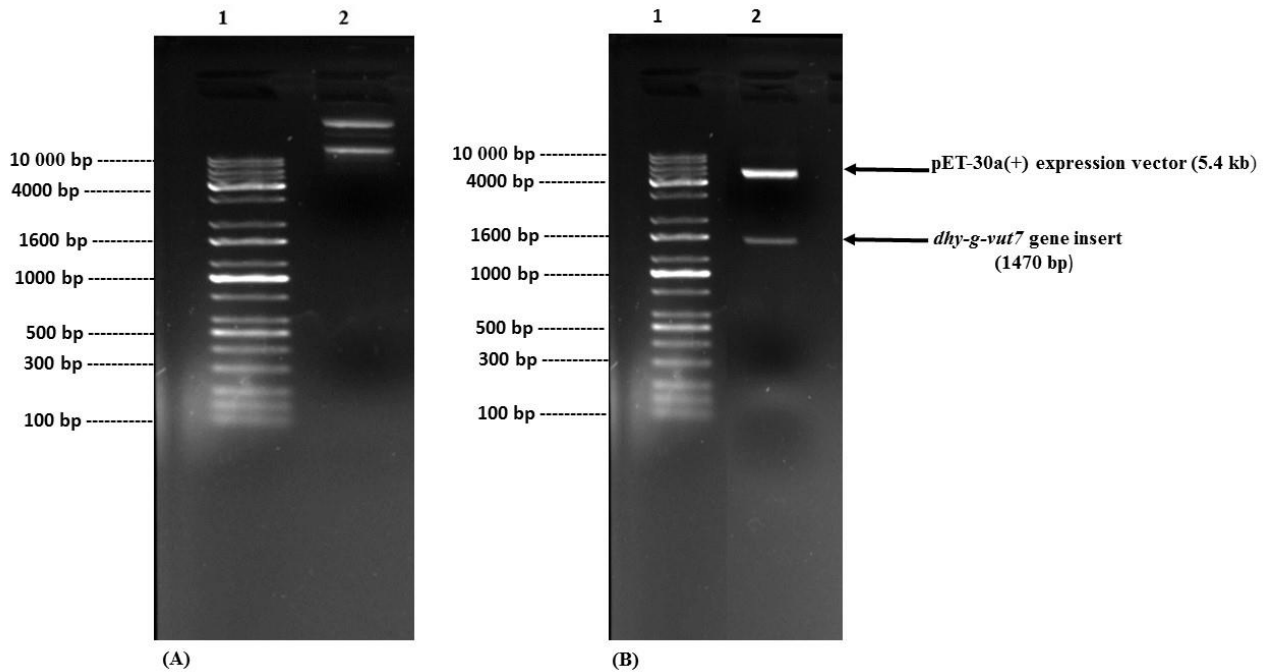
The plasmid DNA was extracted and separated on an agarose/EtBr gel (**Figure 7.4(A)** and **Figure 7.5(A)**) followed by restriction digest with *XbaI* and *HindIII* (**Figure 7.4(B)** and **Figure 7.5(B)**) to confirm the ligation reaction. The extracted plasmid DNA were found to be of high quality and the restriction digests showed the successful cloning of the insert DNA with the expected band sizes of 1233 bp (pET30a(+)\_dhy-sc-vut5) and 1470 bp (pET30a(+)\_dhy-g-vut7).



**Figure 7.4:** Analysis for the confirmation of cloning *dhy-sc-vut5* gene insert.

(A) Agarose gel electrophoresis of extracted plasmid DNA from pET30a(+)\_dhy-sc-vut5. Lane 1: KAPA™ Biosystems Universal ladder (10 kb). Lane 2: Extracted plasmid DNA from transformed *E. coli* BL21 (DE3) pLysS competent cells. (B) Lane 1: KAPA™ Biosystems Universal ladder (10 kb) and Lane 2: Restricted plasmid DNA with *XbaI* and *HindIII* showing the successful cloning of the *dhy-sc-vut5* gene insert (1233 bp).



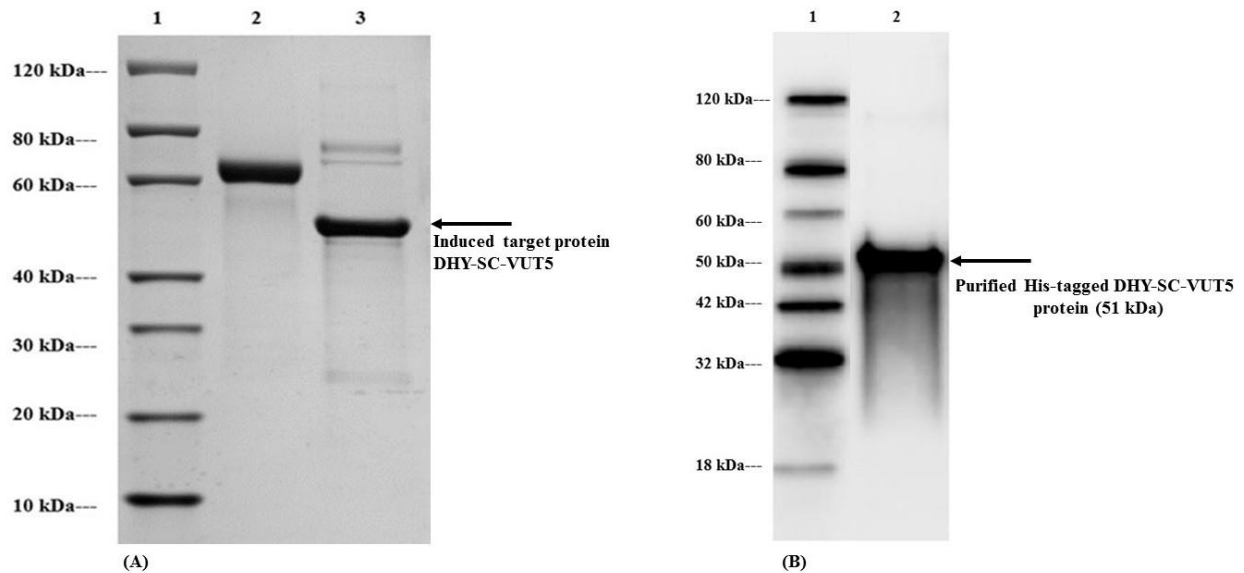


**Figure 7.5:** Analysis for the confirmation of cloning *dhy-g-vut7* gene insert.

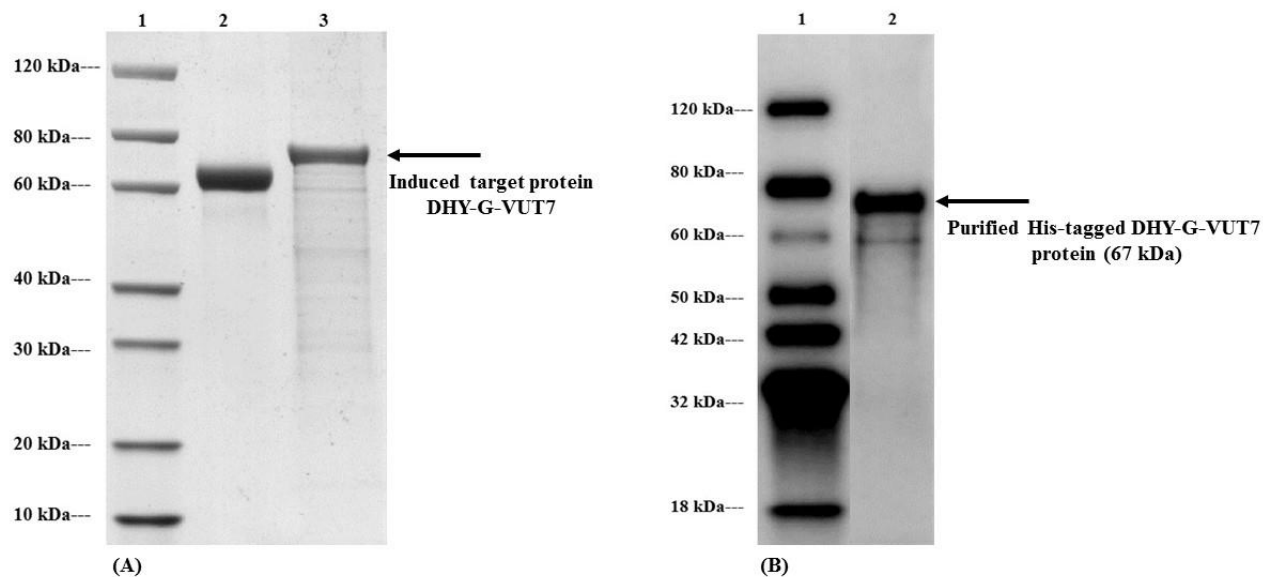
(A) Agarose gel electrophoresis of extracted plasmid DNA from pET30(+)<sub>dhy-g-vut7</sub>. Lane 1: KAPA™ Biosystems Universal ladder (10 kb). Lane 2: Extracted plasmid DNA from transformed *E. coli* BL21 (DE3) pLysS competent cells. (B) Lane 1: KAPA™ Biosystems Universal ladder (10 kb) and Lane 2: Restricted plasmid DNA with *XbaI* and *HindIII* showing the successful cloning of the *dhy-g-vut7* gene insert (1470 bp).

The soluble and insoluble fractions of the cell extracts of the dehydrogenase proteins were analysed on SDS-PAGE to determine the expression level of the proteins and the protein profile following varying conditions of induction. The estimated molecular mass of the individual proteins were calculated from the respective amino acid sequences to compare to the standard protein marker. This was found to be 50.6 kDa and 67.6 kDa for DHY-SC-VUT5 and DHY-G-VUT7 respectively. Varying experimental conditions for induction such as time and temperature were setup to maximize recombinant protein expression levels. The images of SDS PAGE gels for induced short-chain dehydrogenase (DHY-SC-VUT5) and glycine dehydrogenase (DHY-G-VUT7) proteins expressed in BL21 (DE3) upon purification are represented in

**Figure 7.6(A)** and **Figure 7.7(A)** respectively. The results showed high quality and successful protein induction and expression.



**Figure 7.6:** (A) SDS PAGE analysis of novel aldehyde dehydrogenase (DHY-SC-VUT5 ) with the protein band corresponding to a size of 51 kDa after overexpression in BL21 (DE3) *E.coli* cells following induction at 15°C for 16 hours with 0.5 mM IPTG and purification using the Ni column. Lane 1: Protein Marker, GenScript, Lane 2: BSA (2 μg) and Lane 3: Overexpressed purified DHY-SC-VUT5 protein (2 μg). (B) Western blot analysis of DHY-SC-VUT5 protein probed with anti-His antibody. Lane 1: Protein Marker, GenScript, Lane 2: purified His-tagged DHY-SC-VUT5 protein of approximately 51 kDa.



**Figure 7.7:** (A) SDS PAGE analysis of novel aldehyde dehydrogenase (DHY-G-VUT7) with the protein band corresponding to a size of 67 kDa after overexpression in BL21 (DE3) *E.coli* cells following induction at 15°C for 16 hours with 0.5 mM IPTG and purification using the Ni column. Lane 1: Protein Marker, GenScript, Lane 2: BSA (2 µg) and Lane 3: Overexpressed purified DHY-G-VUT7 protein (2 µg). (B) Western blot analysis of DHY-G-VUT7 protein probed with anti-His antibody. Lane 1: Protein Marker, GenScript, Lane 2: purified His-tagged DHY-G-VUT7 protein of approximately 67 kDa.

The overexpressed purified proteins exhibited a band of ~67 kDa and ~51 kDa for the induced DHY-G-VUT7 and DHY-SC-VUT5 proteins respectively from the cell lysate. The bands were confirmation that both proteins were successfully overexpressed, purified and corresponded to the calculated molecular mass. The optimal induction conditions were achieved at 15 °C upon incubation period of 16 hours for both proteins using 0.5 mM IPTG. The current findings were consistent with those found in other literature (Choi and Geletu 2018, Fazaeli *et al.* 2018). The protein concentrations were achieved using the Bradford protein assay which revealed 0.62 mg/ml for DHY-SC-VUT5 and 0.53 mg/ml for DHY-G-VUT7 respectively.

Western blot analysis was carried out to confirm the purity, molecular weight and expression of the target proteins observed on the SDS PAGE gel. The images of Western blot analysis (

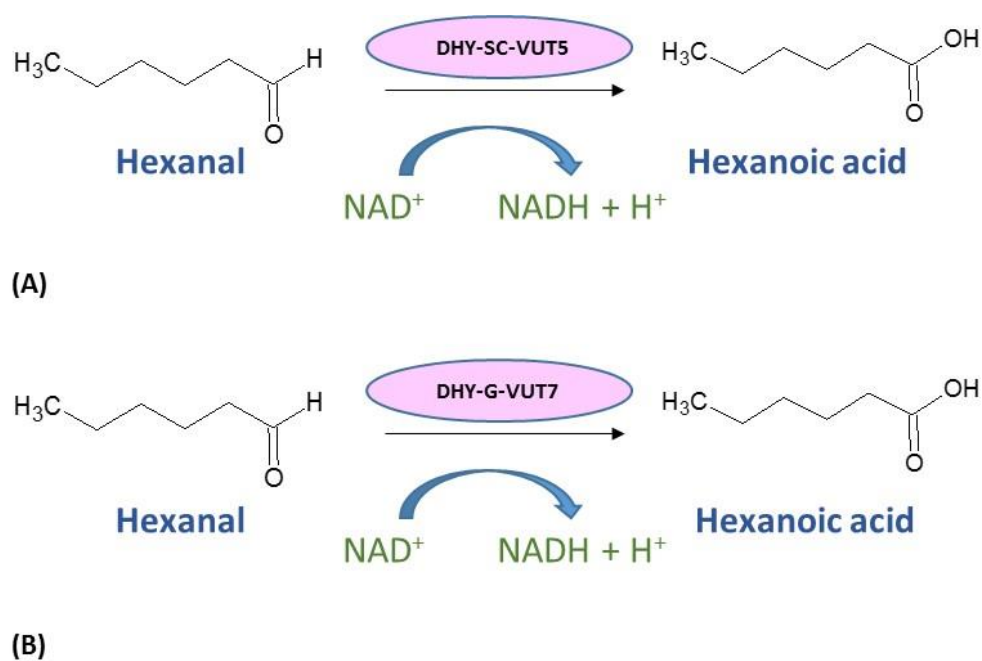
**Figure 7.6(B)** and **Figure 7.7(B)**) showed the binding of the anti-His antibody to both short-chain dehydrogenase band (DHY-SC-VUT5) and glycine dehydrogenase band (DHY-G-VUT7) containing 6x His-tags with an expected band size of 51 kDa and 67 kDa, respectively. Thus, confirming the successful expression and purification of both the target proteins. This confirmation method is has been used successfully by many researchers for the purification of the heme-containing cytochrome P450 oxygenases (Aslantas and Surmeli 2019), cell surface esterase (Bach *et al.* 2003), biosurfactant protein 1 (MBSP1) (Araújo *et al.* 2020) and bi-functional 3-hydroxybutanal dehydrogenase (Frey *et al.* 2016).

#### **7.3.4 Optimal aldehyde dehydrogenase (DHY-SC-VUT5 and DHY-G-VUT7) concentration**

Varying concentrations of the purified DHY-SC-VUT5 and DHY-G-VUT7 enzymes (0.001 mg/ml, 0.003 mg/ml, 0.007 mg/ml and 0.014 mg/ml) were tested individually using the ADH assay as stipulated by Li *et al.* (2010). This was carried out to determine the optimal concentration of enzyme to be used for the kinetics study and cocktail of enzymes designed in Chapter 8, section 8.1.2. The effect of varying enzyme concentrations in the assay was achieved by change in absorbance value at a wavelength of 340 nm indicative of the amount of NADH formed by the reduction of the coenzyme NAD<sup>+</sup> for individual enzyme concentrations of DHY-SC-VUT5 and DHY-G-VUT7. Absorbance data and graphs for the assays shown in Appendix C.

The gradual amount of NAD<sup>+</sup> reduced to NADH was recorded over the time interval, indicated by an increase in absorbance values. The optimal enzyme concentration for DHY-SC-VUT5 and DHY-G-VUT7 were observed to be 0.003 mg/ml. DHY-SC-VUT5 displayed higher NADH production in comparison to DHY-G-VUT7 with an absorbance value of 0.207 a.u and 0.131 a.u respectively. Thus, the concentration was used for subsequent enzymatic kinetic studies. Maximum NADH production took place within the first 120 sec suggesting that the reaction rate

was very fast. The findings were in agreement with a previous study (Li *et al.* 2010) which showed the reduction of  $\text{NAD}^+$  to  $\text{NADH}$  exhibits strong UV absorption at 340 nm and increase in absorbance value is indicative of the amount of  $\text{NADH}$  produced when catalysed by the enzyme. From the current study, it can be concluded that the class of dehydrogenase enzyme identified for DHY-SC-VUT5 and DHY-G-VUT7 proteins are indeed novel aldehyde dehydrogenases with the ability to convert hexanal to hexanoic acid shown by the reduction of coenzyme  $\text{NAD}^+$  to  $\text{NADH}$ . This reaction has been schematically represented in **Figure 7.8** for both proteins.



**Figure 7.8** Schematic representation of the enzymatic reaction of the two identified novel aldehyde dehydrogenase DHY-SC-VUT5 (A) and aldehyde dehydrogenase DHY-G-VUT7 (B).

Furthermore, the general trend observed from the data agrees with the literature with regards to the functioning of enzymes (Bashir *et al.* 2014). Enzyme molecules function by binding to its substrate, catalysing the reaction and lastly releasing the product. The following steps each require a specific amount of time and the more enzyme molecules that are available, the more product can

be produced. Therefore, the more enzyme there is available, the more quickly the substrate can be converted into product. Once all the enzyme has been used up, the reaction will be complete. It can be clearly seen that the enzyme catalyses the reaction rapidly when it can be assumed that the enzyme has been exhausted and all possible product has been made.

The results showed that an increase in enzyme concentration has a direct effect on the amount of  $\text{NAD}^+$  converted to NADH as indicated by an increase in the relative activity. The results were analyzed using two-way analysis of variance (ANOVA) in a Graph Pad Prism ver. 8.4.1 (460) (San Diego, California USA). The concentration of DHY-SC-VUT5 and DHY-G-VUT 7 novel enzymes accounted for 42.88 % ( $p < 0.0001$ ) and 22.3 % ( $p = 0.0003$ ) of the total variance respectively. Thus, the effect of enzyme concentration on NADH production can be considered very significant.

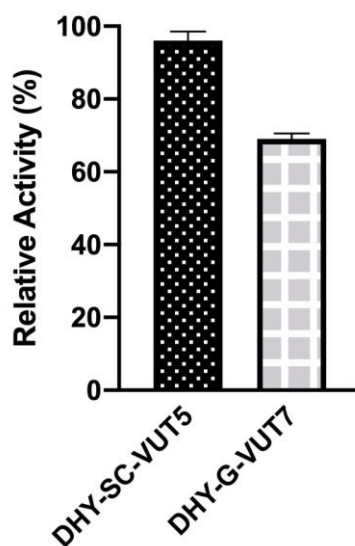
### **7.3.5 Optimal hexanal substrate concentration and enzyme kinetics**

Varying concentrations of the substrate hexanal were tested in the ALDH assay as stipulated by Li *et al.* (2010). This was carried out to determine the optimal concentration of substrate to be used for the kinetics study and the cocktail of enzymes designed in Chapter 8. The effect of varying substrate concentrations in this assay is depicted by change in absorbance value at a wavelength of 340 nm. This is indicative of the amount of NADH formed by the reduction of the coenzyme  $\text{NAD}^+$  hence the direct conversion of hexanal to hexanoic acid. Absorbance data and graphs for the assays shown in Appendix C. The highest relative activity was observed at 150 mM and 50 mM for DHY-SC-VUT5 and DHY-G-VUT7 respectively. The recorded optimal concentration for each enzyme was used for the kinetic studies.

Two-way ANOVA analysis using Graph Pad Prism confirmed that the concentration of hexanal affects the relative activity and accounts for 12.29 % ( $p = 0.0010$ ) and 22.46 % ( $p = 0.0001$ ) total variance for DHY-SC-VUT5 and DHY-G-VUT7 respectively. Thus, the effect of substrate is

considered extremely significant to the change in relative activity observed for both enzymes. This is in agreement with similar results observed for an aldehyde-alcohol dehydrogenase (AdhE), a key enzyme in bacterial fermentation, converting acetyl-CoA to ethanol was observed when kinetics studies were carried out (Kim *et al.* 2019).

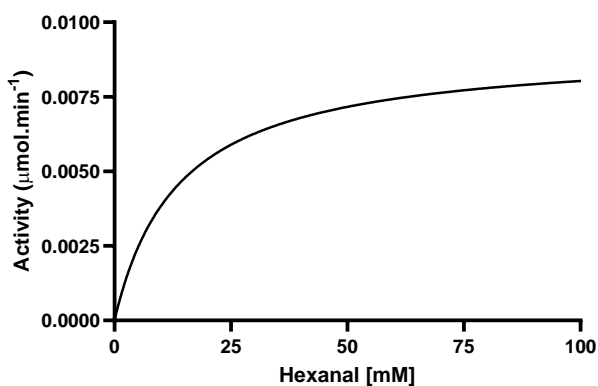
The relative activity of both enzymes were subjected to comparison at their optimal concentrations of 150 mM (DHY-SC-VUT5) and 50 mM (DHY-G-VUT7) respectively at a time interval of 3 minutes using hexanal as a substrate (**Figure 7.9**). The results showed marked difference between the relative activity of DHY-SC-VUT5 and DHY-G-VUT7 with the former exhibiting the highest activity for the same substrate. The difference can be attributed to the relative activity of the individual enzymes which accounted for 28.85 % of the total variance and the effect which was considered extremely significant ( $p < 0.0001$ ) as revealed by 2-way ANOVA.



**Figure 7.9:** Comparison of the relative activities of two novel aldehyde dehydrogenase DHY-SC-VUT5 and DHY-G-VUT7 using hexanal as a substrate over a time interval of 3 minutes at 30 °C. The data was analysed using GraphPad Prism Version 8.4.1 (460).

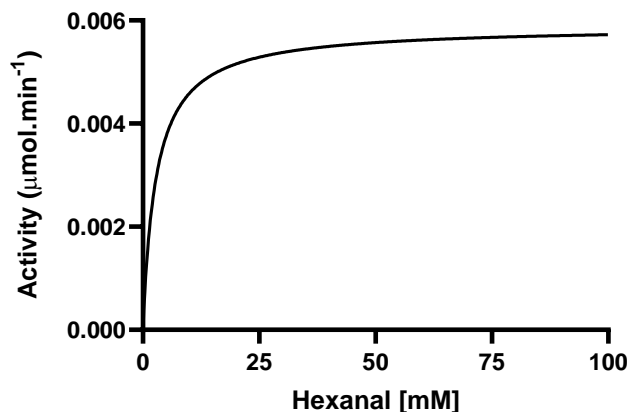
The absorbance values for the varying hexanal substrate concentrations were used to calculate the velocity and to construct the Michaelis-Menten graphs using Graph Pad Prism Version 8.4.1 (460)

for DHY-SC-VUT5 (**Figure 7.10**) and DHY-G-VUT7 protein (**Figure 7.11**). The graphs describe the rate of the enzymatic reaction by relating reaction rate (rate of formation of product) to the concentration of a substrate. It can be deduced that the enzymatic reactions of both graphs followed a typical first order for Michaelis-Menten function when hexanal was used as a substrate in the presence of  $\text{NAD}^+$ . Thus, the rate of the enzymatic reactions are directly proportional to the concentration of the reacting substance. The kinetics data clearly demonstrates that both these discovered proteins, DHY-SC-VUT5 and DHY-G-VUT7, function as aldehyde dehydrogenases.



**Figure 7.10:** Enzymatic activity of novel aldehyde dehydrogenase DHY-SC-VUT5 at different concentration of hexanal at 30 °C. Absorbance values were recorded spectrophotometrically at 340 nm wavelength. The kinetic data was fitted into Michaelis-Menten model using GraphPad Prism Version 8.4.1 (460).





**Figure 7.11:** Enzymatic activity of novel aldehyde dehydrogenase DHY-G-VUT7 at different concentration of hexanal at 30 °C. Absorbance values were recorded spectrophotometrically at 340 nm wavelength. The kinetic data was fitted into Michaelis-Menten model using GraphPad Prism Version 8.4.1 (460).

The  $K_m$  value was found to be 13.77 mM with a  $V_{max}$  of 0.009135  $\mu\text{mol}\cdot\text{min}^{-1}$  and 2.832 mM with a  $V_{max}$  of 0.005886  $\mu\text{mol}\cdot\text{min}^{-1}$  for DHY-SC-VUT5 and DHY-G-VUT7, respectively. The relationship between rate of reaction and concentration of substrate depends on the affinity of the enzyme for its substrate. Since a small  $K_m$  value indicates that the enzyme has a high affinity for the substrate, it can be concluded that both enzymes displayed a high affinity for hexanal as a substrate. By using 13.77 mM (DHY-SC-VUT5) and 2.832 mM (DHY-G-VUT7) hexanal, the proteins reached half of  $V_{max}$  and thus, acts at a more or less constant rate, regardless of variations in the concentration of substrate within the physiological range. However, DHY-G-VUT7 importantly shows a higher affinity for hexanal than DHY-SC-VUT5 due to its much lower  $K_m$  value thus making DHY-G-VUT7 a more potent and efficient enzyme for alkyl aldehyde degradation. The  $K_m$  values obtained for both aldehyde dehydrogenases are similar to a membrane bound aldehyde dehydrogenase complex AldFGH involved in acetic acid fermentation isolated from *Acetobacter pasteurianus* SKU1108 (Yakushi *et al.* 2018).

Hexanal is an intermediate in the transformation of toxic hexanol to the non-toxic carboxylic acid (Ławniczak *et al.* 2020). This intermediate reaction would not be possible without aldehyde dehydrogenases (Zhu *et al.* 2017). Accumulation of hexanal in the soil as a result of the slow transformation process by indigenous microorganisms is a contributing factor in soil pollution (Lopachin and Gavin 2014). Due to this slow process and the toxic nature of this compound, there is a need to develop an enzyme-based solution to speed up the degradation process by identifying novel aldehyde dehydrogenases as successfully done in this study. Furthermore, the resulting carboxylic acids are beneficial for promoting soil fertility and plant growth (Nigam and Srivastava 2005) besides its recorded industrial application.

There have been aldehyde dehydrogenases identified and purified for application in hydrocarbon break down. Specifically a long-chain aldehyde dehydrogenase (Ald1) which participates in *n*-alkane utilization was isolated from the soluble fraction of *Acinetobacter* *Acinetobacter* sp. Strain M-1 cells grown on *n*-hexadecane as a sole carbon source (Ishige *et al.* 2000). The open reading frame of ald1 was 1,512 bp long, corresponding to a protein of 503 amino acid residues (molecular mass, 55,496 Da). The deduced amino acid sequence showed high similarity to those of various aldehyde dehydrogenases. This gene was stably expressed in *Escherichia coli* showed high enzyme activity toward *n*-alkanals (C4 to C14), with a preference for longer carbon. The highest activity was obtained with tetradecanal. Also, Frey *et al.* (2016), cloned and functionally expressed a bi-functional 3-hydroxybutanal dehydrogenase/reductase involved in acetone metabolism by *Desulfococcus biacutus*. The highest specific activities were observed for reduction of C3 - C5-aldehydes with NAD<sup>+</sup>.

Another study by Kamimura *et al.* (2017) identified for the first time a bacterial aromatic aldehyde dehydrogenase (ALDH) obtained from 20 putative ALDH genes of *Sphingobium* sp. strain SYK-6 critical for the efficient catabolism of syringaldehyde. Syringaldehyde is obtained from lignin and are essential intermediates for the production of basic chemicals using microbial cell factories. Furthermore, a study by Panday *et al.* (2016) showed the elevated expression of ALDH in earthworms in heavy metal polluted soils. Hence, this was used as a bioindicator to monitor heavy

metal pollution in soil. The ALDH activity was found to be highest in worms cultured in 5 ppm heavy metal contaminated soils. Their study showed an accumulation in earthworms in these soils since they ensure that the toxic aldehydes are neutralized to carboxylic acids as a stress responsive method. All these studies further highlight the importance of our newly discovered aldehyde dehydrogenases for application in detoxifying various pollutants.

## 7.4 Conclusion

In conclusion, we have discovered two novel dehydrogenases (DHY-SC-VUT5 and DHY-G-VUT7) from the metagenome of hydrocarbon contaminated soils. The hydrogenases showed better catalytic activity. The DHY-G-VUT7 dehydrogenase showed higher substrate affinity than DHY-G-VUT5. The high affinity of both enzymes for hexanal as a substrate is valuable for their application in the breakdown of alkyl aldehyde compounds that persist in the soil and water environments. Hence, these enzymes can be used as part of an enzyme cocktail that works in synergy to completely degrade alkanes. Moreover, since the discovered enzymes are novel it is important to conduct crystallographic studies of such invaluable enzymes to solve their structures and explore further down stream applications.

## 7.5 References

- Araújo, S.C. da S., Silva-Portela, R.C.B., de Lima, D.C., da Fonsêca, M.M.B., Araújo, W.J., da Silva, U.B., Napp, A.P., Pereira, E., Vainstein, M.H., and Agnez-Lima, L.F., 2020. MBSP1: a biosurfactant protein derived from a metagenomic library with activity in oil degradation. *Scientific Reports*, 10 (1), 1–13.
- Aslantas, Y. and Surmeli, N.B., 2019. Effects of N-terminal and C-terminal polyhistidine tag on the stability and function of the thermophilic P450 CYP119. *Bioinorganic Chemistry and Applications*, 2019.
- Bach, H., Berdichevsky, Y., and Gutnick, D., 2003. An Exocellular Protein from the Oil-Degrading Microbe. *Society*, 69 (5), 2608–2615.
- Bashir, Y., Pradeep Singh, S., Kumar Konwar, B., Bashir, Y., Pradeep Singh, S., and Kumar Konwar, B., 2014. Metagenomics: An Application Based Perspective. *Chinese Journal of Biology*, 2014, 1–7.
- Berini, F., Casartelli, M., Montali, A., Reguzzoni, M., Tettamanti, G., and Marinelli, F., 2019. Metagenome-sourced microbial chitinases as potential insecticide proteins. *Frontiers in Microbiology*, 10 (JUN), 1–12.
- Berini, F., Casciello, C., Marcone, G.L., and Marinelli, F., 2017. Metagenomics: Novel enzymes from non-culturable microbes. *FEMS Microbiology Letters*, 364 (21), 1–19.
- Bouhajja, E., McGuire, M., Liles, M.R., Bataille, G., Agathos, S.N., and George, I.F., 2017. Identification of novel toluene monooxygenase genes in a hydrocarbon-polluted sediment using sequence- and function-based screening of metagenomic libraries. *Applied Microbiology and Biotechnology*, 101 (2), 797–808.
- Bradford, M.M., 1976. A Rapid and Sensitive Method for the Quantitation Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Analytical Biochemistry*, 72, 248–254.
- Chen, R., Li, C., Pei, X., Wang, Q., Yin, X., and Xie, T., 2014. Isolation an Aldehyde

- Dehydrogenase Gene from Metagenomics Based on Semi-nest Touch-Down PCR. *Indian Journal of Microbiology*, 54 (1), 74–79.
- Choi, T.J. and Geletu, T.T., 2018. High level expression and purification of recombinant flounder growth hormone in *E. coli*. *Journal of Genetic Engineering and Biotechnology*, 16 (2), 347–355.
- DeCastro, M.E., Rodríguez-Belmonte, E., and González-Siso, M.I., 2016. Metagenomics of thermophiles with a focus on discovery of novel thermozyms. *Frontiers in Microbiology*, 7 (SEP), 1–21.
- Fazaeli, A., Golestani, A., Lakzaei, M., Rasi Varaei, S.S., and Aminian, M., 2018. Expression optimization of recombinant cholesterol oxidase in *Escherichia coli* and its purification and characterization. *AMB Express*, 8 (1).
- Frey, J., Rusche, H., Schink, B., and Schleheck, D., 2016. Cloning, functional expression and characterization of a bifunctional 3-hydroxybutanal dehydrogenase /reductase involved in acetone metabolism by *Desulfococcus biacutus*. *BMC Microbiology*, 16 (1), 1–9.
- Garrido-Sanz, D., Redondo-Nieto, M., Guirado, M., Pindado Jiménez, O., Millán, R., Martín, M., and Rivilla, R., 2019. Metagenomic Insights into the Bacterial Functions of a Diesel-Degrading Consortium for the Rhizoremediation of Diesel-Polluted Soil. *Genes*, 10 (6), 456.
- Gregson, B.H., Metodieva, G., Metodiev, M. V., and McKew, B.A., 2019. Differential protein expression during growth on linear versus branched alkanes in the obligate marine hydrocarbon-degrading bacterium *Alcanivorax borkumensis* SK2T. *Environmental Microbiology*, 21 (7), 2347–2359.
- Gurusinghe, S., Brooks, T.L., Barrow, R.A., Zhu, X., and Weston, L.A., 2019. Technologies for the Selection , Culture and Metabolic. *Molecules*, 24, 1–20.
- Gutiérrez-Lucas, L., 2014. Strategies for the Extraction, Purification and Amplification of Metagenomic DNA from Soil Growing Sugarcane. *Advances in Biological ...*, (June), 281–289.

- Hårdeman, F. and Sjöling, S., 2007. Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS Microbiology Ecology*, 59 (2), 524–534.
- Hausjell, J., Halbwirth, H., and Spadiut, O., 2018. Recombinant production of eukaryotic cytochrome P450s in microbial cell factories. *Bioscience Reports*, 38 (2), 1–13.
- Iqbal, H.A., Craig, J.W., and Brady, S.F., 2014. Antibacterial enzymes from the functional screening of metagenomic libraries hosted in *Ralstonia metallidurans*. *FEMS Microbiology Letters*, 354 (1), 19–26.
- Ishige, T., Tani, A., Sakai, Y., and Kato, N., 2000. Long-chain aldehyde dehydrogenase that participates in n-alkane utilization and wax ester synthesis in *Acinetobacter* sp. strain M-1. *Applied and Environmental Microbiology*, 66 (8), 3481–3486.
- Itoh, N., Kariya, S., and Kurokawa, J., 2014. Efficient PCR-based amplification of diverse alcohol dehydrogenase genes from metagenomes for improving biocatalysis: Screening of gene-specific amplicons from metagenomes. *Applied and Environmental Microbiology*, 80 (20), 6280–6289.
- Jung, J., Philippot, L., and Park, W., 2016. Metagenomic and functional analyses of the consequences of reduction of bacterial diversity on soil functions and bioremediation in diesel-contaminated microcosms. *Scientific Reports*, 6 (February), 1–10.
- Kamimura, N., Goto, T., Takahashi, K., Kasai, D., Otsuka, Y., Nakamura, M., Katayama, Y., Fukuda, M., and Masai, E., 2017. A bacterial aromatic aldehyde dehydrogenase critical for the efficient catabolism of syringaldehyde. *Scientific Reports*, 7 (October 2016), 1–12.
- Kim, G., Azmi, L., Jang, S., Jung, T., Hebert, H., Roe, A.J., Byron, O., and Song, J.J., 2019. Aldehyde-alcohol dehydrogenase forms a high-order spiroosome architecture critical for its activity. *Nature Communications*, 10 (1).
- Laemmli, U.K., 1970. 227680a0. *Nature*, 227, 680–685.
- Ławniczak, Ł., Woźniak-Karczewska, M., Loibner, A.P., Heipieper, H.J., and Chrzanowski, Ł., 2020. Microbial degradation of hydrocarbons—basic principles for bioremediation: A

- review. *Molecules*, 25 (4), 1–19.
- Li, P., Wang, L., and Feng, L., 2013. Characterization of a novel rieske-type alkane monooxygenase system in *Pusillimonas* sp. strain T7-7. *Journal of Bacteriology*, 195 (9), 1892–1901.
- Li, X., Li, Y., Wei, D., Li, P., Wang, L., and Feng, L., 2010. Characterization of a broad-range aldehyde dehydrogenase involved in alkane degradation in *Geobacillus thermodenitrificans* NG80-2. *Microbiological Research*, 165 (8), 706–712.
- Liu, X., Dong, Y., Zhang, J., Zhang, A., Wang, L., and Feng, L., 2009. Two novel metal-independent long-chain alkyl alcohol dehydrogenases from *Geobacillus thermodenitrificans* NG80-2. *Microbiology*, 155 (6), 2078–2085.
- Liu, Y.F., Galzerani, D.D., Mbadinga, S.M., Zaramela, L.S., Gu, J.D., Mu, B.Z., and Zengler, K., 2018. Metabolic capability and in situ activity of microorganisms in an oil reservoir. *Microbiome*, 6 (1), 1–12.
- Lopachin, R.M. and Gavin, T., 2014. Molecular mechanisms of aldehyde toxicity: A chemical perspective. *Chemical Research in Toxicology*, 27 (7), 1081–1091.
- Lu, C., Hong, Y., Liu, J., Gao, Y., Ma, Z., Yang, B., Ling, W., and Waigi, M.G., 2019. A PAH-degrading bacterial community enriched with contaminated agricultural soil and its utility for microbial bioremediation. *Environmental Pollution*, 251, 773–782.
- Nigam, R. and Srivastava, M.M., 2005. Effects of carboxylic and amino acids on Cd uptake by *Lycopersicon esculentum*. *Chemical Speciation and Bioavailability*, 17 (1), 19–26.
- Niranjan, V., Wong, M., Reddy, J., Prasanna, A., Deshpande, S., and Kong, H., 2016. A metagenomic analysis of soil samples to find the distribution of microflora in different soil types, (February), 1–7.
- Panday, R., Bhatt, P.S., Bhattarai, T., Shakya, K., and Sreerama, L., 2016. Aldehyde dehydrogenase expression in *Metaphire posthuma* as a bioindicator to monitor heavy metal pollution in soil. *BMC Research Notes*, 9 (1), 1–5.



- Pedron, R., Esposito, A., Bianconi, I., Pasolli, E., Tett, A., Asnicar, F., Cristofolini, M., Segata, N., and Jousson, O., 2019. Genomic and metagenomic insights into the microbial community of a thermal spring. *Microbiome*, 7 (1), 1–13.
- Rashamuse, K., Ronneburg, T., Hennessy, F., Visser, D., Van Heerden, E., Piater, L., Litthauer, D., Möller, C., and Brady, D., 2009. Discovery of a novel carboxylesterase through functional screening of a pre-enriched environmental library. *Journal of Applied Microbiology*, 106 (5), 1532–1539.
- Shinde, P., Musameh, M., Gao, Y., Robinson, A.J., and Kyratzis, I. (Louis), 2018. Immobilization and stabilization of alcohol dehydrogenase on polyvinyl alcohol fibre. *Biotechnology Reports*, 19, e00260.
- Silva, N.M., De Oliveira, A.M.S.A., Pegorin, S., Giusti, C.E., Ferrari, V.B., Barbosa, D., Martins, L.F., Morais, C., Setubal, J.C., Vasconcellos, S.P., Da Silva, A.M., De Oliveira, J.C.F., Pascon, R.C., and Viana-Niero, C., 2019. Characterization of novel hydrocarbon-degrading *Gordonia paraffinivorans* and *Gordonia sihwensis* strains isolated from composting. *PLoS ONE*, 14 (4), 1–16.
- Uchiyama, T. and Miyazaki, K., 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Current Opinion in Biotechnology*.
- Yakushi, T., Fukunari, S., Kodama, T., Matsutani, M., Nina, S., Kataoka, N., Theeragool, G., and Matsushita, K., 2018. Role of a membrane-bound aldehyde dehydrogenase complex AldFGH in acetic acid fermentation with *Acetobacter pasteurianus* SKU1108. *Applied Microbiology and Biotechnology*, 102 (10), 4549–4561.
- Yan, W., Li, F., Wang, L., Zhu, Y., Dong, Z., and Bai, L., 2017. Discovery and characterization of a novel lipase with transesterification activity from hot spring metagenomic library. *Biotechnology Reports*, 14, 27–33.
- Yang, S., Wen, X., Zhao, L., Shi, Y., and Jin, H., 2014. Crude oil treatment leads to shift of bacterial communities in soils from the deep active layer and upper permafrost along the China-Russia Crude Oil Pipeline route. *PLoS ONE*, 9 (5), 12–14.

Zhu, J., Lu, K., Xu, X., Wang, X., and Shi, J., 2017. Purification and characterization of a novel glutamate dehydrogenase from *Geotrichum candidum* with higher alcohol and amino acid activity. *AMB Express*, 7 (1).

---

# Chapter 8

## Engineering of Enzyme Cocktails for Complete Transformation of the Toxic Hexanol into a Useful Product

---

Article submitted for publication

### *Abstract*

Usually more than one enzyme is involved to achieve complete or near complete degradation of hydrocarbons or their transformation into a less toxic or useful compounds. Hexanol, a hydrocarbon which belongs to alkanol can cause serious health effect. Short-term exposure to hexanol causes irritation to the eyes, respiratory tract and skin. If this liquid is swallowed, aspiration into the lungs may result in chemical pneumonitis, which is inflammation of the lungs or breathing difficulty that could be acute or chronic. Thus, here we report engineering of two enzyme cocktails based on combination of novel enzymes, namely an alcohol dehydrogenase AOL-VUT3 with each of DHY-SC-VUT5 and DHY-G-VUT7 dehydrogenases for complete transformation of hexanol. The activity of the cocktails was indirectly assessed by measuring the amount NADH released as a result of reduction of the co-factor  $\text{NAD}^+$  following action of the pair of enzymes on hexanol and converting it to carboxylic acid. As a result, AOL-VUT3/DHY-SC-VUT5 and AOL-VUT3/DHY-G-VUT7 cocktails were engineered. Accordingly, AOL-VUT3/DHY-SC-VUT5 cocktail exhibited the highest activity relative to AOL-VUT3/DHY-G-VUT7. Therefore, the successful enzymatic transformation of the toxic hexanol into carboxylic acid is a major breakthrough since, carboxylic acids and their derivatives are used in the production of polymers, biopolymers, coatings, adhesives, and pharmaceutical drugs. They can also be used as solvents, food additives, antimicrobials, and flavorings. Thus, the developed enzyme cocktails

provide two-fold advantage as they remove the toxic hexanol and produce the very useful carboxylic acid. Hence, future study should focus on immobilization and stabilization of the engineered enzyme cocktails for large scale bioremediation of alkanols and an industrial scale production of carboxylic acid from hexanol.

**Keywords:** Microorganisms, alcohol dehydrogenase, aldehyde dehydrogenase, carboxylic acid, enzyme cocktails, hydrocarbons, bioremediation, hexanol, hexanal.

<b>TABLE OF CONTENTS</b>	<b>PAGES</b>
8.1 Introduction .....	289
8.2 Materials and Methods .....	291
8.2.1 Expressed and purified proteins for enzyme cocktails .....	291
8.2.2 Designed enzyme cocktails.....	292
8.2.3 ADH enzyme activity assay.....	293
8.3 Results and Discussion.....	294
8.3.1 Assessing the activity of the enzyme cocktails.....	294
8.4 Conclusion.....	300
8.5 References .....	301

## 8.1 Introduction

Microorganisms are found ubiquitously and have colonized diverse environments for thousands of years even those considered to be extreme (Jin *et al.* 2019). They play a major role in food webs, biogeochemical cycles, and maintenance and survival of plants, animals, and other organisms through symbiotic relationships (Kamagata and Narihiro 2016). Several microorganisms inhabiting petroleum hydrocarbon contaminated environments such as soils are directly involved in degradation reactions to regenerate different forms of elements needed by other organisms (Jacoby *et al.* 2017). A contributing factor to their metabolic versatility is their genetic resourcefulness for the transformation of contaminants into less-toxic, final end products. Besides, their ubiquitous nature extends their biotechnological potential with many possible applications by either using a microorganism consortium or the enzymes they produce (Peixoto *et al.* 2011). The current study focuses on the use of enzyme cocktails sourced from metagenome of hydrocarbon contaminated soils for the breakdown of alkanols.

Enzymes are referred to as biocatalysts that speed up biochemical reactions in living organisms (Robinson 2015). They are classified as proteins and are made up of the building blocks, amino acids that range in size from 100 to more than 2000 amino acid residues. Enzymes are usually extracted from cells and used to catalyze a wide range of commercially important processes. The use of enzymes for bioremediation strategies is simpler in comparison to working with the whole microorganisms (Peixoto *et al.* 2011). Thus, the use of isolated enzymes does not generate toxic by-products and elimination of whole-cell competitiveness.

In order to accomplish the transformation of hydrocarbon pollutants in both soil and water environments, it is imperative that enzyme cocktails are developed and incorporated into bioremediation strategies. Truskewycz *et al.* (2019), developed and observed that enzymes in cocktails work in synergy to ensure a higher degradation efficiency at lower costs in comparison to the current bacterial strains and consortiums. Besides, enzyme cocktails would prove more efficient since specific enzymes for both recalcitrant and toxic compounds can be applied together

(Peixoto *et al.* 2011). The synergistic effect also considered as a cascade of reactions ensures that the product from the first enzymatic reaction feeds as a substrate to the second reaction, and this to the third and fourth (Aalbers and Fraaije 2017). Accordingly, each step is dependent on the step before in order to ensure the continuation of the enzymatic pathway. This is basically mimicking the activity of the bacterial consortium that are involved in degrading a particular substrate in the soil environment hence achieving ultimately complete degradation (Tiralerdpanich *et al.* 2018).

Limited or no studies have developed and implemented tailor-made enzyme mixtures working in synergy for complete alkane degradation (Ji *et al.* 2019). Recently, Ji *et al.* (2019) managed to develop an enzyme cocktail containing a NADH regeneration system for oily sludge bioremediation. They prepared intracellular enzymes of oil-degrading strain *Acinetobacter calcoaceticus* and applied together with formate dehydrogenase from *Candida boidinii* to determine the degradation potential and found to be 35.6 % after 12 hours. However, the enzyme cocktail developed in their study was a crude enzyme extract used in varying concentrations together with formate dehydrogenase and did not reflect the traditional synergistic cocktail. Besides, the enzymes in a crude mixture remain unknown since individual enzymes have not been purified and their enzyme activity not assessed individually.

With the observed gap in research and lack of appropriate enzyme cocktails designed to target the transformation of hexanols, greater emphasis should be placed on their development. This will contribute immensely to the already existing bioremediation strategies in place. Thus, the current study's main objective was to prepare two enzyme cocktails (AOL-VUT3/DHY-SC-VUT5 and AOL-VUT3/DHY-G-VUT7) both targeted at alkanol degradation using expressed and purified enzymes alcohol dehydrogenase and aldehyde dehydrogenases discussed previously in Chapter 6 and 7. These orchestrated mixtures of enzymes can be implemented in future bioremediation studies involving different classes of hydrocarbons to determine their substrate affinity and relative activities.

## 8.2 Materials and Methods

### 8.2.1 Expressed and purified proteins for enzyme cocktails

A fosmid library constructed using metagenomic DNA isolated from hydrocarbon contaminated soil samples were functionally screened to identify fosmid clones involved in hydrocarbon degradation pathways. Fifteen candidate fosmid clones were sent for PACBIO SMRT® sequencing at Inqaba Biotech™ (Pretoria, South Africa) followed by *de novo* assembly of the sequences to extract the longest contigs. Three gene sequences (one alcohol dehydrogenase and two aldehyde dehydrogenases) were identified using ORF finder on the NCBI database (<https://www.ncbi.nlm.nih.gov/orffinder/>).

The ORF sequences designated *aol-vut3*, *dhy-sc-vut5* and *dhy-g-vut7* were codon optimised for codon compatibility in *E. coli* to increase protein expression and His tags were added for downstream purification. The optimised gene sequences were sent to GenScript® USA Inc for synthesis. They were then cloned into pET-30a(+) expression vectors (Novagen, USA) and transformed into BL21(DE3) *E. coli* host cells. The method for the expression and purification of the three proteins have been discussed in more detail in Chapters 6 and 7 and their related information are shown in **Table 8.1**. Full-length sequences of *aol-vut3*, *dhy-sc-vut5* and *dhy-g-vut7* were deposited in the GenBank with accession numbers MT606178, MT606179, MT606180 and MT606181, respectively.



**Table 8.1:** Some details of expressed and purified proteins used to design the enzyme cocktails.

Designated enzyme name	Class	Size (kDa)	Gene similarity	% <sup>*</sup>	Organism	Substrate
AOL-VUT3	Alcohol dehydrogenase	25	Histidinol dehydrogenase	24.7	<i>Mycobacterium bovis</i> (strain ATCC BAA-935 / AF2122/97)	Hexanol
DHY-SC-VUT5	Aldehyde dehydrogenase	50.6	Short-chain dehydrogenase	96.7	<i>E. coli</i>	Hexanal
DHY-G-VUT7	Aldehyde dehydrogenase	67.6	Glycine dehydrogenase (decarboxylating)	23.9	<i>Caulobacter vibrioides</i> (strain NA1000 / CB15N) ( <i>Caulobacter crescentus</i> )	Hexanal

\* A protein Basic Local Alignment Search Tool (BLAST) called Uniprot-Swiss prot (<https://www.uniprot.org/>) was carried out on individual ORF sequences to find regions of local similarity between known sequences.

### 8.2.2 Designed enzyme cocktails

The expressed and purified proteins AOL-VUT3, DHY-SC-VUT5 and DHY-G-VUT7 were used to design two cocktail combinations (AOL-VUT3/DHY-SC-VUT5 and AOL-VUT3/DHY-G-VUT7) to determine their level of degradation activity on hexanol. The AOL-VUT3 protein should work on the hexanol to produce hexanal and thereby reduce the cofactor NAD<sup>+</sup> to NADH and DHY-SC-VUT5 and DHY-G-VUT7 proteins catalyse the reaction of the hexanal to a carboxylic acid and further reduce the NAD<sup>+</sup> to NADH.

### 8.2.3 Alcohol dehydrogenase (ADH) enzyme activity assay

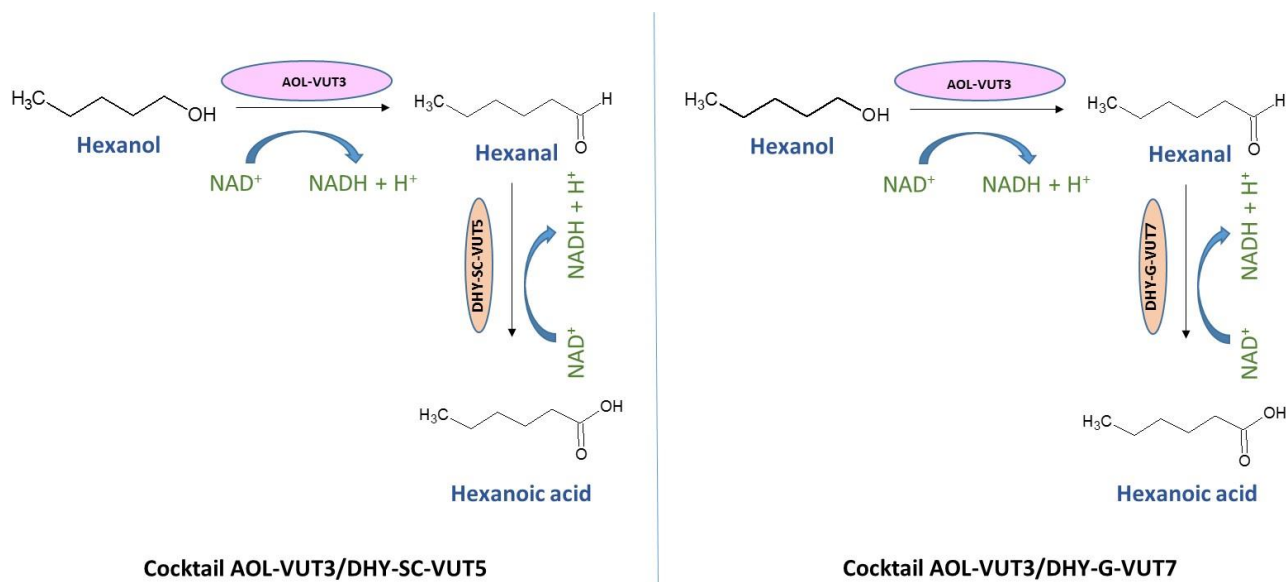
The activity of the enzyme cocktails (AOI-VUT3/DHY-SC-VUT5 and AOL-VUT3/DHY-G-VUT7) were carried out according to the protocol by Li *et al.* (2010) for assessing alcohol dehydrogenase (ADH) activity. The standard reaction mixture for the oxidation reaction (200  $\mu$ l) contained 150 mM hexanol (Merck, USA) as the substrate, 50 mM Nicotinamide adenine dinucleotide (NAD) in 50 mM Tris-HCl (pH 8.0) and 0.003 mg/ml enzymes. Varying concentrations of hexanal substrate (25 mM, 50 mM, 100 mM, 150 mM and 200 mM) and enzyme concentrations (0.001 mg/ml, 0.003 mg/ml, 0.007 mg/ml and 0.014 mg/ml) were previously tested in Chapter 6 and 7 in order to determine the individual optimal concentrations. Enzymatic activity was assayed spectrophotometrically at 340 nm in a 96 well plate by determining the Nicotinamide adenine dinucleotide phosphate (NADH) produced or oxidised, using the BioTek<sup>®</sup> Epoch 2 microplate reader (Agilent Technologies, USA). Both enzymes were added to the reaction mixture to start the reaction. This was mixed thoroughly and immediately placed in the microplate reader. The absorbance values were recorded at twenty second intervals for 3 minutes at 30 °C. The enzymatic activity of the individual enzymes in the cocktail were measured by the increase in  $A_{340}$  due to the formation of NADH.

## 8.3 Results and Discussion

### 8.3.1 Assessing the activity of the enzyme cocktails

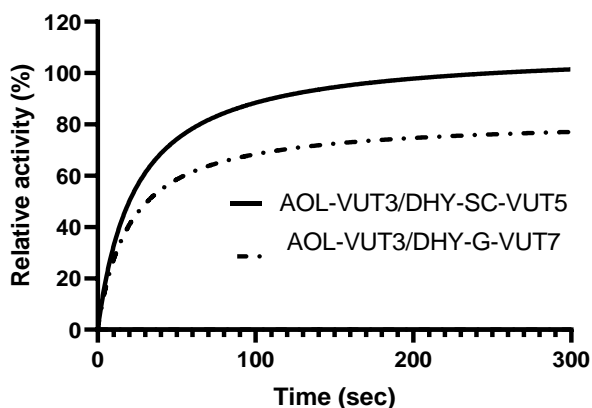
The activity of the two enzyme cocktails were assessed using the protocol developed by Li *et al.* (2010) for determining ADH activity. The activity was assayed spectrophotometrically at 340 nm by determining the amount of  $\text{NAD}^+$  converted to NADH. The absorbance values were recorded over a period of 3 minutes following the addition of both enzymes for a particular cocktail. The results obtained for both cocktails displayed a similar trend of activity for AOL-VUT3/DHY-SC-VUT5 cocktail and AOL-VUT3/DHY-G-VUT7 cocktail respectively. Absorbance data and graphs for the assays shown in Appendix C. This trend involved an initial increase in absorbance values between the 20 – 60 sec time interval. This was indicative of the production of NADH followed by a further spike in absorbance between the 60 – 120 sec intervals. Thereafter, the graph reached a plateau and no further increase in absorbance value was observed.

From the observation, it can be concluded that the initial spike in absorbance values was due to the activity of AOL-VUT3 catalysing the reaction of converting hexanol to hexanal. The second spike is as a result of the enzymes DHY-SC-VUT5 and DHY-G-VUT7 respectively working on the product from the initial reaction, converting hexanal to hexanoic acid. These spikes in absorbance readings are as a result of the conversion of  $\text{NAD}^+$  to NADH when both enzymes commence to work and confirm the transformation of hexanol to carboxylic acid. Therefore, both enzyme cocktails have shown the ability to work in synergy in the complete breakdown of hexanol to carboxylic acid. The complete depiction of the enzymatic reaction carried out by each cocktail is represented in **Figure 8.1**.



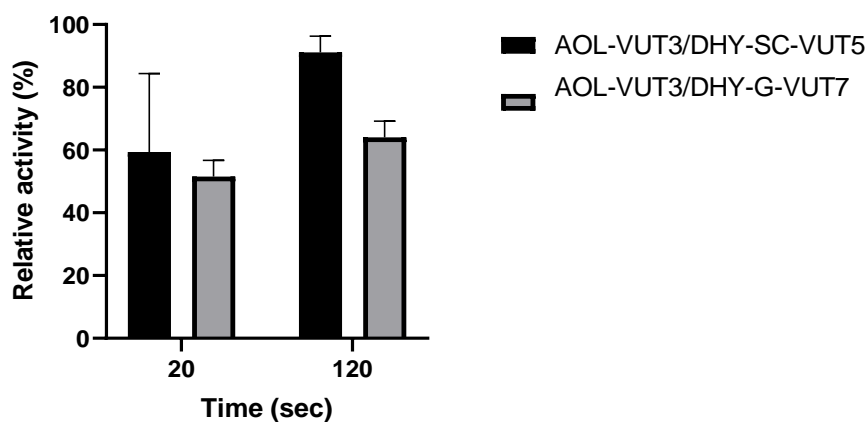
**Figure 8.1:** Diagrammatic representation of the enzymatic activity of two designed cocktails containing an alcohol dehydrogenase and an aldehyde dehydrogenase (AOL-VUT3/DHY-SC-VUT5 and AOL-VUT3/DHY-G-VUT7).

The relative enzymatic activities of both cocktails were calculated and plotted over a time interval 3 minutes as shown in **Figure 8.2**. The relative activities were compared in order to determine the most efficient combination and the cocktail of enzymes with the highest enzymatic activity when working in synergy. The results showed that AOL-VUT3/DHY-SC-VUT5 cocktail displayed the highest activity in comparison to AOL-VUT3/DHY-G-VUT7 with a percentage difference of 8 % and 27 % respectively. Additionally, both enzymes showed transformation of hexanol (20 sec) and hexanal (120 sec) resulting in carboxylic acid. Even though, both cocktails exhibited substantially high relative activities and ability to work in synergy, AOL-VUT3/DHY-SC-VUT5 cocktail was the most efficient and powerful (**Figure 8.2**).



**Figure 8.2:** Comparison of the relative activity of the enzyme cocktails (AOL-VUT3/DHY-SC-VUT5 and AOL-VUT3/DHY-G-VUT7) on the degradation of hexanol as a substrate. The kinetic data was fitted into Michaelis-Menten model using GraphPad Prism Version 8.4.1 (460).

The relative activities were used to plot **Figure 8.3** below after incubation of the enzyme cocktails along with hexanol for 20 and 120 seconds at 30 °C. This further shows the higher relative activity of AOL-VUT3/DHY-SC-VUT5 at both time intervals. Thus, confirming the findings in **Figure 8.2** above that AOL-VUT3/DHY-SC-VUT5 cocktail is more effective.



**Figure 8.3:** Comparison of the relative activity of the enzyme cocktails (AOL-VUT3/DHY-SC-VUT5 and AOL-VUT3/DHY-G-VUT7) on the degradation of hexanol as a substrate after 20 and 120 sec of incubation at 30 °C. Data was analysed using GraphPad Prism Version 8.4.1 (460).

There is limited study to date to successfully describe and implement an enzyme cocktail for efficient bioremediation of hydrocarbons. Recent study conducted by Ji *et al.* (2019) developed an enzyme cocktail containing a NADH regeneration system for efficient bioremediation of oil sludge contamination. This study showed a short fall of using unknown proteins derived from a crude mixture of intracellular enzymes from an oil-degrading strain *Acinetobacter calcoaceticus* 21 enhanced with formate dehydrogenase gene *Cbfdh* from *Candida boidinii*, cloned and functionally expressed in *E. coli* BL21. Their findings showed the degradation rate achieved for sludge oil to be 35.6 % after 12 hours however their cocktail posed identification problems for the specific enzymes involved in the degradation and enzymatic activity assessments could not be carried out on individual purified enzymes. This therefore makes the current study valuable since the derived cocktails contain identified, purified and formulated proteins to work in synergy. Thus, allowing for their enzymatic activities to be assessed individually and in the cocktails.

Another recent study by Gregson *et al.* (2019) involving assessing the differential protein expression on linear versus branched alkanes by the obligate marine hydrocarbon-degrading bacterium *Alcanivorax borkumensis* SK2<sup>T</sup>, identified an AlmA-type monooxygenase, alcohol dehydrogenase and an aldehyde dehydrogenase. The combination of these enzymes was confirmed to make up a complete pathway for the terminal oxidation of branched alkanes. Furthermore, Dhiman *et al.* (2014) implemented the use of an enzyme cocktail of xylanase from *Bacillus stearothermophilus* and laccase from *Ceriporiopsis subvermispora* to reduce the acute ecotoxicity and pollution of paper mill effluent showing a reduction of biological oxygen demand (BOD) and color of effluents by 60.1 % and 25.8 %, respectively.

A number of synergistic enzyme cocktails have been developed and successfully used for different applications at an industrial level. A study by Agrawal *et al.* (2018) successfully implemented the use of an enzyme cocktail ( $\beta$ -glucosidase, xylanase and laccase) for efficient hydrolysis of lignocellulosic biomass since no wild type organism is capable of producing all enzymes at desired levels. Another study by Falade *et al.* (2018) involved the use of ligninolytic enzyme cocktails (laccase, manganese peroxidase, and peroxidase) as biocatalysts for the elimination of endocrine-

disrupting chemicals in wastewater. Such examples show the potential of implementing enzyme cocktails rather than bacterial consortiums for faster and more efficient strategies.

Hexanol is known to be a toxic pollutant in different soil environments that have been accumulating due to anthropogenic practices through the years. Short-term exposure to hexanol causes irritation to the eyes, respiratory tract and skin. If this liquid is swallowed, aspiration into the lungs may result in chemical pneumonitis, which is inflammation of the lungs or breathing difficulty that could be acute or chronic. Furthermore, long-term exposure by individuals can also affect the nervous system (Hazardous substance fact sheet 2007). Therefore, the first of its kind enzyme cocktails developed in this study are invaluable tools in transforming these contaminants into less harmful compounds that can be used by microorganisms and made more readily available for uptake by plants. The enzyme cocktails will speed up the transformation time that naturally takes place over long periods by the natural microbiota in the soil. Thus, generating carboxylic acids such as hexanoic acids back into the soil allows for faster remediation of hydrocarbon polluted soil environments.

Furthermore, the value of carboxylic acids or organic acids in soil offers advantages at both soil and plant level since they can be mobilized and solubilized more easily as compared to the complex hydrocarbon classes (Nigam and Srivastava 2005, López *et al.* 2012). This helps with soil permeability, desalination of soils, reduction of nutrient leaching, stimulation of the plant root system and pH regulation of the rhizosphere to absorb more nutrients (Lan *et al.* 2019). Thus, promoting plant growth due to the increase in soil quality and speeding up the bioremediation process of polluted soils and allowing for reclamation of infertile land.

Most importantly, carboxylic acids and their derivatives are used in the production of polymers, biopolymers, coatings, adhesives, and pharmaceutical drugs (Thomsen *et al.* 2008). They can also be used as solvents, food additives, antimicrobials, and flavorings (Murali *et al.* 2017). Thus, the

developed enzyme cocktails can be used to remove the toxic hexanol from hydrocarbon polluted soils and further, use the carboxylic acid formed for industrial applications.



## 8.4 Conclusion

The current chapter focused on assessing the activity of two enzyme cocktails (AOI-VUT3/DHY-SC-VUT5 and AOL-VUT3/DHY-G-VUT7) which were designed for possible and complete conversion of hexanol to carboxylic acid via a two-step reaction. The three proteins, namely alcohol dehydrogenase (AOL-VUT3) and two aldehyde dehydrogenases (DHY-SC-VUT5 and DHY-G-VUT7) were obtained from a large insert fosmid library previously prepared using metagenomic DNA isolated from hydrocarbon contaminated soil samples. The study focused on established concentrations of enzymes and substrates identified in Chapter 6 and 7 to depict the optimal activity of the enzyme combinations. Additionally, the study further compared the relative activity of the two designed cocktails using hexanol as a substrate to determine if these enzymes can work in synergy to obtain complete transformation of hexanol into carboxylic acid. Therefore, the successful enzymatic transformation of the toxic hexanol into carboxylic acid is a major breakthrough since, carboxylic acids and their derivatives are used in the production of polymers, biopolymers, coatings, adhesives, and pharmaceutical drugs. They can also be used as solvents, food additives, antimicrobials, and flavorings. Thus, the developed enzyme cocktails provide two-fold advantage as they remove the toxic hexanol and produce the very useful carboxylic acid. Hence, future study should focus on immobilization and stabilization of the engineered enzyme cocktails for large scale bioremediation of alkanols and an industrial scale production of carboxylic acid from hexanol.

## 8.5 References

- Aalbers, F.S. and Fraaije, M.W., 2017. Coupled reactions by coupled enzymes: alcohol to lactone cascade with alcohol dehydrogenase–cyclohexanone monooxygenase fusions. *Applied Microbiology and Biotechnology*, 101 (20), 7557–7565.
- Agrawal, R., Semwal, S., Kumar, R., Mathur, A., Gupta, R.P., Tuli, D.K., and Satlewal, A., 2018. Synergistic enzyme cocktail to enhance hydrolysis of steam exploded wheat straw at pilot scale. *Frontiers in Energy Research*, 6 (NOV), 1–11.
- Dhiman, S.S., Garg, G., Sharma, J., Kalia, V.C., Kang, Y.C., and Lee, J.K., 2014. Reduction in acute ecotoxicity of paper mill effluent by sequential application of xylanase and laccase. *PLoS ONE*, 9 (7).
- Falade, A.O., Mabinya, L. V., Okoh, A.I., and Nwodo, U.U., 2018. Ligninolytic enzymes: Versatile biocatalysts for the elimination of endocrine-disrupting chemicals in wastewater. *MicrobiologyOpen*, 7 (6), 1–17.
- Gregson, B.H., Metodieva, G., Metodiev, M. V., and McKew, B.A., 2019. Differential protein expression during growth on linear versus branched alkanes in the obligate marine hydrocarbon-degrading bacterium *Alcanivorax borkumensis* SK2T. *Environmental Microbiology*, 21 (7), 2347–2359.
- Hazardous substance fact sheet, 2007. HOW TO DETERMINE IF YOU ARE BEING Medical Testing HEALTH HAZARD INFORMATION Acute Health Effects Cancer Hazard Reproductive Hazard Other Long-Term Effects.
- Jacoby, R., Peukert, M., Succurro, A., Koprivova, A., and Kopriva, S., 2017. The role of soil microorganisms in plant mineral nutrition—current knowledge and future directions. *Frontiers in Plant Science*, 8 (September), 1–19.
- Ji, L., Fu, X., Wang, M., Xu, C., Chen, G., Song, F., Guo, S., and Zhang, Q., 2019. Enzyme cocktail containing NADH regeneration system for efficient bioremediation of oil sludge contamination. *Chemosphere*, 233, 132–139.

- Jin, M., Gai, Y., Guo, X., Hou, Y., and Zeng, R., 2019. Properties and applications of extremozymes from deep-sea extremophilic microorganisms: A mini review. *Marine Drugs*, 17 (12).
- Kamagata, Y. and Narihiro, T., 2016. Symbiosis studies in microbial ecology. *Microbes and Environments*, 31 (3), 201–203.
- Lan, X., Du, H., Peng, W., Liu, Y., Fang, Z., and Song, T., 2019. Functional diversity of the soil culturable microbial community in eucalyptus plantations of different ages in Guangxi, South China. *Forests*, 10 (12).
- Li, X., Li, Y., Wei, D., Li, P., Wang, L., and Feng, L., 2010. Characterization of a broad-range aldehyde dehydrogenase involved in alkane degradation in *Geobacillus thermodenitrificans* NG80-2. *Microbiological Research*, 165 (8), 706–712.
- López, L., Hernández, M., Ruiz, J., Carcaño, M., Medina, G., Portillo, R., and Muñoz, J., 2012. Adsorption of Plant and Bacterial Carboxylic Acids in Agricultural Soil. *Terra Latinoamericana*, 30 (3), 261–270.
- Murali, N., Srinivas, K., and Ahring, B.K., 2017. Biochemical production and separation of carboxylic acids for biorefinery applications. *Fermentation*, 3 (2), 1–25.
- Nigam, R. and Srivastava, M.M., 2005. Effects of carboxylic and amino acids on Cd uptake by *Lycopersicon esculentum*. *Chemical Speciation and Bioavailability*, 17 (1), 19–26.
- Peixoto, R.S., Vermelho, A.B., and Rosado, A.S., 2011. Petroleum-degrading enzymes: Bioremediation and new prospects. *Enzyme Research*, 2011 (1).
- Robinson, P.K., 2015. Enzymes: principles and biotechnological applications. *Essays in Biochemistry*, 59, 1–41.
- Thomsen, A., Malmstrom, E., and Hvilsted, S., 2008. Stability and Utility of Pyridyl Disulfide Functionality in RAFT and Conventional Radical Polymerizations. *Journal of Polymer Science: Part A: Polymer Chemistry*, 46 (I), 7207–7224.
- Tiralerdpanich, P., Sonthiphand, P., Luepromchai, E., Pinyakong, O., and Pokethitiyook, P.,

2018. Potential microbial consortium involved in the biodegradation of diesel, hexadecane and phenanthrene in mangrove sediment explored by metagenomics analysis. *Marine Pollution Bulletin*, 133 (June), 595–605.

Truskewycz, A., Gundry, T.D., Khudur, L.S., Kolobaric, A., Taha, M., Aburto-Medina, A., Ball, A.S., and Shahsavari, E., 2019. Petroleum hydrocarbon contamination in terrestrial ecosystems—fate and microbial responses. *Molecules*, 24 (18), 1–20.

## GENERAL CONCLUSION AND FUTURE PROSPECTS

The primary objective for this study was to isolate metagenomic DNA from different hydrocarbon contaminated soil samples and construct a metagenomic DNA library to recover novel genes that could possibly code for novel enzymes involved in the hydrocarbon degradation pathways. The discovery of these enzymes are crucial for the bioremediation of hydrocarbon contaminated soil and water environments due to the high levels of anthropogenic practices and high demand of petroleum as a major energy source in everyday life.

In this study, total metagenomic DNA was successfully isolated from hydrocarbon contaminated soil samples and ligated to PCC2FOS to create fosmid libraries that were subsequently screened for hydrocarbon degrading potential. Microbial profiling studies showed that within the bacterial domain, *Proteobacteria* and *Actinobacteria* were found to be the predominant phyla in the metagenomic DNA and detected at roughly 84 % and 8 % phylum abundance respectively. Sequence and functional-based metagenomics recovered an AMO  $\alpha$ -subunit (amoA1) of 36 kDa, an ADH of 25 kDa and two ALDHs of 50.6 kDa and 67.6 kDa. Kinetic studies showed the high affinity of the enzymes for their respective substrates. Two enzyme cocktails designed displayed high relative activity, working in synergy for alkanol degradation. This is a valuable finding due to the limitation of enzyme cocktails on the market for bioremediation.

In conclusion, the outcome of this work proves that hydrocarbon contaminated soils are a potential source of hydrocarbon degrading enzymes. The *in-silico* analysis reported very low similarities with known or studied AMO  $\alpha$ -subunit (amoA1), ADH and ALDH identified in this study. Hence, it could be concluded that the enzymes isolated from this study were indeed novel. Functional metagenomics proved to be an invaluable technique for the discovery of hydrocarbon degrading enzymes for its potential application in bioremediation strategies. Furthermore, the development of two enzyme cocktails working in synergy has added immensely to the knowledge base for application of enzyme mixtures to oil spills to attain complete degradation.

Future work would include assessing the enzyme cocktails on different classes of organic compounds to determine their degradation range. Furthermore, identifying all the subunits for the monooxygenase protein from future metagenomics work can help to derive an active enzyme for future application. The crystallization of the enzymes identified and purified could also provide additional information on the mechanism of action and folding pattern involved.

## APPENDIX

### Appendix A: Materials used in the study

#### Appendix A1: List of Buffers and solutions used in the study.

Buffers and solutions	Composition	pH
10X TBE	108 g Tris base; 55 g Boric acid; 7.45 g EDTA In 1000 ml of dH <sub>2</sub> O	~8.3
10X TGS	30.0 g of Tris base; 144.0 g of glycine; 10.0 g of SDS In 1000 ml of dH <sub>2</sub> O.	~8.3
6X DNA Loading dye	30 % (v/v) glycerol 0.25 % (w/v) bromophenol blue 0.25 % (w/v) xylene cyanol FF	-
Binding Buffer	20 mM sodium phosphate; 0.5 M NaCl; 20-40 mM imidazole In 1000 ml of dH <sub>2</sub> O	7.4

Coomassie staining solution	10 % (V/V) Acetic acid;  50 % (V/V) Methanol;  0.1 % (W/V) coomassie blue R250;  40 % dH <sub>2</sub> O	-
CTAB	100 mM Tris-HCl;  1.4 M NaCl;  20 mM EDTA;  2 % CTAB  In 1000 ml of dH <sub>2</sub> O	8.0
De-staining solution	10 % (V/V) Acetic acid;  50 % (V/V) Methanol;  40 % dH <sub>2</sub> O	-
1 M EDTA	186.1 g of disodium EDTA•2H <sub>2</sub> O;  In 400 mL of dH <sub>2</sub> O  Adjust the pH to 8.0 with NaOH	8.0
Elution Buffer	20 mM sodium phosphate;  0.5 M NaCl;  500 mM imidazole  In 1000 ml of dH <sub>2</sub> O	7.4
End-Repair 10X Buffer	330 mM Tris-acetate [pH 7.5];	7.5



	660 mM potassium acetate; 100 mM magnesium acetate; 5 mM DTT In 100 ul	
Laemmli sample buffer	65.8 mM Tris-HCl; 26.3 % (w/v) glycerol; 2.1 % SDS; 0.01 % Bromophenol blue	6.8
1 M PBS	610 mL of 1M Na <sub>2</sub> HPO <sub>4</sub> ; 390 mL of 1M NaH <sub>2</sub> PO <sub>4</sub>	7.0
Phage Dilution Buffer	10 mM Tris-HCl [pH 8.3]; 100 mM NaCl; 10 mM MgCl <sub>2</sub> In 1000 ml of dH <sub>2</sub> O	8.3
TE	10 mM Tris-HCl (pH 8.0); 1 mM EDTA (pH 8.0)	8.0
1 M Tris-HCl	121.1 g of Tris base in 800 ml of H <sub>2</sub> O; Adjust pH with 32% HCl	8.0
Tfbl solution	30 mM Potassium acetate 100 mM Rubidium chloride	8.0

	10 mM Calcium chloride 50 mM Manganese chloride 15 % (v/v) Glycerol	
TfbII solution	10 mM (3-(N-morpholino)propanesulfonic acid) 75 mM Calcium chloride 10 mM Rubidium chloride 15 % (v/v) Glycerol	pH 6.5
Wash buffer (Protein purification)	50 mM KH <sub>2</sub> PO <sub>4</sub> -K <sub>2</sub> HPO <sub>4</sub> 300 mM NaCl 40 mM Imidazole	pH 8.0
Elution buffer (Protein purification)	50 mM KH <sub>2</sub> PO <sub>4</sub> -K <sub>2</sub> HPO <sub>4</sub> 300 mM NaCl 300 mM Imidazole	pH 8.0

**Appendix A2: List of antibiotics and inducers used in the study.**

<b>Reagents</b>	<b>Preparation</b>
Chloramphenicol (34 mg/ml): antibiotic	0.680 g of chloramphenicol;  20 ml of 100 % ethanol  Filter-sterilised and aliquoted in 2 ml sterile microcentrifuge tubes.  12.5 µg/ml was the final concentration used.
Kanamycin (100 mg/ml): antibiotic	2 g of kanamycin;  20 ml of distilled water  Filter-sterilised and aliquoted into 2 ml sterile microcentrifuge tubes.  50 µg/ml was the final concentration used.
Isopropyl-β-D-thiogalactopyranoside (IPTG) 1M: inducer	4.77 g of IPTG  20 ml of distilled water  Filter-sterilised and aliquoted in 2 ml sterile microcentrifuge tubes.  0.1 M is the final concentration used

**Appendix A3: List of media used in the study.**

<b>Media</b>	<b>Formulation per Liter</b>	<b>Preparation</b>
LB agar	10 g Peptone; 5 g Yeast Extract; 10 g Sodium Chloride; 12 g Bacteriological Agar	All the components are mixed and autoclaved at 121 °C for 15 minutes. Medium was cooled at 50-60 °C to be poured in sterile petri dishes.  (pH ~7.5)
LB broth	10 g Peptone; 5 g Yeast Extract; 10 g Sodium Chloride	All the components are mixed and autoclaved at 121 °C for 15 minutes.  (pH ~7.5)
BH agar	0.2 g Magnesium Sulphate 0.02 g Calcium Chloride 1 g Monopotassium Phosphate 1 g Dipotassium Phosphate 1 g Ammonium Nitrate 0.05 g Ferric Chloride 20 g Agar	Suspended 23.27 g of BH media in 1000 ml distilled water, boil until completely dissolved. Sterilise by autoclaving at 121 °C for 15 minutes.  (pH 7.0)
SOC agar	10 mM magnesium chloride, 10 Mm magnesium sulfate,	All components are mixed and autoclaved at 121 °C for 15 minutes. The medium was cooled and 20 mM

	2.5 mM potassium chloride, 10 mM sodium chloride, 2 % tryptone, 0.5 % yeast extract, 20 mM glucose (to be added after autoclaving)	glucose added after autoclaving.
--	---------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------

**Appendix A4: List of all microorganisms used in the study.**

Strains	Genotype/Features	Source
EPI300™-T1R Phage T1-resistant <i>Escherichia coli</i> Plating strain	[F- <i>mcrA</i> Δ( <i>mrr-hsdRMS-mcrBC</i> ) ( <i>StrR</i> ) φ80 <i>dlacZ</i> Δ <i>M15</i> Δ <i>lacX74</i> <i>recA1 endA1 araD139</i> Δ( <i>ara</i> , <i>leu</i> )7697 <i>galU galK</i> λ- <i>rpsL nupG</i> <i>trfA tonA dhfr</i> ]	Epicentre, USA
<i>Escherichia coli</i> BL21 (DE3)pLysS competent cells	F- <i>ompT gal dcm lon hsdS<sub>B</sub></i> ( <i>r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>) λ(<i>DE3</i> [<i>lacI lacUV5-T7p07 ind1</i> <i>sam7 nin5</i>]) [<i>malB<sup>+</sup></i>]<sub>K-12</sub>(λ<sup>S</sup>) <i>pLysS</i>[<i>T7p20 ori<sub>p15A</sub></i>](<i>Cm<sup>R</sup></i>)</i>	Lucigen

**Appendix A5: List of all vectors used in the study.**

<b>Vectors</b>	<b>Features</b>	<b>Selective marker</b>	<b>Source</b>
pCC2FOS	Copy controlled vector, linearized and dephosphorylated at Eco72I restriction site.  Requires EPI300™-T1R E. coli strain for high copy number induction, used for construction of fosmid library.	Chloramphenicol	Epicentre, USA
pCC2pFos_D8	pCC2FOS derived constructs from the hydrocarbon contaminated soil metagenomic library.	Chloramphenicol	This study
pCC2pFos_D3	pCC2FOS derived constructs from the hydrocarbon contaminated soil metagenomic library.	Chloramphenicol	This study
pCC2pFos_B2	pCC2FOS derived constructs from the hydrocarbon contaminated soil metagenomic library.	Chloramphenicol	This study
pCC2pFos_A2	pCC2FOS derived constructs from the hydrocarbon contaminated soil metagenomic library.	Chloramphenicol	This study
pET30a(+)	Expression vector of N and C-terminally His-tagged protein.	Kanamycin	Merck, USA

pET30a(+)-amo-vut1	pET30a(+) derived expression vector construct containing alpha subunit (amoA1) synthesized gene.	Kanamycin	This study
pET30 (+)_aol-vut3	pET30a(+) derived expression vector construct containing alcohol dehydrogenase synthesized gene.	Kanamycin	This study
pET30a(+)_dhy-sc-vut5	pET30a(+) derived expression vector construct containing aldehyde dehydrogenase synthesized gene.	Kanamycin	This study
pET30a(+)_dhy-g-vut7	pET30a(+) derived expression vector construct containing aldehyde dehydrogenase synthesized gene.	Kanamycin	This study

**Appendix A6: Preparation of 15 % resolving gel and 5 % stacking gel for SDS-PAGE.**

Solution components	Volume (10 ml)	
	15 % Resolving gel	5 % stacking gel
Distilled water	2.3	6.8
30 % Bis-acrylamide mix	5.0	1.7
1.5M Tris-HCl (pH 8.8)	2.5	1.25
10 % SDS	0.1	0.1
10 % Ammonium persulfate (APS)	0.1	0.1
TEMED	0.004	0.01

**Appendix B: Supplementary results for the molecular part of the study.**

**Appendix B1: Summary of statistical analysis of the *De novo* assembly and alignments of DNA sequences of pCC2pFos\_A2, pCC2pFos\_B2 and pCC2pFos\_D8 clones.**

▪ **pCC2pFos\_A2 clone**

	Count	Average length	Total bases
Reads	2,847	5,456.85	15,535,658
Matched	1,976	3,757.62	7,425,052
Not matched	871	9,311.83	8,110,606
Contigs	16	7,819	125,118

▪ **pCC2pFos\_B2**

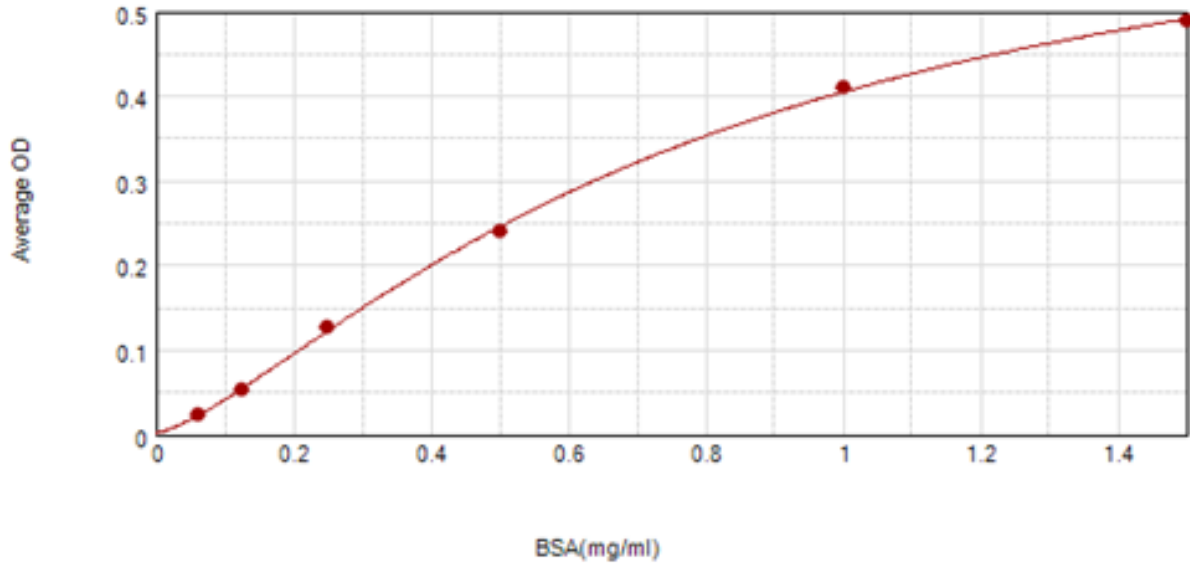
	Count	Average length	Total bases
Reads	2,796	5,321.11	14,877,828
Matched	1,913	3,651.58	6,985,466
Not matched	883	8,938.12	7,892,362
Contigs	31	6,668	206,725

▪ **pCC2pFos\_D8**

	Count	Average length	Total bases
Reads	4,052	5,369.75	21,758,217
Matched	2,953	4,006.67	11,831,682
Not matched	1,099	9,032.33	9,926,535
Contigs	35	6,385	223,476



**Appendix B2: Standard curve for Bradford assay at wavelength of 595 nm.**



● STD#1 (Standards@Bradford: AvgOD vs Standard Value) Weighting: Fixed

Curve Fit Results ▲

Curve Fit : 4-Parameter Logistic  $y = D + \frac{A - D}{1 + \left(\frac{x}{C}\right)^B}$

	Parameter	Estimated Value	Std. Error	Confidence Interval
STD#1	A	0.002	0.010	[-0.043, 0.047]
R <sup>2</sup> = 1.000	B	1.363	0.145	[0.739, 1.987]
EC50 = 0.779	C	0.779	0.102	[0.341, 1.218]
	D	0.692	0.061	[0.431, 0.953]

**Appendix B3: Summary of BLAST search results for the four ORFs**

ORF	Nucleotide Sequences	Amino acids sequences	Family	Best HIT
<i>amo-vut1</i>	<p>CTGATGAGTAAATCACACCCGCGCTGGCGCTTAGCAAAGAAGATCCTCAC</p> <p>CTGGCTGTTTTTATCGCGGTGATTGTGTTGCTGGTGGTCTACGCCAAAAA</p> <p>AGTGGACTGGGAAGAGGTCTGGAAGGTCATCCGCGACTACAATCGCGTT</p> <p>GCGCTGCTTAGTGCGGTCTGGGCTGGTGGTCGTCAGCTATCTGATTTACGG</p> <p>CTGCTATGACCTGCTCGCCGCTTTACTGCGGTACAAACTGGCGAAGCG</p> <p>CCAGGTGATGCTGGTGTCTTTATCTGCTACGCCTTCAACCTGACGCTCAG</p> <p>TACCTGGGTCGGCGGCATTGGTATGCGCTATCGTTTGTACTCTCGTCTGGG</p> <p>GTTGCCCGTAGCACTATTACGCGGATTTTCTCGCTCAGTATTACCACCAA</p> <p>CTGGCTGGGCTACATTTTACTGGCAGGGATTATCTTTACCGCAGGCGTGGT</p> <p>GGAGTTGCCGGATCACTGGTATGTCGATCAAACACTACGCTGCGTATTCTCGG</p> <p>CATTGGCTTACTGATGATTATCGCGGTTTATTTGTGGTTTTGCGCTTTCGCG</p> <p>AAGCACCGCCATATGACCATCAAAGGGCAAAGCTGGTGTGCCTTCATG</p> <p>GAAATTCGCCCTCGCTCAGATGCTGATTTCCAGCGTAACTGGATGGTAAT</p> <p>GGGGGCGATTATCTGGCTGTTACTTGGTCAAAGCGTGAACATTTCTTTGT</p> <p>ACTGGGCGTGTTACTGGTTAGTAGTATTGCTGGCGTCATCGTGCATATTCC</p> <p>GGCGGGGATCGGTGTGCTGGAAGCGGTGTTTATCGCACTACTGGCTGGG</p> <p>GAGCATACCTCAAAGGTACAATTATCGCCGCCCTACTCGCTTACCGTGTG</p> <p>CTGTATTACTTTATCCACTGCTGCTGGCGCTGATCTGTTACCTGTTGCTGG</p> <p>AAAGCCAGGCGAAGAAGCTACGGGCGAAAAATGAAGCGGCGATGTGA</p>	<p>MMSKSHPRWRLAKKILTWLF</p> <p>FIATIVLLVYAKKVDWEEVW</p> <p>KVIRDYNRVALLSAVGLVVVSY</p> <p>LIYGCYDLLARFYCGHKLAKRQ</p> <p>VMLVSFCYAFNLTSTWVGGI</p> <p>GMRYRLYSRLGPLGSTITRIFSL</p> <p>SITTNWLGYILLAGIIFTAGVVE</p> <p>LPDHWYVDQTTLRILIGILLMI</p> <p>IAVYLVFCAFAKHRHMTIKGQ</p> <p>KLVLPWKFALAQMLISSVNW</p> <p>MVMGAIWLLLGQSVNYFFVL</p> <p>GVLLVSSIAGVIVHIPAGIGVLE</p> <p>AVFIALLAGEHTSKGTIIAALLA</p> <p>YRVLYYFIPLLLALICYLLESQA</p> <p>KKLRAKNEAAM</p>	<p>Ammonia</p> <p>monooxygenase</p>	<p>AMO <math>\alpha</math>-subunit</p> <p>(<i>amoA1</i>) exhibited</p> <p>similarity in sequence</p> <p>to that found in</p> <p><i>Nitrosomomas</i></p> <p><i>europaea</i> (strain ATCC</p> <p>19718 / CIP 103999 /</p> <p>KCTC 2705 / NBRC</p> <p>14298)</p> <p>(EC:1.14.99.39).</p> <p>Displayed 21.3 %</p> <p>similarity.</p>

<p><i>aol-vut3</i></p>	<p>CTGGGTATTGTTTGTCCCTGAGCGCGGTTGGTGCTGTTGGCGTTCGGT  CAGTGCCGACTGTCCAGTGGGCTTTTCTGTTCTGTTTCATCCATTACCACT  TAACCGCCTTTGGCGTTGCAGCAAGCGTTTCAGACGTGCTGTTGGTTGCAC  TGCTGAGCTGCACTATCCCCTTTCTCGTTGTGTCCGCATCCTCAAGCGCGAC  AGCTGAAGCTATATCTTCTGCACGTTTTGCCGAATTTTCTGCACGTATTGCC  GCCGCTTCTGCCGCACTTTTGTCTGCGATGCTGATACCGCACTTCCCGCA  GCCTCTGTGCCTTCGTGGATGCCGTTGACGCACTCCCCGCCGCGCTGTT  TTTGCGTCTGCCGCGCAGAGGCGCTCCGTTCCGCTGCTGTTTCAGATGAC  CTGGCATTGCTCTCGGACGTTTTTCCCGCCCTGGCAGAATTTTCTGCCGCC  GTTGCCGAGGAAGCTGCACGACCGGCACTTGATGATGCGTTCGTTTCTGA  TGATTTTGTGCCTCTTTTGTAGGCCACCGCATCTCGTGTGAAGTGGCGGC  CTCTGACGCTTTCGTGGCCGCGGTGGAGGCAGACGTGGCGGCTGATTGTT  GTGACGCTGCAGCATTCTGTTTCTGACGTTTTTCCCGCACCGGCACTGGTGG  CCGCCGCTTTTTTGTAGGACTCTGCGGCTGCGGCACTTTTTTCCGCTTCAGT  GGCCTTTGCTGATGCCGCTTCTGCGCCGGAGGACGCTTCTCTGA</p>	<p>MGIVCSPERGWCCWRSQVC  RTVQWAFLVSSITTLTAFGVA  ASVSDVLLVALLSCTIPFLVSA  SSSATAEAISSARFAEFSARIAA  ASAALLLCDADTALPAASVAFV  DAVDALPAAAVFASAAAALR  SAAVSDDLAFVSDVFAALAEFS  AAVAEEAARPALDDAFVSDDF  AASFEATASRAEVAASDAFVA  AVEADVAADCCDAAAFVSDV  FAAPALVAAAFEDSAAAALFS  ASVAFADAASAPEDAS</p>	<p>Alcohol  dehydrogenase</p>	<p>Histidinol  dehydrogenase (HDH)  exhibited similarity in  sequence to that  found in  <i>Mycobacterium bovis</i>  (strain ATCC BAA-935  / AF2122/97)  (EC:1.1.1.23).  Displayed 24.7 %  similarity.</p>
<p><i>dhy-sc- vut5</i></p>	<p>ATGGCAGTAAAGATTTTCAGGAGTCTGAAAGACGGCACAGGAAAACCGG  TACAGAACTGCACCATTTCAGTGAAGCCAGACGTAACAGCACCGGTG  GTGGTGAACACGGTGGGCTCAGAGAATCCGGATGAAGCCGGGCGTTACA  GCATGGATGTGGAGTACGGTCAGTACAGTGCATCCTGCAGGTTGACGGT  TTTCCACCATCGCACGCCGGGACCATCACCGTGTATGAAGATTCACAACCG  GGGACGCTGAATGATTTTCTCTGTGCCATGACGGAGGATGATGCCCCGGCC  GGAGGTGCTGCGTCGTCTTGAAGTATGATGGTGAAGAGGTGGCGCGTAAC  GCGTCCGTGGTGGCACAGAGTACGGCAGACGCGAAGAAATCAGCCGGCG</p>	<p>MAVKISGVLKDGTKPVQNC  TIQLKARRNSTTVVNTVGSE  NPDEAGRYSMDVEYGQYSVIL  QVDGFPPSHAGTITVYEDSQP  GTLNDFLCAMTEDDARPEVLR  RLELMVEEVARNASVVAQSTA  DAKKSAGDASASAAQVAALVT  DATDSARAASTSAGQAASSA</p>	<p>Dehydrogenase</p>	<p>Short-chain  dehydrogenase  exhibited similarity in  sequence to that  found in <i>E. coli</i>.  Displayed 96.7 %  similarity.</p>

	<p>ATGCCAGTGCATCAGCTGCTCAGGTGCGGGCCCTTGACTGATGCAACG  ACTCAGCACGCGCCGCCAGCACGTCCGCGGACAGGCTGCATCGTCAGCT  CAGGAAGCGTCTCCGGCGCAGAAGCGGCATCAGCAAAGGCCACTGAAG  CGGAAAAAAGTGCCGAGCCGAGAGTCTCAAAAAACGCGGGCGCCAC  CAGTGCCGGTGCGGGCGAAAACGTCAGAAACGAATGCTGCAGCGTCACAA  CAATCAGCCGCCACGTCTGCCTCCACCGCGGCCACGAAAGCGTCAGAGGC  CGCCACTTCAGCACGAGATGCGGTGGCTCAAAAGAGGCAGCAAATCAT  CAGAAACGAACGCATCATCAAGTGCCGGTTCGTGCAGCTTCTCGGCAACG  GCGGCAGAAAATTCTGCCAGGGCGGCAAAAACGTCCGAGACGAATGCCA  GGTCATCTGAAACAGCAGCGGAACGGAGCGCCTCTGCCGCGGCAGACGC  AAAAACAGCGGGCGGGGAGTGCGTCAACGGCATCCACGAAGGCGACA  GAGGCTGCGGGAAGTGCGGTATCAGCATCGCAGAGCAAAAGTGCGGCAG  AAGCGGCGGCAATACGTGCAGAAAATTCGGCAAAACGTGCAGAAGATAT  AGCTTCAGCTGTCGCGCTTGAGGATGCGGACACAACGAGAAAGGGGATA  GTGCAGCTCAGCAGTGCAACCAACAGCACGTCTGAAACGCTTGCTGCAAC  GCCAAAGGCGGTTAAGGTGGTAATGGATGAAACGAACAGAAAAGCCAC  TGGACAGTCCGGCACTGA</p>	<p>QEASSGAEASAKATEAEKSA  AAAESSKNAAATSAGAAKTSE  TNAASQQAATSASTAATKA  SEAATSARDAVASKEAAKSSET  NASSAGRAASSATAAENSAR  AAKTSETNARSSETAAERSASA  AADAKTAAAGSASTASTKATE  AAGSAVSASQSKSAAEAAAIR  AENSAKRAEDIASAVALEDAD  TTRKGIVQLSSATNSTSETLAA  TPKAVKVVMDETNRKAHWTV  RH</p>		
<i>dhy-g-vut7</i>	<p>TTGCCCAGCAGGTCCGTGAAACGATGGAGCGCCGTGCAGCCGGTCTTAAA  CCGCCCCCTGGGCGGCTGCAGCATTTGAATCCGGGCTGCGACAATCAAC  AGAGGAGGAGAAGAGTGACAGCAGAGCTGCGTAATCTCCCGCATATTGC  CAGCATGGCCTTTAATGAGCCGCTGATGCTTGAACCCGCTATGCGCGGG  TTTTCTTTGTGCGCTTGCAAGCCAGCTTGGGATCAGCAGCCTGACGGATG</p>	<p>MPSRSVKRWSAVQPVLNRPP  GRLQHLNPGCDNQRRRRVT  AELRNLPHIASMAFNEPLMLE  PAYARVFFCALAGQLGISSLTD  AVSGDSLTAQEALATLALSGD  DDGPRQARSYQVMNGIAVLP</p>	Dehydrogenase	<p>Glycine  dehydrogenase  (decarboxylating)  exhibited similarity in  sequence to that  found in <i>Caulobacter</i></p>

<p>CGGTGTCCGGCGACAGCCTGACTGCCAGGAGGCACTCGCGACGCTGGC  ATTATCCGGTGATGATGACGGACCACGACAGGCCCGCAGTTATCAGGTCA  TGAACGGCATCGCCGTGCTGCCGGTGTCCGGCACGCTGGTCAGCCGGACG  CGGGCGCTGCAGCCGTACTIONCGGGGATGACCGGTTACAACGGCATTATCGC  CCGTCTGCAACAGGCTGCCAGCGATCCGATGGTGGACGGCATTCTGCTCG  ATATGGACACGCCCGCGGGATGGTGGCGGGGGCATTGACTGCGCTGA  CATCATCGCCCGTGTGCGTGACATAAAACCGGTATGGGCGCTTGCCAACG  ACATGAACTGCAGTGCAGGTGAGTTGCTTGCCAGTGCCGCTCCCGGCGT  CTGGTCACGCAGACCGCCCGGACAGGCTCCATCGGCGTCATGATGGCTCA  CAGTAATTACGGTGCTGCGCTGGAGAAACAGGGTGTGGAAATCACGCTG  ATTTACAGCGGCAGCCATAAGGTGGATGGCAACCCCTACAGCCATCTTCC  GGATGACGTCCGGGAGACACTGCAGTCCCGGATGGACGCAACCCGCCAG  ATGTTTGCAGCAGAAGGTGTGCGCATATAACCGGCTGTCCGTGCAGTTGT  GCTGGATACCGAGGCTGCAGTGTACAGCGGTCAGGAGGCCATTGATGCC  GGACTGGCTGATGAACTTGTTAACAGCACCGATGCGATCACCGTCATGCG  TGATGCACTGGATGCACGTAAATCCCGTCTCTCAGGAGGGCGAATGACCA  AAGAGACTCAATCAACAACTGTTTCAGCCACTGCTTCGAGGCTGACGTTA  CTGACGTGGTGCCAGCGACGGAGGGCGAGAACGCCAGCGCGGCGCAGCC  GGACGTGAACGCGCAGATCACCGCAGCGGTTGCGGCAGAAAACAGCCGC  ATTATGGGGATCCTCAACTGTGAGGAGGCTCACGGACGCGAAGAACAGG  CACGCGTGCTGGCAGAAACCCCGGTATGACCGTGAAAACGGCCCGCCGC  ATTCTGGCCGACACCACAGAGTGCACAGGCGCGCAGTGACACTGCGCT</p>	<p>VSGTLVSRTRALQPYSGMTGY  NGIARLQQAASDPMVDGILL  DMDTPGGMVAGAFDCADIIA  RVRDIKPWWALANDMNCSAG  QLLASAASRRLVTQTARTGSIG  VMMAHSNYGAALEKQGVET  LIYSGSHKVDGNPYSHLPDDV  RETLQSRMDATRQMFAQKVS  AYTGLSVQVVLDEAAVYSGQ  EAIDAGLADELVNSTDAITVM  RDALDARKSRLSGGRMTKETQ  STTVSATASQADVTDVVPATE  GENASAAQPDVNAQITAAVA  AENSRLMILNCEEAHGREEQ  ARVLAETPGMTVKTARRILAA  APQSAQARSDTALDRLMQGA  PAPLAAGNPASDAVNDLLNTP  V</p>		<p><i>vibrioides</i> (strain  NA1000 / CB15N)  (<i>Caulobacter  crescentus</i>) (EC  1.4.4.2).  Displayed 23.9 %  similarity.</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	-----------------------------------------------------------------------------------------------------------------------------------------------------------

	GGATCGTCTGATGCAGGGGGCACCGGCACCGCTGGCTGCAGGTAACCCG GCATCTGATGCCGTTAACGATTTGCTGAACACACCAGTGTA			
--	------------------------------------------------------------------------------------------------	--	--	--



	<p>AGCGGCGCGCCCGGCGCTGGATGATGCGTTTGTGAGCGATGATTTTGCGGCGAGCTTTGAAGCGACCCGCGAGCCGCGCGGAAGTGGCGGC  GAGCGATGCGTTTGTGGCGGCGGTGGAAGCGGATGTGGCGGCGGATTGCTGCGATGCGGCGGCGTTTGTGAGCGATGTGTTTGCGGCGCC  GGCGCTGGTGGCGGCGGCGTTTTTTGAAGATAGCGCGGCGGCGGCGCTGTTTAGCGCGAGCGTGGCGTTTGCGGATGCGGCGAGCGCGCC  GGAAGATGCGAGCCATCATCATCATCATTAATAAAGCTT</p>
<i>dhy-sc-vut5</i>	<p>NdeI--ATG--SS3--His tag--Stop codon--HindIII</p> <p>CATATGCGGGTAAAATTAGCGGCGTGCTGAAAGATGGCACCGGCAAACCGGTGCAGAACTGCACCATTAGCTGAAAGCGCGCCGCAAC  AGCACCCCGTGGTGGTGAACACCGTGGGCGAGCGAAAACCCGGATGAAGCGGGCCGCTATAGCATGGATGTGGAATATGGCCAGTATAGC  GTGATTCTGCAGTGGATGGCTTTCCGCCGAGCCATGCGGGCACCATACCGTGTATGAAGATAGCCAGCCGGGCACCCTGAACGATTTTCT  GTGCGCGATGACCGAAGATGATGCGCGCCCGGAAGTGTGCGCCGCTGGAAGTGTGGAAGAAGTGGCGCGCAACGCGAGCGTGG  TGGCGCAGAGCACCGCGGATGCGAAAAAAGCGCGGGCGATGCGAGCGCGAGCGGCGCAGGTGGCGGCGCTGGTGACCGATGCGACC  GATAGCGCGCGCGGCGAGCACAGCGCGGGCCAGGCGGCGAGCAGCGCGCAGGAAGCGAGCAGCGGCGCGAAGCGGCGAGCGCGA  AAGCGACCGAAGCGGAAAAAAGCGCGGCGGCGGCGGAAAGCAGCAAAAACGCGGCGGCGACCGAGCGGGCGCGGCGAAAAACAGCGA  AACCAACGCGGCGGCGAGCCAGCAGAGCGCGGCGACCAGCGCGAGCACCGGCGGACCAAAGCGAGCGAAGCGGCGACCAGCGCGCGC  GATGCGGTGGCGAGCAAAGAAGCGGCGAAAAGCAGCGAAACCAACGCGAGCAGCAGCGGGCCGCGGCGGAGCAGCGCGACCGCGG  CGGAAAACAGCGCGCGCGGCGGAAAACAGCGAAACCAACGCGCGCAGCAGCGAAACCGCGGCGGAACGAGCGCGAGCGCGGCGGC  GGATGCGAAAACCGCGGCGGCGGGCAGCGGAGCACCGCGAGCACCAAAGCGACCGAAGCGGCGGGCAGCGCGGTGAGCGCGAGCCAG  AGCAAAAGCGCGGCGGAAGCGGCGGCGATTGCGCGGAAAACAGCGCGAAACGCGCGGAAGATATTGCGAGCGCGGTGGCGCTGGAAG  ATGCGGATACCACCGCAAAGGCATTGTGAGCTGAGCAGCGGACCAACAGCACCGAGCGAAACCCTGGCGGCGACCCCGAAAGCGGTGA  AAGTGGTGGATGAAACCAACCGCAAAGCGCATTGGACCGTGCGCCATCATCATCATCATTAATAAAGCTT</p>



*dhy-g-vut7*

NdeI--ATG--SS4--His tag--Stop codon--HindIII

CATATG CCGAGCCGCAGCGTGAAACGCTGGAGCGCGGTGCAGCCGGTGTCTGAACCGCCCGGGCCGCCTGCAGCATCTGAACCCGGGC  
TGCGATAACCAGCAGCGCCGCCCGCGTACC CGGAACTGCGCAACCTGCCGCATATTGCGAGCATGGCGTTTAACGAACCGCTGATGC  
TGGAACCGGCGTATGCGCGCGTGT TTTTTT GCGCGCTGGCGGGCCAGCTGGGCATTAGCAGCCTGACCGATGCGGTGAGCGGCGATAGCCT  
GACCGCGCAGGAAGCGCTGGCGACCCTGGCGCTGAGCGGCGATGATGATGGCCCGGCCAGGCGCGCAGCTATCAGGTGATGAACGGCAT  
TGCGGTGCTGCCGGTGAGCGGCACCCTGGTGAGCCGCACCCGCGCGCTGCAGCCGTATAGCGGCATGACCGGCTATAACGGCATTATTGCG  
CGCTGCAGCAGGCGGCGAGCGATCCGATGGTGGATGGCATTCTGCTGGATATGGATACCCGGGCGGCATGGTGGCGGGCGCGTTTGAT  
TGCGCGGATATTATTGCGCGCGTGC GCGATATTAACCGGTGTGGGCGCTGGCGAACGATATGAACTGCAGCGCGGGCCAGCTGCTGGCG  
AGCGCGGCGAGCCGCCGCTGGTGACCCAGACCGCGCGCACCGGCAGCATTGGCGTGATGATGGCGCATAGCAACTATGGCGCGGCGCTG  
GAAAAACAGGGCGTGAAATTACCCTGATTTATAGCGGCAGCCATAAAGTGGATGGCAACCCGTATAGCCATCTGCCGATGATGTGCGCG  
AAACCCTGCAGAGCCGCATGGATGCGACCCGCCAGATGTTTGC GCGAGAAAGTGAGCGCGTATACCGGCCTGAGCGTGCAAGTGGTGCTGG  
ATACCGAAGCGGCGGTGTATAGCGGCCAGGAAGCGATTGATGCGGGCCTGGCGGATGAACTGGTGAACAGCACCGATGCGATTACCGTGA  
TGCGCGATGCGCTGGATGCGCGCAAAAGCCGCCTGAGCGGCGGCCGCATGACCAAAGAAACCCAGAGCACACCCTGAGCGCGACCGCGA  
GCCAGGCGGATGTGACCGATGTGGTGCCGGCGACCGAAGGCGAAAACGCGAGCGCGGCGCAGCCGATGTGAACGCGCAGATTACCGCG  
GCGGTGGCGGCGGAAAACAGCCGCATTATGGGCATTCTGAACTGCGAAGAAGCGCATGGCCGCGAAGAACAGGCGCGCGTGCTGGCGGA  
AACCCCGGGCATGACCGTGAAAACCGCGCGCCGATTCTGGCGGCGGCGCCGAGAGCGCGCAGGCGCGCAGCGATAACCGCGCTGGATCG  
CCTGATGCAGGGCGCGCCGGCGCGCTGGCGGCGGGCAACCCGGCGAGCGATGCGGTGAACGATCTGCTGAACACCCCGGTGCATCATCA  
TCATCATCATTAATAAAGCTT

**Appendix B5: Sequences of all modified protein sequences for each ORF.**

Modified ORF	Protein sequence
<i>amo-vut1</i>	<p>MMSKSHPRWRLAKKILTWLFFIAVIVLLVVYAKKVDWEEVWKVIRDYNRVALLSAVGLVVVSYLIYGCYDLLARFYCGHKLAKRQVMLVSFICYAFN  LTLSTWVGGIGMRYRLYSRLGLPGSTITRIFSLSTTNWLGYILLAGIIFTAGVVLPDHWYVDQTTLRILGIGLLMIIAVYLVFCAFACAKHRHMTIKGQK  LVLPQWKFALQMLISSVNWVMGAIWLLGQSVNYFFVLGVLLVSSIAGVIVHIPAGIGVLEAVFIALLAGEHTSKGTIIAALLAYRVLYYFIPLLLAL  ICYLLESQAKKLRAKNEAAMHHHHHH..</p>
<i>aol-vut3</i>	<p>MEAISSARFAEFSARIAAASAALLLCDADTALPAASVAFVDAVDALPAAVAFASAAAEALRSAAVSDDLAFVSDVFAALAEFSAAVAEEAARPALDD  AFVSDDFEASFEATASRAEVAASDAFVAAVEADVAADCCDAAAFVSDVFAAPALVAAFFEDSAAAALFSASVAFADAASAPEDASHHHHHH..</p>
<i>dhy-sc-vut5</i>	<p>MAVKISGVLKDGTKPVQNTIQLKARRNSTTVVNTVGSNPDEAGRYSMDVEYQYQVILQVDGFPPSHAGTITVYEDSQPGTLNDFLCAMTE  DDARPEVLRRELMVEEVARNASVVAQSTADAKKSAGDASASAAQVAALVTDATDSARAASTSAGQAASSAQEASSGAEASAKATEAEKSAAA  AESSKNAATSAGAAKTSETNAAASQQAATSASTAATKASEAATSARDAVASKEAAKSETNASSAGRAASSATAAENSARAAKTSETNARSSE  TAAERSASAAADAKTAAAGSASTASTKATEAAGSAVSASQKSAAEAAAIRAENSAKRAEDIASAVALADADTTRKGIVQLSSATNSTSETLAATPKA  VKVVMDETNRKAHWTVRHHHHHH..</p>
<i>dhy-g-vut7</i>	<p>MPSRSVKRWSAVQPVLNRPPGRLQHLNPGCDNQRRRRVTAELRNLPHIASMAFNEPLMLEPAYARVFFCALAGQLGISSLTDAVSGDSLTAQE  ALATLALSGDDDGPQRQARSYQVMNGIAPVSGTLVSRTRALQPYSGMTGYNGIARLQQAASDPMVDGILLDMTPGGMVAGAFDCADIIARV  RDIKPVWALANDMNCSAGQLLASAASRRLVTQTARTGSIGVMMMAHSNYGAALEKQGVETLIYSGSHKVDGNPYSHLPDDVRETLSRMDATR  QMFAQKVSAYTGLSVQVLDTEAAVYSGQEIDAGLADELVNSTDAITVMRDALDARKSRLSGGRMTKETQSTTVSATASQADVTDVVPATEGE  NASAAQPDVNAQITAAVAENSRRIMGILNCEEAHGREQARVLAETPGMTVKTARRILAAAPQSAQARSDTALDRLMQGAPAPLAAGNPASDA  VNDLLNTPVHHHHHH..</p>

**Appendix C: Supplementary data for functional analysis of recombinant proteins**

**Appendix C1: Absorbance values for varying AOL-VUT3 protein concentration.**

Time(sec)	0.001 mg/ml		0.003 mg/ml		0.007 mg/ml		0.014 mg/ml	
	Read 1	Read 2	Read 1	Read 2	Read 1	Read 2	Read 1	Read 2
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.342	0.352	0.36	0.36	0.362	0.353	0.316	0.308
40	0.347	0.35	0.36	0.362	0.356	0.35	0.319	0.308
60	0.346	0.349	0.358	0.361	0.354	0.348	0.32	0.307
80	0.345	0.35	0.357	0.359	0.352	0.347	0.319	0.306
100	0.344	0.35	0.356	0.358	0.351	0.345	0.319	0.305
120	0.345	0.348	0.355	0.357	0.349	0.345	0.318	0.304
140	0.346	0.348	0.354	0.356	0.347	0.344	0.317	0.304
160	0.347	0.349	0.353	0.356	0.347	0.343	0.316	0.303
180	0.349	0.349	0.352	0.354	0.345	0.342	0.315	0.302
200	0.351	0.348	0.351	0.353	0.345	0.341	0.314	0.302
220	0.353	0.349	0.35	0.353	0.343	0.34	0.314	0.301
240	0.354	0.35	0.35	0.352	0.344	0.339	0.313	0.301
260	0.354	0.352	0.349	0.352	0.342	0.339	0.313	0.301
280	0.355	0.353	0.348	0.351	0.341	0.339	0.312	0.301
300	0.354	0.354	0.347	0.35	0.34	0.338	0.311	0.301

**Appendix C2: Absorbance values for varying DHY-SC-VUT5 protein concentration.**

Time(sec)	0.001 mg/ml		0.003 mg/ml		0.007 mg/ml		0.014 mg/ml	
	Read 1	Read 2	Read 1	Read 2	Read 1	Read 2	Read 1	Read 2
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.475	0.424	0.452	0.527	0.486	0.465	0.374	0.37
40	0.477	0.416	0.425	0.558	0.48	0.465	0.387	0.376
60	0.476	0.444	0.427	0.547	0.479	0.464	0.388	0.376
80	0.481	0.429	0.425	0.555	0.478	0.463	0.387	0.374
100	0.478	0.43	0.442	0.562	0.476	0.462	0.387	0.374
120	0.473	0.433	0.456	0.558	0.475	0.462	0.385	0.373
140	0.484	0.43	0.45	0.553	0.474	0.461	0.384	0.371
160	0.475	0.431	0.441	0.552	0.473	0.461	0.383	0.371
180	0.471	0.41	0.433	0.552	0.471	0.461	0.381	0.37
200	0.469	0.419	0.441	0.554	0.47	0.464	0.38	0.369
220	0.469	0.44	0.442	0.553	0.469	0.464	0.379	0.368
240	0.47	0.434	0.452	0.551	0.468	0.464	0.378	0.368
260	0.472	0.42	0.44	0.553	0.467	0.466	0.377	0.368
280	0.471	0.429	0.441	0.552	0.466	0.466	0.376	0.366
300	0.482	0.41	0.436	0.552	0.465	0.466	0.376	0.366

**Appendix C3: Absorbance values for varying DHY-G-VUT7 protein concentration.**

Time(sec)	0.001 mg/ml		0.003 mg/ml		0.007 mg/ml		0.014 mg/ml	
	Read 1	Read 2	Read 1	Read 2	Read 1	Read 2	Read 1	Read 2
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.366	0.35	0.53	0.333	0.483	0.369	0.374	0.355
40	0.36	0.354	0.531	0.332	0.494	0.373	0.371	0.353
60	0.363	0.351	0.526	0.332	0.495	0.371	0.369	0.352
80	0.362	0.348	0.522	0.331	0.492	0.369	0.367	0.355
100	0.36	0.348	0.523	0.329	0.492	0.365	0.365	0.35
120	0.357	0.348	0.524	0.328	0.488	0.362	0.365	0.347
140	0.354	0.347	0.531	0.329	0.481	0.362	0.363	0.346
160	0.357	0.347	0.527	0.328	0.485	0.36	0.361	0.343
180	0.357	0.347	0.529	0.327	0.483	0.356	0.36	0.343
200	0.356	0.351	0.524	0.325	0.482	0.356	0.359	0.342
220	0.352	0.354	0.519	0.324	0.482	0.356	0.358	0.341
240	0.354	0.355	0.524	0.325	0.482	0.354	0.357	0.346
260	0.352	0.357	0.515	0.323	0.482	0.353	0.356	0.339
280	0.351	0.357	0.523	0.323	0.479	0.353	0.356	0.339
300	0.35	0.357	0.513	0.322	0.476	0.352	0.355	0.343

**Appendix C4: Absorbance values for varying hexanol concentration using AOL-VUT3 protein at a wavelength of 340 nm.**

Time(sec)	25mM		50 mM		100 mM		150 mM		200 mM	
	Read1	Read2	Read1	Read2	Read1	Read2	Read1	Read2	Read1	Read2
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.374	0.384	0.402	0.354	0.39	0.364	0.41	0.4	0.404	0.345
40	0.382	0.385	0.399	0.361	0.389	0.362	0.409	0.402	0.408	0.351
60	0.383	0.382	0.404	0.362	0.39	0.362	0.41	0.404	0.409	0.354
80	0.383	0.378	0.405	0.361	0.39	0.362	0.41	0.403	0.408	0.356
100	0.381	0.377	0.404	0.36	0.39	0.363	0.41	0.402	0.408	0.359
120	0.381	0.377	0.404	0.36	0.391	0.36	0.409	0.4	0.408	0.359
140	0.381	0.378	0.404	0.36	0.393	0.361	0.408	0.4	0.407	0.358
160	0.381	0.377	0.405	0.36	0.395	0.362	0.407	0.399	0.407	0.357
180	0.381	0.377	0.405	0.361	0.396	0.363	0.406	0.397	0.406	0.356
200	0.382	0.376	0.404	0.362	0.397	0.365	0.405	0.397	0.406	0.356
220	0.382	0.376	0.404	0.362	0.398	0.364	0.404	0.396	0.405	0.357
240	0.382	0.375	0.404	0.363	0.398	0.365	0.404	0.395	0.405	0.357
260	0.383	0.375	0.403	0.364	0.398	0.364	0.403	0.395	0.405	0.358
280	0.383	0.375	0.403	0.364	0.397	0.365	0.401	0.393	0.405	0.358
300	0.384	0.375	0.404	0.365	0.397	0.367	0.4	0.393	0.405	0.357

**Appendix C5: Absorbance values for varying hexanal concentration using DHY-SC-VUT5 protein at a wavelength of 340 nm.**

Time(sec)	25mM		50 mM		100 mM		150 mM		200 mM	
	Read1	Read2	Read1	Read2	Read1	Read2	Read1	Read2	Read1	Read2
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.344	0.387	0.394	0.35	0.378	0.367	0.368	0.423	0.368	0.415
40	0.346	0.385	0.396	0.344	0.376	0.363	0.369	0.422	0.366	0.424
60	0.347	0.386	0.395	0.345	0.376	0.361	0.369	0.421	0.364	0.424
80	0.346	0.389	0.395	0.351	0.375	0.361	0.369	0.42	0.365	0.419
100	0.345	0.39	0.394	0.346	0.374	0.361	0.369	0.419	0.364	0.417
120	0.343	0.391	0.394	0.347	0.373	0.362	0.368	0.418	0.365	0.419
140	0.347	0.392	0.394	0.347	0.373	0.362	0.368	0.418	0.365	0.42
160	0.343	0.394	0.393	0.345	0.372	0.363	0.368	0.417	0.366	0.418
180	0.344	0.394	0.393	0.347	0.372	0.363	0.367	0.417	0.366	0.417
200	0.345	0.395	0.393	0.35	0.371	0.363	0.368	0.417	0.366	0.42
220	0.344	0.395	0.392	0.349	0.371	0.363	0.367	0.416	0.365	0.417
240	0.346	0.395	0.392	0.349	0.371	0.363	0.367	0.416	0.366	0.421
260	0.347	0.395	0.392	0.351	0.371	0.364	0.367	0.416	0.365	0.417
280	0.346	0.395	0.392	0.352	0.37	0.364	0.367	0.415	0.365	0.415
300	0.346	0.396	0.391	0.354	0.371	0.364	0.367	0.414	0.365	0.416

**Appendix C6: Absorbance values for varying hexanal concentration using DHY-G-VUT7 protein at a wavelength of 340 nm.**

Time(sec)	25mM		50 mM		100 mM		150 mM		200 mM	
	Read1	Read2	Read1	Read2	Read1	Read2	Read1	Read2	Read1	Read2
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.344	0.355	0.365	0.369	0.347	0.334	0.362	0.349	0.349	0.372
40	0.348	0.361	0.363	0.365	0.355	0.337	0.365	0.363	0.354	0.377
60	0.352	0.362	0.366	0.364	0.357	0.34	0.369	0.368	0.363	0.379
80	0.353	0.362	0.368	0.364	0.357	0.34	0.369	0.373	0.367	0.378
100	0.354	0.363	0.372	0.365	0.356	0.338	0.368	0.373	0.369	0.378
120	0.354	0.362	0.371	0.364	0.356	0.339	0.368	0.378	0.372	0.38
140	0.353	0.363	0.371	0.364	0.356	0.339	0.367	0.38	0.372	0.376
160	0.353	0.362	0.372	0.362	0.356	0.339	0.367	0.38	0.374	0.379
180	0.352	0.362	0.371	0.363	0.356	0.339	0.366	0.381	0.374	0.379
200	0.352	0.361	0.37	0.362	0.358	0.338	0.367	0.381	0.374	0.375
220	0.353	0.365	0.37	0.361	0.357	0.337	0.367	0.38	0.374	0.372
240	0.352	0.364	0.371	0.362	0.357	0.338	0.367	0.383	0.373	0.378
260	0.353	0.364	0.371	0.361	0.357	0.339	0.367	0.383	0.373	0.377
280	0.353	0.364	0.369	0.362	0.357	0.338	0.367	0.383	0.374	0.378
300	0.352	0.365	0.37	0.361	0.357	0.339	0.367	0.381	0.373	0.374



**Appendix C7: Absorbance values for cocktails AOL-VUT3/DHY-SC-VUT5 and AOL-VUT3/DHY-G-VUT7 at a wavelength of 340 nm.**

Time (sec)	Cocktail AOL-VUT3/DHY-SC-VUT5		Cocktail AOL-VUT3/DHY-G-VUT7	
	Read 1	Read 2	Read 1	Read 2
<b>0</b>	0.000	0.000	0.000	0.000
<b>20</b>	0.374	0.34	0.353	0.346
<b>40</b>	0.372	0.344	0.354	0.338
<b>60</b>	0.372	0.35	0.356	0.339
<b>80</b>	0.375	0.36	0.358	0.345
<b>100</b>	0.38	0.376	0.361	0.355
<b>120</b>	0.384	0.391	0.365	0.358
<b>140</b>	0.384	0.394	0.373	0.359
<b>160</b>	0.385	0.395	0.379	0.359
<b>180</b>	0.384	0.396	0.38	0.359
<b>200</b>	0.385	0.396	0.38	0.359
<b>220</b>	0.385	0.396	0.38	0.359
<b>240</b>	0.385	0.396	0.381	0.359
<b>260</b>	0.386	0.395	0.382	0.359
<b>280</b>	0.386	0.394	0.382	0.359
<b>300</b>	0.386	0.393	0.382	0.36

**Appendix C8: Graphs plotted using Graph Pad Prism to determine optimal enzyme and substrate concentration for AOL-VUT3, DHY-SC-VUT5 and DHY-G-VUT7 at a wavelength of 340 nm.**

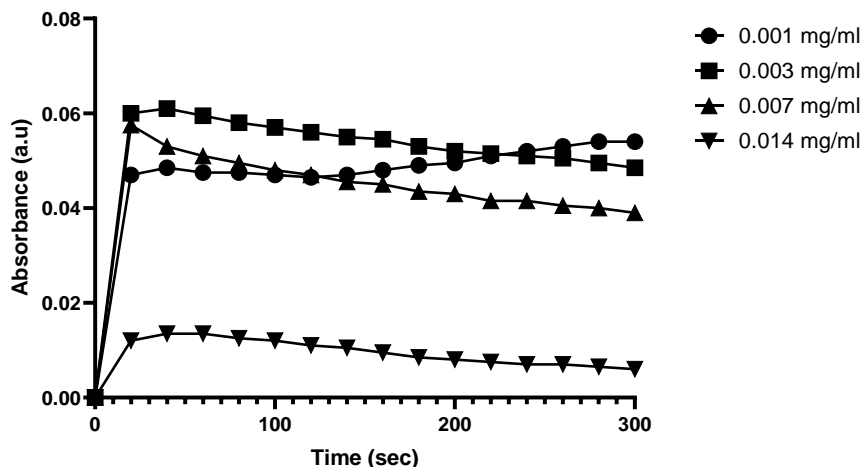


Figure C8.1: The effect of varying AOL-VUT3 enzyme concentrations 0.001 mg/ml, 0.003 mg/ml, 0.007 mg/ml and 0.014 mg/ml on NADH production via the reduction of NAD<sup>+</sup> by hexanol in the ADH assay. Data was analysed using GraphPad Prism Version 8.4.1 (460).

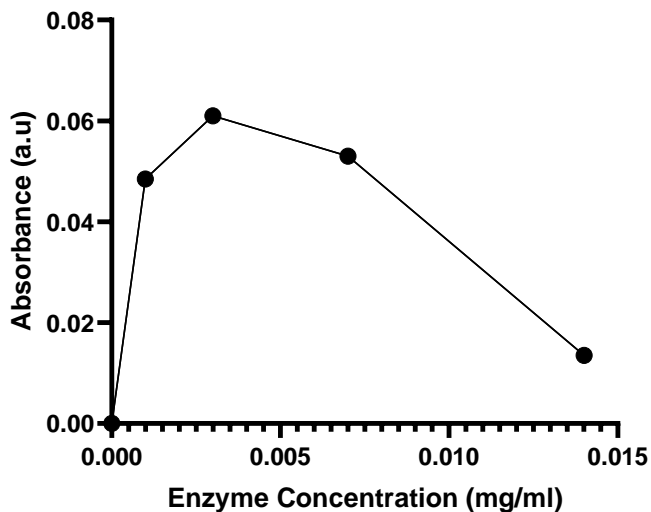


Figure C8.2: The effect of increasing AOL-VUT3 enzyme concentration on the reduction of NAD<sup>+</sup> to NADH at time interval 20 seconds using hexanol. Data was analysed using GraphPad Prism Version 8.4.1 (460).

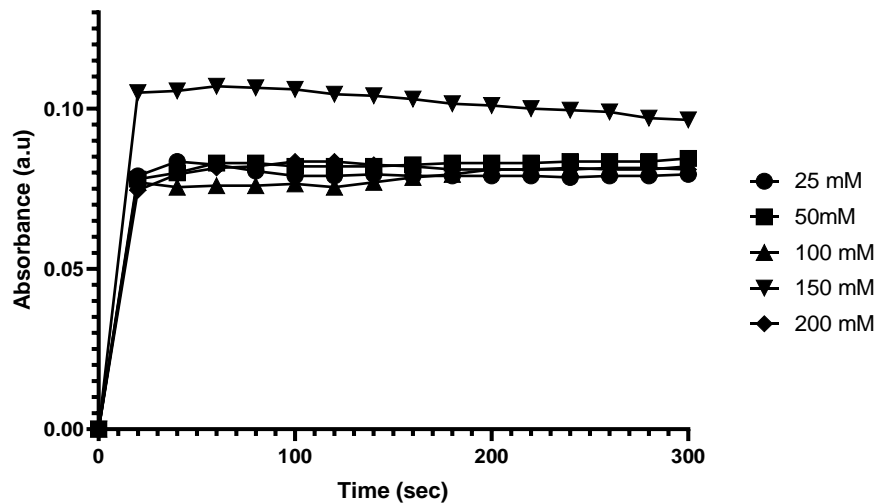


Figure C8.3: The effect of various hexanol concentrations 25 mM (A), 50 mM (B), 100 mM (C), 150 mM (D) and 200 mM (E) on NADH production via the reduction of  $\text{NAD}^+$  using AOL-VUT3 enzyme in the ADH assay. Data was analysed using GraphPad Prism Version 8.4.1 (460).

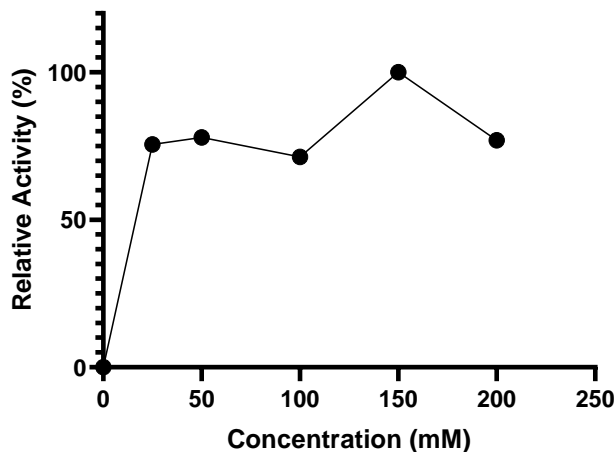


Figure C8.4: The effect of increasing hexanol concentration on the reduction of  $\text{NAD}^+$  to NADH at time interval 20 seconds using the AOL-VUT3 enzyme at a concentration of 0.003 mg/ml. Data was analysed using GraphPad Prism Version 8.4.1 (460).

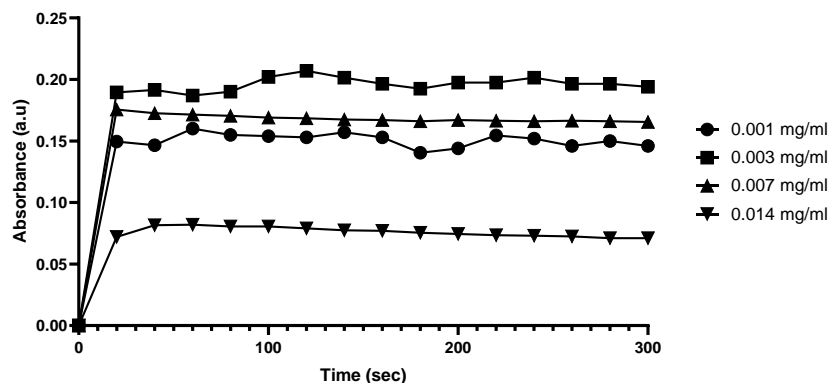


Figure C8.5: The effect of varying DHY-SC-VUT5 enzyme concentrations (0.001 mg/ml, 0.003 mg/ml, 0.007 mg/ml and 0.014 mg/ml) on NADH production via the reduction of  $\text{NAD}^+$  by hexanal in the ALDH assay. Data was analysed using GraphPad Prism Version 8.4.1 (460).

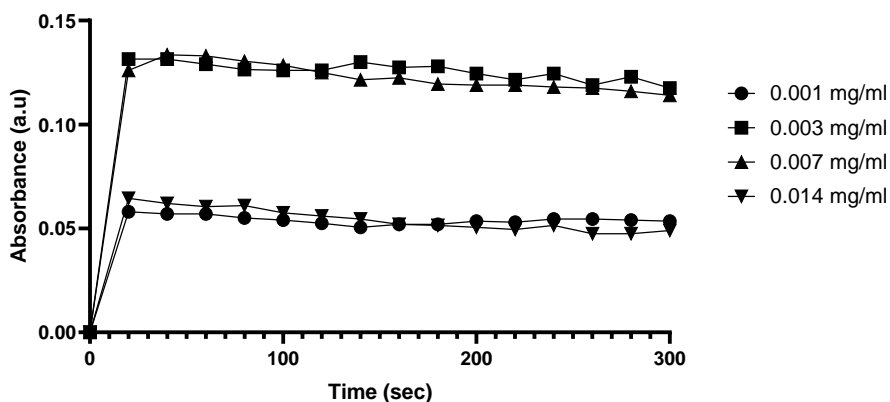


Figure C8.6: The effect of varying DHY-G-VUT7 enzyme concentrations (0.001 mg/ml, 0.003 mg/ml, 0.007 mg/ml and 0.014 mg/ml) on NADH production via the reduction of  $\text{NAD}^+$  by hexanal in the ALDH assay. Data was analysed using GraphPad Prism Version 8.4.1 (460).

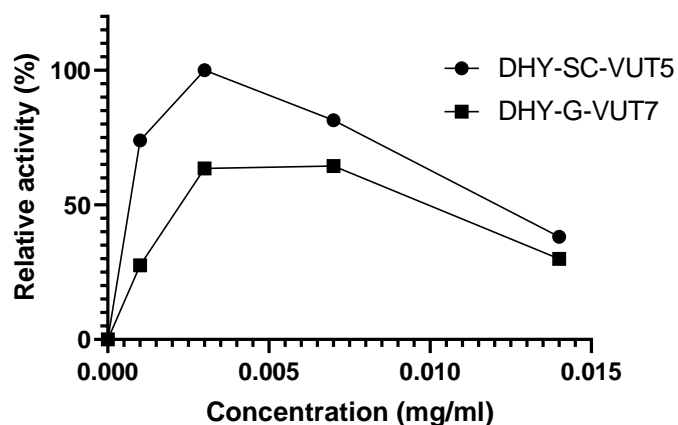


Figure C8.7: The effect of increasing DHY-SC-VUT5 and DHY-G-VUT7 enzyme concentration on the reduction of  $\text{NAD}^+$  to NADH at time interval 20 seconds using hexanal. Data was analysed using GraphPad Prism Version 8.4.1 (460).

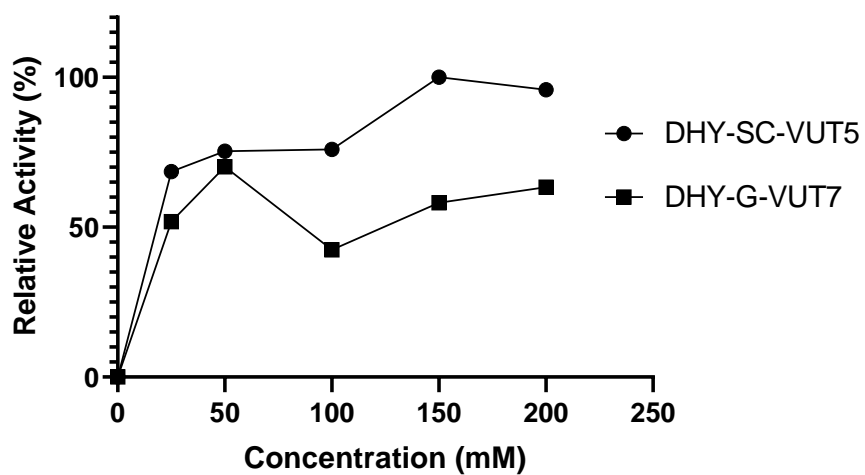


Figure C8.8: Relative activity of the DHY-SC-VUT5 and DHY-G-VUT7 enzyme on NADP production using varying concentrations of hexanal (25 mM, 50 mM, 100 mM, 150 mM and 200 mM) at 20 sec. time interval. Data was analysed using GraphPad Prism Version 8.4.1 (460).

**Appendix C9: Graphs for assessing activity of enzyme cocktails.**

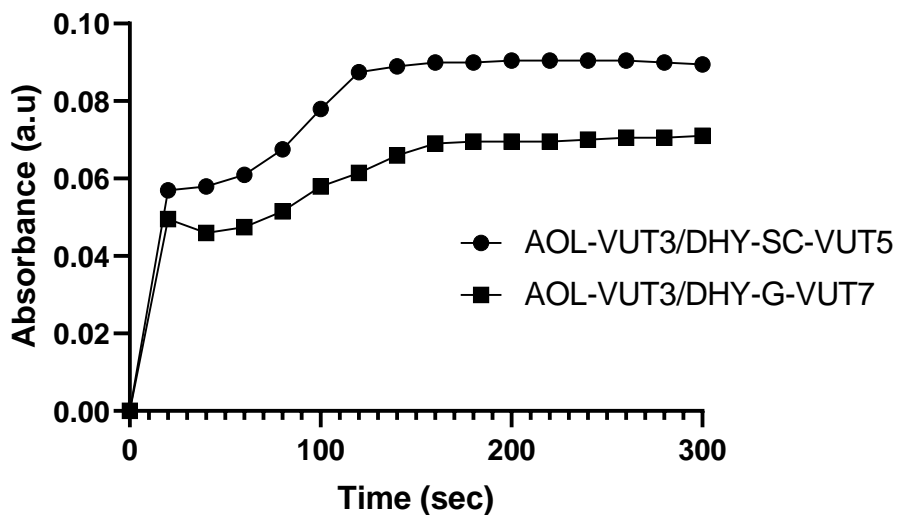


Figure C9.1: Comparison of enzymatic activity of the AOL-VUT3/DHY-SC-VUT5 and AOL-VUT3/DHY-G-VUT7 enzyme cocktails catalyzing the complete breakdown of hexanol substrate to the final product of carboxylic acid. Data was analysed using GraphPad Prism Version 8.4.1 (460).

## Appendix C10: Submitted articles for publication

**Biotechnology Letters**  
**Cloning, Expression, Purification and Kinetic Study of a Novel Alcohol Dehydrogenase (AOL-VUT3) Mined from Metagenome of Hydrocarbon Contaminated Soils**  
 –Manuscript Draft–

Manuscript Number:	
Full Title:	Cloning, Expression, Purification and Kinetic Study of a Novel Alcohol Dehydrogenase (AOL-VUT3) Mined from Metagenome of Hydrocarbon Contaminated Soils
Article Type:	Original Research Paper
Section/Category:	Microbial and Enzyme Technology
Keywords:	Hydrocarbons, hexanol, hexanal, alkanols, enzyme kinetics, alcohol dehydrogenase, fosmid clones, reduction, bioremediation.
Corresponding Author:	Naser Aliye Feto, Ph.D. Vaal University of Technology Vanderbijlpark-Johannesburg, Gauteng SOUTH AFRICA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Vaal University of Technology
Corresponding Author's Secondary Institution:	
First Author:	Cindy Baburam
First Author Secondary Information:	
Order of Authors:	Cindy Baburam Naser Aliye Feto, Ph.D.
Order of Authors Secondary Information:	
Funding Information:	South African Bio-Design Initiative (SABDI) (420/01 SABDI 16/1021.) Dr. Naser Aliye Feto
Abstract:	Microorganisms from hydrocarbon contaminated soils have adapted to harsh environments by utilizing specific enzymatic pathways to catabolise different classes of hydrocarbon. Over the years enzymes have become highly sought-after biomolecules for environmental, pharmaceutical and industrial and other applications. Alcohol dehydrogenase lies at the heart of aerobic hydrocarbon degradation since it is involved in the second step of catalysing the oxidation of alcohols to aldehydes. Fifteen candidate fosmid clones with hydrocarbon degrading potential were screened to identify alcohol dehydrogenase genes by comparison to sequences available in the Uniprot database. An open reading frame of 759 bp encoding for alcohol dehydrogenase (AOL-VUT3) with 24.7 % sequence similarity to histidinol dehydrogenase (HDH) from <i>Mycobacterium bovis</i> (strain ATCC BAA-935 / AF2122/97) (EC:1.1.1.23) was found and the low sequence similarity shows the novelty of the aol-vut3. The gene sequence was codon-optimized as per <i>Escherichia coli</i> codon preference and synthesized. The synthesised gene was successfully cloned into a pET-30a(+) expression vector and transformed into chemically competent <i>Escherichia coli</i> Rosetta™ (DE3) cells. The protein expression was induced using 1 mM Isopropyl-β-D-thiogalactoside (IPTG) and incubated at 15 °C for 16 hours. The protein was expressed, purified and its kinetic parameters studied. The enzyme successfully converted hexanol to a less complex hexanal. Accordingly, kinetics study showed a low Km value of 2.875 mM with a Vmax of 0.008614 μmol.min <sup>-1</sup> , showing the high affinity of the enzyme to its substrate (hexanol) thus, making it a valuable candidate for degradation of alkanol. Since, the discovered enzyme is novel it is important to conduct crystallographic study of such indispensable protein and solve its structure.

## BMC Biotechnology - Receipt of Manuscript 'Cloning, Expression, Purification...'

---

From: BMC Biotechnology (bmcbiotechnology@biomedcentral.com)

To: anaser22@yahoo.com

Date: Tuesday, July 21, 2020, 07:56 PM GMT+2

---

Ref: Submission ID 53bc2c7a-97b6-43a3-90d2-7ade4f00c197

Dear Dr Feto,

Thank you for submitting your manuscript to BMC Biotechnology.

Your manuscript is now at our initial Quality Check stage, where we look for adherence to the journal's submission guidelines, including any relevant editorial and publishing policies. If there are any points that need to be addressed prior to progressing we will send you a detailed email. Otherwise, your manuscript will proceed into peer review.

You can check on the status of your submission at any time by using the link below and logging in with the nature.com account you created for this submission:

[https://researcher.nature.com/your-submissions?  
utm\\_source=submissions&utm\\_medium=email&utm\\_campaign=confirmation-email&journal\\_id=12896](https://researcher.nature.com/your-submissions?utm_source=submissions&utm_medium=email&utm_campaign=confirmation-email&journal_id=12896)

To help with your article processing charge at acceptance, and to see if you're eligible for any waivers or BMC membership discounts, please click here:

<https://payment.springernature.com/quote/identify/submission/53bc2c7a-97b6-43a3-90d2-7ade4f00c197>

Kind regards,

Peer Review Advisors  
BMC Biotechnology

---

Springer Nature offers an open access support service to make it easier for our authors to discover and apply for APC funding. For further information please visit <http://www.springernature.com/gp/open-research/funding>

“Our flexible approach during the COVID-19 pandemic”

If you need more time at any stage of the peer-review process, please do let us know. While our systems will continue to remind you of the original timelines, we aim to be as flexible as possible during the current pandemic.



## Biodegradation

### Engineering of Enzyme Cocktails for Complete Transformation of the Toxic Hexanol into a Useful Product –Manuscript Draft–

<b>Manuscript Number:</b>	
<b>Full Title:</b>	Engineering of Enzyme Cocktails for Complete Transformation of the Toxic Hexanol into a Useful Product
<b>Article Type:</b>	Original research article
<b>Keywords:</b>	Microorganisms; alcohol dehydrogenase; aldehyde dehydrogenase; carboxylic acid; enzyme cocktails; hydrocarbons; bioremediation; hexanol; hexanal
<b>Corresponding Author:</b>	Naser Aliye Feto Vaal University of Technology SOUTH AFRICA
<b>Corresponding Author Secondary Information:</b>	
<b>Corresponding Author's Institution:</b>	Vaal University of Technology
<b>Corresponding Author's Secondary Institution:</b>	
<b>First Author:</b>	Cindy Baburam
<b>First Author Secondary Information:</b>	
<b>Order of Authors:</b>	Cindy Baburam Tsepo Tsekoa Naser Aliye Feto
<b>Order of Authors Secondary Information:</b>	
<b>Funding Information:</b>	frican Bio-Design Initiative (SABDI) (420/01 SABDI 16/1021)      Not applicable
<b>Abstract:</b>	Usually more than one enzyme is involved to achieve complete or near complete degradation of hydrocarbons or their transformation into a less toxic or useful compounds. Hexanol, a hydrocarbon which belongs to alkanol can cause serious health effect. Short-term exposure to hexanol causes irritation to the eyes, respiratory tract and skin. If this liquid is swallowed, aspiration into the lungs may result in chemical pneumonitis, which is inflammation of the lungs or breathing difficulty that could be acute or chronic. Thus, here we report engineering of two enzyme cocktails based on combination of novel enzymes, namely an alcohol dehydrogenase AOL-VUT3 with each of DHY-SC-VUT5 and DHY-G-VUT7 dehydrogenases for complete transformation of hexanol. The activity of the cocktails was indirectly assessed by measuring the amount NADH released as a result of reduction of the co-factor NAD <sup>+</sup> following action of the pair of enzymes on hexanol and converting it to carboxylic acid. As a result, AOL-VUT3/DHY-SC-VUT5 and AOL-VUT3/DHY-G-VUT7 cocktails were engineered. Accordingly, AOL-VUT3/DHY-SC-VUT5 cocktail exhibited the highest activity relative to AOL-VUT3/DHY-G-VUT7. Therefore, the successful enzymatic transformation of the toxic hexanol into carboxylic acid is a major breakthrough since, carboxylic acids and their derivatives are used in the production of polymers, biopolymers, coatings, adhesives, and pharmaceutical drugs. They can also be used as solvents, food additives, antimicrobials, and flavorings. Thus, the developed enzyme cocktails provide two-fold advantage as they remove the toxic hexanol and produce the very useful carboxylic acid. Hence, future study should focus on immobilization and stabilization of the engineered enzyme cocktails for large scale bioremediation of alkanols and an industrial scale production of carboxylic acid from hexanol.
<b>Suggested Reviewers:</b>	Jeremy Burton, PhD Lawson Health Research Institute, University of Western Ontario