Mining a South African deep mine metagenome for the discovery of novel biocatalysts

by

Nathlee Samantha Abbai

Submitted in accordance with the requirements for the degree Philosophiae Doctor

in the
Department of Microbial, Biochemical and Food Biotechnology
Faculty of Natural and Agricultural Sciences
University of the Free State
Bloemfontein
South Africa

Promoter: **Prof. D. Litthauer**
Co-promoters: **Prof. E. van Heerden**
Dr. L.A. Piater

May 2009

Declaration

I declare that this thesis hereby submitted by me for the Doctor of Philosophy degree at the University of the Free State is my own independent work and has not previously been submitted by me at another university or faculty. I further cede copyright of the thesis in favour of the University of the Free State

---

Nathlee Samantha Abbai

Acknowledgements

The author wishes to thank the following people and organizations.

My Lord, Jesus Christ who is my source of strength.

This study was supported by the NRF (National Research Foundation, South Africa), the Ernst, Ethel Erickson Trust and the Metagenomics Platform.

I would like to express my appreciation to my promoter, Prof. D. Litthauer, for his support for the duration of this study.

I also thank Dr. LA Piater, and Prof. E van Heerden, for their input.

Inqaba Biotechnology for the 454 sequencing data, Prof J. Albertyn and Molecular Biology group (UOVS) for the Sanger sequencing data.

A special thanks to Prof Rolf Daniel and the students at the Department of Microbiology, University of Göttingen, for their assistance, especially the use of lab facilities and for making my stay in Germany a pleasant one.

Thank you to Dr. C. Hugo, Prof M. Smit, and the students at the Biotransformation group (UOVS) for their assistance and providing some of the material used in this study.

To my amazing parents and my brother, thank you very much for all your support, words cannot express my gratitude.

To all my friends and colleagues at the Extreme Biochemistry Group who have become very special to me. Thank you.

Thank you to Kamini, Landi, Godfrey and Walter for always lending a helping hand and for the value that you have added to my life and the success of this project.

Chapter 1
Review of Literature

1.1 Introduction 1

1.1.1 DNA extraction from environmental samples 4

1.2.1 DNA extraction from Soil 4

1.2.1.1 Chemical Lysis **6**

1.2.1.2 Enzymatic Lysis **6**

1.2.1.3 Physical Lysis **6**

1.2.2 DNA purification 7

1.3 DNA extraction from Water 7

1.4 DNA extraction from Biofilms 8

1.5 Metagenomic Library Construction 9

1.5.1. Small-insert library construction 11

1.5.2 Large-insert library construction 13

1.5.2.1 Phage vectors **13**

1.5.2.2 Cosmid vectors **14**

1.5.2.3 Fosmid vectors **15**

1.5.2.4 Bacterial Artificial Chromosomes (BAC) vectors **17**

1.6 Heterologous gene expression 19

1.7. Screening of Metagenomic Libraries 20

1.7.1 Sequence-based screening 20

1.7.2 Function-based screening 22

1.8 Scope of the present study 24

Chapter 2

Construction of genomic libraries with a South African deep mine isolate and screening for the presence of lipases

2.1. Introduction 26

2.2 Materials and Methods 28

2.2.1 Genomic DNA extraction 28

2.2.2 Confirmation of culture identity 29

2.2.3 Small-insert library construction 29

2.2.3.1 Preparation of genomic DNA for small insert library construction 30

2.2.3.1.1. Partial digestion of GE-7 genomic DNA 30

2.2.3.1.2 Physical fragmentation of GE-7 genomic DNA 31

2.2.3.1.3 End repair of fragmented DNA 31

2.2.3.1.4 Ligation reactions 31

2.2.3.1.5 Bacterial transformation 32

2.2.3.1.6 Plasmid DNA extraction and restriction analysis 32

2.2.4 Large insert library construction 32

2.2.4.1 Preparation of genomic DNA for large-insert library construction 33

2.2.4.1.1 Partial digestion of GE-7 gDNA 33

2.2.4.1.2 Ligation into pCC1FOS fosmid vector, packaging and tittering 33

2.2.4.1.3 Fosmid DNA extractions and restriction analysis 34

2.2.5 Library screening

2.2.5.1 Confirmation of lipase activity 35

2.2.6 Ultrafast sequencing of GE-7 fosmid and plasmid clones using the Genome sequencer 20 FLX system 36

2.2.6.1 DNA sequencing with the ABI 3730XL Automated Sequencer 38

2.2.6.2 Gap closure using primer walking 40

2.2.6.3 ORF predictions 42

2.3. Results and Discussion 43

2.3.1 Genomic DNA extraction 43

2.3.2 Culture confirmation 43

2.3.3 Small-insert library construction 45

2.3.3.1 Partial digestion of GE-7 genomic DNA 45

2.3.3.2 Physical fragmentation of GE-7 genomic DNA 46

2.3.3.3 Sticky-ended library 47

2.3.3.4 Blunt-ended library 48

2.3.4 Library screening (Small-inserts) 50

2.3.4.1 Confirmation of lipase activity 50

2.3.5 Large-insert library construction 51

2.3.5.1 Digestion of GE-7 gDNA 51

2.3.5.2 Fosmid library production 52

2.3.5.3 Screening of Fosmid library 55

2.4 Ultrafast sequencing of GE-7 fosmid and plasmid clones using the Genome sequencer 20 FLX system 56

2.5 Conclusion 59

Chapter 3

What lies beneath? Assessment of diversity of a Beatrix mine biofilm

3.1 Introduction **64**

3.2 Materials and Methods **65**

3.2.1 Environmental DNA extraction **65**

3.2.1.1 Method 1 **65**

3.2.1.2 Method 2 **67**

3.2.1.3 Method 3 **67**

3.2.2 Diversity studies **68**

3.2.2.1 PCR conditions **68**

3.2.2.2 Denaturing Gradient Gel electrophoresis (DGGE) analysis **68**

3.2.3 Cloning and sequencing of the 16S bacterial and archaeal rRNA and the 18S eukaryotic rRNA genes **69**

3.2.3.1 Library construction **69**

3.2.3.2 Plasmid DNA extraction and restriction analysis **70**

3.2.3.3 Sequencing of randomly selected clones **70**

3. 3 Results and Discussion **70**

3.3.1 Environmental DNA extraction **70**

3.3.2 Diversity studies **72**

3.3.3 Cloning and sequencing of the 16S bacterial and archaeal rRNA and 18S rRNA eukaryotic rRNA **77**

3.3.4 Phylogenetic analysis of selected bacterial clones **78**

3.4. Conclusion **86**

Chapter 4

Sequenced-based screening for cytochrome P450 monoxygenases by a PCR based approach

4.1 Introduction 87

4.2 Materials and Methods 89

4.2.1 PCR screening for the presence of cytochrome P450 alkane hydroxylases from the metagenome 89

4.2.2. Cloning of the CYP153 fragment 90

4.2.2.1 Plasmid DNA extraction and restriction digestion 90

4.2.2.2 Sequencing of randomly selected clones 90

4.3 Results and Discussion 91

4.3.1 Sequenced-based screening for the presence cytochrome P450 alkane hydroxylases in the metagenome 91

4.3.2 Cloning and sequencing of the CYP 153 fragments 92

4.4 Conclusion 96

Chapter 5

Mining the metagenome for biocatalysts

5.1 Introduction 98

5.2 Materials and Methods 99

5.2.1 Small-insert library construction 99

5.2.1.1 Ligation 99

5.2.1.2 Bacterial transformation 99

5.2.1.3 Plasmid DNA extraction and restriction analysis 99

5.2.1.4 Library screening using the function-driven approach 99

5.2.1.4.1 Plate assay for lipolytic enzymes 100

5.2.1.4.2 Protease plate assay 101

5.2.1.4.3 Amylase plate assay 101

5.2.1.4.4 Beta-lactamase plate assay 102

5.3. DNA sequencing and analysis 102

5.4 Heterologous expression of the patatin in *E.coli* 103

5.4.1 Bacterial strains, plasmids and growth conditions 103

5.4.2 Construction of expression plasmids 103

5.4.2.1 PCR amplification of patatin 103

5.4.2.2 Constructs for expression in E. coli 104

5.4.2.3 Expression and purification of the patatin 105

5.4.2.4 Sodium dodecyl sulphate polyacrylamide gel electrophoresis 105

5.4.2.5 Protein concentration determination 106

5.4.2.6 Biochemical analysis 106

5.4.2.6.1 Effects of temperature on enzyme activity 107

5.4.2.6.2 Effects of pH on enzyme activity 107

5.5 Large-insert library construction 107

5.5.1 Ligation into pCC1FOS fosmid vector and phage packaging 107

5.5.2 Fosmid DNA extractions and restriction analysis 108
5.5.3 Fosmid library screening 108
***5.5.3.1 Antibacterial assays* 108**
***5.5.3.2 Detection of antibiotic resistance* 108**

5.6 Results and Discussion 110
5.6.1 Small-insert library construction 110
***5.6.1.1 Partial digestion of environmental DNA* 110**
***5.6.1.2 Small-insert library* 110**
***5.6.1.3 Library screening using the function-driven approach* 111**
***5.6.1.4 Sequencing reactions* 112**

5.7 Constructs for expression in *E. coli* 123
5.7.1 Expression and purification of the patatin 125
5.7.2 Biochemical analysis of the expressed patatin 129
***5.7.2.1 Effects of temperature and pH on enzyme activity* 129**

5.8 Large-insert library construction 130
5.8.1 Fosmid Library Screening 131

Conclusion 132

Chapter 6

Summary

Summary/Opsomming 134
References 139

List of Figures

Figure 1.1: Construction and screening of metagenomic libraries. Schematic representation of construction of libraries from environmental samples. The images at the top from left to right show bacterial mats at Yellowstone, soil from a boreal forest in Alaska, cabbage white butterfly larvae, and a tubeworm (taken from Handelsman, 2004). **3**

Figure 1.2: Process of fosmid cloning (Epicentre Biotechnologies). **17**

Figure 2.1: Map of pZerO-2 plasmid vector (Invitrogen). **30**

Figure 2.2: Map of pCC1FOS and pCC1BAC (Epicentre Biotechnologies). **34**

Figure 2.3: Schematic illustration of the Pyrosequencing method. Pyrosequencing is a non-electrophoretic real-time DNA sequencing method that uses the luciferase-luciferin light release as the detection signal for nucleotide incorporation into target DNA. The four different nucleotides are dispensed iteratively to a four-enzyme mixture. The pyrophosphate (PPi) released in the DNA polymerase-catalyzed reaction is quantitatively converted to ATP by ATP sulfurylase, which provides the energy to firefly luciferase to oxidize luciferin and generate light *(ho).* The light is detected by a photon detection device and monitored in real time by integrated software in a format called program. Finally, apyrase catalyzes degradation of nucleotides that are not incorporated and the sequencing reaction will be ready for the next nucleotide addition (taken from Gharizadeh *et al.,*2007). **38**

Figure 2.4: Template display of contig00098. The black arrows represent the vector sequences; red arrows block (1) represent Sanger sequencing from the vector. Green and orange arrows represent individual 454 reads in their respective orientation. The yellow block (2) represents the primer designed using the parameter of the Gap4 program, block (3) represents primer walking using the specific primer designed from that region. **41**

Figure 2.5: Assembled reads making up contig00098. The sequence highlighted in yellow represents the primer that was designed using Staden. **42**

Figure 2.6: Diagramatic representation of ORFs predicted by Artemis in all six reading frames. **43**

Figure 2.7: PCR amplification of the 16S rRNA gene from GE-7. M: MassRuler DNA Ladder (SM#0403); lane 1: positive control (*E. coli*); lane 2: negative control (sterile water); and lane 3: GE -7. **44**

Figure 2.8: Restriction analysis of the 16S rRNA gene from GE-7. M: MassRuler DNA Ladder (SM#0403); lane 1: *Eco*RI digestion; lane 2: *Sma*I digestion; and lane 3: *Pst*I digestion. **45**

Figure 2.9: Partial digestion of GE-7 genomic DNA. Lane1:*Xho*I digestion; M: MassRuler DNA Ladder (SM #0403). **46**

Figure 2.10: Fragmentation by nebulization. Lane 1: fragmented gDNA; M: MassRuler DNA Ladder (SM#0403). **47**

Figure 2.11: Restriction analysis of clones from sticky library. M: MassRuler DNA Ladder (SM #0403); lanes 1-21: randomly selected clones. **48**

Figure 2.12: Restriction analysis of clones from blunt library. M: MassRuler DNA Ladder (SM # 0403); lanes 1-21: randomly selected clones. **49**

Figure 2.13: The 9 positive clones obtained when screened on LB tributyrin plates supplemented with Kanamycin (50 µg/ml). S = clones from sticky library and B = clones from blunt library. **50**

Figure 2.14: PCR amplification of the lipase gene from clones positive by functional screening. M- MassRuler DNA Ladder (SM#0403); lane 1: positive control (*Geobacillus kaustophilus* gDNA); lane 2: negative control (sterile water); lane 3: clone1; lane 4: clone 2; lane 5: clone 3; and lane 6: clone 4. **51**

Figure 2.15: Partial digestion of gDNA with *Eco*RV. Lane 1: GE7 gDNA; and lane 2: fosmid control standard (36 kb). **52**

Figure 2.16: Restriction digests of fosmid clones. Lane 1: MassRuler 1kb ladder; lanes 2-11: fosmid clones: and lane 12: Marker III (Lambda DNA digested with *Hin*dIII. **53**

Figure 2.17: Restriction digests of electroporated clones.
Lane 1: MassRuler 1kb ladder; lanes 2-21: fosmid clones: and lane 22: Marker III (Lambda DNA digested with *Hin*dIII). **54**

Figure 2.18: The eight positive clones obtained when screened on LB tributyrin plates supplemented with chloramphenicol (12.5 µg/ml). **55**

Figure 2.19: (a) PCR amplification of the lipase gene from clones found positive by functional screening. M- MassRuler DNA Ladder (SM#0403); lane 1: positive control (*Geobacillus kaustophilus* gDNA); lane 2: negative control (sterile water); and lane 3: positive clone. **56**

Figure 2.19: (b) A lipase positive clone obtained from the infection library when screened on LB olive oil containing rhodamine B and supplemented with chloramphenicol (12.5 µg/ml). A clone showing no lipase activity was used as the control. **56**

Figure 2.20: Linear representation of ORFs present in plasmid and fosmid clones. **61**

Figure 3.1: Genomic DNA extractions from the Beatrix Mine biofilm. M: MassRuler DNA 1kb Ladder (SM#0403-Fermentas); lane 1: DNA extracted with chemical and enzymatic lysis (method1); lane 2: DNA extracted after pretreatment with aluminium sulfate (method 2); and lane 3: DNA extracted with the FastDNA soil Kit (method 3). **71**

Figure 3.2: Domain specific PCR. A: Bacterial PCR; B: Archaeal PCR; and C: Eukaryotic PCR. M: MassRuler 1 kb ladder (SM#0403-Fermentas); lane 1: pos control; lane 2: neg control: and lane 3: biofilm sample (Beatrix Mine). **73**

Figure 3.3: DGGE analysis of the archaeal and bacterial population present in the biofilm. **74**

Figure 3.4: Restriction analysis of selected 16S bacterial clones. M: MassRuler DNA Ladder

(SM #0403); lanes: 1-10 -randomly selected clones. **75**

Figure 3.5: Phylogenetic tree representing bacterial 16S rRNA gene sequences constructed using the ARB program. Maximum parsimony, maximum likelihood and neighbour joining analysis produced highly similar tree topologies. The scale represents a 10% sequence divergence. **76**

Figure 3.6: Graphical representation of the phylogenetic distribution of the bacterial clonal library based on 16S rRNA gene sequences from the Beatrix gold mine. **77**

Figure 3.7: Restriction analysis of selected 18S eukaryotic and 16S archaeal clones. M: MassRuler DNA Ladder (SM#0403); lanes 1-10: randomly selected eukaryotic clones and lanes 11-20: randomly selected archeal clones. **81**

Figure 4.1: Schematic representation of different cytochrome P450 systems. (A): class I, bacterial system; (B): class I, mitochondrial system; and (C): class II microsomal system (taken from Hannemann *et al.,* 2007). **82**

Figure 4.2: PCR amplification of the smaller fragment of the CYP153 gene. M: MassRuler DNA Ladder (SM#0403); lane 1: positive control (*Pseudomonas putida)*; lane 2: negative control; and lane 3: metagenomic DNA. **83**

Figure 4.3: PCR amplification of the larger fragment of the CYP153 gene. M: MassRuler DNA Ladder (SM #0403); lane 1: positive control (*Pseudomonas putida*); lane 2: negative control; and lane 3: metagenomic DNA. **88**

Figure 4.4: Restriction analysis of the cloned 800 bp product of the CYP153 gene. M: MassRuler DNA Ladder (SM # 0403); lanes 1-20: randomly selected clones. **91**

Figure 4.5: Phylogenetic tree of CYP153 cytochrome P450 family homologues to P450 clones from the Beatrix mine. The phylogenetic tree was constructed by the neighbour-joining method with MEGA version 4.1 software. The accession numbers of the aligned sequences are shown in parenthesis. The numbers associated with the branches refer to the bootstrap values (confidence limits). **92**

Figure 5.1: Standard curve for the BCA protein assay kit (Pierce) at 37°C (enhanced method) using BSA as protein standard. Error bars indicate standard deviation after performing the experiment in triplicate. **93**

Figure 5.2: Partial digestion of biofilm gDNA. M:MassRuler DNA Ladder (SM#0403-Fermentas) and lane 1: -*Bam*HI digestion (Fermentas). **109**

Figure 5.3: Restriction analysis of clones from the Beatrix library. M: MassRuler DNA Ladder (SM # 0403- Fermentas); lanes: 1-20 -randomly selected clones. **110**

Figure 5.4: A clone showing activity on LB tributyrin as indicated by the red arrow. **111**

Figure 5.5: Multiple alignments of selected bacterial isochorismatase proteins using ClustalW (*P. entomophila* YP_6065771; *P. putida* YP_001747104; *Ochrobactrum* YP_001372521; *P. fluorescens* YP_0028749491; *N.europea* NP_842301; and *P. aeruginosa* YP_0020814351). The residues forming the catalytic site are highlighted in green. **110**

Figure 5.6: Multiple sequence alignments using ClustalW of selected bacterial sulfatases (*Verminephrobacter* YP_997301; *Parvibaculum* YP_ 001414401; and *Sinorhizobium* YP_001327736) containing the conserved catalytic motif (C-X-P-X-R). Only conserved amino acids are present in the alignment. **114**

Figure 5.7: ClustalW multiple alignments of selected bacterial phospholipase, patatin family proteins (*Roseiflexus* YP_001276494; *Chloroflexus* YP_001636863; *Herpetosiphon* YP_001543920; *Salinibacter* YP_4467821; and *Aeromonas* YP_ 001141850) containing the conserved esterase/lipase domains. **116**

Figure 5.8: Phylogenetic tree of phospholipase, patatin family proteins and different classes of lipolytic enzyme families. The phylogenetic tree was constructed by the neighbour-joining method with MEGA version 4.1software. The accession numbers of the aligned sequences are shown in parenthesis. The numbers associated with the branches refer to the bootstrap values (confidence limits). The scale represents a 20% sequence divergence. **119**

Figure 5.9: PCR amplification of the patatin *ORF*. M: MassRuler DNA Ladder (SM#0403-Fermentas) and lane 1: negative control (sterile water); and lane 2: plasmid DNA of clone pNS6. **121**

Figure 5.10: Vector map of pET-28b(+) indicating the kanamycin resistance gene, ColE1 origin of plasmid replication, *lacI* coding sequence and the multiple cloning site under the T7 promoter. Sequence of the pET-28b(+) cloning region showing the ribosome binding site and configuration for the N-terminal His-Tag and thrombin cleavage site. **122**

Figure 5.11: Restriction analysis of selected pET28b(+) clone. M- MassRuler 1kb ladder (SM #0403-Fermentas): and lane 2: positive clone. **123**

Figure 5.12: Expression of the patatin in *E. coli.* Lane 1: Molecular marker (BioRad), lane 2: uninduced control, lane 3: expressed patatin after 4 hrs induction at 30°C with 0.5 mM IPTG. **124**

Figure 5.13: Purification of the expressed patatin through Ni-affinity chromatography. **125**

Figure 5.14: Partially purified patatin. Lane 1: Molecular marker (BioRad), lane 2: uninduced control, lane 3: expressed patatin after 4 hrs induction at 30°C with 0.5 mM IPTG, lane 4: partially purified protein. **126**

Figure 5.15: Western blot analysis of the partially purified patatin. **126**

Figure 5.16: Substrate specificity for *p*-nitrophenyl esters with varying carbon lengths. Error bars indicate standard deviations after performing the experiment in triplicate. **127**

Figure 5.17: Temperature profile of the patatin. Error bars indicate standard deviations after performing the experiment in triplicate. **128**

Figure 5.18: pH profile of the patatin. Error bars indicate standard deviations after performing the experiment in triplicate. **129**

Figure 5.19: Restriction analysis of selected fosmid clones. M1-Marker II (Lambda DNA digested with *Hin*dIII), lanes 1-16: randomly selected fosmid clones, M2-MassRuler 1kb ladder (SM #0403-Fermentas). **130**

List of Tables

Table 1.1: Pros and cons of small-insert and large-insert libraries (Daniel, 2005) **11**

Table 1.2: Biocatalysts and bioactive compounds isolated from metagenomic libraries (Sharma *et al.,* 2005) **19**

Table 2.1: 16S rRNA primer sequences (Lane, 1991) **29**

Table 2.2: LipA specific primers (Barnard, 2005) **36**

Table 2.3: ABI-Plasmid-Cycle program **39**

Table 2.4: ABI-Fosmid-Cycle program **40**

Table 2.5: Assembly results using the Newbler software **57**

Table 2.6: Lipolytic activity conferring plasmids and fosmids and sequence similarities **62**

Table 3.1 PCR primers used in this study **69**

Table 3.2: DNA concentration and purity readings **72**

Table 3.3: Sequence data of selected 16S rRNA bacterial clones **79**

Table 3.4: Sequence data of selected archaeal 16S rRNA clones **84**

Table 3.5: Sequence data of selected eukaryote 18S rRNA clones **85**

Table 4.1: Primers used for the amplication of the CYP153 gene (van Beilen *et al.,* 2006)

90

Table 4.2: Sequence data of CYP153 clones **94**

Table 5.1: Bacterial strains and plasmids used in this study **103**

Table 5.2: Primer set used for expression cloning **104**

Table 5.3 Lipolytic activity conferring plasmids and sequence similarities **112**

Abbreviations

ATP Adenosine triphosphate
BCA Bicinchoninic acid
BLAST Basic Logical Alignment Search Tool
bp base pairs
BSA Bovine serum albumin
°C Degrees Celsius
DGGE Denaturing Gradient Gel Electrophoresis
DNA Deoxyribonucleic acid
dNTPs Deoxyribonucleoside triphosphates
EDTA Ethylenediaminetetraacetate
e.g for example
FAD Flavin adenine dinucleotide
FMN Riboflavin 5’-monophosphate
Ga One billion years
gDNA Genomic DNA
HEPES 4-(2-Hydroxyethyl)piperazine-1- ethanesulfonic acid sodium salt
i.e that is
IPTG Isopropyl -D-thiogalactoside
KB Kilo bases
kDa Kilo Daltons
LB Luria-Bertani broth
ml Milliltres
MOPS 3-(N-morpholino) propanesulfonic acid
NaCl Sodium chloride
NADH Nicotinamide adenine dinucleotide (reduced)
NADPH Nicotinamide adenine dinucleotide phosphate (reduced)
ng Nanogram
nm Nanometer
OD Optical density
ORF Open reading frame
OTU Operational taxonomic unit

PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PFGE	Pulsed Field Gel Electrophoresis
psi	Pounds per square inch
rRNA	Ribosomal Ribonucleic acid
rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
TAE	Tris, Acetic acid, EDTA
TE	Tris, EDTA
TLB	LB tributyrin
U	Units
µg	Microgram
µl	Microlitres
X-Gal	5-bromo-4-chloro-3-indolyl -D-galactoside

Abstract

The construction and screening of gene libraries prepared from DNA directly isolated from environmental samples is a recent and powerful tool for the discovery of new enzymes of biotechnological interest (Gabor *et al.,* 2004). Standard methods based on the screening of isolated microorganisms are inherently limited to the tiny fraction of cultivable microbial species (<1%); environmental gene banks in principle provide access to the entire sequence space present in nature (Handelsman *et al.,* 1998). Environmental libraries allow the screening of functional classes of genes from thousands of organisms and research in this area will provide an essential backdrop for understanding evolution and biochemical pathways (Rajendran *et al.,* 2008). The metagenomics approach has been shown to be an efficient method for obtaining novel biocatalysts and useful genes from uncultured microorganisms from diverse environments.

Before proceeding to the metagenome analysis, we constructed genomic libraries from a South African deep mine isolate *Geobacillus thermoleovorans* GE-7. The library was screened for lipolytic activity on LB tributyrin (TLB). Active clones were sequenced using 454 technologies, and the sequencing results revealed the presence of the lipA and GDSL lipases, of which the latter has not yet been characterized in this organism. In addition genes associated with fatty acid degradation, different glycolytic activities, lipolytic activity, spore germination; proper protein folding, antibiotic resistance and the cell wall were also identified in the active clones. Some of the genes identified may also aid in understanding how this organism had adapted to the environment from which it was isolated from.

Biofilm collected from the Beatrix gold mine was selected for the metagenomic studies. We performed a diversity assessment of the biofilm by cloning and sequencing of the 16S (bacterial and archaeal) and 18S eukaryotic ribosomal RNA. A further phylogenetic study was performed on the 16S rRNA clonal library. Based on the phylogenetic analysis, we decided to screen the metagenome using the sequenced-based approach for cytochrome P450 monooxygenases in particular the CYP153 family, a family of terminal hydroxylases and long chain alkane degraders. Cloning and sequencing of the CYP153 PCR products, revealed the presence of this family of enzymes in the metagenome.

For the function-based approach, both small and large-insert metagenomic libraries were constructed. The libraries were screened for lipolytic, amylase, protease as well as antibacterial and antibiotic resistant genes. Only lipolytic active clones were obtained. Sequence analysis of selected TLB active clones revealed the presence of three different lipolytic enzymes (isochorismatase, sulfatase and phosholipase, patatin family protein). Only the phospholipase, patatin protein was further characterized. The patatin was heterologously expressed in *E.coli.* Biochemical analysis of the partially purified protein showed that the enzyme had a preference for shorter carbon chained substrates, indicating that patatin displays esterase rather than lipase activity and functioned optimally at 30°C and pH 8.

Chapter 1
Literature Review

1.1 Introduction

The total number of prokaryotic cells on earth has been estimated at 4-6X10^{30} comprising between 10^6 and 10^8 separate genospecies (distinct taxonomic groups based on gene sequence analysis). This diversity represents an enormous (and largely untapped) genetic and biological pool that can be exploited for the recovery of novel genes, entire metabolic pathways and their products (Cowan *et al.,* 2005). Most definitive microbiological studies have been conducted in laboratories using pure cultures. Such studies have been critical to the development of microbiology, and provide the basis for our understanding of the microbial world. However, the microbial species and interactions that really count in nature do not occur in pure culture (DeLong, 2002). More than 99% of bacteria in the environment cannot be cultured using conventional methods (Yun and Ryu, 2005). The classical cultivation techniques require that the different organisms derived from an environmental sample be cultured on appropriate growth medium and then separated. Separation of bacterial communities and growing them on different media, however, results in loss of major portions of the microbial community, because of the different growth requirements of many different microbes (Entcheva *et al.,* 2001). A new frontier of science has emerged that unites biology and chemistry for the exploration of natural products from previously uncultured soil microorganisms (Handelsman *et al.,* 1998). Norman Pace and colleagues were the first to propose the use of cultivation-independent approaches to study natural microbial populations (DeLong, 2002).

The application of culture-independent nucleic acid technology has greatly advanced the detection and identification of microorganisms in natural environments (Hurt *et al.,* 2001). The use of molecular biology techniques with environmental samples has allowed researchers to examine facets of natural microbial communities that were previously inaccessible (Mumy and Findlay, 2004). Microbial ecologists, systematicists, and population geneticists have become increasingly interested in methods for complete, unbiased isolation of DNA from the environment because such procedures promise to make the genomes of uncultured indigenous microorganisms available for molecular analysis (Mo e *et al.,* 1994).

Among the methods designed to gain access to the physiology and genetics of uncultured organisms, *metagenomics*, the genomic analysis of a population of microorganisms, has emerged as a powerful centerpiece (Handelsman, 2004). The term metagenomics is derived from the statistical concept of *meta*-analysis (the process of statistically combining separate analyses) and genomics (the comprehensive analysis of an organism's genetic material (Schloss and Handelsman, 2003). Metagenomics describes the functional and sequence based analysis of the collective microbial genomes contained in an environmental sample. The past few years have witnessed an explosion of interest and activity in metagenomics, accompanied by advances in technology that have facilitated studies at a scale that was not feasible when the field began (Riesenfeld *et al.,* 2004).

Metagenomic analysis involves isolating DNA directly from an environmental sample. Numerous nucleic acid extraction methods have been developed. DNA fragmentation is a significant problem when constructing metagenomic libraries because vigorous extraction methods results in DNA shearing which affects ligation reactions (Cowan *et al.,* 2005). DNA fragments are cloned into a suitable vector, transformed into a host bacterium, and screened (Handelsman, 2004). The ability to clone large DNA fragments allows entire functional operons to be targeted with the possibility of recovering entire metabolic pathways.

Two strategies are generally used to screen and identify novel biocatalysts or genes from metagenomic libraries: function-based and sequence-based screening (Yun and Ryu, 2005). The function-driven analysis is initiated by identification of clones that express a desired trait followed by characterization of the active clones by sequence or biochemical analysis. This approach identifies clones that have potential applications in medicine, agriculture or industry by focusing on natural products or proteins that have useful activities. Sequence-driven analysis relies on the use of conserved DNA sequences for the design of hybridization probes or PCR primers to screen metagenomic libraries for clones that contain sequences of interest. Significant discoveries have also resulted from random sequencing of metagenomic clones and metagenomic DNA (Schloss and Handelsman, 2003; Edwards *et al.,* 2006).

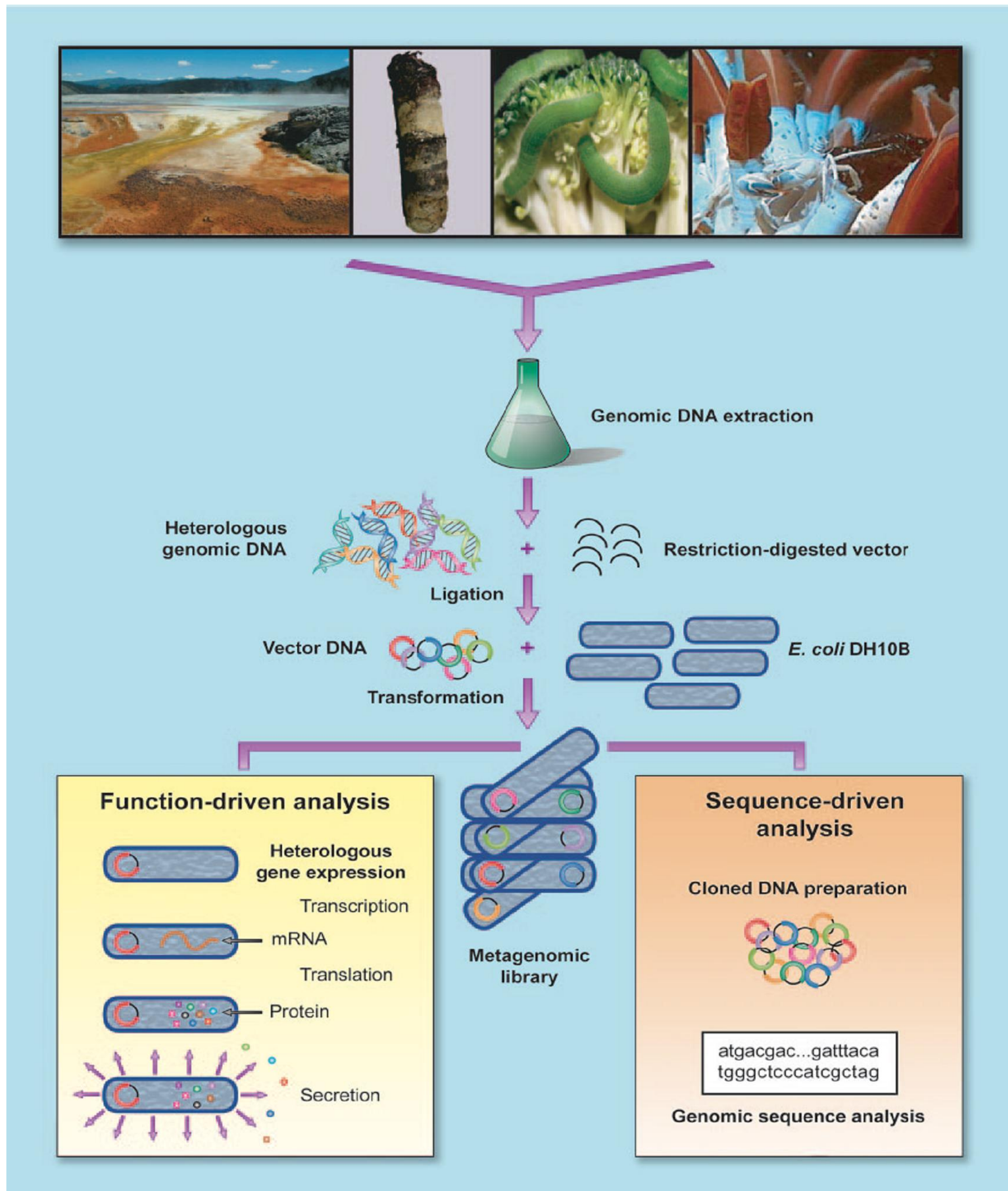


Figure 1.1: Construction and screening of metagenomic libraries. Schematic representation for the construction of libraries from environmental samples. The images at the top from left to right show bacterial mats at Yellowstone, soil from a boreal forest in Alaska, cabbage white butterfly larvae, and a tubeworm (taken from Handelsman, 2004).

1.2 DNA extraction from environmental samples

Molecular microbiology studies rely heavily on methods of DNA extraction from environmental samples with complex composition (He *et al.,* 2005). DNA extraction from environmental samples has three requirements: extraction of high molecular weight DNA; extraction of DNA free of inhibitors for subsequent molecular biological manipulations to be performed; and representative lysis of microorganisms within the sample (Yeates *et al.,* 1998). Thus, the application of a proper DNA extraction protocol is crucial.

1.2.1 DNA extraction from Soil

According to Voget *et al.* (2003), one gram of soil may contain up to 4 000 different species therefore soil appears to be a major reservoir of microbial genetic diversity and may be considered as a complex environment. This complexity results from multiple interacting parameters including pH, climatic variations and biotic activity (Robe *et al.,* 2003). Two factors that can complicate DNA extraction from soils are acidity and soil-DNA interactions. DNA is unstable under acidic conditions, owing to depurination-induced degradation of DNA. Several soil components can bind DNA thereby making it difficult to extract (Henneberger *et al.,* 2006).

There are two main approaches for the isolation of microbial DNA from soil and sediment samples: (1) the cell extraction method, and (2) the direct lysis method (Lipthay *et al.,* 2004). Cell extraction is based on the isolation of the microbial cells from soils, prior to lysis to release microbial DNA (He *et al.,* 2005). Bacterial cells are separated from the soil matrix by blending in a washing buffer followed by differential centrifugation (Jacobsen and Rasmussen, 1992). A major limitation of the cell extraction method is that it is time consuming and that spreading of organisms by aerosolization and other spills during the repeated blending and centrifugation steps is practically unavoidable. Dispersion of soil particles by cation-exchange resins (CER) has been the basis for the development of several methods to extract microorganisms from soil (Jacobsen and Rasmussen, 1992). CER extraction is partly the chemical (removal of divalent cations) and partly mechanical due to the applied shear (Frølund *et al.,* 1995). Although cell extraction methods result in the extraction of purer DNA, low yields of DNA are obtained with such methods (Gabor *et al.,* 2003).

The direct lysis method, lyses the microbial population within the soil matrix and then separates the DNA from the mixture (Zhou *et al.,* 1996). Direct extraction exposes the cellular nucleic acids to contaminating compounds such as humic and fulvic acids (Purdy, 2005). Humic and fulvic acids are formed by the polycondensation of soil organic matter derived from the remains of plants, animals and microbes. Because of their chemical nature, humic and fulvic acids are three-dimensional structures that have the ability to bind other compounds to their reactive functional groups and absorb water, ions and organic molecules (Fortin *et al.,* 2004). The humic acids in soil have similar size and charge characteristics to DNA resulting in their co-purification. Humic contaminants also interfere in DNA quantification since they exhibit absorbance at both 230 nm and 260 nm, the latter used to quantitate DNA (Yeates *et al.,* 1998). Humic and fulvic acids have been reported to interfere with restriction endonucleases, DNA-transforming enzymes, decreasing efficiencies in DNA-DNA hybridization and inhibiting Taq polymerase (Tebbe and Vahjen, 1993; Kuske *et al.,* 1998; England *et al.,* 2001).

Despite the above mentioned disadvantages, the direct lysis method has been widely used during the last decade because high yields of DNA are obtained with this method (Robe *et al.,* 2003). It is assumed that direct procedures access larger fractions of indigenous microbial populations and recover nucleic acids of larger genetic diversity than indirect methods (Gabor *et al.,* 2003). Direct lysis can be divided into the following steps: (1) washing the material to remove soluble components that may impair manipulation of the isolated DNA; (2) disruption of the cells in the soil matrix to release DNA from the cells; (3) separation of the DNA from the soil; and (4) isolation and purification of the released DNA so that it can be used in various molecular procedures. A variety of methods integrating most or all of these steps have been published (Rajendran *et al.,* 2008).

The critical step in any nucleic acid extraction method is the lysis step. Lysis methods can be divided into 3 types: (1) chemical; (2) enzymatic; and (3) physical disruption (Robe *et al.,* 2003). In most widely utilized extraction methods, a combination of lysis techniques is usually employed (Purdy, 2005).

1.2.1.1 Chemical Lysis

A number of chemicals are used to lyse cells (Purdy, 2005). The lysis mixtures can be categorized into mixtures that contain a detergent (either sodium dodecyl sulfate [SDS] (Bruce *et al.,* 1992) or Sarkosyl mixtures that contain NaCl, and mixtures that contain various buffers [usually Tris or phosphate, pH 7 to 8] (Miller *et al.,* 1999). SDS has been the most widely used cell lysis treatment for DNA extraction from pure cultures, soils and sediments (Zhou *et al.,* 1996). The modifications of the basic chemical lysis techniques include high-temperature [60°C to boiling] (Bruce *et al.,* 1992) incubation, a phenol or chloroform extraction step, and incorporation of chelating agents (EDTA and Chelex 100) to inhibit nucleases and disperse soil particles (Miller *et al.,* 1999). Chemical lysis can also select for certain taxa by exploiting their unique biochemical characteristics (Cowan *et al.,* 2005).

1.2.1.2 Enzymatic Lysis

Specific enzymes can be used to break down the cell walls of defined types of microbes: lysozyme is used to lyse gram-negative bacteria while achromopeptidase lyses gram-positive cells and lyticase lyses fungal cells. Thus, the obvious problem with enzymatic lysis is that it is selective (Purdy, 2005). Proteinase K is also often included in enzymatic extraction methods to degrade proteins in the samples to facilitate nucleic acid release (Zhou *et al.,* 1996).

1.2.1.3 Physical Lysis

Physical treatments, which destroy soil structure, tend to give the greatest access to the whole bacterial community, including bacteria deep within soil microaggregates (Robe *et al.,* 2003). The most commonly used physical disruption methods are freeze-thaw cycles, freezing in liquid nitrogen (Kuske *et al.,* 1998), followed by grinding or bead beating (Yeates *et al.,* 1998). Bead beating is based on the physical disruption of cells by glass or ceramic beads under rapid agitation and the protection of the DNA by use of a stabilizing lysis buffer. The efficiency of cell disruption, but also damage to DNA strands, depends on the energy input during beating, as well as on the type and speed of the beads (Bürgmann *et al.,* 2001). Bead beating often results in significant DNA shearing (Robe *et al.,* 2003). However, it has been reported to be the most effective lysis method presently available (Purdy, 2005).

1.2.2 DNA purification

Following nucleic acid extractions, it is usually necessary to purify the product. Contaminants may include protein, or other compounds such as humic acids (Purdy, 2005). Many different procedures for the purification of DNA has been applied including cesium chloride-ethidium bromide (CsCl-EtBr) gradient centrifugation, hydroxyapatite columns, polyvinylpolypyrrolidone (PVPP), silica matrix or magnetic capture hybridization PCR (Lipthay *et al.,* 2004). Cesium chloride-ethidium bromide density gradients are time-consuming and limit the number of samples that can be analyzed. Additionally, they often result in significant losses of extracted DNA and decreased recovery rates (Tebbe and Vahjen, 1993; Gabor *et al.,* 2003).

Protein is often co-extracted with DNA and can be removed using classical methods of protein separation, such as phenol/chloroform extraction. Phenol/chloroform/isoamylalcohol partitions nucleic acids into the aqueous phase and precipitates proteins at the aqueous/organic interface (Sambrook and Russell, 2001). Typically, inhibitory substances are removed using spin columns packed with various resins. Gel filtration (also known as size exclusion) resins have been widely applied (Miller, 2001). Cullen and Hirsch (1998) described the use of Sephadex G-75 spin columns to purify DNA. According to Miller (2001), Sepharose resins are more efficient than Sephadex resins at purifying humic acids from soil and sediment extracts. They are also easier to use because of their higher gravity-flow rates.

1.3 DNA extraction from Water

The major issue with many water samples is low biomass (Purdy, 2005). The problem is not a trivial one, because these organisms are very small (<0.6 μm) and dilute ($<10^9$/l), making nonselective collection of a sufficient number of cells and quantitative DNA extraction and purification difficult (Fuhrman *et al.,* 1988). Due to low bacterial abundance, it is necessary to concentrate large volumes (several litres) of the samples by filtration (Bej *et al.,* 1991). Filtration methods are typically used to concentrate microorganisms for analysis requiring detection of <1 microorganism per ml. Methods used previously to collect picoplankton for bulk analysis include direct filtration through cylindrical membrane filters and vacuum filtration onto fluorocarbon-based filters. The volume of water that can be filtered, typically tens of litres, limits these methods of collection (Giovannoni *et al*., 1990).

Tangential flow filtration offers the opportunity to collect large quantities of biomass from up to thousands of litres of water (O'Brien *et al.,* 1998). Giovannoni *et al*. (1990) described the use of tangential flow filtration for the concentration of marine picoplankton. The tangential flow filtration apparatus is made up of an intake prefilter (Nytex, normally 10 µm pores) and a tangential flow filter (Fluorocarbon membrane, normally 0.1 µm pores). However, numerous drawbacks of this type of filtration were highlighted: (i) possibilities of cell losses due to incomplete recovery of concentrated cells; (ii) breakage of delicate cells due to shear forces generated by repeated passage of cells through the filter unit; and (iii) unbiased collection of particular types of cells, and cell losses due to grazing of picoplankton by phagic organisms (Giovannoni *et al.,* 1990).

Several methods for DNA extraction from soil and sediments have been described but there is no widely used method suitable for environmental water (Petit *et al.,* 1999). The DNA extracted from water samples should meet the following criteria: - (i) the final DNA should be representative of the total DNA within the naturally occurring microbes at the time of sampling; (ii) the final yield should be >25 µg, (iii) the DNA should be of large molecular weight (minimum 10 kb, but preferably 50 kb or larger); and (iv) DNA should be of sufficient purity (Schmitz *et al.,* 2008). The freeze-thaw lysis extraction method was used for the efficient lysis of cells from water samples. Bacterial cells were collected by filtration using Fluoropore filters (Millipore) and subjected to six cycles of freeze-thaw lysis to release nucleic acids from the filter surface (Bej *et al.,* 1991).

1.4 DNA extraction from Biofilms

Biofilms are the product of adhesion and growth of microorganisms on surfaces. On one hand, biofilms act as biological filters by mineralizing biologically degradable material from the water and forming locally immobilized biomass. On the other hand, biofilms may unpredictably emerge in distribution systems and cause diverse problems in terms of bacterial contamination with hygienically relevant bacteria (Schwartz *et al.,* 2003). The majority of bacteria in freshwater are found growing on biofilms on the surfaces of submerged substrata or sediments, and these biofilms can be complex communities with intricate architectural organization. In natural waters, biofilms are complex heterogeneous structures composed of bacteria, algae and other microorganisms within an extracellular matrix (Jackson *et al.,* 2001). These organisms present in the biofilm are more resistant to

environmental and chemical stresses (Trachoo, 2004). The exopolymer matrix of biofilms restricts the diffusion of large molecules and bind antimicrobials. The negatively charged exopolysaccharides are also efficient in protecting cells from positively charged biocides by restricting their permeation through binding (Schwartz *et al.,* 2003). The biofilm can be collected in sterile bottles containing phosphate-buffered saline (PBS), pH 7 and transported to the laboratory (Neria-González *et al.,* 2006).

The ultra deep mines of South Africa offer access to the terrestrial deep subsurface. During normal mining operations, the advancing tunnels intersect water-bearing features or boreholes that are left to drain. Most of these water-bearing features will become the host of large-scale biofilms or mine slimes (Wanger *et al.*, 2008). Biofilms can be removed from surfaces by scraping off attached cells using a Teflon scraper. Although this method is simple and requires an inexpensive device, it may not be suitable on samples with irregular shape and rough surfaces (Trachoo, 2004).

Biofilms that were collected from an acid mine drainage site at Iron Mountain, USA, were made up of mostly an extracellular polymeric substance infused with slime cells and small cocci. Methods for the extraction of DNA from this biofilm included bead beating and freeze-thaw lysis. It was observed that the freeze-thaw lysis method produced greater quantity and less sheared DNA when compared to bead beating (Bond *et al.,* 2000). Lyautey *et al.* (2005) isolated DNA from an epilithic biofilm using a combination of the chemical and enzymatic lysis procedure. The biofilm was recovered as a homogenous suspension by using a tissue homogenizer. The extraction was done according to Zhou *et al.* (1996), and involved grinding in liquid nitrogen, freeze-thawing, and an extended hot lysis treatment with SDS. DNA obtained with this method was readily amplifiable. Samples that showed any signs of humic acid contamination (a yellowish brown colour) were purified using Sepharose columns.

1.5 Metagenomic Library Construction

The construction and screening of gene libraries prepared from DNA directly isolated from environmental samples is a recent and powerful tool for the discovery of new enzymes of biotechnological interest (Gabor *et al.,* 2004). Standard methods based on the screening of isolated microorganisms are inherently limited to the tiny fraction of cultivable microbial species (<1%); environmental gene banks in principle provide access to the entire sequence

space present in nature (Handelsman *et al.,* 1998). Environmental libraries will allow the screening of functional classes of genes from thousands of organisms. Research in this area will provide an essential backdrop for understanding evolution and biochemical pathways (Rajendran *et al.,* 2008).

The basic steps of DNA library construction involve: generation of suitably sized DNA fragments, cloning of fragments into an appropriate vector, and screening for the gene of interest (Cowan *et al.,* 2005). One of the challenges of environmental cloning is the large number of transformants that needs to be produced and screened. It has been estimated that more than 10^7 plasmid clones (5 kb inserts) or 10^6 (bacterial artificial chromosome) (100 kb inserts) would be required in order to represent collective genomes *i.e*. the metagenome of several thousand different species as typically present in a soil sample (Handelsman *et al.,* 1998).

In the first step of library construction, DNA is fragmented. Fragmentation is achieved either mechanically or enzymatically (partial digests). DNA is randomly sheared using a nebulizer to produce large fragments (~25 kb) (Goldberg e*t al.,* 2006). During nebulization, DNA solutions are squeezed through small pores that cause DNA strands to break (Gabor *et al.,* 2004). Enzyme-based methods to fragment DNA are non-random; digestion is dependant on restriction sites or methylation patterns, which may be a bias in the genome representation of the produced gene bank (Oefner *et al.,* 1996). Furthermore, enzymatic restriction may be inhibited by contaminants in the DNA extract, which is particularly a problem when working with nucleic acids isolated directly from environmental samples (Gabor *et al.,* 2004). The choice of the vector selected depends on the type of library that is constructed. Two types of libraries can be constructed (small and large-insert), a description of each of theses libraries is discussed below, the advantages and disadvantages of each library are highlighted in Table 1.1 and Table 1.2 presents some of the products that have been obtained from small and large-insert metagenomics libraries.

Table 1.1: Pros and cons of small-insert and large-insert libraries (Daniel, 2005)

Advantages	***Disadvantages***
Small-inert library (plasmids)	
High copy number allows detection of weakly-expressed foreign genes	Small insert size
Expression of foreign genes from promoter is feasible	Large numbers of clones need to be screen to obtain positives
Cloning of sheared DNA is possible	Not suitable for cloning when screening for activities and pathways that are encoded by large gene clusters
Technically simple	
Large-insert library (cosmids, fosmids, BACs)	
Large insert size	Low copy number might prevent detection of weakly expressed foreign genes
Small numbers of clones can be screened to obtain positives	Limited expression of foreign genes by vector promoters
Suitable for cloning when screening for activities and pathways that are encoded by large gene clusters	Requires high molecular weight DNA
Suitable for partial genomic characterization of uncultured soil microorganisms	Technically difficult

1.5.1. Small-insert library construction

For the construction of small-insert libraries, plasmid vectors are employed. Plasmid vectors are typically ~3-8 kb, and stable plasmid inserts are typically less than 10 kb (Laib *et al.,* 2006). Plasmid vectors can be of three main types: general purpose cloning vectors, expression vectors, and promoter probe or terminator probe vectors (Chauthaiwale *et al.,* 1992). Cloning of foreign DNA fragments in general purpose cloning vectors (*e.g.* pBR322) selectively inactivates one of the markers (insertional inactivation) or derepressers a silent

marker (positive selection) so as to differentiate the recombinants from the native phenotype of the vector (Chauthaiwale *et al.,* 1992).

Focusing on positive selection vectors, these vectors are efficient tools simplifying *in vitro* DNA recombination procedures. A variety of plasmid vectors for positive selection has been described. They rely on the inactivation of a lethal gene, a lethal site, or a dominant function conferring the cell sensitive to metabolites, or a repressor of an antibiotic resistance function (Yazynin *et al.,* 1999). The plasmid-encoding lethal genes designed to construct positive selection vectors involve (i) colicin encoding genes and (ii) a coupled cell division (*ccdB)* gene. Colicin encoded by an *E. coli* plasmid is one of the bacteriocins that affect the growth of host cells (Young-Jun *et al.,* 2002). The *ccdB* gene is a cytotoxic gene which poisons topoisomerase II (DNA gyrase) resulting in DNA damage that cannot be repaired (Matin and Hornby, 2000). To construct a positive selection vector, the *ccdB* was fused with a -galactosidase gene (*LAC*) and then inserted under the control of the *LAC* promoter, which is induced by IPTG but repressed by *LACI*^{q} repressor. With the induction of the *LAC* promoter, the *ccdB* gene is expressed resulting in the death of the host cells. However, when a DNA fragment is inserted into any one of the multiple cloning sites between the *LAC* promoter and the fused gene fragment, the toxic function of the *ccdB* gene is relieved (Young-Jun *et al.,* 2002).

Several metagenomic libraries have been constructed using plasmid vectors. Studies conducted by Henne *et al.* (1999) on soil samples collected from 3 different locations in Germany, described the identification of plasmids carrying inserts capable of utilizing 4-hydroxybuytrate as a sole carbon and energy source, in addition the clones exhibited 4-hydroxybuytrate dehydrogenase activity. Lipolytic enzymes, amylases, phosphatases and dioxygenases were identified in a number of clones obtained from a library constructed using the plasmid vector pJOE930 (Lämmle *et al.,* 2007).

Suicide vectors have also been used for the construction of metagenomic libraries. Gabor *et al.* (2004) reported on the isolation of recombinant *E.coli* strains expressing amidase activity with distinctive substrate profiles. The genomic libraries were constructed with marine sediment and soil samples using the pZERO-2 positive selection vector. In addition, novel thermophilic and thermostable lipolytic enzymes have been identified in a metagenome hot spring library constructed using the same vector (pZERO-2) (Tirawongsaroj *et al.,* 2008). The advantage of using such a vector system is the elimination

of false positives *i.e*. plasmids carrying no inserts, when working with a large clone library it is not possible to screen every single clone before storing the clones as an indexed library.

1.5.2 Large-insert library construction

For the construction of large-insert libraries, a number of vectors (phages, cosmids, fosmids and bacterial artificial chromosomes [BACs]) may be used.

1.5.2.1 Phage vectors

Phages, viruses that infect bacteria, are commonly used vector systems (She, 2003). A number of phage vectors are used in DNA and cDNA library construction. Perhaps the most widely used are phage lambda (λ) vectors. Early cloning strategies used phage lambda as a vector to archive natural population DNA (DeLong, 2002). Phage vector systems are of the gene-replacement type, the phage's genome cleavage and packaging machinery makes specific nucleolytic cleavages at the cohesive ends between concatemeric genomes (so-called **cos sites**). This releases the genomic unit-length molecules for packaging. The phage head has a tight constraint on the amount of DNA that it will accommodate (~25 kb) thereby providing a limitation for the use of this vector for the construction of large insert libraries. Despite the above mentioned limitation, studies have been conducted using both single and double-stranded phages for library construction; some of these studies are discussed below. Since phages were employed before the advent of newer cloning vectors (large insert constructs), the available literature is generally old with a few exceptions.

Phage M13, a single stranded phage has been widely characterized and its genome sequenced (Geider, 1986). The phage infects cells *via* F pili, with the appearance of a mature phage within 15 min. M13 is used in nucleotide sequencing and site-directed mutagenesis due to the fact that its genome can exist either in a single-stranded form inside a phage coat or as a double-stranded replicative form within the host cell. The expression of polypeptides fused to the surface of a filamentous phage (fusion phage vector) has been used has a powerful method for recovering particular sequences from clone libraries (Maruyama *et al.,* 1994). In addition Maruyama *et al.* (1994) constructed a fusion expression vector (λfoo), which allows foreign proteins such as *E. coli* -galactosidase and plant lectin *Bauhinia purpurea* agglutinin to be expressed on the phage surface and this vector can be used for the construction of cDNA libraries.

With regards to the double-stranded phage, bacteriophage lambda-derived vectors were widely used for the following reasons: (i) acceptance of a large fragment of foreign DNA (25 kb) by the phage, thereby increasing the chances of obtaining a complete gene; (ii) development of techniques that reduce problems of background due to non-recombinants; and (iii) ease with which the phage library can be stored at 4°C (Chauthaiwale *et al.,* 1992). Metagenomic DNA lambda libraries were constructed with samples obtained from lakes in different parts of Africa. Screening of the libraries revealed the presence of clones expressing esterase/lipase, cellulase and mannanase activity (Rees *et al*., 2003).

An improvement to the lamda vector system has been described by Pierce *et al.* (1992); the P1 cloning system allows *in vitro* packaging of foreign DNA as large as 95 kb. The DNA can be replicated as a low copy number plasmid in *E. coli,* and then induced to high copy number by the addition of isopropyl -D-thiogalactopyranoside to the medium. The cloning efficiency of the P1 system is comparable with that of λ-cosmid system (Sternberg, 1990). To overcome the problem of a high background of non-recombinants Pierce *et al.* (1992) constructed a positive selection P1 vector that contains the *Bacillus amyloliquefaciens sacB* gene. Expression of the *sacB* gene kills *E.coli* that is grown in the presence of sucrose. This cloning system has been used to construct complete *Drosophila* and mouse libraries.

Apart from the phage vectors, other vectors (cosmids, fosmids and BAC's) that contain a combination of the desirable traits *e.g.* induction from low to high copy number and stable maintance of inserts >40 kb (Kim *et al.,* 1992; Moon and Magor, 2004) have also been employed and each of these vectors are discussed below.

1.5.2.2 Cosmid vectors

Cosmids have been instrumental as recombinant vehicles for introducing large DNA inserts into *E. coli* and other gram-negative bacteria (Connell *et al.,* 1995). Cosmids are conventional plasmids that contain one or more copies of a small region of bacteriophage λ, the cohesive end site (*cos*), which contains all the *cis*-acting elements required for packaging of viral DNA into bacteriophage λ particles (Sambrook and Russell, 2001). Cosmids combine some of the features of lambda cloning (efficiency of transfection with packaged "phage" particles) with some of the advantages of using a plasmid replicon [acceptance of a larger segment of foreign DNA] (Cattaneo *et al.,* 1981). Cosmids have been shown to accommodate inserts of >30 kb (Entcheva *et al.,* 2001). The possibility of

cloning large segments of DNA in cosmid vectors offers distinct advantages, in particular for the study of multigene families. Large size fragments of mouse embryo DNA were successfully cloned in the cosmid pHC79 (Cattaneo *et al.,* 1981).

Metagenomic libraries have been constructed using cosmid vectors (Cowan *et al.,* 2005). Eland *et al.* (2006) prepared genomic DNA libraries using the cosmid vector pWE15. The library contained clones harbouring inserts of approximately 25 to 40 kb. Esterases with unique substrate specificities making them useful for biotechnological applications were identified in the cosmid library. Genes associated with hydrolytic activities were identified in a cosmid soil metagenome library. The library contained clones exhibiting agarolytic and proteolytic activity. In addition, clones also possessing cellulase, α-amylase and pectate lyases were identified (Voget *et al.,* 2003). In another study, 150 Mb of cloned environmental DNA was obtained using a cosmid vector. Sequencing analysis revealed the existence of a novel deltaproteobacterial group. In addition clones carrying a variety of genes involved in informational (DNA polymerase I subunit) and operational functions were identified (Moreira *et al.,* 2006). López-Garc a *et al.* (2004) reported on the use of a multicopy cosmid for constructing environmental libraries. However, the reasoning behind using a multicopy cosmid for constructing well represented and stable environmental libraries are not well understood, as single-copy fosmid and BAC vectors are commercially available (Béjà, 2004).

1.5.2.3 Fosmid vectors

The introduction of fosmids (F1 origin-based cosmid vector) as cloning vectors in 1992 has improved genomic cloning efforts (Béjà, 2004). Fosmids are modified plasmids that contain the F' factor origin of replication derived from *E. coli* (Kim *et al.,* 1992). Fosmids are capable of stably propagating complex inserts, and segments of genomes from higher organisms that have previously been unclonable to *E. coli* due to extreme instability of the inserts. Therefore, fosmid vectors can maintain and propagate previously unstable or unclonable genomic segments allowing for the construction of libraries with fuller representation of genomes (Kim *et al.,* 1992). The strategy of cloning into a fosmid is described in (Figure 1.3). The illustration depicts cloning using the CopyControl fosmid, a feature of this vector is inducible copy number, which allows maintenance and storage of constructs at single copy for improved stability of cloned inserts, as well as culture at high copy number for efficient DNA purification for downstream applications such as sequencing (Moon and Magor, 2004).

The latest technologies developed for the Human Genome Project using a fosmid vector were combined with oceanographic methods to produce the first environmental genomic library (Shizuya *et al.,* 1992; Stein *et al.,* 1996).

Fosmids are good vectors for constructing metagenomic libraries because of their high cloning efficiency, improved stability in *E. coli*, and optimum (40 kb) insert size (Lee *et al.,* 2004). Environmental 16S rRNA libraries constructed using fosmids have enabled researchers to gain insight into the physiology of unknown organisms, due to the large insert size that is cloned into a fosmid, sequences flanking the 16S rRNA gene yielded clues about the physiology of the organisms (Vergin *et al.,* 1998). Chung *et al.* (2008) reported on the selection of an antifungal clone from a metagenomic fosmid library prepared from forest soil, using *Saccharomyces cerevisiae* as target yeast. Antibacterial active clones obtained from a fosmid library were described in a study conducted by Lim *et al.* (2005). One clone showed antibacterial activity against *B. subtilis,* the clone also conferred a purple pigmentation on the *E. coli* colony. Analysis of the clone revealed that the pigmentation was due to the presence of isomeric compounds, indigo blue and indirubin. A fosmid library containing inserts of approximately 35 kb was constructed from tidal flats with the CopyControl fosmid pCC1FOS; this was the first report of a library of this kind, and a novel lipase was identified in that library (Lee *et al.,* 2006).

Many other studies using this type of fosmid vector have been employed. Some of these studies are discussed below. Hårdeman and Sjöling (2007) identified and isolated a novel low-temperature active lipase from a low-temperature sediment of the Baltic Sea. Low temperature lipases could be used as detergent additives, or in the processing of volatile substances, reducing energy costs, as well as for the bioremediation of low-temperature soils or water. Sulfatases, esterase enzymes that catalyze the hydrolysis of sulfate esters, were also identified in metagenomic fosmid libraries (Woebken *et al.,* 2007). Clones displaying the genetic and physiological characteristics of chitin degrading enzymes have been identified in a pCC1FOS library constructed from an alkaline, hypersaline lake and estuary (LeCleir *et al.,* 2007). One of the latest reports of a metagenome library constructed with the pCCIFOS vector is by Huang *et al.* (2009), who described the presence of clones expressing water-soluble melanin in *E. coli*. The advantage of using the CopyControl fosmid for library construction is the fact that replication can be switched from single to multiple copy number. For stable maintenance of inserts it is important that the copy number within

the host be low, however in many metagenomic studies, active clones are sequenced to determine which proteins are responsible for the observed activity. In order to do this, a high yield of DNA is needed to provide a good template for sequencing reactions and this is obtained by inducing the clones to high copy number.

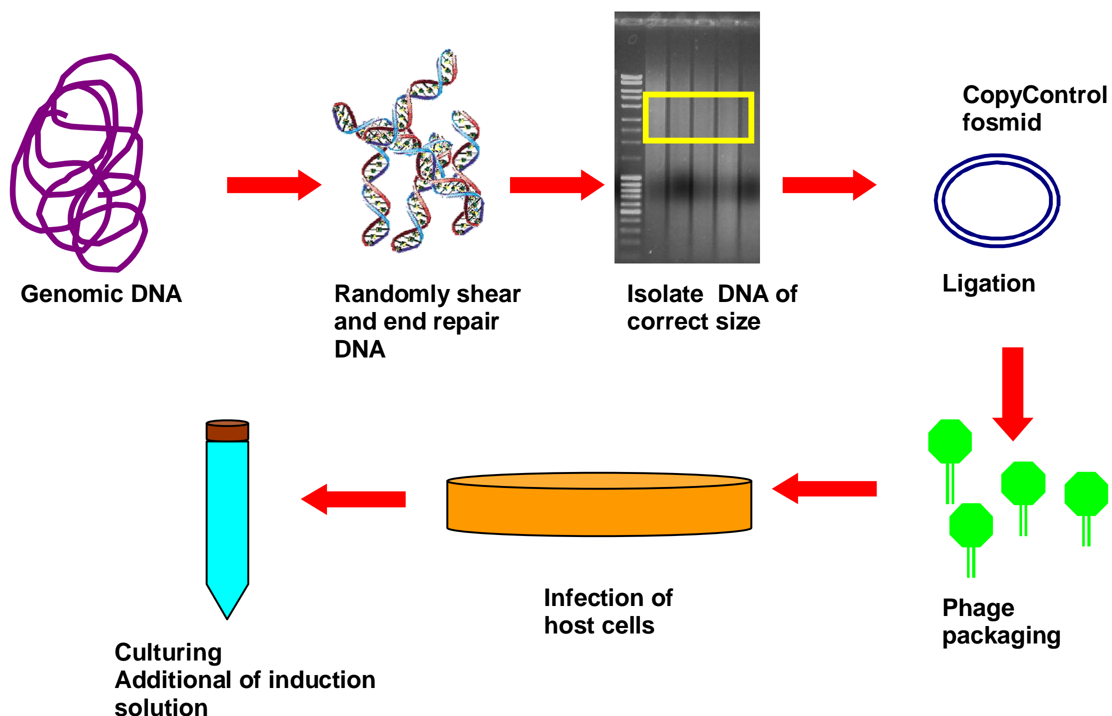


Figure 1.2: Process of fosmid cloning (Epicentre Biotechnologies).

1.5.2.4 Bacterial Artificial Chromosomes (BAC) vectors

BACs are modified plasmids that contain an origin of replication derived from *E. coli* F' factor. The replication of the BAC vector is strictly controlled, keeping the replication at one or two copies per cell (Béjà *et al.,* 2000). The advantage of BAC vectors is that they maintain very large DNA inserts [>100 kb] (Rondon *et al.,* 2000). The technical challenge in this approach, is maintaining the large size of the DNA fragments in the host cell

(Handelsman *et al.,* 1998). BAC cloning systems have been used to generate high quality genomic libraries for gene mapping, identification and sequencing (Al-Hasani *et al.,* 2003). According to Li and Qin (2005), BAC vectors are widely used in metagenomic projects for library construction because of their capacity to clone larger sized inserts and stably maintain clones of prokaryotic and eukaryotic origins.

The first application of BAC vectors to recover DNA from a mixed microbial population was reported in 1996. In this study, a 40 kb DNA fragment from a planktonic archaeon was identified in a BAC library and partially sequenced, providing the first glimpse into the gene arrangement, content and identity of marine archaea (Stein *et al.,* 1996). Rondon *et al.* (2000) described clones expressing DNase, amylase, antibacterial and lipase activity in a BAC library containing clones with an average insert size of 44.5 kb, prepared from soil samples. MacNeil *et al.* (2001) constructed, and expressed in *E. coli*, a BAC library containing genomic fragments 5-120 kb, isolated directly from soil organisms. Screening of the library resulted in the identification of several antimicrobial activities expressed by different recombinant clones. One clone has been partially characterized and found to express several small molecules related to and including indirubin. Gillespie *et al.* (2002) demonstrated the direct cloning and heterologous expression of environmentally derived DNA as a means to access new biologically active small organic compounds. They isolated and characterized two broad-spectrum triaryl cation antibiotics present at elevated levels in three color-producing clones from a BAC library containing soil DNA. Although environmental BAC libraries enable screening of larger inserts as compared to fosmids, a larger amount of starting material is required to construct the library. In addition the library can contain smaller inserts because there is no selection against small fragments in this cloning system (Béjà, 2004).

Table 1.2: Biocatalysts and bioactive compounds isolated from metagenomic libraries (Sharma *et al.,* 2005)

Biocatalyst and bioactive compound	***Vector***	***Environmental Sample***
Agarase	Cosmid	Soil
Alcohol oxidoreductase	Plasmid	Soil, Sediment
Amidase	Cosmid	Soil
Amylase	Cosmid, BAC	Various samples, soil
Cellulase	Cosmid	Soil
Chitinase	Lamda- ZAP	Marine water
DNase	BAC	Soil
Esterase	Plasmid	Soil
Glycerol/ diol dehydratase	Plasmid	Soil, Sediment
4-Hydroxybutyrate dehydrogenase	Plasmid	Soil
Lipase	Plasmid, cosmid, BAC	Soil
Nitrilase	Lamda- ZAP	Various samples
Pectate lyase	Cosmid	Soil
Xylanase	Lamda- ZAP	Insects gut
Indirubin	BAC	Soil
Long chain N-acyl aromatic amino acids	Cosmid	Soil
Polyketide synthase	Cosmid	Soil
Turbomycin	BAC	Soil
Violacein, deoxyviolacein	Cosmid	Soil

1.6 Heterologous gene expression

Since 1977 when Genetech scientists produced the first human protein (somatostatin) in a bacterium, the expression of proteins in heterologous hosts played a critical role in the launch of the biotechnology industry (Gustafsson *et al.,* 2004). In order to achieve efficient expression of a foreign gene in a bacterial cell it is necessary to put that gene under the control of the transcriptional and translational machinery of the host cell (Makrides, 1996). The choice of the host used for library construction is guided in part by the requirements for transformation and the stability of the vector (Handelsman *et al.,* 2002). *E. coli* is by far the most widely employed host, due to the vast body of knowledge about its genetics, physiology and complete genomic sequence which facilitates gene cloning and cultivation (Rai and Padh, 2001).

However, as a prokaryote expression system, *E. coli* cannot perform post-translational modifications that are required for the proper folding of the secondary, tertiary and

quaternary structures of the protein of interest (Jung and Williams, 1997). *E. coli* expressed proteins tend to retain their amino terminal methionine which may affect protein stability (Chaudhuri *et al.,* 1999). Some targeted proteins are also folded incorrectly due to the lack of chaperon molecules in the host cells (Yin *et al.,* 2007). Codon bias problems are highly prevalent in recombinant expression systems. However, by increasing the copy number of the limiting tRNA species, *E.coli* can be controlled to match the codon usage frequency in heterologous genes (Yin *et al.,* 2007). Despite the drawbacks associated with heterologous expression in *E. coli* as mentioned above, this organism is still widely used as a cloning host and this has been facilitated by many improvements that have resulted from serendipitous discoveries and this trend is likely to continue. Although certain post-translational modifications will probably remain beyond the reach of *E. coli*, suitable robust engineered strains for cloning in *E. coli* have become available (Novy *et al.,* 2001).

1.7. Screening of Metagenomic Libraries

The microbial world seems to offer the greatest natural resource of molecular diversity. The classical approach of cultivating and characterizing isolates on the strain level prior to screening and gene isolation is valid and powerful, yet severely restricted in scope (Lorenz and Schleper, 2002). Since its introduction, metagenomics has identified a significant number of novel genes encoding for biocatalysts or molecules with high potential for use in pharmaceutical products or production processes (Streit and Schmitz, 2004). Microorganisms serve as a potential source for new biocatalysts, as they have adapted to a wide range of different environmental conditions based on the development of highly optimized enzymes, which are suited to perform optimally under the physiochemical conditions of their habitats (Lämmle *et al.,* 2006). As mentioned previously, two types of screening strategies can be employed: sequence-based or function-based screening.

1.7.1 Sequence-based screening

The tremendous amount of genomic sequence information contained within metagenomic libraries can provide a window into the ‘black box’ of natural environments (Handelsman, 1998). According to Streit *et al.* (2004), both complete and random sequencing of metagenomic DNA can be useful strategies for identifying novel genes with potentially interesting characteristics. Sequence-dependent metagenomic approaches can be divided into two sub-groups (i) PCR-dependent methods that focus on the identification of single genes, and rely on the use of gene-specific consensus primers, or alternatively, (ii) direct

sequencing of metagenomic clone libraries utilizing a range of methods including end-sequencing and random sequencing (Sjöling and Cowan, 2008). The advantages of sequence-based screening is that it is not dependent on the expression of cloned genes by foreign hosts, and well established screening techniques such as PCR and colony hybridization can be employed for different targets (Daniel, 2004). Sequence-based screening involves the design and use of PCR primers and hybridization probes, which are based on conserved regions of already known genes and therefore only novel variants of known classes of proteins, can be identified. In addition, another drawback of this technique of screening is that it is not selective for full-length genes (Daniel, 2004; Knietsch *et al.,* 2003).

Despite the above mentioned drawbacks of this screening approach, genes encoding different enzymes have been isolated by PCR screening. Knietsch *et al.* (2003) isolated genes encoding B_{12}-dependent glycerol and diol dehydratases, from metagenomic libraries screened by colony hybridization using dehydratase-specific DNA fragments as probes. Dehydratases can be used for the biotechnological production of 1, 3 propanediol which is used for the industrial production of polyester fibres, polyurethanes and cyclic compounds. Roh *et al.* (2007), described the isolation of a novel soluble heme-binding domain of cyctochrome b_5 (cyt b_5) gene from the environment by PCR screening of the total metagenomic DNA. According to the author's knowledge this was the first report of a cyt b_5 gene isolated from the environment. Focusing on cytochrome genes, self-sufficient cytochrome P450 monoxygenases are useful catalysts for oxidation reactions. Kim *et al.* (2007) isolated a new self-sufficient cytochrome (clone SYK181 which formed a distinct phylogenetic line when compared to known bacterial self-sufficient ctyochromes) from a soil metagenome. The translated product of this gene showed hydroxylase activity towards fatty acids and aromatic compounds.

The future prospect of sequenced-based technology will involve DNA microarray or microchip technology. This is a powerful tool for studying gene expression and regulation on a genomic scale and detecting genetic polymorphisms (Wu *et al.,* 2001). High density arrays can be effective for the quantitative detection of genes from complex environments such as soil and water (Sebat *et al.,* 2003). In addition, in the past few years researchers have witnessed an explosion in total genomic and metagenomic sequenced-based projects in order to gain insight into the microbial diversity that is present in the respective

environments (Keller and Ramos, 2008; Raes *et al.,* 2007). Currently, high throughput sequencing technologies, such as pyrosequencing have been employed for the sequencing of complete genomes of single organisms as well as metagenomes; the 454 Life Sciences GS20 sequencing platform has been employed for such applications (Wang *et al.,* 2007). Pyrosequencing technology relies upon enzyme cascades and CCD (charge coupled device) luminescence detection capabilities to measure the release of inorganic pyrophosphate with every nucleotide that is incorporated (Huse *et al.,* 2007). Pyrosequencing has been successful for both confirmatory sequencing and *de novo* sequencing (Ronaghi, 2001) with the only limitation being shorter read lengths (100 bp) as compared to Sanger sequencing. The GS FLX system which is a modification of the GS20 system has shown promise in the *de novo* sequencing of genomes. This is due to the read length obtained (200 bp) which is high enough to allow the assembly of whole genome sequences (Pühler, 2008). To date, projects based on sequencing of both whole genomes, plasmids and metagenomes have highlighted the advantages of this technology (Moore *et al.,* 2006; Smith *et al.,* 2007; Szczepanowski *et al.,* 2008).

1.7.2 Function-based screening

Functional expression is commonly used as a method to screen for specific gene classes. Functional searches for novel genes in metagenomic libraries have often been performed using highly sophisticated colony picking and pipetting robots. In many recent publications, large clone-libraries have been screened. However, often several hundred thousand clones have to be analyzed to detect less than ten active clones in a single screen (Streit and Schmitz, 2004).

Despite the above mentioned drawback of this type of screening, this approach still enables the rapid acquisition of clones that have the potential for direct application in industry. Moreover, this screening method can detect genes with completely novel DNA sequences, which may have functions distinct from known biocatalysts (Yun and Ryu, 2005). Diverse activities have been discovered by functional analysis of metagenomic libraries. The diversity of functionally active clones discovered in metagenomic libraries validates the use of functionally active screens as one means to characterize the libraries (Riesenfeld *et al.,* 2004). The advantage of this approach is that it does not require that the genes of interest be recognizable by sequence analysis, making it the only approach in metagenomics that has the potential to identify entirely new classes of genes with new or known functions

(Handelsman, 2004). Even though there are some limitations in screening, such as the functional expression of foreign genes in a heterologous screening host, many novel biocatalysts, such as esterases, lipases, proteases, oxidoreductases, nitrilases, and amylases have been isolated using this approach (Kim *et al.,* 2006; Lorenz and Schleper, 2002).

Environmental DNA libraries that recover functional genes from uncultivated bacteria provide a promising drug discovery tool (Courtois *et al.,* 2003). Genes involved in natural product synthesis can be cloned directly from environmental samples such as soil, and expressed in a heterologous host, supporting the idea that this technology has the potential to provide novel natural products from the wealth of environmental microbial diversity present, and is a potentially important new tool for drug discovery (MacNeil *et al.,* 2002). Healy *et al. (*1995), constructed gene libraries ("zoolibraries") in *E. coli* using DNA isolated from the mixed liquor of thermophilic, anaerobic digesters, which were in continuous operation with lignocellulosic feedstocks for over 10 years. Clones expressing cellulase and xylosidase were readily recovered from these libraries. Sodium proton antiporters play a role in transporting Na^+ across the cytoplasmic membrane of all living cells. The Na/$^+$ H^+ antiporters have several functions, such as maintenance of intracellular pH homeostasis, detoxification of cells from Na^+, regulation of cell volume, and establishment of an electrochemical potential of Na^+. Environmental DNA libraries prepared from three different soils were screened for genes conferring $Na^+(Li^+)/H^+$ antiporter activity on antiporter-deficient *E. coli* (Majern k, 2001).

Chitin, a (1 4) -linked homopolymer of N-acetyl-D-glucosamine is an abundant structural polysaccharide produced by many marine organisms. Cottrell *et al.* (1999), constructed genomic DNA libraries from uncultivated marine bacteria. Screening of the libraries revealed the presence of clones that were capable of hydrolyzing a fluorogenic analogue of chitin. The ability to degrade chitin would be an important attribute of marine bacteria given the presumed high input of detrital chitin into the sea.

1.8 Scope of the present study

This study focuses on the construction and screening of a South African deep mine metagenome with the aim of identifying novel biocatalysts. Many challenges are faced when constructing metagenomic libraries due to the complexity of some of the sample material tested. In this study, we used a pure culture served as a control experiment to test techniques already in place as well as techniques that had not yet been employed in our laboratory. This thesis also highlights the assessment of the diversity of the deep mine, sequenced-based screening for a selected gene using a PCR based approach and mining the metagenome for the discovery of novel biocatalysts.

Chapter 1: focuses on literature surrounding the topic of this project, and the selection criteria for some of the techniques used in this study.

Chapter 2: focuses on library construction (both small and large-insert libraries) using a pure culture, *Geobacillus thermoleovorans* GE7. The constructed libraries were screened for the presence of lipolytic activity. All active lipolytic clones were sequenced using pyrosequencing by 454 Life Sciences. Sequencing of this magnitude provided insight into some of the genes carried within the GE7 genome since this organism's genome has not yet been sequenced.

Chapter 3: provideds an assessment of the diversity of a deep mine sample. Different extraction methods were evaluated and we observed that although extraction kits yield DNA of high purity, the DNA yields are lower when compared with manual extraction methods. We selected to use a manual method that incorporates a pre-treatment step to remove some of the potential inhibitors that may affect downstream applications. The microbial population of the deep mine was investigated by cloning and sequencing of the 16S and 18S rRNA genes for bacteria, archaea and eukarya. A further phylogenetic study was performed on the bacterial 16S clonal library. The types of organisms present in the metagenome indicated the possible presence of alkanes in the deep mine. Based on this, we decided to screen for the presence of alkane degraders *.i.e.* cytochrome P450 monooxygenases using the sequence-based approach.

Chapter 4: the presence of cytochrome P450 monooxygenases (alkane degraders) in the deep mine was investigated in this chapter. The CYP153 (terminal hydroxylase and long chain alkane degraders) was isolated using a PCR based approach. The gene was cloned and sequenced. According to the sequence analysis, this family of enzymes is present in the biofilm and the biotechnological application of this family of enzymes is also highlighted.

Chapter 5: describes the construction of small and large-insert metagenomic libraries. The libraries were screened for clones displaying lipolytic, amylase, protease, antibacterial and antibiotic resistance. Only lipolytic active clones were obtained. The active clones were sequenced to determine which proteins were responsible for the displayed activity. Sequence analysis revealed the presence of 3 different esterase-type enzymes (isochorismatase, sulfatase and phospholipase, patatin). The phospholipase, patatin was heterologously expressed in *E. coli*. The purified protein was further characterized.

Chapter 2

Construction of genomic libraries with a South African deep mine isolate and screening for the presence of lipases

2.1. Introduction

Studies have shown that the biosphere extends at least 2.8 km beneath the Earth's surface (Fredrickson and Onstott, 1996). The extreme conditions encountered in the subsurface include high temperature, high pressure, high salinity, high ambient radiation, low availability of energy sources, and in the case of natural gases, low water activity. However, microorganisms, specifically bacteria, have been discovered living under such extreme environmental conditions (DeFlaun *et al.,* 2007).

Rod-shaped bacteria which differentiate into endospores under aerobic conditions have traditionally been assigned to the genus *Bacillus* and a very diverse collection of species have been assigned to this genus (Ash *et al.,* 1993). The *Bacillus* rRNA group 5, which comprised thermophilic *Bacillus* strains was transferred into a new genus *Geobacillus* which represented a phenotypically and phylogenetically coherent group of thermophilic bacilli with high levels of 16S rRNA sequence similarity [98.5-99.2%] (Poli *et al.,* 2006; Feng *et al.,* 2007). The members of this new genus have been reported to be present in various thermophilic and mesophilic geographic areas (Rahman *et al.,* 2007). Less than five years ago an obligate thermophilic bacterium designated GE-7 was isolated from a 3.2 km depth from a South African gold mine. The *in situ* rock temperature at that respective depth was ~50°C, and the pH of the fissure water was 8. Phylogenetic analysis of the 16S rRNA revealed that GE-7 belonged to the genus *Geobacillus*, and was closely related to *Geobacillus thermoleovorans* (99.6%) and thus named *Geobacillus thermoleovorans* GE-7 (DeFlaun *et al.,* 2007).

Industrial interest in *Geobacillus* species have arisen from their potential applications in biotechnological processes (McMullen *et al.,* 2004). Thermophiles represent a potential source of thermostable enzymes. Enzymes isolated from these microorganisms have been shown to be active at high temperatures and often resistant to, as well as active in the presence of organic solvents and detergents (Hawumba *et al.,* 2002). Thermostable

enzymes, such as amylases and proteases, have been investigated from thermophilic bacteria (Abdel-Fattah, 2002). However, according to Li and Zhang (2005) thermophiles are also good candidates for producing thermostable lipases. These enzymes catalyze the hydrolysis of ester bonds of triacylglycerols at the interface between an insoluble substrate and water. This feature distinguishes lipases from esterases which act on water soluble substrates (Soliman *et al.,* 2007). The three dimensional structures of esterases and lipases show the characteristic alpha/beta fold composed of alpha-helices and beta-sheets which is also found in haloperoxidases and epoxide hydrolases. Most of these hydrolases contain the conserved pentapeptide G-X-**S**-X-G (X- denotes any amino acid) around the active site serine and function *via* a Ser-Asp-His catalytic triad and the oxyanion hole formed by the backbone amides of two conserved residues (Fischer and Pleiss, 2003; Rhee *et al.,* 2005). Lipolytic enzymes have significant biotechnological importance because of their ability to catalyze regio- and stereo-selective organic reactions; they have no requirements for co-factors, are stable in organic solvents and have broad substrate specificity (Lee *et al.,* 2004; Ranjan *et al.,* 2005).

Studies conducted by DeFlaun *et al.* (2007), have postulated the presence of lipolytic enzymes (lipases and esterases) in GE-7. We therefore selected *Geobacillus thermoleovorans* GE-7 as the test culture, for constructing genomic libraries, and screened for lipases/esterases before proceeding to the environmental sample. We decided to construct both small-insert (plasmid) and large-insert (fosmid) libraries. For the small-insert library our aim was to evaluate sticky-ended cloning *versus* blunt-ended cloning. Although creating a sticky library is a more practical choice, we decided to test both strategies for the following reasons: - (i) DNA extracted from the environment is not of the highest purity, and (ii) contaminants present in the extraction may affect the functioning of restriction enzymes tested, making the DNA difficult to digest and this will have a negative impact on library construction. If we obtain such a sample when proceeding to the construction of metagenome libraries, then we would have to fragment the DNA by physical methods such as nebulization. In addition, the DNA would have to be end-repaired thereby following the process of blunt-end cloning. Therefore, it would be important to also have the blunt-end cloning optimized.

With reference to library screening, the number of plasmid clones that need to be screened to obtain a positive hit are extremely large due to the small size of inserts cloned into

plasmids, therefore making the screening process technically difficult. We constructed a fosmid library to alleviate the problem of tedious screening procedures. With the fosmid vector the insert sizes that can be cloned are up to 40 kb, therefore allowing for a fewer number of clones to be screened in order to obtain a positive hit.

2.2 Materials and Methods

2.2.1 Genomic DNA extraction

Large scale genomic DNA extraction was performed using the hot phenol method as described by Towner, 1991. A GE-7 culture, grown for 2 days at 55°C in R2A broth (Merck), was used for the extraction. Cells (100 ml) were centrifuged at 5 000 rpm for 10 min at 4°C. The mass of the cell pellet was determined. Pellets were then resuspended in TE buffer (10 mM Tris-HCl; 100 mM EDTA, pH 8, Merck) [40ml for every 0.5 g of cells] and centrifuged at 5 000 rpm for 5 min at 4°C. This was followed by the addition of the resuspension buffer (50 mM Tric-HCl, pH 8; 0.7 M sucrose, Merck) [3.2 ml for every 0.5 g of cells]. Lysozyme (10 mg/ml Sigma) [600 μl for every 0.5 g of cells] was added and the cells were placed on ice for 5 min. This was followed by the addition of 0.5 M EDTA pH 8 (600 μl for every 0.5 g of cells) and 10% SDS (Merck) (500 μl for every 0.5 g of cells). The suspension was gently mixed and placed on ice for 5 min. Digestion buffer (1% SDS; 0.1 M EDTA; 0.05 M Tris-HCl; 0.2 M NaCl; Proteinase K (Sigma) to a final concentration of 5 mg/ml) was then added. The suspension was incubated for 16 h at 55°C with mild shaking. An equal volume of phenol:chloroform:isopropanol (25:24:1) (Merck) was added to the suspension and gently mixed for 3 h at room temperature. This was followed by centrifugation at 6 000 rpm for 10 min at 25°C. The supernatant was transferred to a new tube and an equal volume of chloroform:isopropanol (24:1) was added to the suspension and gently mixed. The suspension was centrifuged at 5 000 rpm for 10 min at 4°C. The supernatant was transferred to a new tube, and placed on ice for 5 min. This was followed by the addition of 0.1X the volume 5 M NaCl, pH 5.7 and 10 ml of 100% cold ethanol. The suspension was placed at -20°C for 10 min. The DNA was spooled and transferred to sterile 1.5 ml eppendorf tubes. Nucleic acids were pelleted by centrifugation for 1 min at 4°C. The resultant pellets were washed with 1 ml of 70% ethanol, centrifuged for 1 min at 4°C and air dried. The pellet was then resuspended in TE buffer, incubated at 55°C for 1 h and stored at 4°C.

2.2.2 Confirmation of culture identity

Before proceeding to library construction, it was important to confirm that we were working with GE-7. This was done by firstly amplifying the 16S ribosomal RNA (rRNA) using universal bacterial primers (Table 2.1). Restriction analysis of the 16S rRNA was performed using *Eco*RI; *Sma*I and *Pst*I (Fermentas) that were selected to give fragments of approximately *Eco*RI (855 bp and 663 bp), *Sma*I (1377 bp and 141 bp), and *Pst*I (850 bp and 668 bp). Each of these enzymes digests the 1500 bp PCR product once. The restriction digests were performed according to the manufacturer's instructions for each of the enzymes.

Table 2.1: 16S rRNA primer sequences (Lane, 1991)

NAME	***SEQUENCE***
27F	5´-AGAGTTTGATCMTGGCTCAG-3´
1492R	5´-TACGGYTACCTTGTTACGACTT-3´

2.2.3 Small-insert library construction

Small-insert libraries were constructed with the pZERO-2 suicide vector (Invitrogen). The vector contains a lethal gene (*ccdb*) (Figure 2.1). The *ccdB* gene is cytotoxic and functions by poisoning topoisomerase II (DNA gyrase) resulting in DNA damage that is irrepairable (Matin and Hornby, 2000). However, when a DNA fragment is inserted into the multiple cloning sites between the *LAC* promoter and the fused gene fragment, the toxic function of the *ccdB* gene is relieved due to disruption of the gene (Young-Jun *et al.,* 2002). Both sticky and blunt-end cloning was performed with this vector.

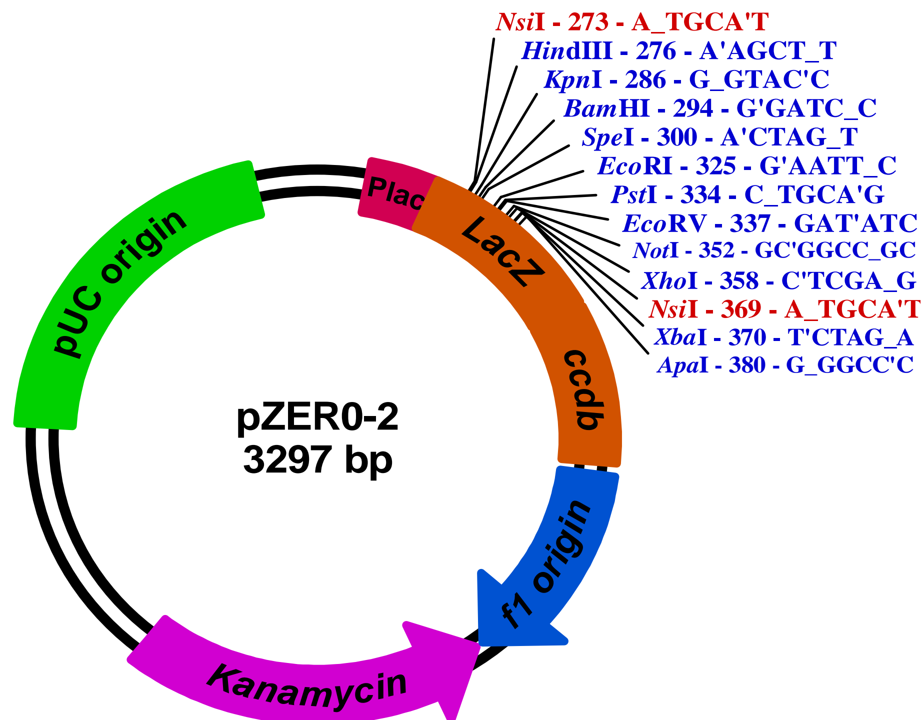


Figure 2.1: Map of pZerO-2 plasmid vector (Invitrogen).

2.2.3.1 Preparation of genomic DNA for small insert library construction

2.2.3.1.1 Partial digestion of GE-7 genomic DNA

For the construction of the sticky-ended library, the GE-7 gDNA was partially digested with *Xho*I (Fermentas). Partial digests were done using time dependant digestion for the respective enzyme. The digests at different time intervals were run on 0.8% agarose gels stained with SYBR Gold (Invitrogen) as recommended by the vector instruction manual. The size range selected for cloning was 2-6 kb. This region was excised from the gel using a sterile scalpel and DNA was purified using the GFX Gel Purification Kit (Amersham Biosciences) according to the manufacturer’s instructions.

2.2.3.1.2 Physical fragmentation of GE-7 genomic DNA

For blunt-end cloning the gDNA was fragmented by nebulization, 50 µg of DNA was added to 750 µl of Shearing Buffer (TE buffer + 10% glycerol) and pipetted into the bottom of the nebulizer placed on ice (Invitrogen). The nebulizer was connected to a compressed gas or air source and the DNA sheared for 2 min at 10 psi. Following nebulization, 700 µl of the sheared DNA was transferred to a sterile microcentrifuge tube and 80 µl 3 M sodium acetate, pH 5.2 (Merck) and 700 µl 100% isopropanol (Merck) was added, mixed well and incubated overnight at -20°C. Following incubation, the sample was centrifuged at 14 000 rpm for 15 min at 4°C. The supernatant was removed and the DNA pellet washed with 800 µl of cold 100% ethanol (Merck), centrifuged at 14 000 rpm for 5 min and the ethanol decanted. The sample was centrifuged again at 14 000 rpm for 1 min. The pellet was air dried and resuspended in 50 µl of sterile water and stored at -20°C until further use.

2.2.3.1.3 End repair of fragmented DNA

The sample was firstly polished with BAL31 Nuclease [Fermentas] (Goldberg *et al.,* 2006). The reaction was incubated at 30°C for 5 min and purified using the GFX Purification kit (Amersham Biosciences). This was followed by treatment with T4 Polymerase and Klenow [Fermentas] (Goldberg *et al.,* 2006). During treatment the sample was incubated at room temperature for 35 min and thereafter extracted with a equal volume of phenol: chloroform:isoamylalcohol (25:24:1), centrifuged at 14 000 rpm for 5 min. The upper layer was transferred to a new tube and extracted with a equal volume of chloroform: isoamylalcohol (24:1), again the upper layer was transferred to a new tube. DNA was precipitated overnight with 100% ethanol, thereafter centrifuged at 14 000 rpm for 15 min to pellet the DNA. The pellet was washed (5X) with 80% ethanol to remove all traces of phenol, air dried and resuspend in 50 µl of sterile water.

2.2.3.1.4 Ligation reactions

Ligation reactions using T4 ligase and ligase buffer (Fermentas) were incubated overnight at 16°C. For sticky-ended cloning a vector: insert ratio of 1:2 was used. The fractionated gDNA was ligated into the *Xho*I digested pZERO-2 vector. For blunt-ended cloning a vector: insert ratio of 1:10 was used and the partially digested *Eco*RV gDNA was ligated into the *Eco*RV (Fermentas) digested pZERO-2 vector.

2.2.3.1.5 Bacterial transformation

Ligation mixtures were used to transform *E. coli* TOP10 competent cells (Invitrogen) by heat shock according to Sambrook *et al.* (1989). To 50 µl of chemically competent cells, 10 µl of the ligation mix containing ~100 ng of DNA was added, mixed gently and incubated for 30 min on ice. Following incubation the tube was transferred to a water bath set at 42°C for 40s. After heat shock, the tubes were immediately placed on ice for 2 min. This was followed by the addition of 800 µl of LB broth and incubation of the cells at 37°C for 1 h with mild shaking. After the 1 h incubation period the transformation mix was plated out onto LB kanamycin (50 µg/ml) (Merck) agar plates to select for transformants. Plates were incubated overnight at 37°C.

2.2.3.1.6 Plasmid DNA extraction and restriction analysis

Random colonies (20) were selected and inoculated into 5 ml of LB broth containing kanamycin (50 µg/ml). The tubes were incubated overnight at 37°C with mild shaking. Plasmid DNA was then extracted from the cultures using the lysis by boiling method described by Sambrook *et al.* (1989) with slight modifications. One ml of culture was pelleted by centrifugation at 14 000 rpm for 2 min. The supernatant was discarded and the pellet resuspended in 350 µl STET buffer (0.1 M NaCl; 10 mM Tris-HCl [pH 8]; 1 mM EDTA [pH 8]; 5% Triton X-100) and briefly vortexed. This was followed by the addition of 25 µl lysozyme (Sigma) (10 mg/ml) to the suspension and vortexing for 3 s. The tubes were placed in a boiling water bath for 40 s and thereafter centrifuged for 10 min at 14 000 rpm at room temperature. Following centrifugation, the pellet containing the cellular debris was removed using a sterile toothpick. To the supernatant, 420 µl of cold isopropanol was added, the suspension was mixed by inversion of the tubes and the DNA was precipitated by placing the tubes at -20°C for 30 min. Following precipitation, the tubes were centrifuged at 4°C for 5 min. The supernatant was removed and 1 ml of cold 70% ethanol was added to the pellet. The tubes were centrifuged at 4°C for 2 min. The ethanol was removed and the pellet was dried in a rotary evaporator (Savant). The pellet was then resuspended in 50 µl sterile water.

2.2.4 Large-insert library construction

A large-insert library was constructed using the pCC1FOS fosmid vector (Epicentre Biotechnologies). Fosmids are good vectors for constructing metagenomic libraries because of their high cloning efficiency, improved stability in *E. coli*, and optimum (40 kb) insert size

(Lee *et al.,* 2004). Fosmids can maintain and propagate previously unstable or unclonable genomic segments allowing for the construction of libraries with fuller representation of genomes (Kim *et al.,* 1992).

2.2.4.1 Preparation of genomic DNA for large-insert library construction

2.2.4.1.1 Partial digestion of GE-7 gDNA

Genomic DNA from GE-7 was partially digested with *Eco*RV. Partial digests were done according to the manufacturer's instructions for the respective enzyme and the digest was run on a 1% agarose gels stained with SYBR Gold. A 36 kb fosmid control DNA was used as the size marker. The size range selected for cloning was a region ~1 cm above and below the fosmid control DNA. This region was excised from the gel using a sterile scalpel and DNA was purified using the GFX Gel Purification Kit (Amersham Biosciences) according to the manufacturer's instructions.

2.2.4.1.2 Ligation into pCC1FOS fosmid vector, phage packaging and titering

Ligation reactions were performed at room temperature for 2 h according to the manufacturer's instructions (Epicentre Biotechnologies). Before the packaging, EPI300-$T1^R$ host cells (provided with kit) were plated out on a LB plate and the plate was incubated overnight at 37°C. This was followed by inoculating 50 ml LB broth + 10 mM $MgS0_4$ with a single colony of culture and incubating the broth at 37°C overnight with shaking. On the day of packaging, 50 ml LB broth + 10 mM $MgS0_4$ was inoculated with 5 ml of the overnight culture and incubated at 37°C to an A_{600} - 0.8-1.0. The cells were stored at 4°C until needed. For the packaging reactions, one tube of the MaxPlax Lambda Packaging Extract (provided with kit) was thawed on ice. When thawed, 25 µl (one-half) of each packaging extract was immediately transferred to a second sterile tube and placed on ice and the remaining 25 µl was returned to -70°C for later use. This was followed by the addition of 10 µl of the ligation mix to 25 µl of the MaxPlax Lambda Packaging Extract held on ice. The sample was mixed and incubated at 30°C for 90 min. After the 90 min incubation, the remaining 25 µl of the MaxPlax Lambda Packaging Extract was added and the incubation procedure was repeated as mentioned above. Following incubation, Phage Dilution Buffer (10 mM Tris-HCl [pH8.3]; 100 mM NaCl; 10 mM $MgCl_2$) was added to a final volume of 1 ml gently mixed and 25 µl of chloroform was added.

For the phage titering procedure serial dilutions of the 1 ml packaged phage particles were made using the Phage Dilution Buffer (PDB). This was followed by the addition of 10 µl of each of the above dilutions to 100 µl of EPI300-T1^R host cells. The cells were incubated at 37°C for 1 h and plated out on an LB agar plate supplemented with chloramphenicol (12.5 µg/ml). We attempted to follow the cloning strategy for the BAC vector with the fosmid (transformation by electroporation). The reasoning behind this was as follows, comparison of the pCC1FOS vector map with that of the pCC1BAC vector map, showed that the maps were almost identical (Figure 2.2). Following infection and transformation the plates were incubated overnight at 37°C to select for fosmid clones. Random clones were inoculated into 5 ml LB broth containing chloramphenicol (12.5 µg/ml). Fosmids are low copy number vectors (1-10 copies per cell), therefore it was important to add an induction solution to the inoculated tubes in order to increase the copy number prior to fosmid DNA extractions. The tubes were then incubated overnight at 37°C with mild shaking.

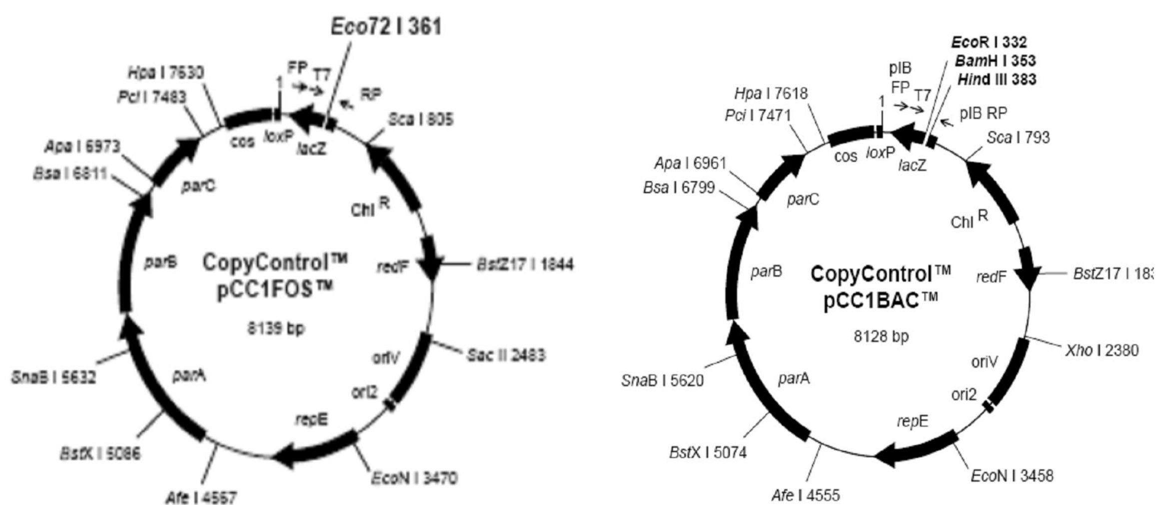


Figure 2.2: Vector maps of pCC1FOS and pCC1BAC (Epicentre biotechnologies)

2.2.4.1.3 Fosmid DNA extractions and restriction analysis

Fosmid DNA extractions were performed using the GeneJet Plasmid Miniprep Kit (Fermentas). Prior to extraction, the clones were induced to high copy number using the induction solution provided with the cloning kit according to the instruction manual. The

bacterial cells were pelleted by centrifugation at 14 000 rpm for 2 min, the cells were then resuspended in 250 µl of resuspension solution and vortexed until no cell clumps were visible. Thereafter, 250 µl of the lysis solution was added and the sample was mixed properly by inverting the tubes 4-6 times, this was then followed by the addition of 350 µl of the neutralization solution, the sample was mixed properly by inverting the tubes and centrifuged at 14 000 rpm for 5 min to pellet cell debris and chromosomal DNA. Following centrifugation, the supernatant was transferred to a spin column provided with the kit and centrifuged at 14 000 rpm for 1 min. This was followed by the addition of 500 µl of wash solution, the sample was centrifuged at 14 000 rpm for 1 min, the flow through discarded and the washing and centrifugation step was repeated. An additional centrifugation step was incorporated to ensure the removal of residual ethanol; the sample was centrifuged at 14 000 rpm for an additional 1 min. The spin column was transferred to a sterile 1.5 ml tube and 50 µl of pre-warmed sterile water was added to the column, the column was incubated for 2 min at room temperature and thereafter centrifuged at 14 000 rpm for 2 min.

Due to the large size of expected inserts, restriction digests were separated by PFGE. Digests were run on a 1% agarose gel, for 18 h in a CHEF DRII pulsed field gel box (BioRad Laboratories) in 1X TAE at 14°C with a voltage gradient of 3.63 V/cm; and the switch interval ramped from 1-6 s. The gel was stained with ethidium bromide for 20 min and viewed under a UV transilluminator.

2.2.5 Library screening

Lipolytic activity was measured by the tributyrin plate halo assay. The plate assay was performed using an adapted protocol described by Ro *et al.* (2004), 1% tributyrin and 1% gum arabic (Sigma) was dissolved in 20 ml of sterile water and mixed by sonication until a milky homogenous suspension was obtained. The suspension was then added to LB agar, autoclaved and cooled before the addition of kanamycin (50 µg/ml). Clones were replica plated onto LB tributyrin (TLB) plates. Plates were incubated at 37°C until activity was detected.

2.2.5.1 Confirmation of lipase activity

Assaying for activity on LB tributyrin does not distinguish lipase from esterase activity due to the fact that both lipases and esterases are capable of hydrolyzing tributyrin, thus it was important to confirm lipase activity. A second plate assay using LB olive oil and rhodamine B (Sigma) (Kouker and Jaeger, 1987) was performed on the positive TLB clones. Lipases only are able to hydrolyze olive oil because of their affinity for longer carbon chained substrates (Chahinian *et al.,* 2002), 1% olive oil and 1% gum arabic was dissolved in 20 ml sterile water and mixed by sonication to obtain a homogenous suspension. Rhodamine B (1 mg/ml) was dissolved in sterile water and filter sterilized. The rhodamine B solution (0.001%) and the olive oil was then added to LB agar, autoclaved and cooled before the addition of kanamycin (50 μg/ml). Agar plates containing the olive oil and rhodamine B appear pink and opaque. Positive clones were replica plated onto the agar plates and incubated at 37°C until activity was detected. Lipase production was monitored by irradiating plates with UV light at 350 nm. An orange fluorescence formed around a colony was indicative of a positive result. In addition to the plate assays whole cell PCR was done on the TLB positive clones using lipase specific primers (Integrated DNA Technologies) which amplify a 500 bp fragment internal to the lipA gene (Table 2.2). The primers were designed based on the lipA in *Geobacillus kaustophilus* HTA426 and this organism was used as the positive control for the PCR reactions.

Table 2.2: LipA specific primers (Barnard, 2005)

NAME	***SEQUENCE***
GTL2F	5'- ATC CAT ATC ATC GCC CAC AGC CAA G-3'
GTL2R	5' CGT CCG TTC TGT GGC AAA GCT CAA A- 3'

2.2.6 Ultrafast sequencing of GE-7 fosmid and plasmid clones using the Genome sequencer 20 FLX system

Fosmid and plasmid DNA was extracted from *G. thermoleovorans* active clones grown overnight at 37°C in LB broth with the respective antibiotic. Ten micrograms of pooled plasmid and fosmid DNA (100 ng of each fosmid and plasmid) was sequenced at Inqaba

Biotechnologies, South Africa using pyrosequencing methodology. GS 20 FLX library construction and sequencing was performed as described by Marguiles *et al.* (2005). Briefly, the pooled fosmid and plasmid DNA was sheared by nebulization to a size range of 300-800 bp. DNA fragment ends were repaired and phosphorylated using T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides “A” and ”B” supplied with the 454 Life Sciences sequencing reagent kit were ligated to the DNA fragments using T4 DNA ligase. Purified DNA fragments were hybridized to DNA capture beads and clonally amplified by emulsion PCR (emPCR). DNA capture beads containing amplified DNA were deposited onto a 70 X 75 mm PicoTitrePlate. The raw reads obtained were assembled into contigs using the Newbler Assembly software.

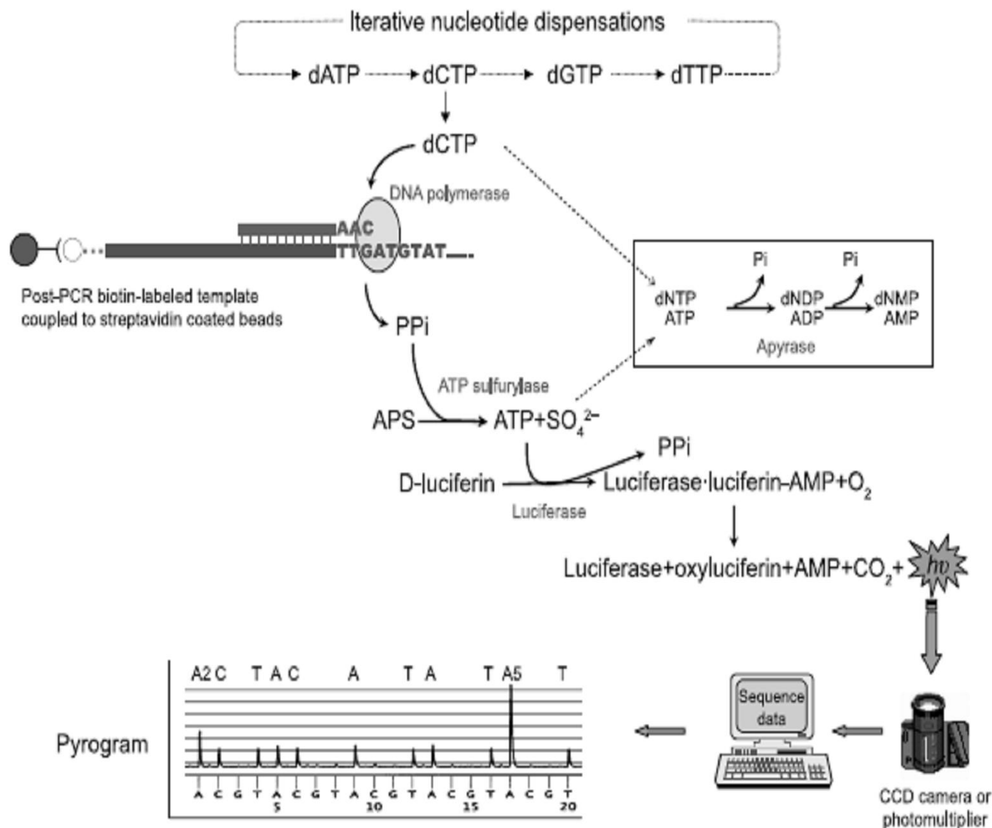


Figure 2.3: Schematic illustration of the Pyrosequencing method. Pyrosequencing is a non-electrophoretic real-time DNA sequencing method that uses the luciferase-luciferin light release as the detection signal for nucleotide incorporation into target DNA. The four different nucleotides are dispensed iteratively to a four-enzyme mixture. The pyrophosphate (PPi) released in the DNA polymerase-catalyzed reaction is quantitatively converted to ATP by ATP sulfurylase, which provides the energy to firefly luciferase to oxidize luciferin and generate light (ho). The light is detected by a photon detection device and monitored in real time by integrated software in a format called program. Finally, apyrase catalyzes degradation of nucleotides that are not incorporated and the sequencing reaction will be ready for the next nucleotide addition (taken from Gharizadeh *et al.,* 2007).

2.2.6.1 DNA sequencing with the ABI 3730XL Automated Sequencer (Applied Biosystems)

Single Sanger sequencing runs in both directions were performed on individual plasmids and fosmids in both directions. This was done in order to close the gaps between the contigs and assemble the individual clones. The sequencing reactions were performed with the ABI 3730XL Automated Sequencer (Applied Biosystems). The sequence reaction contained the following:

Plasmid Sequencing reactions

DNA	5 µl
Big Dye	0.5 µl
5X Seq Buffer	2 µl
Primer (Sp6 /T7) 3.2 pmol	1 µl
Sterile water	1.5 µl
Total	**10 µl**

Table 2.3: ABI-Plasmid-Cycle program

ABI cycle sequencing conditions		
Initial denaturation	96°C for 1min	
(Denaturation)	96°C for 10 s	x25 cycles
(Annealing)	50°C for 5 s	x25 cycles
(Elongation)	60°C for 4min	x25 cycles
Temp	4°C hold	

Fosmid Sequencing reactions

DNA	12 µl
Big Dye	4 µl
5X Seq Buffer	2 µl
Primer (pCC1F/pCC1R), 5 pmol	1 µl
Sterile water	1 µl
Total	**20 µl**

Table 2.4: ABI-Fosmid-Cycle program

ABI cycle sequencing conditions		
Initial denaturation	96°C for 4min	
(Denaturation)	95°C for 30 s	x25 cycles
(Annealing)	59°C for 25 s	
(Elongation)	60°C for 4min	
Temp	20°C hold	

2.2.6.2 Gap closure using primer walking

The Sanger reads were incorporated into the Gap4 [which is part of the Staden package] (Staden, 1996) database and contigs representing the ends of the plasmid or fosmid were identified. To close the gaps of the individual clones and to identify middle contigs of the inserts if the gap was large, sequencing was performed using plasmid and fosmid DNA of individual clones as templates, which spanned the region containing the gaps using primers designed from the ends of the contigs. Staden selects the most suitable primer which usually comprises a 100 bp overlap. The consensus sequences obtained after gap closure for each clone, were used for the ORF predictions.

Figure 2.4: Template display of contig00098. The black arrows represent the vector sequences; red arrows block (1) represent Sanger sequencing from the vector. Green and orange arrows represent individual 454 reads in their respective orientation. The yellow block (2) represents the primer designed using the parameter of the Gap4 program, block (3) represents primer walking using the specific primer designed from that region.

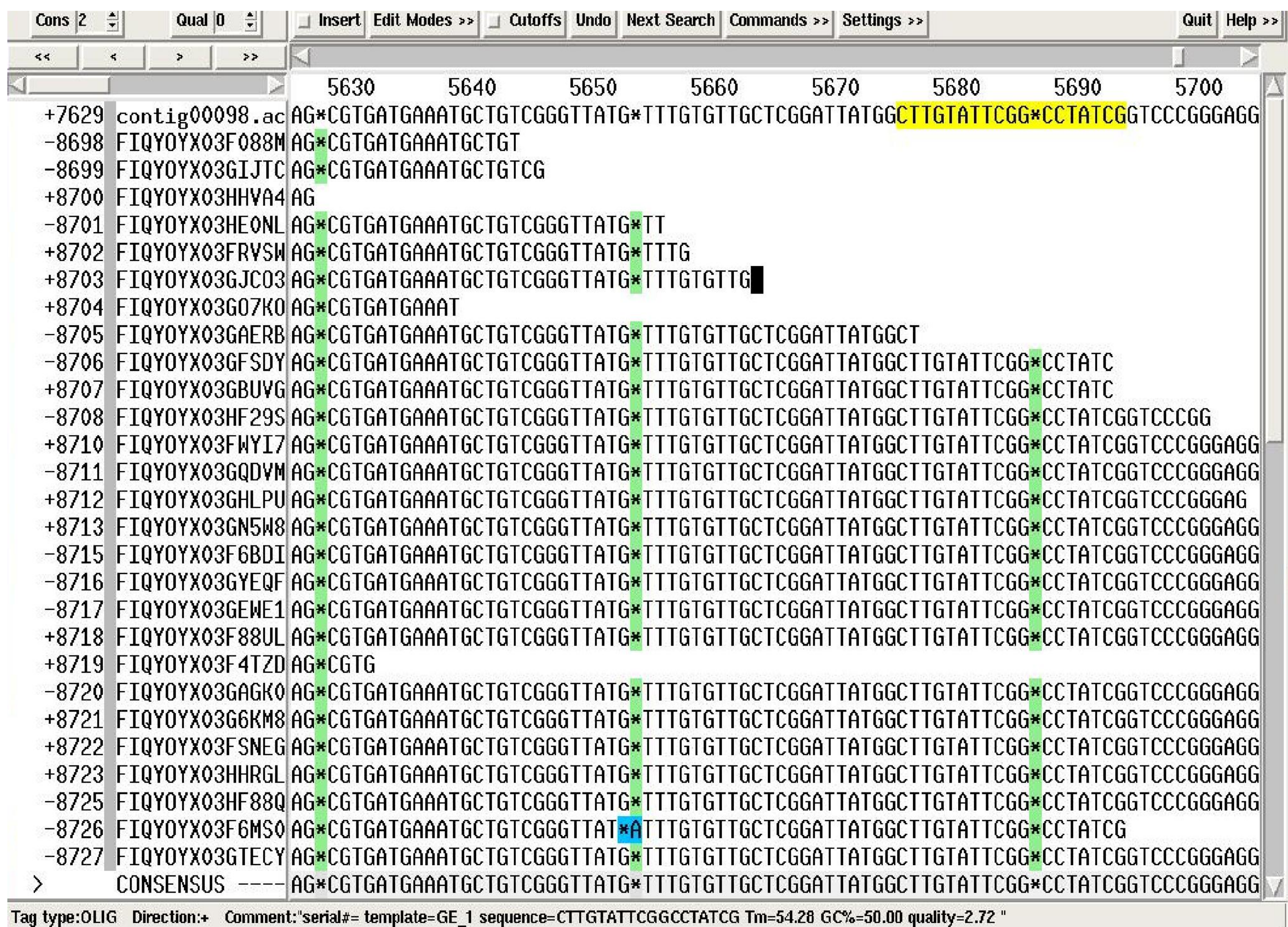


Figure 2.5: Assembled reads making up contig00098. The sequence highlighted in yellow represents the primer that was designed using Staden.

2.2.6.3 ORF predictions

Sequences were loaded into Artemis for manual checking and correction of open reading frames (ORFs) (Berriman and Rutherford, 2003). Artemis is a free genome viewer (Sanger Centre website: http://www.sanger.ac.uk/Software/Artemis/) and annotation tool that allows visualization of sequence features and the results of analyses within the context of the sequence, and its six-frame translation (Figure 2.6). ORFs were compared to known sequences deposited in the non-redundant protein databases of the National Centre for Biotechnology Information (NCBI, USA) using standard protein-protein BLAST (*blastp*) (Altschul *et al.,* 1997).

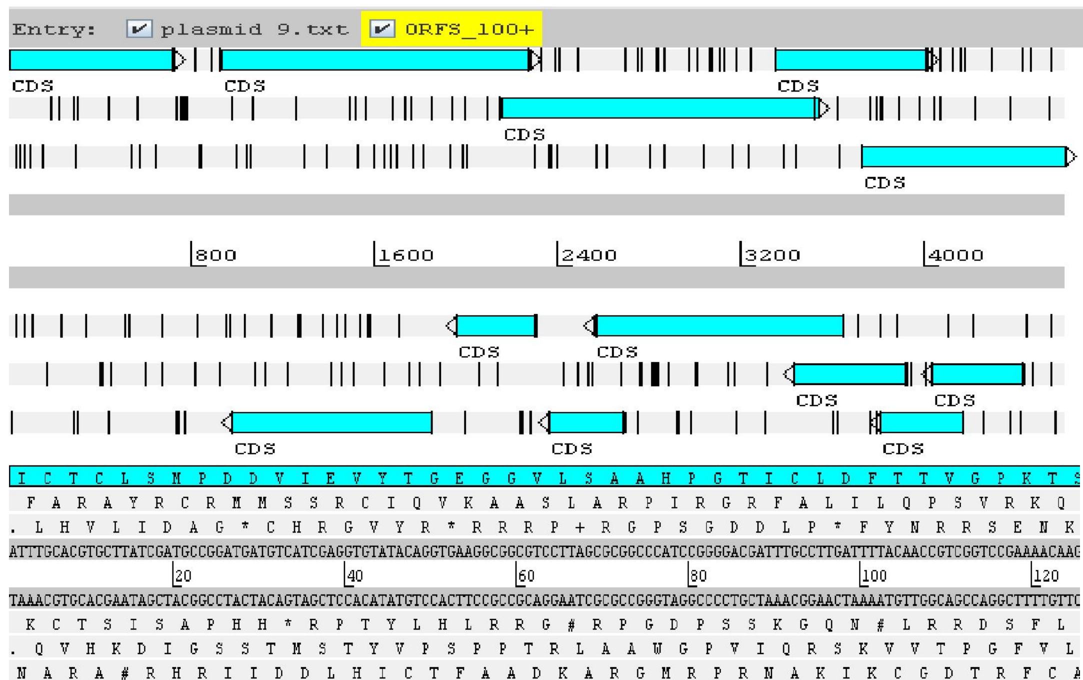


Figure 2.6: Diagramatic representation of ORFs predicted by Artemis in all six reading frames.

2.3. Results and Discussion

2.3.1 Genomic DNA extraction

DNA of high yield (39 µg/g cells) with minimal shearing was obtained using the hot phenol method.

2.3.2 Culture confirmation

Amplification of the 16S rRNA with universal bacterial primers yielded a PCR product of the expected size (1500 bp) (Figure 2.7).

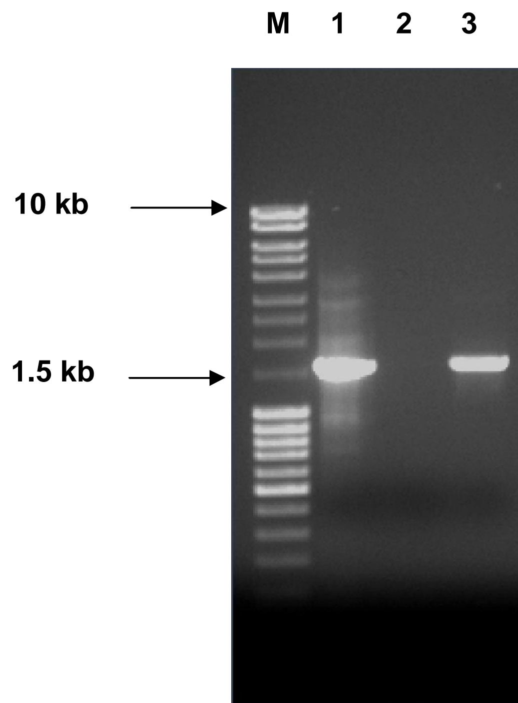


Figure 2.7: PCR amplification of the 16S rRNA gene from GE-7. M: MassRuler DNA Ladder (SM#0403-Fermentas); lane 1: positive control (*E. coli*); lane 2: negative control (sterile water); and lane 3: GE -7.

Restriction analysis of the digested 16S rRNA with *Eco*RI (855 bp and 663 bp), *Sma*I (1377 bp and 141 bp) and *Pst*I (850 bp and 668 bp) revealed fragment sizes of the expected lengths for GE-7.

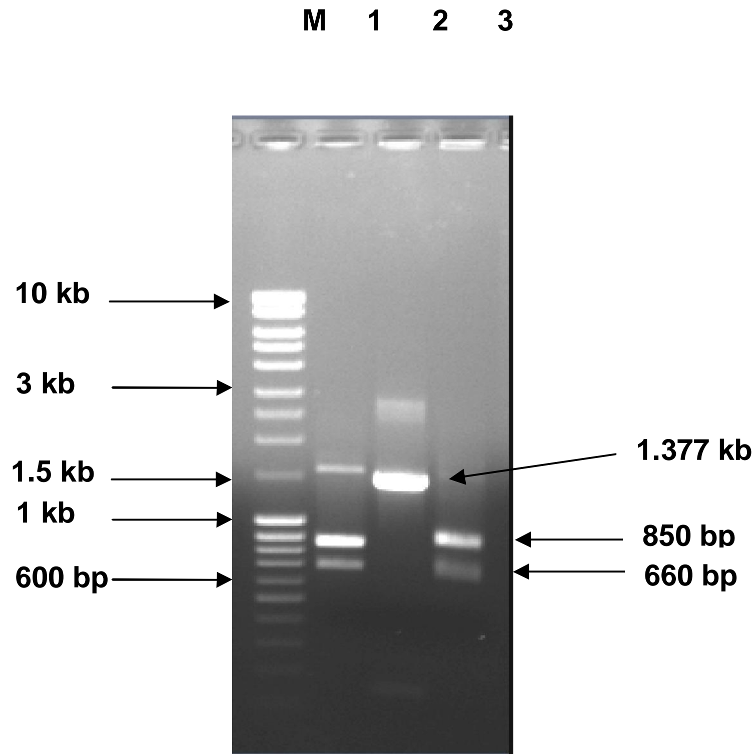


Figure 2.8: Restriction analysis of the 16S rRNA gene from GE-7. M: MassRuler DNA Ladder (SM#0403); lane 1: *Eco*RI digestion ; lane 2: *Sma*I digestion; and lane 3: *Pst*I digestion.

2.3.3 Small-insert library construction

2.3.3.1 Partial digestion of GE-7 genomic DNA

The gDNA was size fractionated in the size range of 2-6 kb. However, in order to obtain complete digestion, the 1 h digestion time recommended by the manufacturer was inadequate. Restriction times were increased to 3 h.

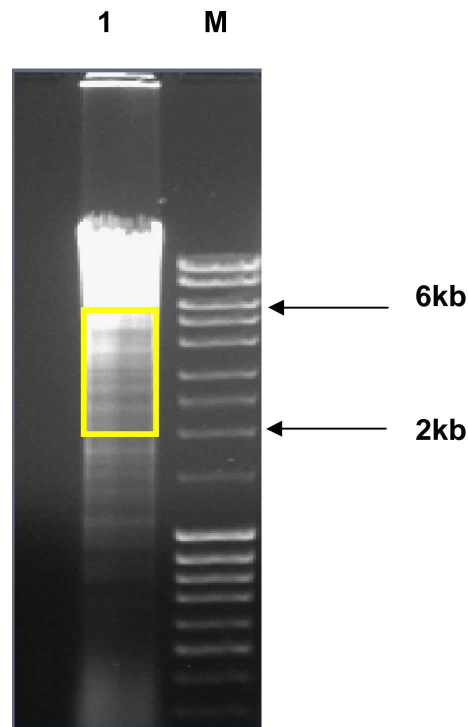


Figure 2.9: Partial digestion of GE-7 genomic DNA. Lane 1: *Xho*I digestion; and M: MassRuler DNA Ladder (SM# 0403).

2. 3.3.2 Physical fragmentation of GE-7 genomic DNA

A longer shearing time (2 min) and lower pressure (10 psi) yielded DNA in the appropriate size range 2-6 kb for the cloning. The excised DNA was purified, polished and treated with enzymes used in section 2.2.3.1.3.

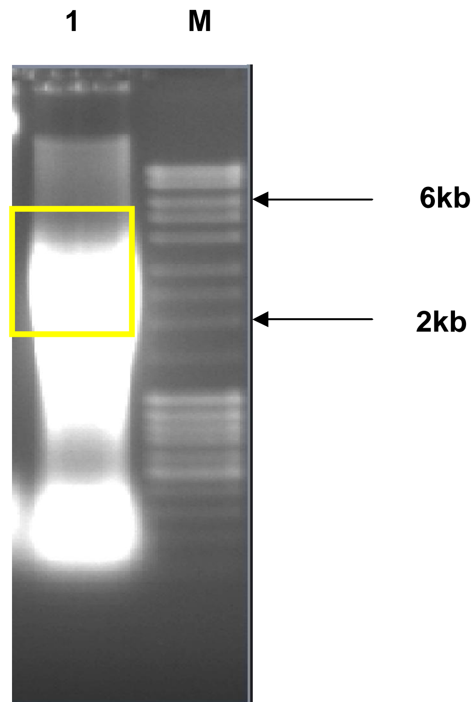


Figure 2.10: Fragmentation by nebulization. Lane 1: fragmented gDNA; and M: MassRuler DNA Ladder (SM# 0403).

2.3.3.3 Sticky-ended library

A total of ~1 600 clones were obtained after overnight ligation with the *Xho*I digested vector. Random clones (20) were inoculated into 5 ml LB broth containing 50 µg/ml kanamycin and the tubes were incubated overnight at 37°C with shaking. Plasmid DNA was then extracted from the clones and restriction analysis was performed using *Xho*I and *Hin*dIII to release the 3.3 kb plasmid backbone and determine the insert sizes.

With the sticky-ended library no background colonies of self-ligated vector was observed. According to the restriction analysis 69% of the inserts were 2-4 kb and the remaining 31% were >4-6 kb (Figure 2.11). The average insert size was 4 kb.

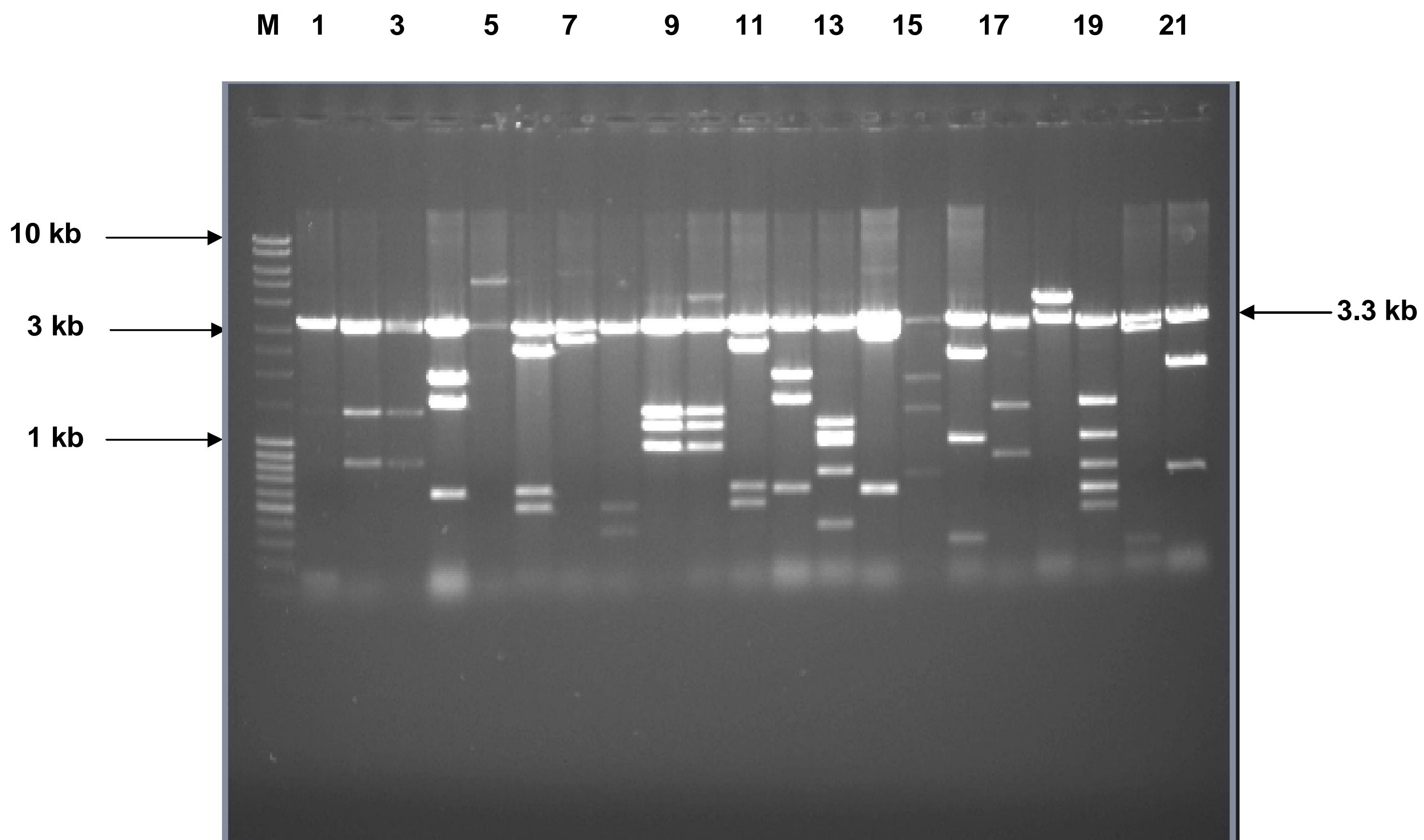


Figure 2.11: Restriction analysis of clones from sticky library. M: MassRuler DNA Ladder (SM#0403); and lanes 1-21: randomly selected clones.

2.3.3.4 Blunt-ended library

Few transformants (<50) were obtained after cloning the nebulized, repaired DNA. The control insert worked therefore we knew that the ligation and transformation was working. The only possible explanation was that probably the DNA was not efficiently end repaired thereby decreasing the transformation efficiency. There was no way of checking if this was the problem. BAL31 Nuclease removes between 100-300 bp of DNA during polishing; therefore running the repaired DNA on a gel is not conclusive evidence to suggest that the repair of the DNA was optimal.

We then decided to partially digest the DNA with *EcoR*V and clone into the *EcoR*V digested vector. For the blunt-ended cloning we decided to add 50% PEG 6000 to the ligation reactions. It was observed that the addition of PEG significantly enhanced the ligation

efficiency. For blunt-end cloning PEG is known to stimulate the formation of ligation products at very low insert and enzyme concentrations as well as accelerate the rate of ligation by 1-3 orders of magnitude (Sambrook *et al.,* 1989). We compared the library size of ligations performed with and without PEG and found that the addition of PEG greatly enhanced the number of clones obtained (>10 fold) and this was comparable to the number of clones obtained with the sticky-ended library.

Restriction analysis for the blunt-ended library revealed that 80% of the insert sizes were 2-4 kb and 20% were >4-6 kb (Figure 2.12).

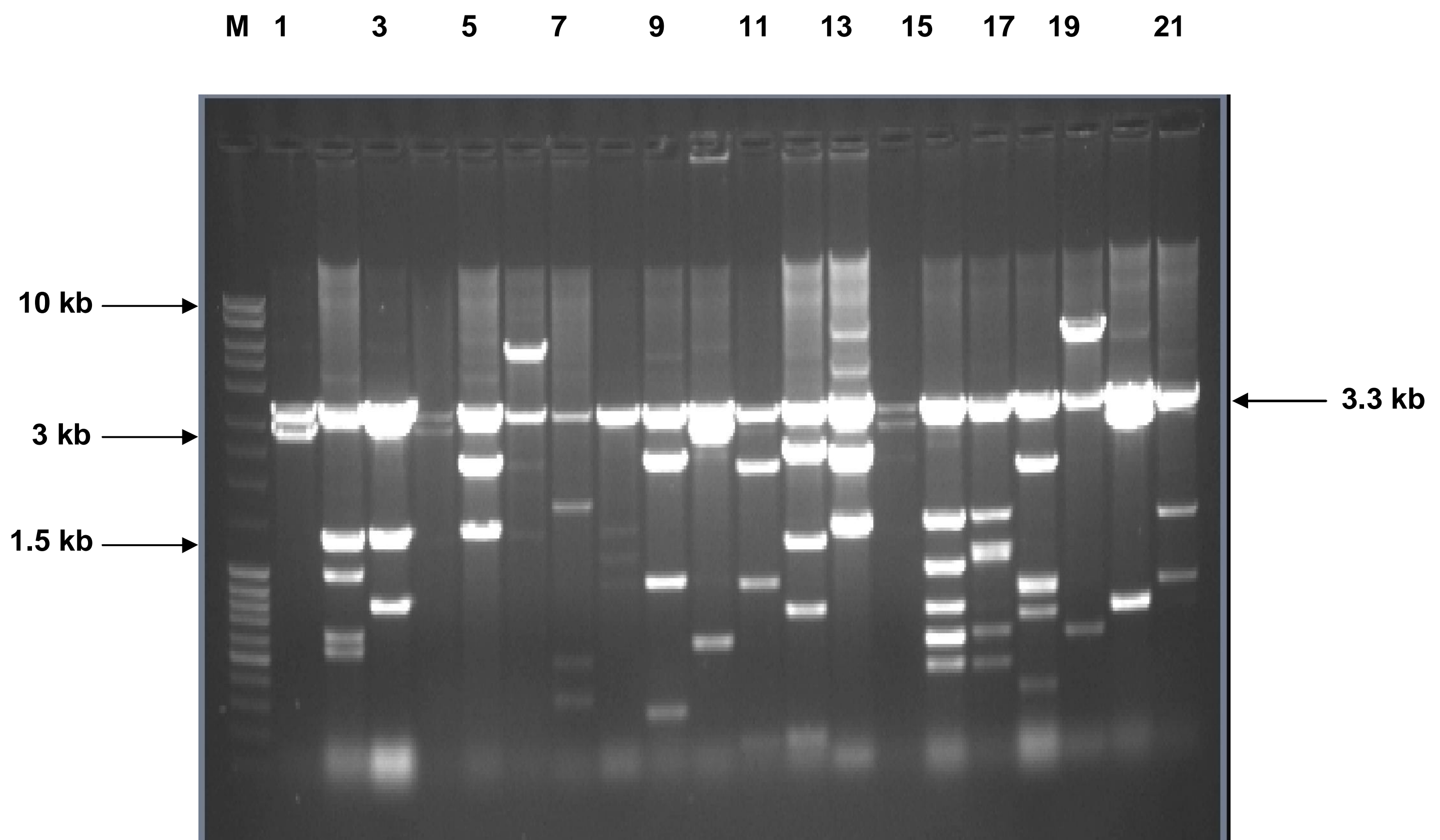


Figure 2.12: Restriction analysis of clones from blunt library. M-MassRuler DNA Ladder (SM#0403); and lanes 1-21: randomly selected clones.

2.3.4 Library screening (small-inserts)

Positive activity was observed for 9 clones (3 sticky + 6 blunt) when screened on LB

tributyrin plates (Figure 2.13). We found that even after the clones had been sub-cultured they were still active.

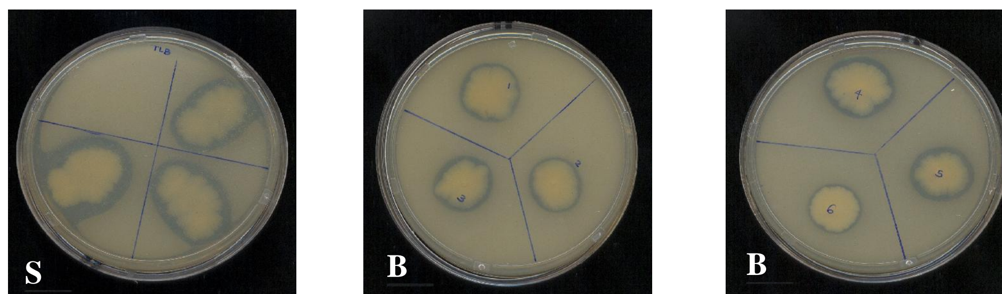


Figure 2.13: The 9 positive clones obtained when screened on LB tributyrin plates supplemented with kanamycin (50 µg/ml). S = clones from sticky library and B = clones from blunt library.

2.3.4.1 Confirmation of lipase activity

As mentioned previously, assaying on tributyrin does not distinguish lipase from esterase activity; therefore we decided to do a lipase PCR using specific primers designed to amplify and internal fragment of the lipA gene. With the amplification reactions only one clone (pGT1) yielded the expected 500 bp fragment (Figure 2.14). At this point we could not draw the conclusion that the PCR negative clones were not lipases, because GE-7 may contain more than one type of lipase for which no literature was available because the genome of GE-7 is not sequenced. We then decided to perform the second plate assay using LB olive oil and rhodamine B. Based on our observations, none of the remaining clones yielded positive results when viewed under UV light. From these results we were able to confirm that the majority of the clones that we had obtained possibly expressed esterase rather than lipase activity. However this had to be confirmed by sequencing.

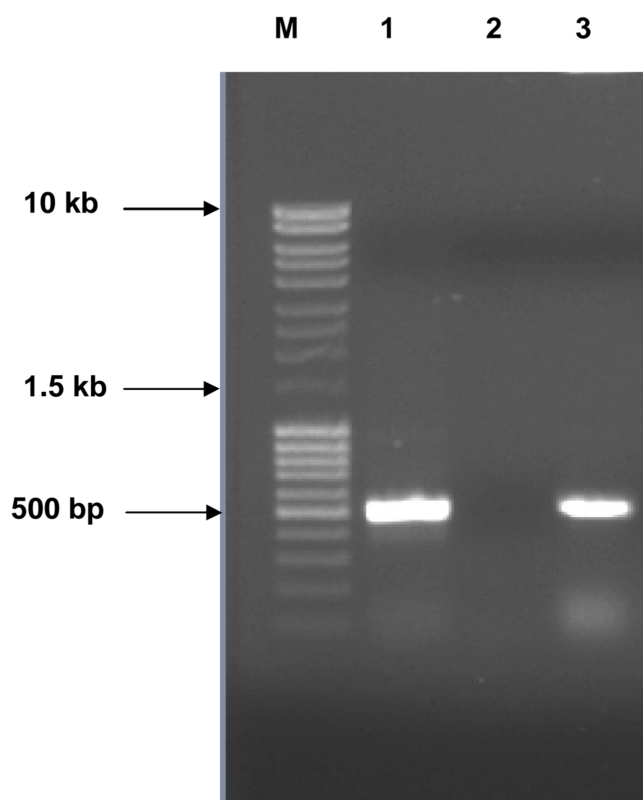


Figure 2.14: PCR amplification of the lipase gene from clones found positive by functional screening. M-MassRuler DNA Ladder (SM#0403); lane 1: positive control (*Geobacillus kaustophilus* HTA426 gDNA); lane 2: negative control (sterile water); and lane 3: positive clone.

2.3.5 Large-insert library construction

2.3.5.1 Digestion of GE-7 gDNA

DNA was size fractionated in the range of ~36 kb (Figure 2.15). The band of the appropriate size was excised from the gel and purified using the GFX kit (Amersham Biosciences).

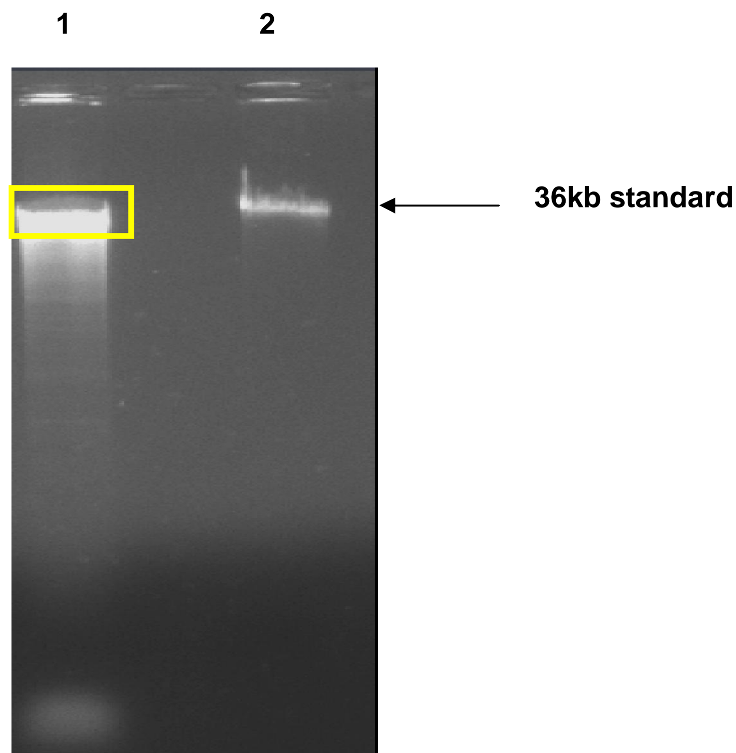


Figure 2.15: Partial digestion of gDNA with *Eco*RV. Lane 1: GE-7 gDNA; and lane 2: fosmid control standard (36 kb).

2.3.5.2 Fosmid library production

The efficiency of phage packaging and infection was compared with that of electroporation. We observed a 10-fold increase in the number of clones obtained with electroporation when compared with infection. It was important to obtain a large library and screen many clones in order to ensure that the library was representative of the genome of this organism. In order to determine the number of clones that had to be screened to obtain a positive hit, the following formula was employed (Gabor *et al.,* 2004):

$$N_P = \frac{\ln(1-P)}{\ln\left(1-\frac{n_{active}}{n_{total}}\right)} = \frac{\ln(1-P)}{\ln\left(1-\frac{(I-X)}{c \cdot G}\right)}$$

Np = number of clones that need to be screened

P = probability that a target gene will be recovered at least once (95% probability)

I = insert size (40 kb)

X = size of the gene of interest (500 bp)

G = genome size (3 Mb)

c = correction factor (a correction factor of **1** was selected – assuming independent expression)

According to the formula the number of fosmid clones that needed to be screened was 693 using an insert size of 40 kb, however we had only obtained 285 clones with infection. Therefore, it seemed as if electroporation was producing the desired library size. This correlated with previous studies that indicated that high transformation efficiencies were obtained using electroporation especially when cloning larger inserts (>100 kb) (Sheng *et al.,* 1995). Fosmid DNA extractions were done on random clones and restriction analysis was performed to determine insert sizes. We found that the average insert sizes observed with the packaging and infection reactions were ±35 kb, and that the fosmid backbone that was recovered was 1.3 kb larger than the expected size (Figure 2.16).

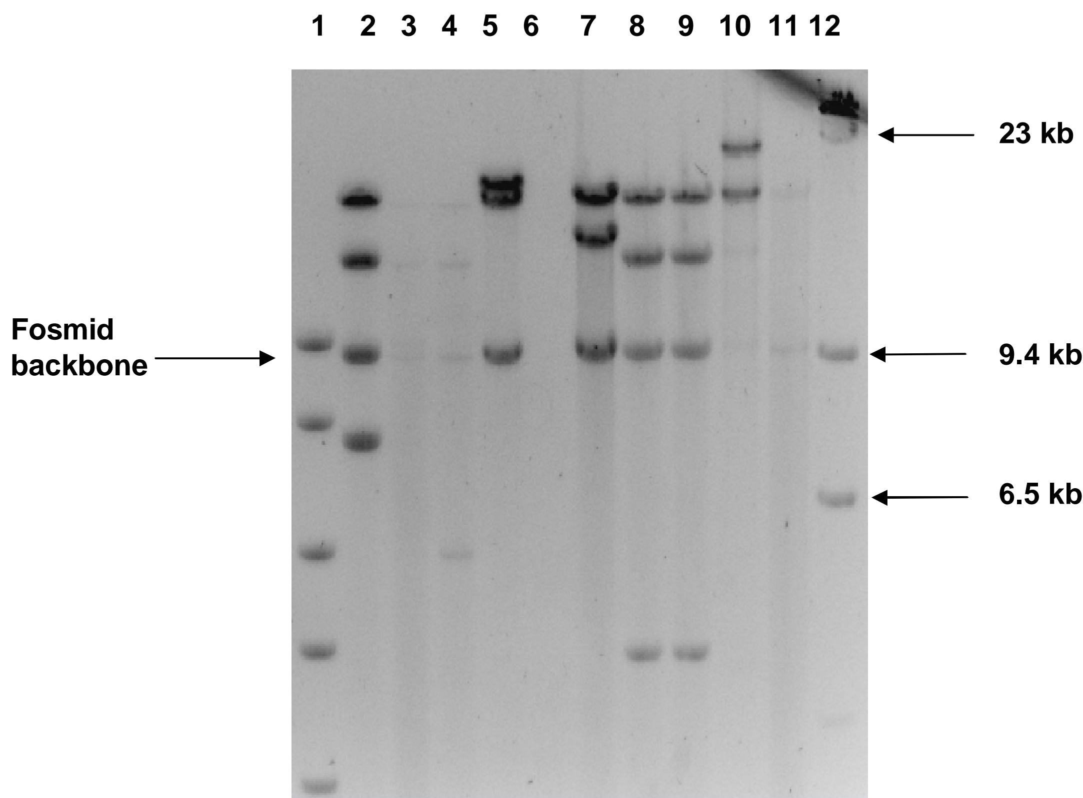


Figure 2.16: Restriction digests of fosmid clones. Lane 1: MassRuler 1kb ladder; lanes 2-11: fosmid clones: and lane 12: Marker III (Lambda DNA digested with *Hin*dIII.

With the transformation reactions (electroporation) we found that less than 50% of the colonies that we selected contained inserts (Figure 2.17) and the insert sizes ranged from ~4-10 kb which was significantly smaller when compared to those obtained by packaging and infection. However, with electroporation the correct fosmid backbone size was obtained. Based on this we suspected that there was something present in the packaging extract that contributed to the larger backbone size. According to Hohn and Murray (1977) when working with a two component packaging system as in this case, it is possible that native phage DNA that is present in the extracts either competes with the foreign DNA or is packaged together with the foreign DNA. The other possibility is that the native DNA was integrated into the vector because the fosmid vector has features similar to that of the lambda phage vector.

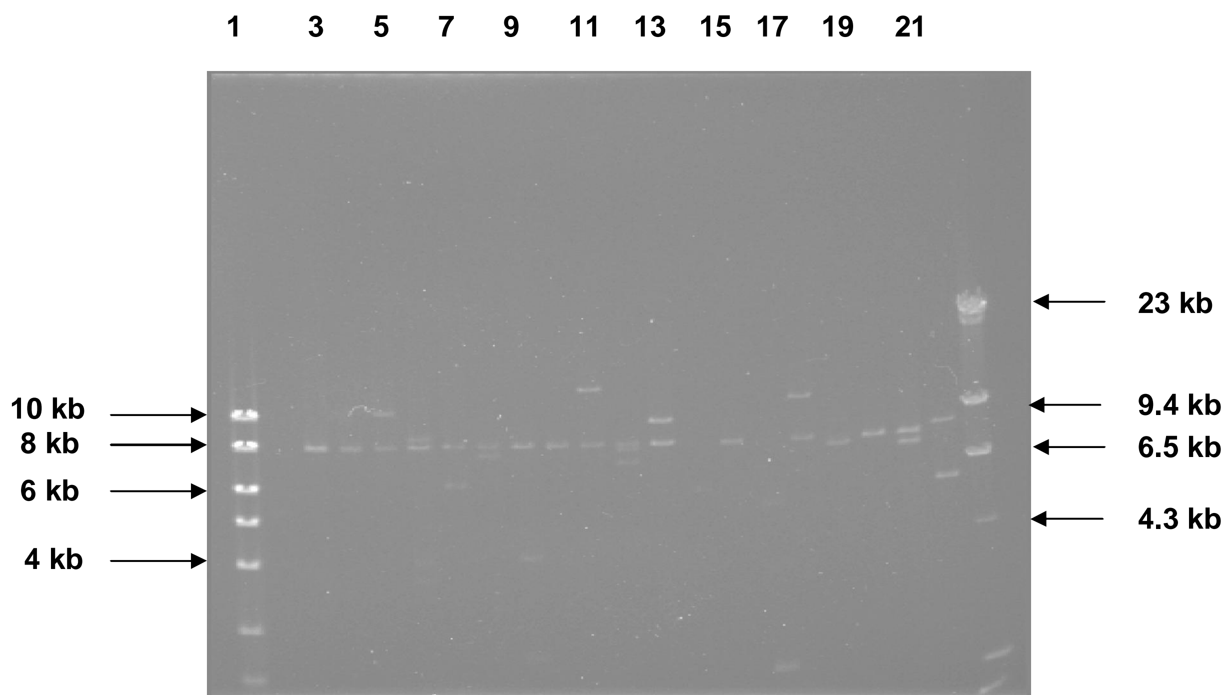


Figure 2.17: Restriction digests of transformed (electroporation) clones. Lane 1: MassRuler 1kb ladder; lanes 2-21: fosmid clones: and lane 22: Marker III (Lambda DNA digested with *Hin*dIII).

2.3.5.3 Screening of Fosmid library

Nine clones (1 clone obtained by infection and 8 clones obtained by transformation) displaying activity were detected on tributyrin plates after 24-36 h incubation at 37°C (Figure 2.18). Of the nine clones, 5 carried the same insert as revealed by end sequencing of the individual clones. One clone (pGTFos3) yielded a positive amplification product (500 bp) with the lipase primers (Figure 2.19 a). The presence of the lipase was confirmed by screening the clones on LB olive oil containing rhodamine B for which a positive result was observed for pGTFos3 (Figure 2.19 b) when viewed under UV light, this was consistent with the PCR results. Clones showing no activity on LB olive oil and rhodamine B plates possibly display esterase activity; however this had to be determined by sequence analysis of the active clones.

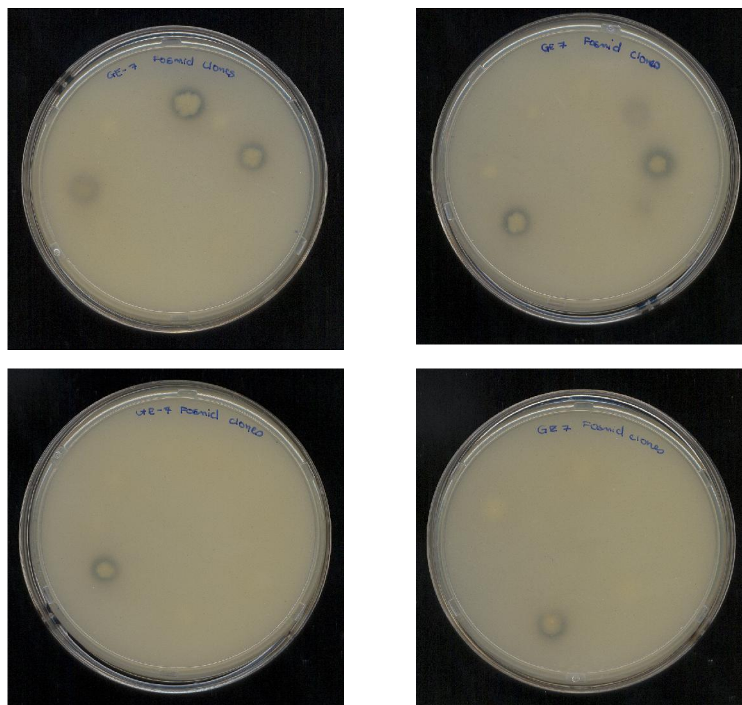


Figure 2.18: The nine positive clones obtained when screened on LB tributyrin plates supplemented with chloramphenicol (12.5 μg/ml).

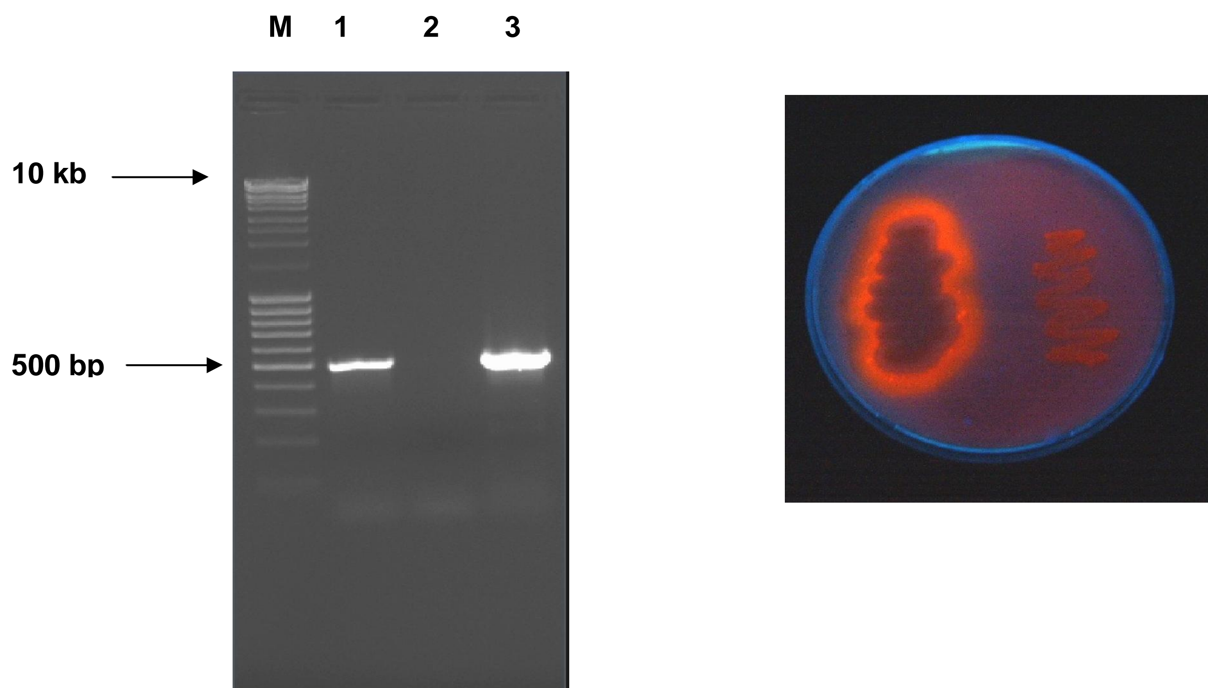


Figure 2.19: (a) PCR amplification of the lipase gene from clones found positive by functional screening. M-MassRuler DNA Ladder (SM#0403); lane 1: positive control (*Geobacillus kaustophilus* gDNA); lane 2: negative control (sterile water); and lane 3: positive clone.

Figure 2.19: (b) A lipase positive clone obtained from the infection library when screened on LB olive oil containing rhodamine B and supplemented with chloramphenicol (12.5 µg/ml). A clone showing no lipase activity was used as the control.

2.3.6 Ultrafast sequencing of GE-7 fosmid and plasmid clones using the Genome sequencer 20 FLX system

The assembly analysis of the FLX 454 sequencing data after assembly using the Newbler software is represented in Table 2.5.

Table 2.5: Assembly results using the Newbler software

***Large Contig Metrices* (>500 bp)**	
Number of contigs	32
Number of bases	76745
Average contig size	2398
N50 contig size	4623
Largest contig size	10843
Q40 plus bases	97.43%
Q40 minus bases	2.57%
All Contig Metrices	
Number of contigs	745
Number of bases	231056

BLASTP results revealed that the predicted proteins had closest homologs in *Geobacillus* (Table 2.6). Genes associated with fatty acid degradation, different glycolytic activities, lipolytic activity, spore germination, proper protein folding, antibiotic resistance and the cell wall were identified. The sequencing results confirmed the screening results as clone pGT1, pGTFos1 and pGTFos3 were shown to harbour lipolytic genes. As expected pGT1 and pGTFos3 contained the lipA gene because amplification products were obtained when screened with primers designed internal to the lipA. The lipA gene was PCR amplified from *G. thermoleovorans* YN, a bacterium isolated from desert soil in Egypt, and biochemical analysis of the 43 kDa protein revealed the ability of this enzyme to function above 60°C and at an alkaline pH [9-9.5] (Soliman *et al.,* 2007). Studies conducted at our institution on the lipA from GE-7 revealed the ability of this enzyme to function at an optimum temperature of 65°C and pH 9-10. In addition, the enzyme was stable for 1 h at 70°C and 23 h at 65°C (Barnard, 2005). This highlights the potential of this enzyme as a valuable candidate for application in the detergent, oil and fat, dairy and pharmaceutical industries. Furthermore, running bioprocesses at elevated temperatures will lead to higher diffusion rates, increased solubility of lipids and other hydrophobic substrates in water, and the reduced risk of contamination (Leow *et al.,* 2007). Clone pGTFos1 was shown to harbour a carboxylesterase. Esterases have potential use in the production of chiral drugs because

they are able capable of enantioselective hydrolysis and esterification in a cheap and environmentally friendly way (Maqbool *et al.,* 2006).

The second lipase present in clone pGTFos3 was shown to belong to a new subfamily of lipolytic enzymes, the GDSL family. This family displays the characteristic G-D-S-L motif instead of the conventional G-X-S-X-G (X- denotes any amino acid) motif. GDSL lipases are hydrolytic enzymes with multifunctional properties such as broad substrate specificity and regiospecificity. They have potential for use in the hydrolysis and synthesis of ester compounds that are of interest in pharmaceutical, food, biochemical and the biological sector (Akoh *et al.,* 2004; Lämmle *et al.,* 2007). Although the focus of this part of the study was to identify lipases in the GE-7 libraries, sequencing the clones enabled us to have an undersyanding some of the genes that flank the lipase ORF in this organism since its genome has not yet been sequenced. In addition, we were able to co-relate some of the identified proteins to the environment from which GE-7 was isolated. This part of the study also highlighted the potential of obtaining false positives during functional-based screening. The reason for this is not clear and further reference will be made to this in section 5.6.1.4. Therefore it was important to sequence the inserts of the active clones in order to determine which genes were responsible for the observed phenotype on the activity plate. The presence of false positives when screening with TLB was also observed with the metagenome. In the future it may be feasible to implement an alternative media that can be used in conjunction with TLB for screening metagenomic libraries for lipolytic activity thereby reducing the number of false positives that could be obtained

In the deep mine the presence of organic acids were detected and thought to be important for bacterial survival (Feng *et al.,* 2007; Onstott *et al.,* 2006). We identified a putative transporter gene for benzoate (aromatic carboxylic acid). Like thermophilic bacteria, GE-7 responds to heat induced stresses by induction of heat shock proteins, which remove or refold damaged proteins. A gene encoding the ATP-dependant heat shock responsive protease Clp was identified in one of the clones (Shih and Lai, 2007). Relating to survival in harsh environments, genes associated with spores (spore germination protein GerKB and transglutiminase [associated with the spore coat]) were identified. The presence of endospores in *Bacillus* related species have allowed these organisms to survive in a wide range of environments, pH 2-12; temperatures between 5 and 78°C; salinity from 0 to 30%; and pressures from 0.1 MPa to 30 MPa (Takami *et al.,* 2004). Likewise, according to

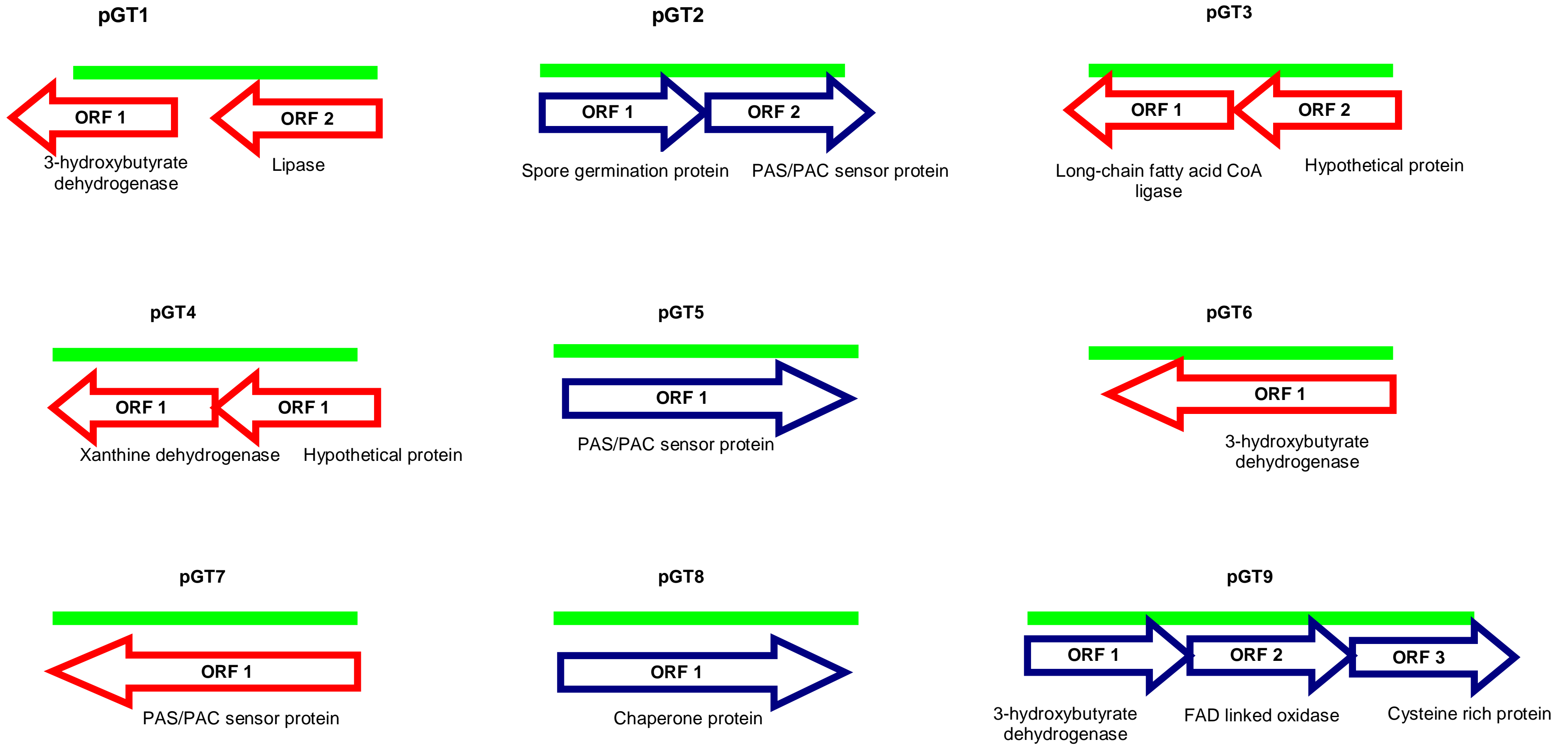
DeFlaun *et al.* (2007), GE-7 grows optimally at 65°C, pH 6.5 and salinity of 0 to 21% NaCl concentrations. In addition, a beta-lactamase gene was also identified in one of the clones. According to Rhazi *et al* (1999), beta-lactamases are capable of hydrolyzing ester bonds; therefore it was not unusual to detect it using a substrate that detects lipolytic activity. In addition a beta-lactamase gene flanking an esterase ORF was also identified while screening a metagenome library for lipolytic activity on TLB, the authors suggested that the metagenome derived esterase was part of ancient antibiotic resistance cluster and this was supported by the ability of the beta-lactamase to hydrolyze DD-peptidases and ester bonds (Eland *et al.,* 2006).

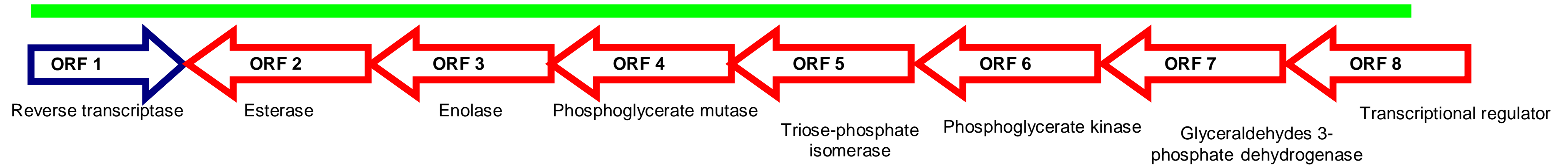
2.4 Conclusion

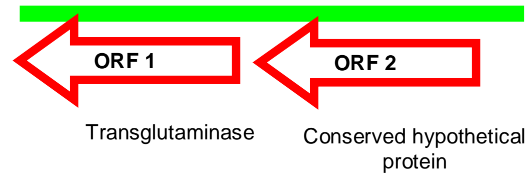
Most biocatalysts are inherently labile, and their stability during processing is of utmost importance (Leow *et al.,* 2007). Thermophiles produce unique biocatalysts that function under extreme conditions comparable to those prevailing in various industrial processes and are therefore of great interest for industrial applications. Amongst these enzymes, lipases are the most versatile (Li and Zhang, 2005). Lipases or acylglycerol hydrolases are enzymes that catalyze the hydrolysis of long chain triglycerides with the formation of diacylglycerides, monoglycerides, glycerol and free fatty acids (Leow *et al.,* 2003). Many lipases from animals, plants and microorganisms have been purified, characterized and cloned for industrial uses, such as digestive food aids, food additives for flavour, reagents for the synthesis of useful compounds and treatment of domestic sewage (Cho *et al.,* 2000). Microbial extracellular lipases are usually more thermostable than animal or plant lipases therefore they received much more attention for their potential use in industries (Leow *et al.,* 2003).

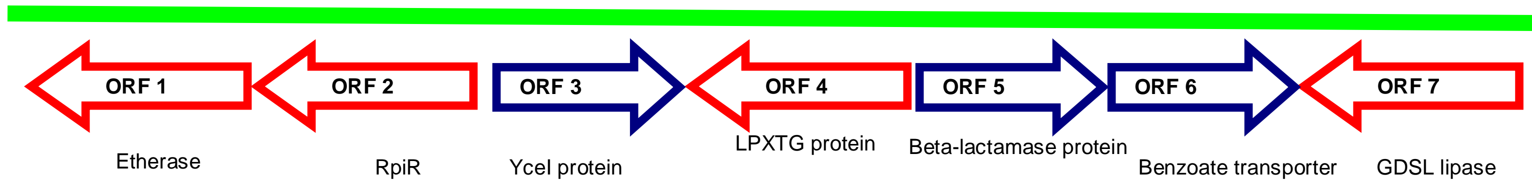
Apart from identifying lipases in our sequence data, we also attempted to discuss some of the genes that flank the lipase ORF. We found that many of the genes identified contribute to the survival of this organism in the extreme environment from which it was identified. In addition to the lipA gene, we were also able to identify a second lipase that falls under a different subfamily (GDSL). It was interesting to observe that from the limited number of clones and sequence data that we had obtained we were able to identify different lipases. The number of different lipases present in GE-7 remains unknown. This can be determined by whole genome sequencing of this organism using the 454 technology, or creating more fosmid libraries and sequencing the TLB active clones. However, the latter has its

drawbacks, in that in order to handle a large number of clones high-throughput screening strategies that employ robotics are required. On the other hand, the 454 technology has the advantage of providing a wealth of data in a shorter period of time as compared to primer walking of individual fosmid and plasmid clones and this could be a future endeavour in our lab.


pGT1
ORF 1
ORF 2
3-hydroxybutyrate dehydrogenase
Lipase
pGT2
ORF 1
ORF 2
Spore germination protein
PAS/PAC sensor protein
pGT3
ORF 1
ORF 2
Long-chain fatty acid CoA ligase
Hypothetical protein
pGT4
ORF 1
ORF 1
Xanthine dehydrogenase
Hypothetical protein
pGT5
ORF 1
PAS/PAC sensor protein
pGT6
ORF 1
3-hydroxybutyrate dehydrogenase
pGT7
ORF 1
PAS/PAC sensor protein
pGT8
ORF 1
Chaperone protein
pGT9
ORF 1
ORF 2
ORF 3
3-hydroxybutyrate dehydrogenase
FAD linked oxidase
Cysteine rich protein







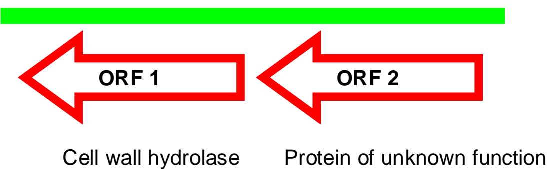


Figure 2.20: Linear representation of ORFs present in plasmid and fosmid clones.

Table 2.6: Lipolytic activity conferring plasmids and fosmids and sequence similarities

Plasmid/Fosmid		Function	Organism	E value	Accession No.	% Similarity
pGT1	ORF 1	3-hydroxybutyrate dehydrogenase	*Geobacillus thermodenitrificans* NG80-2	5e-40	YP_001125984	85
	ORF 2	Lipase	*Geobacillus thermoleovorans*	1e-90	CAL36912	98
pGT2	ORF 1	Spore germination protein GerKB	*Geobacillus thermodenitrificans* NG80-2	5e-162	YP_001125588	77
	ORF2	Putative PAS/PAC sensor protein	*Geobacillus sp.* Y412MC61	0.0	ZP_03558509	94
pGT3	ORF1	Xanthine dehydrogenase	*Geobacillus sp.* Y412MC61	0.0	ZP_03557275	96
pGT4	ORF1	Long-chain fatty-acid-CoA ligase	*Geobacillus kaustophilus* HTA426	8e-72	YP_147344	98
pGT5	ORF1	Putative PAS/PAC sensor protein	*Geobacillus sp.* Y412MC61	0.0	ZP_03558509	94
pGT6	ORF1	3-hydroxyisobutyrate dehydrogenase	*Geobacillus sp.* Y412MC61	1e-118	ZP_03558757	100
pGT7	ORF1	Putative PAS/PAC sensor protein	*Geobacillus sp.* Y412MC61	0.0	ZP_03558509	94
pGT8	ORF 1	ATP-dependent chaperone ClpB	*Geobacillus sp.* Y412MC61	0.0	ZP_03557880	96
pGT9	ORF1	3-hydroxyisobutyrate dehydrogenase	*Geobacillus sp.* Y412MC61	1e-118	ZP_03558757	100
	ORF2	FAD linked oxidase domain protein	*Geobacillus sp.* Y412MC61	0.0	ZP_03558756	100
	ORF3	Cysteine-rich region domain protein	*Geobacillus sp.* Y412MC61	0.0	ZP_03558755	99
pGTFos1	ORF1	Reverse transcriptase	*Geobacillus sp.* WCH70	6e-73	ZP_02912279	99
	ORF2	Carboxylesterase	*Geobacillus sp.* Y412MC61	0.0	ZP_03556921	95
	ORF3	Enolase	*Geobacillus kaustophilus* HTA426	0.0	YP_148907	100
	ORF4	Phosphoglyceromutase	*Geobacillus kaustophilus* HTA426	0.0	YP_148908	100
	ORF 5	Triose-phosphate isomerase	*Geobacillus kaustophilus* HTA426	4e-138	YP_148909	99

	ORF 6	Phosphoglycerate kinase	*Geobacillus kaustophilus* HTA426	0.0	YP_148910	99
	ORF 7	Glyceraldehyde-3-phosphate dehydrogenase	*Geobacillus kaustophilus* HTA426	1e-176	YP_148911	99
	ORF 8	Transcriptional regulator (central glycolytic gene regulator)	*Geobacillus kaustophilus* HTA426	0.0	YP_148912	97
pGTFos 2	ORF1	Transglutaminase	*Geobacillus kaustophilus* HTA426	2e-132	YP_148767	99
	ORF2	Conserved hypothetical protein	*Anoxybacillus flavithermus*	0.0	YP_002316710	97
pGTFos3	ORF1	Etherase	*Geobacillus thermodenitrificans* NG80-2	2e-162	YP_001125981	95
	ORF2	RpiR family transcriptional regulator	*Geobacillus thermodenitrificans* NG80-2	7e-154	YP_001125982	92
	ORF3	Ycel family	*Geobacillus sp.* Y412MC61	2e-95	ZP_03559104	98
	ORF4	Lipase	*Geobacillus thermoleovorans*	0.0	AF134840	98
	ORF5	LPXTG cell wall protein	*Geobacillus sp.* Y412MC61	4e-125	ZP_03148398	79
	ORF6	Metallo beta-lactamase protein	*Geobacillus thermodenitrificans* NG80-2	2e-162	YP_001125999	85
	ORF7	Putative benzoate transporter	*Geobacillus thermodenitrificans* NG80-2	3e-154	YP_001125991	87
	ORF 8	Lipolytic protein- GDSL lipase	*Geobacillus sp.* Y412MC61	4e-125	ZP_03148398	79
pGTFos4	ORF1	Cell wall-associated hydrolase	*Bacillus megaterium*	4e-50	AAO52795	95
	ORF2	Protein of unknown function UPF0236	*Geobacillus sp.* WCH70	0.0	ZP_02913434	96

Chapter 3

What lies beneath? Assessment of diversity of a Beatrix mine biofilm

3.1 Introduction

The overwhelming majority of life on earth is microbial, both in terms of the number of organisms and phylogenetic diversity (Bourret, 2006). Microorganisms are extremely difficult to study in nature, owing to their small size and morphological simplicity. Although culturing techniques have been used to analyze microorganisms, a great majority remain uncultured in the laboratory (Rondon *et al.,* 2000). According to Chaudhuri *et al.* (2006) only 1 to 4% of microbes can be cultured. One of the most important challenges to microbial ecologists is to elucidate the physiological properties, energy conservation pathways, and the ecological significance of recently discovered, uncultured microorganisms (Elshahed *et al.,* 2005). The rapidly expanding field of metagenomics has facilitated the understanding and accessibility of microbial genomes and their functions (Hårdeman and Sjöling, 2007). Metagenomic research has been useful in exploiting the unknown bacterial diversity (Pontes *et al.,* 2007). The quality of metagenomic analysis is directly correlated to the quality of DNA obtained, and several extraction methods have been developed for the extraction of environmental DNA (Green and Keller, 2006). However, there is no universal method of extraction; the type of method selected depends on the sample being analyzed (Chaudhuri *et al.,* 2006). The differences in cell wall structures, adhesion of the microorganisms to the sample matrix as well as physical, chemical and biological characteristics of the sample affect the extraction procedure (Niemi *et al.,* 2001). Luna *et al.* (2006) observed that different DNA extraction methods can affect microbial diversity estimates.

The fundamental unit of biological diversity and the basis of taxonomic hierarchy is the species. However, knowledge about natural bacterial diversity is still limited To date, the total number of validly published bacterial species names is 9300 of which about 1541 are synonyms belonging to 1786 different named genera of which about 97 are synonyms (http://www.dsmz.de/microorganisms/main.php?contentleft_id=14). Due to the advent of molecular techniques, the full extent of molecular diversity can now be surveyed (Fierer *et al.,* 2007). Woese (1987) proposed a phylogenetic classification system based on the

divergence of small subunit ribosomal RNA sequences (16S rRNA for bacteria and archaea [prokaryotes] and 18S rRNA for eukaryotes). Methods based on direct PCR amplification and analysis of rRNA genes were developed and allowed a more comprehensive analysis of microbial communities (Boon *et al.,* 2002). 16S rRNA sequence analysis has been used as a tool for assessing the genetic diversity of environmental samples (Baker *et al.,* 2003). Ribosomal RNA genes can occur in variable numbers in different organisms, they are unaffected by horizontal gene transfer thereby making these genes suitable for studies on microbial evolution and phylogeny (Amann and Ludwig, 2000).

More recently, techniques dependent on DNA melting behaviour have been developed for analyzing the dynamics of microbial populations in the environment (Liu *et al.,* 1997). One of the techniques is denaturing gradient gel electrophoresis (DGGE), which allows for the separation of DNA fragments of the same length according to the differences in nucleotide sequences (Watanabe *et al.,* 2001). Many researchers use DGGE to study microbial communities in various environments such as biofilms and activated sludge (Zhang and Fang, 2000). DGGE has been used to separate 16S rRNA fragments, which are directly amplified from community DNA by PCR using universal primers (Watanabe *et al.,* 2001).

However, sequences obtained from DGGE bands are short (~200 bp). The shorter the sequence of the DGGE fragments, the less refined the phylogenetic inference. Sequencing of DGGE bands is only sufficient to determine broad phylogenetic affiliations but is inadequate to perform a precise phylogenetic analysis (Dìez *et al.,* 2001). Sequencing of the entire 16S rRNA gene is useful for understanding phylogenetic relationships among prokaryotes above the species level (Pontes *et al.,* 2007).

South African mines provide ready access to some of the world's deepest terrestrial extreme environments (Pfiffner *et al.,* 2006). Life in extreme environments has been studied by intensively focusing attention on the diversity of organisms and molecular and regulatory mechanisms involved (Satyanarayana *et al.,* 2005). A motile, spore formimg, sulfate-reducing, chemoautotrophic thermophile was isolated from a South African gold mine. The organism, *Candidatus Desulforudis audaxviator*, was shown to be capable of fixing its own nitrogen and carbon by using machinery shared with archae (Chivian *et al.,* 2008). The diversity of sulfate-reducing bacteria associated with South African gold mine boreholes has also been characterized. Sequencing of the Dissimilarity sulphite reductase (Dsr) and the

16S rRNA amplified from DNA extracted from the water sample suggested that gram positive sulfate reducing bacteria are widely distributed in the deep mine (Baker *et al.,* 2003). In this study we examined the diversity (in terms of 16S and 18S rRNA analysis as well as DGGE analysis for only the domain prokarya) of a biofilm collected from the Beatrix gold mine in the Northern Free State region. Beatrix is located in the Witwatersrand Basin, and the mines in this region follow reefs from shallow levels to great depths (Pfiffner *et al.,* 2006). The biofilm used was collected from a depth of 808.5 m and the temperature, pH and conductivity readings that were taken on site were as follows; 31.3°C, 7.31 and 5.03 mS/cm.

3.2 Materials and Methods

3.2.1 Environmental DNA extraction

Three methods of DNA extraction were evaluated. The methods were (i) a combination of chemical and enzymatic lysis (Zhou *et al.,* 1996; Bond *et al*., 2000); (ii) a pretreatment with aluminium sulfate prior to physical lysis (Dong *et al.,* 2006; Labuschagne and Albertyn, 2007; Towner, 1991) and (iii) the FastDNA® SPIN Kit (Qbiogene). Following extraction, the DNA was quantified using a NanoDrop ND-1000 spectrophotometer (Inqaba Biotech).

3.2.1.1 Method 1

To 5 g of biofilm, 13.5 ml of DNA extraction buffer (100 mM Tris-HCl (pH 8); 100 mM Na_2EDTA (pH8); 100 mM sodium phosphate (pH 8); 1.5 ml NaCl; 1% CTAB) was added and mixed. This was followed by the addition of 100 µl Proteinase K [10 mg/ml] (Sigma). The sample was incubated for 30 min at 37°C in an orbital shaker. After incubation 1.5 ml of 10% SDS was added and the sample was incubated for 2 h at 65°C and gently mixed by end-over-end inversion every 15 min. Following incubation, the sample was subjected to 3 cycles of freeze-thaw lysis, freezing in liquid nitrogen and thawing at 65°C. Following freeze-thaw lysis the sample was centrifuged at 6 000 rpm for 10 min at room temperature. The supernatant was transferred to a new tube; the pellet was resuspended in 4.5 ml of extraction buffer and 0.5 ml of 20% SDS. The sample was incubated at 65°C for 20 min, and centrifuged at 6 000 rpm for 10 min at room temperature. This extraction step was done twice. To the pooled supernatants an equal volume of chloroform-isoamylalcohol (24:1) was added. The sample was centrifuged at 6 000 rpm for 10 min at room temperature; and the aqueous (upper) phase was transferred to a new tube. Isopropanol (0.6 volume) was then added, and the sample was incubated overnight at -20°C for the precipitation of the nucleic

acids. Following overnight incubation, the sample was centrifuged at 9 500 rpm for 20 min at 4°C. The supernatant was discarded and the pellet was washed with 70% ethanol. This was followed by centrifugation at 9 500 rpm for 20 min at 4°C. The supernatant was discarded and the pellet was resuspended in 2-4 ml sterile water. The volume used for resuspension depended on the size of the pellet.

3.2.1.2 Method 2

Prior to DNA extraction, the sample was treated with aluminum sulfate according to Dong *et al.* (2006) to remove inhibitors that may have been present. The treatment involved mixing 2 g of the biofilm sample with 300 µl of phosphate buffer (0.1 M NaH_2PO_4-$NaHPO_4$, pH 6.6), followed by the addition of 100 µl of 100 mM aluminum sulfate. The pH of the sample was adjusted to 8 by the addition of 1 M sodium hydroxide (Merck). Following treatment, DNA extraction was performed according to Labushagne and Albertyn (2007). The sample was mixed with 500 µl of DNA isolation buffer (100 mM Tris-HCl pH 8; 50 mM EDTA; 1% SDS) and vortexed. To the suspension, 200 µl of glass beads was added and the sample was vortexed for 4 min with immediate cooling on ice and addition of 275 µl 7 M ammonium acetate (pH 7), incubated for 5 min at 65°C followed by cooling on ice for 5 min. To the suspension 500 µl of chloroform was added, vortexed and centrifuged at 14 000 rpm for 5 min at 4°C. The supernatant was removed and precipitated with 1 volume of isopropanol overnight at -20°C. Following precipitation, the DNA was spooled and centrifuged for 1 min at 4°C. The pellet was washed with 70% ethanol and centrifuged at 14 000 rpm for 5 min at 4°C, dried and dissolved in 50 µl sterile water containing RNase A (Fermentas). The DNA was stored at -20ºC.

3.2.1.3 Method 3

With the FastDNA® SPIN Kit (Qbiogene) the amount of starting material was 0.5 g, the biofilm was added to the Lysing Matrix E tube (a mixture of silica and ceramic particles). This was followed by the addition of 978 µl of sodium phosphate buffer and 122 µl of MT Buffer. The sample was vortexed for 30 s. The samples were centrifuged at 14 000 rpm for 30 s and the supernatant was transferred to a new tube. To the supernatant 250 µl of PPS (Protein Precipitation Solution) reagent was added and mixed by shaking the tube 10X by hand. The sample was centrifuged for 5 min at 14 000 rpm and the supernatant was transferred to a 15 ml falcon tube to which 1 ml of the Binding Matrix Suspension was

added. The tube was inverted by hand for 2 min to allow binding of DNA to matrix. The tube was then allowed to stand at room temperature for 3 min to allow settling of the silica matrix. This was followed by the removal of 500 µl of the supernatant which was discarded. The Binding Matrix was resuspended in the remaining supernatant. Approximately 600 µl of the mixture was added to a SPIN™ Filter and centrifuged at 14 000 rpm for 1 min. The catch tube was emptied and the remaining supernatant was added and centrifuged as above. Following centrifugation 500 µl SEWS-M (Salt / Ethanol Wash Solution) was added to the filter and centrifuged at 14 000 rpm for 1 min. The flow-through was decanted and centrifuged for an additional 2 min to remove residual SEWS-M wash solution. The filter was removed and placed in a new kit supplied catch tube and allowed to air dry for 5 min at room temperature. The DNA was eluted in 50 µl DES (DNA Elution Solution- Ultra Pure Water) and centrifuged at 14 000 rpm for 1 min. The DNA was stored at -20°C.

3.2.2 Diversity studies

3.2.2.1 PCR conditions

Polymerase Chain Reaction (PCR) using universal primers (Integrated DNA Technologies) specific for each of the domains (bacteria, archaea and eukarya) was performed. The primer sets for each domain are represented in Table 3.1. PCR amplification with respective primers was performed as described in section 2.2.2. The PCR programme for the bacterial and archaeal GC-clamped primers included an initial denaturation at 95°C for 5 min, followed by 25 cycles of denaturation at 95°C for 45 s; annealing at 55°C for 45 s; elongation at 72°C for 1 min and a final elongation at 72°C for 10 min.

3.2.2.2 DGGE

DGGE was performed at 60°C with a D-Code system (BioRad Laboratories) according to the manufacturer's instruction. A 8% polyacrylamide gel with a denaturant gradient from 30% to 60% was used for analyzing fragments amplified using 341FGC/517 and 344FGC/517 primer sets (Table 3.1). The gel was run for 3 h at 200 V, and stained with ethidium bromide.

Table 3.1 PCR primers used in this study

Primer name	***Sequence***	***Reference***
27F (Bacteria)	5'-AGA GTT TGA TCC TGG CTC AG -3'	Lane, 1991
1492R (Bacteria)	5'-GGT TAC CTT GTT ACG ACT T -3'	Lane, 1991
A2F (Archaea)	5'-ATT ACC GCG GCT GCT GG -3'	Reysenbach and Pace, 1995
20b (Archaea)	5'-YTC CSG TTG ATC CYG CSR GA -3'	Rincon *et al.,* 2006
EUK A (Eukarya)	5'-AAC CTG GTT GAT CCT GCC AGT -3'	Dìez *et al.,* 2001
EUK B (Eukarya)	5'-TGA TCC TTC TGC AGG TTC ACC TAC -3'	Dìez *et al.,* 2001
341F(Bacteria)	5'- CCT ACG GGA GGC AGC AG-3'	Watanabe *et al.,* 2001
341FGC-Clamped	5'-CGC CCG CCG CGC GCG GCG GGC GGG -3'	Watanabe *et al.,* 2001
517R	5'-ATT ACC GCG GCT GCT GG -3'	Bano *et al.,* 2004
344FGC-Clamped (Archaea)	5'- ACG GGG CGC AGC AGG CGC GA-3'' (40 bp GC clamp) 5'-CGC CCG CCG CGC CCC GCG CCC GTC CCG CCG CCC C CG CCC C-3'	Bano *et al.,* 2004

3.2.3 Cloning and sequencing of the 16S bacterial and archaeal rRNA and the 18S eukaryotic rRNA genes

3.2.3.1 Library construction

In order to have an idea of which organisms were present in the biofilm sample, the 16S rRNA gene amplicons from domains bacteria and archaea, and the 18S rRNA gene amplicon from the domain eukarya were cloned into pGEM®T-Easy (Promega). Ligation

reactions were performed for 2 h at room temperature. The ligation products were then used to transform *E.coli* TOP10 chemically competent cells as described in section 2.2.3.1.5. Transformants were selected on LB agar plates containing 100 µg/ml of ampicillin (Sgma), IPTG and X-gal (PeqLab). The plates were incubated overnight at 37°C.

3.2.3.2 Plasmid DNA extraction and restriction analysis

Random colonies were picked up and inoculated into 5 ml of LB containing ampicillin (100 µg/ml). The tubes were incubated overnight at 37°C with mild shaking. Plasmid DNA was then extracted from the cultures using the lysis by boiling method described in section 2.2.3.1.6.

3.2.3.3 Sequencing of randomly selected clones

Fifty bacterial, 20 archaeal and 5 eukaryote clones were selected for sequencing at Macrogen (Korea). Electropherograms generated from the sequences were inspected with FinchTV software (Geospiza). Edited sequences were then compared with the gene databases with BLASTN (National centre for Biotechnology Information). The bacterial clones were subjected to further phylogenetic analysis using the ARB program (Ludwig *et al.,* 2004). Alignments were done using ARB based on conserved primary and secondary structural elements within the 16S rRNA sequences. The resulting alignments were used to position the sequences in the ARB tree. Phylogenetic analyses were carried out with the maximum parsimony, neighbour joining; and maximum likelihood programs in the ARB software package. Trees were bootstrapped 100X and the Felsentein correction type was used (Ludwig *et al.,* 2004). The number of operational taxonomic units (OTUs) was determined for the bacterial 16S rRNA library using DOTUR, a computer program for defining operational taxonomic units and estimating species richness (Schloss and Handelsman, 2005). An OTU was defined as a group with ≥97% identity in their small-subunit rRNA gene sequences following the conventional definition of a microbial species (Fierer *et al.,* 2007). A rarefaction curve was produced using the DOTUR programme.

3. 3 Results and Discussion

3.3.1 Environmental DNA extraction

Based on our observations, we found that the three extraction methods performed yielded DNA in the size range of >10 kb (Figure 3.1). DNA obtained with the chemical and

enzymatic lysis method (method 1) showed a high degree of inhibitory substances as observed by its intense brown colour indicating the presence of contaminants. Although a large volume of DNA was obtained with this method (2 ml), the DNA had to be serially diluted (1:100) before concentration readings could be taken due to the high degree of contaminants that affected the readings (Table 3.2). Pre-treating the sample with aluminium sulfate prior to the extraction process removed only a small fraction of the contaminants (method 2). The FastDNA® SPIN Kit (method 3) yielded DNA of highest purity when compared with the other methods; however, a lower concentration of DNA was obtained. Our results are consistent with studies conducted by Luna *et al.* (2006) indicating that lower yields of DNA are obtained using commercial kits when compared to *in situ* lysis procedures.

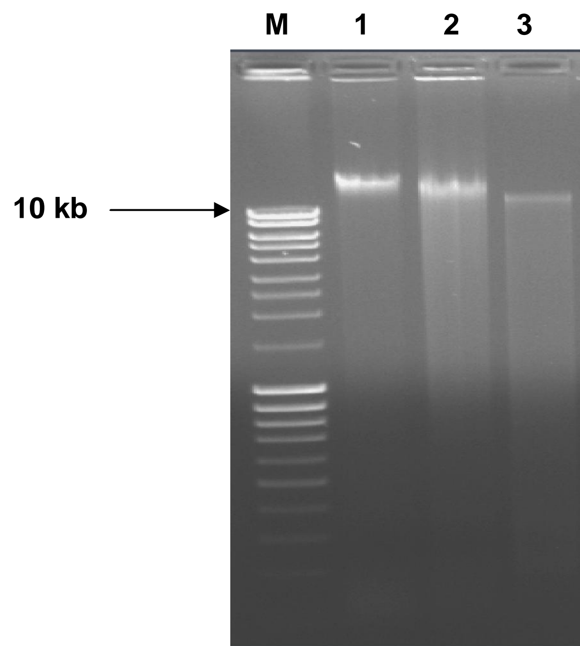


Figure 3.1: Genomic DNA extractions from the Beatrix Mine biofilm. M: MassRuler DNA 1kb Ladder (SM#0403-Fermentas); lane 1: DNA extracted with chemical and enzymatic lysis (method1); lane 2: DNA extracted after pretreatment with aluminium sulfate (method 2); and lane 3: DNA extracted with the FastDNA soil Kit (method 3).

Table 3.2: DNA concentration and purity readings

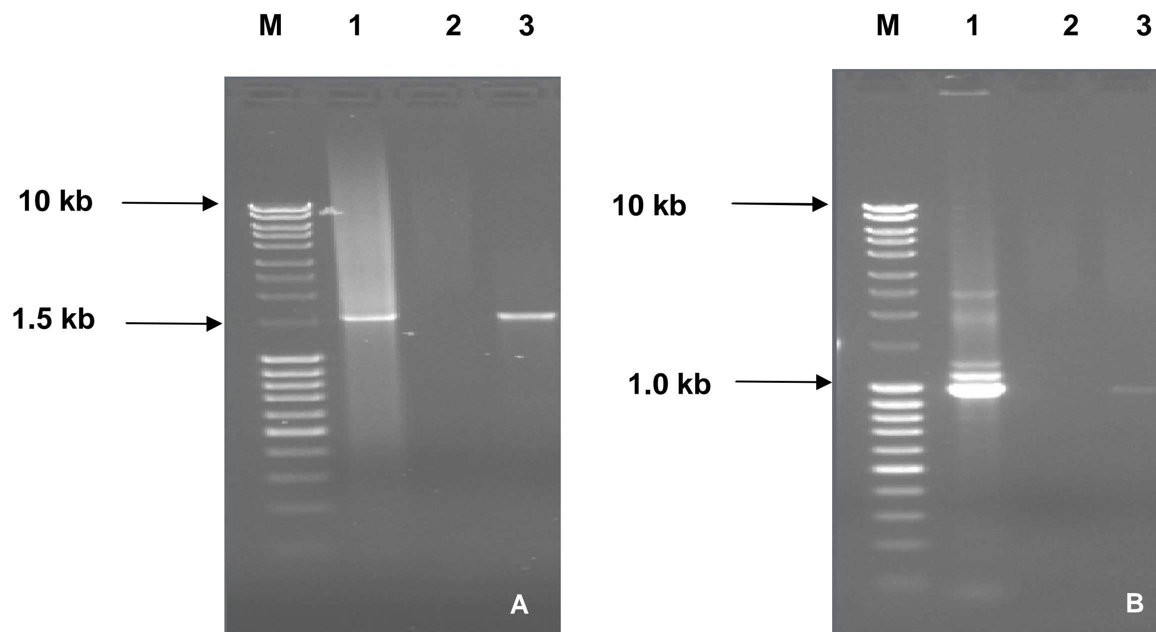
Extraction Method	***DNA Concentration (μg/g) of biofilm***	***$A_{260/280}$***
Zhou *et al.* (1996)	2631	1.35 (1:100 dilution)
Dong et al, (2006); and Labuschagne and Albertyn (2007); (Towner, 1991)	23	1.18
FastDNA® SPIN Kit (Qbiogene)	4.8	1.90

For downstream processing such as library construction it is important to have a high concentration of DNA to begin with, because losses of DNA occur during certain processes such as gel extraction. Based on this we had to select between methods 1 and 2 to continue with. DNA obtained with method 1 did not digest with various restriction enzymes tested; this was probably due to the presence of inhibitory substances that affected the restriction enzymes. The sample was then purified using a commercially available purification kit as well as serially diluted before performing the restriction digests. However, neither purification nor dilution seemed to remove the inhibitory substances in order to make the DNA readily digestible. DNA obtained with method 2 was readily digestible and therefore we selected this method of extraction to continue with.

3.3.2 Diversity studies

The biofilm sample was shown to contain bacterial, archaeal and eukaryotic DNA (Figure 3.2 A, B, and C). This is expected with metagenomic samples especially when using direct methods of extraction that allow for a larger fraction of the metagenome to be accessed as compared with an indirect method of extraction. The disadvantage of obtaining eukaryotic DNA is that it can account for more than ~60% of the DNA yield, thereby under-representing

the prokaryotes that are present in the sample material (Gabor *et al.,* 2003). However, this was not really detrimental to the future of our study considering the inability of the eukaryotic genes to be expressed in the library due to the presence of their non-coding sequences. It could however mean that the majority of the clones would harbour DNA which would not be functional and thus complicating the screening process.



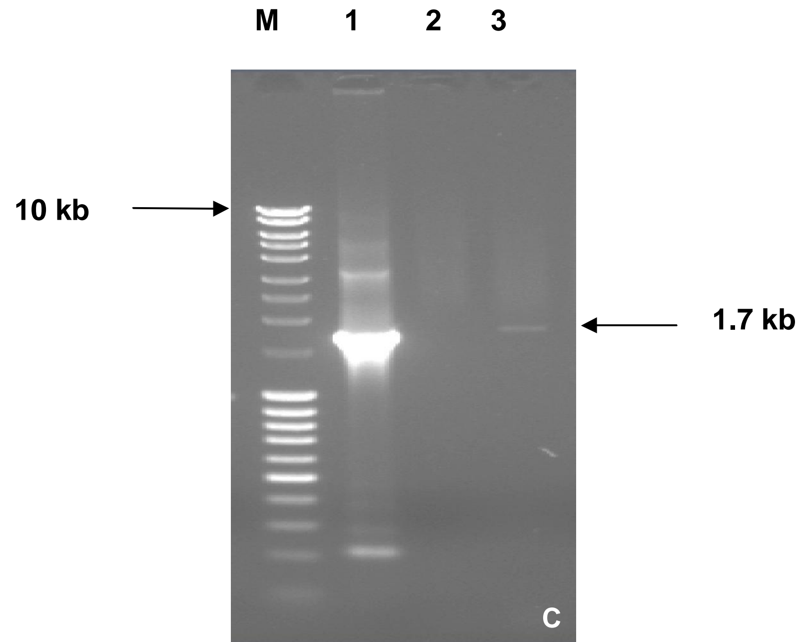


Figure 3.2: Domain specific PCR. A: Bacterial PCR; B: Archeal PCR; and C: Eukaryotic PCR. M: MassRuler 1kb ladder (SM#0403- Fermentas); lane 1: pos. control; lane 2: neg. control: and lane 3: biofilm sample (Beatrix Mine).

Focusing mainly on archaea and bacteria, we were interested in understanding the depth and evenness of the archaeal and bacterial populations present in the biofilm. The 200 bp products generated by PCR using primers containing a GC clamp yielded banding patterns that suggested the presence of at least 6 possible archaeal and 22 bacterial populations present in the biofilm (Figure 3.3). Our results are consistent with previous studies that have shown that archaea tend to have a lower diversity than bacteria in the same environment. A possible reason for archaea having a lower diversity than bacteria in a particular environment is due to the fact that the metabolic processes of archaea requires so much energy that their phylogenetic diversification is limited when compared to that of bacteria (Aller and Kemp, 2008). In addition according to Gonzalez-Toril *et al.* (2003), archaea generally account for only a minor fraction (<5%) of the total prokaryotic community. The DGGE analysis was also important for library screening, because the number of different bands present on the gel provides a diversity value which is incorporated into a formula that determines the number of clones that need to be screened in order to determine if the library is representative of the metagenome. However, DGGE only provides an estimate of the

number of members present in a community. Furthermore, only the dominant members of a community will be visualized by DGGE thereby masking other minor members which may be present.

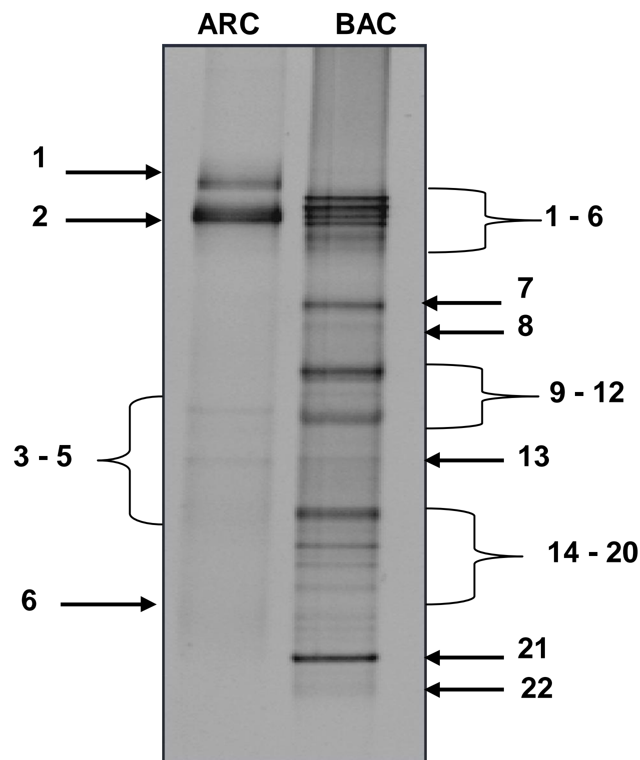


Figure 3.3: DGGE analysis of the archaeal and bacterial population present in the biofilm. The numbers indicate a consecutive numbering of the bands.

3.3.3 Cloning and sequencing of the 16S bacterial and archaeal rRNA and 18S rRNA eukaryotic rRNA amplicons

Restriction analysis of selected clones (16S rRNA bacteria; 16S rRNA archaea; and 18S rRNA eukaryotic) revealed fragments of the expected sizes for each domain: bacteria (1.5 kb); archaea (1.0 kb); and eukaryotes (1.7 kb) (Figure 3.4 and 3.8).

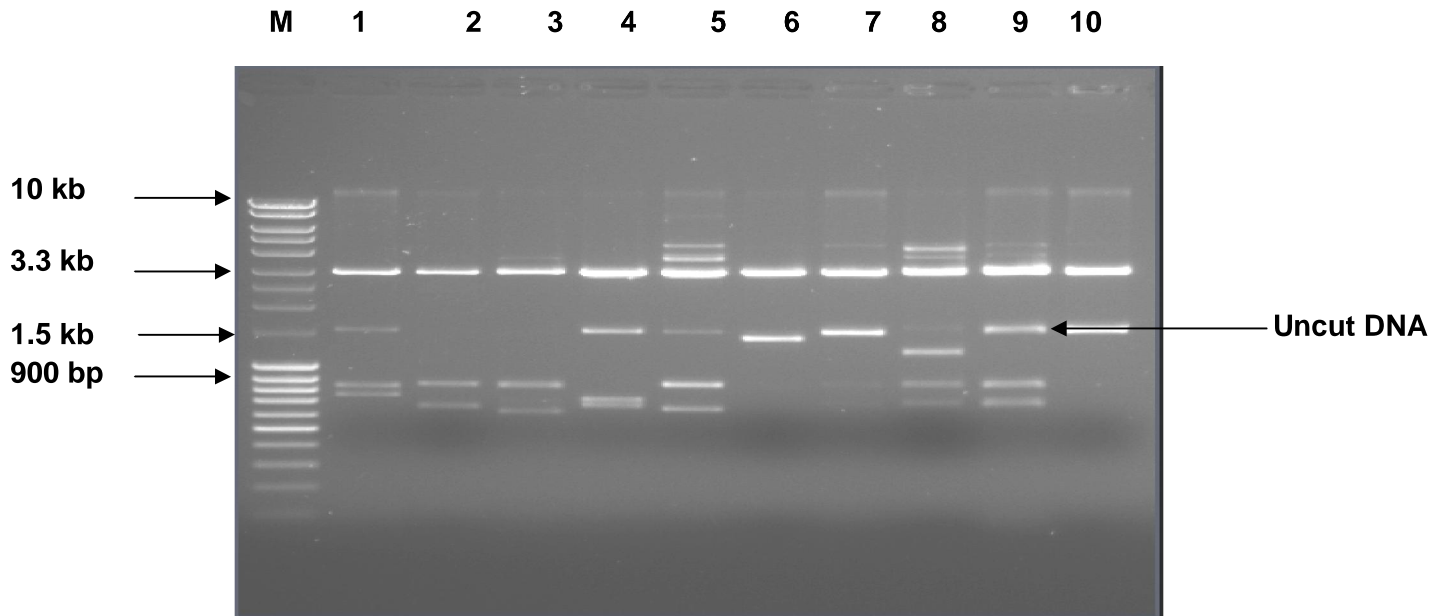


Figure 3.4: Restriction analysis of selected 16S bacterial clones. M: MassRuler DNA Ladder (SM#0403); and lanes: 1-10: randomly selected clones.

3.3.4 Phylogenetic analysis of selected bacterial clones

Since the main purpose of this study was to assess the presence of the different types of organisms in a Beatrix mine biofilm before construction of a metagenome library, one-directional sequencing of selected 16S clones was performed. The sequences were trimmed to ~800 bp in length, and the problematic parts of the reads were edited out. Of the 50 clones sequenced 13 clones were discarded from the phylogenetic analysis because some of the clones were chimeras. Table 3.3 represents the 50 clones with their significant hits, whereas the constructed phylogenetic tree (Figure 3.5) represents 37/50 clones. The alignments of the edited sequences were checked by confirming complementary base-pairing in known regions of the alignment. The phylogenetic analysis allowed us to easily identify the position of our clones within known bacterial phyla.

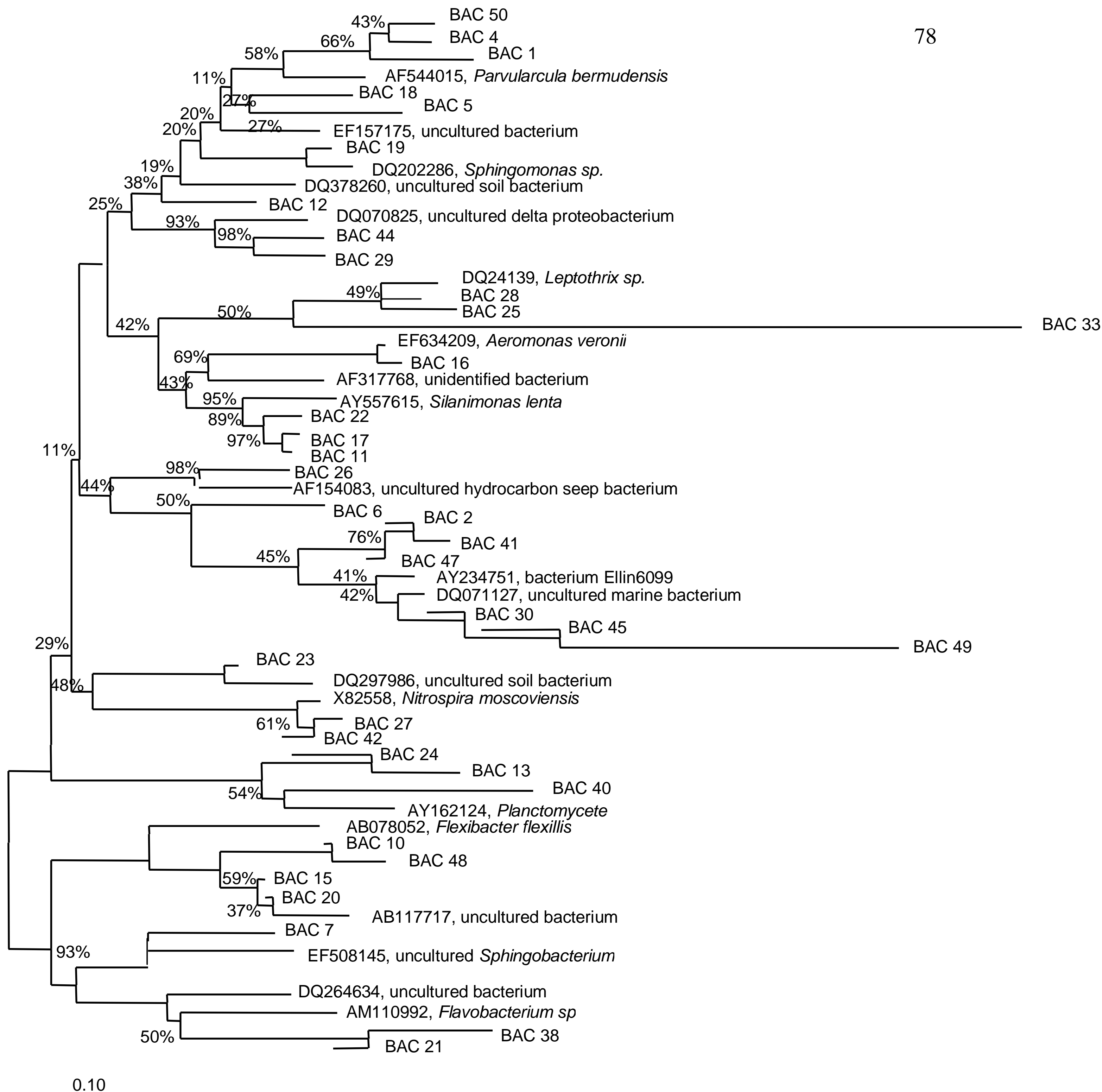


Figure 3.5: Phylogenetic tree representing bacteria l16S rRNA gene sequences constructed using the ARB program. Maximum parsimony, maximum likelihood and neighbour joining analyses produced highly similar tree topologies. The scale represents a 10% sequence divergence.

On the basis of the phylogenetic analysis, the clonal sequences were affiliated with at least 5 phyla of the domain bacteria (Figure 3.6). The dominant group of our clone library

was allocated to the phylum *Proteobacteria* (43% of the total number of clones) with the organisms being distributed within the alpha, beta, gamma and delta classes. The second dominant group of the clone library, represented by 8 clones (21%), was classified into the phylum *Acidobacteria*. The other groups of the library were determined to be, in order of abundance, in the phyla *Bacteriodetes, Nitrospirae,* and *Planctomycetes*. Sequences allocated to the phyla that are present in our clone library have been reported to be isolated from various environments including marine environments *e.g.* the Sargasso Sea and sea floor basalts, sub-surface ground water, limestone caves, tar pits, alkaliphilic hot springs, mine drainage sites and biofilms (Cho and Giovanni, 2003; Kim and Crowley, 2007; Ikner *et al.,* 2007). We found that clones BAC 28, BAC 25 and BAC 33 clustered with an organism capable of transforming arsenic and various other metals. Numerous studies have been conducted in our laboratory on the metal transforming properties of organisms isolated from South African deep mines. Thermostable enzymes capable of transforming iron, chromium, uranium and gold have been identified in the organism *Thermus scotoductus* SA-01, a thermophile isolated from a South African gold mine (van Heerden *et al.,* 2008).

An interesting finding was that clone BAC 26 clustered with an uncultured hydrocarbon seep bacterium from oil contaminated soil. Organisms residing in such an environment need to be able to degrade hydrocarbon substrates. Studies conducted by Spangenberg and Frimmel (2001) and Ward *et al.* (2004) indicate the presence of substantial quantities of hydrocarbon gases and long chain alkanes in the Witwatersrand basin, of which the Beatrix mine is a part. The *n*-alkanes in the C_{15}-C_{27} range (long chains) were shown to be the dominant resolvable compounds in the gas chromatograms of the saturated hydrocarbons from the carbon seams in the basin. Enzymes such as cytochrome P450 monooxygenases have the ability to utilize long chain alkanes. Based on this, the next chapter focuses on the detection of cytochrome P450 monooxygenases from the metagenome through sequenced-based PCR screening. Therefore by identifying which organisms are present in the sample and the properties of the site from which the sample was obtained, we can co-relate this to the types of enzymes or proteins that we expect to find when screening is employed.

Table 3.3: Sequence data of selected 16S rRNA bacterial clones

Clone	***Organism***	***% Identity***	***E value***	**Accession No.**
BAC1	Uncultured sludge bacterium 16S rRNA	89	0	AF234706.1
BAC2	Uncultured *Acidobacteria* bacterium 16S rRNA	95	0	EU223942
BAC3	Uncultured *Nitrospira* sp clone 16S rRNA	85	0	EU084880
BAC4	*Brucella suis*, complete genome	88	0	CP00912.1
BAC5	*Spingomonas yunnanensis* strain YIM 003 partial 16S rRNA	94	0	AY894691.1
BAC6	Bacterial species 16S rRNA	88	0	ZA5709.1
BAC7	Uncultured *Bacteroidetes* bacterium	83	0	EU229455.1
BAC8	*Methlocella tundrae* partial 16S rRNA	78	2e-172	AJ563928.1
BAC9	*Marinobacter sp.* 16S rRNA	80	0	AJ429499.1
BAC10	Uncultured bacterium gene16S rRNA	90	0	AB280279.1
BAC11	Uncultured *gamma proteobacterium* partial 16S rRNA gene	97	0	AJ534675.1
BAC12	Uncultured *alpha proteobacterium* partial 16S rRNA gene	98	0	EF612403.1
BAC13	Uncultured bacterium clone 16S rRNA	88	0	EF522844.1
BAC14	Uncultured bacterium gene16S rRNA	94	4e-117	AB240487.1
BAC15	Uncultured soil bacterium partial 16S rRNA	92	0	EU589303.1
BAC16	*Aeromonas veronii* partial 16S rRNA	97	0	AM184224.1
BAC17	Uncultured bacterium clone partial 16S rRNA	93	0	DQ532277.1
BAC18	Uncultured bacterium clone partial 16S rRNA	89	0	EF125428.1
BAC19	*Sphingomonas sp.* 16S rRNA	96	0	AB047364.1
BAC20	Uncultured soil bacterium partial 16S rRNA	93	0	EU589303.1
BAC21	Uncultured *Cytophagales* bacterium clone, 16S rRNA	90	0	AF361197.1
BAC22	*Lysobacter taiwanensis* 16S rRNA	94	0	DQ314555.1
BAC23	Unidentified bacterium clone 16S rRNA	92	0	EF219552
BAC24	Uncultured bacterium clone 16S rRNA	92	0	EU083503.1

BAC25	*Beta-proteobacterium* 16S rRNA	89	0	AF236010.1
BAC26	Uncultured bacterium 16S rRNA	93	0	AM176872.1
BAC27	*Nitrospira sp.* 16S rRNA	97	0	Y14644.1
BAC28	*Methylibium aquaticum* 16S rRNA	97	0	DQ664244.1
BAC29	Uncultured *deltaproteobacterium* 16S rRNA	90	0	AJ581552.1
BAC30	Uncultured F*irmicutes* 16S rRNA	97	0	EF665782.1
BAC31	Uncultured bacterium clone 16S rRNA	84	3e-119	AF392636.1
BAC32	Uncultured *Cytophagales* bacterium clone, 16S rRNA	83	2e-90	AF361197.1
BAC33	Uncultured bacterium clone 16S rRNA	92	0	DQ463246.1
BAC34	Uncultured *Bacillus sp. c*lone 16S Rrna	82	7e-21	AY876909.1
BAC35	Uncultured bacterium clone 16S rRNA	83	1e-106	EU015107.1
BAC36	Uncultured sludge bacterium 16S Rrna	89	1e-131	AF234706.1
BAC37	Uncultured bacterium clone 16S rRNA	94	1e-17	EF173339.1
BAC38	Uncultured bacterium clone 16S rRNA	87	0.0	EF632776.1
BAC39	Uncultured bacterium clone 16S rRNA	86	1e--157	EU386115.1
BAC40	*Plantomycete* 16S Rrna	84	0.0	AY1621229.11
BAC41	Uncultured bacterium gene 16S rRNA	96	0.0	AB286376.1
BAC42	*Nitrospira moscoviensis* 16S rRNA	98	0.0	X82558.1
BAC43	Uncultured bacterium clone 16S rRNA	88	0.0	EU617826.1
BAC44	Uncultured bacterium clone 16S rRNA	97	0.0	EU160003.1
BAC45	Uncultured bacterium clone	93	0.0	EU015113.1
BAC46	Uncultured bacterium vlone 16S rRNA	95	1e-174	AF392636.1
BAC47	Uncultured bacterium isolate	96	0.0	AY703460.1
BAC48	Uncultured bacterium 16S rRNA	86	0.0	AB280279.1
BAC49	Uncultured bacterium clone	96	0.0	EU148614.1
BAC50	Uncultured *Phaeobacter* 16S rRNA	86	0.0	EU375177.1

Phylogenetic distribution of bacterial clonal library

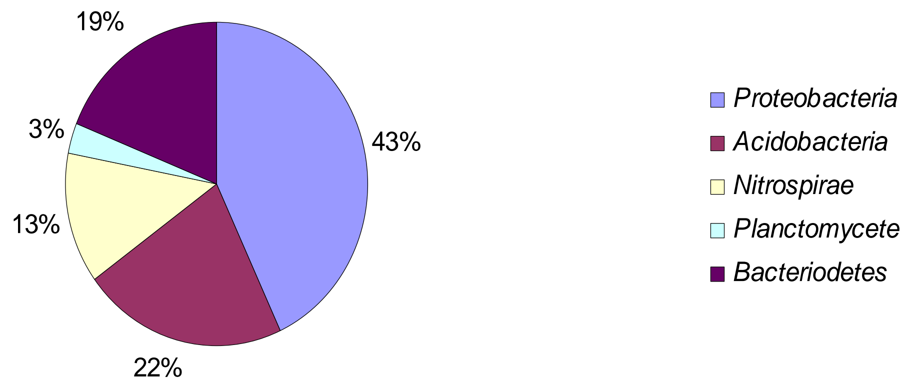


Figure 3.6: Graphical representation of the phylogenetic distribution of the bacterial clonal library based on 16S rRNA gene sequences from the Beatrix gold mine.

There was no clone whose 16S rRNA gene sequence was 100% identical to those of known bacterial species. Four clones showed ≥97% sequence identity to known bacterial species. Sequences with >97% identity were treated as identical (Tamaki *et al.,* 2005). The remaining clones showed ≤95% sequence identity to other identified species present in the NCBI database. Our results indicate that the bacterial community in the deep mine biofilm is primarily composed of unknown bacterial species. This made the biofilm an interesting sample to work with for the preparation of metagenome libraries because of the possibility of obtaining unique classes of certain enzymes.

According to the rarefaction analysis of the 37 clones analyzed 29 different OTUs were observed. However, the rarefaction results indicate that only a small portion of the richness in the bacterial community (at the ≥97% sequence identity level) was surveyed with the 16S clones sequenced as the curve did not reach an asymptote (Figure 3.7). Nonparametric estimators *e.g.* Chao I and ACE are frequently used to estimate the total number of OTUs in a given community (Schloss and Handelsman, 2005). However, in our study the

nonparametric estimates of the total number of OTU richness failed to stabilize or reach an asymptote. This meant that millions of clones would have to be screened if functional screening were to be performed and to adequately cover the genomic diversity in the biofilm.

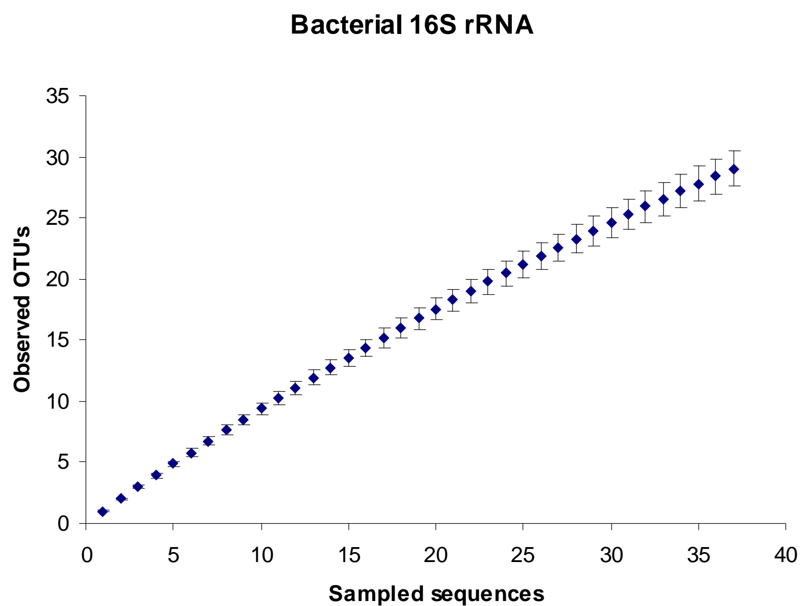


Figure 3.7: Rarefaction curve for the bacterial 16S rRNA library from the Beatrix gold mine. The curve was generated using the DOTUR programme. There is no apparent asymptote in the curve suggesting that the clones sequenced do not encompass the full extent OTU richness.

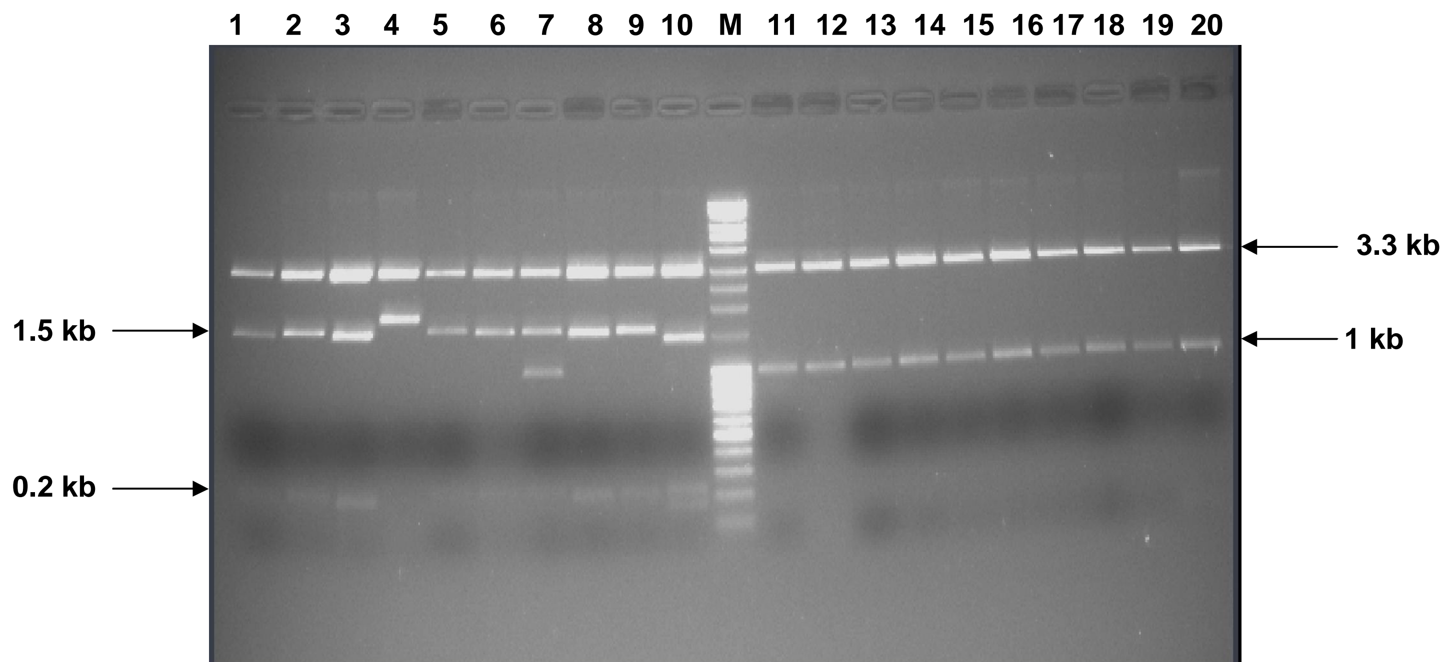


Figure 3.8: Restriction analysis of selected 18S eukaryotic and 16S archaeal clones. M: MassRuler DNA Ladder (SM #0403); lanes: 1-10 -randomly selected eukaryotic clones and lanes 11-20: randomly selected archeal clones.

Table 3.4: Sequence data of selected archaeal 16S rRNA clones

Clone	***Organism***	***% Identity***	***E value***	***Accession No.***
ARC1	Uncultured archeon clone ZES-28 partial 16S rRNA	99	0.0	EF367466.1
ARC2	Uncultured archeon clone ZES-28 partial 16S rRNA	99	0.0	EF367466.1
ARC3	Uncultured archeon clone ZES-28 partial 16S rRNA	99	0.0	EF367466.1
ARC4	Uncultured crenarchaeote partial 16S rRNA	82	1e-142	AJ870317.1
ARC5	Uncultured archeon clone ZES-28 partial 16S rRNA	99	0.0	EF367466.1
ARC6	Uncultured archeon clone ME-59 16S rRNA	94	0.0	DQ641778.1
ARC7	Uncultured crenarchaeote clone partial 16S rRNA	94	0.0	DQ641760.1
ARC8	Uncultured archeon clone ZES-28 partial 16S rRNA	98	0.0	EF367466.1
ARC9	Uncultured archeon clone ME-41 16S rRNA	99	0.0	DQ641760.1
ARC10	Uncultured archeon clone ZES-47 partial 16S rRNA	98	0.0	EF367485.1
ARC11	Uncultured archeon clone ZES-28 partial 16S rRNA	99	0.0	EF367466.1
ARC12	Uncultured archeon clone ZES-28 partial 16S rRNA	99	0.0	EF367466.1
ARC13	Uncultured archeon clone ZES-28 partial 16S rRNA	99	0.0	EF367466.1
ARC14	Uncultured archeon clone ZES-28 partial 16S rRNA	99	0.0	EF367466.1
ARC15	Uncultured archeon clone ZES-28 partial 16S rRNA	96	0.0	EF367466.1
ARC16	Uncultured archeon clone ZES-28 partial 16S rRNA	99	0.0	EF367466.1
ARC17	Uncultured archeon clone ZES-28 partial 16S rRNA	97	0.0	EF367466.1
ARC18	Uncultured archeon clone ZES-28 partial 16S rRNA	97	0.0	EF367466.1
ARC19	Uncultured archeon clone ZES-28 partial 16S rRNA	96	0.0	EF367466.1
ARC20	Uncultured archeon clone ZES-28 partial 16S rRNA	96	0.0	EF367466.1

Table 3.5: Sequence data of selected eukaryotic 18S rRNA clones

Clone	***Organism***	***% Identity***	***E value***	***Accession No.***
EUK1	*Vorticella fusca* SSU rRNA	96	0.0	DQ190468.1
EUK2	Uncultured eukaryote clone partial 18s rRNA	83	0.0	AY664993.1
EUK3	*Theileria sp.* 18s rRNA	77	0.0	AF162431.1
EUK4	Uncultured eukaryote gene, SSU rRNA	90	0.0	AB275108.1
EUK5	*Stenostomum leucops* 18s rRNA	95	0.0	D85095.1

Sequence data obtained from the archaeal clonal library showed sequence identities with that of uncultured archaea present in the database, in particular a single uncultured archaeal library from a marine sample. This could be attributed to the fact that there is limited sequence data on archaea present in the databases because only 4 classified taxonomic groups of the domain are known thus far as described by Baker *et al.,* (2003) and an additional number of unclassified and environmental archaea are known (http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Undef&name=Archaea&lvl=3&srchmode=1&keep=1&unlock).

Studies conducted on the fracture water of the Beatrix mine indicated the presence of bromide and chloride in the fracture water. It is also suggested that a saline water invasion could have occurred after the 2.0 Ga hydrothermal episode. The Wits Basin water is a mixture of hydrothermal and paleo-meteoric water which ranges in age from ~10k to >1.5 Ma, is of low salinity. This mixing between two components probably occurred for at least the past ~100 Myr and could be a mechanism by which microorganisms colonised these environments and this could explain the presence of diversity related to marine isolates (Onstott *et al.,* 2006). Although this study does not really focus on eukaryotes, we did clone and sequence the 18S rRNA from eukaryotes to determine if the presence of eukaryal DNA would complicate our screening. Two of the five clones that were sequenced showed hits with that of uncultured eukaryotes. The disadvantage of obtaining eukaryotic

DNA is that it can account for most of the DNA yield, thereby under-representing the prokaryotes that are present in the sample material (Gabor *et al.,* 2003).

3.4. Conclusion

Microorganisms present in different environments have different susceptibilities to cell lysis methods, and the sequences present in the isolated DNA is dependant on the type of extraction method used (Daniel, 2005). In this study, we evaluated 3 direct extraction methods, based on the direct extraction of DNA from sample matrices. We excluded the possibility of using an indirect method for extraction due to previous reports that state, the inability of obtaining high yields of DNA with such methods (Gabor *et al.,* 2003), thereby making it impractical for the future of this study. We found that a combination of aluminium sulfate treatment coupled with the use of glass beads for physical disruption of the cells yielded DNA of adequate concentration for downstream processing such as library construction. According to Dong *et al.* (2006) small quantities (1-2 µl of 100 mM aluminium sulfate) sufficiently removed humic acids from the sample without resulting in a loss of DNA yield. In addition the use of a physical method of extraction allows for a greater accessibility of the microbial population present in the sample material.

We live in a world dominated by microorganisms. However, little is known about the role they play in the environment (Manichanh *et al.,* 2008). Knowledge of microbial diversity has increased due to the ability to survey genes directly in environmental samples. The field of metagenomics has opened a new era of microbiology, since most microbial resources have not been really explored as well as the potential to discover novel microbial resources from uncultivated bacteria (Lee, 2005). The 16S rRNA gene has been a useful prokaryotic phylogenetic marker (Osborne *et al.,* 2005). While 16S rRNA gene libraries are ideally representative of the diversity present in the source environment, it is likely that the primers employed are not truly “universal” and therefore there is no single set that will successfully target respective microorganisms in a particular environment (Aller and Kemp, 2008). Despite these drawbacks, Quaiser *et al.* (2002) showed that the characterization of the 16S rRNA genes revealed the presence of a clade of crenarchaeota (archaea) among many other novel lineages of uncultivated bacteria. In this study, we were able to identify all three domains in our sample material, indicating that our extraction method was successful in accessing a majority of the microorganisms present in the sample matrix. Sequence

analysis of the selected clones indicated the majority of the microorganisms present in the deep mine are uncultured, thereby increasing the potential for identifying unique genes and biocatalysts from these organisms. The limited assesment of diversity was performed to assess the potential for the discovery of novel genes. Clearly, the biofilm contains an extensive population of microbes, including novel genospecies, which could provide novel genetic material. Exhaustive screening of the biofilm to adequately cover the entire metagenome would therefore be impractical in our situation as this would require high throughput systems which could handle millions of clones.

Chapter 4

Sequenced-based screening for cytochrome P450 monoxygenases by a PCR based approach

4.1 Introduction

A ubiquitous superfamily of heme-containing enzymes, cytochrome P450 monooxygenases (CYP's) are useful catalysts for oxidation reactions that reductively cleave molecular oxygen and introduce one oxygen atom into activated and nonactivated carbon bonds to yield regio and stereo-chemically pure oxygenated compounds (Kim *et al.,* 2007; Funhoff *et al.,* 2007). Cytochromes P450 are so named due to their character as hemoproteins and from their unusual spectral properties displaying a typical absorption maximum of the reduced CO-bound complex at 450 nm: cytochrome stands for a hemoprotein, P for pigment and 450 reflects the absorption peak of the CO complex at 450 nm (Bernhardt, 2006).

P450 enzymes generally fall into two broad classes. Class I P450 systems contain most bacterial P450s as well as the mitochondrial P450 systems from eukaryotes. Both groups, although not phylogenetically related, are composed of three-component systems comprising a NAD(P)H binding flavoprotein reductase, which transfers reducing equivalents from a pyridine nucleotide (NADH or NADPH) to the second component of the system, a ferrodoxin, which in turn reduces the cytochrome P450 itself. Class II (microsomal cytochromes) enzymes are two component systems comprising a FAD and FMN containing NADPH reductase (in which FAD and FMN are in equimolar ratio) in the P450. This type of P450 is found almost exclusively in eukaryotes (Roberts *et al.,* 2003; Bernhardt, 2006; Hannemann *et al.,* 2007) (Figure 4.1).

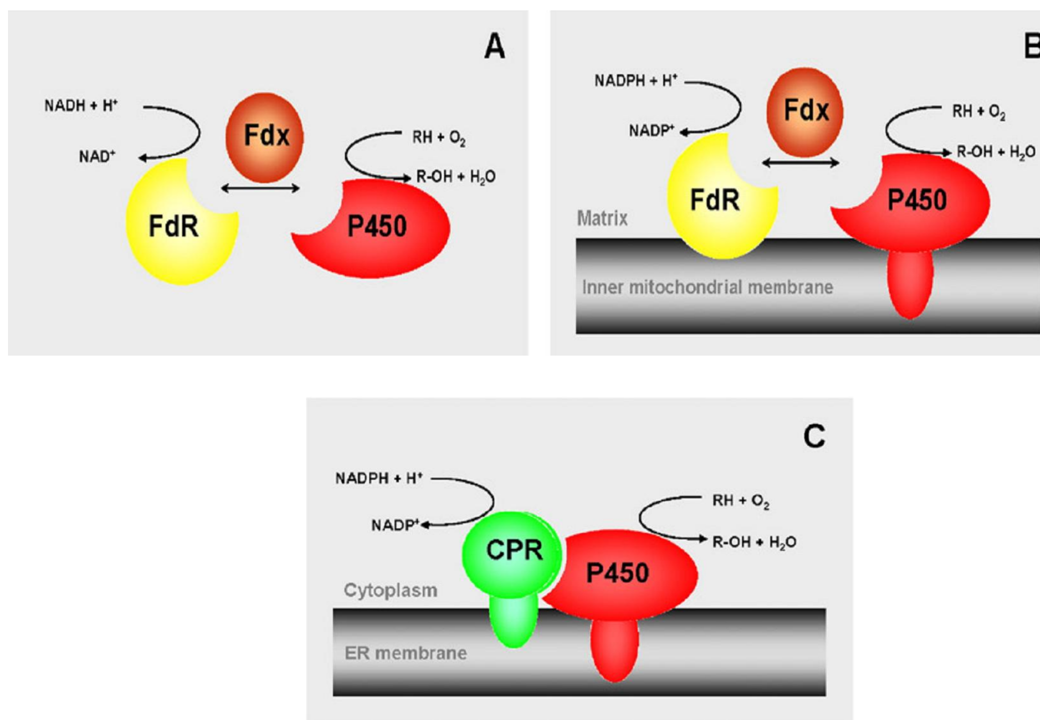


Figure 4.1: Schematic representation of different cytochrome P450 systems. (A): class I, bacterial system; (B): class I, mitochondrial system; and (C): class II microsomal system (taken from Hannemann *et al.,* 2007).

There are however, examples of cytochrome P450s that are not readily defined as a member of either of the two classes mentioned above. P450 BM3 (CYP102A1) from *Bacillus megaterium,* which oxidizes long chain fatty acids, comprises a diflavin reductase fused to the P450 to form a catalytically self-sufficient single polypeptide enzyme and is an example of a self-sufficient P450 (Roberts *et al.,* 2003; Kim *et al.,* 2007;). P450's are involved in a plethora of metabolic processes, both anabolic and catabolic, and collectively interact with an enormous variety of substrates (De Mot and Parret, 2002). CYP's in bacteria are involved in the biosynthesis of secondary metabolites such as antibiotics and in the utilization of hydrophobic low molecular weight compounds such as alkanes and aromatics (Kubota *et al.,* 2005). In plants, P450's are involved in numerous biosynthesis pathways of both central and secondary metabolites and they are key enzymes for detoxification or activation of herbicides (O'Keefe *et al.,* 1994). During the past decade, microbial P450s and recently mammalian P450s have been investigated with respect to the production of fine chemicals, fragrances, pharmaceutical compounds and for their use in bioremediation (Urlacher and Eiben, 2006).

Thus far only a few alkane-hydroxylating P450 enzymes have been identified and characterized. Cytochrome P450 enzymes of the CYP153 family have been shown to be responsible for terminal hydroxylation of aliphatic, alicyclic and alkyl-substituted compounds. CYP153's are the best soluble CYP enzymes for terminal alkane hydroxylation (van Beilen *et al.,* 2006; Funhoff *et al.,* 2007). An intensively studied area of pharmaceutical research is the search for new anticancer drugs. The biocatalytic production of the anticancer drug perillyl alcohol from limonene has been reported to involve a CYP153 cytochrome from *Mycobacterium sp.* thereby indicating the potential application of this enzyme in the clinical setting (Urlacher and Eiben, 2006).

Focusing mainly on this family of P450s, studies conducted at our institution are in the process of developing cassette PCR's for nucleotide sequenced-based screening by isolating DNA directly from the environment with the intention of constructing chimeric genes encoding a hybrid protein that can be expected to have enzymatic activity. The ultimate goal is to reduce the complexity of the P450 systems by constructing self-sufficient systems which can function as catalysts without the need for costly co-factors thereby increasing their efficiency and reducing costs when applied to industry. However, this chapter focuses only on the retrieval of new DNA fragments belonging to this family from the Beatrix biofilm. The fragments that are obtained will be subjected to further experiments involving the construction of chimeric genes by a research group focusing mainly on the biocatalysis of P450s at our institution.

4.2 Materials and Methods

4.2.1 PCR screening for the presence of cytochrome P450 alkane hydroxylases from the metagenome

The properties of the sample material and DNA extraction procedures employed were discussed in Chapter 3, section 3.2.1. The extracted metagenome DNA was screened for the presence of alkane hydroxylases, using highly degenerate primers based on the conserved sequence motifs of CYP153 homologs. The primer pair CF and P450rv3 (Intergrated DNA Technologies) amplifies the large fragment (800 bp) of the CYP153 P450. An additional primer pair (P450FW1 and P450rv3) amplifies a highly conserved fragment (~339 bp) that is internal to the larger fragment. Both primer sets were used in this study.

The following cycling conditions were used: 4 min at 95°C; 25 cycles of 45 s at 95°C, 1 min at 58°C, and 1 min at 72°C; 5 min at 72°C; and 4°C until further use.

Table 4.1: Primers used for the amplication of the CYP153 gene (van Beilen *et al.,* 2006)

Primer	***Sequence***	***Reference***
CF	5`- ATGTTYATHGCNATGGAYCCNC-3´	Kubota *et al.* 2005
P450FW1	5`-GTSGGCGGCAACGACACSAC- 3´	van Beilen *et al.* 2006
P450rv3	5`- GCASCGGTGGATGCCGAAGCCRAA-3´	van Beilen *et al.* 2006

4.2.2 Cloning of the CYP153 fragment

The CYP153 fragment was cloned into pGEM®T-Easy (Promega). Ligation reactions were performed for 2 h at room temperature. The ligation products were then used to transform *E. coli* TOP10 chemically competent cells as described in section 2.2.3.1.5. Transformants were selected on LB agar plates containing 100 μg/ml of ampicillin, IPTG and X-gal. The plates were incubated overnight at 37°C.

4.2.2.1 Plasmid DNA extraction and restriction digestion analysis

Random colonies (20) were selected and inoculated into 5 ml of LB containing ampicillin (100 μg/ml). The tubes were incubated overnight at 37°C with mild shaking. Plasmid DNA extractions and restriction digests to confirm the presence of the cloned fragments were performed as described in section 2.2.3.1.6.

4.2.2.2 Sequencing of randomly selected clones

Insert DNA of recombinant clones were sequenced using an ABI 3130XL genetic analyzer (Applied Biosystems, Foster City, CA), incorporating the ABI Big Dye Terminator Cycle Sequencing kit version 3.1 (Applied Biosystems, Foster City, CA). Sequencing was performed at Inqaba Biotechnical Industries Pty. Ltd., South Africa. Electropherograms

generated from the sequences were inspected with FinchTV software (Geospiza). Nucleotide sequences were compared to known sequences deposited in the nucleotide database of the National Centre for Biotechnology Information (NCBI, USA) using a standard nucleotide BLAST (Altschul *et al.,* 1997).

4.3 Results and Discussion

4.3.1 Sequenced-based screening for the presence of cytochrome P450 alkane hydroxylases in the metagenome

We observed positive amplification for both the small (339 bp) and larger (800 bp) fragment of the CYP153 gene (Figure 4.1 and 4.2). The 800 bp amplification product was purified using the BioSpin Gel Extraction Kit (BioFlux) and used as template for the amplification of the 339 bp fragment which served as a confirmatory reaction because the 339 bp fragment is highly conserved within the CYP153 family of enzymes. The larger fragment was sub-cloned and the inserts sequenced to determine if we had obtained any P450 alkane hydroxylases.

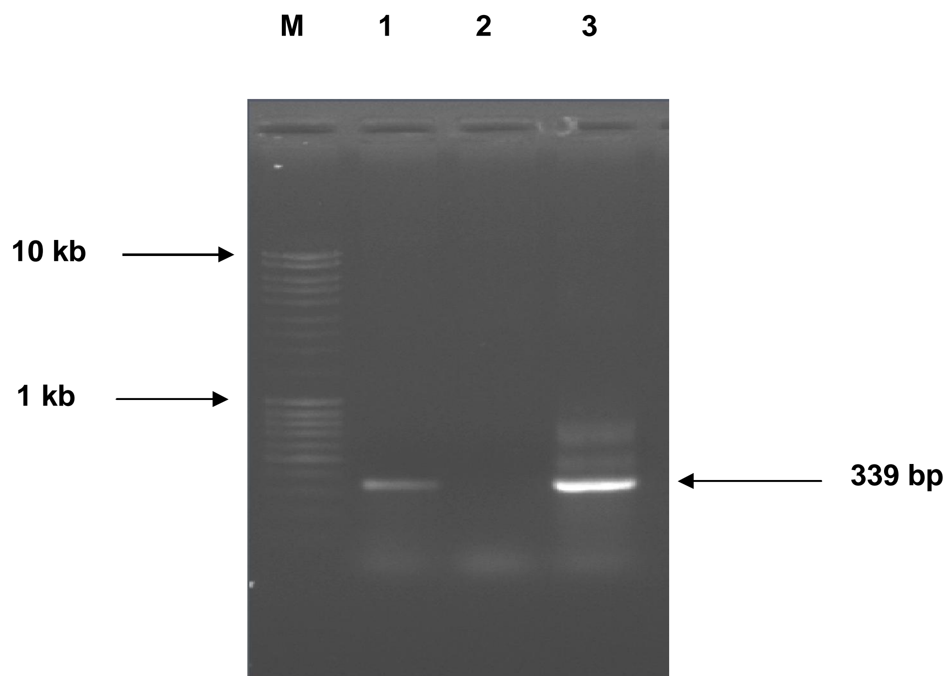


Figure 4.2: PCR amplification of the smaller fragment of the CYP153 gene. M: MassRuler DNA Ladder (SM# 0403); lane 1: positive control (*Pseudomonas putida)*; lane 2: negative control; and lane 3: metagenomic DNA.

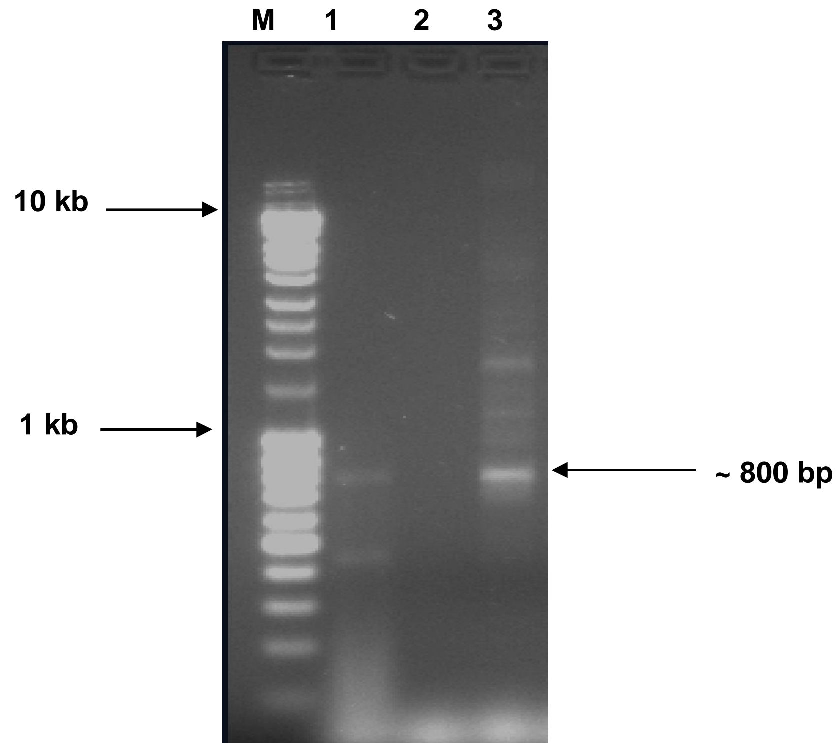


Figure 4.3: PCR amplification of the larger fragment of the CYP153 gene. M: MassRuler DNA Ladder (SM#0403); lane 1: positive control (*Pseudomonas putida*); lane 2: negative control; and lane 3: metagenomic DNA.

4.3.2 Cloning and sequencing of the CYP153 fragments

Restriction analysis of selected clones revealed fragments of the expected sizes for some of the clones (~800 bp) (Figure 4.4). Twenty clones carrying inserts of the correct size were sequenced.

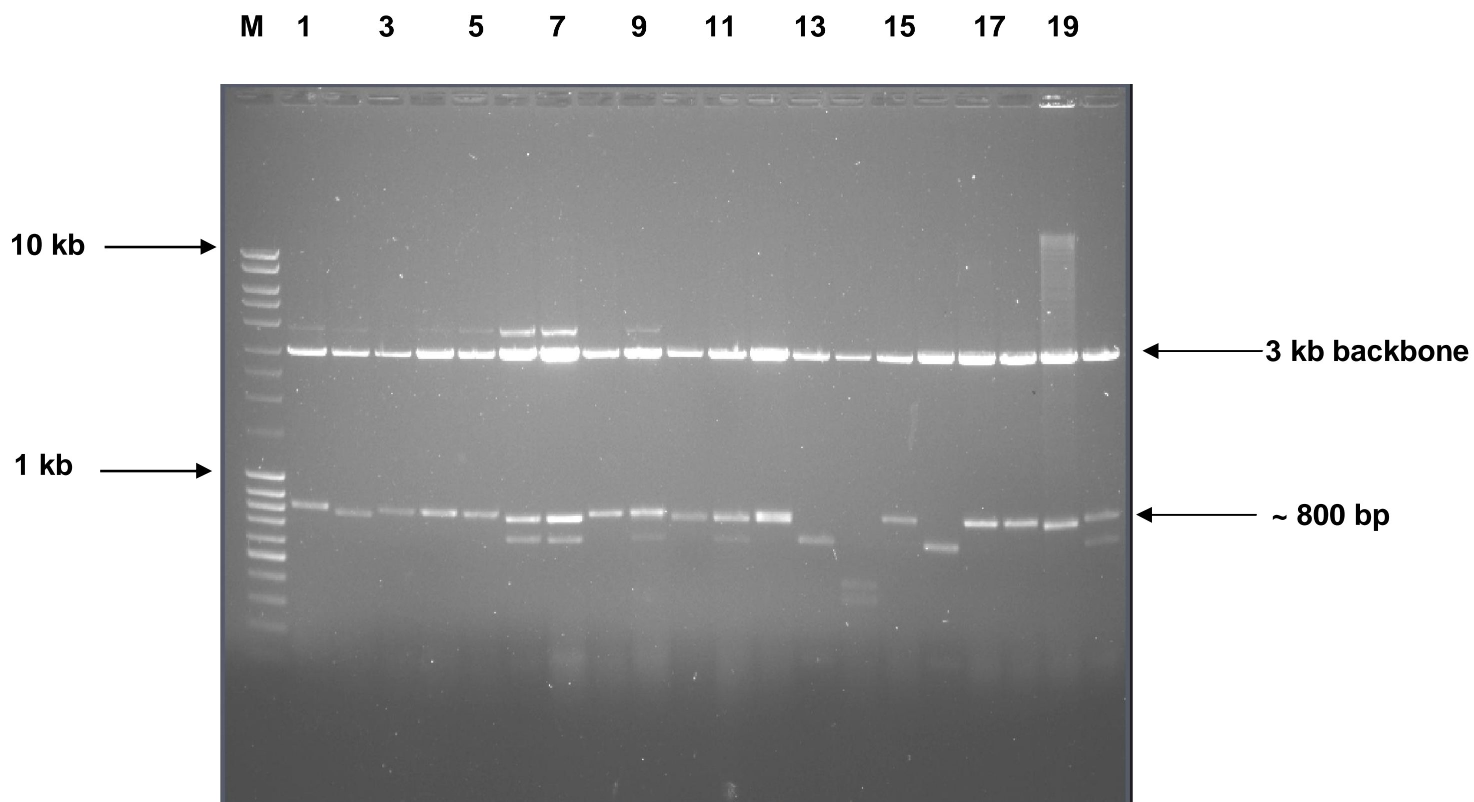


Figure 4.4: Restriction analysis of the cloned 800 bp product of the CYP153 gene. M: MassRuler DNA Ladder (SM #0403); and lanes 1-20: randomly selected clones.

Table 4.2: Sequence data of CYP153 clones

Clone	***Organism***	***% Identity***	***E value***	***Accession No.***
pCYP1	*Parvibaculum lavametivorans,* cytochrome P450	80 (644/800)	4e-160	CP000774
pCYP3	*Caulobacter sp*, cytochrome P450	76 (621/803)	3e-61	CP000927
pCYP4	*Caulobacter crescentus* cytochrome P450	76 (674/814)	3e-102	CP001340
pCYP5	*Bradyrhizobium sp*. cytochrome P450	77(525/680)	2e-93	CU234118
pCYP7	*Sphingopyxis macrogoltabida,* ahpJ gene (alkane hydroxylase)	74 (457/617)	1e-155	AJ850057
pCYP8	*Sphingopyxis macrogoltabida,* ahpJ gene (alkane hydroxylase)	76 (328/431)	5e-49	AJ850057
pCYP14	Uncultured cytochrome P450, Niigata 007	77 (645/837)	6e-114	AB206804
pCYP15	*Parvibaculum lavametivorans,* cytochrome P450	69 (451/649)	2e-58	CP000774
pCYP16	*Parvibaculum lavametivorans,* cytochrome P450	78 (647/820)	1e-141	CP000774
pCYP17	*Phenylbacterium zucinecum*	75 (629/835)	2e-88	CP000747

NB: *the numbers in parenthesis represents the number of nucleotides which are identical between the query and the subject sequences.*

According to the sequence analysis only 50% of the clones sequenced showed hits with that of cytochrome P450 genes. The remaining 50% of the clones showed hits with genes and enzymes completely unrelated to P450s. We found that 9/10 clones positive for P450s showed hits with that of known bacterial P450s, while only one clone hit with that of an uncultured P450. The known bacterial P450s to whichthe clones showed hits with were isolated from both an environmental as well as a clinical setting. *Parvibaculum lavametivorans* is an environmental isolate that utilizes many commercial anionic and non-ionic surfactants as well as alkanes as sole carbon sources. This organism requires a biofilm on a solid surface when utilizing alkyl surfactants (Schleheck and Cook, 2005). This co-relates with the type of sample material that was used (biofilm) for the detection of the P450 enzymes in this study. *Sphingopyxis macrogoltabida* is a sphingomonad, classified as an alpha-proteobacterium, and is present in diverse environments. In chapter 3, we identified the organisms present in the metagenome by performing diversity assessments. Some of

the 16S rRNA bacterial clones clustered with sphingomonads, indicating the presence of this group of bacteria in the sample material, therefore detecting a P450 showing identity to a P450 from this type of bacterium is not unusual. In addition certain *Sphingomonas* species have been shown to hydroxylate substituted pyrrolidines, pyrrolidones, piperidines and azetidines (all intermediates for interesting pharmaceuticals), and these organisms contain P450s belonging to the CYP153 family (Funhoff *et al.,* 2007).

Two of the clones showed hits with *Caulobacter* species. *Caulobacter crescentus* is known for its ability to survive in low-nutrient environments and heavy metal-contaminated sites. This bacterium was also isolated from a deep subsurface gold mine and has been shown to form biofilms with the potential for use in bioreactors for bioremediation (Hu *et al.,* 2005). *Bradyrhizobium sp*, a bacterium associated with plant root nodules has been shown to contain P450s that may be important for the survival of the microbe during nodule senescence or for persistence in soil (Keister *et al.,* 1999). One of our clones hit with a P450 from *Phenylbacterium zucinecum*, this bacterium is considered a human pathogen and was isolated from a clinical sample (Luo *et al.,* 2008). This indicates the diversity of the P450s present in the metagenome.

Phylogenetic analysis of the positive clones was performed in order to determine how the clones cluster when compared to other members of the CYP153 family. The known bacterial sequences that were included in the analysis were from the groups of bacteria that were used for the design of the CYP153 degenerate primers used in this study. In addition, uncultured bacterial P450 alkane hydroxylases isolated from oil contaminated soil in Japan were also included in the phylogenetic analysis (Kubota *et al.,* 2005). The phylogenetic tree was constructed using the bootstrap analysis, neighbour joining method which is part of the MEGA version 4.1 software package (Kumar *et al.,* 2008). The clones clustered on their own when compared to known and uncultured bacterial CYP153 P450s, indicating that they are quite different in their sequences (Figure 4.5). It is possible that we have identified novel variants of this family of enzymes during our screening process.

The ultimate aim of any study that employs metagenomic screening is to try and find the "ideal biocatalyst" that can be employed in the industrial setting. The positive P450 clones that we have obtained from the metagenome will be subject to further experiments. Firstly, attempts to obtain the full length CYP153 fragments will be employed; this will be performed

by using cassette PCR as described by Kubota *et al.* (2005). The full length genes will be used in expression studies and the catalytic functions of the respective chimeras will be explored. Once functional full length genes are obtained, the next step is to construct a self-sufficient P450 that has the hydroxylating and electron transfer subunits joined in a single polypeptide, indicating that the subunit docking step which is possibly rate-limiting in catalysis is eliminated thereby increasing the efficiency of this family of enzymes in industry. This is currently being investigated at our institution.

4.4 Conclusion

This study represents the first report of cytochrome P450 enzymes namely the CYP153 family, present in South African gold mines. According to Funhoff *et al.* (2007) CYP153's are the best, soluble CYP enzymes for terminal aliphatic hydroxylation, displaying high regio-specificity and activity compared with other P450 enzymes. The structural characterization and engineering of these enzymes will provide more insight into the enzymatic alkane activation and will support the development of alkane oxygenases for industrial application. Studies on alkane hydroxylases are no longer limited to a small number of well characterized enzymes. New and promising P450 alkane hydroxylases have been discovered by metagenomic screening.

We were able to identify P450 enzymes in the metagenome that did not cluster phylogenetically with any of the known bacterial P450s, indicating the vast potential for the identification of novel alkane hydroxylases. However, further studies that exploit the catalytic potential of the clones in the degradation of various alkanes will be of interest. Furthermore a diversity assessment of the CYP153 P450's in the deep mine will also serve as an interesting study.

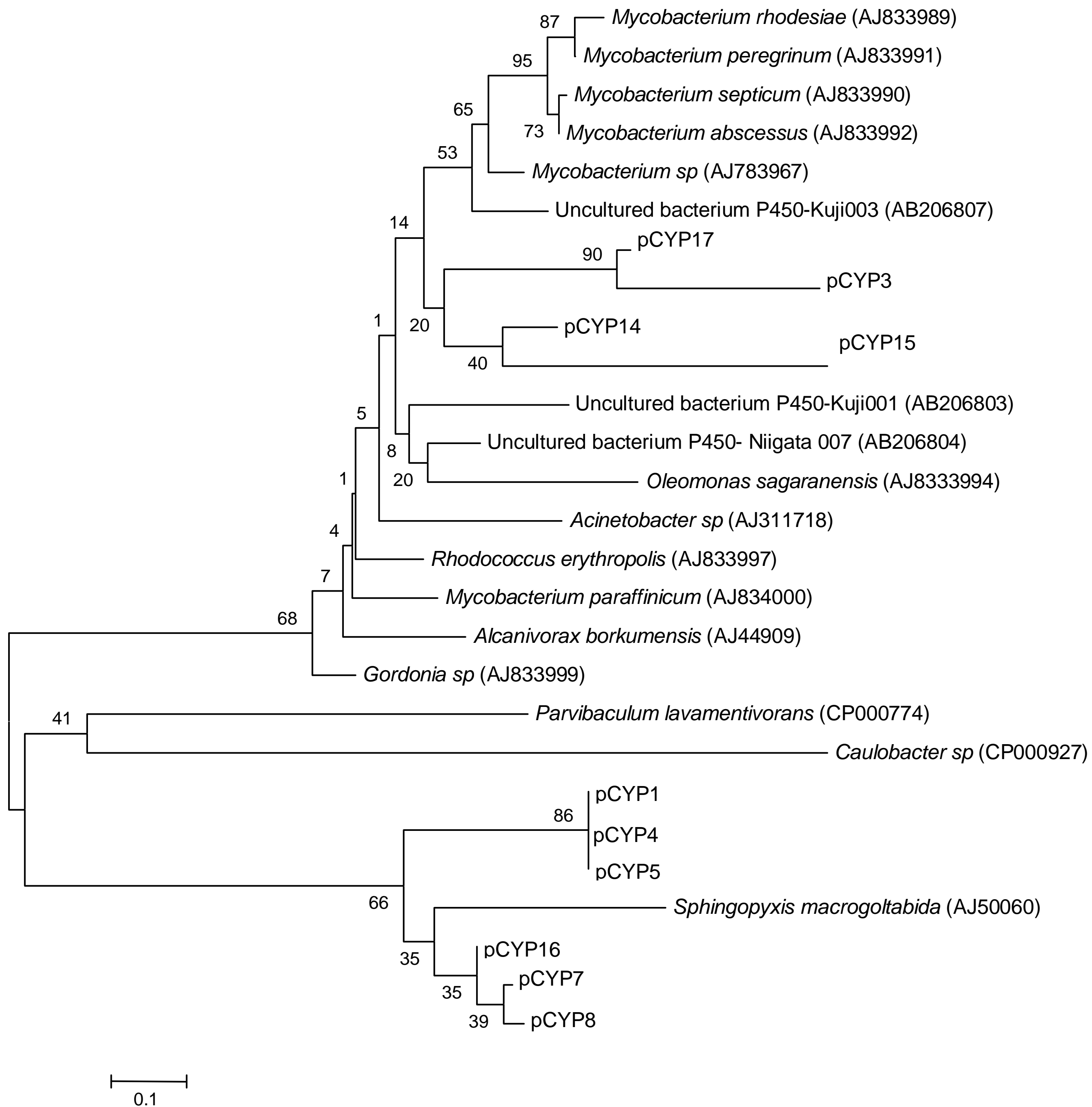


Figure 4.5: Phylogenetic tree of CYP153 cytochrome P450 family homologues to P450 clones from the Beatrix mine. The phylogenetic tree was constructed by the neighbour-joining method with MEGA version 4.1 software. The accession numbers of the aligned sequences are shown in parenthesis.

Chapter 5

Mining the metagenome for biocatalysts

5.1 Introduction

The great majority of prokaryotic species present in natural environments, termed the "metagenome" have never been cultured, identified or analyzed, thereby prompting a revolution in our thinking of the natural prokaryotic world and its relationship with the environment (Li, *et al.,* 2005). Although the analysis of 16S rRNA sequences gives insight into previously unsuspected biodiversity, it yields little information on the yet uncultured microorganisms. This shortfall has been partially addressed by the construction and analysis of DNA libraries, made directly from the environment, with no prior culturing (Schmeisser *et al.,* 2003). However, the success of this technology depends on the ability to clone fragments of DNA into appropriate vectors and the ability to express heterologous DNA effectively (MacNeil *et al.,* 2001). Random cloning of microbial DNA from the environment and subsequent screening of expression libraries for the presence of a desired enzyme activity has become a useful tool for the discovery of novel biocatalysts (Handelsman *et al.,* 1998).

Functional screening requires gene expression and proper folding of the resulting protein in a heterologous host, most frequently *E. coli,* which is not always easily achieved. The minimal set of requirements for gene expression includes the presence of a promoter for transcription, and a ribosome binding site (rbs) in the -20 to -1 region upstream of the start codon for iniation of translation (Makrides *et al.,* 1996). The formation of an active protein may also rely on *trans* factors that need to be provided by the host organism such as special transcription factors, inducers, chaperones, co-factors, protein-modifying enzymes, or proper secretion machinery (Gabor *et al.,* 2004). Although the common *E. coli* expression systems have been predicted to be able to successfully express up to 40% of genes with sequences that are available in the public databases, other alternative hosts for library construction and screening with different expression capabilities are currently under development [*Bacillus subtilis, Pseudomomas putida, Streptomyces lividans*] (Ferrer *et al.,* 2007). In this study three different esterases (isochorismatase, sulfatase and phospholipase patatin) were identified by metagenomic library construction and screening. Of the three enzymes, the

phospholipase patatin was further characterized. Patatin displays broad esterase activity but the physiological role of this enzyme remains unclear.

5.2 Materials and Methods

5.2.1 Small-insert library construction

The metagenomic DNA was digested with 1 U *Bam*HI (New England Biolabs) for 1 h at 37°C. The size selected for cloning was 2-6 kb. DNA in the correct size range was excised from the gel stained with SYBR Gold and purified using the GFX Gel Purification Kit (Amersham Biosciences) according to the manufacturer's instructions.

5.2.1.1 Ligation

Ligation reactions using T4 DNA ligase and ligase buffer (Fermentas) were incubated overnight at 16°C. For sticky-ended cloning a vector: insert ratio of 1:2 was used. The fractionated gDNA was ligated to the *Bam*HI digested pZerO-2 vector.

5.2.1.2 Bacterial transformation

Ligation mixtures were used to transform *E. coli* TOP10 competent cells by heat shock as described in section 2.2.3.1.5.

5.2.1.3 Plasmid DNA extraction and restriction analysis

Random colonies (20) were picked up and inoculated into 5 ml of LB containing kanamycin (50 µg/ml). The tubes were incubated overnight at 37°C with mild shaking. Plasmid DNA extractions and restriction digests were performed as described in section 2.2.3.1.6.

5.2.1.4 Library screening using the function-driven approach

Plate assays to detect lipolytic enzymes, amylases, proteases and beta-lactamases were conducted. Indexed libraries stored in 96-well microtitre plates were replica plated onto the selective screening media.

The following formula (Gabor *et al.,* 2004) was employed in order to determine the number of clones that needed to be screened to obtain a positive hit.

$$N_P = \frac{\ln(1 - P)}{\ln\left(1 - \frac{(I - X)}{G \cdot c \cdot z}\right)}$$

Np = number of clones that need to be screened

P = probability that a target gene will be recovered at least once (50% probability)

I = insert size

X = size of the gene of interest

G = genome size (3100kb according to Institute for genomic research)

c = correction factor

z = diversity (no. of bands present on DGGE gel)

5.2.1.4.1 Plate assay for lipolytic enzymes

Lipolytic activity was measured by the tributyrin plate halo assay. The plate assay was performed using an adapted protocol described by Ro *et al.* (2004), 1% tributyrin and 1% gum arabic (Sigma) was dissolved in 20 ml of sterile water and mixed by sonication until a milky homogenous suspension was obtained. The suspension was then added to LB agar, autoclaved and cooled before the addition of kanamycin (50 µg/ml). Clones were replica plated onto LB tributyrin (TLB) plates. Plates were incubated at 37°C until activity was detected.

Assaying for activity on LB tributyrin does not distinguish lipase from esterase activity due to the fact that both lipases and esterases are capable of hydrolyzing tributyrin, it was important to confirm lipase activity for the clones that produced positive results on the tributyrin plates. A second plate assay was done using LB olive oil and rhodamine B (Sigma) (Kouker and Jaeger, 1987). Lipases only are able to hydrolyze olive oil because of their affinity for longer carbon chained substrates (Chahinian *et al.,* 2002), 1% olive oil and 1% gum arabic was dissolved in 20ml sterile water and mixed by sonication to obtain a homogenous suspension. Rhodamine B (1 mg/ml) was dissolved in sterile water and filter sterilized. The rhodamine B solution (0.001%) and the olive oil was then added to LB agar,

autoclaved and cooled before the addition of kanamycin (50 μg/ml). Agar plates containing the olive oil and rhodamine B appear pink and opaque. Positive clones were replica plated onto the agar plates and incubated at 37°C until activity was detected. Lipase production was monitored by irradiating plates with UV light at 350 nm. An orange fluorescence formed around a colony was indicative of a positive result.

5.2.1.4.2 Protease plate assay

The protease plate assay was performed using an adapted protocol described by Vazquez *et al.* (1995), the skim milk medium used for screening of protease activity on plates was composed of 10 g skim milk powder (Oxoid, Basingstoke, United Kingdom). The skim milk was dissolved in 100 ml of sterile water and autoclaved. LB agar was prepared and autoclaved, the skim milk suspension was then added and mixed properly, and the media was allowed to cool before the addition of kanamycin (50 μg/ml). The clones were replica plated onto the skim milk plates and incubated at 37°C until activity was detected. A clear halo around a colony is indicative of a positive result.

5.2.1.4.3 Amylase plate assay

The amylase plate assay was performed according to Omemu *et al.* (2005). Soluble starch was prepared in house. Soluble starch (2 g) was resuspended in 60 ml sterile water and dissolved at room temperature. Remazol Brilliant Blue (0.5 g) (Sigma) was then added to the mixture and dissolved after the drop wise addition of Na_2SO_4 (1 mg/ml), and NaOH, 7.5% w/v. The mixture was stirred for 1.5 h at room temperature, followed by the addition of ethanol (98%) to precipate the conjugated starch. The mixture was incubated at -20ºC for 1h to complete precipitation. The product was collected on a vacuum filter and washed with 100 ml aliquots of 67% buffered ethanol (1 volume 0.05 M sodium acetate buffer, pH 5.4 and 2 volumes absolute ethanol) until the filtrate was free from all traces of blue colour. The product was collected by centrifugation (5 000 rpm, 5 min) and washed several times with 200 ml aliquots of 100% ethanol until the filtrate was free from all traces of blue colour. Larger clumps were broken up to ensure effective washing. This step was important, since large amounts of unbound dye may interfere with the identification of active clones. After a colourless washing filtrate was obtained, the product was dehydrated by rinsing with 100 ml 80% ethanol, followed by 95% ethanol and finally acetone. After drying under a vacuum overnight, approximately 2 g of conjugated starch was obtained; the product was pulverized in a mortar, and stored at room temperature.

The screening medium for amylase activity contained LB agar and 01% blue starch. One gram of conjugated starch was added to 100 ml of hot sterile water, and resuspended. The blue starch was then added to the LB agar, mixed properly and autoclaved. The media was allowed to cool before the addition of kanamycin (50 μg/ml). The clones were replica plated onto the starch plates and incubated at 37°C until activity was detected. A clear halo around a colony is indicative of a positive result.

5.2.1.4.4 Beta-lactamase plate assay

The beta-lactamase assay was performed according to Gabor *et al.* (2003). LB agar was prepared and autoclaved. The media was allowed to cool before the addition of kanamycin (50 μg/ml), and 100 μg/ml **ampicillin**). The clones were replica plated onto agar plates and incubated at 37°C. Clones that produce beta-lactamases will grow on the media supplemented with **ampicillin** (100 μg/ml) due to their ability to cleave the beta-lactam ring of the antibiotic, rendering it inactive.

5.3.1 DNA sequencing and analysis

Insert DNA of positive recombinant clones were sequenced by primer walking using an ABI 3130XL genetic analyzer (Applied Biosystems, Foster City, CA), incorporating the ABI Big Dye Terminator Cycle Sequencing kit version 3.1 (Applied Biosystems, Foster City, CA). Sequencing was performed by Inqaba Biotechnical Industries Pty. Ltd., South Africa. Electropherograms generated from the sequences were inspected with FinchTV software (Geospiza) and Vector NTI (Invitrogen) under a private license. Translated ORFs were compared to known sequences deposited in the non-redundant protein databases of the National Centre for Biotechnology Information (NCBI, USA) using standard protein-protein BLAST (*blastp*) (Altschul *et al.,* 1997). Multiple sequence alignments were carried out by ClustalW (Thompson *et al.,* 1994).

5.4 Heterologous expression of the patatin in *E. coli*

5.4.1 Bacterial strains, plasmids and growth conditions

All bacterial strains and plasmids used in this study are listed in Table 5.1. *E. coli* TOP10 and BL21 (DE3) pLysS competent cells were used as hosts for genetic manipulation and expression of proteins. *E. coli* strains were grown in LB broth supplemented with kanamycin

(50 µg/ml) overnight at 37°C with shaking for the selection of the plasmids carrying inserts. Plasmid pET-28b(+) (Novagen) was used for the expression of the protein in *E. coli* BL21(pLysS) competent cells.

Table 5.1: Bacterial strains and plasmids used in this study

Strain/plasmid	***Description***	***Reference***
Escherichia coli TOP10	One Shot TOP10 chemically competent cells FP-P *mcr*A (*mrr-hsd*RMS-*mcr*BC) 80*lac*Z M15 *lac*X74 *rec*A1 *ara*D139 (*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G	Invitrogen
Escherichia coli BL21 (DE3) pLysS	*E.cloni* EXPRESS BL21(DE3) chemically competent cells F^- *ompT* $hsdS_B$ (r_B- m_B^-) *gal dcm* (DE3)	Lucigen
pGEM®T- Easy	Amp^r, T7 and SP6 promoter, *Lac*Z, ori	Promega
pET-28b(+)	Kan^r, T7 promoter, *Lac*I, N-terminal His-Tag and Thrombin configuration, ori	Novagen

5.4.2 Construction of expression plasmids

5.4.2.1 PCR amplification of patatin

The complete patatin gene was amplified by PCR using plasmid DNA from clone pNS6 as the template. The Expand High Fidelity PCR system (Roche) was used for the amplification reaction. PCR reactions were performed in a total reaction volume of 50 µl using a Thermal Cycler (PxE 0.2, Thermo Electron Corporation). Reaction mixtures consisted of 10X Expand High Fidelity Buffer with 15 mM $MgCl_2$ (5 µl), dNTP's (100 mM), Expand High Fidelity Enzyme mix (0.75 µl), 200 ng of plasmid DNA and 300 ng of both the forward and reverse primer. The primer set is given in (Table 5.2). Reaction conditions consisted of an initial denaturing step at 94°C for 2 min, followed by 30 cycles of denaturing at 94°C (30 s),

annealing at 58°C (30 s) and elongation at 72°C (2 min). A final elongation step of 10 min at 72°C was added to ensure complete elongation of the amplified product.

Table 5.2: Primer set used for expression cloning

Primer	***Sequence***	T_M
PATA-F	5'-GGCATATGGGCCTGACGCTGATC3'	62.2°C
PATA-R	5'- TCAAGCTTTCAGGCGCTCTCG-3'	60°C

Underlined sequences indicate introduced restriction sites for NdeI and HindIII

PCR products were excised from a 1% agarose gel stained with GoldView and purified using the BioSpin Gel Extraction Kit (Bioflux) according to manufacturer's instructions.

5.4.2.2 Constructs for expression in E. coli

The purified PCR product was ligated into pGEM®T- Easy vector, the ligation reaction was performed at 22°C for 2 h. The ligated product (50 ng of DNA) was then used to transform *E. coli* TOP10 competent cells by heat shock transformation (Invitrogen) as described in section 2.2.3.1.5. Plasmid DNA extractions and restriction digests were performed as described in section 2.2.3.1.6. For ligation into the pET-28b(+) vector, inserts were double digested with *Nde*I (10 U) and *Hind*III (5 U) at 37°C for 4 h. The pET-28b(+) vector is designed for expression of N-terminal 6XHis tagged protein. The digested product was excised from a 1% agarose gel stained with GoldView and purified using the BioSpin Gel Extraction Kit (Bioflux) according to manufacturer's instructions. The ligated product was used to transform *E. coli* TOP10 competent cells. Transformants were selected on LB agar plates supplemented with kanamycin (50 µg/ml) and incubated overnight at 37°C. Positive clones were identified through plasmid DNA isolations and restriction digestion using *Nde*I (10 U) and *Hind*III (5 U) at 37°C for 2 h. The positive plasmid was introduced into *E. coli* BL21 (DE3) pLysS competent cells and transformants were selected on LB agar plates

supplemented with kanamycin (50 μg/ml), chloramphenicol (34 μg/ml) and incubated overnight at 37°C.

5.4.2.3 Expression and purification of the patatin

A culture of *E. coli* (transformed with the patatin gene containing pET vector) was grown to (A_{600} 0.2-0.4) and thereafter induced by the addition of 0.5 mM IPTG and grown for an additional 4 h at 30°C. Cells were harvested by centrifugation and treated with the B-PER solution (Pierce, USA). Cells were harvested by centrifugation and resuspended in 20 mM MOPS (pH 7.4) containing 50 mM imidazole and 0.5 M NaCl [approximately 1 g cells (wet weight) in 10 ml]. Cells were broken by sonication for 5 min (5 cycles at 80 W), unbroken cells and debris were removed through centrifugation (8 000 rpm for 10 min). The soluble fraction was separated from the insoluble fraction by ultracentrifugation at 30 000 rpm for 90 min. The soluble fraction was loaded onto a HisTrap FF column (5 ml, Amersham Biosciences) and unbound proteins eluted (5 $ml.min^{-1}$) using 20 mM MOPS (pH 7.4) containing 50 mM imidazole (Sigma) and 0.5 M NaCl (Merck). Bound proteins were then eluted in the same buffer using a linear gradient (100 ml) of imidazole up to 0.5 M. Fractions were subjected to SDS-PAGE analysis and visualized with Coomassie Blue R-250 (Sigma).

5.4.2.4 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Electrophoresis under denaturing conditions or sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed using the “Mighty Small” miniature slab gel electrophoresis unit, SE 200, from Hoefer Scientific Instruments. The protocol used was that described by Laemmli (1970) using a 10% resolving gel and 4% stacking gel. Precision Plus Protein Standards (BioRad) was used as the molecular weight marker for the SDS–PAGE, and proteins were visualized by staining polyacrylamide gels with Coomassie Blue R- 250.

5.4.2.5 Protein concentration determination

Protein concentration was determined using the bicinchoninic acid (BCA) method (Smith *et al.*, 1985). BCA Protein Assay Kit from Pierce (Rockford, IL, USA) was used according to the manufacturer’s instructions with bovine serum albumin (BSA) as the standard (supplied with kit).

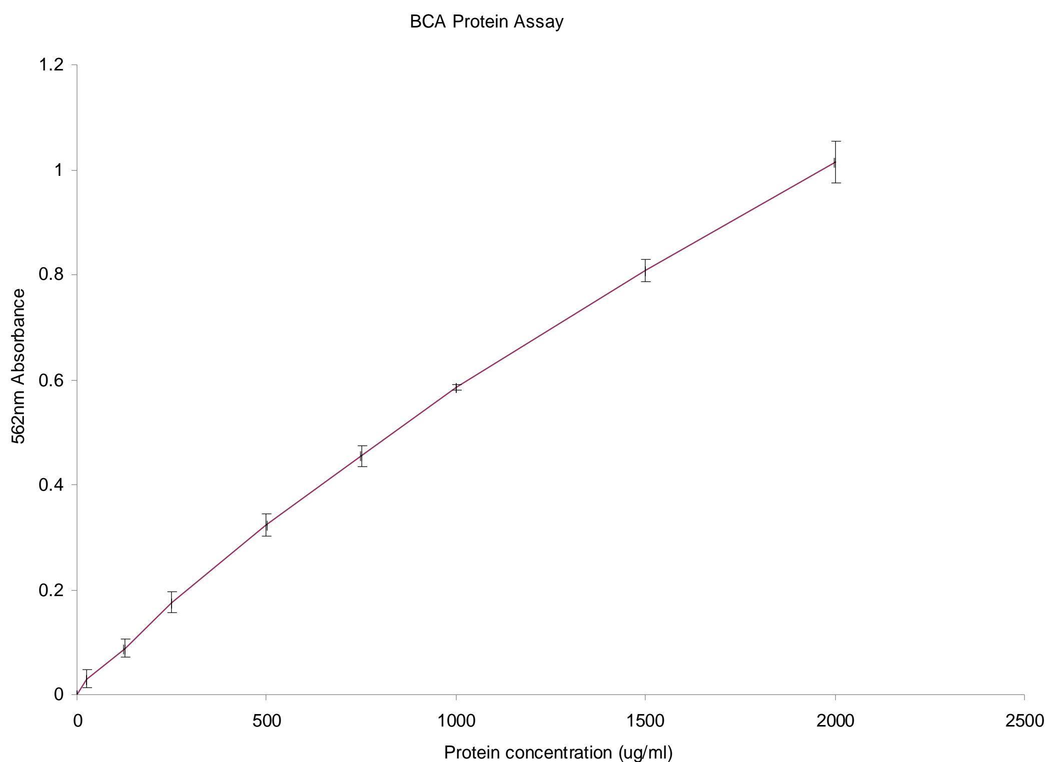


Figure 5.1: Standard curve for the BCA protein assay kit (Pierce) at 37°C (enhanced method) using BSA as protein standard. Error bars indicate standard deviation after performing the experiment in triplicate.

5.4.2.6 Biochemical analysis

To examine substrate specificity and to find the best substrate for activity of the patatin we tested various *para*-nitrophenol ester substrates with C_4-C_{16} acyl chain lengths (Sigma). The production of *p*-nitrophenol was continuously monitored at 405 nm by use of a DU-650 spectrophotometer (Beckman) with temperature control at 25°C, for 5 min. The assay contained 600 µl of 100 mM Tris/HCl (pH 8) buffer and 0.5 mM substrate. Substrate stocks of *p*-NP-butyrate, caproate, caprylate, caprate, and palmitate were dissolved in isopropanol. Blank rates were determined in order to subtract the appropriate values for non-enzymatic hydrolysis of substrates from the results. Activity was defined as the amount of enzyme required to release 1 µmol of *p*-nitrophenol/min from the *p*-nitrophenyl ester.

5.4.2.6.1 Effects of temperature on enzyme activity

The optimal temperature for enzyme activity was determined for a temperature range of 20 to 40°C under standard assay conditions.

5.4.2.6.2 Effects of pH on enzyme activity

For the determination of the optimum pH of the enzyme, activity was measured for a pH range of 7 to 10 under standard assay conditions. The buffers used were 100 mM HEPES (Sigma) (pH 7 to 7.5) and 100 mM Tris-HCl (pH 8 to 10).

5.5 Large-insert library construction

Several attempts to construct a large-insert fosmid library with the biofilm were unsuccessful using the instruction manual provided with the pCC1FOS vector, even with slight modifications of the protocol which worked perfectly when we constructed the *G. thermoleovorans* GE-7 fosmid library. We then obtained a protocol of the DOE Joint Genome Institute, Department of Energy, Office of Science (USA) for the construction of the fosmid libraries. We then decided to use this protocol and adapted it for deep mine sample because it was a protocol that was specifically used to construct fosmid libraries using pure culture. This part of the project was done at the Georg-August Institute of Microbiology and Genetics, University of Göttingen, Germany. The genomic DNA was blunt-end repaired using the repair enzymes provided with the kit, the reaction was incubated at room temperature for 2 h. Following incubation the sample was purified using the SureClean solution (Bioline, USA).

5.5.1 Ligation into pCC1FOS fosmid vector and phage packaging

For the ligation reactions a primary and secondary ligation was performed using ~100 ng/µl of DNA. The primary ligation was performed overnight at 16°C using the FastLink Ligase and buffer provided with the Kit. After the overnight ligation step, an additional 1 U of FastLink Ligase was added to the sample and incubated at room temperature for 2h. Thereafter the packaging reactions were performed at 30°C for 2 h (2 cycles of 2 h). Prior to packaging the infection cells were prepared; the cells used were phage resistant *E. coli* EPI 300 and the pre-inoculum was obtained by inoculation 0.5 µl of the host cells into 50 ml of LB broth containing 10 mM $MgSO_4$. The cells were grown overnight at 37°C with shaking. To a second flask of 50 ml of LB broth containing 10 mM $MgSO_4$, 5 ml of the pre-inoculum was

added, the flask was incubated at 37°C with shaking until the cells reached A_{600} 0.8-1.0. To 100 µl of infection cells 10 µl of the packaged extract was added and incubated at 37°C for 1.5 h (no shaking). The library was then plated out onto LB + chloramphenicol (12.5 µg/ml). The plates were incubated overnight at 37°C.

5.5.2 Fosmid DNA extractions and restriction analysis

Fosmid DNA extractions and restriction digests were performed as described in section 2.2.4.1.3.

5.5.3 Fosmid library screening

The fosmid library was screened for the presence of clones exhibiting antibacterial properties as well as for the presence of clones harbouring antibiotic resistance genes.

5.5.3.1 Antibacterial assays

For antibacterial assays, clones were grown on LB+chloramphenicol (12.5 µg/ml) plates overnight at 37°C. The plates were then overlaid with top agar containing exponentially grown *Pseudomonas putida* (KOB 1), a chloramphenicol resistant organism. The plates were incubated at 30°C for 48 h to facilitate the development of a lawn for the *P. putida* culture. The plates were incubated for a further 5 days in order to detect activity (MacNeil *et al.,* 2001).

5.5.3.2 Detection of antibiotic resistance

The antibiotic screens were carried out using the disc diffusion method. The clones were grown on LB+chloramphenicol (12.5 µg/ml) plates overnight at 37°C. Antibiotic discs (penicillin G [2 U]; cefepime [30 µg]; ciprofloxacin [5 µg]; and streptomycin [10 µg]) (Davies Diagnostic) were placed on the plates containing the clones. The plates were then incubated overnight at 37°C. Clones containing resistant genes for the respective antibiotic would be able to grow in the vicinity of the antibiotic disc, whereas for clones that were sensitive a zone of inhibition would be observed around the respective antibiotic disc.

5.6 Results and Discussion

5.6.1 Small-insert library construction

5.6.1.1 Partial digestion of environmental DNA

The gDNA extracted from the biofilm as described in section 3.2.1 was readily digested with restriction enzymes indicating that potential contaminants that inhibit the activity of the restriction enzymes were removed during the extraction process.

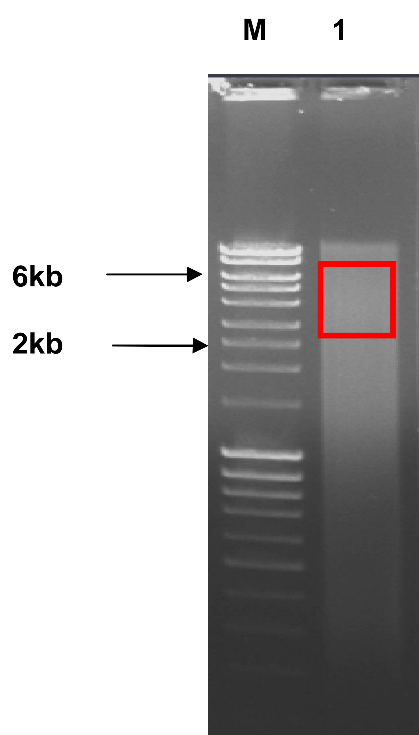


Figure 5.2: Partial digestion of biofilm gDNA. M:MassRuler DNA Ladder (SM#0403- Fermentas) and lane 1: - *Bam*HI digestion (Fermentas).

5.6.1.2 Small-insert library

A metagenomic library was successfully constructed. The library contained ~10 000 clones with an average insert size of 4 kb (Figure 5.3), 3 400 clones were stored as an indexed library in sterile microtitre plates, and the remaining clones were stored as a glycerol stock.

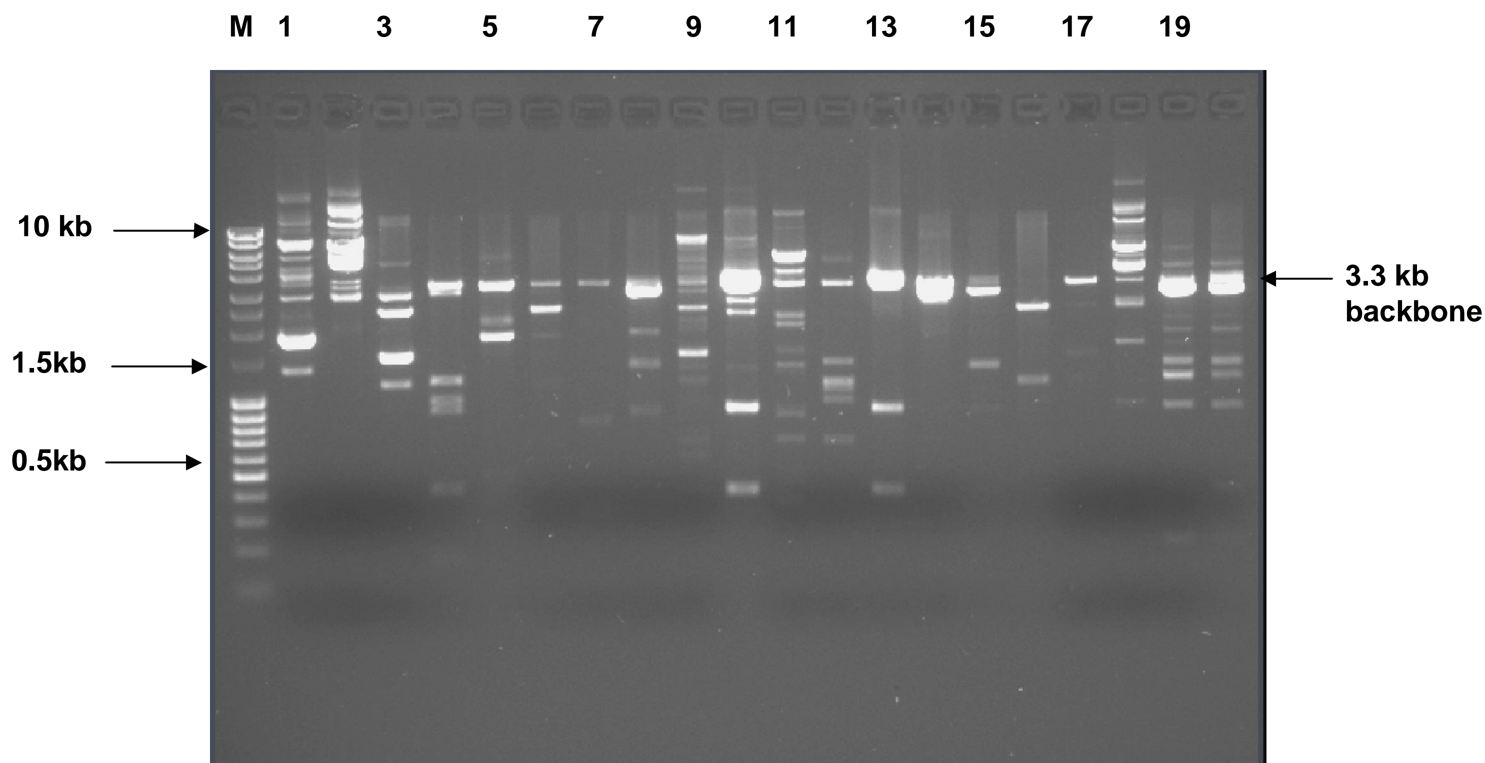


Figure 5.3: Restriction analysis of clones from the Beatrix library. M: MassRuler DNA Ladder (SM#0403-Fermentas); and lanes: 1-20 -randomly selected clones.

5.6.1.3 Library screening using the function-driven approach

No activity was observed for the protease, amylase and beta-lactamase plate screens. This could be attributed to the fact that according to the formula by Gabor *et al.* (2004) the number of clones that needed to be screened was >50 000 and our library only contained ~10 000 clones. The other possible reason could be that the gene/s responsible for protease, amylase and beta-lactamase activity was not present in the metagenome, which is unlikely,or that the proteases may have been inactive or toxic to the host cells. Despite our library being smaller than the required size as stipulated by the formula we were still able to identify 44 positive clones on TLB, which all showed the characteristic clear halo on the screening plate (Figure 5.4). After isolation of the plasmids, re-transformation and re-screening 18 of the 44 clones were still active. The zone of hydrolysis on tributyrin by different clones varied from 5 mm to 13 mm in diameter after 3 days of incubation at 37°C. To further confirm that the observed phenotype could be attributed to the metagenomic DNA

insert, recombinant plasmid DNA was isolated and the presence of an insert was confirmed by restriction analysis for all clones isolated. Based on the zone diameters, 10 out of 18 clones (large clearance zones) were sequenced. The insert size in these plasmids (pNS1-pNS10) varied from 1.5-5 kb. The sizes of the inserts after re-transformation were consistent with the previous restriction analysis of the respective plasmids. As no IPTG was added to the screening plates, the activity obtained was from the native promoter in the clones (Ranjan *et al.*, 2005). In addition, the clones were plated out onto LB olive oil and rhodamine B containing agar plates; however none of the clones produced any fluorescent halo indicating that the enzymes produced by the clones were possibly esterases.

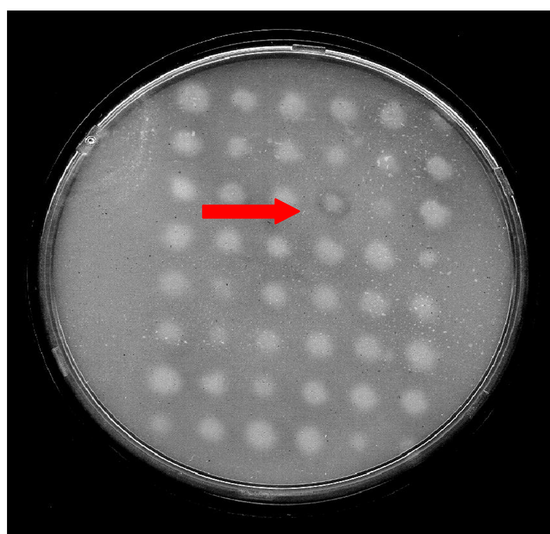


Figure 5.4: A clone showing activity on LB tributyrin as indicated by the red arrow.

5.6.1.4 Sequencing reactions

Analysis of the sequencing data revealed the presence of 3 different esterase type enzymes (isochorismatase, sulfatase and phospholipase, patatin) from the metagenome library (Table 5.3). However the remaining clones showed hits with enzymes completely unrelated to that of lipolytic enzymes (esterases) indicating the drawback of screening with tributyrin because of the large number of false positives. The reason for the presence of false positives is still unclear; there was no loss of sections of the insert during the re-transformation and re-screening. This was confirmed by performing RFLP analysis of selected clones before and after re-transformation and re-screening, the RFLP patterns were consistent. The presence

of a large number of false positives during screening of a metagenomic library using TLB was also reported by Jung *et al* (2003), who observed no close correlation between halo formation and whole cell enzyme activity. This was also observed with the *G.thermoleovorans* genomic libraries that were screened for lipolytic activity.

Table 5.3 Lipolytic activity conferring plasmids and sequence similarities

Plasmid	*Function and Organism*	*% Similarity*	*E value*	*Accession No.*
pNS1	**Isochorismatase family protein, *Geobacter sulfurreducens PCA***	**92**	**3e-73**	**NP_953629**
pNS2	**Sulfatase, *Parvibaculum lavamentivorans DS-1***	**74**	**2e-52**	**YP_001414401**
pNS3	Molybdopterin oxidoreductase Fe4S4*, Alcanivorax sp.*	45	2e-149	DX90729
pNS4	Type I restriction-modification system, M subunit*, Brucella ovis*	47	7e-04	ABQ62383
pNS5	DNA-dependent ATPase, SNF2 family protein, *Chlorobium tepidum TLS*	87	8e-153	NP_661799
pNS6	**Phospholipase, patatin family protein, *Plesiocystis pacifica SIR-1***	**55**	**2e-119**	**ZP_01911272**
pNS7	Primosomal protein, *Acidovorax sp. JS42*	66	1e-96	YP_988045
pNS8	Oxidoreductase, *Anaeromyxobacter dehalogenans 2CP-1*	57	3e-41	ZP_02325211
pNS9	Formate dehydrogenase, *Rhodococcus sp. RHA1*	49	1e-107	YP_700405
pNS10	Putative RHS-related transmembrane protein, *Ralstonia solanacearum*	50	5e-31	NP_523144

***NB:** plasmids marked in bold were identified to be responsible for the observed enzymatic activity.*

Clone pNS1 encoded a protein designated as an isochoristmatase. Isochoristmatase is a cysteine hydrolase that is involved in bacterial siderophore and phenazine biosynthesis. This enzyme is shown to be remarkably similar to the alpha-beta hydrolase family of enzymes (Parsons *et al.,* 2003; Rusnak *et al.,* 1990). BLASTP analysis revealed a moderate similarity (60%) between pNS1 and isochorismatases from various bacteria, including isochorismatase family protein (56%) from *Geobacter sulfurreducens* (NP_953629); isochorismatase family protein (55%) from *Nitrosomonas europaea* (NP_842301); isochorismatase family hydrolase (60%) from *Pseudomonas putida* (YP_001747104). Further sequence analysis revealed a putative purine rich ribosomal binding site located 7bp upstream of the ATG start codon. Multiple sequence alignment comparisons of the pNS1 to other isochorismatases from various bacteria revealed the presence of two (R; D) of the three residues (K/R; C; D) forming the catalytic site of this enzyme (Figure 5.5). Alignment of the best 100 hits with pNS1 showed that this catalytic Cys is conserved in all but one sequence, the isochorismatase from *Methylobacterium chloromethanicum* CM4 (YP_002418956). A similar observation was made by Caruthers *et al.* (2005), two isochorismatase proteins 1J2R (GI: 46014891) and 1NF8 (GI: 33357446), both of which belong to the CSHase superfamily did not contain the catalytic cysteine when compared to other isochorismatases. However, in our study no further characterization could be performed due to the unavailability of the substrate chorismic acid or isochorismic acid that is required for kinetic studies. If activity were illustrated, this would probably imply a catalytic mechanism differing from the accepted one, indicating the novelty of the Isochorismatase isolated from the deep mine metagenome.

Isochorismatase is an example of a group of enzymes in which an enol ether group is involved in an enzyme-catalyzed group transfer reaction. These enzymes are classified into two categories, which are referred to as type I and type II. Type I enzymes use phosphoenolpyruvate as the donor to effect the addition of the enolpyruvyl group to an alcohol function to form vinyl ether bond. Type II enzymes, of which isochorismatase is an example, function as hydrolases to cleave a vinyl ether bond, releasing both alcohol and carbonyl products (Parsons *et al.,* 2003; Rusnak *et al.,* 1990).

Phenazines exhibit antimicrobial activity that provides *Pseudomonas* with a competitive advantage in certain environments and may enhance their potential as opportunistic

pathogens in humans and other animals. Chorismic acid has been shown to be a precursor for phenazine biosynthesis (Parsons *et al.,* 2003). In addition phenazine may be used for the production of antitumor drugs and interest in phenazines has been associated with the development of synthetic analogues tailored to particular therapeutic applications (Blankenfeldt *et al.,* 2004). Thus far this is the only known application of isochoristmatases.

```
Pseudomonasentomophila      MSKFT-YNRLNKDDAAVLLVDHQAGLLSLVRDIEPDRFKNNVLALADLAK 49
Pseudomonasputida           MTHFK-YNRLNKDDAAVLLVDHQAGLLSLVRDIEPDAFKNNVLALADLAK 49
Ochrobactrum                MTKP--YVRLNKDDVAVLLVDHQTGLLSLVRDFDPDKFKNNVLALADLAK 48
Pseudomonasfluorescens      MSTPT-YNRLNKDDAVVLLVDHQTGLISLVQDFSPNEFKNNVLALADLAK 49
Nitrosomonaseuropaea        MSKPFKYSRLSKDDAALLLVDHQAGLISLVQDFGPSEFKNNVLAVGACGK 50
pNS1                        MRDSFQYNRLSKDDAALLLVDHQSGLISLVQDFSPGEFKNNVLAVAACGA 50
Pseudomonasaeruginosa       -------MLIRAATSTLLVVDIQERLLPAIDDGP--ALVEYSQWLLRVAR 41
                                     :     .:*:** *  *:. : *     : :    :   .

Pseudomonasentomophila      YFNLPTILTTSFEQGPNGPLVPELKALFPDAPYIARPGQINAWDNEDFVK 99
Pseudomonasputida           FFSLPTILTTSFEQGPNGPLVPELKALFPDAPYIARPGQINAWDNEDFVK 99
Ochrobactrum                YFKLPTVLTTSFENGPNGPLVPELKETFPDAPYIARPGQINAWDNEEFVK 98
Pseudomonasfluorescens      FFELPTILTTSFEQGPNGPLVPELKEMFPDAPYIARPGQINAWDNEDFVK 99
Nitrosomonaseuropaea        YFKLPTILTTSFEEGPNGPLVPELKEMFPNAPYIARPGNINAWDNEDFVN 100
pNS1                        YFELPTILTTSFEEGPNGPLVPELKEAFPDAPYIAAAGEHQCLGQRGLRR 100
Pseudomonasaeruginosa       ALDVPVLASEQYSKG-LGPTVAALRDEL-DATQILEKLDFSAAADGALLR 89
                             :.:*.: : .:.:*  ** *. *:  : :*. *    : ..  :  : .

Pseudomonasentomophila      AVKATGKKQLIIAGVVTEVCVAFPALAALEEEFEVFVVTDASGTFNEMTR 149
Pseudomonasputida           AVKATGKKQLIIAGVVTEVCVAFPALAALEEEFEVFVVTDASGTFNAMTR 149
Ochrobactrum                AVKATGKKQLLIAGVVTEVCVAFPALSAIEAGFDVFVVTDASGTFNEVTR 148
Pseudomonasfluorescens      AIKATGRKQIIIAGVVTDVCVAFPTLSALAEGFDVFVVTDASGTFNTTVQ 149
Nitrosomonaseuropaea        AVKNTGRKQLIIAGVVTEVCVAFPALSALEQGYEVFVITDASGTFNEVTR 150
pNS1                        GRQEDRAPAADPRGRRDGGLRGVSALSASRGG--VRGVRDSPDRFGNVQR 148
Pseudomonasaeruginosa       APGGD-RRQFVVCGSEAHVCVLQTVLDLLGRGREVFVVEEAIGSRRPSDK 138
                            .            *         ..*        *  : :: .      :

Pseudomonasentomophila      DAAHNRMSQAGAQLMTWFGVACELHRDWRNDVEGLAALFSNHIPDYRNLI 199
Pseudomonasputida           DAAHDRMSRAGAQLMTWFGVACELHRDWRNDIEGLAALCSNHIPDYRNLM 199
Ochrobactrum                HSAWDRMSAAGVQLMTWFGVACELHRDWRNDIEGLGTLFANHIPDYRNLI 198
Pseudomonasfluorescens      QAAWNRMTQAGAQMMNWFSVACELHRDWRNDIEGLGNLLSQRIPNYRNLM 199
Nitrosomonaseuropaea        HTAWLRMQAAGVQLINWFAMACELHRDWRNDIEGLGELFSNHIPNYRNLM 200
pNS1                        GDASRGMAADAGGRR----AADELVRDVV--------------------- 173
Pseudomonasaeruginosa       ALAVERMRQAGAMIVSREMVAFEWMERAGSD-------------RFREIS 175
                              *   *   .         * *  .

Pseudomonasentomophila      TSYNALTAGK- 209
Pseudomonasputida           TSYNAFNAGK- 209
Ochrobactrum                TSYNTMTSGK- 208
Pseudomonasfluorescens      NGYAALTAHQK 210
Nitrosomonaseuropaea        TSYFTITGKK- 210
pNS1                        -----------
Pseudomonasaeruginosa       RNFIR------ 180
```

Figure 5.5: Multiple alignments of selected bacterial isochorismatase proteins using ClustalW (*P. entomophila* YP_6065771; *P. putida* YP_001747104; *Ochrobactrum* YP_001372521; *P. fluorescens* YP_0028749491; *N.europea* NP_842301; and *P. aeruginosa* YP_0020814351). The residues forming the catalytic site are highlighted in green.

BLASTP analysis of pNS2 showed that it shared similarities with sulfatases. Sulfatases are widespread enzymes found in both prokaryotes and eukaryotes. They are involved in various metabolic processes such as sulphate starvation in bacteria and hormone biosynthesis in mammals (Berteau *et al.,* 2006). Sulfatases are a heterogenic group of enzymes that catalyze the hydrolytic cleavage of sulfate esters by liberating inorganic sulfate and the corresponding alcohol. Sulfatases act on a broad diversity of substrates. Despite their heterogeneity the primary and tertiary structures of sulfatases are highly conserved (Berteau *et al.,* 2006). They can either contain a serine or cysteine as the nucleophilic residue making up the active site of this enzyme (Benjdia *et al.,* 2007). The application of microorganisms for enzymatic desulfurization of coal as been explored. The use of specific hydrolases (*e.g*. arylsulfatase) will reduce the organic sulfur content. The sulfur is effectively isolated from the organic starting components and is retrievable as water-soluble sulfate. Apart from the issues of biocatalyst selectivity and operating efficiency, the use of an arylsulfatase for desulfurization of coal has been reported to be a relatively straightforward process. The use of microorganisms in the desulfurization process of coal therefore circumvents the problem of increased costs associated with traditional coal cleaning processes (free patentsonline.com/ EP0396832.html).

The sequence analysis revealed that pNS2 did not contain the entire sulfatase ORF but only the N-terminal region of the protein, however we were still able to identify the sulfatase motif [C-X-P-X-R] (Figure 5.6) within our sequence. Several attempts were made to try and obtain the complete sulfatase gene. A primer (SUL1) was designed from the sequence, the primer sequence was blasted against the NCBI database and showed significant hits with sulfatase genes. Using this primer, the following sequencing reactions were performed:- (i) primer (SUL1) was used to primer walk into the metagenome; however we did not recover the rest of the gene; (ii) plasmid DNA (maxi prep) of all plasmid clones pooled together was used as a template for the sequencing reaction; however we did not recover the rest of the gene; and (iii) fosmid DNA (maxi prep) all the fosmid clones pooled together was used as a template for sequencing reaction; however once again we were unsuccessful in our attempts to retrieve the entire sulfatase gene from the metagenome, indicating that only a part of this gene could be accessed from the metagenome.

```
Verminephrobacter      MNQHPAASLTRPRNAVVILLDSLNRHLLGAYGATEFETPQIDRFCASALRFDRHYAGSLP 60
Parvibaculum           MTNAEDGTGDQPRNAVVILLDSLNRHMIGAYGGREFATPNLDRFAARSTRFTRHFTGSLP 60
pNS2                   ----------MPRNLIVVLLDSLNRHLLGAYGGAEFETPHIDRFARRALRFDRHFTGSLP 50
Sinorhizobium          ------------MRAIFVLFDSLNRTAVGRYGANAVKTPNFDRFAERATTFDSHFVGSLP 48
                                    . :.:*:*****  :* **.  . **::***.  :  *  *:.****

Verminephrobacter      CMPARHDILCGALDFLWRPWGSIEVWEDAITYWLRNAGVVTQLISDHPHLFESGGENYHA 120
Parvibaculum           CMPARHDILCGALDFLWRPWGSVELWEDAITYELRKKGVVTQLISDHPHLFETGGENYHV 120
pNS2                   CMPARHDILCGALDFLWKPWGSIELWEEPITVPLRAAGVTTMLVTDHPHLFESGGENYHT 110
Sinorhizobium          CMPARRDLHTGRLNFMHRSWGPLEPFDNSFPELLGKCGVHSHLITDHLHYFEDGGSTYHT 108
                       *****:*:  * *:*: :.**.:* :::.:.  *   ** : *::** * ** **..**.

Verminephrobacter      DFQGWDYLRGHESDPWKTAQSECAIGAP--LHQVLPGPFP----HEYDTNRTWFKREEDF 174
Parvibaculum           DFTAWDYQRGHEGDPWKTRPDPSWAGAPNFMRKHMP----------YDDSRGYFRGEEDF 170
pNS2                   EFRAWEYLRGHENDPWKTRADPLVTAAS-----VL------------------------- 140
Sinorhizobium          RFRTWDFIRGQEDDPWKAMVQPPLERFKEMYSEKHYDFDDPWKRMQSAVNREFVRGEHEY 168
                        *  *:: **:*.****:  .

Verminephrobacter      PGPQTMASAARWIDENAGRHQRFFLMIDEFDPHEPFDTPQPWACRYRQAQGADEHQPLLV 234
Parvibaculum           PGPRTMGAAARWLNENAGHHGRFMLFVDEFDPHEPFDTPEPYASMY----DPDWEGAHLI 226
pNS2                   ---------------------EFCRYPSHWRPLE-------------------------- 153
Sinorhizobium          PGPRCFKSALEFLDLN-RAADDWFLMVECFDPHEPFAAPERFKEQY----ATGWEGGVLD 223
                                             :    . : * *

Verminephrobacter      WPPYAVDAIERGVLTAAQAQELRNNYGAKLSMIDHWLGRVLDAIERNRLAADTAVILCTD 294
Parvibaculum           WPPYVNGGIEKSVITERQARQIRASYGGKLTMIDKWFGKILDELDAKDLWKDTLVILCTD 286
pNS2                   ------------------------------------------------------------
Sinorhizobium          WPKY-----EKVVDSPEEIAEIRANYAALVTMCDEYFGRLLDYFDEHDLWKDTAIILSTD 278

Verminephrobacter      HGHYLGERDIFGKPGVPLYQPMAHIPLMIRWPGMAPG---RRDMLTTSVDIHATIADIFG 351
Parvibaculum           HGHYLGEKDIWGKPGVPVYEPLGHIPLMIAHPDVAPG---TCDALTTSVDLFATLAELFG 343
pNS2                   ------------------------------------------------------------
Sinorhizobium          HGFLLAEHDWWGKNRMPYYAEISQIPLIIYHPEHAGGGGTRRSALTQTIDLMPTFLDLFG 338

Verminephrobacter      VSAAHRTHGRSLLPAIADPGQQVREHLLAGVWGREVHYIDRSHKYVRAPAQANAPLSMWS 411
Parvibaculum           VEARQRTHGRSLLPLMRKEKPGIRDWLLTGVWGREVHYIDNRFKYARGPAGDNAPLTMMS 403
pNS2                   ------------------------------------------------------------
Sinorhizobium          IDVPQEVQGHSLLPLLKEDRS-MRDVAIFGVFGGPIGSTDGRYTYYLYPEDLYGPD---L 394

Verminephrobacter      NRWSTMPQHHVPGRRLLP-PDRRARIDFMPGSQVPVLRQPFVEGDLLPLWARNLRFSGNH 470
Parvibaculum           NRWSTMPTHFLTREQELPLPDDRAFLDRMPGSGVPVIHQQWDRDDPVPFWAR-TRFAGHH 462
pNS2                   ------------------------------------------------------------
Sinorhizobium          HEYTLMPMHMTSLFTPEELKTSALTAGFNFTKNMPVLRIDALRD------ARRIPNNDRV 448

Verminephrobacter      LWNLDADPREQTDLAGSALEAEYAHKLHAALRAIEAPDDQAIRLGLGV 518
Parvibaculum           LYDLTEDPAEERNLAGTSAEADLAERLRAALVEIEAPKSQLERLGLN- 509
pNS2                   ------------------------------------------------
Sinorhizobium          GWSVDLGTNLVRSSSGPNADAALPG-FGDRAPPVRGNPECAYR----- 490
```

Figure 5.6: Multiple sequence alignments using ClustalW of selected bacterial sulfatases (*Verminephrobacter* YP_997301; *Parvibaculum* YP_ 001414401; and *Sinorhizobium* YP_001327736) containing the conserved catalytic motif (C-X-P-X-R).

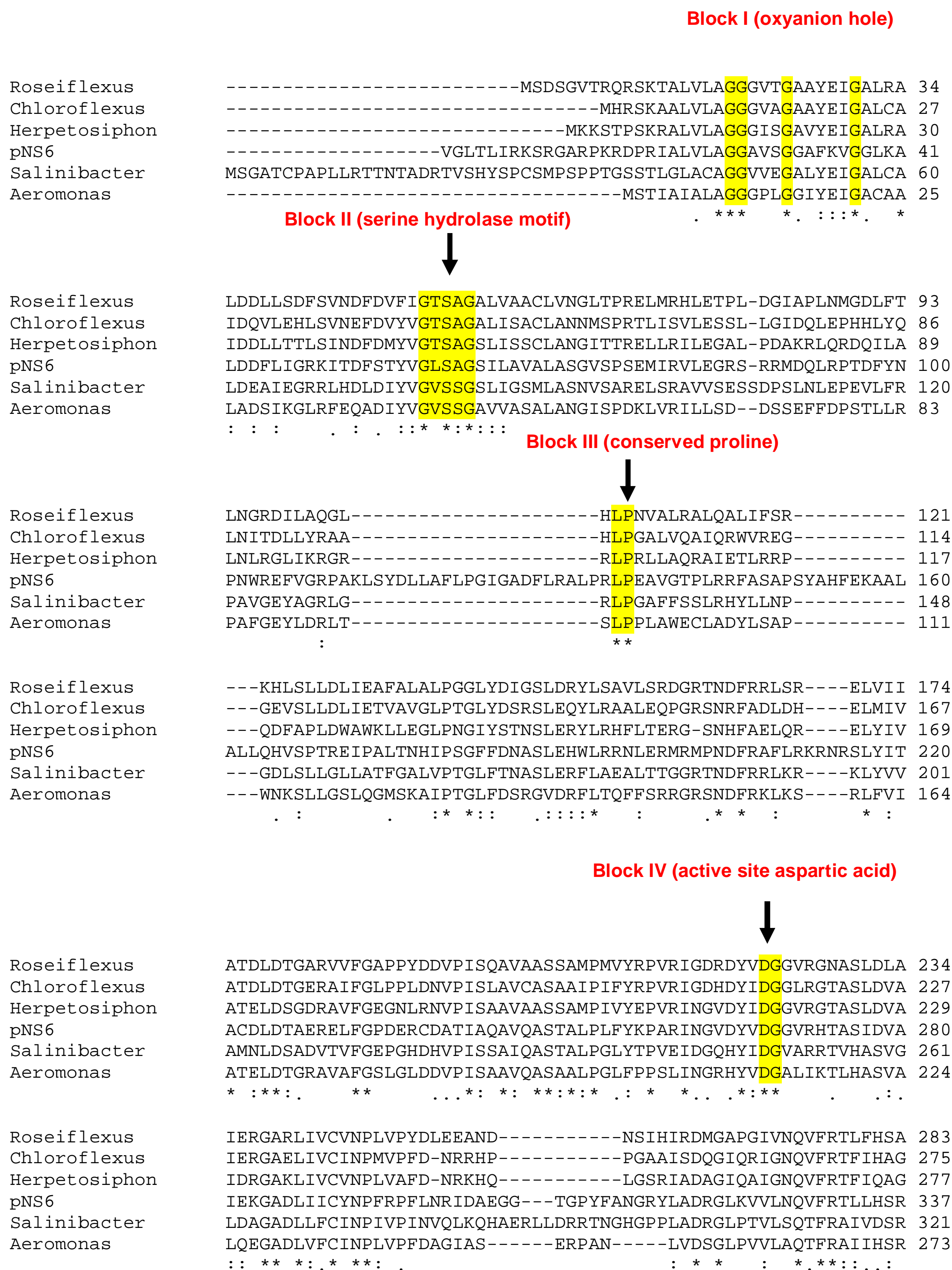
Of the three lipolytic enzymes obtained, pNS6 (phospholipase, patatin family protein-[PLP]) held the greatest interest for us. Patatin is the major protein constituent of potato tubers and displays broad esterase activity (Hirschberg *et al.,* 2001). Patatin has been reported to display lipid acyl hydrolase activity but the physiological role of this enzyme remains unclear (Barta and Bartova, 2008). Patatin B2 and phosholipase A2 share conserved domains, both these proteins contain the classical lipase/esterase motifs and active sites (G-X-S-X-G) (X-denotes any amino acid) and the conserved aspartic acid and histidine. Patatin forms part of the defense mechanisms in potato crops (Barta and Bartova, 2008). *Phytophora infestans,* the causal agent of the late blight epidemic remains agriculture's most destructive disease as new mutations and migrations confound control measures. Studies conducted by Sharma *et al.* (2004) have shown that patatin has an antifungal effect on spore germination of *P. infestans* and can be applied in the biological control of this pathogen due to its antimicrobial activity. Patatins have been shown to display insecticidal activity against corn rootworm, an economically destructive insect pest in corn (Rydel *et al.,* 2003). In South Africa corn is one of our major crops, therefore this enzyme could be beneficial to our agricultural industry. This is being further investigated.

Patatins can be used for the modification of food lipid ingredients and especially phospholipids. The only commercial source of phospholipase A2 activity is from porcine pancreas and alternative biocatalysts with unique or different selectivities would broaden the opportunities for phospholipid ingredient modification for specific applications (Anderson *et al.,* 2002).

BLASTP analysis of pNS6 revealed moderate similarity (<55%) between pNS6 and patatin proteins from other bacteria including phospholipase, patatin family protein (55%) from *Plesiocystis pacifica* (ZP_01911272); patatin (34%) from *Roseiflexus sp* (YP_001276494); patatin (35%) from *Herpetosiphon aurantiacus* (YP_001543920). According to the blast analysis (55% similarity to the closest hit), we have obtained a new patatin protein. Further sequence analysis revealed a putative purine rich ribosomal binding site located 5bp upstream of the ATG start codon.

Multiple sequence alignment comparisons of the pNS6 to other phospholipase patatin proteins from various bacteria revealed the presence of five conserved domains (blocks I-V) (Figure 5.7). The classical lipase/esterase conserved motif and active sites residues [G-X-S-

X-G] (X-denotes any amino acid) and the conserved aspartic acid and histidine were identified in pNS6.



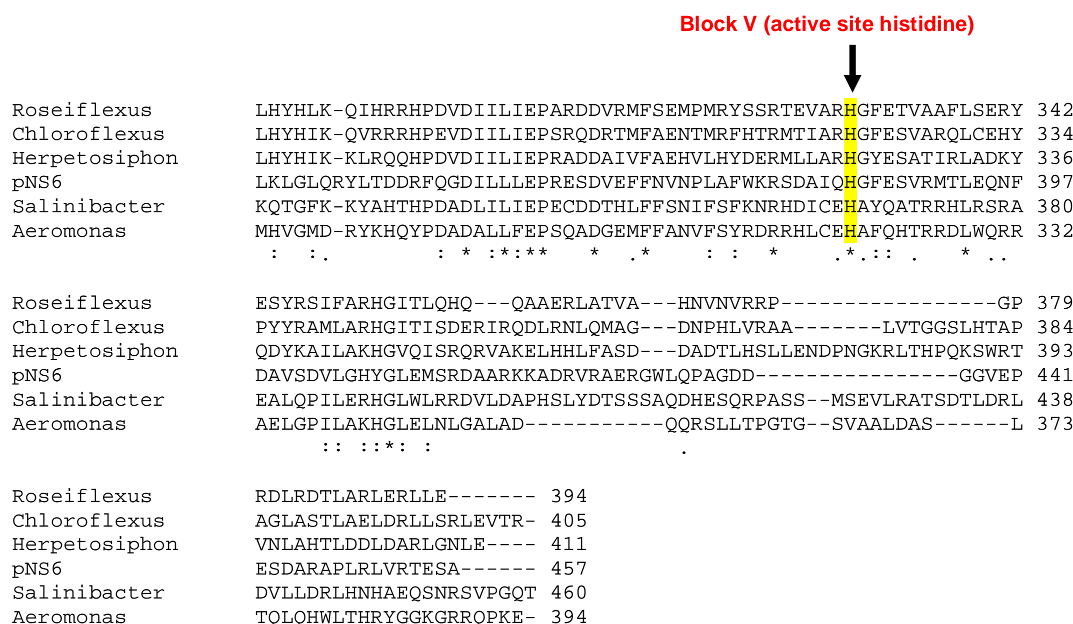


Figure 5.7: ClustalW multiple alignments of selected bacterial phospholipase, patatin family proteins (*Roseiflexus* YP_001276494; *Chloroflexus* YP_001636863; *Herpetosiphon* YP_001543920; *Salinibacter* YP_4467821; and *Aeromonas* YP_ 001141850) containing the conserved esterase/lipase domains.

In addition, we aligned pNS6 with other bacterial lipolytic enzymes however there was no significant alignment. We then decided to construct a phylogenetic tree in order to observe where pNS6 clusters in respect to other lipolytic enzymes; patatin proteins were included in the analysis. Bacterial lipases representing the 8 different families of lipolytic enzymes were used for the phylogenetic analysis. According to the phylogenetic tree (Figure 5.8), pNS6 clustered with the phospholipase, patatin family of proteins rather than with any of the bacterial species containing lipolytic enzymes. Our findings were supported by literature which cites that bacterial PLPs do not show any homology to known groups of bacterial lipases as evaluated by Blast (Altschul *et al.,* 1997) or PFAM domain search (Banerji and Flieger, 2004), indicating that the amino acid sequences of PLPs and established groups of lipolytic enzymes are not closely related. Secondly, although bacterial PLPs contained the G-X-S-X-G, lipase motif which is common even in distinct lipase families, bacterial PLPs do not exhibit the same features as found for other bacterial lipases (Arpigny and Jaeger, 1999). Since the prokaryotic PLPs appear to be more related to their eukaryotic counterparts than to any other group of bacterial lipases, it is proposed that they comprise a new group of bacterial lipolytic enzymes (Banerji and Flieger, 2004).

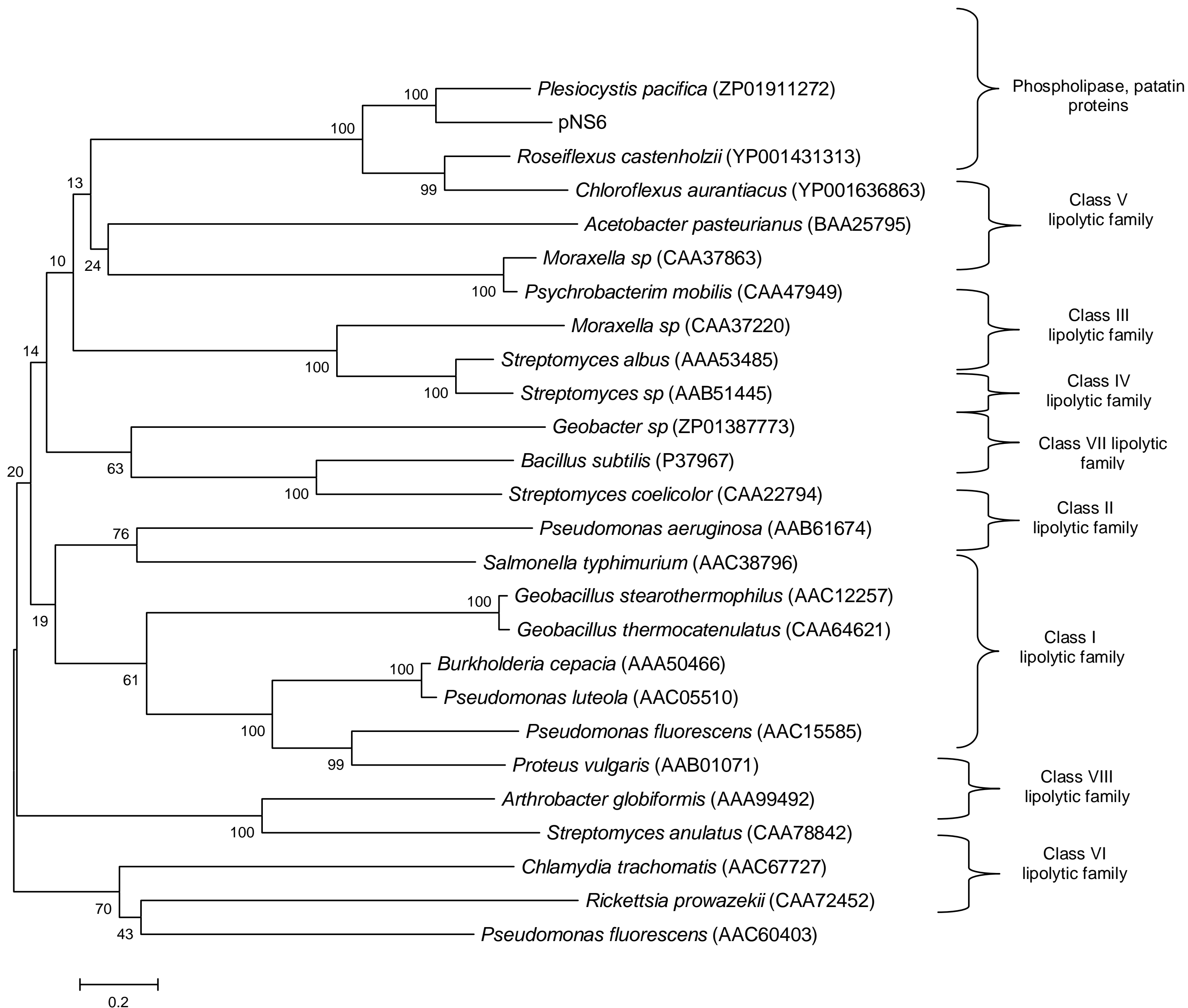


Figure 5.8: Phylogenetic tree of phospholipase, patatin family proteins and different classes of lipolytic enzyme families. The phylogenetic tree was constructed by the neighbour-joining method with MEGA version 4.1software. The accession numbers of the aligned sequences are shown in parenthesis. The numbers associated with the branches refer to the bootstrap values (confidence limits). The scale represents a 20% sequence divergence.

5.7 Constructs for expression in *E. coli*

The complete ORF of the patatin was PCR amplified from plasmid DNA of clone pNS6 yielding a single product of the expected size (Figure 5.9). The primers were designed to incorporate a *Nde*I restriction site at the 5' end and a *Hin*dIII restriction site at the 3' end for directional cloning (Table 5.3) into the pET-28b(+) vector.

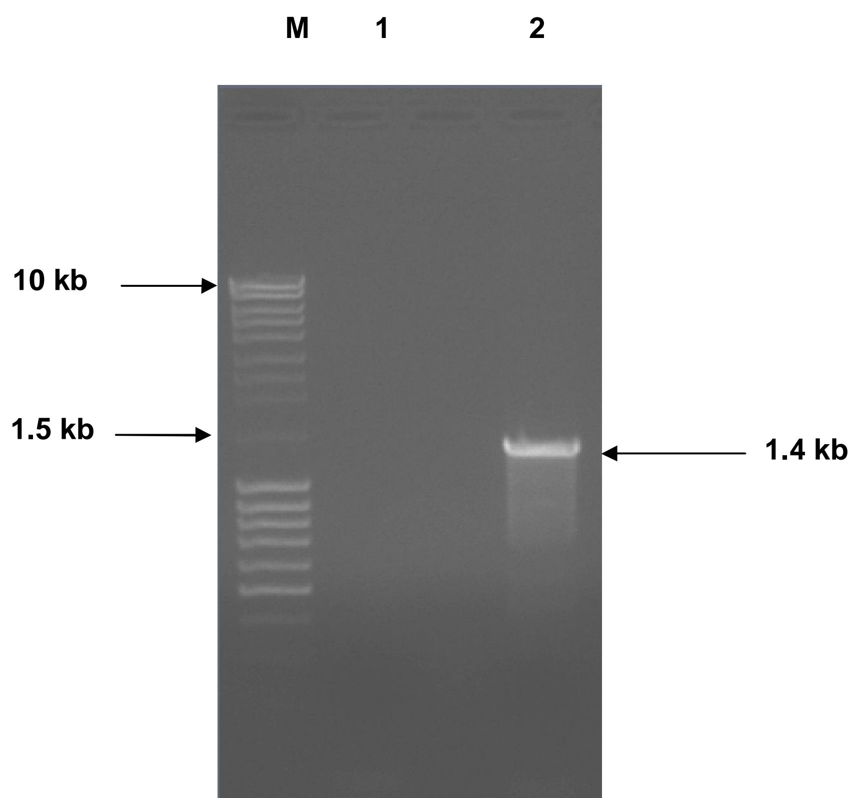
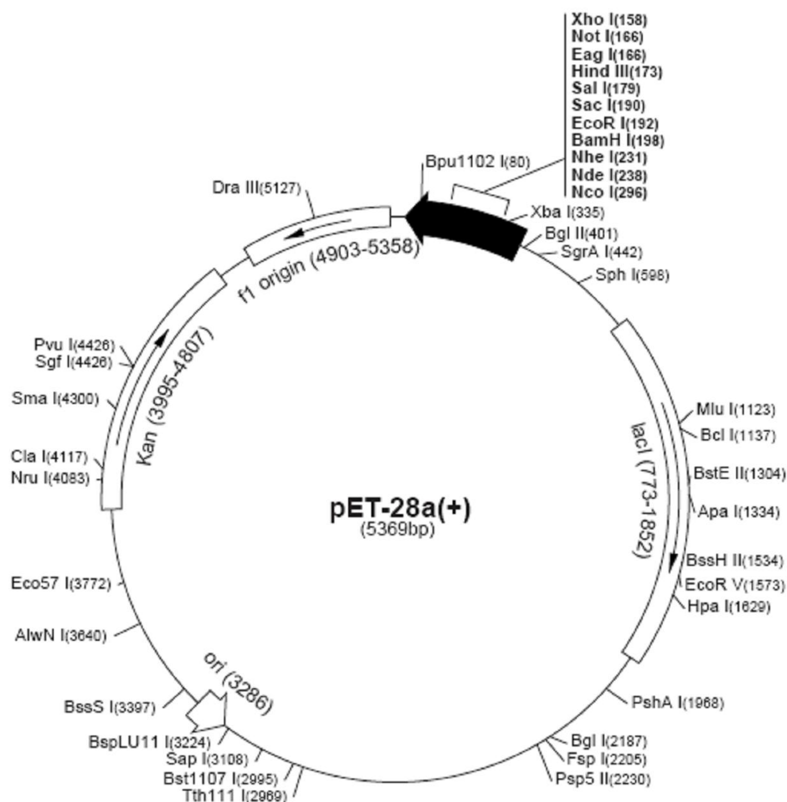


Figure 5.9: PCR amplification of the patatin *ORF*. M: MassRuler DNA Ladder (SM#0403- Fermentas) and lane 1: negative control (sterile water); and lane 2: plasmid DNA of clone pNS6.

The PCR amplified products were ligated into pGEM®-T Easy vector. Inserts of positive clones were double digested with *Nde*I and *Hin*dIII. The pET-28b(+) vector was also digested with the same restriction enzyme combination to yield compatible cohesive ends and the patatin was cloned into the expression vector. pET-28b(+) would result in a poly(6) histidine tag and a thrombin cleavage site fused to the N-terminus (Figure 5.10) of an introduced protein.



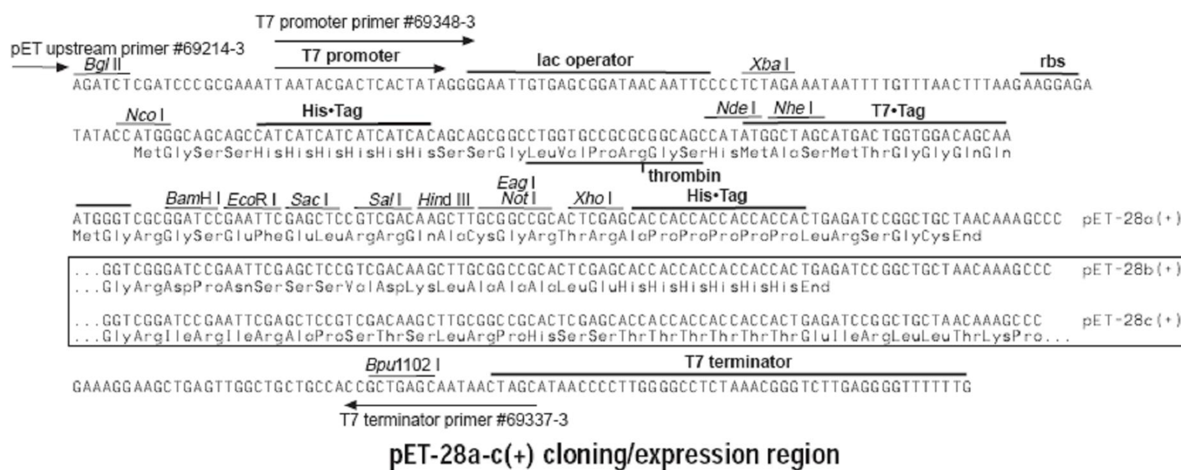


Figure 5.10: Vector map of pET-28b(+) indicating the kanamycin resistance gene, ColE1 origin of plasmid replication, *lacI* coding sequence and the multiple cloning site under the T7 promoter. Sequence of the pET-28b(+) cloning region showing the ribosome binding site and configuration for the N-terminal His-Tag and thrombin cleavage site (taken from Novagen vector manual).

Restriction digests of randomly selected clones confirmed the presence of the insert in the pET-28b(+) vector (Figure 5.11).

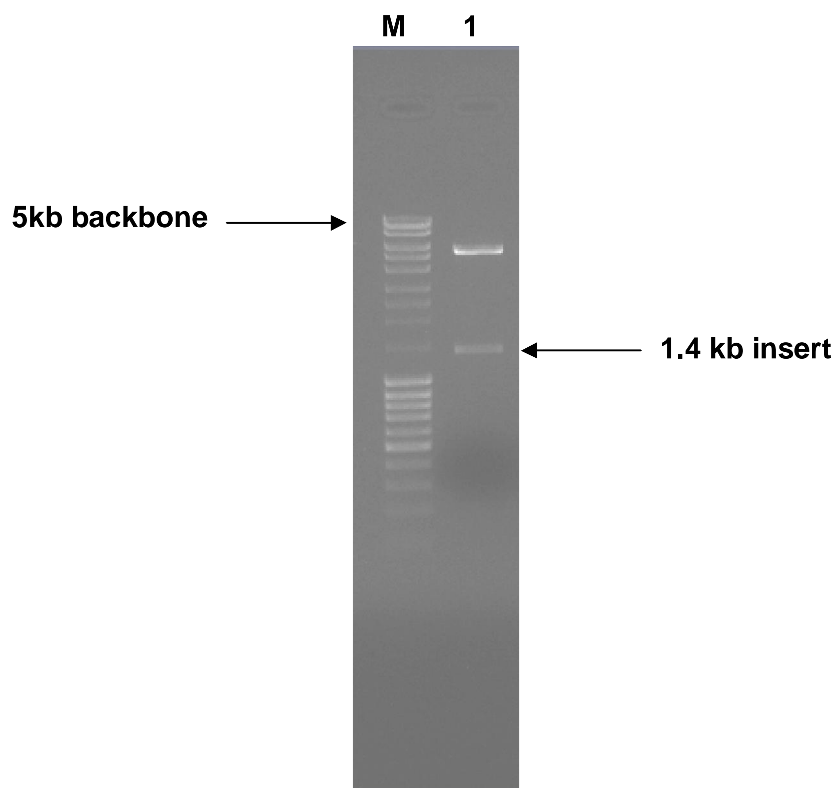


Figure 5.11: Restriction analysis of selected pET28b(+) clone. M- MassRuler 1kb ladder (SM#0403-Fermentas): and lane 2: positive clone.

5.7.1 Expression and purification of the patatin

The patatin was over-expressed in *E. coli* BL21(DE3) pLysS at 30°C by induction through the addition of 0.5 mM IPTG to exponentially growing cells, and subsequent growth for another 4 h. SDS-PAGE analysis showed a protein band of the expected size of approximately 60 kDa in the soluble fraction of *E. coli* (Figure 5.12) for the recombinant patatin.

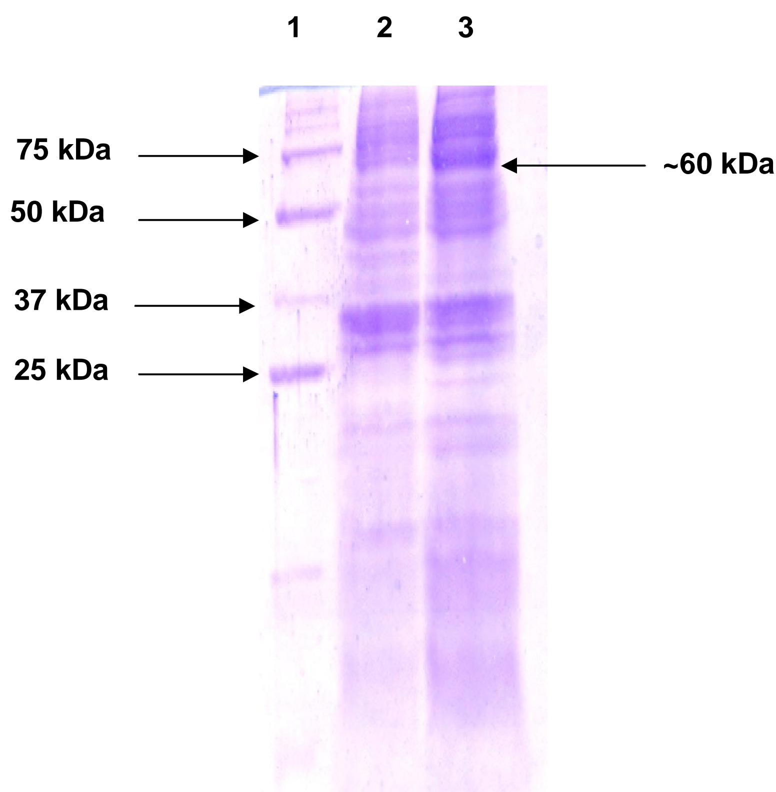


Figure 5.12: Expression of the patatin in *E. coli.* Lane 1: Molecular marker (BioRad), lane 2: uninduced control, lane 3: expressed patatin after 4 h of induction at 30°C with 0.5 mM IPTG.

The recombinant protein was purified by taking advantage of the N-terminal poly (6)-histidine tag using Ni-affinity (Ni-NTA) chromatography. The patatin eluted as single activity peak (Figure 5.13) from the His-Trap FF (Amersham Biosciences) column through the use of linear imidazole concentration gradients. Fractions containing the patatin, as visualized by SDS-PAGE were pooled and loaded onto a second His-Trap FF column for further purification. A shallower gradient was used for the second purification in order to try and separate proteins with similar binding strengths. However, after the second purification step a purified protein was not obtained, and there were still several other bands present on the gel. The partially purified patatin was estimated on SDS-PAGE analysis to be approximately 60 kDa (Figure 5.14). We then decided to confirm the presence of the partially purified patatin by western blot analysis. Western blotting was performed using the SuperSignal®West HisProbe™ kit (Pirece, USA) according to the manufacture's instructions. The presence of the partially purified protein was confirmed and estimated to be

approximately 60 kDA (Figure 5.15) according to the size of the prestained molecular marker (Fermentas) that was run. Before further characterization was performed, the imidazole present in the sample was removed by dialysis, to prevent any inhibitory effect that it may have on enzyme activity.

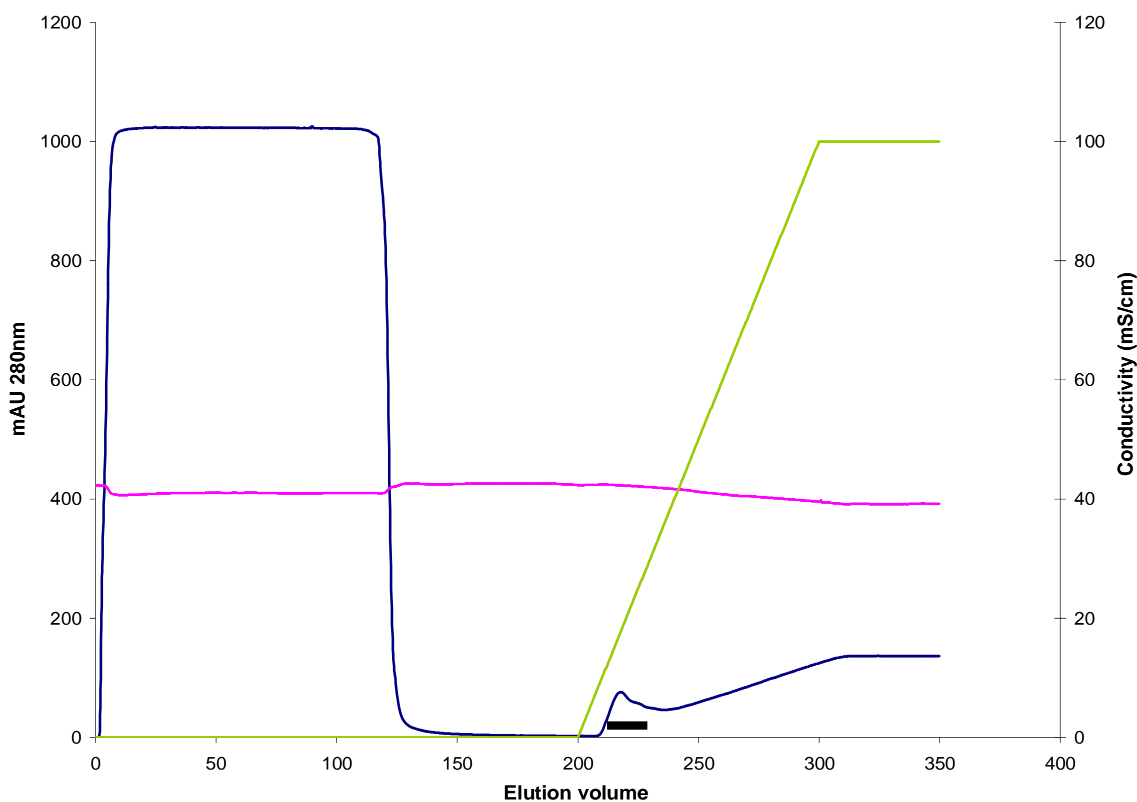


Figure 5.13: Purification of the expressed patatin through Ni-affinity chromatography. The green line indicates the salt gradient. The blue line represents the elution profile of the purified protein and the black horizontal line denotes the volume which was pooled.

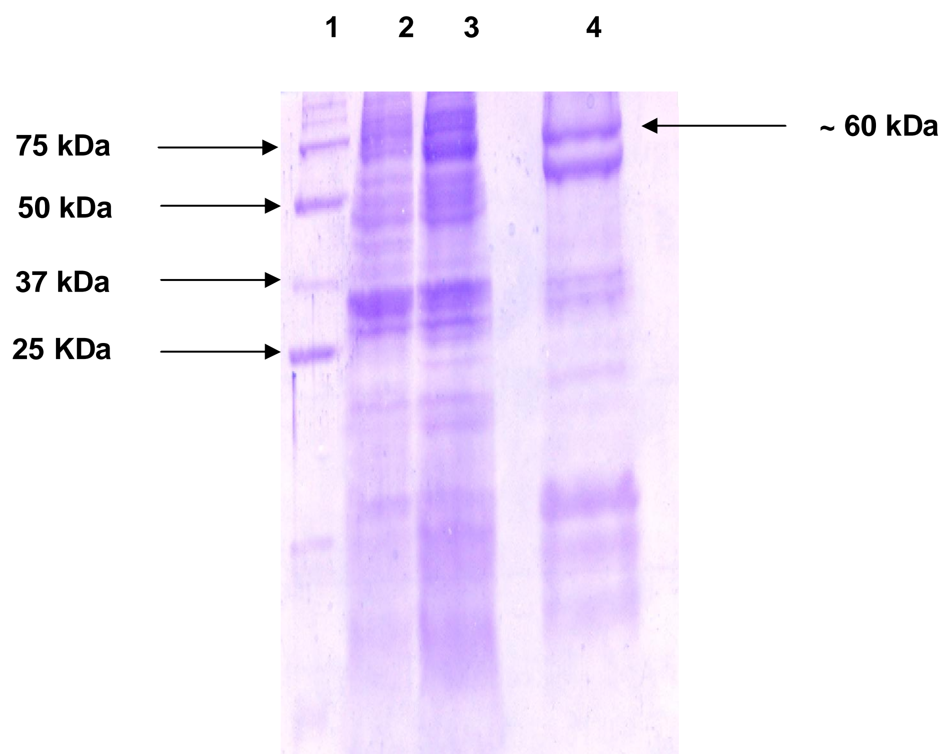


Figure 5.14: Partially purified patatin. Lane 1: Molecular marker (BioRad), lane 2: uninduced control, lane 3: expressed patatin, lane 4: partially purified protein after 4 h of induction at 30°C with 0.5 mM IPTG.

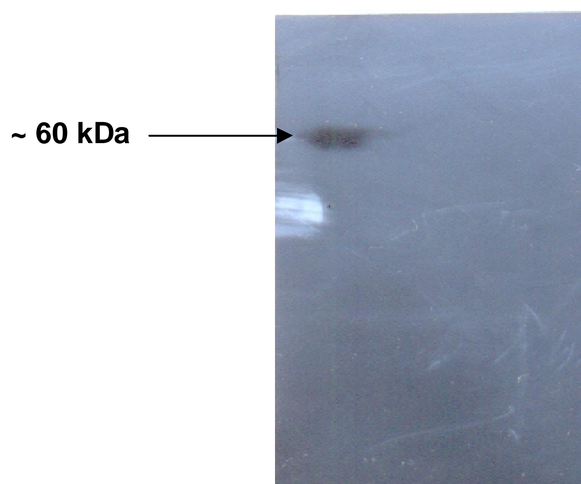


Figure 5.15: Western blot analysis of the partially purified patatin after 4 h of induction at 30°C with 0.5 mM IPTG. The size of the band was determined from the position of prestained marker proteins before Western blotting.

5.7.2 Biochemical analysis of the expressed patatin

Various *para*-nitrophenol esters (C_4-C_{16}) were used for the determination of activity of the patatin. Enzymatic activities were measured spectrophotometrically at 405 nm. The assay contained 600 µl of 100 mM Tris/HCl (pH 8) and 0.5 mM substrate. Substrate stocks of *p*-NP-butyrate, caproate, caprylate, caprate, and palmitate were prepared in isopropanol. Due to the presence of contaminating proteins after purification, we included a negative control in the assay, protein extracted from the pET vector+host cell only served as the control. This was to check that the displayed activity was due to the patatin only and no other contaminating proteins were contributing to the activity. No activity was observed with the negative control. The substrate specificity analysis revealed that the patatin could efficiently hydrolyze short and medium chain (C_4 to C_{10}) *p*-NP esters, experimentally suggesting that this enzyme is an esterase. Highest activity was observed towards *p*-NP caproate [C_6] (Figure 5.16).

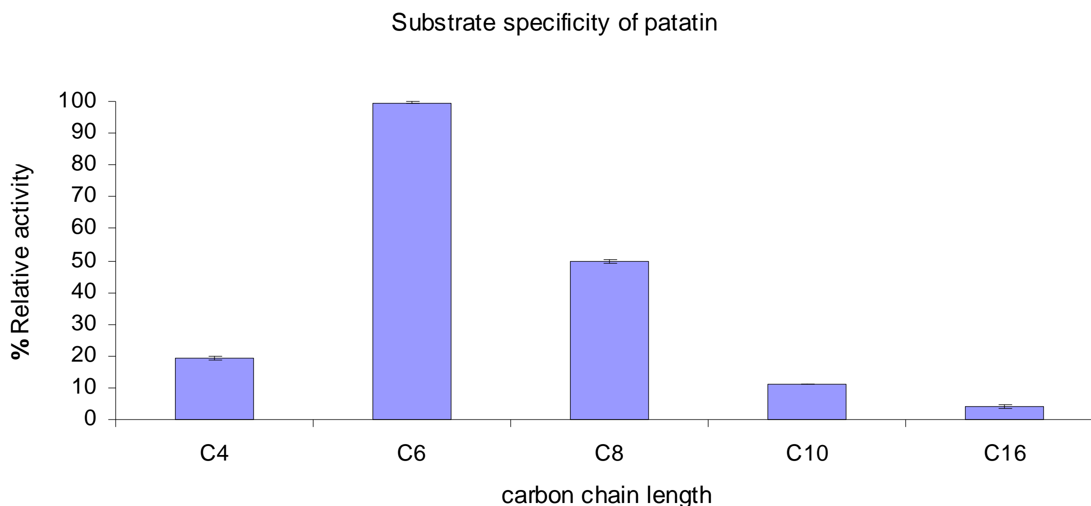


Figure 5.16: Substrate specificity for *p*-nitrophenyl esters with varying carbon lengths. Error bars indicate standard deviations after performing the experiment in triplicate.

5.7.2.1 Effects of temperature and pH on enzyme activity

The patatin displayed activity over a temperature range of 20-40°C (Figure 5.17) and showed highest activity at 30°C. This was expected as this enzyme was isolated from a mesophilic environment. Optimum activity was observed at pH 8 (Figure 5.18), which is

consistent with esterases, since the majority of lipolytic enzymes are functional at alkaline pH. Similar results were obtained by Tirawongsaroj *et al.* (2008) who reported on the expression and characterization of a novel bacterial phospholipase patatin isolated from a hotspring in Thailand.

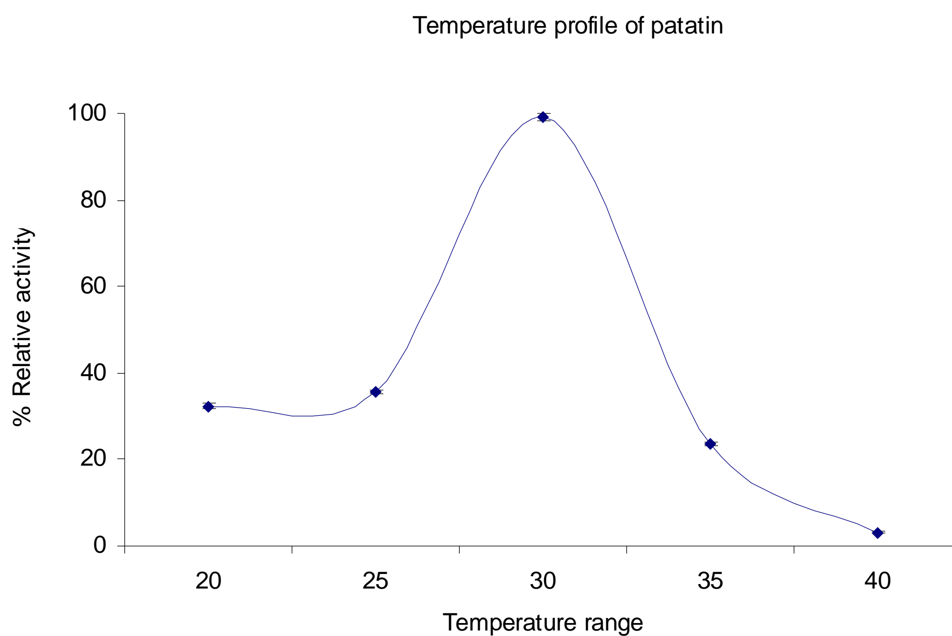


Figure 5.17: Temperature profile of the patatin. Error bars indicate standard deviations after performing the experiment in triplicate.

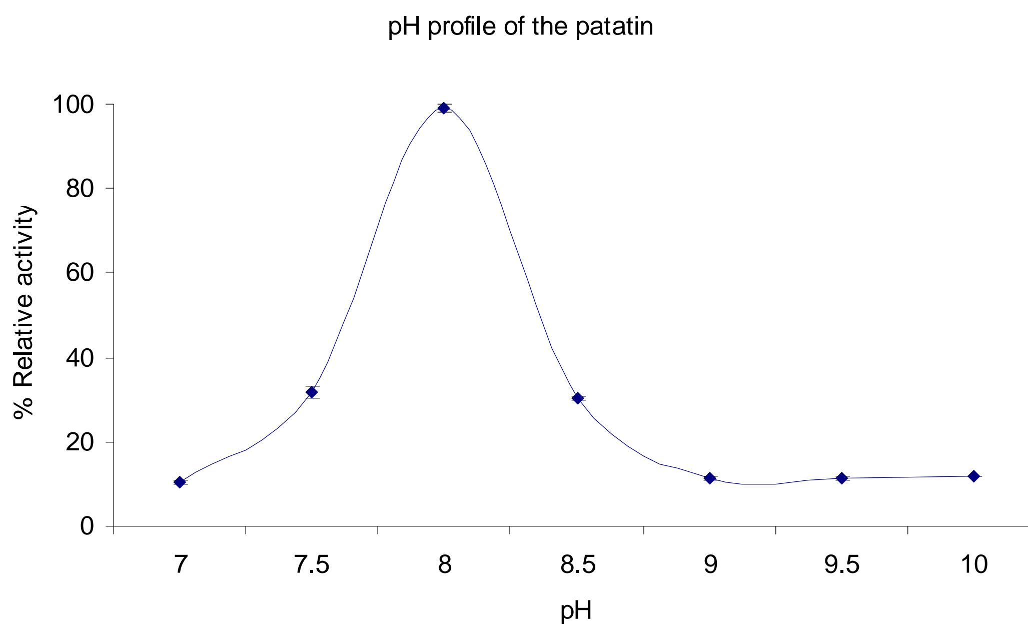


Figure 5.18: pH profile of the patatin. Error bars indicate standard deviations after performing the experiment in triplicate.

5.8 Large-insert library construction

A library of ~30 000 clones was obtained. According to the restriction analysis by PFGE, the inserts sizes of the clones ranged from 25-35 kb (Figure 5.19). The restriction analysis also highlights the rich diversity of the clones.

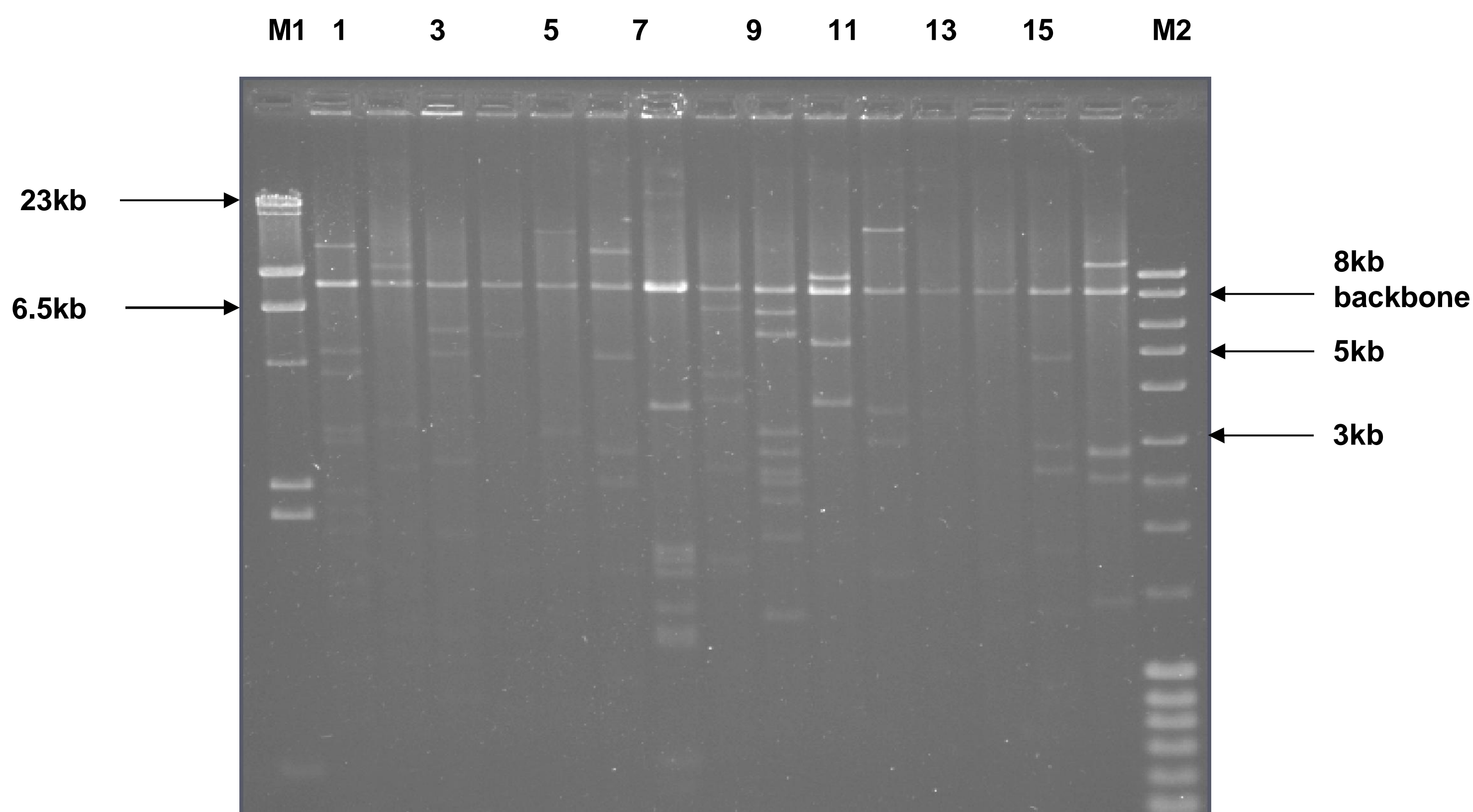


Figure 5.19: Restriction analysis of selected fosmid clones. M1- MarkerII (Lambda DNA digested with *Hin*dIII), lanes 1-16: randomly selected fosmid clones, M2-MassRuler 1kb ladder (SM #0403-Fermentas).

5.8.1 Fosmid Library Screening

The fosmid library was screened for clones expressing antibacterial activity and antibiotic resistance genes. However, no positive clones were obtained for either of the plate screens. The results that we obtained does not necessary indicate that the clones are completely devoid of such genes, it may be possible that they will be active against other reference cultures and antibiotics that have not been tested in this study and this could be a future endeavour for upcoming studies.

5.9 Conclusion

The total number of prokaryotic cells on earth has been estimated at 4-6X10^{30} comprising between 10^6 and 10^8 separate genospecies (distinct taxonomic groups based on gene sequence analysis). This diversity presents an enormous (and largely untapped) genetic

and biological pool that can be exploited for the recovery of novel genes, entire metabolic pathways and their products (Cowan *et al.,* 2005). Metagenomic libraries have been exploited for identification of novel genes and biocatalysts (Rondon *et al.,* 2000). We have identified three different esterase type enzymes (isochorismatase, sulfatase and phospholipase, patatin) from a metagenome library constructed with biofilm obtained from a South African gold mine. Of the three enzymes we selected the pNS6 (phospholipase, patatin family protein) for further characterization. For most of the lipolytic enzymes, the catalytic triad consists of serine in a conserved pentapeptide G-X-S-X-G, aspartate, and a highly conserved histidine. We were able to identify the catalytic residues in pNS6, however phylogenetic analysis revealed that the phospholipase, patatin proteins donot cluster in any of the known 8 classes representative of lipolytic enzymes, and therefore PLP's constitute a new family of lipolytic enzymes.

Patatins can be used in industry for the modification of food lipid ingredients especially, phospholipids. The only commercial source of phospholipase A2 activity is from porcine pancreas and alternative biocatalysts with unique or different selectivities would broaden the opportunities for phospholipid ingredient modification for specific applications (Anderson *et al.,* 2002). Patatins also have application in agriculture, *Phytophora infestans,* the causal agent of the late blight epidemic remains agriculture's most destructive disease, as new mutations and migrations confound control measures. Studies conducted by Sharma *et al.* (2004) have shown that patatin has an antifungal effect on spore germination of *P. infestans* and can be applied in the biological control of this pathogen due to its antimicrobial activity. Patatins have been shown to display insecticidal activity against corn rootworm, an economically destructive insect pest in corn (Rydel *et al.,* 2003). In South Africa corn is one of our major crops, therefore this enzyme will be beneficial to our agricultural industry. The metagenomes of natural microbial communities contain an immense pool of genes that potentially harbour novel biocatalyts as illustrated in this study.

Chapter Six

Summary

The construction and screening of gene libraries prepared from DNA directly isolated from environmental samples is a recent and powerful tool for the discovery of new enzymes of biotechnological interest (Gabor *et al.,* 2004). Standard methods based on the screening of isolated microorganisms are inherently limited to the tiny fraction of cultivable microbial species (<1%); environmental gene banks in principle provide access to the entire sequence space present in nature (Handelsman *et al.,* 1998). Environmental libraries allow the screening of functional classes of genes from thousands of organisms and research in this area will provide an essential backdrop for understanding evolution and biochemical pathways (Rajendran *et al.,* 2008). The metagenomics approach has been shown to be an efficient method for obtaining novel biocatalysts and useful genes from uncultured microorganisms from diverse environments.

Before proceeding to the metagenome analysis, we constructed genomic libraries from a South African deep mine isolate *Geobacillus thermoleovorans* GE-7. The library was screened for lipolytic activity on LB tributyrin (TLB). Active clones were sequenced using 454 technology, and the sequencing results revealed the presence of the lipA and GDSL lipases, of which the latter has not yet been characterized in this organism. This family displays the characteristic G-D-S-L motif instead of the conventional G-X-S-X-G (X-denotes any amino acid) motif. GDSL lipases are hydrolytic enzymes with multifunctional properties such as broad substrate specificity and regiospecificity. They have potential for use in the hydrolysis and synthesis of ester compounds that are of interest in pharmaceutical, food, biochemical and the biological sector (Akoh *et al.,* 2004; Lämmle *et al.,* 2007). In addition, genes associated with fatty acid degradation, different glycolytic activities, lipolytic activity, spore germination, proper protein folding, antibiotic resistance and the cell wall were also identified in the active clones. Some of the genes identified may also aid in understanding how this organism had adapted to the environment from which it was isolated from.

Biofilm collected from the Beatrix gold mine was selected for the metagenomic studies. We performed a diversity assessment of the biofilm by cloning and sequencing of the 16S (bacterial and archaeal) and 18S eukaryotic ribosomal DNA. Phylogenetic assessment of

the bacterial clones indicated that clonal sequences were affiliated with at least 5 phyla of the domain bacteria. Sequences allocated to the phyla that are present in the bacterial library, have been reported to be isolated from various environments including marine environments *e.g.* the Sargasso Sea and sea floor basalts, sub-surface ground water, limestone caves, tar pits, alkaliphilic hot springs, mine drainage sites and biofilms (Cho and Giovanni, 2003; Kim and Crowley, 2007; Ikner *et al.,* 2007; and Stepanauskas and Sieracki, 2007). According to the rarefaction analysis, of the 37 clones analyzed 29 different OTUs were observed. However, the rarefaction results indicate that only a portion of the richness in the bacterial community (at the ≥97% sequence identity level) was surveyed with the 16S clones sequenced as the curve did not reach an asymptote. Sequence data obtained from the archaeal clonal library showed sequence identities with that of uncultured archaea present in the database, in particular a single uncultured archaeal library from a marine sample. This could be attributed to the fact that there is limited sequence data on archaea present in the databases because only 4 taxonomic groups of the domain are known thus far (Baker *et al.,* 2003).

The metagenome was screened by the sequenced-based approach for cytochrome P450 monooxygenases, in particular the CYP153 family (terminal hydroxylases and long chain alkane degraders). Cloning and sequencing of the CYP153 PCR products, revealed the presence of this family of enzymes in the metagenome P450's are involved in a plethora of metabolic processes, both anabolic and catabolic, and collectively interact with an enormous variety of substrates (De Mot and Parret, 2002). CYP's in bacteria are involved in the biosynthesis of secondary metabolites such as antibiotics and in the utilization of hydrophobic low molecular weight compounds such as alkanes and aromatics (Kubota *et al.,* 2005). The biocatalytic production of the anticancer drug perillyl alcohol from limonene has been reported to involve a CYP153 cytochrome from *Mycobacterium sp.* thereby indicating the potential application of this enzyme in the clinical setting (Urlacher and Eiben, 2006).

For the function-based approach, both small and large-insert metagenomic libraries were constructed. The libraries were screened for lipolytic, amylase, protease as well as antibacterial and antibiotic resistant genes. Only lipolytic active clones were obtained. Sequence analysis of selected TLB active clones revealed the presence of three different lipolytic enzymes (isochorismatase, sulfatase and phosholipase, patatin family protein).

Only the phospholipase, patatin protein was further characterized. Sequence analysis revealed the presence of the classical esterase motif (G-X-S-X-G) and conserved aspartatic acid and histidine residues in the patatin. According to phylogenetic analysis phospholipase, patatin proteins constitute a new family of lipolytic enzymes, since they do not form part of any of the eight classes representative of lipolytic enzymes. The patatin was heterologously expressed in *E. coli.* Biochemical analysis of the partially purified protein showed that the enzyme had a preference for shorter carbon chained substrates, indicating that patatin displays esterase rather than lipase activity and functioned optimally at 30°C and pH 8 which was expected considering that the enzyme was isolated from a mesophilic environment.

Keywords Metagenome, diversity assessment, sequenced-based screening, cytochrome P450 alkane hydroxylases, library construction, functional-based screening, lipolytic activity, phospholipase patatin, heterologous expression, biochemical characterization.

Opsomming

Die konstruksie en sifting ('screening') van genoombiblioteke wat voorberei is vanaf DNS, geïsoleer direk uit die omgewing is 'n eietydse en kragtige tegniek vir die ontdekking van nuwe ensieme met biotegnologiese waarde (Gabor *et al*., 2004). Standaard metodes gebaseer op die sifting van geïsoleerde mikro-organismes is inherent beperk tot die klein breukdeel van kweekbare mikrobiese spesies (<1%). Omgewings genoombiblioteke, daarenteen, lewer meer toegang tot die totale DNS basispaar opeenvolging wat daardie omgewing bied (Handelsman *et al*. 1998). Omgewings genoombiblioteke bied die navorser die geleentheid om te sif vir funksionele klasse van ensieme vanaf duisende organismes en navorsing in hierdie studieveld sal bydra tot ons huidige insig en begrip van die evolusie van metaboliese we (Short, 1997; Rajendran *et al*., 2008). Dit is bewys dat die metagenoom aanslag 'n effektiewe tegniek is vir die isolering van nuwe biokataliste en bruikbare gene van onkweekbare mikro-organismes afkomstig vanaf diverse omgewings.

Voordat daar met die metagenoomanalise begin is, het ons 'n genoombiblioteek berei uit 'n bakteriese isolaat afkomstig van 'n diep myn: *Geobacillus thermoleovorans* GE-7. Die biblioteek is gesif vir lipolitiese aktiwiteit m.b.v. LB tributirien plate (TLB). Klone wat aktiwiteit getoon het, se DNS basispaar opeenvolging is bepaal d.m.v. 454 tegnologie en die data het getoon dat die lipA en GDSL lipases teenwoordig was. Die laasgenoemde

lipase is nog nie gekaraktiriseer in hierdie organisme nie. Hierdie lipase familie vertoon die kenmerkende G-D-S-L aminosuur motief in plaas van die konvensionele G-X-S-X-G motief (X = dui enige amino suur aan). GDSL lipases is hidrolitiese ensieme met multi-funksionele eienskappe soos bv. ‘n wye substraatspesifisiteit en regiospesifisiteit. Hierdie ensieme het ook die potensiaal om die hidrolise en sintese van ester verbindings te bewerkstellig wat van belang is in die farmaseutiese, voedsel, biochemiese en biologiese sektore (Akoh *et al*., 2004; Lämmle *et al*., 2007). Aktiewe klone van hierdie genoom biblioteek het ook gene opgelewer wat ge-assosieer word met vetsuur afbraak, verskillende glikolitiese aktiwititeite, lipolitiese aktiwiteite, spoor ontkieming, protein vouïng, antibiotika weerstandbiedendheid en die selwand. Sommige van hierdie gene se funksies mag lei tot beter insig oor hoe hierdie organisme aangepas het in die omgewing waarvan dit geïsoleer was.

Biofilm vanaf die Beatrix goud myn is gekies vir metagenomiese studies. ‘n Diversiteits studie van hierdie biofilm is gedoen m.b.v. die klonering en basispaar volgorde bepaling van 16S (baketerie en archaea) en 18S eukariotiese ribosomale DNS-gene. Filogenetiese analise van die bakteriële klone het getoon dat hulle DNS basispaar opeenvolging gegroepeer het met ten minste 5 fila vanaf die bakeri le koninkryk. Die 5 fila is afkomstig vanaf baie diverse omgewings soos bv. die Sargasso see en see basaltvloere, ondergrondse water, kalksteen grotte, teerputte, ‘n alkaliese warmwater bron, dreinerings areas van myne en biofilms (Cho en Giovanni, 2003; Kim en Crowley, 2007; Ikner *et al*., 2007 en Stepanauskas en Sieracki, 2007). Analise (‘Rarefaction analysis’) van 37 klone het getoon dat daar 29 verskillende OTE (‘operasionele taksonomiese eenhede’) was. Dit moet gemeld word dat die analise wel aangetoon het dat slegs ‘n klein gedeelte van die bakteri le gemeenskap se diversiteit (gebaseer op ‘n 97 % DNS basispaar opeenvolgingidentiteit) bestudeer is, m.b.v. die 16S klone, aangesien die grafiek nooit die asimptoot bereik het nie. DNS basispaar opeenvolgingdata afkomstig vanaf die archaea genoombiblioteek, het identiteite getoon met dié van ‘n onkweekbare archaea in die databasis en in besonder die van ‘n enkele onkweekbare archaea genoombiblioteek vanaf ‘n seewater monster. Hierdie kan toegeskryf word aan die feit dat daar tot op hede slegs beperkte archaea DNS basispaar opeenvolgingdata beskikbaar is en dat daar nog net 4 taksonomiese groepe uit hierdie koninkryk aan ons bekend is (Baker et al., 2003).

Die metagenoom was gesif m.b.v. ‘n DNS basispaar opeenvolging-gebaseerde benadering vir sitokroom P450 mono-oksigenases (CYP450s) met spesiale klem op die CYP153 familie, wat verantwoordelik is vir die terminale hidroksilering van lang-ketting alkane. Klonering en opeenvolgingbepaling van hierdie CYP153 PKR (‘PCR’) produk het die teenwoordigheid van hierdie ensiem familie in die metagenoom getoon en dat hulle betrokke is by ‘n wye verskeidenheid metaboliese prosesse, diverse substraatvoorkeure het en beide kataboliese en anaboliese reaksies kataliseer (De Mot en Parret, 2002). CYP450s in bakterie is verantwoordelik vir die biosintese van sekondêre metaboliete soos bv. antibiotika en die verbruik van lae molekulere massa, hidrofobiese verbindings soos alkane en aromatiese verbindings (Kubota *et al*., 2005). Daar is bewyse dat die CYP153 vanaf *Mycobacterium* sp. betrokke is by die biokatalise van die teenkanker middel: periliel alkohol vanaf limonien wat dui op die potensi le toepassing van hierdie ensiem in ‘n kliniese milieu (Urlacher en Eiben, 2006).

Funksie-gebaseerde sifting was m.b.v. genoombiblioteke wat klein en groot invoegings bevat het, uitgevoer. Die biblioteke was gesif vir lipolitiese, amilase, protease en anti-bakteriese gene. Slegs lipolities-aktiewe klone kon geïdentifiseer word. Klone wat aktiwiteit op TLB-plate getoon het se basispaar opeenvolging was bepaal. Die basispaar opeenvolgingdata het getoon dat daar 3 verskillende lipolitiese ensieme teenwoordig was (isokorismatase, sulfatase en fosfolipase patatien familie proteïne). Slegs die fosfolipase patatein proteïen is verder bestudeer. Getransleerde DNS basispaar opeenvolging data het getoon dat hierdie ensiem die klassieke esterase motief (G-X-S-X-G) sowel as die gekonserveerde histidien en aspartiensuur aminosure het. Filogenetiese analises het getoon dat fosfolipase patatien proteïene ‘n nuwe familie lipolitiese ensieme is, aangesien hulle nie deel uitmaak van die huidige 8 klasse lipoliliese ensieme nie. Die patatiengeen is heteroloog uitgedruk in *Escherichia coli* en biochemiese analise van die semi-suiwer ensiem het getoon dat die ensiem ‘n voorkeur het vir kort koolstof-ketting subtrate. Hierdie was ‘n sterk aanduiding dat patatien esterase aktiwiteit toon en nie lipase aktiwiteit nie. Die ensiem het ook optimaal gefunksioneer by 30˚C en by ‘n pH van 8. Hierdie was te wagte gewees aangesien die ensiem afkomstig was vanaf ‘n mesofiliese omgewing.

.

References

Abdel-Fattah, Y.R. (2002) Optimization of thermostable lipase production from a thermophilic *Geobacillus* sp. using Box-Behnken experimental design. *Biotechnology Letters* **24:** 1217-1222.

Al-Hasani, K., Simpfendorfer, K., wardan, H., Vadolas, J., and Zaibak, F. (2003). Development of a novel bacterial artificial chromosome cloning system for functional studies. *Plasmid* **49:** 184-187.

Aller, J.Y., and Kemp, P.F. (2008) Are archaea inherently less diverse than bacteria in the same environments? *FEMS Microbiology Ecology* **65:** 74-87.

Akoh, C.C., Lee, G. C., Liaw, Y. C., Huang, T. H., and Shaw, J. F. (2004) GDSL family of serine esterases/lipases. *Progress in Lipid Research* **43:**534–552.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25:** 3389-3402.

Amann, R., and Ludwig, W. (2000) Ribosomal RNA-targeted nucleic acid probes for studies in microbial ecology. *FEMS Microbiology Reviews* **24:** 555-565.

Anderson, C., Pinsirodom, P., and Parkin, K.L. (2002) Hydrolytic selectivity of patatin (lipid acyl hydrolase) from potato (*Solanum tuberosum*) tubers toward various lipids. *Journal of Food Biochemistry* **26:** 63-74.

Arpigny, J. I., and Jaeger, K. E. (1999) Bacterial lipolytic enzymes: classification and properties. *Biochemistry Journal* **343:** 177-183.

Ash, C., Priest, F.G., and Collins, D. (1993) Molecular identification of rRNA group 3 bacilli using a PCR probe. *Antonie van Leeuwenhoek* **64:** 253-260.

Baker, G. C., Smith, J, J., and Cowan, D. A. (2003) Review and re-analysis of domain

specific 16S primers. *Journal of Microbiological Methods* **55:** 541-555.

Baker, B.J., Moser, D.P., MacGregor, B.J., Fishbain, S., Wagner, M., Fry, N.K., Jackson, B., Speolstra, N., Loos, S., Takai, K., Lollar, B.S., Fredrickson, J., Balkwill, D., Onstott, T.C., Wimpee, C.F., and Stahl, D.A. (2003) Related assemblages of sulphate-reducing bacteria associated with ultradeep gold mines of South Africa and deep basalt aquifers of Washington State. *Environmental Microbiology* **5(4):** 267-277.

Banerji, S., and Flieger, A. (2004) Patatin-like proteins: a new family of lipolytic enzymes present in bacteria? *Microbiology* **150:** 522-525.

Baneyx, F. (1999) Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology* **10:** 411-421.

Bano, N., Ruffin, S., Ransom, B., and Hollibaugh, J.T. (2004) Phylogenetic composition of Artic ocean assemblages and comparison with Antartic Assemblages. *Applied and Environmental Microbiology* **70(2):** 781-789.

Barnard, T.G. (2005) A kinetic and molecular study of the lipase from *Geobacillus thermoleovorans* GE-7 (unpublished thesis). Department of Microbial, Biochemical and Food Biotechnology, University of the Free state, Bloemfontein.

Barta, J., and Bartova, V. (2008) Patatin, the major protein of potato (*Solanum tuberosum*) tubers, and its occurrence as genotype effect: processing versus table potatoes. *Czech Journal of Food Science* **26(5):** 347-359.

Bej, A.K., Mahbubani, M.H., Dicesare, J.L., and Atlas, R.M. (1991) Polymerase Chain Reaction-gene probe detection of microorganisms by using filter concentrated samples. *Applied and Environmental Microbiology* **57(12):** 3529-3534.

Béjà, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A., Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C., Jovanovich, S.B., Feldman, R.A., and DeLong, E.F. (2000) Construction and analysis of bacterial artificial libraries from a marine microbial assemblage. *Environmental Microbiology* **2:** 516-529.

Béjà, O. (2004) To BAC or not to BAC: marine ecogenomics. *Environmental Microbiology* **15:** 187-190.

Bernhardt, R. (2006) Cytochromes P450 as versatile biocatalysts. *Journal of Biotechnology* **124:** 128-145.

Benjdia, A., Subramanian, S., Leprince, J., Vaudry, H., Johnson, M. K., and Berteau, O. (2008) Anaerobic Sulfatase-maturing enzymes, first dual substrate radical S-adenosylmethionine enzymes. *Journal of Biological Chemistry* **283(26):** 17815-17826.

Berriman, M., and Rutherford, K. (2003) Viewing and annotating sequence data with Artemis. *Briefings in bioinformatics* **4(2):** 124-132.

Berteau, O., Quillot, A., Benjdia, A., and Rabot, S. (2006) A new type of bacterial sulfatase reveal a novel maturation pathway in prokaryotes. *The Journal of Biological Chemistry* **281(32):** 22464-22470.

Blankenfeldt, W., Kuzin, A.P., Skarina, T., Korniyenko, Y., Tong, L., Bayer, P., Janning, P., Thomashow, L.S., and Mavrodi, D.V. (2004) Structure and function of the phenazine biosynthetic protein PhzF from *Pseudomonas fluorescens*. *Proceedings of the National Academy of Science* **101(47):** 16431-16436.

Bond, P.L., Smriga, S.P., and Banfield, J.F. (2000) Phylogeny of microorganisms populating a thick, subaerial, predominantly lithotropic biofilm at an extreme acid mine drainage site. *Applied and Environmental Microbiology* **66(9):** 3842-3849.

Boon, N., De Windt, W., Verstraete, W., and Top, E.M. (2002) Evaluation of nested PCR-DGGE (denaturing gradient gel electrophoresis) with group-specific 16S rRNA primers for the analysis of bacterial communities from different wastewater treatment plants. *FEMS Microbiology Ecology* **39:** 101-112.

Bourret, R. A . (2006) Census of Prokaryotic Senses. *Journal of Bacteriology* **188(12):** 4165-4168.

Bruce, K.D., Hiorns., W.D., Hobman, J.L., Osborn, A.M., Strike, P., and Ritchie, D.A. (1992) Amplification of DNA from native populations of soil bacteria by using polymerase chain reaction. *Applied and Environmental Microbiology* **58:** 3413-3416.

Burgmann, H., Pesaro, M., Widmer, F., and Zeyer, J. (2001) A strategy for optimizing quality and quantity of DNA extracted from soil. *Journal of Microbiological Methods* **45:** 7-20.

Caruthers, J., Zucker, F., Worthey, E., Myler, P.J., Buckner, F., van Voorhuis, W., Mehlin, C., Boni, E., Feist, T., Luft, J., Gulde, S., Lauricella, A., Kaluzhniy, O., Anderson, L., Le Trong, I., Holmes, M.A., Earnest, T., Soltis, M., Hodgson, K.O., Hol, W.G.J., and Merritt, E.A. (2005) Crystal structures and proposed structural/functional claasification of three protozoan proteins from the isochorismatase superfamily. *Protein Structure* **14:** 2887-2894.

Cattaneo, R., Gorski, J., and Mach, B. (1981) Cloning of multiple copies of immunoglobulin variable kappa genes in cosmid vectors. *Nucleic Acids Research* **9(12):** 2777- 2790.

Chahinian, H., Nini, L., Boitard, E., Dubes, J.P., Comeau, L., and Sarda, C. (2002) Distinction between esterases and lipase: a kinetic study with vinyl esters and TAG. *Lipids* **37:** 653-662.

Chaudhuri, T.K., Horii, K., Yoda, T., Arai, M., Nagata, S., Terada, T.P., Uchiyama, H., ikura, T., Tsumoto, K., Kataoka, H., Matsushima, M., Kuwajima, K., and Kumagai, I. (1999) effect of the extra N-terminal methionine residue on the stability and folding of recombinant alpha-lactalbumin expressed in *Escherichia coli*. *Journal of Molecular Biology* **285:** 1179-1194.

Chaudhuri, S.R., Pattanayak, A.K., and Thakur, A.R. (2006) Microbial DNA extraction from samples of varied origin. *Current Science* **91(12):** 1697-1700.

Chauthaiwale, V.M., Therwath, A., and Deshpande, V.V. (1992) Bacteriophage Lambda as a cloning vector. *Microbiological Reviews* **56(4):** 577-591.

Chivian, D., Brodie, E.L., Alm, E.J., Culley, D.E., Dehal, P.S., Desantis, T.Z., Gihring, T.M., Lapidus, A., Lin, L.H., Lowry, S.R., Moser, D.P., Richardson, P.M., Southam, G., Wanger, G., Pratt, L.M., Anderson, G.L., Hazen, T.C., Brockman, F.J., Arkin, A.P., and Onstott, T.C. (2008) Environmental genomics reveals a single species ecosystem deep within the Earth. *Science* **10;322(5899):** 275-278.

Cho, A., Yoo, S.K., and Kim, E.J. (2000) Cloning, sequencing and expression in *Escherichia coli* of a thermophilic lipase from *Bacillus thermoleovorans* ID-1. *FEMS Microbiology Letters* **186:** 235-238.

Cho, J.C., and Giovanni, S, J. (2003) *Parvuarcula bermudensis* gen.nov., sp. nov., a marine bacterium that forms a deep branch in the *alpha-Proteobacteria*. *International Journal of Systemetic Microbiology* **53(4):** 1031-1036.

Chung, E.J., Lim, H.K., Kim, J-C., Choi, G.J., Park, E.J., Lee, M.H., Chung, Y.R., and Lee, S-W. (2008) Forest soil metagenome gene cluster involved in antifungal activity expression in *Escherichia coli*. *Applied and Environmental Microbiology* **74(3):** 723-730.

Connell, T.D., Martone, A.J., and Holmes, R.K. (1995) A new mobilizable cosmid vector for use in *Vibrio cholerae* and other gram-bacteria. *Gene* **153:** 85-87.

Cottrell, M.T., Moore, J.A., and Kirchman, D.L. (1999) Chitinases from uncultured marine microorganisms. *Applied and Environmental Microbiology* **65(6):** 2553-2557.

Courtois, S., Cappellano, C.M., Ball, M., Francou, F.X., Normand, P., Helynck, G., Martinez, A., kolvek, S.J., Hopke, J., Osburne, M.S., August, P.R., Nalin, R., Guérineau, M., Jeannin, P., Simonet, P., and Perndot, J. L. (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Applied and Environmental Microbiology* **69:** 49-55.

Cowan, D., Q. Meyer, W., Stafford, S., Muyanga, R., Cameron, and P. Wittwer. (2005) Metagenomics gene discovery: past, present and future. *Trends in Biotechnology* **23(6):** 321-329.

Cullen, D.W., and Hirsch, P.R. (1998) Simple and rapid method for direct extraction of microbial DNA from soil for PCR. *Soil Biology and Biochemstry* **30(8/9):** 983-993.

Daniel, R. (2004) The soil metagenome-a rich resource for the discovery of novel natural products. *Environmental Microbiology* **15:** 199-204.

Daniel, R. (2005) The metagenomics of soil. *Nature reviews* **3:** 470- 478.

Deflaun, M.F., Fredrickson J.K., Dong, H., Pfiffner, S.M., Onstott, T.C., Balkwill, D.L., Streger, S.H., Stackebrandt, E., Knoessen, S., and van Heerden, E. (2007) Isolation and characterization of a *Geobacillus thermoleovorans* strain from an ultra-deep South African gold mine. *Systematic and Applied Microbiology.* **30:** 152-164.

DeLong, E.F. (2002) Microbial population genetics and ecology. *Current Opinion in Microbiology* **5:** 520-524.

De Mot, R., and Parret, A. H. A. (2002) A novel class of self-sufficient cytochrome P450 monooxygenases in prokaryotes. *Trends in Microbiology* **10(11):** 502-508.

Díez, B., Pedrós-Alió, C., and Massana, R. (2001) Study of Genetic Diversity of Eukaryotic Picoplankton in Different Oceanic Regions by Small-Subunit rRNA Gene Cloning and Sequencing. *Applied and Environmental Microbiology* **67(7):** 2932–2941.

Dong, D., Yan, A., Liu, H., Zhang, X., and Xu, Y. (2006) Removal of humic substances from soil using aluminium sulfate. *Journal of Microbiological Methods* **66:** 217-222.

Edwards, R.A., Rodriquez-Brito, B., Wegley, L., Haynes, M., Breibart, M., Peterson, D.M., Saar, M.O., Alexander, S., Alexander E.C., and Rohwer, R. (2006) Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* **7:** doi: 10.1186/1471-2164-7-57.

Eland, C., Schmeisser, C., Leggewie, C., Babiak, P., Carballeira, J.D., Steele, H.L., Reymond, J-L., Jaeger, K-E., and Streit, W.R. (2006) Isolation and biochemical characterization of two novel metagenome-derived esterases. *Applied and Environmental Microbiology* **72(5):** 3637-3645.

Elshahed, M.S., Najar, F.Z., Aycock, M., Qu, C., Roe, B.A., and Krumholz, L.R. (2005) Metagenomic analysis of the microbial community at Zodletone Spring (Oklahoma): Insights into the genome of a member of a novel candidate division OD1. *Applied and Environmental Microbiology* **71(11):** 7598-7602.

England, L.S., Trevors, J.T., and Holmes, S.B. (2001) extraction and detection of baculoviral DNA from lake water, detritus and forest litter. *Journal of Applied Microbiology* **90:** 630-636.

Entcheva, P., Liebl, W, Johann, A, Hartsch, T, and Streit, W. R. (2001) Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Applied and Environmental Microbiology* **67(1):** 89-99.

Feng, L., Wang, W., Cheng, J., Ren, Y., Zhao, G., Tang, Y., Liu, X., Han, W., Liu, R., and Wang, L. (2007) Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Proceedings of the National Academy of Science* **104(13):** 5602-5607.

Ferrer, M., Golyshin, O., Beloqui, A., and Golyshin, P.N. (2007) Mining enzymes from extreme environments. *Current Opinion in Microbiology* **10(3):** 207-214.

Fischer, M., and Pleiss, J. (2003) The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Research* **31(1):** 319-321.

Fierer, N., Breibart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R.A., Felts, B., Rayhawk, S., Knight, R., Rohwer, F., and Jackson, R.B. (2007) Metagenomic and small-subunit rRNA analysis reveal the genetic diversity of bacteria, archaea, fungi and viruses in soil. *Applied and Environmental Microbiology* **73(21):** 7059-7066.

Fortin, N., Beaumier, D., Lee, K., and Greer, C.W. (2004) Soil washing improves the recovery of total DNA from polluted and high organic content sediments. *Journal of Microbiological Methods* **56:** 181-191.

Fredrickson, K., and Onstott, T.C. (1996) Microbes deep inside the earth. *Scientific American* **275:** 68-73.

Frølund, B., Palmgren, R., Keiding, K., and Nielsen, P.H. (1995) extraction of extracellular polymers from activated sludge using a cation exchange resin. *Water research* **30(8):** 1749-1758.

Funhoff, E.G., Salzmann, J., Bauer, U., Witholt, B., and van Beilen, J.B. (2007) Hydroxylation and epoxidation reactions catalyzed by CYP153 enzymes. *Enzyme and Microbial Technology* **40:** 806-812.

Furham, J.A., Comeau, D.E., Hagstrom, A., and Chan, A.M. (1988) Extraction from natural planktonic microorganisms of DNA suitable for molecular biological studies. *Applied and Enviromental Microbiology* **54(6):** 1426-1429.

Gabor, E.M., de Vries, E.J., Janssen, D.B. (2003) Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. *FEMS Microbiology Ecology* **44:** 153-163.

Gabor, E.M., de Vries, E.J., and Janssen, D.B. (2004) Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel *amidases*. *Environmental Microbiology* **6(9):** 948-958.

Gabor, E.M., Alkema, W.B., and Janssen. D.B. (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environmental Microbiology* **6(9):** 879-886.

Gadler, P., and Faber, K. (2006) New enzymes for biotransformations: microbial alkyl sulfatases displaying stereo-and enantioselectively. *Trends in Biotechnology* **25(2):** doi: 10.1016.

Geider, K. (1986) DNA cloning vectors utilizing replication functions of the filamentous phages of *Escherichia coli. Journal of general Virology* **67:** 2287-2303.

Gharizadeh, B., Ghaderi, M., and Nyren, P. (2007) Pyrosequencing Technology for Short DNA Sequencing and Whole Genome Sequencing. *Seibutsu Butsuri* **47:** 129-132.

Gillespie, D.E., Brady, S.F., Bettermann, A.D., Cianciotto, N.P., Liles, M.R., Rondon, M.R., Clardy, J., Goodman, R.M., and handelsman, J. (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Applied and Environmental Microbiology* **68(9):** 4301-4306.

Giovannoni, S.J., deLong, E.F., Schmidt, T.M., and Pace, N.R. (1990) Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. *Applied and Environmental Microbiology* **56(8):** 2572-2575.

Goldberg, S.M., Johnson, J., Busam, D., Feldblyum, T., Ferriera, S., Friedman, R., Halpern, A., Khouri, H., Kravitz, S.A., Lauro, F.M., Li, K., Rogers, Y.H., Strausberg, R., Sutton, G, Tallon, L., Thomas, T., Venter, E., Frazier, M., and Venter, J. (2006) A sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine genomes. *Proceedings of the National Academy of Science* **103(30):** 11240-11245.

González-Toril, E., Llobet-Brossa, E., Casamayor, E.O., Amann, R., and Amils, R. (2003) Microbial Ecology of an Extreme Acidic Environment, the Tinto River. *Applied and Environmental Microbiology* **69(8):** 4853-4865.

Green, B. D., and M. Keller. (2006) Capturing the uncultivated majority. *Current Opinion in Biotechnology* **17:**236-240.

Gustafsson, C., Govindarajan, S., and Minshull, J. (2004) Codon bias and heterologous protein expression. *Trends in Biotechnology* **22(7):** 346-353.

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):** 245-249.

Handelsman, J., Liles, M., Mann, D., Reisenfeld, C., and Goodmann, R.M. (2002) Cloning

the metagenome: culture-independent access to the diversity and functions of the uncultivated microbial world. *Methods in Microbiology* **33:** 241-255.

Handelsman, J. (2004) Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews* **68(4):** 669-685.

Hannemann, F., Bichet, A., Ewen, K. M., and Bernhardt, R. (2007) Cytochrome P450 systems-biological variations of electron transport chains. *Biochimica et Biophysica Acta* **1770:** 330-344.

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 Microbiol Ecology* **59:** 524-534.

Hawumba, J. F., Theron, J., and Brozel, V.S. (2002) Thermophilic protease-producing *Geobacillus* from Buranga hot springs in western Uganda. *Current Microbiology* **45 (2):**144-150.

He, J., Xu, Z., and Hughes, J. (2005) Pre-lysis washing improves DNA extraction from a forest soil. *Soil Biology and Biochemistry* **37:** 2337-2341.

Healy, F.G., Ray, R.M., Aldrich, H.C., Wilkie, A.C., Ingram, L.O., and Shanmugam, K.T. (1995) Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Applied Microbiology of Biotechnology* **43:** 667-674.

Henne, A., Daniel, R., Schmitz, R.A., and Gottschalk, G. (1999) Construction of Environmental DNA Libraries in *Escherichia coli* and Screening for the presence of genes conferring utilization of 4-Hydroxybutyrate. *Applied and Environmental Microbiology* **65(9):** 3901-3907.

Henneberger, R.M., Walter, M.R., and Anitori, R.P. (2006) Extraction of DNA from acidic, hydrothermally modified volcanic soils. *Environmental Chemistry* **3:** 100-104.

Hirschberg, H. J. H.B., Willem. J., Simons, F. A., Dekker, N., and Egmond, M.R. (2001) Cloning, expression, purification and characterization of patatin, a novel phospholipase A. *European Journal of Biochemistry* **268:** 5037-5044.

Hohn, B., and Murray, K. (1977) Packaging recombinant DNA molecules into bacteriophages particles *in vitro*. *Proceedings of the National Academy of Science* **74(8):** 3259-3263.

Hu, P., Brodie, E.L., Suzuki, Y., McAdams, H.H., and Anderson, G.L. (2005) Whole-genome transcriptional analysis of heavy metals stresses in *Caulobacter crescentus*. *Journal of Bacteriology* **187(24):** 8437-8449.

Huang, Y., Lai, X., He, X., Cao, L., Zeng, Z., Zhang, J., and Zhou, S. (2009) Characterization of a deep-sea sediment metagenomic clone that produces water-soluble melanin in *Escherichia coli. Marine Biotechnology* **11:** 124-131.

Hurt, R.A., Qiu, X., wu, L., Roh, Y., Palumbo, A.V., Tiedje, J.M., and Zhou, J. (2001) Simultaneous recovery of RNA and DNA from soils and sediments. *Applied and Environmental Microbiology* **67(10):** 4495-4503.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Welch, D.M. (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology* **8:** doi: 10.1186.

Ikner, L.A., Toomey, R.S., Nolan, G., Neilson, J.W., Pryor, B.M., Maier R.M. (2007) Culturable microbial diversity and the impact of tourism in Kartchner Caverns, Arizona. *Microbial Ecology* **53(1):** 30-42.

Jackson, C.R., Churchill, P.F., and Roden, E.E. (2001) Successional changes in bacterial assemblage structure during epilithic biofilm development. *Ecology* **82(2):** 555-566.

Jacobsen, C.S., and Rasmussen, O.F. (1992) Development and application of a new method to extract bacterial DNA from soil based on separation of bacteria from soil with cation-exchange resin. *Applied and Environmental Microbiology* **58:** 2458-2462.

Jung, E., and Williams, K.L. (1997) The production of recombinant glycoproteins with special reference to simple eukaryotes including *Dictostelium discoideum*. *Biotechnology and Applied Biochemistry* **25:** 3-8.

Jung, H, C., Ko, S., Ju, S.J., Kim, E.J., Kim, M,.K., and Pan, J.G. (2003) Bacterial cell surface display of lipase and its randomly mutated library facilitates high-throuput screening of mutants showing higher specific activities. *Journal of Molecular Catalysis B: Enzymatic* **26:** 177-184.

Keister, D.L., Tully, R.E., and van Berkum, P. (1999) A cytochrome P450 cluster in the *Rhizobiaceae*. *Journal of General and Applied Microbiology* **45:** 301-303.

Keller, M., and Ramos, J.L. (2008) Microbial goods from single cells and metagenomes. *Current Opinion in Microbiology* **11:** 195-197.

Kim, U-J., Shizuya, H., de Jong, P.J., Birren, B., and Simon. M. I. (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Research* **20(5):** 1083-1085.

Kim, Y-J., Choi, G-S., Kim, S-B., Yoon, G-S., Kim, Y-S., and Ryu, Y-W. (2006) Screening and characterization of a novel esterase from a metagenomic library. *Protein Expression and Purification* **45:**315-323.

Kim, J.S., and Crowley, D.E. (2007) Microbial diversity in natural asphalts of the La Brea Tar pits. *Applied and Environmental Microbiology* **73(14):** 4579-4591.

Kim, B.S., Kim, S.Y., Park, J., Hwang, K.Y., Yoon, Y.J., Oh, W.K., Kim, B.Y., and Ahn, J.S. (2007) Sequence-based screening for self sufficient P450 monoxygenases from a metagenome library. *Journal of Applied Microbiology* **102:** 1392-1400.

Knietsch, A., Bowien, S., whited, G., Gottschalk, G., and Daniel, R. (2003) Identification and characterization of coenzyme B_{12}-dependent glycerol dehydratase-and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Applied and Environmental Microbiology* **69(6):** 3048-3060.

Kubota, M., Nodate, M., Yasumoto-Hirose, M., Uchiyama, T., Kagami, O., Shizuri, Y., and Misawa, N. (2005) Isolation and functional analysis of cytochrome P450 CYP153A genes from various environments. *Bioscience Biotechnology Biochemistry* **69 (12):** 2421-2430.

Kumar S., Dudley J., Nei M., and Tamura K. (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics* **9:** 299-306.

Kuske, C.R., Banton, K.L., Adora, D.L., Stark, P.C., Hill, K.K., and Jackson, P.L. (1998) Small-scale DNA sample preparation method for field PCR detection of microbial cells and spores in soil. *Applied and Environmental Microbiology* **64:** 2463-2472.

Kouker, G., and Jaeger, K-E. (1987) Specific and sensitive plate assay for bacterial lipases. *Applied and Environmental Microbiology* **53(1):** 211-213.

Labuschagne. M., and J. Albertyn. (2007) Cloning of an epoxide hydrolase-encoding gene from *Rhodotorula mucilaginosa* and functional expression in *Yarrowia lipolytica. Yeast* **24:** 69-78.

Laib, S., Robertson, R.M., and Smith, D.E. (2006) Preparation and characterization of a set of linear DNA molecules for polymer physics and rheology studies. *Macromolecules* **39:** 4115-4119.

Lämmle, K., Zipper, H., Breuer, M., Hauer, B., Buta, C., Brunner, H., and Rupp, S. (2007) Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *Journal of Biotechnology* **127:** 575-592.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680–685.

Lane, D.J. (1991) 16S/23S rRNA sequencing In: Nucleic Acid Techniques in Bacterial Systematics (Stackebrandt, E. and Goodfellow, M. Eds), pp. 115-175. Wiley, Chichester, UK.

Le Cleir, G.R., Buchan, A., Maurer, J., Moran, M.A., and Hollibaugh, J.T. (2007) Comparison of chitinolytic enzymes from an alkaline. Hypersaline lake and an estuary. *Environmental Microbiology* **9(1):** 197-205.

Lee, S.W., Won, K., Lim, H.K., Kim, J. C., Choi G. J., and Cho, K. Y. (2004) Screening for novel lipolytic enzymes from uncultured soil microorganisms. *Applied Microbiology and Biotechnology* **65:** 720-726.

Lee, M-H., Lee, C-H., Oh, T-K., Song, J. K., and Yoon, J-H. (2006) Isolation and characterization of a novel lipase from a metagenomic library of tidal flat sediments: Evidence for a new family of bacterial lipases. *Applied and Environmental Microbiology* **72(11):** 7406-7409.

Lee, S.W. (2005) Metagenome, the untapped microbial genome, toward discovery of novel microbial resources and application into the plant pathology. *Plant Pathology Journal* **21(2):** 93-98.

Leow, T.C., Rahman, R.N.Z.R.A., Basri, M., and Salleh, A.B. (2003) High Level Expression of Thermostable Lipase from *Geobacillus* sp. Strain T1. *Bioscience Biotechnology Biochemistry* **68(1):** 96-103.

Leow, T.C., Rahman, R.N.Z.R.A., Basri, M., and Salleh, A.B. (2007) A thermoalkaliphilic lipase of *Geobacillus* sp. T1. *Extremophiles* **11:** 527-535.

Li, H., and Zhang, X. (2005) Characterization of thermostable lipase from thermophilic *Geobacillus* sp. *TW1*. *Protein Expression and Purification* **42:** 153-159.

Li, Y., Wexler, M., Richardson, D.J., Bond, P., Johnston, L., and Andrew, W.B. (2005) Screening a wide host-range water metagenomic library in trytophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of clone *trp* genes. *Environmental Microbiology* **7(12):** 1927-1936.

Li, X., and Qin, L. (2005) Metagenomics-based drug discovery and marine microbial diversity. *Trends in Biotechnology* **23(11):** 539-543.

Lim, H.K., Chung, E.J., Kim, J-C., Choi, G.J., Jang, K.S., Chung, J.R., Cho, K.Y., and Lee, S-W. (2005) Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. *Applied and Environmental Microbiology* **71(12):** 7768-7777.

Lipthay, J.R., Enzinger, C., Johnsen, K., Aamand, J., and Sørenson, S.J. (2004) Impact of DNA extraction method on bacterial community composition measured by denaturing gradient gel electrophoresis. *Soil Biology and Biochemistry* **36:** 1607-1614.

Liu, W.T., Marsh, T.L., Cheng, H., and Forney, L.J. (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Applied and Environmental Microbiology* **63(11):** 4516-4522.

López-Garc a, P., Brochier, C., Moreira, D., and Rodriguez-Valera, F. (2004) Comparative analysis of a genome fragment of an uncultivated mesopelagic. *Environmental Microbiology* **6:** 19-34.

Lorenz, P., and Schleper, C. (2002) Metagenome- a challenging source of enzyme discovery. *Journal Molecular Catalysts B: Enzymology* **19-20:** 13-19.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Kumar, Y., Buchner, A., Lai, T., Steppi, S., Jobb, G., Förster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., König, A., Liss, T., Lüßmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A., and Schleifer, K-H. (2004) ARB: a software environment for sequence data. *Nucleic Acids Research* **32(4):** 1363-1371.

Luna, G.M., Dell'Anno, A., and Danovaro, R. (2006) DNA extraction procedure: a critical issue for bacterial diversity assessment in marine sediments *Environmental Microbiology* **8(2):** 308-320.

Luo, Y., Xu, X., Ding, Z., Liu, Z., Zhang, B., Yan, Z., Sun, J., Hu, S., and Hu, X. (2008) Complete genome of *Phenylobacterium zucineum* - a novel facultative intracellular bacterium isolated from human erythroleukemia cell line K562. *BMC Genomics* **9:** 386.

Lyautey, E., Lacoste, B., Ten-Hage, L., Rols, J-L., and Garabetian. (2005) Analysis of bacterial diversity in river biofilms using 16S rDNA PCR-DGGE: methodological settings and fingerprints interpretation. *Water Research* **39:** 380-388.

MacNeil, I.A., Tiong, C.L., Minor, C., August, P.R., Grossman, T.H., Loiacono, K.A., Lynch, B.A., Philips, T., Narula, S., Sundaramoorthi, R., Tyler, A., Alredge, T., Long, h., Gilman, M., Holt, D., and Osburne, M.S. (2001) Expression and isolation of antimicrobial small molecules from soil DNA libraries. *Journal of Molecular Microbiology and. Biotechnology* **3(2):** 301-308.

Majernik, A., Gottschalk, G., and Daniel R. (2001) Screening of environmental DNA libraries for the presence of genes conferring $Na^+(Li^+)/H^+$ antiporter activity on *Escherichia coli* : characterization of the recovered genes and the corresponding gene products. *Journal of Bacteriology* **183:** 6645-6653.

Makrides, S. C. (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiological Reviews* **60(3):** 512-538.

Manichanh, C., Chapple, C. E., Frangeul, L., Gloux, K., Guigo, R., and Dore, J. (2008) A comparison of random sequence reads versus 16S rDNA sequences for estimating the biodiversity of a metagenome library. *Nucleic Acids Research* **36(16):** 5180-5188.

Marguiles, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, K., Braverman, M.S., Chen, Y-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J-B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Pu, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasen, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S,H., Wang, Y., Weiner, M.P., Yu, P., Begly, R.F., and Rothberg, J.M. (2005)

Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437(7057):**376-380.

Maqbool, Q.A., Johri, S., Rasool, S., Hassan, S.R., Verma, V., Nargotra, A., Koul, S., and Qazi, G.N. (2006) Molecular cloning of carboxylesterase gene and biochemical characterization of encoded protein from *Bacillus subtilis* (RRL BB1). *Journal of Biotechnology* **125:** 1-10.

Maruyama, I.N., Maruyama, H.I., and Brenner, S. (1994) λfoo: A λ phage for the expression of foreign proteins. *Proceedings of the National Academy of Science* **91:** 8273-8277.

Matin, M.M., and Hornby, D.P. (2000) A positive selection vector combining tetracycline resistance that eliminates the need for bacterial plating. *Analytical Biochemistry* **278:** 46-51.

McMullan, G., Christie, J.M., Rahman, T.J., Banat, I.M., Ternan, N.G., and Marchant, R. (2004) Habitat, applications and genomics of the aerobic, thermophilic genus *Geobacillus. Biochemical Society Transactions* **32(2):** 215- 217.

Miller, D.N., Bryant, J.E., Madsen, E.L., and Ghiorse, W.C. (1999) Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Applied and Environmental Microbiology* **65:** 4715-4724.

Miller, D.N. (2001) Evaluation of gel filtration resins for the removal of PCR- inhibitory substances from soils and sediments. *Journal of Microbiological Methods* **44:** 49-58.

Moreira, D., Rodríguez-Valera, F., and López-García, P. (2006) Metagenomic analysis of mesopelagic Antarctic plankton reveals a novel deltaproteobacterial group. *Microbiology* **152:** 505-517.

Moon, D.A., and Magor, K.E. (2004) Construction and characterization of a fosmid library for comparative analysis of the duck genome. *Animal Genetics* **35:** 408-423.

Moore, M.J., Dhingra, A., Soltis, P.S., Shaw, R., Farmerie, W.G., Folta, K.M., and Soltis, D.E. (2006) Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biology* **6:** doi: 11.86/1471-2229-6-17.

More, M.I., Herrick, J.B., Silva, M.C., Ghiorse, W.C., and Madsen, E.L. (1994) Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. *Applied and Environmental Microbiology* **60(5):** 1572-1580.

Mumy, K.L., and Findlay, R.H. (2004) Convenient determination of DNA extraction efficiency using an external DNA recovery standard and quantitative- competitive PCR. *Journal of Microbiological Methods* **57:** 259-268.

Niemi, R.M., Heiskanen, I., Wallenius, K., and Lindström. (2001) Extraction and purification of DNA in rhizosphere soil samples for PCR-DGGE analysis of bacterial consortia. *Journal of Microbiological Methods* **45:** 155-165.

Neria-González, I., Wang, E.T., Ramírez, F., Romero, J.M., and Rodríguez. (2006) characterization of bacterial community associated to biofilms of corroded oil pipelines from southeast of Mexico. *Anaerobe* **12:** 122-133.

Novy, R., Drott, D., Yaeger, K., and Mierendorf, R. (2001) Overcoming the codon bias of *E.coli* for enhanced protein expression. *inNovations* **12:** 1-3.

O'Brien, D., Carton, M., Eardly, D., and Patching, J.W. (1998) In situ filtration and preliminary molecular analysis of microbial biomass from the Rainbow hydrothermal plume at 36˚15'N on the Mid-Atlantic Ridge. *Earth and Planetary Science Letters* **157:** 223-231.

Oefner, P.J., Hunicke-Smith, S.P., Chiang, L., Dietrich, F., Mulligan, J., and Davis, R.W. (1996) Efficient random subcloning of DNA sheared in a recirculating point-sink flow system. *Nucleic Acids Research* **24:** 3879-3886.

O' Keefe, D. P., Tepperman, J. M., Dean, C., Leto, K. J., Erbes, D. L., and Odell, J. T., (1994) Plant expression of a bacterial cytochrome P450 that catalyzes activation of a sulfonylurea pro-herbicide. *Plant Physiology* **105:** 473-482.

Omemu, A.M., Akpan, I., Bankole, M.O., and Teniola, O.D. (2005) Hydrolysis of raw starches by *amylase* of *Aspergillus niger* AM07 isolated from the soil. *African Journal of Biotechnology* **4:** 19-25.

Onstott, T.C., Lin, L, H., Davidson, M., Mislowack, B., Borcsik, J., Slater, G., Ward, J., Lollar, B.S., Lippmann-Pipke, J., Boice, E., Pratt, L.M., Pniffner, S., Moser, D., Gihring, T., Kieft, T.L., Phelps, T.J., van Heerden, E., Litthauer, D., Deflaun, M., Rothmel, R., Wanger, G., and Southam, G. (2006) The origin and age of biogeochemical trends in deep fracture water of the Witwatersrand basin, South Africa. *Geomicrobiology Journal* **23:** 369-414.

Osborne, C.A., Galic, M., Sangwan, P., and Janssen, P.H. (2005). PCR-generated artefact from 16S rRNA gene-specific primers. *FEMS Microbiology Letters* **248:** 183-187.

Parsons, J.F., Calabrese, K., Eisenstein, E., and Ladner, J.E. (2003) Structure and mechanism of *Pseudomonas aeruginosa* PhzD, an isochorismatase from the phenazine biosynthetic pathway. *Biochemistry* **42:** 5684-5693.

Petit, F., Craquelin, S., Guespin-Michel, J., and Buffet-Janvresse, C. (1999) Nucleic acid extraction from polluted estuarine water for detection of viruses and bacteria by PCR and RT-PCR analysis. *Research Microbiology* **150:** 143-151.

Pfiffner, S.M., Cantu, J.M., Smithgall, A., Peacock, A.D., and White, D.C. (2006) Deep subsurface microbial biomass and community structure in Witswatersrand basin mines. *Geomicrobiology Journal* **23:** 431-442.

Pierce, J.C., Sauer, B., and Sternberg, N. (1992) A positive selection vector for high molecular weight DNA by the bacteriophage P1 system: improved cloning efficiency. *Proceedings of the National Academy of Science.* **89:** 2056-2060.

Poli, A., Romano, I., Caliendo, G., Nicolaus, G., Orlando, P., de Falco, A., Lama, L., Gambacorta, A., and Nicolaus, B. (2006) *Geobacillus toebii* subsp. *decanicus* subsp.nov., a hydrocarbon-degrading, heavy metal resistant bacterium from hot compost. *Journal of General and Applied Microbiology* **52:** 223-234.

Pontes, D.S., Lima-Bittencourt, C.I., Chartone-Souza, E., and Nascimento, A.M.A. (2007). Molecular approaches: advantages and artifacts in assessing bacterial diversity. *Journal of Indian Microbiology and Biotechnology* **34:** 463-473.

Pühler, A. (2008) Genome research in the light of ultrafast sequencing technologies. *Journal of Biotechnology* **136:** 1-2.

Purdy, K. (2005) Nucleic acid recovery from complex environmental samples. *Methods in Enzymology* **397:** 271-292.

Quaiser, A., Ochsenreiter, T., Klenk, H.P., Kletzin, A., Treusch, A.H., Meurer, G., Eck, J., Sensen, C.W., and Schleper, C. (2002) First insight into the genome of an uncultivated crenarchaeote from soil. *Environmental Microbiology* **4(10):** 603-611.

Raes, J., Foerstner, K.U., and Bork, P. (2007) Get the most out of your metagenome: computational analysis of environmental sequence data. *Current Opinion in Microbiology* **10:** 1-9.

Rahman, R.N.Z.A., Leow, T.C., Salleh, A.B and Basri, M. (2007) *Geobacillus zalihae* sp.nov, a thermophilic lipolytic bacterium isolated from palm oil mill effluent in Malaysia. *BMC Microbiology* **7(77):** doi: 10.1186/1471-2180-7-77.

Rajendran, J., and Gunasekaran, P. (2008) Strategies for accessing soil metagenome for desired applications. *Biotechnology Advances* **26:** 576-590.

Rai, M., and Padh, H. (2001) Expression systems for production of heterologous proteins. *Current Science* **80(9):** 1121-1128.

Ranjan, R., Grover, A., Karardar., R. K., and Sharma, R. (2005) Isolation of novel lipolytic genes from uncultured bacteria of pond water. *Biochemical and Biophysical Research Communications* **335:** 57-65. *Journal of Biochemistry* **341:** 409-413.

Rees, H. C., Grant, S., Jones, B., Grant, W. D., and Heaphy, S. (2003) Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA

libraries. *Extremophiles* **7:** 415-421.

Reisenfeld, C.S., Scholss, P.D., and Handelsman, J. (2004) Metagenomics: Genomic analysis of microbial communities. *Annual Reviews Genetics* **38:** 525-552.

Reysenbach, A. L., and Pace, N.R. (1995) In Archaea: A Laboratory Manual- Thermophiles (Robb, F.T and Place, A.R., eds), pp. 101-107, Cold Spring Harbor Press, Cold Spring Harbor.

Rhazi, N., Galleni, M., Page, M.I., and Frère, J-M. (1999) Peptidase activity of beta-lactamases. *Biochemistry Journal* **341:** 409-413.

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 Enivironmental Microbiology* **71(2):** 817-825.

Rincón, B., Rapsoo, F., Borja, R., Gonzalez, J.M., Portillo, M.C., and Saiz-Jimenez, C. (2006) Performance and microbial communities of a continuous stirred tank anaerobic reactor treating two-phases olive mill solid wastes at low organic loading rates. *Journal of Biotechnology* **121:** 534-543.

Ro, H.-S., Hong, H.P., Kho, B.H., Kim, S., and Chung, B.H. (2004) Genome-wide cloning and characterization of microbial esterases, *FEMS Microbiology Letters* **233:** 97–105.

Robe, P., Nalin, R., Capellano, C., Vogel, T.M., and Simonet, P. (2003) Extraction of DNA from soil. *European Journal of Soil Biology* **39:** 183-190.

Roberts, G. A., Celik, A., Hunter, D. J. B., Ost, T. W. B., White, J. H., Chapman, S. K., Turner, N. J., and Flitsch, S. L. (2003) A self-sufficient cytochrome P450 with a primary structural organization that includes a flavin domain and a [2Fe-2S] redox centre. *The Journal of Biological Chemistry* **278(49):** 48914-48920.

Roh, C., villatte, F., Kim, B.G., and Schmid, R.D. (2007) Screening and purification for novel cytochrome b_5 from uncultured environmental micro-organisms. *Letters in Applied Microbiology* **44:** 475-480.

Ronaghi, M. (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Research* **11:** 3-11.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., MacNeil, I.A., Minor, C., Tiong, C.L., Gilman, M., Osburne, M.S, Clardy, J., Handelsman, J., and Goodmann, R.M. (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology* **66(6):** 2541-2547.

Rusnak, F., Liu, J., Quinn, N., Berchtold, G. A., and Walsh, C. T. (1990) Subcloning of the enterobactin biosynthetic gene *entB*: expression, purification, characterization, and substrate specificity of isochorismatase. *Biochemistry* **29:** 1425-1435.

Rydel, T. J., Williams, J. M., Krieger, E., Moshiri, F., Stallings, W. C., Brown, S. M., Pershing, J. C., Purcell, J. P., and Alibhai, M. F. (2003) The crystal structure, mutagenesis, and activity studies reveal that patatin is a lipid acyl hydrolase with a Ser-Asp catalytic dyad. *Biochemistry* **42:** 6696-6708.

Sambrook, J., Fritsch, E.F., and Maniatis T. (1989) *Molecular Cloning: A Laboratory Manual*. Vol 2. Cold Spring Harbor Laboratory Press. Plainview, New York.

Sambrook, J., and Russell, D.W. (2001) Molecular cloning: A laboratory manual (second Ed), volume 2, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Satyanarayana, T., Raghukumar, C., and Shivaji, S. (2005) Extremophilic microbes: Diversity and perspectives. *Current Science* **89(1):** 78-89.

Schleheck, D., and Cook, A.M. (2005) ω-Oxygenation of the alkyl sidechain of linear alkylbenzenesulfonate (LAS) surfactant in *Parvibaculum lavamentivorans. Archives in Micobiology* **183:** 369-377.

Schloss, P.D., and Handelsman, J. (2003) Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology* **14:** 303-310.

Schloss, P.D., and Handelsman, J. (2005) Introducing DOTUR, a computer programme for defining operational taxonomic units and estimating species richness. *Applied and Environmental Microbiology* **71(3):** 1501-1506.

Schmeisser, C., Stöckigt, C., Raasch, J., Wingender, K.N., Timmis, O.F., Wenderoth, H.-C., Flemming, H., Ziesegang, R.A., Schmitz, K.E., Jaeger, K., and W.R. Streit. (2003) Metagenome survery of biofilms in drinking-water networks. *Applied and Environmental Microbiology* **69:** 7298–7309.

Schmitz, J. E., Daniel, A., Collin, M., Schuch, R., and Fischetti, V.A. (2008) Rapid DNA library construction for functional genomic and metagenomic screening. *Applied and Environmental Microbiology* **74(5):** 1649-1652.

Schwatrz, T., Hoffmann, SD., and Obst, U. (2003) Formation of natural biofilms during chlorine dioxide and u.v. disinfection in a public drinking water distribution system. *Journal of Applied Microbiology* **95:** 591-601.

Szczepanowski, R., Bekel, T., Goesmann, A., Krause, L., Krömeke, H., Kaiser, O., Eichler, W., Pühler , A., and Schlüter, A. (2008) Insight into the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to antimicrobial drugs analysed by the 454-pyrosequencing technology. *Journal of Biotechnology* **136:** 54–64.

Sebat, J.L., Colwell, F.S., and Crwaford, R.L. (2003) Metagenomic profiling: Microaaray analysis of an environmental genomic library. *Applied and Environmental Microbiology* **69(8):** 4927-4934.

Sharma, N., Gruszewski, H.A., Park, S-W., Holm, D.G., and Vivanco, J.M. (2004) purification of an isoform of patatin with antimicrobial activity against *Phytophora infestans*. *Plant Physiology and Biochemistry* **42:** 647-655.

She, K. (2003) So you want to work with giants: The BAC vector. *Bio teach Journal* **1:** 69-74.

Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proceedings of the National Academy of Science* **89:** 8794-8797.

Shih, C-J., and Lai, M-C. (2007) Analysis of the AAA + chaperone *clpB* gene and stress-response expression in the halophilic methanogenic archaeon *Methanohalophilus portucalensis*. *Microbiology* **153:** 2572-2583.

Sheng. Y., Mancino, V., and B. Birren. (1995) Transformation of *Escherichia coli* with large DNA molecules by electroporation. *Nucleic Acids Research* **23 (11):** 1990-1996).

Sjöling, S., and Cowan, D. (2008) Metagenomics: Microbial community genomes revealed. In Psychrophiles: from biodiversity to biotechnology. Margesin, R., Schinner, F., Marx, J., and Gurday. G (eds). Berlin Heidelberg: Springer-verlag, pp.313-330.

Soliman, N.A., Knoll, M., Abdel-Fattah, Y.R., Schmid, R.D., and Lange, S. (2007) Molecular cloning and characterization of thermostable esterase and lipase from *Geobacillus thermoleovorans* YN isolated from desert soil in Egypt. *Process Biochemistry* **42:** 1090-1100.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J. and Klenk, D.C. (1985) Measurement of protein using bicinchoninic acid. *Analytical Biochemistry* **150:**76-85.

Spangenberg, J.E., and Frimmel, H.E. (2001) Basin-internal derivation of hydrocarbons in the Witwatersrand basin, South Africa: evidence from bulk and molecular ^{13}C data. *Chemical Geology* **173:** 339-355.

Staden, R. (1996). Staden package. *Molecular Biology* 5(3): 233-241.

Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H., and DeLong, E.F. (1996) Characterization of uncultivated prokaryotes: Isolation and Analysis of a 40-kilobase pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology* **178(3):** 591-599.

Sternberg, N. (1990) Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. *Proceedings of the National Academy of Science* **87:** 103-107.

Streit, W.R., and Schmitz, R.A. (2004) Metagenomics- the key to the uncultured microbes. *Current Opinions in Microbiology* **7:** 492-498.

Streit, W.R., Daniel, R., and Jaeger, K-E. (2004) Prospecting for biocatalysts and drugs in the genomes of non-cultured microorganisms. *Current Opinion in Biotechnology* **15:** 285-290.

Stepanauskas, R., and Sieracki, M.E. (2007) Matching phylogeny and metabolisms in the uncultured marine bacteria, one cell at a time. *Proceedings of the National Academy of Science* **104(21):** 9052-9057.

Tamaki, H., Sekiguchi, Y., Hanada, S., Nakamura, K., Nomura, N., Matsatoshi., and Kamagata, Y. (2005) Comparative analysis of bacterial diversity in freshwater sediment of a shallow eutrophic lake by molecular and improved cultivation-based techniques. *Applied and Environmental Microbiology* **71(4):** 2162-2169.

Tebbe, C.C., and Vahjen, W. (1993) Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Applied and Environmental Microbiology* **59(8):** 2657-2665.

Tirawongsaroj, P., Sripang, R., Harnpicharnchai, P., Thonggaram, T., Champreda, V., Tanapongpipat, S., Pootanakit, K., and Eurwilaichitr, L. (2008) Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. *Journal of Biotechnology* **133:** 42-49.

Trachoo, N. (2004) Biofilm removal techniques using sands as a search tool for accessing microbial attachment on surface. *Songklanakarin Journal of Science and Technology* **26(1):** 109-115.

Thompson, J.D., Higgins, D.G., and Gibson T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighing, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22:** 4673-4680.

Towner, P., 1991. Isolation of DNA by SDS-*proteinase* K treatment. In: T.A.
Brown (ed.) Essential molecular biology, A practical approach Vol 1, Oxford University
Press, UK, pp.52-53.

van Beilen, J.B., Funhoff, E.G., van Loon, A., Just, A., Kaysser, L., Bouza, M., Holtackers, R., Ro¨thlisberger R., Li, Z., and Witholt, B. (2006) Cytochrome P450 Alkane Hydroxylases of the CYP153 Family Are Common in Alkane-Degrading Eubacteria Lacking Integral Membrane Alkane Hydroxylases. *Applied and Enviromental Microbiology* **72 (1):** 59-65.

Van Heerden, E., Oppermann, D.J., Bester, A.P., Van Marwijk, J., Cason, E.D., Litthauer, D., Piater, L.A., and Onstott, T.C. (2008) Metabolic promiscuity from the deep subsurface: a story of survival or superiority. *Proceedings of SPIE* **7097:** 1-12.

Vazquez, S.C., Rios-Merino, L.N., MacCormack, W.P., and fraile, E.R. (1995) *Protease*-producing psychrotrophic bacteria isolated from Antartica. *Polar Biology* **15:** 131-135.

Vergin, K.L., Urbach, E., Stein, J.L., DeLong, E.F., Lanoil, B.D., and Giovannoni, S.J. (1998) screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order *Planctomycetales*. *Applied and Environmental Microbiology* **64(8):** 3075-3078.

Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K-E., and Streit, W.R, (2003) Prospecting for novel biocatalysts in a soil metagenome. *Applied and Environmental Microbiology* **69(10):** 6235-6242.

Urlacher, V.B., and Eiben, S. (2006) Cytochrome P450 monoxygenases: perspectives for synthetic application. *Trends in Biotechnology* **24(7):** 325-330.

Wang, C., Mitsuya, Y., Gharizadeh, B., Ronaghi, M., and Shafer, R.W. (2007) Characterization of mutation spectra with ultra-deep Pyrosequencing: Application to HIV-1 drug resistance. *Genome Research* **doi:** 10.1101/gr.6468307.

Wanger, G., Onstott, T. C., and Southam, G. (2008) Stars of the terrestrial subsurface: A novel ‘star-shaped’ bacterial morphotype from a South African platinum mine. *Geobiology* **6:** 325-330.

Ward, J.A., Slater, G.F., Moser, D.P., Lin, L.H., Lacrampe-Couloume, G., Bonin, A.S., Davidson, M., Hall, J.A., Mislowack, B., Bellamy, R.E.S., Onstott, T.C., and Lollar, B.S. (2004) Microbial hydrocarbon gases in the Witwatersrand basin, South Africa: implications for the deep biosphere. *Geochemica et Cosmochimica Acta* **68(15):** 3239-3250.

Watanabe, K., Kodama, Y., and Harayama, S. (2001) Design and evaluation of PCR primers to amplify bacterial 16S ribosomal fragments used for community fingerprinting. *Journal of Microbiological Methods* **44:** 253-262.

Woebken, D., Teeling, H., Wecker, P., Dumitriu, A., Kostadinov, I., DeLong, E.F., Amann, R., and Glöckner, F.O. (2007) Fosmids of novel marine *Planctomycetes* from the Namibian and Oregon xcoast upwelling systems and their cross-composition with planctomycete genomes. *International Society for Microbial Ecology* **1:** 419-435.

Woese, C.R. (1987) Bacterial evolution. *Microbiology Reviews* **51:** 221-271.

Wu, L., Thompson, D.K., Li, G., Hurt, R.A., Tiedje, J.M., and Zhou J. (2001) Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Applied and Environmental Microbiology* **67(12):** 5780-5790.

Yazynin, S., Lange, H., Mokros, T., Deyev, S., and Lemke, H. (1999) A new phagemid vector for positive selection of recombinants based on a conditionally lethal barnase gene. *FEBS Letters* **452:** 351-354.

Yeates, C., Gillings, M.R., Davison, A.D., Altavilla, N., and Veal, D. A. (1998) Methods for microbial DNA extraction from soil for PCR amplification. *Biological Procedures Online* **1(1):** 40-47.

Yin, J., Li, G., Ren, X., and Herrler, G. (2007) Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *Journal of Biotechnology* **127:** 335-347.

Young-Jun, C., Tsung-Tsan, W, and Lee, B.H. (2002) Positive selection vectors. *Critical Reviews in Biotechnology* **22(3):** 225-244.

Yun, J., and Ryu, S. (2005) Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. *Microbial Cell Factories* **4(8):** 4-8.

Zhang, T., and Fang, H.P. (2000) Digitization of DGGE (denaturing gradient gel electrophoresis) profile and cluster analysis of microbial communities. *Biotechnology Letters* **22:** 399-405.

Zhou, J., Bruns, M.A., and Tiedje, J.M. (1996) DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology* **62(2):** 316-322.