Microbial diversity and gene mining in Antarctic Dry Valley mineral soils

Jacques J Smith

A thesis submitted in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY
Department of Biotechnology,
University of the Western Cape
Bellville

Supervisor: Prof. D.A. Cowan

February 2006
ACKNOWLEDGEMENTS

My heavenly Father for His guiding hand and infinite wisdom.

My family, for their love, support and understanding through all the years of study

For Daleen, whose steadfast patience and support carried me through the trying times

Professor Don Cowan for having given me the opportunity and necessary guidance

The National Research Foundation and Ryoichi Sasakawa for providing the necessary funding

Everyone in ARCAM, for humour and assistance and a memorable experience
ABSTRACT

Microbial diversity and gene mining in Antarctic Dry Valley mineral soils

JACQUES J SMITH

PhD thesis, Department Biotechnology,
University of the Western Cape

Soil communities are regarded as among the most complex and diverse assemblages of microorganisms with estimated bacterial numbers in the order of $10^9$ cells.g$^{-1}$. Studies on extreme soils however, have reported lower cell densities, supporting the perception that the so-called extreme environments exhibit low species diversity. To assess the extent of microbial diversity within an extreme environment, the mineral soils of the Dry Valleys, Ross Dependency, Eastern Antarctica were investigated using 16S rDNA analysis. Three mineral soils designated MVG, PENP and BIS were analysed, each differing with respect to altitude, protein, lipid, water and DNA content. The mid-altitude sample, MVG, yielded the highest levels of DNA and the low altitude BIS soil contained the highest levels of protein, lipid and water. 16S clone libraries were constructed and 60 unique clones were identified and sequenced. BLASTn analysis revealed eight phylogenetic groups with Cyanobacteria, Actinobacteria and Acidobacteria representing the majority. The Cyanobacterial phylotypes were unique to the desiccated high-altitude soils of the PENP sample, suggesting a soil-borne Cyanobacterial population. 21% of the phylotypes identified were assigned as ‘uncultured’.

DNA isolated from the Antarctic mineral soils was also used to construct a metagenomic clone library consisting of 90700 clones with an average insert size of
3.5 kb, representing an estimated 3.4% of the available metagenome. Activity-based screening of the library for genes conferring lipolytic activity yielded no positive clones. It is suggested that the failure to produce positive clones might be a result of insufficient nucleotide coverage of the metagenomic DNA.

The metagenomic DNA extracted from the Dry Valley mineral soils was further analyzed using PCR. Two sets of degenerate primers based on conserved regions within lipolytic genes were used to target lipase and esterase genes. One set of primers was selected from a previous study. A second primer set was designed manually from amino acid alignments of true lipase genes from family I, sub-families I-VI. PCR analysis resulted in nine partial gene fragments varying between 240 bp and 300 bp. Bioinformatic analysis revealed that all nine partial gene fragments harboured α/β-hydrolase motifs, putatively identifying two esterases and three lipases from both bacterial and fungal origin.
DECLARATION

I declare that Microbial diversity and gene mining in Antarctic Dry Valley mineral soils is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Jacques J Smith 24 February 2006

--------------------------------------
## TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** ................................................................. ii

**ABSTRACT** .................................................................................... iii

**DECLARATION** .............................................................................. v

**LIST OF ABBREVIATIONS** .......................................................... x

**LIST OF FIGURES** ......................................................................... xiii

**LIST OF TABLES** ........................................................................... xiv

**CHAPTER ONE: LITERATURE REVIEW**

1.1 Introduction .................................................................................. 1

1.2 Antarctica .................................................................................... 4

1.2.1 Antarctic Dry Valleys ............................................................... 5

1.3 Culture-independent analysis of environmental (metagenomic) source 7

1.3.1 16S rDNA sequencing ............................................................. 7

1.3.2 DNA denaturing and reassociation kinetics .......................... 8

1.3.3 G+C analysis ........................................................................... 11

1.4 Metagenomic gene discovery ..................................................... 12

1.4.1 Enrichment ............................................................................. 13

1.4.1.1 Sample enrichment ............................................................. 13

1.4.1.2 Genome enrichment ........................................................... 14

1.4.1.3 Gene targeting ................................................................... 16

1.4.2 Total community DNA extraction ......................................... 17

1.4.3 Metagenomic DNA libraries ................................................. 18

1.5 Psychrophiles .............................................................................. 20

1.5.1 Molecular aspects associated with cold adaptation .......... 22

1.5.1.1 Lipid composition ............................................................... 22

1.5.1.2 Protein stability ................................................................. 23

1.5.2 Biotechnological applications of psychrophiles .............. 25

1.6 Lipases ....................................................................................... 27

1.6.1 Classification and taxonomy .................................................... 27

1.6.2 True lipases ............................................................................ 28

1.6.3 The GDSL family .................................................................. 29

1.6.4 Hormone sensitive lipase (HSL) family ............................ 30

1.6.5 Regulation of lipase expression ............................................ 31
1.6.6 Mechanism of secretion 32
1.6.6.1 Type I secretion 32
1.6.6.2 Type II secretion 33
1.6.6.2.1 Secretion across the inner membrane 33
1.6.6.2.2 Secretion across the outer membrane 33
1.6.6.3 The autotransporter pathway 34
1.6.7 Periplasmic folding 34
1.6.8 Psychrotrophic lipases 36
1.7 Biotechnological applications of lipases and esterases 38
1.7.1 Lipases in detergents 38
1.7.2 Biodiesel 39
1.7.3 Food industry 39
1.8 Aims of this study 41

CHAPTER TWO: MATERIALS AND METHODS

2.1 Reagents 42
2.1.1 Chemicals 42
2.1.2 Antarctic soil samples (collection and storage) 42
2.1.3 Antibiotics 42
2.1.4 Enzymes 44
2.2 Culture Media 44
2.2.1 Luria-Bertani (LB) broth 44
2.2.2 LB agar 44
2.2.3 Terrific Broth (TB) 44
2.2.4 GYT Medium 45
2.2.5 Lipase specific medium 45
2.2.6 Esterase specific medium 45
2.3 Metagenomic DNA isolation 45
2.3.1 Modified Zhou method 45
2.3.2 Miller protocol 46
2.3.3 Bead beating protocol 47
2.4 Soil analysis 47
2.4.1 Dry weight assessment and water content 47
2.4.2 Protein assessment 48
2.4.3 Lipid analysis
2.5 DNA quantification
2.6 PCR amplification using 16S rDNA primers
2.7 Agarose gel electrophoresis
2.8 GFX™ DNA purification
2.9 pMOSBlue blunt ended cloning
2.10 Preparation of electrocompetent *E. coli* cells
2.11 Preparation of chemically competent *E. coli* cells
2.12 Transformation of *E. coli* cells by:
  2.12.1 Electroporation
  2.12.2 Chemical
2.13 Colony PCR
2.14 Restriction endonuclease digestion
2.15 ARDRA analysis of 16S rDNA amplicons
2.16 Plasmid DNA extraction
2.16.1 Alkaline lysis
2.16.2 Talent kit
2.17 Nucleic acid sequencing
2.18 Preparation of metagenomic libraries
  2.18.1 Metagenomic DNA digestion
  2.18.2 A-tailing of 3’ termini of restriction digested metagenomic DNA
  2.18.3 Recovery of restriction digested metagenomic DNA from agarose gel
  2.18.4 5'-dephosphorilation of metagenomic DNA
  2.18.5 Metagenomic library construction
2.19 Amplification of metagenomic library
2.20 Activity-based screening of the metagenomic library
2.21 PCR-based screening of the metagenomic library
2.22 PCR amplification of partial lipase fragments

**CHAPTER THREE: BACTERIAL DIVERSITY IN ANTARCTIC DRY VALLEY MINERAL SOILS**

3.1 Introduction
3.2 Aims
3.3 Results
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>ATP-binding cassette</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>BAC</td>
<td>Bacterial artificial chromosome</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BrdU</td>
<td>5-bromo-2-deoxyuridine</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CaCl$_2$</td>
<td>calcium chloride</td>
</tr>
<tr>
<td>cm</td>
<td>centimeter</td>
</tr>
<tr>
<td>CP</td>
<td>cloud point</td>
</tr>
<tr>
<td>CTAB</td>
<td>cetyl-trimethyl-ammonium bromide</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>d</td>
<td>days</td>
</tr>
<tr>
<td>dNTPs</td>
<td>deoxyribonucleic-5'-triphosphate</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxyribonucleoside triphosphate</td>
</tr>
<tr>
<td>DTT</td>
<td>ditriothritol</td>
</tr>
<tr>
<td>°C</td>
<td>degrees Celsius</td>
</tr>
<tr>
<td>EtOH</td>
<td>ethanol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylene diamine tetraacetic acid</td>
</tr>
<tr>
<td>Fig.</td>
<td>figure</td>
</tr>
<tr>
<td>g</td>
<td>gram</td>
</tr>
<tr>
<td>GSP</td>
<td>general secretory pathway</td>
</tr>
<tr>
<td>GYT</td>
<td>glycerol yeast extract and tryptone</td>
</tr>
<tr>
<td>× g</td>
<td>centrifugal force</td>
</tr>
<tr>
<td>μg</td>
<td>microgram</td>
</tr>
<tr>
<td>H$_2$SO$_4$</td>
<td>Sulfuric acid</td>
</tr>
<tr>
<td>h</td>
<td>hour</td>
</tr>
<tr>
<td>HSL</td>
<td>Homoserine lactone</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl β-D-thiogalactosidase</td>
</tr>
<tr>
<td>kb</td>
<td>kilo basepairs</td>
</tr>
<tr>
<td>KCl</td>
<td>potassium chloride</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo Dalton</td>
</tr>
</tbody>
</table>
kV  kilo volt
\lambda  lambda
l  liter
LB  Luria-Bertani
Lif  lipase specific foldase
\mu l  micro liter
M  molar
MCS  multiple cloning site
\mu F  micro Farad
ml  milliliter
min  minute
MgCl$_2$  magnesium chloride
MgSO$_4$  Magnesium sulfate
mM  millimolar
MnCl$_2$  manganese chloride
m.s$^{-1}$  meters per second
NaCl  sodium chloride
NaH$_2$PO$_4$  Sodium dihydrogen orthophosphate
NaI  sodium iodide
NaOH  sodium hydroxide
ng  nano gram
ng.$\mu$l$^{-1}$  nano gram per micro liter
NH$_4$AOc  Ammonium acetate
(NH$_4$)$_2$SO$_4$  Ammonium sulphate
nt  nucleotide
\Omega  Ohm
OD  optical density
PCR  polymerase chain reaction
PEG  poly ethylene glycol
pmol  picomole
rpm  revolutions per minute
RBS  ribosome-binding site
rDNA  ribosomal deoxyribonucleic acid
s  second
SAP  Shrimp alkaline phosphatase
LIST OF FIGURES

CHAPTER 1

1.1  *E.coli* 16S ribosomal gene sequence and variable regions  9
1.2  Catalytic action of lipases  28

CHAPTER 3

3.1  Metagenomic DNA isolated from Antarctic Dry Valley mineral soils  66
3.2  PCR amplification of 16S rDNA genes from metagenomic DNA  66
3.3  Colony PCR analysis of recombinant pMOSBlue plasmids harbouring 16S rDNA genes  67
3.4  ARDRA analysis of 16S rDNA genes  67
3.5  Collectors curve of Antarctic Dry Valley mineral soil phylotypes  68
3.6  Pie chart representing Antarctic Dry Valley mineral soil phylotype distribution  77

CHAPTER 4

4.1  Efficacy of metagenomic DNA extraction using alternate protocols  85
4.2  Metagenomic DNA smear generated following *Alu*I restriction endonuclease digestion  86
4.3  *Eco*RI Restriction endonuclease digestion of recombinant PCR®-XL-TOPO® clones  87
4.4  PCR analysis of recombinant PCR®-XL-TOPO® clones  89

CHAPTER 5

5.1  Restriction endonuclease digestion of recombinant pGEM® – T easy plasmids  96
5.2  Amino acid alignment of lipases used in degenerate primer design  98
5.3  PCR amplification using primers OXF1 and ACR1  100
5.4  PCR amplification using primers LipF and LipR2  100
5.5  Restriction endonuclease digestion of recombinant pTZ57T/A  101
# LIST OF TABLES

## CHAPTER 1

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Growth temperatures of psychrophiles and psychrotrophs</td>
<td>21</td>
</tr>
<tr>
<td>1.2</td>
<td>Commercially relevant enzymes from psychrophiles</td>
<td>26</td>
</tr>
<tr>
<td>1.3</td>
<td>Cold-active lipases isolated from microorganisms</td>
<td>36</td>
</tr>
</tbody>
</table>

## CHAPTER 2

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Relevant plasmids, primers and strains used in this study</td>
<td>43</td>
</tr>
</tbody>
</table>

## CHAPTER 3

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Characteristics of the Dry Valley mineral soils</td>
<td>64</td>
</tr>
<tr>
<td>3.2</td>
<td>Unique 16S rDNA clone sequences identified in Antarctic mineral soils</td>
<td>72</td>
</tr>
<tr>
<td>3.3</td>
<td>Table of diversity indices</td>
<td>76</td>
</tr>
</tbody>
</table>

## CHAPTER 5

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Conserved motifs of true lipases from family I, subfamily I-VI</td>
<td>99</td>
</tr>
<tr>
<td>5.2</td>
<td>Bioinformatic analysis of the partial lipase gene fragments obtained during PCR analysis</td>
<td>102</td>
</tr>
<tr>
<td>5.3</td>
<td>Nucleotide and protein sequences of the partial lipolytic gene fragments</td>
<td>103</td>
</tr>
</tbody>
</table>
Chapter 1
Literature Review

1.1 Introduction

It is widely acknowledged that soil microbial diversity in the environment (also referred to as the metagenome) is severely underestimated. Initial studies to isolate and identify the bacterial diversity in soils involved culturing without realizing the bias introduced by selective culturing conditions. Identifying the bacteria present in various soil communities would enable ecologists to cluster the microorganisms into operational taxonomic units thereby defining which types are present (composition), number of types (richness) and the frequency or relative abundance of each type (structure).

The paradigm of culture-based identification of bacteria in a given environmental sample was first shifted fifteen years ago. Using reassociation kinetics on microbial community DNA, Torsvik and co-workers (1990) estimated the bacterial diversity at $10^3$ different genomes per sample, a figure 200 times larger than that obtained for culturing alone. It was concluded that standard plating techniques only accessed a minute fraction of the soil bacterium flora. The inadequacy of soil community analysis by culturing was further demonstrated by Amann et al. (1995) where they determined that microscopic visualization of an environmental sample yielded cell counts two orders of magnitude higher compared to culturing. Recently reanalysis of reassociation kinetics data of an environmental sample indicated that the diversity was underestimated a thousand fold assuming a lognormal distribution (Gans et al., 2005).
On the basis of rRNA gene sequence comparison, Woese (1987) showed that bacteria could be classified into 12 divisions based on familiar culturable organisms such as cyanobacteria, spirochetes and Gram-positive bacteria. During the past decade, these divisions have expanded to 40, primarily due to culture-independent phylogenetic surveys of microbial communities (Hugenholtz et al. 1998). The limitations associated with standard plate isolation strategies is that the specific culturing conditions required to successfully isolate the majority of bacteria do not exist (Janssen et al., 2002).

Comparing the data from culturable and non-culturable bacteria, four cosmopolitan divisions make up 90% of those bacteria readily cultured from soils (Proteobacteria, Cytophagales, Actinobacteria and low G+C Gram positive bacteria). Other cosmopolitan divisions such as Actinobacteria, Verrucomicrobia, GNS bacteria and OP11 are either absent or poorly represented, even though division OP11 constitutes a major bacterial group (Hugenholtz et al., 1998).

Resolving the total number of prokaryotic species on Earth is hindered by problems regarding the definition of species as well as practical limitations associated with counting prokaryotic species (Konstantinidis and Tiedje, 2004). Most of Earth’s prokaryotes are found in the open ocean and in soil, where the numbers of cells are believed to be in the order of $10^{29}$ to $10^{30}$, with the total number of prokaryotic cells on Earth estimated at $6 \times 10^{30}$ (Whitman et al., 1998). Problems in identifying and cataloguing all the prokaryotic species stems from the fact that the prokaryotic species concept is not comparable to the eukaryotic system (Konstantinidis and Tiedje, 2004). Of the 22 species concepts established for eukaryotes, none is applicable to prokaryotes because prokaryotes lack diagnostic morphological characteristics, proliferate asexually and exchange genetic
information in ways unusual to that of higher organisms (Brenner et al., 2000; Stackebrandt et al., 2002). In addition, the metabolic and physiological properties of prokaryotes are too poorly understood to accurately define phenotypic characteristics required for species description (Vandamme et al., 1996). Additional limitations include limited sampling of environments followed by attempts to extrapolate the data to a global scale especially when the level of endemicity is still undetermined. High degrees of endemicity would greatly expand Earth’s diversity. For example, a high level of endemism has been recorded in the deep sea (Bull and Stach, 2004). Additionally, the description of species based on ssu rRNA analysis is constraining mostly due to the 16S rDNA molecule being too conserved to resolve species (Stackebrandt and Goebel, 1994).

Another obstacle facing culture-dependent analysis is the viable but not culturable (VBNC) state of some bacteria. The VBNC state is believed to be either a process of self-preservation whereby the bacterium ensures survival during environmental stress, or an end-of-life-cycle process (McDougald et al., 1998). The argument is that if the former is true, then resuscitation of the bacteria in the dormant VBNC state should occur when conditions become favourable. Resuscitation has, however, not been conclusively proven because it is unclear whether these ostensibly ‘uncultured’ cells have permanently lost culturability (are dead), are killed by or are unable to grow on standard isolation media or, are in a dormant state from which they might be recovered only if the correct methodology is in place (Kell et al., 2004; Barer et al., 1998). Nevertheless, the presence of bacteria in an environmental sample and the inability to culture all of them, adds to the bias of culture-dependent studies (McDougald et al., 1998).
This chapter aims to highlight and discuss the current knowledge pertaining to the biodiversity and environmental factors of the Antarctic Dry Valley mineral soils located in the Ross Dependency, Eastern Antarctica. The review discusses the ongoing technologies employed to assess global biodiversity within diverse environments, which include culture-based analysis and constraints, 16S rDNA analysis, and reassociation kinetics. The discussion concludes with approaches to exploit biodiversity within communities using metagenomic gene discovery tools such as metagenomic enrichment strategies, gene targeting and metagenomic libraries.

As a major section of this study centers on the diversity of cold active lipolytic enzymes (EC 3.1.1.3), the final section of this review discusses the general characteristics of lipolytic enzymes followed by the specific adaptations of these cold-active enzymes and their biotechnological applications.

1.2 Antarctica

Antarctica is considered to be the last pristine continent. It has a land area of approximately $4 \times 10^6$ km$^2$ but expands its surface area 5.5 times to $22 \times 10^6$ km$^2$ during the austral winter due to ice formation (Friedmann, 1993). Of the 12.38 million km$^2$ of rock, 11.97 million km$^2$ is covered in an ice sheet 2 km thick giving it the highest mean elevation (2.3 km) of any continent. Extreme temperature fluctuations varying between $-50^\circ$C to $20^\circ$C and very high gravity-driven winds, referred to as katabatic winds, low atmospheric humidity and low water potential are some of the factors which render Antarctica one of the least hospitable environments on earth.
1.2.1 Antarctic Dry Valleys

Less than 2% of the Antarctic continent is permanently ice free (Cowan and Ah Tow, 2004). The largest ice-free area is the McMurdo Dry Valleys, situated in the Eastern region of Antarctica (Doran et al., 2002). Other ice-free regions are found on the Antarctic peninsula, Vestfold Hills, Bunger Hills, some coastal fringes and various locations in the Transantarctic mountains. The Dry Valleys, collectively called the Ross Desert, are the coldest and driest deserts on Earth and are subjected to harsh environmental conditions (Wilson, 1970). The mean annual air temperature is -20°C, with temperatures fluctuating around 0°C during summer months and dropping as low as -55°C during winter. Ground surface temperatures during periods of direct sunlight average 15°C but the air temperature seldom rises above 0°C (Wynn-Williams, 1990). The upper mineral soil surface has a low water content (0.5%-2% wt) due to the desiccating atmosphere. During winter months the high katabatic winds reduce humidity to <10% which result in long periods of desiccation (Horowitz et al., 1972). Although precipitation averages 15 g cm\(^{-1}\) year\(^{-1}\) in the form of snow, very little moisture reaches the valley floor because of the high sublimation rate and low humidity. In addition, the mineral soils are burdened by mineral salts such as sodium, calcium, magnesium, chloride, sulfate and nitrate. The accumulation of these salts is due to upward translocation from the substratum by capillary action and seaspray blown inland by onshore winds (Vishniac, 1993).

The Dry Valley mineral soils are very nutrient poor as there is no plant life to liberate organic matter. The presence of organic material would facilitate water retention and aid to curb the high sublimation rate (Smith and Tearle, 1985). The aridity and low water
content of the Ross desert soils suggest low microbial abundance, as water is the limiting factor in bacterial growth (Horowitz et al., 1972). Although temperate mineral soils contain between 5% and at most 15% organic carbon, the Ross Desert soils contain significantly less (≤0.05%) (Cameron, 1970). Sources of organic carbon in the Ross Deserts are presumed to be exogenous as very few chemoautotrophs have been documented. Cyanobacteria and algae have also been isolated, but photosynthetic productivity appears to be confined to cryptoendolithic habitats in close proximity to temporary or permanent water (Friedmann, 1993). For this reason it is presumed that organic matter influx in the Ross Desert is airborne-driven from aquatic habitats. The aerial dispersal of cyanobacterial mats from ice-covered lakes in Taylor Valley has been estimated at $2.93 \times 10^4$ kg.y$^{-1}$ (Kappen and Straka, 1988).

Antarctic soils are largely aerobic and the occurrence of anaerobic bacteria is considered rare (Line 1988). Early culture-based analysis by Baker and Smith (1972) indicated that seventy-one percent of the bacteria present in Antarctic soil were related to coryneforms within the genera *Arthrobacter, Brevibacterium, Cellulomonas, Corynebacterium* and *Kurthia*. Some *Bacillus* and *Micrococcus* species as well as *Pseudomonas, Flavobacterium* and other Gram-negative aerobic rods such as *Alcaligenes* and *Agrobacterium* were also identified (Baker and Smith, 1972). In the Ross Desert the highest numbers of culturable isolates are present in the permafrost layer and towards the surface. Chromogenic bacteria are most prevalent at the soil surface whereas non-pigmented bacteria are mainly located below the soil surface (Cameron et al., 1972). Bacteria isolated from the McMurdo Dry Valleys most commonly belong to the genera: *Achromobacter, Arthrobacter, Bacillus, Corynebacterium, Flavobacterium, Micrococcus, Planococcus,*
Pseudomonas, Streptomyces, and Nocardia of which Coryneforms represent the most abundant bacteria present (Cowan and Ah Tow, 2004).

1.3 Culture independent analysis of environmental (metagenomic) sources

1.3.1 16S rDNA sequencing

Initial attempts at determining microbial diversity relied on the direct extraction and sequencing of 5S rRNA molecules from environmental samples (Stahl et al., 1985). However, the mere 120 bp of the 5S rRNA molecule provided limited information due to the paucity of independently varying nucleotides, limiting its usefulness to simple ecosystems. Subsequently, an approach using the 16S rRNA gene for phylotype identification was suggested (Pace et al., 1986; Olsen et al., 1986). The 16S rRNA gene has an average length of 1500 nucleotides, yields substantially more information than the 5S rRNA gene, and contains highly conserved sequence domains interspersed with nine variable regions (Fig 1.1) (Van de Peer et al., 1996). This allowed for more detailed analysis and identification of signature sequences over a wide taxonomic range.

Advantages in using rRNA genes for analysis include: 1) They occur in high copy numbers (from 1 to 10 per cell), 2) sequence data can be obtained without prior cultivation and 3) 16S rDNA sequences have been determined for a large number of bacteria (Amann et al., 1995).

The most widely adopted method for obtaining 16S rDNA sequences is through the use of PCR. Total community DNA is isolated from the environment and the 16S rRNA genes amplified using universal primers designed to target the conserved regions within the
rRNA genes (Fig 1.1). The PCR amplicons are cloned and recombinant clones are sequenced and compared to one other and sequences within the 16S rDNA database (Altschul et al., 1997). Although PCR analysis of community DNA bypasses culturing and is a rapid process for generating vast amounts of data, several factors have been identified which can skew phylogenetic analysis. Rainey et al. (1994) suggested that high percentage G+C DNA templates are discriminated against due to low efficiency of strand separation during denaturation. Target sequences that are in higher abundance are preferentially amplified and cloning sticky-end products may produce better results than blunt-end products (Rainey et al., 1994). Amplification of rRNA genes using standard PCR conditions have been found to exclude other important environmental taxa due to primers designed to target only known bacteria (Reysenbach et al., 1992) and Suzuki and Giovannoni (1996) indicated that the amount and origin of background DNA in a sample could affect results by out competing the 16S rDNA target sequences. The formation of chimeras has also been reported (Maidak et al., 2000). Although, computer programs have been developed to eliminate chimeric sequences, chimeras with greater than 85% homology are not easily identified (Theron and Cloete, 2000).

1.3.2 DNA denaturation and reassociation kinetics

DNA denaturation and reassociation kinetics have been used to determine the heterogeneity of a given environmental sample and so provide a broad-scale analysis of community structure (Torsvik et al., 1990; Øvreås et al., 2003). It does not, however, provide information relating to the phylogenetic origin of the microorganisms present. As the isolated community DNA is denatured, the optical density of the DNA is recorded over time and plotted against temperature (usually between 65°C and 95°C) and
displayed as a sigmoidal curve. Using the information gathered upon DNA denaturation the first derivative of the melting curve is calculated and the melting profiles converted to mol %G+C as described by Mandel et al. (1970). The first derivative of the melting curve provides information concerning the %GC distribution within the community. A less complex sample for instance, might display a single melting point which results in a steep slope on the sigmoidal curve. A more complex community containing a variety of genomes will have different melting points and produce a shallower slope (Øvreås et al., 2003). Genomic DNA of known mol %G+C is used as standard in both DNA denaturation and reassociation studies.

Following denaturation, the DNA is allowed to reassociate. The principle of reassociation kinetics is based on the rate of DNA hybridization within the sample, where the rate of hybridization is proportional to the concentration of complementary DNA sequences and inversely proportional to the length of different sequences (Theron and Cloete, 2000). DNA reassociation is measured over time and the fraction of reassociated DNA (C/C₀) is expressed as a function of C₀t, where C₀ is the initial molar concentration of single-
stranded DNA nucleotides. The reaction rate constant, $k$, can be expressed as $1/C_0t_{1/2}$, where $t_{1/2}$ denotes the time in seconds required for 50% reassociation. The genomic DNA standards are then used to calculate the size of the community genome relative to that of the standard (Britten et al., 1974)

### 1.3.3 G+C analysis

When only a coarse level of resolution of a community is required, G+C analysis is typically used. The technique is based on the fact that prokaryotic genomic DNA varies between 24% and 76% with respect to G+C vs. A+T and that particular taxonomic groups do not vary more that 5% in G+C content (Vandamme et al., 1996). G+C analysis is considered to be a coarse analysis because several taxonomic groups may have a similar G+C range. G+C range does, however, correlate with certain physiological attributes. Microorganisms with a high G+C range (60-75%) are generally obligate aerobes with an oxidative metabolism, while organisms with a fermentative metabolism largely have low G+C content (Santo Domingo et al., 1998). The method is based on the principle that bis-benzimidazole binds to adenine and thymidine thereby altering the buoyant density of the DNA molecule in proportion to its A+T content (Holben and Harris, 1995). A gradient of DNA fragments of different G+C concentrations is subsequently established using equilibrium density-gradient (CsCl) centrifugation. The G+C content of each fraction is established by using a standard curve relating G+C content to density. G+C analysis can be applied to all the extracted DNA in the environmental sample and it is not subject to the biases of PCR-based methods. The method is quantitative and provides a means for detecting sparse members of a community since it separates low biomass fractions from dominant fractions (Tiedje et al., 1999).
1.4 Metagenomic gene discovery

Strategies to find and isolate novel enzymes from the environment involve enrichment, isolation and screening of a wide variety of organisms in pure culture for the desired activity. To increase the probability of obtaining the desired gene a range of culturing and enrichment conditions need to be set up. This includes incubation at various temperatures, levels of pH, and supplementing media with different carbon sources under both aerobic and anaerobic conditions (Wilkinson, 2002). Upon identification of the microorganism possessing the desired activity, the gene is then isolated. Although this method is a common route to discovering new enzymes a substantial fraction of the gene diversity could be lost due to inefficient culturing conditions and VNBC organisms (McDougald et al., 1998; Janssen et al., 2002).

To overcome culturing constraints, several DNA-based molecular methods have been developed, primarily focusing on 16S RNA genes. Although 16S RNA analysis provides information regarding the species present, little to no information regarding the functionality or phylogenetic relationship of the bacteria can be determined (Streit and Schmitz, 2004).

Metagenomic gene discovery circumvents culture-based enzyme screening techniques by targeting DNA directly in the environment as outlined below. The generation of large insert libraries allows cloning of large (40–150 kb) DNA fragments from environmental samples, theoretically isolating genes from any origin, sequence and function (Streit and Schmitz, 2004). Direct genomic cloning not only provide access to biotechnologically useful genes, but also to operons or genes encoding pathways that direct the synthesis
of complex molecules such as antibiotics. Furthermore, sequence space flanking a particular gene of interest can easily be obtained and/or the phylogenetic origin of the functional gene can be determined (Streit and Schmitz, 2004). Examples of metagenomic gene discovery include the cloning of DNA from uncultured soil microorganisms and the identification of five clones conferring 4-hydroxybutyrate dehydrogenase (Henne et al., 1999) and three clones conferring lipolytic activity (Henne et al., 2000). Rondon et al. (2000) used a BAC vector to isolate genes conferring DNase, lipase and amylase activity from soil. Eleven clones exhibiting lipolytic activity towards tributyrin were identified following the construction of a metagenomic library from pond water (Ranjan et al., 2005) and Majerník and co-workers (2001) identified a single gene conferring Na⁺/H⁺ antiporter activity by screening metagenomic libraries using a Na⁺/H⁺ antiporter deficient E. coli host.

1.4.1 Enrichment

1.4.1.1 Sample enrichment

Sandaa and co-workers (1999) determined, using reassociation kinetics, that heavy metal contaminated soils contained approximately 2000 genomes.g⁻¹ soil whereas non-contaminated soils could contain as much as 16 000 genomes.g⁻¹. Thus, a gene of interest represents a very small fraction of the nucleic acid pool due to the high degree of genetic diversity present in soils. This diminishes the likelihood of obtaining that gene upon expression screening of metagenomic libraries. Pre-enrichment of the sample enhances the probability of detecting a desired gene. One such strategy is culture enrichment, which is aimed at the culturable fraction of microorganisms in an environment. Selection pressure is based on nutritional, physical or chemical criteria.
Two separate studies used selective enrichment strategies to isolate genes conferring alcohol oxidoreductase (Knietsch et al., 2003) or dehydratase activity (Knietsch et al., 2003). Using indicator agar-plate assays, these investigators identified sixteen different clones utilizing 1,2-propanediol and two clones possessing the glycerol or diol dehydratase genes, respectively.

Other forms of enrichment include specific whole-cell enrichment where, for example, size selective filtration was used during the Saragasso Sea genome sequencing project to omit eukaryotic cells (Venter et al., 2004). Similarly, enrichment for Buchnera aphidicola was achieved by differential centrifugation, followed by whole genome sequencing (Schloss and Handelsman, 2003).

1.4.1.2 Genome enrichment

Genomic enrichment is also aimed at the active components of a microbial community. During a process called stable-isotope probing (SIP) the genomic DNA or RNA of specific members of a community is selectively labeled (Boschker and Middelburg, 2002). Isotopes of $^{13}$C, $^{18}$O and $^{15}$N are used to label substrates such as phenol, methanol, ammonia and methane. The isotopes are incorporated in the DNA and RNA of the actively growing microorganisms and the labeled nucleic acids are separated using density gradient centrifugation. SIP was first used to identify $^{13}$C-enriched phospholipid fatty acids signature profiles in experiments aiming to identify microbial populations responsible for acetate oxidation in sediments (Boschker et al., 1998). A previously unknown group of phenol degraders, Thauera, was identified using $^{13}$C-phenol in an anaerobic bioreactor by enriching populations with RNA-SIP (Manefield et al., 2002). Pelz
et al. (2001) also employed SIP to identify sulfate-reducing bacteria in hydrocarbon-contaminated aquifers using $^{13}$C-toluene as substrate. SIP substrates are expensive and limited in availability and alternatives such as 5-bromo-2-deoxyuridine (BrdU) can be used. BrdU is a structural analogue of thymidine and the uptake of $[^3]$Hthymidine has routinely been used for measuring the in situ growth of bacteria in various environments (Smalla, 2004). BrdU incubation was used to detect metabolically active bacteria in lake water (Borneman, 1999). BrdU labelled DNA or RNA can also be separated using density gradient centrifugation or immunocapture. Although these methods prove effective, loss of specific enrichment may occur due to cross feeding and recycling of the label within the community (Smalla, 2004).

Another approach, suppressive subtractive hybridisation (SSH), identifies genetic differences between microorganisms. SSH can be used to successfully identify specific genes in closely related bacteria such as genes responsible for pathogenesis or xenobiotic degradation (Cowan et al., 2005). Two separate restriction digested fractions are generated termed ‘driver’ and ‘tester’, each ligated to unique adaptors. The ‘driver’ is present in excess (usually 100 fold) and is used to trap and eliminate the background in the ‘tester’ by associating with identical DNA sequences during hybridisation. Unique heteroduplex DNA association between ‘driver’ and ‘tester’ represents the DNA sequences of interest (Felske, 2002). Galbraith et al. (2004) demonstrated that SSH could also be used to identify differences between complex DNA samples.
### 1.4.1.3 Gene targeting

Targeting genes with specific metabolic or biodegradative capabilities using PCR has been extensively used. Sheu *et al.* (2000) detected polyhydroxyalkanoate synthases from nineteen PHA-positive strains by designing degenerate primers. Gene-specific PCR can also be used in environmental samples enabling the detection of specific genes. Bell *et al.* (2002) designed degenerate primers using conserved amino acid regions within lipase genes. Using these degenerate primers they obtaining partial lipase gene fragments from a metagenomic soil sample, and obtained the full-length lipase genes using genome-walking PCR (Morris *et al.*, 1995). Although proven to work well (Hallin and Lindgren 1999; Henckel *et al.*, 2000), this approach is limiting with respect to biocatalyst discovery. Firstly, primer design is limited to available sequences. Functionally similar genes from distant families or genes resulting from convergent evolution are unlikely to be detected using a single gene-family-specific set of primers. Very distant gene families tend to require separate family-specific primers as primers with excessive degeneracy (>256 fold) seldom yield results. Gene-specific PCR only yields a partial gene fragment, requiring additional steps to obtain the up- and down-stream flanking regions. In such instances, the partial gene fragment can be used as a probe to identify possible full-length genes in a metagenomic library or enzyme restricted metagenomic DNA, which can be excised and cloned. More direct approaches for obtaining up- or down-stream sequences are PCR-based and include strategies such as genome walking PCR (Kilstrup and Kristiansen, 2000), Thermal Asymmetric Interlaced PCR (TAIL) (Liu and Whittier, 1995), panhandle PCR (Megonigal *et al.*, 2000) and inverse PCR (Hartl and Ochman 1994).
1.4.2 Total community DNA extraction

The first step toward gene targeting is nucleic acid extraction. Although numerous nucleic acid extraction methods exist, extraction is achieved using two types: cell extraction and direct lysis (Roose-Amsaleg et al., 2001). During cell extraction, whole cells are separated from the sample prior to lysis and DNA retrieval. It is a time consuming process and generally includes laborious methods such as differential centrifugation and sucrose gradients to obtain whole cells (Steffan and Atlas, 1988). Direct lysis methods do not require cell separation. The DNA yield is much higher and more representative of the microbial community compared to the yield obtained using cell extraction (Steffan et al., 1988). This is largely attributed to the fact that a greater number of the microorganisms are subjected to lysis, notably those embedded in soil particles (Roose-Amsaleg et al., 2001). In addition, non-bacterial DNA, such as from fungal or plant material is also isolated. A major drawback of direct lysis is that other organic components such as humic and fulvic acids are co-extracted. These contaminants inhibit downstream reactions, especially PCR, and require additional purification such as gel extraction or PVPP treatment (Steffan et al., 1988). The mainstream methods of choice in direct lysis are the Miller (Miller et al., 1999) and Zhou (Stach et al., 2001) protocols. The Miller protocol is a mechanical process which involves bead beating and tends to access a higher degree of diversity because Gram-positive and other bacteria such as Micromonospora sp. are insensitive to chemical lysis treatment (Niemi et al., 2001). However, chemical lysis has shown to be more efficient in extracting high molecular weight DNA (Stach et al., 2001). Total DNA extracted directly from environmental samples does not contain equal distribution of the communities’ genomes, which generally leads to bias in downstream applications such as PCR (Suzuki and Giovannoni, 1996).
1.4.3 Metagenomic DNA libraries

The same principles for single genome DNA library construction (generation of DNA fragments via shearing or enzyme digestion, cloning of DNA fragments into a suitable vector and screening for genes of interest) apply for metagenomic library construction. The first study in which metagenomic cloning was reported involved the cloning of a 40 kb genome fragment from a planktonic marine archaeon to characterize uncultivated prokaryotes (Stein et al., 1996). However, the intended downstream application(s) dictates the vector and host systems used. Protocols that yield highly fragmented DNA (0.5-5.0 kb) severely limits applications of the isolated DNA, especially when downstream processing involves restriction digestion to generate sticky ends, which will result in significant loss of the total DNA complement. To accommodate randomly fragmented DNA, cloning systems designed for blunt-end or T/A ligation could be used. Where restriction digestion is a requisite, high molecular DNA is paramount, for example partial digestion using Sau3AI and subsequent ligation to BamHI prepared vector. The latter approach was successfully applied in screening and detecting single gene products permitting between $10^4$ and $10^6$ clones to be screened (Henne et al., 1999). Other cloning systems such as cosmids, fosmids and bacterial artificial chromosomes (BACs) are widely used in library construction and accommodate large inserts (25-45 kb for cosmids and fosmids, >40 kb for BACs). Large inserts enable capture of more sequence space and allow for detection of whole metabolic pathways, such as the biosynthetic gene cluster for the synthesis of the antibiotic violacin (Brady et al., 2001).

Construction and screening of single or metagenomic DNA libraries typically relies on *E. coli* as expression host. This host is favoured because it is the best characterized in terms
of biochemistry and genetics, having served as a biological model for cellular processes such as DNA replication, transcription, metabolic pathways and signal transduction (Thompson and Zhou, 2004). However, the limitations associated with *E. coli* as an expression host are underlined by the low number of positive clones obtained during a single round of screening, usually less than 0.01% (Cowan *et al*., 2005). The best chance for recovering heterologously expressed genes in *E. coli* is from native promoters or read-through transcription from vector based promoters. The probability, however, of expressing a heterologous gene due to a translational fusion is extremely low and would statistically involve the screening of >10^7 clones (Gabor *et al*., 2004). This suggests that the discovery of target genes from a complex metagenomic sample without sample enrichment is technically challenging.

*E. coli* transcriptional machinery is known to be relatively promiscuous in recognizing foreign expression signals, but with a strong bias towards Firmicutes genes (Gabor *et al*., 2004). Other important factors for efficient heterologous gene expression include the efficiency with which the –35 and –10 DNA elements of the promoter correlate with the *E. coli* consensus sequences TTGACA and TATAAT, respectively (Gross *et al*., 1992). If heterologous expression relies on transcription from a vector-encoded promoter, the foreign gene must be inserted in the correct orientation with respect to the promoter. Conversely, if expression is dependent on a native promoter it must be recognized by the host RNA polymerase holoenzyme (Old and Primrose, 1994). Factors governing translation such as initiation codons and ribosome-binding site (RBS) are also important in successful expression. However, given that the characteristics of the genes cloned during metagenomic library construction are unknown, translational factors could
contribute to lack of expression. The introduction and expression of foreign DNA in a host organism often change the metabolism of the organism in ways that may impair normal cellular function due to increased metabolic load. This typically occurs upon over expression of the foreign DNA due to high copy number of the plasmid or transcription from a strong plasmid-based promoter (Glick, 1995). Such a high level of expression of a foreign protein can initiate a cellular stress response, including increased synthesis of cellular proteases, so that the recombinant protein is rapidly degraded (Glick, 1995). Careful choice of expression plasmid and promoter would circumvent such a problem and enhance the probability of obtaining positive clones.

1.5 Psychrophiles

Psychrophilic organisms are widely distributed and inhabit both terrestrial and aquatic environments. Such habitats exist in polar and alpine regions, much of the world's oceans, subterranean systems such as caves, the upper atmosphere, and man-made systems (Cavicchioli et al., 2002). Due to huge debate surrounding the distinction between psychrophilic and psychrotolerant, one of the earliest criteria used to define psychrophilic microorganisms is the ability to form visible colonies on solid media within 1 to 2 weeks at 0°C. Although this is no longer a definitive characteristic it illustrates that psychrophiles are defined by their minimum growth temperature and not by their optimum or maximum growth temperature as with mesophiles and thermophiles. Currently, the definition of psychrophile by Morita (1975) is widely accepted (Helmke and Weyland, 2004) (Table 1.1). Determining the minimum growth temperatures of psychrophiles and psychrotrophs is difficult because of their very slow growth rates. Nutrient availability is also important as the media composition can alter the minimum
growth temperature. The lowest temperature of bacterial growth is thought to be -12°C (Russell, 1992) for at lower temperatures the intracellular ice formation in the cytoplasm and an increase in solute concentrations would prevent growth.

Table 1.1 Minimum, optimal and maximum growth temperatures of psychrophilic and psychrotrophic microorganisms

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Optimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychrophilic</td>
<td>&lt;0°C</td>
<td>&lt;15°C</td>
<td>&lt;20°C</td>
</tr>
<tr>
<td>Psychrotrophic</td>
<td>3-5°C</td>
<td>&gt;15°C</td>
<td>&gt;20°C</td>
</tr>
</tbody>
</table>

The ability to grow at low temperatures is widespread amongst microbial flora, which include representatives from the eubacteria, yeasts, fungi, algae and Archaea. Like the ability to perform photosynthesis, it seems probable that the ability to grow at low temperatures has evolved independently in several microorganisms (Russell, 1992). Additionally, species within a single genus are known to be psychrophilic, mesophilic or thermophilic, for example the genera *Clostridium* and *Bacillus* amongst the eubacteria and the cyanobacterial genus *Phormidium* (Russell, 1992).

Psychrophilic and psychrotolerant microorganisms are not only widespread in the environment. Many well-known food-spoilage and food-poisoning bacteria are known to thrive in cold environments. Food-spoilage bacteria include *Brochothrix thermosphacta* and *P. fragi*. *B. thermosphacta* is favoured in packed meat and meat products due to their tolerance to low pH, curing salts and chill-storage. Listeriosis is caused by *Listeria monocytogenes*, which grow at temperature around 10°C, particularly in cold raw milk,
soft cheese and processed poultry (Glass and Doyle, 1989). *Yersinia enterocolitica* is capable of growth and toxin production at temperatures as low as 4°C and causes yersiniosis, which manifests as gastroenteritis in adults and peritonitis, ileitis, pseudoappendicitis in children following the ingestion of contaminated milk (Wannet *et al*., 2001).

### 1.5.1 Molecular aspects associated with cold adaptation

#### 1.5.1.1 Lipid composition

Temperature has a major influence on both the fluidity and phase behaviour of membrane lipids. As microorganisms cannot insulate themselves, they are required to regulate their lipid composition to maintain the activity of their membrane-associated proteins. The primary thermal responses of membrane lipids are alterations in fatty acyl composition, which in the event of a growth temperature drop result in increased unsaturation, reduction in average chain length, increase in methyl branching and an increase in the ratio of *anteiso* branching relative to *iso* branching (Russell and Hamamoto, 1998). The most common response to a decrease in temperature is the accumulation of unsaturated fatty acids in membrane lipids (Nedwell, 1999). The level of unsaturation is controlled by cold-shock fatty acid desaturases thereby regulating the homeostasis of the membrane fluidity. Desaturases are encoded by four genes, *desA-D*, are membrane bound in bacteria and algae and target intact acyl lipids as substrates (Foot *et al*., 1983; Los, 2004). The introduction of double bonds into fatty acids by desaturase action allows for rapid response upon a decrease in membrane fluidity allowing adaptation to a sudden drop in temperature. Cell membranes are essentially anisotropic solutions of phospholipids and proteins in a fluid phase and it is only within...
the fluid phase that they are biologically functional. As temperature decreases, membranes become increasingly viscous which leads to a decrease in membrane fluidity (Sinensky, 1974) and at a certain temperature, depending on the physiology of the microorganism, the membrane phase will shift to solid and biological function lost. The alterations in membrane lipids thus allow maintenance of biological function, a feature called homeoviscous adaptation (Sinensky, 1974).

1.5.1.2 Protein stability
The stability and flexibility required for enzymes to adapt to various temperature ranges are largely controlled by amino acid substitutions. Although past research mainly focused on mesophilic and thermophilic enzymes, studies on psychrophilic enzymes revealed that the same general changes in amino acid composition apply. Obtaining the flexibility required for catalysis in cold environments is achieved by reversing the stabilizing factors required for rigidity in thermophilic enzymes. These include fewer ionic interactions and intramolecular hydrogen bonds (Cavicchioli et al., 2002). To aid in destabilization cold-active enzymes contain more polar and/or charged side groups, which promote interaction with solvents (Fields, 2001). Additional strategies to enhance entropy in psychrophilic enzymes involve a reduction in proline residues while increasing the number of glycines (Haney et al., 1999b). The increase in flexibility and lowering of stability in cold-active enzymes results in increased activity at low temperatures and low thermostability at elevated temperatures (Gerike et al., 2001). Although the alterations listed above are used by enzymes to adapt to all ranges of temperature, one potential region for change, the active site, however, remains conserved across broad taxonomic groups and all temperature ranges. It appears that substrate binding is so specific that
any changes might lead to significant reduction or outright loss in catalytic activity, inferring that the altered or substituted residues must occur some distance from the active site.

Homology modeling of *Arthrobacter* species’ citrate synthase, a well studied psychrophilic enzyme, exhibits an additional surface loop compared to its mesophilic counterpart from *M. smegmatis*. The loop consists of nine amino acids of which six are charged. On another surface loop, present in both enzymes, the *Arthrobacter* enzyme has nine substitutions of neutral with charged amino acids to further enhance surface interaction (Gerike *et al.*, 2001). Except for the latter changes, no additional charged amino acids are present in the psychrophilic enzyme. Other expected changes in thermostability such as increased number of isoleucine residues or changes in the arginine:lysine ratio were not observed. Although the proline distribution differed between the two enzymes, nothing could be deduced based on homology modeling. Crystal structure analysis revealed that the overall structures of the two enzymes were similar. The active site of the psychrophilic enzyme had however, a larger opening due to shortening of a loop on one side and the substitution of an arginine with an alanine. As was expected, the structure and integrity of the active sites were retained. However, the surface surrounding the entrance was more negatively charged, aiding in forcing substrates into the pocket. Overall, the surface of the cold-active enzyme was more hydrophobic, causing destabilization by means of entropy-driven ordering of the surrounding solvent water molecules thus contributing to low temperature activity (Russell *et al.*, 1998; Russell, 2000; Gerike *et al.*, 2001).
1.5.2 Biotechnological applications of psychrophiles

Table 1.2 lists some of the biotechnologically relevant enzymes from psychrophilic prokaryotes. Only within the past few years has it been recognized that psychrophilic microorganisms and their enzymes provide a reservoir for biotechnological exploitation (Gounot, 1991; Margesin and Schinner, 1994). Possible applications include:

- Enzymes such as lipases, cellulases and proteinases used in detergents
- Enzymes as flavour-modifying agents
- Biosensors for environmental applications
- Environmental bioremediation
- Food processing, such as cheese manufacture and meat tenderizing

Advantages of using psychrophiles and/or their enzymes in biotechnological applications include:

- Heat lability of psychrophilic enzymes allows for rapid and economic termination of processes using moderate heat treatment
- Less expensive processing due to elimination of expensive heating and/or cooling processes
Table 1.2 Examples of commercially relevant enzymes from psychrophiles. Data obtained from Cavicchioli et al. (2002), Gerike et al. (2001) and Bhat (2000)

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Organism</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triose phosphate isomerase</td>
<td><em>Vibrio marinus</em></td>
<td>Biotransformation</td>
</tr>
<tr>
<td>Lipase</td>
<td><em>Aspergillus nidulans</em></td>
<td>Food, detergents, cosmetics</td>
</tr>
<tr>
<td>α-Amylase</td>
<td><em>Alteromonas haloplanktus</em></td>
<td>Pulp bleaching, starch hydrolysis</td>
</tr>
<tr>
<td>Pectinase</td>
<td><em>Sclerotinia borealis</em></td>
<td>Cheese ripening, wine industry</td>
</tr>
<tr>
<td>Alkaline phosphatase</td>
<td><em>Vibrio</em> sp. I5</td>
<td>Molecular biology</td>
</tr>
<tr>
<td>Cellulase</td>
<td><em>Fibrobacter succinogenes</em></td>
<td>Animal feed, textiles</td>
</tr>
<tr>
<td>Xylanase</td>
<td><em>Cryptococcus adeliae</em></td>
<td>Wine industry, fermentation</td>
</tr>
<tr>
<td>Nitrile hydolase</td>
<td><em>Rhodococcus</em> sp. N-774</td>
<td>Low-temperature acrylamide synthesis</td>
</tr>
<tr>
<td>DNA ligase</td>
<td><em>Pseudoalteromonas haloplanktis</em></td>
<td>Molecular biology</td>
</tr>
</tbody>
</table>

Although a wide range of potential applications and benefits can be gained from cold-active enzymes, very few of the benefits have been exploited. This is primarily due to a lack of understanding of the cellular behavior and molecular structure of psychrophiles compared with thermophilic counterparts (Russell and Hamamoto, 1998).

One aspect of psychrophilic proteins and their biotechnological potential which have been studied is that of ice-nucleation. Ice-nucleating proteins produced by *P. fluorescens*, *X. campenstris* and *Erwinia herbicola* have substantial biotechnological potential. These plant pathogens induce frost damage in crops by triggering ice formation on leaf and flower surfaces (Li and Lee, 1995). Ice-nucleating proteins are encoded by *ina* genes, although the mechanism of ice formation remains unclear (Wolber and Warren, 1989; Kajava and Lindow, 1993). Possible biotechnological applications for ice-nucleating
proteins include ice-cream manufacture, synthetic snow, freeze-texturing, freeze-drying and concentrating (Russell and Hamamoto, 1998). Using natural or genetically engineered bacteria deficient in ice formation as frost protectants for sensitive plants to out-compete natural ice-nucleating pathovars is another application which has shown potential (Gurian-Sherman and Lindow, 1993).

1.6 Lipases

1.6.1 Classification and taxonomy

Until recently, enzymes were classified as true lipases (EC 3.1.1.3) based on two criteria. Firstly, activity increased in the presence of lipid:water interfaces, a phenomenon termed ‘interfacial activation’ (Brzozowski et al., 1991) and secondly, lipases exhibit a ‘lid’ domain, which is a surface loop covering the active site which was displaced upon contact with the interface (Derewenda et al., 1994). These criteria became untenable due to a number of exceptions where enzymes having a ‘lid’ did not exhibit interfacial activation (Verger, 1997). For this reason, lipases are generally defined as carboxylesterases capable of catalyzing both the hydrolysis and synthesis of long-chain acylglycerols (Fig. 1.2). Glycerol esters with an acyl chain of ≥10 carbon atoms are regarded as lipase substrates and those of ≤10 carbon atoms as esterase substrates, with trioleoglycerol being the standard lipase substrate under laboratory conditions (Jaeger et al., 1999). The presence of esterases is normally indicated by the hydrolysis of glycerolesters such as tributyrylglycerol (tributyrin) (Jensen, 1983). It should be noted, however, that most lipases are also capable of hydrolyzing esterase substrates.
1.6.2 True lipases

Lipolytic enzymes form part of the α/β-hydrolase superfamily (Ollis et al., 1992) and are distinguished by a catalytic triad Ser-Asp-His. The serine residue usually occurs in a conserved pentapeptide Gly-Xaa-Ser-Xaa-Gly. Lipolytic enzymes are classified into 8 families broadly based on size, function and manner of secretion. Family I is further categorized into six subfamilies of which subfamilies 1, 2 and 3 constitute true lipases, and subfamilies 4, 5 and 6 lipases and phospholipases from Gram-positive bacteria. *P. aeruginosa* is the prototypical lipase of subfamily I.1 and has a molecular mass ranging between 30-32 kDa. Other lipases belonging to subfamily I.1 that display high amino acid similarity to *P. aeruginosa* are produced by *V. cholerae*, *A. calcoaceticus*, *P. wisconsinensis* and *P. vulgaris*. Subfamily I.2 lipases have slightly higher molecular weights (33 kDa) owing to an amino acid insertion forming an anti-parallel double β-strand at the surface of the molecule (Nobel et al., 1993). Lipases of both family I.1 and I.2 require the co-expression of a chaperone protein termed lipase-specific foldase (Lif) to correctly fold into an active state (Rosenau et al., 2004). Exceptions are lipases from...
Another distinguishing feature of subfamilies I.1 and I.2 is the two aspartic residues involved in a Ca\(^{2+}\)-binding site are conserved. Two cysteine residues, which form a disulphide bridge, are conserved in the majority of lipase sequences. The residues involved in Ca\(^{2+}\)-binding and disulphide bridge formation are located near the catalytic His and Asp residues and are believed to be important in active site stabilization (Kim et al., 1997). Secretion of subfamilies I.1 and I.2 lipases is mediated through the general secretory pathway (Gsp). Subfamily I.3 contains enzymes from two species, \textit{P. fluorescens} and \textit{S. marcescens}, with molecular weights of 50 kDa and 65 kDa, respectively. Not only are they larger than those of subfamilies I.1 and I.2, but they do not contain cysteine residues and have no N-terminal signal peptide for proteolytic cleavage during type II secretion. Due to the lack of an N-terminal signal peptide, secretion is mediated through the ATP-binding-cassette (ABC) transporter system (Duong et al., 1994).

1.6.3 The GDSL family

GDSL lipases of family II do not exhibit the signature pentapeptide Gly-Xaa-Ser-Xaa-Gly. These enzymes contain a Gly-Asp-Ser-(Leu) [GDS(L)] motif, incorporating the active site serine residue. The serine residue in these enzymes is situated much closer to the N-terminus than in other lipolytic enzymes (Upton and Buckley, 1995). The esterase from \textit{S. scabies} is also included in this family due the 30% similarity it shows towards the \textit{Aeromonas hydrophila} esterase. Crystal structures of the \textit{S. scabies} GDSL esterase show a catalytic dyad instead of a triad in the catalytic center (Wei et al., 1995). Both the enzymes from \textit{S. scabies} and \textit{A. hydrophila} have an \(\alpha/\beta\) tertiary fold substantially
different from that of the $\alpha/\beta$-hydrolase family (Arpigny and Jaeger, 1999). The latter two enzymes share conserved sequence blocks with three other bacterial esterases from *P. aeruginosa*, *S. typhimurium* and *P. luminescens*. Secretion of these lipases are mediated through the autotransporter system where the C-terminal of the protein mediates secretion by the formation of 12 amphipathic $\beta$-sheets to form a pore in the outer membrane.

### 1.6.4 Hormone sensitive lipase (HSL) family

Lipases belonging to families IV-VI are all classified as HSL lipases. The most striking characteristic of family IV lipases is their high amino acid similarity to mammalian HSL (Hemilä *et al.*, 1994). Family V enzymes originate from across the temperature spectrum and include lipases from psychrophilic (*Moraxella* sp., *Psychrobacter immobilis*), mesophilic (*Pseudomonas oleovorans*, *H. influenzae*, *A. pasteurianus*) and thermophilic (*S. acidocaldarius*) bacteria.

Family V share significant amino acid similarity (20-25%) to various bacterial non-lypolytic enzymes such as epoxide hydrolases and dehalogenases, which also possess the typical $\alpha/\beta$-hydrolase fold and a catalytic triad (Misawa *et al.*, 1998). Family VI enzymes represent some of the smallest esterases known and range from 23-26 kDa. The 3D structure of *P. fluorescens* carboxylesterase indicated that the active form is a dimer (Kim *et al.*, 1997) and hydrolyses small substrates exhibiting broad specificity. Limited studies have been performed on the other enzymes in this family as their amino acid sequences were derived from whole-genome sequences. Family VII esterases are large (55 kDa) and share significant amino acid sequence homology (30% identity) to eukaryotic acetylcholine esterases and intestine/liver carboxylesterases. The esterase
from *Arthrobacter oxydans* is active against phenylcarbamate herbicides and is plasmid borne enabling possible transfer to other species or strains (Pohlenz *et al.*, 1992). Family VIII is represented by three enzymes which show striking similarity to class C $\beta$-lactamases. A 150 amino acid stretch is 45% similar to an *Enterobacter cloacae ampC* gene suggesting that the esterases in this family possess an active site more reminiscent of that found in class C $\beta$-lactamases, which involve a Ser-Xaa-Xaa-Lys motif (Galleni *et al.*, 1988).

### 1.6.5 Regulation of lipase gene expression

The regulation of lipase gene expression has been studied in *S. aureus*, *Streptomyces* species and *A. calcoaceticus*. Lipase production is thought to be mediated by quorum sensing and two-component transduction. Quorum sensing is a cell density response system associated with a cascade of other genes, primarily genes associated with virulence (Weingart *et al.*, 1999). Many Gram-negative bacteria produce extracellular signaling molecules called autoinducers which belong to the class of acylated homoserine lactones (AHL’s). AHL’s bind to regulator proteins which induce or repress certain specific target genes. The best-studied example of lipase gene regulation is the lipase operon of *P. aeruginosa*. Extracellular lipase is detected when *P. aeruginosa* enters stationary phase, which is suggestive of cell density-dependant regulation (Stuer *et al.*, 1986). Gene fusion studies with *LipA::LacZ* revealed that lipase gene expression was controlled by the quorum sensing activator RhlR and the autoinducer $N$-butyryl-homoserine lactone (BHL). RhlR activated the transcription of the two-component regulatory system LipR. In turn, LipR regulated lipase transcription by binding to an upstream activating sequence preceding the $\sigma^{54}$-dependant promoter of the lipase operon (Jaeger *et al.*, 1999;
Rosenau and Jaeger, 2000). The global regulator protein GacA has also been implicated in lipase regulation where overproduction of GacA led to increased levels of extracellular lipase (Reimmann et al., 1997).

1.6.6 Mechanism of secretion

All known bacterial lipases are extracellular enzymes. This requires translocation across the single cytoplasmic membrane of Gram-positive bacteria or both the inner and outer membranes, enclosing the periplasmic space of Gram-negative bacteria. Different mechanisms of protein export have evolved to direct bacterial enzymes to their final destination of which three (type I, II and V) are involved in lipase secretion. Efficient secretion of enzymatically active enzymes is tightly coupled to correct folding and involve specific and unspecific periplasmic folding mediated by enzyme-specific chaperones.

1.6.6.1 Type I secretion

Type I secretion pathways, also referred to as ATP-binding cassette (ABC) protein export systems, are employed by a wide range of different Gram-negative bacteria and are responsible for the secretion of toxins, proteases and lipases (Fath and Kolter, 1993; Binet et al., 1997). The proteins secreted by the type I pathway are not subject to proteolytic cleavage and therefore lack cleavable N-terminal leader peptide sequences (Binet et al., 1997). Instead, the secretion signal is located within the C-terminal 60 amino acids of the secreted protein (reviewed by Binet et al., 1997; Duong et al., 1996).
1.6.6.2 Type II secretion

1.6.6.2.1 Secretion across the inner membrane

Type II secretion involves a two step process whereby proteins first cross the inner membrane with the aid of Sec proteins (Pugsley, 1993), followed by translocation across the outer membrane via either Xcp proteins (Filloux et al., 1998) or autotransporter proteins (Henderson et al., 1998).

The Sec secretion system of *E. coli* has been extensively studied and consists of a cytoplasmic secretion-specific chaperone (SecB), a protein translocation ATPase (SecA) and an integral membrane protein complex formed by at least six different protein subunits (SecY, SecE, SecD, SecF, SecG and YajC) (reviewed by Pugsley et al., 1997;).

1.6.6.2.2 Secretion across the outer membrane

Following export of proteins, which are to be secreted, into the periplasm via the Sec translocase system, the proteins may undergo further modifications before they are finally translocated across the outer membrane. This final translocation step requires several accessory proteins, collectively referred to as the type II secretion apparatus or secreton. This apparatus is highly specific and is capable of distinguishing proteins to be secreted from resident periplasmic proteins and, with a few exceptions, it can discriminate between its own secreted proteins and those induced from other species (Lindeberg et al., 1996; Filloux et al., 1998).
The type II secretion systems are widely distributed and appear to be the primary pathway for the secretion of extracellular degradative enzymes by Gram-negative bacteria (Hobbs and Mattick, 1993; Pugsley et al., 1997; Russel, 1998).

### 1.6.6.3 The autotransporter pathway

Bacterial proteins that are targeted to the microbial surface or released into the environment often depend on periplasmic proteins and almost always require outer membrane proteins to promote their secretion. The autotransporter family of Gram-negative bacterial proteins is a unique subset of secreted proteins that do not rely on other proteins for transit from the periplasm to the bacterial surface (reviewed by Henderson et al., 1998; Jacob-Dubuisson et al., 2001). The esterase, EstA, produced by *P. aeruginosa* is an enzyme secreted by the type V autotransporter pathway and contains the characteristic GDSL active site consensus motif (Arpigny and Jaeger, 1999). EstA and other type V proteins possess a C-terminal domain that mediates targeting to and translocation across the outer membrane. All presently known autotransporter proteins originate from pathogenic bacteria and encompass diverse functionalities and include proteases, toxins, adhesins and invasins (Henderson et al., 1998).

### 1.6.7 Periplasmic folding

Lipases secreted via the secreton-mediated pathway fold into an active conformation in the periplasm before they are translocated across the outer membrane. To achieve this secretion-competent conformation, lipases require specific intermolecular folding catalysts, Lif proteins. These foldases are encoded in the same operon as their cognate lipases and have been identified in several bacteria. Lifs represent a unique family of
proteins with no significant homology towards other classes of proteins and are grouped into four families (Rosenau et al., 2004). Algorithms used in protein production reveals that all Lifs appear to possess a similar secondary structure, with 70% consisting of α-helical and 30% of random coil elements (Frenken et al., 1993). The N-terminus of Lifs contain a predicted hydrophobic transmembrane section which is thought to anchor the Lif to the inner membrane with almost the entire protein exposed to the periplasm. However, membrane anchoring does not appear vital as Lifs with truncated N-terminal domains or fused with signal sequences to allow translocation into the periplasm, were still able to activate their cognate lipases when expressed in the host strain. The precise role of lipase activation by Lifs is difficult to determine in vivo due to folding and secretion being such a tightly coupled process. Current knowledge is based on expression studies of various lipase/Lif systems in heterologous hosts as well as in vitro experiments using purified lipases and Lifs. Findings demonstrated that Lifs and their cognate lipases form stable complexes that can be co-purified or co-immunoprecipitated (El Khattabi et al., 2000) and that Lifs mediate refolding of chemically denatured lipases in vitro (Ihara et al., 1995). Lipases are not only dependent on these intermolecular chaperones, they show distinct specificity, since the chaperones of different species are not interchangeable (El Khattabi et al., 1999). To date, the mechanism whereby these foldases recognize their cognate lipases is unknown, which prompted the theory that disulfide bonds might play an important role in lipase/foldase recognition. This however is not the case as cysteine-to-serine variants were able to adopt an active conformation, implying that disulfide bonds are not necessary for the recognition of the lipase by its foldase nor for the interaction of both proteins (Liebeton et al., 2001).
1.6.8 Psychrotrophic lipases

Table 1.3 lists the psychrotrophic lipases isolated over the past 15 years. Although lipases are considered the third most important group of enzymes in biotechnology next to proteases and carbohydrases (Hasan et al., 2005), there are currently only a few documented psychrophilic lipases. The latest report on the isolation of a cold-active lipase is that by Kulakova and co-workers (2004) where they isolated a *Psychrobacter* sp. from Antarctic soil exhibiting lipolytic activity. Isolation and characterization of the enzyme indicated the presence of a cold-active esterase. To date all studies involved in the isolation of cold-active lipolytic enzymes relied on isolation of a pure culture exhibiting lipolytic activity on selected media. Following identification of the isolates, they proceeded to isolate and characterize the lipolytic genes by constructing single genome

<table>
<thead>
<tr>
<th>Origin</th>
<th>Microorganism</th>
<th>Enzyme</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antarctica</td>
<td><em>Moraxella</em> TA144</td>
<td>Lipase</td>
<td>Feller et al., 1991</td>
</tr>
<tr>
<td>Antarctica</td>
<td><em>Psychrobacter immobilis</em> B10</td>
<td>Lipase</td>
<td>Arpigny et al., 1993</td>
</tr>
<tr>
<td>Alaska</td>
<td><em>Pseudomonas</em> sp. B11-1</td>
<td>Lipase</td>
<td>Choo et al., 1998</td>
</tr>
<tr>
<td>Soil isolate</td>
<td><em>Aspergillus nidulans</em> WG312</td>
<td>Lipase</td>
<td>Mayordomo et al., 2000</td>
</tr>
<tr>
<td>Water isolate</td>
<td><em>Pseudomonas</em> sp. Strain KB700A</td>
<td>Lipase</td>
<td>Rashid et al., 2001</td>
</tr>
<tr>
<td>Siberia</td>
<td><em>Acinetobacter</em> sp. Strain No. 6</td>
<td>Esterase</td>
<td>Suzuki et al., 2002</td>
</tr>
<tr>
<td>Dairy products</td>
<td><em>Pseudomonas fragi</em></td>
<td>Lipase</td>
<td>Alquati et al., 2002</td>
</tr>
<tr>
<td>Antarctica</td>
<td><em>Psychrobacter</em> Ant 300</td>
<td>Esterase</td>
<td>Kulakova et al., 2004</td>
</tr>
</tbody>
</table>
libraries and screening for expression in *E. coli*. This has led to the isolation of various lipolytic enzymes of which lipases appear to be more frequent (Table 1.3). The cold-active lipolytic enzymes isolated vary with respect to size and range between 32 kDa and 68 kDa and all display the characteristic Gly-Xaa-Ser-Xaa-Gly motif associated with esterase/lipase proteins of the $\alpha/\beta$-hydrolase superfamily (Arpigny and Jaeger, 1999). The esterase from *Psychrobacter* Ant300 shows 44% identity to the *Moraxella* TA144 lipase and groups, together with *Acinetobacter* sp. Strain No. 6, with the HSL lipase family. The optimum catalytic activity of all the enzymes varied with *Moraxella* TA144 lipase exhibiting the lowest optimum at 17°C and *Acinetobacter* sp. Strain No. 6 esterase and *Psychrobacter immobilis* B10 lipase exhibited the highest at 45°C. The pH range of all the cold-active enzymes isolated ranged from neutral to slightly alkaline (pH 6-9). All but one of the enzymes, *Pseudomonas* sp. Strain KB700A lipase, showed activity towards esters of short chain fatty acids ($\leq$C$_6$), whereas Strain KB700A indicated high activity towards long chain fatty acids (C$_{18}$). Structural analysis of the *Pseudomonas fragi* lipase indicated the presence of 24 arginine residues, more than twice the number present in the *P. aeruginosa* and *B. cepacia* mesophilic enzymes. 20 of the 24 arginine residues were found to be evenly distributed across the surface of the protein, thus increasing the amount of charged residues and enhancing flexibility in cold environments (Gerike *et al.*, 2001). Of the cold-active lipolytic enzymes isolated, the lipase from *A. nidulans* is the only one used in industry and is used in food processing, detergents and the manufacture of cosmetics (Mayordomo *et al.*, 2000).
1.7 Biotechnological applications of lipases and esterases

Microbial lipases and esterases constitute an important group of biotechnologically valuable enzymes because of the versatility of their applied properties and ease of mass production. Lipases are valued biocatalysts because they act under mild conditions, are highly stable in organic solvents, have diverse substrate specificity and they display high levels of stereo-, chemo- and regioselectivity in catalysis (Jaeger and Eggert, 2002; Jaeger and Reetz, 1998).

1.7.1 Lipases in detergents

The most commercially important field of application for hydrolytic lipases is in detergents for both household and industrial use. An estimated 1000 tons of lipases are added to the approximately 13 billion ton of detergents annually (Godfrey and West, 1998). To improve the efficacy of detergents, modern heavy-duty detergents usually contain one or more enzymes such as protease, amylase, cellulase and lipase (Jaeger and Reetz, 1998). Using enzyme cocktails in detergents is environmentally friendly due to a reduction in undesirable chemicals, biodegradability of the enzymes, a non-negative impact on sewage treatment processes and the enzymes present no risk to aquatic life (Hasan et al., 2005). Lipases active under alkaline conditions, such as the A. oryzae derived lipase, are preferred because laundering conditions usually occur between pH 10-11 (Satsuki et al., 1990). Other lipases used in detergents originate from Candida sp. and Chromobacterium.
1.7.2 Biodiesel

Another potential application for lipases is the production of biological diesel. The process entails the production of diesel from vegetable oils. Currently, interest in biodiesel is mostly environmental as the fuel exhibits almost total absence of sulphur and does not contribute to new carbon dioxide emissions (Nabi et al., 2006). The production of biodiesel involves the conversion of vegetable oil into methyl- or other short-chain alcohol esters by a single transesterification reaction catalyzed in an organic solvent (Jaeger and Eggert, 2002). This is however, an energy consuming process also requiring the addition of esters drawback mono and diglycerides, glycerol, water and an alkaline catalyst (Salis et al., 2005). Another drawback is the low-temperature property of the fuel. The triacylglycerol source for the production of the diesel plays an important role, as esters from saturated fatty acids crystallize at higher temperatures than that of unsaturated esters (Salis et al., 2005). The ratio of saturated to unsaturated fatty acids determined the temperature at which the fuel starts to freeze or cloud point (CP) of biodiesel. If the biodiesel mixture is inadequate the CP will be too high and additives such as butyl oleate must be added to maintain fluidity (Linko et al., 1998). Although immobilized P. cepacia lipase has shown promise in overcoming most of the drawbacks, industrial scale production has not yet been attained due to the high cost of the biocatalyst (Hasan et al., 2005).

1.7.3 Food industry

Lipases are used to improve the quality of a number of food products such as fruit juices, baked foods, fermented vegetables and dairy products (Zalacain et al., 1995). The position, chain length and degree of saturation greatly influence the nutritional and
sensory value of a given triglyceride (Jaeger and Reetz, 1998). The hydrolysis of milk fat in dairy products results in flavour enhancement in certain cheeses, especially that of soft cheese. The flavour is produced by the free fatty acids released during lipolysis of the cheese. The addition of lipases which release short chain fatty acids (C4 and C6) generate a sharp, tangy flavour, where as the release of medium chain fatty acids (C12 and C14) imparts a soapy taste (Hassan et al., 2005).
1.8 Aims of this study

The mineral soils of Antarctica have been extensively studied using culture-based approaches (Friedmann, 1993). Molecular phylogenetic analysis has allowed for increased access to microbial diversity in all biotopes including other Antarctic niches such as cryptoendolithic (de la Torre et al., 2003), cryoconite holes (Christner et al., 2003) and lake ice (Priscu et al., 1998) and lake (Brambilla et al., 2001) communities. Extreme environments such as the Arctic potentially harbour a novel nucleic acid pool with biotechnological potential. One such important biocatalyst is lipase of which several have been isolated from similar environments (Feller et al., 1991; Choo et al., 1998; Kulakova et al., 2004). All of the cold-active lipolytic enzymes isolated to date, however, have been accessed using culture-based techniques. In an effort to better understand bacterial community structure in Antarctic Dry Valley mineral soils, and probe the mineral soil metagenome using PCR-based methodology the primary aims of this investigation were the following:

(i) Determine and compare the bacterial diversity of Antarctic Dry Valley mineral soil samples.

(ii) Screen the Antarctic Dry Valley mineral soils for the presence of genes conferring lipolytic activity.
2.1 Reagents

2.1.1 Chemicals

Table 2.1 lists all the strains, plasmids and primers used in this study. Unless otherwise stated, all chemicals used were of analytical grade and obtained from MERCK laboratory supplies. Ingredients used for microbial media was obtained from MERCK (biolab diagnostics), Oxoid (Oxoid ltd) or KIMIX laboratory supplies.

2.1.2 Antarctic soil samples (collection and storage)

Antarctic mineral soil samples were collected from three dry mineral soil sites: (i) underneath a Crab-eater seal carcass on Bratina Island (BIS1); (ii) the mid-slopes of Miers Valley (MVG); and (iii) fine gravels from Penance Pass, a high altitude site between the Miers and Shangri La Valleys (PENP) in sterile 500 ml Nalgene® tubes. Samples were recovered under aseptic conditions by removal of a 1 cm surface layer of mineral soil from a 20 x 20cm sample area. All samples (approx. 400g) were mixed thoroughly and resampled before storage at <0°C for transport. Samples were stored at -80°C until required.

2.1.3 Antibiotics

Supplementation of antibiotics to either broth or growth media was performed aseptically. The appropriate filter sterilized antibiotic was added to autoclaved broth or
media following cooling to ~ 45°. Antibiotic final concentrations were: ampicillin, 100 μg.ml⁻¹; kanamycin, 50 μg.ml⁻¹ and tetracycline, 15 μg.ml⁻¹.

Table 2.1 Bacterial strains, plasmids, and primers used in this work.

<table>
<thead>
<tr>
<th>Strains/Plasmids/Primers</th>
<th>Relevant Characteristics</th>
<th>Source/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH5α</td>
<td>recA endA1 hsdR17 supE4 gyrA96 relA1 Δ(lacZYA-argF)U169 (80dlacZΔM15)</td>
<td>Promega</td>
</tr>
<tr>
<td>pMOSBlue</td>
<td>endA1 hsdR17(r12-m12+)supE44 thi -1 recA1 gyrA96 relA1 lac[F' proAB+ lacIqZΔM15:Tn10(Tc®)]</td>
<td>Amersham Biosciences</td>
</tr>
<tr>
<td>Top10</td>
<td>F' mcrA Δ(mrr-hsdRAM-mcrBC) ϕ80lacZΔM15 ΔlacX74 recA1 deoR araD139 Δ(ara-leu)7697 galU galK rpsL (Str·) endA1 nupG</td>
<td>Invitrogen Corporation (Carlsbad, CA, USA)</td>
</tr>
<tr>
<td><strong>Plasmids:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pMOSBlue</td>
<td>lacZ' Ap^®</td>
<td>Amersham</td>
</tr>
<tr>
<td>PCR-XL-TOPO</td>
<td>lacZ'+ ccdB Km^β</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>pTZ57R/T</td>
<td>lacZ' Ap^®</td>
<td>Fermentas</td>
</tr>
<tr>
<td><strong>Primers:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E9F</td>
<td>5' – GAGTTTGATCTGGCTGATT –3'</td>
<td>Farellly <em>et al.</em>, 1995</td>
</tr>
<tr>
<td>U1510R</td>
<td>5' – GGTACCTGGTACGAGTTT –3'</td>
<td>Reysenbach and Pace, 1995</td>
</tr>
<tr>
<td>#OXF1</td>
<td>5' – CCYGTKGTSYTNCGTACG –3'</td>
<td>Bell <em>et al.</em>, 2001</td>
</tr>
<tr>
<td>#ACR1</td>
<td>5' – AGGCCNCCAKNGARTGNSC –3'</td>
<td>Bell <em>et al.</em>, 2001</td>
</tr>
<tr>
<td>#LipF</td>
<td>5' – GACATGRCNCCYWKGCTRT –3'</td>
<td>This study</td>
</tr>
<tr>
<td>#LipR</td>
<td>5' – GACATGRCNCCYWKGCTRT –3'</td>
<td>This study</td>
</tr>
<tr>
<td>LipR2</td>
<td>5' – GCCAGGATCCGCACTGACG –3'</td>
<td>This study</td>
</tr>
<tr>
<td>M13F</td>
<td>5' – GTTTTCCCCAGTCAGGAC –3'</td>
<td>“www.IDT.com”</td>
</tr>
<tr>
<td>M13R</td>
<td>5' – CAGGAAACGTCATGAC –3'</td>
<td>“www.IDT.com”</td>
</tr>
<tr>
<td>M13F sequencing</td>
<td>5' – CGCGGCAATAGTATCGAC –3'</td>
<td>“www.IDT.com”</td>
</tr>
<tr>
<td>M13R sequencing</td>
<td>5' – GAGCCGATAAATTTCACACAGG –3'</td>
<td>“www.IDT.com”</td>
</tr>
<tr>
<td>16F3:16.5</td>
<td>5' – GCCAGCAGCGCCGGAATGAC –3'</td>
<td>UCT</td>
</tr>
</tbody>
</table>

# Designed to target the conserved oxyanion hole region of the lipase gene
^ Designed to target the conserved active site region of the lipase gene
Δ Modification of the OXF1 primer by lessening the degeneracy and so narrow the specificity to true lipase sequences
* Modification of the ACR1 primer by lessening the degeneracy and so narrow the specificity to true lipase sequences
^ Manufactured by Integrated DNA Technologies in the USA
§ Sequencing primer designed at the University of Cape Town sequencing facility to bind internally at bp 517-536 of the 16S rDNA gene (Figure 1.1)
2.1.4 Enzymes

The following enzymes were used in DNA manipulation: Restriction enzymes and T4 DNA ligase (Fermentas), Shrimp Alkaline Phosphatase (SAP) (Roche). Thermostable BIO-X-ACT™ DNA polymerase and 10 × reaction buffer was obtained from Bioline.

2.2 Culture Media

2.2.1 Luria-Bertani (LB) broth

LB broth consisted of 1% (w/v) tryptone, 0.5% (w/v) yeast extract and 1% (w/v) NaCl. All constituents were mixed together with ultra high quality (UHQ) Millipore water and the pH adjusted to 7.0 using 10 N NaOH prior to autoclaving.

2.2.2 LB agar

LB agar consisted of LB broth prepared as above (Section 2.2.1) with the addition of 1.3% bacteriological agar prior to autoclaving.

2.2.3 Terrific Broth (TB)

TB was prepared according to Sambrook et al. (1999) and consisted of 1.2% (w/v) tryptone, 2.4% (w/v) yeast extract, 0.4% glycerol (v/v), 17 mM KH₂PO₄ and 72 mM K₂HPO₄ in UQH Millipore water. Tryptone, yeast extract and glycerol were made up separately in 900 ml UHQ before autoclaving. Following cooling to room temperature all solutions were mixed.
2.2.4 GYT medium

GYT medium was prepared according to Tung and Chow (1995) and contained 10% (v/v) glycerol, 0.125% (w/v) yeast extract and 0.25% (w/v) tryptone. The GYT medium was mixed thoroughly, filtered using a 0.22 μ filter and stored at 4°C until required.

2.2.5 Lipase specific medium

Lipase agar plates were prepared according to Kouker and Jaeger (1987) with slight modification. The medium was prepared by mixing 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% NaCl, 1% (v/v) olive oil, 0.001% rhodamine B and 0.1% gum Arabic after which the pH was adjusted to 7.0 and the medium autoclaved. Where required, the appropriate antibiotic was added after cooling of the medium before it was poured into plates. Plates were stored at 4°C until used.

2.2.6 Esterase specific medium

Agar plates for esterase screening contained the same constituents as lipase plates except the olive oil and rhodamine B was replaced with 1% tributyrin (v/v).

2.3 Total community DNA Isolation

2.3.1 Modified Zhou method

Community DNA extractions were performed according to the modified Zhou protocol (Stach et al., 2001). Aliquots of mineral soils (5 g) were weighed out into sterile 30 ml Nalgene® centrifuge tubes followed by the addition of 6.75 ml soil extraction buffer (1% CTAB [w/v]; 100 mM Tris, pH 8.00; 100 mM NaH₂PO₄, pH 8.00; 100 mM EDTA; 1.5 M NaCl; 0.02% Protease K [w/v]). The tubes were incubated horizontally at 37°C for 30
min with shaking. 750 μl 20% [w/v] SDS was added to each tube followed by a further 2 h incubation at 65°C with gentle inversions every 20 min. Following incubation, the tubes were centrifuged at 3000 × g for 10 min at room temperature and the supernatant pooled into a sterile Nalgene® 30 ml centrifuge tube. An equal volume Phenol/Chloroform/Isoamyl was added and mixed gently followed by centrifugation at 16 000 × g for 10 min. Supernatants were again transferred to sterile Nalgene® 30 ml centrifuge tubes with the addition of an equal volume of chloroform. After careful mixing the tubes were centrifuged at 16 000 × g for 10 min at room temperature and supernatants recovered. Chloroform washes were repeated until the supernatants were clear. Once all washes were complete 0.6 volumes of isopropanol was added to the supernatants and DNA precipitation allowed to take place overnight at room temperature. DNA was pelleted by centrifugation at 10 000 × g for 10 min, washed with 70% ethanol, recentrifuged at 10 000 × g for 5 min, and air dried in a sterile hood. UHQ Millipore water was used to resuspend the DNA pellet and a small fraction was analysed by gel electrophoresis.

2.3.2 Miller protocol

DNA extraction from all three soils were extracted using the Miller protocol (Miller et al., 1999). Between 0.5 and 1 g of soil was added to sterile 2 ml screw cap tubes containing 0.5 g sterile Quartz sand, followed by 300 μl phosphate buffer, pH 8.00, 300 μl lysis solution (0.5 M Tris-HCl, pH 8.00, 10% SDS [v/v], 100 mM NaCl) and 300 μl chloroform. The sample tubes were mixed and either shaken in a bead beater (Bio101 FastPrep FP120, Savant Instruments Inc. Holbrook, NY) at 4.5 m.s⁻¹ for 40 s or vortexed for 1 – 1.5 min at full speed, followed by centrifugation for 5 min at 13 000 × g. Supernatants
were transferred to clean 1.5 ml Eppendorf tubes with the addition of 7 M NH₄AOc to achieve a final concentration of 2 M. Tubes were inverted several times until white flocculates appeared and centrifuged for 5 min at 13 000 × g. The supernatants were recovered and transferred to clean centrifuge tubes after which 0.6 volumes of isopropanol was added, the tubes inverted several times and incubated at room temperature for 15 min. DNA was collected at room temperature by centrifugation at 13 000 × g for 10 min, washed with 70% EtOH and air dried in a sterile hood. UHQ Millipore water was used to resuspend the air-dried DNA pellet and a small fraction was analysed by gel electrophoresis.

2.3.3 Bead beating protocol
Environmental DNA was extracted from mineral soils using the FastDNA spin kit for soil (Bio101 Inc., Vista, CA, USA) and bead beater (Bio101 FastPrep FP120, Savant Instruments Inc. Holbrook, NY). Extraction was performed as described in the manufacturer’s instructions for soil DNA extraction and DNA eluted in UHQ Millipore water.

2.4 Soil analysis
2.4.1 Dry weight assessment and water content
Dry weights were determined by placing 10 g samples of mineral soil sample, in duplicate, in pre-weighed glass petri dishes. The samples were incubated at 100°C and weighed every 24 h for a period of 3 days using a Mettler-Toledo PE 360 balance. Water content was determined the total difference in soil sample weight, expressed as a percentage.
2.4.2 Protein assessment

Total protein of each of the soil samples was determined using the Bio-Rad protein assay kit (Bio-Rad). 200 \( \mu \text{l} \) of the supernatant of DNA extractions was used in the microassay procedure as outlined by the manufacturer. Optical density measurements were performed at 595 nm and plotted against a 1-20 \( \mu \text{g} \) BSA standard curve.

2.4.3 Lipid analysis

Total lipid was extracted according to Folch et al. (1957). 20 ml of chloroform:methanol (2:1 [v/v]) was prepared and added to 0.5 g of each soil sample. Following 15 min shaking at room temperature the samples were centrifuged at 3000 \( \times \text{g} \) for 5 min and the supernatant removed. The solvent phase was then washed with 0.2 volumes 0.9% aqueous NaCl and centrifuged at 1000 \( \times \text{g} \) for 5 min. The upper phase was aspirated and the interphase washed twice with methanol:water (1:1). Following the final wash, centrifugation and removal of the upper phase, the lower chloroform phase containing extracted lipids was allowed to evaporate overnight in the fume hood. Lipid content was determined gravimetrically.

2.5 DNA quantification

Extracted, dried community DNA was resuspended in sterile UHQ H\(_2\)O and allowed to stand overnight. Uncut commercial \( \lambda \)-DNA of known concentration was used as a DNA standard to determine the quantity of total community DNA extracted. \( \lambda \)-DNA was diluted to a concentration of 5 ng.\( \mu \text{l}^{-1} \) and loaded at 5, 10 and 15 ng aliquots into a 1% (w/v) agarose gel. Dilutions of 1:10 and 1:100 of community DNA was loaded adjacent to the
λ-DNA standards and analyzed following agarose gel electrophoresis for 4 min at 80 V. For more sensitive DNA analysis, quantification was performed using the Nanodrop ND-1000. The instrument was blanked using 1 μl of the same UHQ Millipore water as for DNA resuspension or elution. 1 μl of resuspended or eluted DNA was then added to the reading platform and the DNA concentration recorded.

2.6 PCR amplification using 16S rDNA primers

High molecular weight community DNA was used a template for 16S rDNA amplification using primers E9F (Farely et al., 1995) and U1510R (Reysenbach and Pace, 1995). PCR reactions (25 μl) contained ~ 10 ng template DNA, 1 × NEB PCR buffer (20 mM Tris-HCl (pH 8.8), 10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 1% [v/v] Triton X – 100), 0.16 mM of each dNTP, 160 ng BSA, 0.5 pmol of each primer and 0.5 μl Taq DNA polymerase. PCR reaction mixtures were placed in an Applied Biosystems thermocycler Gene Amp® 2700 using the following PCR cycling conditions: Initial denaturation at 95°C for 90 s, followed by 28 cycles of denaturation at 94°C for 20 s, annealing at 50°C for 40 s, and extension at 72°C for 1 min. The final elongation step was performed at 72°C for 3 min. A positive reaction containing E. coli genomic DNA as template DNA was included. For control purposes, a reaction mixture containing all reagents except template was routinely included. An aliquot of each reaction mixture was analysed using gel electrophoresis as described in Section 2.7.

2.7 Agarose gel electrophoresis

Analysis of DNA was performed using agarose gel electrophoresis (Sambrook et al., 1982). Horizontal 0.8% – 2% (w/v) TBE agarose slab gels were cast and electrophoresed
at 100 V in 0.5 × TBE buffer (40 mM Tris-HCl, 1 mM EDTA, 20 mM boric acid, pH 8.5). Where DNA was to be recovered and used in downstream applications, 0.8% – 2% (w/v) TA agarose slab gels were cast and electrophoresed at 100 V in 0.5 × TA (20 mM Tris-HCl, 10 mM glacial acetic acid, pH 8.5). To allow visualization of the DNA on a UV transilluminator, the gels were supplemented with 0.5 μg.ml⁻¹ ethidium bromide. The DNA fragments were sized according to their migration in the gel as compared to that of standard DNA molecular markers (Lamda DNA restricted with PstI; 100 bp HyperladderI, Bioline).

2.8 GFX™ DNA purification
Purification of DNA from either solution or agarose gels were performed using the GFX™ DNA and gel band purification kit (Amersham biosciences) according to manufacturer’s specifications.

2.9 pMOSBlue blunt ended cloning
The pMOSBlue (Amersham Biosciences) cloning plasmid was used in the construction of the 16S rDNA metagenomic libraries. pMOSBlue is supplied as a blunt 5’ – dephosphorylated linear plasmid suitable for cloning blunt ended DNA inserts. GFX™-purified (Section 2.8) 16S rDNA PCR amplicons were blunt ended, ligated and transformed into pMOSBlue competent cells according to the manufacturer’s instructions.

2.10 Preparation of electrocompetent E. coli cells
Electrocompetent DH5α, XL1blue and JM109 E. coli cells were prepared as outlined in Sambrook and Russell (2001), with slight modification. All glassware were thoroughly
acid-washed with 30% H$_2$SO$_4$, rinsed and autoclaved prior to use. A single colony of the 
*E. coli* strain was inoculated into 30 ml of LB-broth and incubated at 37°C with shaking 
until stationary phase. 10 ml of the culture was transferred to two aliquots of 500 ml of 
LB-broth and incubated at 30°C until mid-logarithmic phase (OD$_{600}$ of 0.4). The flasks 
were rapidly cooled in ice-water for 20 min and the cells were collected in polypropylene 
tubes by centrifugation at 1000 × g for 10 min in an Eppendorf 5810 R swing bucket 
centrifuge. The supernatant was decanted and the cells resuspended in equal volume ice-
cold Millipore water. After harvesting the cells as above, the pellets were resuspended in 
250 ml 10% glycerol, collected by centrifugation and the supernatant carefully decanted. 
The cell pellet was resuspended in 1 ml GYT medium and the cell density at OD$_{600}$ 
adjusted to between 2 × 10$^{10}$ to 3 × 10$^{10}$ cells.ml$^{-1}$. The cells were aliquotted into 40 µl 
volumes, and stored at -80°C until required.

### 2.11 Preparation of chemically competent *E. coli* cells

Chemically competent cells were prepared according to the method of Hanahan (1983) 
with slight modification. All glassware were thoroughly acid-washed with 30% H$_2$SO$_4$, 
rinsed and autoclaved prior to use. A single colony of the *E. coli* strain was inoculated 
into 30 ml of LB-broth and incubated at 37°C with shaking until stationary phase. 1 ml of 
the culture was transferred to 100 ml of LB-broth and incubated at 30°C until mid-
logarithmic phase (OD$_{600}$ of 0.5). The flasks were rapidly cooled in ice-water for 20 min 
and 60 ml of the cells were collected in polypropylene tubes by centrifugation at 1000 × g 
for 10 min in an Eppendorf 5810 R swing bucket centrifuge. After discarding the 
supernatant, the cells were resuspended in 0.5 × volume filter sterilized competency 
buffer (0.1 M CaCl$_2$ [w/v], 0.07 M MnCl$_2$ [w/v] and 0.04 M NaOAc [w/v], pH 5.5) and
incubated at 4°C for 30 min. Following incubation the cells were harvested by centrifugation at 1000 × g for 5 min and resuspended in 7.5 ml competency buffer. 575 μl 80% glycerol was added thoroughly mixed and the competent cells dispensed into 100 μl aliquots and stored at -80°C until required.

2.12 Transformation of *E. coli* cells by:

2.12.1 Electroporation

An Eppendorf tube containing 40 μl of electrocompetent cells was removed from -80°C and allowed to thaw on ice. 2μl of ligation mix was added to the thawed cells and gently mixed. The mixture was returned to ice for ~1 min then pipetted into a pre-cooled 0.1 cm sterile electroporation cuvette (Bio-Rad Laboratories, Hercules, CA, USA). Electroporation was performed using the following conditions: 1.25 – 1.8 kV, 25 μF, 200 Ω. Immediately following electroporation, 950 μl TB broth, pre-warmed to 37°C, was added to the cuvette, the cells transferred to a 15 ml Falcon tube and incubated at 37°C for 1 h with agitation. The cells were plated in aliquots of 5 to 50 μl onto LB-agar plates supplemented with the appropriate antibiotic. Where applicable, recombinant transformants were selected by blue/white colour selection based on insertional inactivation of the *lacZ* gene. For this purpose, the cells were spread together with 40 μl of X-gal (2% [v/v] stock solution) and 10 μl IPTG (100 mM stock solution) over the surface of LB-agar plates, supplemented with the appropriate antibiotic and incubated overnight at 37°C.

2.12.2 Chemical
An Eppendorf tube containing 100 μl of chemically competent cells was removed from -80°C and allowed to thaw on ice. 2μl of ligation mix was added to the thawed cells and gently mixed. The mixture was incubated on ice for 30 min then heat-shocked at 42°C for 90 s in a water bath. The Eppendorf tube was returned to ice for 2 min where after 900 μl of sterile LB-broth was added and the Eppendorf tube incubated at 37°C for 1 h with agitation. The cells were plated in aliquots of 100 to 200 μl onto LB-agar plates supplemented with the appropriate antibiotic. Where applicable, recombinant transformants were selected by blue/white colour selection based on insertional inactivation of the lacZ gene. For this purpose, the cells were spread together with 40 μl of X-gal (2% [v/v] stock solution) and 10 μl IPTG (100 mM stock solution) over the surface of LB-agar plates, supplemented with the appropriate antibiotic and incubated overnight at 37°C.

2.13 Colony PCR

Following transformation of *E. coli* cells (Section 2.11), colony PCR was used to screen all putative recombinant clones. The putative recombinants were aseptically inoculated into LB-broth containing the appropriate antibiotic (Section 2.1.2) and incubated overnight at 37°C with agitation. Cultures were further analysed by pipetting 200 μl of each into 0.6 ml PCR tubes. The tubes were centrifuged at 13 000 × g to pellet the cells and the supernatant discarded. The cells were resuspended in 200 μl UHQ Millipore water and lysed by incubation at 98°C for 5 min. Tubes were then centrifuged at 13 000 × g for 5 min to pellet cell debris. The DNA-containing supernatant served as template in 30 μl PCR reactions performed essentially as described in Section 2.6 (the annealing
temperature was lowered to 49\(^\circ\)C). An aliquot of each PCR reaction was analysed by gel electrophoresis as described in Section 2.7.

### 2.14 Restriction endonuclease digestion

All restriction enzyme digestions were performed in sterile Eppendorf tubes in small reaction volumes (10 – 20 \(\mu\)l). The reactions contained the appropriate volume of 10 \(\times\) buffer supplied by the manufacturer for the specific enzyme, and 5 – 10 U of enzyme per \(\mu\)g of plasmid or genomic DNA. Reactions were incubated for either short periods, 0.5 – 1.5 h, or overnight in a water bath at 37\(^\circ\)C, unless specified otherwise. When digestions included two enzymes requiring different salt concentrations for optimal activity, the enzyme requiring a lower salt concentration was used first after which the salt concentration was adjusted and the second enzyme added. The digestion products were analyzed by gel electrophoresis in 1\% or 2\% (w/v) agarose gels as described in Section 2.7.

### 2.15 ARDRA analysis of 16S rDNA amplicons

Following colony PCR (Section 2.12), amplicons of the correct size were subjected to restriction endonuclease digestion (Section 2.13) by taking 5 \(\mu\)l of the PCR reaction and 6 U of the tetranucleotide-specific enzyme \(AfaI\). Restriction digestions (15 \(\mu\)l) were viewed on 2\% agarose gels following gel electrophoresis (Section 2.7).
2.16 Plasmid DNA extraction

2.16.1 Alkaline lysis

Colonies were picked from the agar plates, inoculated into 5 ml of LB-broth supplemented with the appropriate antibiotic, and incubated overnight at 37°C with agitation. Plasmid DNA was isolated from the cultures by the alkaline lysis method (Birnbiom and Doly, 1979), with the following modifications. After incubation, cells from 2 ml of each culture was collected in 2 ml Eppendorf tubes by centrifugation at 10000 × g for 1 min at room temperature. The supernatant was discarded and the bacterial pellet suspended in 400 μl of Solution 1 (50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA, pH 8.0). After incubation at room temperature for 10 min, 400 μl of Solution 2 (1% [w/v] SDS, 0.2 N NaOH) was added and the tubes were incubated on ice for 10 min. Following the addition of 300 μl of 7.5 M ammonium acetate (pH 7.6), the tubes were incubated on ice for 10 min, and then centrifuged at 13 000 × g for 5 min at room temperature. The plasmid DNA was precipitated from the supernatant by the addition of 650 μl isopropanol for 10 min at room temperature. The precipitated plasmid DNA was collected at room temperature by centrifugation at 12 000 × g for 10 min and the supernatant discarded before addition of 100 μl of 2 M ammonium acetate (pH 7.4). The tubes were incubated on ice for 10 min. Following ambient centrifugation at 12 000 × g for 5 min, 110 μl of isopropanol was added to the supernatant and the tubes incubated at room temperature for 10 min. Precipitated DNA was collected and the pellets washed with 70% ethanol to remove residual salts from the DNA. The DNA was air-dried and resuspended in UHQ Millipore water. Plasmid DNA was analyzed on a 1% (w/v) agarose gel as described in Section 2.7.
2.16.2 Talent Kit

Plasmid extractions performed for subsequent nucleotide sequence analysis was performed using the Talent plasmid purification kit.

2.17 Nucleic acid sequencing

Sequencing of cloned insert DNA was performed using the MegaBACE 500 Automated Capillary DNA Sequencing System (Amersham Biosciences). Where 16S library clone inserts were sequenced, the internal 16S rDNA *E. coli* primer F3:16.5, which primes nucleotides 517 – 536, was used. All other sequencing reactions entailed using universal vector derived primers M13 forward and/or reverse (Table 1). Cloned 16S rDNA genes were sequenced by targeting the negative strand allowing $1 \times$ coverage. Both strands of all the partial lipase genes were sequenced to allow minimum ambiguity.

2.18 Preparation of metagenomic libraries

2.18.1 Metagenomic DNA digestion

Approximately 2 μg aliquots of community DNA were used for restriction enzyme digestion using *Alu*I. Digestions were performed using 6 U of *Alu*I at time intervals varying between 45 and 55 min. Following digestion, reaction mixtures and λ-/*Pst*I molecular weight marker were loaded into separate wells of a 1.5% - 0.8% gradient agarose gel and electrophoresed as described in Section 2.7.

2.18.2 A-tailing of 3’ termini of restriction digested metagenomic DNA

The addition of single adenosine nucleotides to the 3’ ends of the partially digested metagenomic DNA was performed using a standard PCR reaction. Following restriction
endonuclease digestion (Section 2.16) reaction mixtures were used directly in A-tailing PCR reactions at a concentration not exceeding 30% (v/v). PCR reactions (60 μl) contained 20 μl restriction digested metagenomic DNA reaction, 1 × NEB PCR buffer (20 mM Tris-HCL (pH 8.8), 10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 1% [v/v] Triton X – 100), 0.1 mM dATP, and 1.5 μl Taq DNA polymerase. The PCR reaction mixtures were placed into an Applied Biosystems thermocycler Gene Amp® 2700 using the following conditions: 72°C for 30 min followed by rapid cooling to 4°C and gel electrophoresis as described in Section 2.7.

2.18.3 Recovery of restriction digested metagenomic DNA from agarose gel

Following electrophoresis and UV visualization of digested A-tailed metagenomic DNA, the 2 – 10 kb fraction was excised using a sterile scalpel blade. The excised gel was placed in a sterile 1.5 ml Eppendorf tube and the DNA recovered using the GFX™-gel extraction kit (Amersham Biosciences).

2.18.4 5’-dephosphorylation of metagenomic DNA

Dephosphorylation of 5’ ends of GFX™-purified genomic DNA was performed using Shrimp Alkaline Phosphatase (SAP), Roche, according to manufacturer’s specifications. The reaction was made up to a final volume of 50 μl by adding 12 μl SAP at 1 U.μl⁻¹ and 5 μl 10 × buffer (50 mM Tris-HCl, 5 mM MgCl₂, pH 8.5) to the eluted GFX™-purified DNA, followed by incubation at 37°C for 1 h.
2.18.5 Metagenomic library construction

The pCR®-XL-TOPO® cloning kit (Invitrogen) was used for the construction of metagenomic libraries. A ligation reaction was set up by combining 5 ng pCR®-XL-TOPO® and 0.87 μl (130 ng.μl⁻¹) dephosphorylated and A’ – tailed metagenomic DNA. The mixture was incubated at 22°C for 5 min and quenched by the addition of 6× TOPO® cloning stop solution. The ligation reaction was immediately used to transform electrocompetent E. coli Top10 cells (Invitrogen.)

2.19 Amplification of metagenomic library

After plating 5 μl of the ligation reaction to determine cloning efficiency, the remainder of the transformation mixture (~995 μl) was removed from 4°C, transferred to 10 ml of LB-broth and grown for 4 – 6 h at 37°C with shaking. After the allotted time, 5 μl of the culture was plated onto kanamycin selective media and allowed to grow at 37°C over night. The remainder of culture was stored at 4°C to stunt growth. The following day, the colonies were counted and the theoretical number of times the original library was amplified determined. Colonies counted following plating after 4 h of shaking at 37°C were eight times more compared to t=0. To the ~ 11 ml of culture, 11 ml of 30% sterile glycerol was added and thoroughly mixed. Aliquots of 2.75 ml, which represented a single copy of the library, was made into labeled 2 ml sterile screw-cap tubes and stored at -80°C until needed.

2.20 Activity-based screening of the metagenomic library

Two copies of the library were thawed on ice and separately plated in aliquots of 60 μl (~2000 clones) onto esterase and lipase specific media and incubated at 37°C until
colonies were visible. The plates were removed and incubated further at 16°C for 7 d and monitored for the presence of lipolytic activity.

2.21 PCR-based screening of the metagenomic library

A single copy of the library was subjected to alkaline lysis to purify the recombinant pCR®-XL-TOPO® plasmid DNA. The isolated plasmid DNA was divided into ten aliquots and used as template DNA for the detection of lipolytic genes during PCR. A neat, 1:10, 1:20 and 1:50 dilution was prepared from each aliquot and each template was added to a separate PCR reaction mix (30 μl) containing 1 × NEB PCR buffer (20 mM Tris-HCL (pH 8.8), 10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 1% [v/v] Triton X – 100), 0.1 mM dATP, 0.5 pmol of each primer LipF and LipR₂ and 1.5 μl Taq DNA polymerase. PCR reaction mixtures were placed in an Applied Biosystems thermocycler Gene Amp® 2700 using the following PCR cycling conditions: Initial denaturation at 95°C for 90 s, followed by 35 cycles of denaturation at 94°C for 20 s, annealing at 50°C for 40 s, and extension at 72°C for 1 min. The final elongation step was performed at 72°C for 3 min. For control purposes, a reaction mixture containing all reagents except template was included. A positive reaction containing a cloned lipase gene of B. multivora as template DNA was included. An aliquot of each reaction mixture was analysed using gel electrophoresis as described in Section 2.7

2.22 PCR amplification of partial lipase fragments

For amplification of partial lipase fragments, primers LipF and LipR₂ (Table 1) were used. Community DNA isolated from Antarctic Dry Valley mineral soils served as template for reactions. The PCR reactions (20 – 30 μl) contained between 10 and 50 ng genomic DNA,
1 × NEB PCR buffer (20 mM Tris-HCl (pH 8.8), 10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 1% [v/v] Triton X – 100), 2 mM MgCl₂, 0.2 mM of each deoxynucleoside triphosphate (dNTP), 0.5 pMol of each primer and 1 U Taq polymerase. The PCR reactions were placed in a Perkin-Elmer 2400 thermocycler using the following conditions: initial denaturation at 94°C for 5 min, followed by 4 cycles of denaturation at 94°C for 30 s, annealing at 65°C for 1 min, and elongation at 72°C for 2 min. Cycling conditions were then altered: the same denaturation and elongation conditions were used, but the annealing temperature was reduced to 64°C with a reduction of 1°C every cycle for 14 cycles. Cycling conditions were again adjusted: the denaturation and elongation conditions were maintained, but the annealing temperature was set at 50°C for 20 cycles. Following the last cycle, a final elongation step at 72°C for 3 min was performed. For control purposes, a reaction mixture containing all reagents except template DNA was included. Aliquots of the PCR reaction mixtures were subsequently analyzed by agarose gel electrophoresis as described in Section 2.7.
Chapter 3

Bacterial diversity in Antarctic Dry Valley mineral soils

3.1 Introduction

Temperate and tropical soil communities are regarded as among the most complex and diverse assemblages of microorganisms (Kuske et al., 1997), with estimated bacterial numbers in the order of $10^9$ cells.g$^{-1}$ and over $10^4$ distinct species, as shown by reassociation kinetics (Dunbar et al., 2002; Torsvik et al., 1990). However, the desiccated mineral soils of the Dry Valleys, Ross Dependency, Eastern Antarctica are generally thought to harbour very low cell densities (Cameron et al., 1972) supporting the perception that these so-called extreme environments exhibit low species diversity and low cell numbers. This is attributed to the imposition of environmental extremes which, for the Antarctic Dry Valleys, include low temperatures, wide temperature fluctuations, low nutrient status, low water availability, high incident radiation and physical disturbance (Wynn-Williams, 1990). Nevertheless, it was recently shown by ATP, lipid and DNA quantitation that Dry Valley mineral gravels may contain between $10^6$ and $10^8$ prokaryotic cells.g$^{-1}$ (Cowan et al., 2002).

The current understanding of Antarctic mineral soil microbiology is based almost exclusively on culture-based studies. These studies have suggested that most Antarctic microbes belong to a restricted number of cosmopolitan taxa and are largely aerobic, with only few reported anaerobic isolates (Friedmann, 1993). Large numbers of coryneform-related bacteria such as Arthrobacter, Brevibacterium, Cellulomonas,
Corynebacterium were reported together with gracilicutean isolates (members of the Gram-negative Eubacteria) such as Pseudomonas and Flavobacterium. Firmicutean bacteria isolated included Bacillus, Micrococcus, Nocardia, Streptomyces, Flavobacterium and pseudomonads (Cameron et al., 1972; Friedmann, 1993). A number of less common genera such as Beijerinckia, which rarely occur outside tropical soils, Xanthomonas, a pathogen associated with higher plants and Planococcus, a marine genus, have also been isolated from Antarctic soils (Friedmann, 1993). Cyanobacteria are also well-documented inhabitants of Antarctic soil biotopes (de la Torre et al., 2003; Taton et al., 2003) but are thought to be restricted to moist habitats (de los Ríos et al., 2004).

It is now widely acknowledged that culture-based community studies inevitably induce a high degree of bias, while important groups of organisms which may be fastidious, co-culture-dependant or in a viable but non-culturable (VBNC) state may be unrepresented (Amann et al., 1995; Holmes et al., 2000; McDougald, 1998; Waterbury et al., 1979). It is therefore probable that historical data from culture-dependant studies do not accurately represent the true microbial species diversity of the Dry Valley mineral soils. A number of important Antarctic Dry Valley microbial biotopes, including cryptoendolithic communities (de la Torre et al., 2003), cryoconite holes (Christner et al., 2003), and lake ice and marine ice flows (Priscu et al., 1998) have been subject to detailed community analyses using modern molecular phylogenetic techniques. However, the supposedly less complex and more ‘extreme’ mineral soils have yet to be investigated in detail.
3.2 Aim

The principle aim of the work reported in this chapter was to assess and compare bacterial diversity of three different Antarctic Dry Valley mineral soils.

To achieve this aim, the following objectives were set forth:

- Isolation of pure, intact high molecular weight DNA from all three sample sites
- PCR amplification of 16S rDNA genes
- Construction of three independent 16S rDNA libraries
- Sequence analysis and identification of unique 16S clones

3.3 Results

3.3.1 Soil properties

The properties for each of the soil samples used in this study are listed in Table 3.1. Sample MVG was a fine particulate gravel collected from the surface on a south-facing slope 225 m above sea level. This mid-altitude sample represents a desiccated, oligotrophic habitat, exposed to the harsh environmental conditions (Wynn-Williams, 1990). DNA yields from multiple extractions averaged 840 ng.g$^{-1}$.

Sample PENP was recovered in a saddle between Miers and Shangri-La Valleys at an altitude of 585 m above sea level. The site was approximately 50 m from the margin of Lake Purgatory. PENP is also desiccated and oligotrophic but likely to be subject to a higher incidence of humidity due to the altitude.
Table 3.1  Site descriptions for Dry Valley mineral soil samples

<table>
<thead>
<tr>
<th>Sample code</th>
<th>Site description</th>
<th>GPS</th>
<th>DNA (ng/g)</th>
<th>Protein (μg/g)</th>
<th>Lipid (μg/g)</th>
<th>Altitude (m)</th>
<th>Altitude (% wt.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PENP</td>
<td>Dry sand from high altitude site</td>
<td>78° 04.762' 165° 52.083'</td>
<td>480</td>
<td>280</td>
<td>86</td>
<td>584</td>
<td>0.7</td>
</tr>
<tr>
<td>MVG</td>
<td>Dry sorted sands and gravels from mid altitude valley slopes</td>
<td>78° 06.140' 165° 48.646'</td>
<td>840</td>
<td>386</td>
<td>66</td>
<td>225</td>
<td>0.8</td>
</tr>
<tr>
<td>BIS</td>
<td>Fine dark particulate soil from underside of seal carcass</td>
<td>78° 00.966' 165° 32.795</td>
<td>320</td>
<td>656</td>
<td>326</td>
<td>2</td>
<td>6.2</td>
</tr>
</tbody>
</table>

Sample BIS was a eutrophic soil collected from the underside of a crab-eater seal carcass. The soil was dark and had a clay-like consistency. BIS showed high levels of both protein and lipid and contained almost 8-fold more water than the other two sites. A higher water content implies the presence of organic matter, as organic matter has been shown to facilitate water retention (Smith and Tearle, 1985). Higher levels of lipid and protein might be attributed to the high levels of organic carbon which provide nutrition facilitating a higher turnover. Higher turnover of nutrients and increased biological activity could possibly explain the lower extracellular DNA levels noted for BIS compared to MVG and PENP. Not only were the DNA yield from BIS low, but also contained high levels of humic acids and other PCR inhibitors. Higher DNA yields obtained from MVG and PENP might be due to low turnover within the soils. DNA isolated from MVG and PENP also showed no humic acid contamination requiring no additional purification prior to PCR. DNA quantities isolated from the mineral soils however were significantly lower compared to DNA yields of between 2 and 50 μg·g⁻¹ from more temperate soils (Stach et al., 2001).
3.3.2 Cloning of the 16S rDNA PCR-amplicons into pMOSBlue

To assess the prokaryotic diversity in the Antarctic cold mineral deserts, universal primers E9F (Farely et al., 1995) and U1510R (Reysenbach and Pace, 1995) were used to target and PCR-amplify 16S rDNA genes. The primers were designed to target the conserved regions at the 3’ and 5’ ends of the 16S gene, and to yield a 1.5 kb PCR product. Using community DNA isolated from all three sample sites as template, PCR was performed as described in Section 2.6 (Fig. 3.1). A single amplicon of 1.5 kb was observed following gel electrophoresis (Fig. 3.2). No amplification products were observed in the negative control. The gel-purified amplicons were polished and phosphorylated as described in Section 2.9 and ligated to the pMOSBlue cloning plasmid. Following transformation of MOSBlue competent cells, recombinant transformants with a Gal− phenotype on X-gal containing indicator plates were subjected to colony PCR (Section 2.12). Colonies yielding 1.5 kb PCR-amplicons (Fig. 3.2) using vector-derived M13 forward and M13 reverse primers were subjected to ARDRA analysis. False positives which failed to yield amplicons or produced amplicons of the incorrect size (Fig. 3.3 lanes 2 and 10) were discarded.

3.3.3 16S rDNA ARDRA analysis

Recombinant pMOSBlue plasmids containing the expected 1.5 kb 16S rDNA sequences were subjected to ARDRA. The number and size of each DNA fragments signified a unique ARDRA pattern (Fig. 3.4). ARDRA analyses were continued until 20 unique 16S rDNA clones from each sample site had been identified. In total, 181 clones were analysed, 81 from PENP, 55 from MVG and 45 from BIS. This implies that 24% of PENP, 36% of MVG and 40% of BIS clones were unique. Collector’s curve data (Fig. 3.5)
Fig. 3.1 Metagenomic DNA extracted from Dry Valley mineral soils prior to 16S PCR. Lane 1, DNA molecular weight marker; Lane 2, MVG metagenomic DNA; Lane 3, PENP metagenomic DNA, Lane 4, BIS metagenomic DNA. The sizes of the DNA molecular weight marker, phage lambda DNA digested with PstI, are indicated to the left of the figure.

Fig. 3.2 Amplicons obtained following 16S PCR amplification using MVG, PENP and BIS metagenomic DNA as template. Lane 1, DNA molecular marker; Lane 2, Sample of the PCR reaction using MVG as template; Lane 3, Sample of the PCR reaction using PENP as template; Lane 4, Sample of the PCR reaction using BIS as template; Lane 5, Negative control lacking template DNA. The sizes of the DNA molecular weight marker, phage lambda DNA digested with PstI, are indicated to the left of the figure.
Fig. 3.3  Amplicons obtained following colony PCR amplification of the 16S rDNA genes using plasmid pMOSBlue DNA as template. Lane 1, DNA molecular weight marker; Lanes 2 – 11, Sample of reaction mixture following PCR; Lane 12, Negative control reaction lacking template DNA. The sizes of the DNA molecular weight marker, phage lambda DNA digested with PstI, are indicated to the left of the figure.

Fig. 3.4  16S rDNA amplicons subjected to ARDRA analysis. Lane 1, DNA molecular weight marker; Lanes 2-9, 16S rDNA amplicons restricted with AlfaI. The sizes of the DNA molecular weight marker, phage lambda DNA digested with PstI, are indicated to the left of the figure.
indicates that all three samples are still rising and that sampling has not been exhausted.

### 3.3.4 16S rDNA analysis and distribution

Partial sequence data was obtained using internal primer F3:S16.5 (Table 2.1). Sequences obtained and used for BLASTn analysis included variable regions V5 and V6 and part of V4 (Fig. 1.1). Bacteria identified from the 60 sequences representing all 181 clones grouped into eight broad phylogenetic groups (Cyanobacteria, Actinobacteria, Acidobacteria, Verrucomicrobia, α-Proteobacteria, β-Proteobacteria, Chloroflexi and Bacteroidetes) (Table 3.2) and showed an average identity of $\geq 91\%$ to known phylotypes. Three phylotypic groups showed distribution in all three sites where the remaining five were either unique to one or two sites. Cyanobacteria (13%), Actinobacteria (26%) and Acidobacteria (16%) represented the majority of the identified phylotypes.
3.4 Discussion

3.4.1 Phylotype coverage

Cyanobacteria are generally associated with moist environments making water availability the principal factor dictating cyanobacterial distribution. For example, cyanobacteria are common in moist soils (Miller and Bebout, 2004) and aquatic habitats including glacial streams (Vincent et al., 1993) and flushes (Horne, 1972), saline lakes (Parker and Wharton, 1985) and cryoconite holes (Christner et al., 2003). Cyanobacteria are known to form thick microbial mats on the upper surface of lake sediments (Broady, 1996). The ability to spread out on various surfaces is conferred by gliding motility of filamentous cyanobacteria together with the secretion of extracellular polysaccharides (EPS). This results in a highly pigmented and structured biofilm covering the substrate (de los Ríos et al., 2004). Their presence in the PENP sample might be indicative of widespread occurrence in the Dry Valleys and their phototrophic ability might contribute to an active community structure.

Acidobacteria is a newly recognized group of microorganisms widespread in soil biotopes. They show diverse physical and chemical properties (Barns et al., 1999) and form a major fraction of non-cultured bacteria (Stevenson et al., 2004). There are only three cultured representatives: *Acidobacterium capsulatum*, *Halophaga foetida* and *Geothrix fermentans*. *A. capsulatum* is a moderately acidophilic aerobic heterotroph (Hiraishi et al., 1995) whereas *H. foetida* and *G. fermentans* are strict anaerobes that ferment aromatic compounds and acetate, respectively (Lonergan et al., 1996; Liesack et al., 1994). The majority of sequences that make up this division are from environmental clones. In fact, of the eight monophyletic subdivisions of the *Acidobacteria*, subdivisions
1, 3, 4 and 6 are represented by environmental clone sequences only (Hugenholtz et al., 1998). The widespread occurrence of members of the Acidobacteria division detected in environmental samples suggests their presence might be ecologically significant. They have been detected in peat bog, acid mine drainage, hot spring, and fresh water lake systems (Stevenson et al., 2004; Barns et al., 1999; Hiraishi et al., 1995).

Actinobacteria are high GC rich Gram-positive heterotrophic bacteria (Basilio et al., 2003). They are one of the better-defined phylogenetic groups and well represented in culture studies. However, a small number of genera are poorly known and are so highly divergent from all the other members of the group that they classify as subclasses (Stackebrandt et al., 1997). These subclasses are small, each represented by a single or a very few strains, and include Rubrobacteridae, Acidimicrobidae, Sphaerobacteridae and Coriobacteridae (Holmes et al., 2000). Rubrobacteria are abundant in a variety of environments and have been documented in agricultural soil (Ueda et al., 1995), peat (Rheims et al., 1996), pasture soil (McCaig et al., 1999) and forest soil (Liesack and Stackebrandt, 1992).

Verrucomicrobia is a newly proposed division of Bacteria represented by a small fraction of isolates: Verrucomicrobium spinosum (Ward-Rainey et al., 1995), four Prosthecobacter species (Hedlund et al., 1997) and three strains of ultramicrobacteria (Janssen et al., 1997). Verrucomicrobia and Prosthecobacter were isolated from fresh water and the ultramicrobacteria were isolated from soil. All of the isolates preferentially use sugars as growth substrates and, like members of the Acidobacteria, are widespread and abundant in the environment, especially soil (Hugenholtz et al., 1998). Of the
several monophyletic subdivisions, only two represent cultured isolates. Interestingly, four species of the Verrucomicrobial genus *Prosthecobacter* might constitute an evolutionary link between the members of Bacteria and Eucarya due to the presence of genes for tubulin production, a cytoskeletal element previously only found in eukaryotes (Jenkins *et al*., 2002; Staley *et al*., 2005).

### 3.4.2 Phylotype distribution

Most prominent within the PENP clone set were two orders of cyanobacteria; *Oscillatoriales* and *Nostocales*. Clone PENP49 showed close relatedness (98%) to *Nostoc* sp. and PENP18 to *Phormidium* sp. Ant-lunch (99%). Clone PENP4 showed similarity to the Oscillatorium, *Phormidium tenue*.

The majority (25%) of the BIS clones grouped with uncultured environmental bacteria (Table 3.2). Nevertheless, the uncultured clones were phylogenetically diverse and primarily mapped to the actinobacteria, acidobacteria and α-proteobacteria. Three of the BIS clones, BIS 31 grouped with unclassified bacteria, rendering their taxonomical status undefined.

Cyanobacteria appeared to be restricted to only the high altitude PENP sample site. The appearance of cyanobacterial phylotypes as major contributors to only one of the three clone libraries supports the consensus that cyanobacterial distribution in the Dry Valleys is non-homogeneous (Vishniac, 1993).
<table>
<thead>
<tr>
<th>Sample site/clones (Accession number)</th>
<th>Phylogenetic group</th>
<th>I.D. of nearest match (Accession number)</th>
<th>% ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>PENP4 (DQ062859)</td>
<td>Cyanobacteria</td>
<td>Leptolyngbya sp. (AY239604)</td>
<td>96</td>
</tr>
<tr>
<td>PENP7 (DQ062860)</td>
<td>Actinobacteria</td>
<td>Arthrobacter agilis (AF134184)</td>
<td>99</td>
</tr>
<tr>
<td>PENP18 (DQ062861)</td>
<td>Cyanobacteria</td>
<td>Phormidium sp. Ant-Lunch (AF263335)</td>
<td>99</td>
</tr>
<tr>
<td>PENP22 (DQ062862)</td>
<td>Cyanobacteria</td>
<td>Uncultured Antarctic bacterium (AF076163)</td>
<td>97</td>
</tr>
<tr>
<td>PENP25 (DQ062863)</td>
<td>Cyanobacteria</td>
<td>Uncultured Antarctic Cyanobacterium (AY151721)</td>
<td>95</td>
</tr>
<tr>
<td>PENP35 (DQ062864)</td>
<td>Environmental samples</td>
<td>Uncultured organism clone (AY897885)</td>
<td>93</td>
</tr>
<tr>
<td>PENP37 (DQ062865)</td>
<td>Bacteroidetes</td>
<td>Uncultured Bacteroidetes (AY689627)</td>
<td>97</td>
</tr>
<tr>
<td>PENP40 (DQ062866)</td>
<td>Actinobacteria</td>
<td>Uncultured Actinobacterium clone FBP460 (AY250884)</td>
<td>99</td>
</tr>
<tr>
<td>PENP42 (DQ062867)</td>
<td>Bacteroidetes</td>
<td>Uncultured bacterial clone (AJ290025)</td>
<td>96</td>
</tr>
<tr>
<td>PENP48 (DQ062868)</td>
<td>Verrucomicrobia</td>
<td>Uncultured soil bacterial clone C019 (AF013522)</td>
<td>95</td>
</tr>
<tr>
<td>PENP49 (DQ062869)</td>
<td>Cyanobacteria</td>
<td>Nostoc sp. (AY566855)</td>
<td>98</td>
</tr>
<tr>
<td>PENP50 (DQ062870)</td>
<td>Bacteria, environmental sample</td>
<td>Uncultured bacterial clone csbio160368 (AY187335)</td>
<td>98</td>
</tr>
<tr>
<td>PENP54 (DQ062871)</td>
<td>Actinobacteria</td>
<td>Uncultured Actinobacterium (AY690206)</td>
<td>92</td>
</tr>
<tr>
<td>PENP55 (DQ062872)</td>
<td>Bacteroidetes</td>
<td>Uncultured Bacteroidetes clone VC5 (AY211071)</td>
<td>96</td>
</tr>
<tr>
<td>PENP62 (DQ062873)</td>
<td>Acidobacteria</td>
<td>Uncultured Acidobacterium (AY571794)</td>
<td>92</td>
</tr>
<tr>
<td>PENP68 (DQ062874)</td>
<td>Bacteria, environmental samples</td>
<td>Uncultured bacterium (AY662047)</td>
<td>93</td>
</tr>
<tr>
<td>PENP75 (DQ062875)</td>
<td>Bacteria, environmental sample</td>
<td>Uncultured bacterial clone sipK9 (AJ307936)</td>
<td>93</td>
</tr>
<tr>
<td>PENP76 (DQ062876)</td>
<td>Cyanobacteria</td>
<td>Oscillatoria sp. Ant-G16 (AF26333)</td>
<td>99</td>
</tr>
<tr>
<td>PENP77 (DQ062877)</td>
<td>Cyanobacteria</td>
<td>Trichormus azollae (AJ630454)</td>
<td>95</td>
</tr>
<tr>
<td>PENP78 (DQ062878)</td>
<td>Cyanobacteria</td>
<td>Uncultured Antarctic cyanobacterium (AY151722)</td>
<td>97</td>
</tr>
<tr>
<td>MVG1 (DQ062879)</td>
<td>Actinobacteria</td>
<td>Geodermatophilus sp. G1S (X92364)</td>
<td>94</td>
</tr>
<tr>
<td>MVG2 (DQ062880)</td>
<td>Bacteria, environmental sample</td>
<td>Uncultured bacterial clone C-F-15 (AF443586)</td>
<td>96</td>
</tr>
<tr>
<td>MVG5 (DQ062881)</td>
<td>Actinobacteria</td>
<td>Uncultured Actinobacterium (AY690226)</td>
<td>94</td>
</tr>
<tr>
<td>MVG9 (DQ062882)</td>
<td>Verrucomicrobia</td>
<td>Uncultured bacterial clone C019 (AF013522)</td>
<td>96</td>
</tr>
<tr>
<td>MVG14 (DQ062883)</td>
<td>Actinobacteria</td>
<td>Bacterium Ellin504 (AY960767)</td>
<td>95</td>
</tr>
<tr>
<td>MVG15 (DQ062884)</td>
<td>Actinobacteria</td>
<td>Conexibacter woesi (AJ440237)</td>
<td>96</td>
</tr>
<tr>
<td>MVG16 (DQ062885)</td>
<td>Acidobacteria</td>
<td>Uncultured Acidobacterium (AY571792)</td>
<td>98</td>
</tr>
<tr>
<td>MVG18 (DQ062886)</td>
<td>Bacteria, environmental sample</td>
<td>Uncultured bacterial clone FBP241 (AY250867)</td>
<td>95</td>
</tr>
<tr>
<td>MVG19 (DQ062887)</td>
<td>Bacteria, environmental sample</td>
<td>Uncultured bacterial clone SO27 (AF013554)</td>
<td>96</td>
</tr>
<tr>
<td>MVG20 (DQ062888)</td>
<td>Bacteria, environmental sample</td>
<td>Uncultured bacterial clone D132 (AY274138)</td>
<td>95</td>
</tr>
<tr>
<td>MVG21 (DQ062889)</td>
<td>Acidobacteria</td>
<td>Uncultured Acidobacterium (AY571794)</td>
<td>94</td>
</tr>
<tr>
<td>MVG23 (DQ062890)</td>
<td>Actinobacteria</td>
<td>Uncultured Actinobacterium (AF234135)</td>
<td>94</td>
</tr>
<tr>
<td>MVG24 (DQ062891)</td>
<td>Chloroflexi</td>
<td>Uncultured bacterium (AY922044)</td>
<td>98</td>
</tr>
<tr>
<td>MVG25 (DQ062892)</td>
<td>Acidobacteria</td>
<td>Uncultured Acidobacterial clone BAC-220H (AY214900)</td>
<td>92</td>
</tr>
<tr>
<td>MVG28 (DQ062893)</td>
<td>Bacteroidetes</td>
<td>Uncultured Bacteroidetes (AY921683)</td>
<td>96</td>
</tr>
<tr>
<td>MVG31 (DQ062894)</td>
<td>Actinobacteria</td>
<td>Actinobacterial strain PB90-4 (AJ229240)</td>
<td>96</td>
</tr>
</tbody>
</table>
There was no visible evidence of an adjacent cyanobacterial source (either algal mat material or hypolithic sites) at the time of sampling. The cyanobacterial phylogenotypes identified in sample PENP (putatively members of the genera *Nostoc, Phormidium* and *Oscillatoria*) are also not indicative of cryptoendolithic origin. The cyanobacterial components of these communities are principally members of the *Gloeocapsa, Anabaena, Chroococcidiopsis, Hemia chloris, Heterococcus* and *Lyngbya* (Nienow and Friedmann,
1993). It is also notable that the hills on the northern flanks of the Miers Valleys appear to lack the sandstone and marble rock strata that principally harbour cryptoendolithic communities (Nienow and Friedmann, 1993). It is thus suggested that the cyanobacterial signals identified in sample PENP constitute free-living cyanobacteria. Although the PENP sample showed a very low water content (0.7% wt. \( \text{H}_2\text{O} \)) as determined by dry weight measurements, the data might suggest that such values are a poor determinant of cyanobacterial distribution. Both meteorological studies (Wynn-Williams, 1990) and surveys of lichen distribution (Pannewitz et al., 2003) indicate that water availability is strongly altitude dependent. Relative humidity measurements from altitudinal transects in the Taylor and Wright Valleys (Horowitz et al., 1972) suggest that atmospheric humidity may be an important determinant of water availability. However, a detailed survey of cyanobacterial distribution in relation to environmental and physical factors (altitude, soil type, soil water content, soil humidity, and atmospheric humidity) is required to more fully understand the factors that dictate the distribution of soil-borne cyanobacterial populations.

Actinobacterial signals were frequent in all three sites, a result that is consistent with the widespread distribution of actinobacteria in various soil types (Basilio et al., 2003). The high frequency of actinobacterial phylotype signals (Table 2) suggests that this group contributes a significant fraction of the soil microbial population, as shown by others (Holmes et al., 2000). However, few of the actinobacterial sequences could be matched to known phylotypes at >95% homology, suggesting that a substantial pool of novel uncultured psychrotrophic actinobacterial species remain to be identified. Given the industrial importance of this group of organisms (Bunch, 1998; Jorgensen et al., 2001),
this observation adds some weight to the importance of developing new isolation strategies (Janssen et al., 2002).

The acidobacterial signals were the most highly populated clade in the nutrient-rich BIS sample (Fig 3.5). 18% of all phylotypic signals obtained were assigned as ‘uncultured’, and were prevalent in all three sites (Fig 3.5). Phylotypic signals identified as members of the acidobacteria were major components of both the desiccated mid-altitude (MVG) and C-enriched (BIS) samples.

While PCR-based analyses of microbial diversity are widely acknowledged to be less than fully representative due to biases introduced by factors such as extraction efficiency (Farely et al., 1995), and hybridisation specificity (Chandler et al., 1997; Suzuki and Gionvannoni, 1996) it is also accepted that community composition is more effectively elucidated by this method than by conventional culturing. The results indicate that a diverse range of prokaryote phylotypes is present in Antarctic Dry Valley cold desert mineral soils. It was noted, as have others (Lipson and Schmidt, 2004), that the highest proportion of sequences identified fall into the so-called ‘unculturable’ class. A significant proportion of the sequences obtained showed relatively low homology to extant sequences (<95%), suggesting that the mineral soils of the Dry Valleys represent a substantial pool of novel species and/or genera. Calculations of diversity indices such as the Shannon-Weiner index support previous suggestions that all three sample sites harbour relatively low species diversity (Table 3.3) (Mckay, 1993; Vincent, 1988). The calculated values, H = 1.598, 1.331, and 1.238 for PENP, MVG, and BIS respectively, are substantially lower than would be expected for temperate soil biotopes, which typically
have values of between 6 and 7 (Dunbar et al., 2000; Hughes et al., 2001). Evenness (E) values calculated are very low for all three samples. At an optimum even distribution, a value of 1 is given, which implies that the mineral soils indicate a rather high degree of unevenness (Zar, 1999). The estimated species richness (S) for all thee samples is low (Table 3.3), supporting the consensus that extreme environments exhibit low species diversity (Sandaa et al., 1999). Accurately determining species richness is also hindered by the fact that large numbers of the phylotypes are not identifiable to species level and that the collectors curves have not reached a plateau (Schloss and Handelsman, 2004).

Table 3.3 Comparison of phylotype richness, diversity and evenness values for the Antarctic mineral soil PENP, MVG and BIS bacterial communities.

<table>
<thead>
<tr>
<th>Index</th>
<th>PENP</th>
<th>MVG</th>
<th>BIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>S(^a)</td>
<td>49</td>
<td>50</td>
<td>57</td>
</tr>
<tr>
<td>H(^b)</td>
<td>1.598</td>
<td>1.331</td>
<td>1.238</td>
</tr>
<tr>
<td>E(^c)</td>
<td>0.410</td>
<td>0.340</td>
<td>0.306</td>
</tr>
</tbody>
</table>

\(^a\) Phylotype richness, S, was calculated as the percentage of the total number of distinct ARDRA patterns to clones.

\(^b\) Shannon-Weiner diversity index (Dunbar et al., 2000) was calculated as follows: \(H = -\sum(\rho_i)(\log_2 \rho_i)\), where \(\rho_i\) is the proportion of an unique ARDRA pattern relative to the sum of all patterns.

\(^c\) Evenness (Dunbar et al., 2000) was calculated from the Shannon-Weiner diversity function as follows: \(E = H/H_{\text{max}}\) where \(H_{\text{max}} = \log_2(S)\)

Richness as indicated by the collection curve (Fig. 3.5) equals twenty, since sampling was continued until only twenty unique clones had been identified for each of the three soil samples. Additional sampling should increase phylotype richness until the approximated values for species richness in Table 3.3 are reached.
Fig 3.6  Relative percentage distribution of the phylotypes identified from the Antarctic Dry Valley mineral soils.
Cyanobacteria and actinobacteria, two of the dominant phyla identified in the oligotrophic mineral soil samples (MVG and PENP), are well represented in early culture-dependent studies of Antarctic Dry Valley mineral soil microbiology (Cameron et al., 1972). However, readily isolatable taxa, such as *Achromobacter, Bacillus, Corynebacterium, Micrococcus, Planococcus* and *Pseudomonas* (Cameron et al., 1972), are not represented in any of the three 16S rDNA clone libraries. Conversely, groups such as Acidobacteria, Verricimicrobia and Bacteroidetes, which are absent from both historical and recent culture-dependent studies appear to be relatively common on the basis of 16S clone distribution. These groups have also been identified as significant components of the microbial communities in two previous molecular phylogenetic studies of specific Antarctic habitats: the cryptoendolithic (de la Torre et al., 2003) and cryoconite hole communities (Christner et al., 2003).

It has not been established whether the desiccated mineral soils of the Dry Valleys of Antarctica constitute stable microbial communities, or merely assemblages of organisms attached to mobile particulates. The mobility of the mineral soils of the Dry Valleys (particularly during the windy austral winter seasons (Wynn-Williams, 1990) argues for the latter. A stable community is presumed to contain the elements of energy capture and turnover of primary nutrient components. Two of the three sites analysed in this study may thus represent putative communities: the phototrophic and N$_2$-fixation capacity of the cyanobacterial phylotypes identified in sample PENP, and the presence of an exogenous nutrient source in sample BIS offer the elements required to maintain a trophic structure. This cannot be said of the desiccated mineral soils of much of the Dry Valleys (represented by sample MVG in this study). Although certain phylotypic groups
identified in sample MVG are known to include chemoautotrophic species (e.g., the β-proteobacteria (Hoeft et al., 2004; Sinigalliano et al., 2001)) the presence of specific chemoautotrophic species cannot be directly inferred from these results. The demonstration of a community structure may nevertheless be inferred by identification of both the presence and activity of key genes, enzymes and/or processes. One possible way to demonstrate metabolically active communities is the isolation of environmental RNA. RNA is an ideal biomarker because it serves as a function of DNA copy number and has a high turnover and would serve as a good indicator of cellular activity (Manefield et al., 2002). RNA isolation is, however only indicative of the genes expressed at a given time and does not provide information regarding the organism of origin. More focused analysis of metabolically active communities could involve stable-isotope-labeled (SIP) substrates (Pelz et al., 2001). With this approach specific genes and organisms involved in the assimilation of the chosen substrates can be identified (Boschker and Middelburg, 2002).

16S rDNA analysis performed on the extracted metagenomic DNA from all three sample sites allowed the documentation of several phylotypes (Fig. 3.6). The probability of having isolated DNA only from live and/or active cells needs some consideration. It is likely that the DNA isolated for 16S PCR analysis might also have originated from sources other than metabolically active cells and could include DNA from VNBC individuals, extracellular DNA (naked DNA) and randomly wind-dispersed cells or particles. Methods typically applied to extract DNA from environmental soil samples are designed to extract total DNA (extra- and intracellular) although robust methods such as bead beating lessens both quality and quantity of the extracellular DNA isolated compared to more
gentle chemical extraction procedures (England et al., 2004). Upon cell death and lysis, DNA is released into the environment and becomes available to bacterial nucleases. Despite the presence of nucleases, naked DNA persists in soil for varying lengths of time (Recorbet et al., 1993). Exogenous DNA is stabilized through the formation of DNA-soil particle complexes by adsorbing to negatively charged soil and organic matter forming cation bridges that shield DNA from degradation (Lorenz et al., 1994). Other factors conducive to the preservation of naked DNA within the Antarctic mineral soils are the extreme cold (especially in the permafrost layer (Stokstad, 2003)), low water availability and high levels of salts. The persistence of naked DNA in soil has been demonstrated in both temperate (forest soils, sediments and caves) and Arctic soils (Stokstad, 2003). Naked DNA survival was demonstrated by seeding Dry Valley mineral soil with S. epidermidis genomic DNA, which remained detectable for 23.5 weeks (Ah Tow and Cowan, 2005). England and co-workers (2004) demonstrated naked DNA survived for three months in forest soils and long-term survival of naked DNA was indicated by recovering plant and animal DNA from Siberian sediments dating back to 400 000 and 30 000 years, respectively, (Stokstad, 2003).

If the survival of naked DNA within Antarctic soils is as persantant as in other cold environments such as the Siberian sediments, it is likely that the 16S rDNA signals observed in this study might be from both active cells and from naked DNA. The presence of culturable organisms has been shown by previous culture-based analysis of the mineral soils (Friedmann, 1993). The possibility that naked DNA from 16S rDNA sequences has been cloned can not be ruled out, which indicates the necessity for further investigation to determine the ratio, if any, of naked DNA to culturable cells.
The phylotypes identified within this study of Antarctic mineral soils depict preliminary findings, and additional sampling at various distance intervals from the original sample sites would give a better indication as to the phylotype composition and richness within the mineral soils. Sampling at different transects and if possible at intervals throughout the year will also yield valuable information regarding the distribution of the phylotypes. Dominant and indigenous phylotypes to each mineral soil should be perpetual throughout the year whereas exogenous phylotypes, for instance wind-blown individuals or communities, would be periodic.

Use of SIP (Pelz et al., 2001) and BrdU (Borneman, 1999) would indicate active communities within the soils. This would aid in differentiating between those phylotypes forming part of the active bacterial community and derived from dead cells and/or extracellular DNA.
Chapter 4
Metagenomic library construction

4.1 Introduction
The cloning of microbial DNA directly from the environment to screen for the presence of desired enzyme activity has become a useful tool in discovering new biocatalysts. DNA libraries have been constructed from various diverse environmental regions such as the Arctic, hotsprings, deserts and both fresh and sea water (Cottrell et al., 1999; Lee et al., 2004; Kim et al., 2006). The libraries contain the collective microbial DNA complement, referred to as the metagenome (Handelsman et al., 1998), of each specific environment. The high level of diversity within the metagenome provide an almost inexhaustible source of new enzymes (Cowan, 2000).

As the minimal requirements for gene expression are a promoter for transcription and a ribosome binding site upstream of the translation initiation codon, the construction of metagenomic libraries in ordinary cloning plasmids such as pUC19 or pSK+ has proved successful. Small insert libraries (2-8 kb) are typically constructed in plasmids such as pSK+, which contain strong inducible promoters such as lacI and T7 and/or T3 upstream of the multiple cloning site (MCS) and support expression in cases where the insert does not contain a promoter (Henne et al., 2000). Shotgun cloning is random and inserts of varying sizes are cloned unidirectionally, which make systems with strong inducible promoters on both DNA strands attractive prospects, thus ensuring transcription of genes on both the positive and negative strands. Inducible promoters however, are not
essential as successful expression screening of a metagenomic library constructed in 
pUC19 yielded 9 positive lipolytic clones (Ranjan et al., 2005).

Although small insert libraries have proved successful, large insert libraries (40-300 kb) 
are preferred. In these systems, specialized plasmids such as cosmids, fosmids and BACs 
which are able to maintain the integrity of large inserts, are used (Rondon et al 2000; 
Lee et al., 2004). Large insert libraries are more informative, allowing access to 
neighbouring genes or cis-elements required for effective expression of target genes, 
which can easily be missed in small insert libraries. They are also more likely to provide 
insight into the evolutionary origin of the functional gene using 16S rDNA analysis for 
example (Streit and Schmitz, 2004). Expression screening using large insert libraries is 
usually entirely reliant on native promoters due to the size of the inserts. The only 
drawback is usually associated with the host cell not recognizing the heterologous 
transcription signals. Another advantage of large insert libraries is the high level of 
sequence coverage, which might allow for the reconstruction of whole novel genomes 
(Venter et al., 2004).

Expression screening for the presence of lipolytic genes has been conducted in a number 
of studies (Henne et al., 2000; Lee et al., 2004; Ranjan et al., 2005). Bacterial lipases 
are typically extracellular enzymes, which enables activity screening on agar plates using 
lipid substrates (Kouker and Jaeger, 1987; Feller et al., 1991; Henne et al., 2000). 
Activity based screens for lipolytic enzymes involve nutrient media and a substrate such 
as tributyrin for esterases or olive oil (trioleoglycerol) for lipases. The difference in 
substrate selection is based upon the acyl chain length of the glyceroesters, where
longer chain lengths (>10) are only hydrolyzed by lipases and shorter chains (<10) are hydrolyzed by both lipases and esterases. Tributyrin causes agar to become opaque and hydrolysis of positive clones produces a distinct zone of clearance as lipolytic enzymes hydrolyze the lipid substrate. Rhodamine dye-containing agar plates are pink. Bacterial clones positive for lipase hydrolyze the trioleolglycerol substrate and the resulting uranyl-fatty acid ion forms a complex with the cationic rhodamine B, which is detected as an orange fluorescence surrounding the colony under UV light (Kouker and Jaeger, 1987).

4.2 Aim

The principle aim of the work reported in this chapter was to detect lipase and esterase genes using expression screening.

To achieve this, the following objectives were set forth:

- Isolation of pure, intact high molecular weight DNA from MVG and PENP
- Partial digestion using a tetranucleotide-specific enzyme
- Construction of a PCR®-XL-TOPO® small insert metagenomic library from two of the Dry Valley soils
- Functional screening of the library for lipase activity

4.3 Results

4.3.1 Partial digestion and cloning of metagenomic DNA

High molecular weight DNA for tetranucleotide-specific digestion was obtained from soil samples MVG and PENP using the modified Zhou protocol (Stach et al., 2001) as outlined
in Section 2.3.1. Routinely, the Zhou DNA extraction protocol yielded more DNA per gram of soil compared to the Miller or BIO 101 DNA extraction methods. The DNA isolated using the Zhou protocol showed little or no degradation, whereas both the Bio 101 and Miller protocols yielded less DNA which was more sheared. This prompted the use of Zhou-extracted DNA for library construction.

Fig. 4.1 Metagenomic DNA from Dry Valley mineral soils. Lane 1, DNA molecular weight marker; Lane 2, metagenomic DNA extracted using the Zhou protocol; Lane 3, metagenomic DNA extracted using the Bio101 protocol; Lane 4, metagenomic DNA extracted using the Miller protocol. The sizes of the molecular weight marker, phage lambda DNA digested with PstI, are indicated to the left of the figure.

Trial digestions using three different tetranucleotide-specific enzymes (AfaI, AluI, HaeI) was performed at time intervals varying between 40 and 60 min to determine optimal restriction conditions. AluI yielded a smearing pattern ranging from intact high molecular weight DNA to ~300 bp. However, smearing below 2 kb appeared less prominent, with much of the partially digested DNA apparently between 2 kb and 10 kb (Fig. 4.2). Following AluI digestion the endonuclease reactions were electrophoresed in a 1.0%
agarose gel. DNA ranging in size from 2 kb to 10 kb was excised and recovered using GFX columns. The purified DNA fragments were A-tailed, dephosphorylated and finally cloned into the PCR®-XL-TOPO® cloning vector as outlined in Chapter 2.

To verify the cloning, a small fraction (5 μl) of the total transformation reaction (1 ml) was plated onto kanamycin selective media followed by incubation at 37°C overnight. The remainder of the transformation mix was stored at 4°C until cloning efficiency could be determined. Following overnight incubation the colonies were tallied and the theoretical number of cfu.ml⁻¹ was calculated. The library size was estimated at 90700 clones.

![Image of agarose gel](image)

**Fig. 4.2**  
*Alu*I restriction endonuclease digestion of MVG and PENP metagenomic DNA. Lane 1, DNA molecular weigh marker; Lanes 1, 2 and 3, MVG metagenomic DNA; Lanes 4 and 5, PENP metagenomic DNA. The sizes of the molecular weight marker, phage lambda DNA digested with *Pst*I, are indicated to the left of the figure.
4.3.2 Sequence space coverage

Assessment of the average insert size of the metagenomic library was performed using restriction endonuclease digestion. Thirty colonies were randomly selected from agar plates following library amplification and inoculated into LB-broth supplemented with the appropriate antibiotic. Plasmid DNA was extracted as described in Section 2.15, followed by restriction endonuclease reactions with EcoRI (Section 2.13). Insert sizes ranged from 1.7 kb to 5 kb with only two clones showing inserts at 0.8 kb (Fig. 4.3). Based on the restriction analysis, the average insert size was determined at 3 kb and the library size was calculated to contain 272 Mb of metagenomic DNA. Assuming the minimal number of genomes (2000) (Sandaa et al., 1999) present in the mineral soil sample and an average genome size of 4 Mbp, the library represented ~ 3.4% of the total metagenomic fraction.

Fig. 4.3  Restriction analysis of PCR®-XL-TOPO® clones constructed by ligating partially digested metagenomic DNA fragments. Lane 1, DNA molecular weight marker; Lanes 2 through 10, recombinant PCR®-XL-TOPO® digested with EcoRI; Lane 11, DNA molecular weight marker. The sizes of the molecular weight markers, phage lambda digested with PstI, and EcoRI and HindIII are displayed to the left and right respectively.
4.3.3 Library screening

4.3.3.1 Activity-based screening

To determine whether functional lipase or esterase genes were present in the metagenomic library, expression screening of the cloned inserts was performed on 15 cm indicator plates, prepared as described in Sections 2.2.5 and 2.2.6. Two copies of the library were separately plated in aliquots of 60 µl (~2000 clones) onto esterase and lipase indicator plates, and incubated at 37°C until colonies were visible. The plates were then removed and incubated at 16°C for a period of 7 days. During the incubation the plates were routinely monitored for the presence of clear zones surrounding the colonies on esterase-specific plates and orange halos around colonies on the lipase-specific plates. Following incubation for 7 days, no esterase or lipase activity was detected and incubation was continued for another 5 days with no further change.

4.3.3.2 PCR-based screening of the metagenomic library

In order to probe the basis of the failure to detect functional lipolytic expression, PCR analysis of the library was performed following the absence of both esterase and lipase activity on indicator plates. A single copy of the complete library was subjected to alkaline lysis (Section 2.16) to isolate the plasmid DNA for PCR analysis. PCR was performed using degenerate primers LipF and LipR2 (Section 2.23). No PCR amplicons were detected following PCR (Fig. 4.4).

4.4 Discussion

To construct an effective library, sufficient high molecular weight DNA needs to be isolated from the environment or organism of interest. For this study, DNA was extracted
Fig. 4.4 Agarose gel electrophoretic analysis following PCR of the PCR\textsuperscript{®}-XL-TOPO\textsuperscript{®} metagenomic DNA library using primers LipF and LipR2. Lane 1, DNA molecular weight marker; Lane 2 through 5, metagenomic library template; Lane 6, negative control containing no template DNA; Lane 7, positive control using a recombinant plasmid harbouring a full-length lipase gene of \textit{B. multivora}; Lane 8, Negative control containing only plasmid DNA. The sizes of the molecular weight marker, phage lambda digested with \textit{PstI}, are displayed to the left of the figure.

from oligotrophic soils using three different techniques, and the quality and yield compared. Two methods, Miller and Bio 101, were mechanical extraction methods, whereas the third (Zhou) was a chemical extraction. Analysis of extracted DNA following agarose electrophoresis revealed that the chemical lysis produced intact high molecular weight DNA consistent with other studies (Stach \textit{et al.}, 2001; Niemi \textit{et al.}, 2001). The robust mechanical extraction methods yielded fragmented DNA with little or no high molecular weight DNA. DNA extracted using the Zhou method was used in partial restriction endonuclease digestion.
Library construction proved successful, as was evident from the average insert size (Fig 4.2), although the number of clones generated was lower than generated in other studies (Henne et al., 2000; Majerník et al., 2001; Ranjan et al., 2005). No lipase activity was detected following activity-based screening, which could be attributed to a number of factors.

One such factor might be low nucleotide coverage of the metagenome, implying that the fraction of DNA captured during library construction was not large enough to include a functional lipase gene, although the average insert size was adequate to represent full-length genes (3 kb). Statistically, libraries of $10^7$ clones need to be screened to ensure a positive hit (Gabor et al., 2004), whereas the metagenomic library only constituted 90700 clones ($<10^4$). For metagenomic libraries, the number of clones ($N_p$) required to recover a target gene at least once with the probability $P$ can be derived from the binomial distribution:

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

Where $I$ is the average insert size and $X$ the size of the gene of interest. $G$ represents the average genome size present in the sample and $z$ is the number of genomes assuming even distribution (Gabor et al., 2004). The correction factor $c$ reflects whether expression of the insert is independent from the plasmid (relies on a native promoter and ribosome binding site) or dependent on a plasmid-borne promoter and/or ribosome-binding site (rbs).
The distribution of species \( (z) \) as given by the equation assumes that all species present are represented equally. This assumption however, might lead to a gross underestimation of the biodiversity within a given environment as shown by Gans and co-workers (2005). By using reassociation kinetics two statistical models were proposed, one assuming equal distribution and the second assuming uneven distribution of species. Uneven distribution calculations indicated higher species abundance compared to even distribution figures. Thus, if the calculated species estimates holds true for uneven distribution, the value of \( z \) according to Gabor et al. (2004) should be much larger implying an even larger theoretical \( N_p \).

Assuming the following for the Antarctic Dry Valley metagenomic library:

Average insert size \( (I) = 3 \) kb; Average lipase gene size \( (X) = 1.5 \) kb; Average genome size \( (G) = 4 \) Mbp; \( c = 1; z = 2000 \); with a probability \( P \) of 0.9 of achieving a positive hit during expression screening, the theoretical library size should constitute \( 1.2 \times 10^7 \) clones, a figure substantially larger than the current library size. The estimated species size \( (z) \) of 2000 is based on findings by Sandaa et al. (1999), where they indicating that extreme metal contaminated soils contain approximately 2000 species. Even if the figure of 2000 species for the Dry Valley soils is overestimated ten fold, then the number of theoretical clones required for a 0.9 probability of obtaining a positive hit still far exceeds the current size of the library.

Although the PCR®-XL-TOPO® vector harbours an inducible T7 promoter on the negative strand, induced transcription was never performed. Instead, expression was dependent on the native promoter and rbs. The use of independent gene expression allowed for
lower levels of gene expression as high levels of subfamilies I and II lipases have been shown to be toxic to heterologous hosts (Bell et al., 2002). Additionally, successful gene expression and isolation of lipases was obtained from small insert libraries (Ranjan et al., 2005).

The metagenomic DNA used in constructing the library might not only be from bacterial origin and could also represent DNA from both fungal and Archaeal taxa. Although the presence of eukaryotes in the Antarctic mineral soils is limited to a few lichens, yeasts, protozoa and bacteria (Friedmann, 1993; Ah Tow and Cowan, 2004), the probability of having included eukaryotic DNA in the metagenomic library cannot conclusively be ruled out. Contrary to other studies (Brambilla et al., 2001), Archaeal-specific PCR did not yield any signals, which might imply that Archaea are either absent from the mineral soils analysed or, alternatively, Archaea are present in levels below detection. The likely presence of DNA sources other than that of bacteria could imply that less than 3.4% of the metagenomic library represents prokaryotic DNA.

To verify whether the absence of lipolytic activity could be attributed to one of the factors mentioned above, PCR analysis was performed on plasmid DNA extracted from an amplified copy of the library. PCR has the advantage in that it does not require complex regulatory mechanisms associated with lipolytic enzyme production, only the presence of a nucleotide sequence. Using degenerate lipolytic primers in PCR analysis of the library enabled the detection of a variety of lipolytic genes. Following PCR analysis of the metagenomic library no signals were obtained in any of the PCR reactions and it was concluded that no lipolytic genes were cloned from the environment.
To ensure sufficient coverage and the probable successful isolation of functional genes using expression screening, the use of fosmids, cosmids or BACs should be considered. These cloning systems allow for large inserts (>40 kb) and provide access to up- or downstream genes or elements, such as chaperones, for functional secretion of target enzymes. This especially holds true for family I subfamily I.1 and I.2 lipases (Arpigny and Jaeger, 1999), as lipase specific foldases (lif) are essential for correct folding and subsequent expression of its cognate lipase (Rosenau et al., 2004). Gene expression is mediated by the native promoter and copy number is usually low (single copy per cell) leading to low levels of enzyme, which is ideal, especially as strong inducible promoters combined with high copy numbers could lead to stress on the host cell (Glick, 1995). In addition, the net effect of large insert libraries further decrease $N_p$ due to larger inserts and independent expression of target genes (Gabor et al., 2004).
Chapter 5
Prospecting for genes conferring lipolytic activity in Antarctic mineral soils

5.1 Introduction

Using metagenomics to discover novel genes, pathways and bacteria instead of studying pure isolated cultures from an environment, allows more efficient access to the available information contained within the nucleic acid gene pool. Additionally, metagenomics could allow for improved access to novel genes and sequence space.

The lack of sequence space coverage is clearly demonstrated by prokaryote community analysis. During early culture-based techniques it was observed that very few of the prokaryotes were cultured compared to the cells observed using other techniques such as staining, microscopy and reassociation kinetics (Torsvik et al., 1990; Amann et al., 1995). Molecular phylogenetic studies targeting the 16S rDNA gene greatly expanded the then narrow window of available prokaryotes allowing both documentation of and access to novel bacteria (Amann et al., 1995). Although molecular techniques continuously expand sequence space, the accessible sequence space remains a fraction when weighed against the total estimates of approximately $10^{30}$ individual microorganisms (Whitman et al., 1998; Gans et al., 2005).
5.2 Aims

The primary aim of this chapter was to investigate whether degenerate primers, based on known lipolytic gene sequences, would be able to detect putative cold-active lipase genes within metagenomic DNA samples.

To successfully achieve this, the following techniques were used:

- The design of degenerate primers to specifically amplify gene fragments harbouring α/β-hydrolase motifs from metagenomic DNA isolated from MVG and PENP samples.
- Cloning and bioinformatic analysis of the gene fragments to ascertain possible novelty.

5.3 Results

5.3.1 Degenerate primer design and PCR

5.3.1.1 Degenerate PCR using primer set OXF1 and ACR1

Probing for lipolytic genes within the Antarctic cold desert mineral soils, involved the use of degenerate primers. Three different sets of degenerate primers were used to target and PCR-amplify partial lipase gene fragments (Table 2.1). One of the primer sets used (OXF1 and ACR1) were originally designed by Bell and co-workers (Bell et al., 2002) to target lipase genes within a thermophilic environment. Primers OXF1 and ACR1 were designed in silico using CODEHOP software (Rose et al., 1998). This resulted in a primer pair with a 3’ degenerate core and a 5’ non-degenerate consensus clamp which would
allow higher annealing temperatures and better primer utilization. Initial PCR trials yielded very faint bands following agarose gel electrophoresis (Section 2.22). Aliquots of the PCR reactions were subjected to additional rounds of thermocycling to enhance the PCR-signal, after which the PCR amplicons were cloned. Fourteen colonies displaying a Gal⁺ phenotype on IPTG/X-gal supplemented selective plates were selected and screened. To verify whether all fourteen colonies were recombinant, the pGEM® – T easy plasmid DNA was isolated and subjected to restriction endonuclease digestion using EcoRI. All fourteen clones were recombinant, containing inserts varying in size between 340 bp and 400 bp (Fig. 5.1). Only one of the clones contained an insert harbouring putative α/β-hydrolase characteristics (Fig. 5.1, lane 3). The resultant recombinant clone was designated Lip3. Further attempts to amplify partial fragments failed and it was concluded that the 512-fold degeneracy of OXF1 and ACR1 might have been too high.

Fig. 5.1 Restriction endonuclease digestion of recombinant pGEM® – T easy plasmids. Lane 1 and 13, DNA molecular weight markers; Lanes 2 to 12 and 14 to 16, recombinant pGEM® – T easy plasmids digested with EcoRI. The sizes of the molecular weight marker, lambda DNA digested with PstI, are displayed to the left of the figure.
5.3.1.2 Design and PCR using primer sets LipF and LipR and LipF and LipR₂

Based on the fact that primer sets OXF1 and ACR1 did not efficiently amplify additional partial lipase gene fragments, a new set of degenerate primers, LipF and LipR, was designed. Design of degenerate primers involved the amino acid alignment of various true-lipase genes (Family I, subfamilies I-VI) (Arpigny and Jaeger, 1999) to identify the most conserved regions. Regions showing the highest level of conservation were identified as the nucleic acids spanning the oxyanion hole and active site of the lipolytic enzymes (Fig 5.2). As the nucleotide sequence of both the oxyanion hole and active site of all the true-lipases are not totally conserved (Table 5.1), both primers LipF and LipR were designed to have a 64-fold degeneracy (Table 1.1) to allow access to a significant portion of sequence space. Primers LipF and LipR were used to PCR amplify partial lipolytic fragments from community DNA isolated from samples MVG and PENP. Although agarose gel analysis revealed improved levels of amplification following PCR analysis with primer set LipF and LipR, multiple PCR amplicons and significant background were observed (Fig. 5.3). Following exhaustive PCR attempts it was determined that LipR was not sufficiently specific, which might have lead to mispriming at the active site causing multiple PCR amplicons. Primer LipR was redesigned to have a 32-fold degeneracy and designated LipR₂. Primer pair lip LipF and LipR₂ yielded a single band of approximately 300 bp following PCR and gel electrophoresis (Fig. 5.4). The bands were recovered and cloned into plasmid pTZ57R/T (Fermentas). Eight recombinant clones with were identified based on blue/white selection and restriction analysis revealed fragments varying in length from 240 bp to 300 bp (Fig. 5.5). To determine the origin of the cloned DNA fragments, all
Fig. 5.2  Clustal W amino acid alignment of 18 true lipase sequences from family I, subfamilies I and II as defined by Arpigny and Jaeger (1999). Similar amino acids are highlighted in yellow and identical amino acids are highlighted in blue. The most conserved regions, the oxyanion hole (aa 6 – 12) and active site (aa 113 – 118), are indicated by arrows and served as the core regions for primer design.
Table 5.1  Conserved motifs of the oxyanion hole and active sites of true lipases from Family I, subfamily I-VI.

<table>
<thead>
<tr>
<th>Bacteria</th>
<th>Accession number</th>
<th>Oxyanion hole (Forward primer)</th>
<th>Active site (Reverse primer)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subfamily I</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. aeruginosa</em></td>
<td>D50587</td>
<td>PIVLAHG</td>
<td>GHSQGG</td>
</tr>
<tr>
<td><em>P. fluorescence C9</em></td>
<td>AF031226</td>
<td>PLVLVPG</td>
<td>GHSQGS</td>
</tr>
<tr>
<td><em>V. cholerae</em></td>
<td>X16945</td>
<td>PIVLVHG</td>
<td>GHSQGA</td>
</tr>
<tr>
<td><em>P. fragi</em></td>
<td>X14033</td>
<td>PILLVHG</td>
<td>GHSQGS</td>
</tr>
<tr>
<td><em>P. wisconsinensis</em></td>
<td>U88907</td>
<td>PIVLVHG</td>
<td>GHSQGS</td>
</tr>
<tr>
<td><em>P. vulgaris</em></td>
<td>U33845</td>
<td>PIVLVHG</td>
<td>GHSQGP</td>
</tr>
<tr>
<td><strong>Subfamily II</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. glumae</em></td>
<td>X70354</td>
<td>PVILVHG</td>
<td>GHSQGG</td>
</tr>
<tr>
<td><em>B. cepacia</em></td>
<td>M58494</td>
<td>PIILVHG</td>
<td>GHSQGG</td>
</tr>
<tr>
<td><em>B. luteola</em></td>
<td>AF050153</td>
<td>PIILVHG</td>
<td>GHSQGG</td>
</tr>
<tr>
<td><strong>Subfamily IV</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. subtilis</em></td>
<td>M74010</td>
<td>PVVMVHG</td>
<td>AHSQGG</td>
</tr>
<tr>
<td><em>B. pumilus</em></td>
<td>A34992</td>
<td>PVVMVHG</td>
<td>AHSQGG</td>
</tr>
<tr>
<td><strong>Subfamily V</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. stearothermophilus</em></td>
<td>U78785</td>
<td>PIVLLHG</td>
<td>AHSQGG</td>
</tr>
<tr>
<td><em>B. thermocatenulatus</em></td>
<td>X95309</td>
<td>PIVLLHG</td>
<td>AHSQGG</td>
</tr>
<tr>
<td><em>S. aureus</em></td>
<td>M12715</td>
<td>PVVFVHG</td>
<td>GHSQGG</td>
</tr>
<tr>
<td><em>S. hyicus</em></td>
<td>X02884</td>
<td>PVVFVHG</td>
<td>GHSQGG</td>
</tr>
<tr>
<td><em>S. epidermidis</em></td>
<td>AF090142</td>
<td>PVVFVHG</td>
<td>GHSQGG</td>
</tr>
<tr>
<td><strong>Subfamily VI</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. acnes</em></td>
<td>X99255</td>
<td>PVILIPG</td>
<td>GHSQGG</td>
</tr>
<tr>
<td><em>S. cinnamoneus</em></td>
<td>U80063</td>
<td>PVVLNVG</td>
<td>GHSQGG</td>
</tr>
</tbody>
</table>
Fig. 5.3  PCR amplification using metagenomic DNA from MVG and PENP as template. Lane 1, DNA molecular weight marker; Lanes 2 and 3, amplification following PCR using MVG metagenomic DNA; Lanes 4 and 5, amplification following PCR using PENP metagenomic DNA; Lane 6, Negative control containing no template DNA. The sizes of the molecular weight marker, phage lambda digested with PstI, are indicated on the left of the figure.

Fig. 5.4  PCR amplification using MVG and PENP metagenomic DNA as template. Lane 1, DNA molecular weight marker; Lane 2, MVG metagenomic DNA; Lane 3, PENP metagenomic DNA; Lane 4, negative control containing no template DNA. The sizes of the molecular weight marker, Hyperladder I (Bioline), are indicated on the left of the figure.
eight clones were sequenced using M13 forward and reverse sequencing primers (Table 1.1)

5.4 Analysis of partial lipase sequences

In total, nine different partial DNA fragments associated with enzyme class 3.1.1 were identified (Table 5.2 and 5.3). Nucleotide sequences were analysed for possible sequence errors using BioEdit Sequence Alignment Editor version 5.0.2 and translated into all six reading frames using the translate tool located at www.expasy.org. The correct reading frame was identified by the oxyanion hole amino acids sequence N’- PIVLVHG – C’ and extended for approximately 240 bp to the active site amino acids sequence N’ – GHS(H/Q)GG – C’. Two of the identified partial lipase sequences were 300 bp in length. After BLASTp analysis, all partial sequences showed varying degrees of similarity to
lipase genes from known bacteria and grouped as a variety of E.C.3.1.1 enzymes. Three of the partial clones (240-1, 240-5, 300-5) were identified as being putative enzymes. All but two of the partial fragments, 240-5 and 300-3, harboured conserved domains. The two domains which featured most commonly amongst the gene fragments were MhpC and PldB. MhpC is a functional prediction indicator and indicates that the amino acid sequence belongs to the \( \alpha/\beta \)-hydrolase superfamily. An enzyme containing an MhpC domain is predicted to be either a hydrolase or acyltransferase.

Table 5.2  Partial lipase amplicons obtained following PCR of community DNA samples using degenerate primer sets OXF1 and ACR1 and LipF and LipR2.

<table>
<thead>
<tr>
<th>Clone</th>
<th>Nearest matches</th>
<th>Organism</th>
<th>Conserved domain</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lip3</td>
<td>Alpha Beta hydrolase fold</td>
<td><em>Sphingopyxis alaskensis</em></td>
<td>Abhydrolase 1, MhpC, PldB</td>
<td>42%</td>
</tr>
<tr>
<td>240-1</td>
<td>Putative Esterase</td>
<td><em>Streptomyces coelicolor</em></td>
<td>MhpC</td>
<td>60%</td>
</tr>
<tr>
<td>240-2</td>
<td>Triacylglycerol lipase</td>
<td><em>Cryptococcus neoformans</em></td>
<td>LipA</td>
<td>39%</td>
</tr>
<tr>
<td>240-3</td>
<td>Hydrolase or acyltransferase</td>
<td><em>Synechococcus elongatus</em></td>
<td>Thioesterase, pldB, MhpC, COG3319, DAP2</td>
<td>35%</td>
</tr>
<tr>
<td>240-4</td>
<td>Class 2 lipase</td>
<td><em>Rubrobacter xylanophilus</em></td>
<td>Lipase 2, LipA, MhpC, PldB</td>
<td>52%</td>
</tr>
<tr>
<td>240-5</td>
<td>Putative lipase</td>
<td><em>Streptomyces cinnamonensis</em></td>
<td>None</td>
<td>39%</td>
</tr>
<tr>
<td>240-8</td>
<td>Hydrolase or acyltransferase</td>
<td><em>Synechococcus elongatus</em></td>
<td>Thioesterase, MhpC, LipA, PldB</td>
<td>50%</td>
</tr>
<tr>
<td>300-3</td>
<td>Esterase 2</td>
<td><em>Acetobacter pasteurianus</em></td>
<td>None</td>
<td>28%</td>
</tr>
<tr>
<td>300-5</td>
<td>Putative hydrolase</td>
<td><em>Symbiobacterium thermophilum</em></td>
<td>DAP2, COG0412, COG4188, PldB</td>
<td>61%</td>
</tr>
</tbody>
</table>

Identity implies the percentage identity the putative lipase gene has with that of the nearest match

The MhpC domain is also related to 13 other conserved domains. The PldB domain suggests amino acid similarity to that of known lysophopholipases which are important in
Table 5.3 DNA and protein sequences of the partial lipolytic fragments isolated from the Antarctic Dry Valley mineral soils

<table>
<thead>
<tr>
<th>Clone</th>
<th>DNA sequence</th>
<th>Protein sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lip3</td>
<td>CGCAATCTCTGTTAGTTACGGGTCATCGGAAGGTCAATTGAGTTCTTGATGTGGATATAGAGGCACGAAAGAGA</td>
<td>PVPLVHGPGSPSLPSYADAI YGEWERDFILVQWDQRGT GKYVRPLVPELSPDLYKSN PLILECMITYGIEIALYLIY</td>
</tr>
</tbody>
</table>
Table 5.2 continued

<table>
<thead>
<tr>
<th>Clone</th>
<th>DNA sequence</th>
<th>Protein sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>240-8</td>
<td>GACCGATCGTGCTGGTACACGGGCCGCAGTGTTTCTGGGGAGCTGGTCATGGAAATAGATGCGACTCGAAGACGGGACGCAATGCGATAATTCTCGACGTGCCCGGTCGAGCCGGAGACGC</td>
<td>PIVLVHAFLGSWNSNKIVSRLQTQGHNAILDVPGRAAGTGTACACGTCGAGTTCAGTTTGGAGCCAAAGAGAGCCCAGTGGTTKT AGTAGGTCACAGCGAGGCGC</td>
</tr>
<tr>
<td>300-3</td>
<td>GACCAATTGTGCTGGTACACGGTTACAATTCAGACAACAGCGCGTTACGCTGGTACACGGTTACAATTCAGACAACAGCGCGTTACG</td>
<td>PIVLVHYNSDANSAWQFSESSPNDLALALREVFPAFDIIPVEYGDRSGAEPDNENTSGAFYALPDDLALQLRTLVEPTSLWHQWAFTRYNIVGHSHGG</td>
</tr>
<tr>
<td>300-5</td>
<td>GACCAATTGTGCTGGTACACGGTTACAATTCAGACAACAGCGCGTTACGCTGGTACACGGTTACAATTCAGACAACAGCGCGTTACG</td>
<td>PIVLVHYNSDANSAWQFSESSPNDLALALREVFPAFDIIPVEYGDRSGAEPDNENTSGAFYALPDDLALQLRTLVEPTSLWHQWAFTRYNIVGHSHGG</td>
</tr>
</tbody>
</table>

lipid metabolism. The LipA domain is indicative of predicted acyltransferase and hydrolase functionalities.

### 5.5 Discussion

Obtaining novel biocatalysts traditionally involves either culture-based or expression library screening. However, these two approaches do not allow access to the complete available DNA complement provided by microorganisms leaving the available microbial nucleic acid pool is greatly underused (Whitman et al., 1998). This is primarily due to the fact that some 99% of the microorganisms appear to be uncultured (Amann et al., 1995). Conversely, the use of metagenomic libraries holds promise of accessing totally
unknown sequence space, yet is compromised by the drawbacks of heterologous gene expression (Lorenz, et al., 2002).

In this study, previously uncharacterized Antarctic mineral soils were probed for the presence of genes that might confer lipolytic activity and broadly classify as $\alpha/\beta$-hydrolase enzymes. To achieve this, degenerate primers based on known lipase genes were designed and the metagenomic DNA from the Antarctic mineral soils probed using PCR. The partial DNA fragments were characterized using bioinformatic analysis.

The design and use of degenerate primers to target specific genes within environmental DNA has been employed successfully (Eschenfeldt et al., 2001; Bell et al., 2002). Using degenerate primers potentially allows access to all the target genes within the community, although the efficacy is dependent on the level of degeneracy of the primers and the known sequences used as template during primer design (Cowan et al., 2005). The use of sequence-based screening is thought to be conservative because the oligonucleotide primers used to identify target genes reflect conserved amino acid sequences motifs (Lorenz et al., 2002). However, sequence-based screening as a means to access sequence space might not be as conservative as predicted. Studies on the adaptation and differences of psychro-, meso-, and thermophilic enzymes all indicate that sites such as the active site remain highly conserved throughout all taxonomic groupings and temperature ranges (Fields, 2001). Rather, substitutions and modifications which have led to enzyme enhancement have all been mapped to regions and motifs located far from the active site as not to disrupt the interaction of the catalytic residues (Fields, 2001). Additionally, studies on xylanases and polyketide synthases
demonstrated that novel enzyme sequence space could be retrieved using sequence-based approaches (Seow et al., 1997; Lorenz et al., 2002).

Combining sequence-based screening with community analysis might assist in obtaining additional and even novel sequence space. 16S rDNA analysis of a community may indicate both the community structure and which fraction of the community is represented by uncultured microorganisms. If a very extreme environment is chosen the diversity might be significantly less compared to a more temperate and less stressed environment such as compost (Sandaa et al., 1999). Additionally, the gene to be targeted will greatly influence the type of environment selected. If, for instance, a protease gene is the target it might be more feasible to probe an environment high in protein. This form of enrichment differs from that of laboratory enrichment in that laboratory enrichment only selects for the culturable fraction within the sample because artificial culturing is dependent on current techniques (Janssen et al., 2002).

The reason for obtaining only one partial lipolytic gene fragment using primer set OXF1 and ACR1 remains unclear. It was revealed that the limited conservedness of the lipase genes restricted primer design and that PCR parameters had to be severely altered in an attempt to optimize conditions (Bell et al., 2002 personal communication). In a separate study (PT Basvi, unpublished) on Chinese hot spring sediment, it was found that the OXF1 and ACR1 primers more readily amplified partial lipolytic fragments using published parameters. Given that the OXF1 and ACR1 primers were originally aimed at targeting lipolytic genes from a thermophilic environment, it might be concluded that this primer set is better suited for thermophilic lipolytic enzymes.
Probing for lipase genes yielded nine partial lipolytic gene fragments (Table 5.2), which was indicative of the presence of lipase genes within the Antarctic mineral soils. Although bioinformatic analysis indicated identities varying between 35% - 61% with known bacterial and fungal lipolytic genes, the true phylogenetic relationship of the fragments remains unclear, as only partial gene sequences were known. This study however, represents the first where cold-active lipolytic genes were targeted from an environmental metagenomic sample using a PCR-based strategy. All previous studies relied on cultured strains from environments such as Antarctica, Siberia and Alaska (Table 1.2), which showed lipolytic activity when sub-cultured on indicator plates (Arpigny et al., 1993; Choo et al., 1998; Suzuki et al., 2002).

Bioinformatic analysis of the partial lipolytic gene fragments indicated that they originated from three bacterial phylotypes (Actinobacteria, Proteobacteria and Cyanobacteria) and one fungal phylotype (Basidiomycota). This is evidence that the LipF and LipR2 primer set was capable of detecting lipolytic genes from both bacterial and fungal origin. Of interest is that upon comparison of the partial gene fragments to database entries none of the fragments showed even remote identity towards cold-active lipases identified in previous studies, including those from Antarctic bacteria Psychrobacter immobilis B10, Moraxella strain TA144 and Psychrobacter sp. Ant300 (Feller et al., 1991; Arpigny et al., 1993; Kulakova et al., 2004). Whether this implies that the putative lipase fragments are novel is unsure as they represent partial genes and any attempts at concluding function and/or possible origin is speculative.
A future strategy involving the partial lipase gene fragments would be to obtain the full-length genes. This can be done by using the known sequence of the partial lipase gene fragments as templates for the design of internal specific primers for each individual gene. The up- and downstream flanking regions of the genes can then be obtained using a variety of PCR-based techniques such as genome walking PCR (Kilstrup and Kristiansen, 2000), TAIL-PCR (Liu and Whittier, 1995), subtractive hybridization (Felske, 2002) and inverse PCR (Hartl and Ochman, 1994). Although these techniques have successfully been used to obtain flanking regions from single genome samples, their efficacy on metagenomic samples is as yet undetermined.
Chapter 6
Summary and Conclusion

The study of microbial diversity is complicated by the fact that microorganisms are microscopic and not easily differentiated morphologically (Schloss and Handelsman, 2004). Although current methods employed to assess microbial diversity remain inadequate, the use of 16S rDNA analysis (Pace et al., 1986) and other fingerprinting techniques such as DGGE and SSCP (Smalla, 2004) have allowed for increased access and better understanding of microbial diversity. Efforts to fully determine microbial diversity have led to the study of both extreme and diverse biotopes, which include seawater, sub-seafloor, deserts, forest soil and snow to list but a few (Carpenter et al., 2000; Dunbar et al., 2002; Webster et al., 2003; Venter et al., 2004). In this investigation, 16S rDNA analysis was used to assess the microbial diversity of three Antarctic mineral soils, each differing with respect to altitude, protein, lipid and water content.

During the first part of the study, the community (metagenomic) DNA from all three mineral soils (PENP, MVG and BIS) was probed using standard 16S rDNA primers E9F (Farely et al., 1995) and U1510R (Reysenbach and Pace, 1995). A 16S rDNA library of each sample site was constructed followed by identification of unique clones using ARDRA analysis. The findings indicated that the bacteria present within all three the sample sites broadly grouped into eight phylogenetic groups, with Cyanobacteria, Actinobacteria and
Acidobacteria representing the majority of the phylotypes. A large fraction (21%) of the sequences identified fall within the previously ‘uncultured’ class.

The current understanding of Antarctic mineral soil microbiology is based almost exclusively on culture-based studies (Friedmann, 1993). However, a number of important Antarctic Dry Valley microbial non-soil biotopes, including cryptoendolithic communities (de la Torre et al., 2003), cryoconite holes (Christner et al., 2003), and lake ice and marine ice flows (Priscu et al., 1998) have been subjected to detailed community analyses using modern molecular phylogenetic techniques. This study represents the first culture-independent phylogenetic analysis of Antarctic Dry Valley mineral soils. There is a notable overlap between the phylotypes identified in the cryoconite hole and Dry Valleys, which might be expected as wind-blown particles would colonise Dry Valley soils and adjacent glaciers. Christner et al. (2003) reported the presence of Proteobacteria (α, β), Actinobacteria, Verricimicrobia, Acidimicrobia and Cyanobacteria, all phylotypes identified in the Dry Valley mineral soils. Cyanobacteria and Actinobacteria both constitute major fractions of the phylotypes identified in both the Dry Valleys and cryoconite hole, but the Acidobacteria are significantly more abundant in the Dry Valley soils, with only a single clone identified in the cryoconite hole community. Other phylotypes unique to the cryoconite hole were Cytophagales, Planctomycetes, Gemmimonas and γ-Proteobacteria. Within the Cyanobacteria Christner et al. (2003) identified Chamaesiphon, Leptolyngbya and Phormidium sp. of which only Phormidium was identified in the mineral soils. Priscu et al., (1998) focused on Cyanobacterial communities in lake ice and found the same three species present as in the cryoconite hole, which might not be surprising as cryoconite holes are formed in ice (Christner et
The cryptoendolithic study of de la Torre et al. (2003) identified several phylotypes overlapping with the findings of this investigation. Cyanobacteria identified within the cryptoendolithic community include Gloeocapsa, Plectonema, Chroococcidiopsis, which represent Cyanobacteria normally associated with cryptoendolithic communities (Nienow and Friedmann, 1993). It is evident that historical data from culture-dependant studies do not accurately represent the true microbial species diversity of the Dry Valley mineral soils and although this is a preliminary study exhibiting small sample size over a narrow range, there is substantial motivation for more elaborate investigations.

The presence of cyanobacteria within the PENP sample suggests further investigation, as cyanobacterial populations are not usually associated with desiccated environments (Miller and Bebout, 2004). In addition, no evidence of cyanobacterial mats were observed near Lake Purgatory and the putative members identified within sample PENP are not typically associated with cryptoendolithic communities, which might suggest the presence of free-living cyanobacterial communities in the mineral soils. PENP is most likely to contain evidence of an active community as cyanobacteria are capable of photosynthesis at -5°C and nitrogenase activity has been shown to be stable at -7°C (Pandey et al., 2004). Future work should include detailed surveys of cyanobacterial distribution in relation to environmental and physical factors to understand the principles which dictate the distribution of these soil-borne cyanobacterial populations.

The molecular analysis performed on the Antarctic mineral soils indicate a substantial reservoir of microorganisms. Early studies relying on culture-based techniques only
identified a small fraction of microorganisms due to culturing constraints. This might have led to the assumption that the mineral soils contain a restricted number of cosmopolitan taxa (Cameron et al., 1972; Friedmann, 1993). Molecular phylogenetic analysis allow more insight into community structures of various biotopes, identifying both known and unknown or ‘uncultured’ microorganisms. These uncultured microorganisms provide potential novel resources and could be used in various metagenomic approaches to identify biotechnologically important enzymes or pathways.

A future investigation would be to determine which fraction of the phylotypes identified form part of the extant population of microorganisms and to what degree they play a role within the active community structure. DNA directly isolated from the mineral soils might potentially yield phylotypic signals representing either metabolically active cells, VNBC individuals or extracellular DNA (naked DNA) or a combination of all three. Although DNA from dead cells released into the environment becomes available for degradation by bacterial DNA nucleases, DNA-soil particle complexes shield naked DNA from digestion for varying lengths of time (Recorbert et al., 1993). Survival of naked DNA in soils has been shown to vary between 3 months, 23.5 weeks and 30 000 years for forest, Dry Valley mineral soils and Siberian sediments, respectively (Stokstad, 2003; England et al., 2004; Ah Tow and Cowan, 2005). Environmental factors such as the extreme cold, low water availability and high levels of salts and desiccation may contribute to survival of naked DNA within the Antarctic Dry Valley mineral soils. Before any conclusions can be drawn, additional studies such as determining what proportion of the bacterial cells in the soils is alive should be conducted. This can readily be ascertained by performing live-dead staining followed by fluorescence microscopy (Savichtcheva et al., 2005). Designing
a study to determine if there is any naked DNA present in the soils should also be conducted and the results can be compared to that of the live-dead staining. Thus, if a large proportion of the cells in the soils are dead, the presence of naked DNA in the soils should be expected. Conversely, if a large proportion of the cells are alive, studies indicating bacterial activity using SIP (Pelz et al., 2001) and/or BrdU (Borneman, 1999) should be performed.

If active cells are detected in the mineral soils additional studies should be performed to determine which proportion of the phylotypes are dominant. To achieve this, sampling areas surrounding the original site of collection and at various intervals throughout the year will also yield valuable information regarding the distribution of the phylotypes. Dominant and indigenous phylotypes to each mineral soil should be perpetual throughout the year whereas exogenous phylotypes, for instance wind-blown individuals or communities, would be periodic.

Another approach to study active cells might be to extract the total RNA complement and to enrich for mRNA by subtractive hybridisation (Poretsky et al., 2005). Using random primers, a cDNA population can be generated through RT-PCR. Genes such as 16S rRNA could be targeted and should be detected more readily as their copy numbers are higher.

In the second part of the investigation a metagenomic library was constructed in the PCR®-XL-TOPO® cloning vector and the library was screened for the presence of lipolytic activity using expression screening in E. coli. Although a library of 10^4 clones was
generated no clones harbouring lipolytic activity were detected. Using statistical analysis of the generated library it was concluded that the library was too small (i.e. represented an estimated 3.4% of the metagenome) to represent even one copy of the desired gene (Gabor et al., 2004). A possible way to increase the probability of obtaining a functional gene during expression screening would be to increase the coverage of the metagenomic DNA cloned. This could be achieved by either generating a library consisting of $\geq 10^7$ clones, each containing an average insert size of 3 kb or by using systems which can accommodate larger inserts such as fosmids, cosmids or BACs. The problem with generating a library of $\geq 10^7$ clones is that it is technically challenging to screen (Gabor et al., 2004; Cowan et al., 2005), whereas using fosmids, cosmids or BACs greater coverage can be achieved by generating a library 10-times smaller. Fosmids and related systems allow the cloning of large (>40 kb) DNA fragments and provide access to up- or downstream genes or elements, such as chaperones, for functional secretion of target enzymes (Brady et al., 2001).

The third component of this study involved the PCR-based probing of a metagenomic DNA sample for the presence of putative cold-active lipolytic genes. Degenerate primers were designed to target the consensus DNA sequences of the oxyanion hole and active site within the lipolytic genes. Initial PCR trials using primers OXF1 and ACR1 designed by Bell et al. (2002) yielded only one partial gene fragment. The OXF1 and ACR1 primer set appeared to work well for prospecting in thermophilic samples as Bell and co-workers (2002) succeeded in isolating a single lipase gene from an olive-oil-enriched percolation conducted at 65°C. Upon redesigning the primers eight partial putative lipolytic gene fragments were identified. Bioinformatic analysis of the partial fragments revealed that
five of the sequences were predicted as either esterases or lipases and the remaining three were more broadly classified as belonging to the hydrolase family.

This study indicated that the degenerate primers designed to probe for the presence of lipolytic genes worked efficiently and it was concluded that the metagenomic DNA from the Dry Valley mineral soils contained lipolytic genes. Although only partial genes were isolated, bioinformatic analysis indicated that they possibly originated from three bacterial and one fungal phylotype. The partial fragments also varied with respect to known lipase genes and from one another. The extent of variation, novelty and possible movement in sequence space remains speculative however, as none of the fragments showed even remote identity towards cold-active lipases identified in previous studies. Such relationships could only be ascertained once full-length genes had been acquired. Future work could include extracting the full-length genes using either genome walking or semi nested PCR. These two strategies have been used successfully in separate studies (Bell et al., 2002; Eschenfeldt et al., 2001).
Portions of this work have been presented both nationally and internationally, and published in peer-reviewed journals:

Congress Contributions

International:


National:


Publications


References


