Investigation of Genome Diversity in Antarctic Dry Valley Soils

A thesis submitted to the University of London for the
Degree of Doctor of Philosophy by

Samantha J. Whiting

September 2004

Department of Biochemistry and Molecular Biology
University College London
Gower Street
London WC1E 6BT

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

The McMurdo Dry Valleys of South Victoria Land, Antarctica contain some of the most extreme biotopes on earth; extreme to the extent they have long been considered valid Martian analogues. Life within Dry Valley mineral soils, where water contents range from 0.2-5.0% w/w and mean annual temperatures fall below <-20°C, must further contend with desiccating winds, diurnal freeze-thaw cycles, and high seasonal UV radiation. Our knowledge to date of the microbiology of this biotope has been restricted to cultivation studies.

This study sought to investigate microbial diversity in Antarctic Dry Valley soils using both cultivation and cultivation-independent techniques. Phylogenetically-informative ribosomal RNA gene libraries were generated from *Archaea*-, *Bacteria*- and *Eukarya*-specific PCR products amplified directly from DNA extracted from Dry Valley soils. This approach permits a more representative assessment of microbial diversity, as it circumvents the need for cultivation; it is estimated that only 0.1-10% of microorganisms in soil can be cultivated *in vitro*. Dry Valley soils were found to support a high diversity of bacterial species based upon the analysis of 16S rRNA gene clone sequences, the majority of which showed little similarity to previously cultivated bacteria. The diversity of *Archaea* and eukaryotes was reduced in comparison, and included members of the non-thermophilic *Crenarchaeota* and species of fungi respectively. To complement these molecular analyses, bacteria were also cultivated *in vitro* from Dry Valley soils.

Additionally, functional gene diversity was investigated in these soils focusing on integrons and their associated gene cassettes. Using a PCR-based strategy, evidence was obtained for the presence of class 1 integrons in Antarctic soils. Furthermore, a unique *aadA* gene cassette encoding a streptomycin/spectinomycin adenyltransferase was recovered in this study.

These data reveal Dry Valley soils to support a microbial community of considerable genetic diversity, and provide fresh insights into the role of integrons in pristine environments.

Table of Contents

Table of Figures

Chapter 1: Introduction

Chapter 2: Materials & Methods

Chapter 3: Bacterial Diversity

Chapter 4: Diversity of *Archaea* and *Eukarya*

Chapter 5: Integron Diversity in Antarctic Environments

Chapter 6: Discussion

Acknowledgements

Firstly, I would like to thank my supervisor Dr. John Ward for accepting the challenge of my PhD, and me, and for his advice and support throughout this work. I would also like to thank Dr. Don Cowan for providing me with the opportunity to carry out field work in Antarctica, an experience that I will never forget, and one which has made even the most trying of PhD moments all worthwhile. Most of all, a huge thank you to Dr. Ken Bruce to whom I am indebted. Thank-you for the wealth of guidance and encouragement you have provided, and for all the discussions that have helped to inspire and enthuse me throughout this work.

I would also like to say a big thank-you to everyone on the ground floor who I have worked with along the way, and for the many 'lab' nights out that were enjoyed. Thank-you to Medhi for answering all my 'quick technical questions', and to Stefan and Andy, who were always the first port of call with my many computer queries. Thanks also to Rory, it was great fun working with you, and to Gill for all your help at the very beginning.

Abbey, thank-you for listening through all the dramas, for the coffee and cake sessions we enjoyed, and the night's out where wine was plentiful and the world was put to right.

Most importantly, I would like to say a big thank-you to my family for their continuous and unfaltering support in so many ways, and for putting up with me at times when I was perhaps a 'bit stressed and highly strung'. With special thanks to my Dad who has always been there at the front line, and consequently, borne the brunt end of the stick on many occasions. For the times when no matter what was said it was the wrong thing, and for those leisurely runs that weren't always quite so leisurely. You have helped enormously.

Finally, my research was supported by a BBSRC Studentship and financial assistance in support of my Antarctic field expedition was provided by the University of Waikato, New Zealand with additional funds being provided by University College London Graduate School, the Central Research Fund (University of London), and VWR International. The

receipt of a FEMS Travel Grant from the Society of General Microbiology to attend the FEMS European Congress of Microbiologists (Slovenia, 2003) is also gratefully acknowledged.

Abbreviations & Symbols

aa	amino acid
ARDRA	Amplified ribosomal DNA restriction analysis
bp	base pair
Brat	Bratina Island
BD	Bed dust
BLAST	Basic local alignment search tool
BOX-PCR	BOX-polymerase chain reaction
BrdU	Bromodeoxyuridine
5'CS	5' conserved segment
3'CS	3' conserved segment
CDZ	Czapek-Dox agar
dNTP	deoxynucleotide triphosphate
DGGE	Denaturing gradient gel electrophoresis
DNA	Deoxyribonucleic acid
EDTA	Ethylene diamine-tetra-acetic acid
g	gram(s)
GFP	Green fluorescent protein
kb	kilobase pairs
In	Integron
Inc	Incompatability group
ITS-RFLP	16S-23S Intergenic spacer-restriction fragment length polymorphism
MV	Miers Valley
MVT	Miers Valley Transect
¼ NB	One-quarter strength nutrient broth agar
NCBI	National Centre for Biotechnology Information
orf	Open reading frame
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
RDP	Ribosomal Database Project
rDNA	ribosomal DNA
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
rRNA	ribosomal ribonucleic acid
RT	Reverse transcriptase
SDS	Sodium dodecyl sulphate
SI	Super-integron
SSU	Small sub unit
ST	starch
T_A	Annealing temperature
Tn	Transposon
UV	Ultra Violet
w/v	weight per volume
X-gal	5-Bromo-4-chloro-3-indoyl-1-β-D-galactopyranoside

Chapter 1
Introduction

1.1 The Study of Microbial Diversity

The investigation of microbial diversity in natural environments can be considered a fundamental starting point for microbial ecological studies. A knowledge of the types of microorganisms inhabiting a particular niche is a prerequisite to understanding their function within, and impact on an environment. DeLong (1996) defined biodiversity as:

"Biodiversity is an attribute of an area and specifically refers to the variety within and among living organisms, assemblages of living organisms, biotic communities and processes, whether naturally occurring or modified by humans."

The prokaryote domains, Archaea and Bacteria, comprise millions of different species ubiquitously distributed across the Earth – diversity is not only illustrated in species terms, but also in habitat range and both metabolic and physiological capabilities, a reflection of 3.8 billion years of evolution. Investigations of community composition, the spatial distribution of different members and the function of populations as a whole, especially with regard to both biotic and abiotic factors, can help to elucidate the important symbioses and processes occurring in natural environments.

Additionally, the abundance of certain types of microorganisms and their function in global chemical cycles is such that the role of prokaryotes in the Earth's biosphere should not be over looked – their activities and the processes they catalyze sustain all life on earth. In order to understand the effect of microbial activities on the properties of an environment through to the workings of the biosphere, a representative survey of microbial diversity is required. As Pace (1997) stated,

"A complete cataloguing of Earth's microbiota is needless, and of course, impossible. A representative survey, however, is worthwhile."

By generating a representative survey, diversity studies can help to establish evolutionary relationships and may reveal the unique innovations that have contributed to the evolution of different microbial groups. It should also not be forgotten that microorganisms represent an important source of novel genes and gene products with huge potential for biotechnological and industrial application. The desire to access the largely untapped and highly novel genetic diversity of the microbial world follows logically from studies of diversity.

1.2 Methods for the Study of Soil Microbial Diversity

1.2.1: Culture versus non-culture-based methods

The disparity in the enumeration of microorganisms in environmental samples when comparing values obtained by direct counting methods versus culture-based techniques is now well recognized, the former being one to two orders of magnitude higher than traditional plate counts. It has been estimated that only 0.1 to 10% of microorganisms observable in nature can be cultivated *in vitro* using standard techniques (Amman *et al.*, 1995). Consequently, culture-based approaches have provided a very biased and significantly limited picture of the extent of microbial diversity in natural environments. Indeed, research examining the re-association kinetics of DNA extracted from soil estimated there were 4000 distinct genomes per gram of soil, of which only a small fraction of this diversity could be cultured *in vitro* (Torsvik *et al.*, 1990). However, with the advent of recombinant DNA technology and the application of molecular phylogenetic techniques to the study of microbial diversity, the requirement for cultivation has been circumvented.

It was the pioneering work of Carl Woese and associates using small subunit ribosomal RNA sequences that led to the development of a modern day framework with which to relate organisms objectively (Woese & Fox, 1977; Woese, 1987). Woese proposed the three primary domains within which all cellular life could be grouped, the *Bacteria*, *Eukarya* and *Archaea*. Comparative sequence analysis of 16S and 18S rRNA gene sequences made it possible to infer relationships among organisms. These molecules possessed the following features that made them ideally suited to such a purpose:

- their universal presence in all organisms
- the existence of highly conserved sequence domains interspersed with more variable regions making it possible to infer both relationships between distantly related organisms and to assess more close relationships
- they are not subject to lateral gene transfer
- they do not undergo rapid sequence evolution due to functional constraints
- their sufficiently large size so as to be informative but also small enough that they could be conveniently analyzed.

Most importantly, analysis of SSU rRNA gene sequences permitted microorganisms from natural environments to be studied without the need for cultivation. The first generation of studies to adopt this approach to characterize soil microbial communities focused on the construction and analysis of 16S rRNA gene clone libraries (Liesack & Stackebrandt, 1992; Borneman *et al.*, 1996; Bintrim *et al.*, 1997; Felske *et al.*, 1997; Kuske *et al.*, 1997). Phylogenetically-informative clone libraries were constructed from PCR amplified 16S rRNA gene fragments recovered directly from community DNA extracted from soil. Domain-specific primers could be used to selectively amplify bacterial and archaeal rRNA gene sequences. rRNA gene sequences were then analysed using molecular phylogenetic techniques. The application of these molecular methods to study prokaryote diversity in soil and other environments indeed revealed a wealth of novel and environmentally significant diversity. Exploitation of 16S rRNA gene sequence information today underpins many of the techniques used to study microbial ecology; these tools provide the opportunity to monitor and track microorganisms in environments, thus allowing community structure to be linked to ecosystem function (Torsvik and Ovreas, 2002).

1.2.2: Nucleic acid-based detection methods

The information amassed from 16S rRNA gene clone library analysis of environmental samples has highlighted a wealth of uncultured genetic diversity in soils, which in turn has facilitated the development of culture-independent tools with which to study soil microbial communities. These methods have focused on the characterization of nucleic acids and have permitted the analysis of discrete parts of the microbial community,

through the targeting of specific genome sequences, either using nucleic acid probes or by PCR (Head *et al.*, 1998).

PCR-based fingerprinting techniques are informative about microbial community structure as a whole; they have revolutionized studies of microbial ecology in permitting rapid comparative analysis of multiple samples in such a way that would be impossible with clone library analysis. These methods utilize electrophoretic techniques to resolve PCR amplicons amplified directly from community DNA and provide a genetic fingerprint of each community analysed. Two such methods include ribosomal intergenic spacer analysis (RISA) and RAPD (random amplified polymorphic DNA), whereby PCR amplicons are separated according to size. In the case of techniques such as amplified ribosomal DNA restriction analysis (ARDRA) and terminal restriction fragment length polymorphism (T-RFLP), amplified rRNA gene sequences are first subject to digestion with restriction enzymes prior to electrophoretic separation of fragments. The resulting community fingerprint is dependent on the restriction enzyme(s) used and sequence variation that exists among the gene sequences amplified (Ranjard *et al.*, 1999; Tiedje *et al.* 1999). The techniques of DGGE and TGGE (denaturing- and thermal-gradient gel electrophoresis) of which the former has been extensively used, generate a community fingerprint by separating PCR amplicons of similar size on the basis of sequence differences. DNA molecules are electrophoresed in a gradient of increasing denaturant or temperature; upon melting of the double stranded DNA molecule electrophoretic mobility is reduced and thus, the final position of each band is dependent upon the sequence of the DNA molecule (Muyzer *et al.*, 1993; Hill *et al.*, 2000). Fingerprinting techniques have primarily used phylogenetic marker genes to profile the community structure but in more recent studies, this has been extended to functional genes (Rosado *et al.*, 1998). These techniques are ideally suited to multiple sample analysis and can be used to compare community structure between samples, to investigate spatial distribution of specific phylotypes, and to monitor temporal changes, particularly those in response to environmental perturbations (Torsvik *et al.*, 1998).

Fingerprinting techniques excel in their high degree of resolution and can be regarded as most informative when used in combination with methods that differentiate the metabolically active proportion of the microbial community. This can be achieved by extraction and analysis of rRNA. Due to their pivotal role in protein synthesis, a positive correlation exists between the ribosome content of a cell and its metabolic activity and growth rate; thus rRNA sequences should reflect metabolically active cells (Wagner, 1994; Felske *et al.*, 1996 & 1997). Alternatively, the DNA of metabolically active members can be distinguished from that of dormant microorganisms and naked DNA through the incorporation of a label. Bromodeoxyuridine (BrdU) labeling involves the incubation of soil with a thymidine nucleotide analogue, BrdU that becomes incorporated into newly synthesized DNA – labeled DNA is then isolated by immunocapture (Borneman, 1999). Stable isotope probing can be used in a similar way with the addition of two advantages: i) SIP can also be used to label RNA, and ii) by incorporating the label into particular substrates, it can be used to identify microorganisms involved in specific metabolic processes (Manefield *et al.*, 2002; Wellington *et al.*, 2003).

Fluorescent *in situ* Hybridization (FISH) has been used to identify and quantify microbial groups directly within environmental samples (Amann *et al.*, 1995; Amann & Ludwig., 2000). The hybridization and microscopic analysis of fluorescently labeled probes targeting the rRNA of fixed cells allows both the distribution and population dynamics of microorganisms to be studied (Weber *et al.*, 2001). In combination with probes targeting mRNAs it can be used to link functional information with phylogenetic markers and has also been shown to enhance the scope of cultivation techniques (Hugenholtz *et al.*, 2001). Furthermore, the development of real-time PCR has enabled specific phylotypes and functional genes to be quantified reliably and accurately, supplementing studies of spatial and temporal dynamics with quantitative information (Suzuki *et al.*, 2000; Ochsenreiter *et al.*, 2003).

Perhaps the most significant advance in accessing the genetic diversity of unculturable soil microorganisms has arisen from metagenome studies. Techniques have been developed to extract and clone high molecular weight total DNA extracted directly from

Fingerprinting techniques excel in their high degree of resolution and can be regarded as most informative when used in combination with methods that differentiate the metabolically active proportion of the microbial community. This can be achieved by extraction and analysis of rRNA. Due to their pivotal role in protein synthesis, a positive correlation exists between the ribosome content of a cell and its metabolic activity and growth rate; thus rRNA sequences should reflect metabolically active cells (Wagner, 1994; Felske *et al.*, 1996 & 1997). Alternatively, the DNA of metabolically active members can be distinguished from that of dormant microorganisms and naked DNA through the incorporation of a label. Bromodeoxyuridine (BrdU) labeling involves the incubation of soil with a thymidine nucleotide analogue, BrdU that becomes incorporated into newly synthesized DNA – labeled DNA is then isolated by immunocapture (Borneman, 1999). Stable isotope probing can be used in a similar way with the addition of two advantages: i) SIP can also be used to label RNA, and ii) by incorporating the label into particular substrates, it can be used to identify microorganisms involved in specific metabolic processes (Manefield *et al.*, 2002; Wellington *et al.*, 2003).

Fluorescent *in situ* Hybridization (FISH) has been used to identify and quantify microbial groups directly within environmental samples (Amann *et al.*, 1995; Amann & Ludwig., 2000). The hybridization and microscopic analysis of fluorescently labeled probes targeting the rRNA of fixed cells allows both the distribution and population dynamics of microorganisms to be studied (Weber *et al.*, 2001). In combination with probes targeting mRNAs it can be used to link functional information with phylogenetic markers and has also been shown to enhance the scope of cultivation techniques (Hugenholtz *et al.*, 2001). Furthermore, the development of real-time PCR has enabled specific phylotypes and functional genes to be quantified reliably and accurately, supplementing studies of spatial and temporal dynamics with quantitative information (Suzuki *et al.*, 2000; Ochsenreiter *et al.*, 2003).

Perhaps the most significant advance in accessing the genetic diversity of unculturable soil microorganisms has arisen from metagenome studies. Techniques have been developed to extract and clone high molecular weight total DNA extracted directly from

soil, termed the soil metagenome (Rondon *et al.* 1999). These environmental DNA libraries have been used to capture and express heterologous gene products (Henne *et al.*, 1999 & 2000). Additionally, the identification and sequence analysis of large fragments containing rRNA genes has enabled significant functional information to be linked to specific phylotypes (Quaiser *et al.*, 2002; Liles *et al.*, 2003).

Thus, there exists a considerable resource of molecular tools with which to investigate every aspect of soil microbial ecology, all of which have been developed from and continue to exploit the wealth of sequence information deposited from diversity studies.

1.2.3: Cultivation of microorganisms

Presently, the cultivation of microorganisms from soil is receiving renewed interest following the results of molecular phylogenetic studies, which have highlighted an abundance of novel genetic diversity in soils from across the globe (Rappe & Giovannoni, 2003). The importance of obtaining cultivated representatives of new bacterial lineages, in order to determine their metabolic and physiological potential, is now widely accepted as means to elucidating the role of these organisms in natural environments (Chin *et al.*, 1999).

Chin *et al.*, (1999) examined the culturability of numerically abundant bacteria from anoxic rice paddy soils using a most-probable-number technique, based upon an approach described by Janssen and coworkers (1997) in which three strains affiliated to the *Verrucomicrobia* were isolated. Previous to that study, the *Verrucomicrobia* were represented by only a single characterized strain. Bacteria were cultivated from the terminal positive tubes of a liquid dilution series using both monomers (glucose), dimers (cellobiose) and polymers (xylan and pectin) as growth substrates. A total of nine strains were isolated, seven of which were most similar to uncultivated bacteria detected in the same environment, and included three further representatives of the *Verrucomicrobia.*

Following on from this, Janssen and colleagues were successful in cultivating bacterial isolates from Australian pasture soil that were affiliated to novel lineages within the

divisions *Actinobacteria*, *Acidobacteria*, *Proteobacteria* and, *Verrucomicrobia* (Janssen *et al.*, 2002; Sait *et al.*, 2002). The objective of this study was to increase the culturable fraction of soil bacteria above the typical limit of 5%, so as to isolate previously uncultivated bacteria. A viable count value that represented 14.1% (of the total cell count determined by microscopy) was achieved using a serial dilution technique and dilute nutrient broth (1/100 normal concentration) solidified with gellan gum as the growth media, along with prior sonication of the soil to disperse clumps. Taking directly into account the number of viable cells within the soil and the occurrence of cell clumps (that would give rise to a single colony), this figure was estimated to represent 22% of the maximum CFU count cultivated in this instance.

Zengler *et al.*, (2002) described the application of a highly novel soil cultivation strategy that was first trialed with sea water. In this study, concentrated cell suspensions obtained from soil were used for encapsulation within agarose gel microdroplets; gel microdroplets (GMDs) harbouring a single encapsulated cell (as determined by microscopy) were cultivated by placement in a growth column fed with diluted soil extract. GMDs containing microcolonies were then sorted by flow cytometry and subject to 16S rRNA gene sequence analysis, from which a number were found to exhibit $\leq$96% identity to database sequences. Furthermore, of 960 microcolony-containing GMDs transferred to rich medium, 67% grew to cell densities of 10^7 cells per ml making them suitable for further molecular and biochemical studies.

Thus, these studies demonstrate that considerable scope exists for increasing the culturable fraction of soil microorganisms, using both strategies based upon traditional cultivation techniques and entirely novel methods. The use of an extended incubation period, polymers as growth substrates and alternative gelling agents for solid media represent simple modifications that can produce significant results (Chin *et al.*, 1999; Janssen *et al.*, 2002; Sait *et al.*, 2002). Moreover, by simulating to greater effect the natural environment (i.e. formulating growth media directly from the environmental sample) and permitting cross-communication between organisms, it is hoped that those

organisms previously only known from 16S rRNA gene sequences may be cultivated *in vitro* (Kaeberlein *et al.*, 2002; Zengler *et al.*, 2002).

1.3: Phylogenetic Reconstruction

Molecular phylogenetics concerns the determination of evolutionary relationships among organisms using molecular data such as DNA sequence. The objective is to reconstruct the evolutionary paths that account for modern sequences. Genealogical ties between a group of organisms are illustrated in the form of a phylogenetic tree, which chronicles the evolutionary sequence of events between lineages (Pace, 1997).

The process of phylogenetic reconstruction begins with sequence selection and alignment. Typically in the case of diversity studies, rRNA sequences retrieved from a sample are combined with their closest relatives identified from database searches and representatives of major phylogenetic groups. Primary sequence comparisons are then made by a pairwise alignment of rRNA sequences. Global alignments that take into account information contained over the entire length of the sequence are the principle methodology used for phylogenetic analysis (Phillips *et al.*, 2000). This contrasts with local methods whereby segments of one sequence are matched to another, a process used to best effect in the identification of database relatives e.g. BLAST (NCBI), Sequence_Match (RDP). Clustal W is one of the most-widely used programs for multiple sequence alignment and does so by a global progressive method (Thompson *et al*, 1994). A pairwise alignment of the two most similar sequences is performed first, from which a consensus sequence is obtained and used to align the next most similar sequence and so forth. Unambiguous alignments are necessary to ensure that site-by-site comparison of homologous nucleotides is performed. This task is more easily achieved when analyzing either closely related rRNA sequences or a representative subset. In the case of distantly related sequences, analysis of secondary structure can improve accuracy of the alignment by virtue of the functional constraints of the rRNA molecule (Van de Peer *et al.*, 1998).

Based on the alignment, a mathematical model is used to calculate evolutionary distances from a pairwise comparison of sequences (Lio & Goldman, 1998). The objective of these

models is to identify the minimum number of operations – substitutions, insertions and deletions, which are necessary to arrive at sequence A from sequence B. Models may vary in the assumptions made (i.e. all base positions are equally subject to change) and the cost functions attached to the different operations (Lio & Goldman, 1998; Phillips *et al.*, 2000). The distances calculated are then illustrated as branch lengths in a phylogenetic tree. Different statistical methods can be used to determine the tree topology and branch lengths that best fit the phylogenetic relationships between sequences. The most widely used methods for inferring phylogenies include the neighbour-joining and maximum likelihood methods (Whelan *et al.*, 2001).

The neighbour-joining method operates by cluster analysis, beginning with the most similar pairs and suceessively linking more distantly related pairs (Saitou & Nei, 1987). This method is computationally fast and results in a single tree, making it a popular choice for the analysis of large datasets. Additionally, it does not assume the same evolutionary rate for all lineages (Van de Peer & Wachter, 1997). However, it is disadvantaged in that the order in which sequences are analyzed can alter the final tree topology. Maximum likelihood methods seek the tree that would most likely result in the sequences observed today and can be considered the most statistically robust method (Whelan *et al.*, 2001). Different models of sequence evolution that incorporate in-built estimates can be applied, thus allowing competing trees and complex models to be compared statistically. The main disadvantage of this method lies in its exhaustive approach whereby each tree is assessed individually, making it computationally slow for large numbers of sequences.

The final phase of phylogenetic reconstruction lies with the interpretation of the results and this can be aided by the use of a bootstrap test that allocates confidence levels to a given phylogeny. Typically, this involves resampling the original data to create 100 replicate datasets of the same size from which new distance matrices are created. The resulting bootstrap trees are then compared with the tree inferred from the original alignment, such that the frequency of a given cluster in the original tree is calculated. Thus, if a specific branch order from the original tree is present in all bootstrap samples, the branch point will have a bootstrap value of 100%. A bootstrap value of 95% is

desirable for each branch point but phylogenies often represent values exceeding a threshold of 50%.

1.4: Prokaryote Diversity

1.4.1: Diversity of *Archaea*

The use of molecular phylogenetic techniques for exploring microbial diversity has had no more a profound impact that on our perspectives of the phylogeny of *Archaea*. This kingdom was previously considered from culture-based studies to be composed of organisms inhabiting extreme environments (those characterized by high temperature, high salinity and extremes of pH) that were considered inhospitable to *Eucarya* and *Bacteria*. However, with the retrieval of new 16S rRNA gene sequences from natural environments, members of the *Archaea* have been found to be far more diverse and widespread than previously suspected. Three distinct kingdoms of the domain *Archaea* have currently been delineated based upon ribosomal RNA analyses – the *Crenarchaeota*, the *Euryarchaeota*, and the *Korearchaeota* (Woese *et al.*, 1990; Barns *et al.*, 1994).

Cultured representatives of the *Crenarchaeota* form a closely related group of organisms characterized by an extremely thermophilic, sulfur-metabolizing phenotype. However, molecular diversity studies have identified crenarchaeotal phylotypes in marine (Delong., 1992; Delong *et al.*, 1994; Massana *et al.* 1997), freshwater (Hershberger *et al.*, 1996; Schleper *et al.*, 1997) and soil (Bintrim *et al.*, 1997; Borneman & Triplett, 1997; Jurgens *et al.*, 1997; Buckely *et al.*, 1998) environments. These organisms, termed the non-thermophilic *Crenarchaeota*, appear to possess a ubiquitous distribution and abundance estimates infer they are of considerable ecological significance. Cultured members of the *Euryarchaeaota* show greater physiological diversity and include methanogens, sulfur and sulfate reducers and extreme halophiles (DeLong, 2003). Once again novel lineages have been identified from rRNA sequences, particularly in marine environments (Delong, 1992; Munson *et al.*, 1997; Vetriani *et al.*, 1998). The recently proposed third kingdom, the *Korearchaeota*, is composed entirely of sequences retrieved from hot spring sediments in Yellowstone national park; it is thought these organisms are the closest link to the ancestral prototype of the *Archaea* (Barns *et al.*, 1994; Reysenbach *et al.*, 2000).

The number of archaeal phylotypes is considerably lower than their prokaryotic counterparts, the *Bacteria* (Figure 2.1). However, whilst *Archaea* have been known to thrive in extreme environments, molecular diversity studies have revealed them to be similarly cosmopolitan and adaptive.

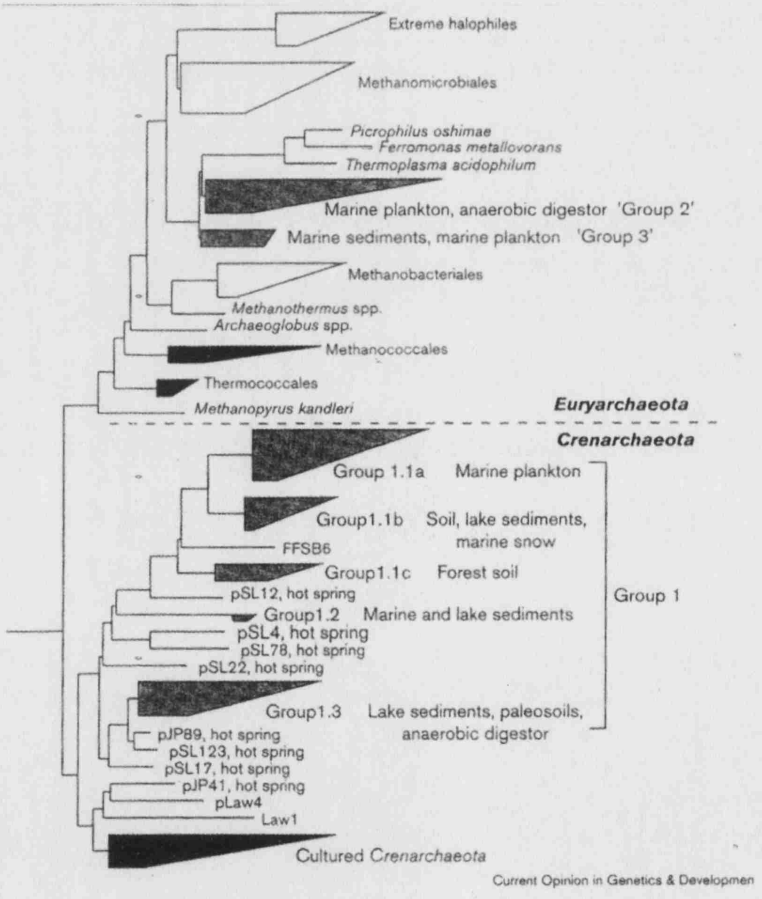


Figure 1.1: Phylogenetic tree illustrating the major lineages of the domain *Archaea*. Taken from Delong, 1998. Thermophilic lineages are represented in black; uncultivated non-thermophilic groups are represented in grey.

1.4.2: Bacterial diversity

Twelve bacterial phyla were initially established by the first comparative phylogenetic analysis of 16S rRNA sequences in 1987 (Woese, 1987). Sequences obtained from well known cultivated organisms formed the basis of this research. However, since the advent of molecular phylogenetic methods our picture of diversity has expanded considerably. The identification of microorganisms without cultivation by sequence analysis of rRNA genes cloned from environmental DNA has revealed a wealth of novel diversity. Major new lineages have been described and diversity within established phyla has grown significantly. In 1998, Hugenholtz *et al.* articulated 36 bacterial divisions at a time when 8000 bacterial 16S rRNA sequences were available; this number was suspected to be over 40 with the inclusion of several candidate divisions represented by single sequences. Recently, this number has increased to 52 major bacterial phyla, only 26 of which contain cultivated representatives (Rappe & Giovannoni, 2003); the remaining phyla are composed entirely of environmental sequences (Figure 1.2). The number of aligned and annotated bacterial rRNA sequences presently deposited in the RDP is 97, 128. This figure illustrates the significant contribution of diversity studies to our perspectives of microbial phylogeny and provides an exceptional resource of information.

The types of microorganisms typically cultivated from environmental samples using standard isolation techniques generally belong to one of four phyla – the *Proteobacteria*, *Firmicutes*, *Actinobacteria*, and *Bacteroidetes* (Hugenholtz, 2002). Each of these phylogenetic groups contains numerous soil-borne species of cosmopolitan distribution, many of which have been subject to extensive genotypic and phenotypic characterization. Consequently, these groups are well represented by rRNA gene sequences obtained from both cultured isolates and environmental clones. In contrast, novel lineages identified from culture-independent analyses are poorly represented (if at all) by cultivated species. Such groups common to soil microbial communities include the *Verrucomicrobia* (O'Farrell & Janssen, 1999; Buckley & Schmidt, 2001), *Acidobacteria* (Kuske *et al.*, 1997; McCaig *et al.*, 1999; Buckley & Schmidt, 2003), *Planctomycetes* (Liesack & Stackebrandt, 1992; Derakshani *et al.*, 2001) and division TM7 (Hugenholtz *et al.*, 2001). Diversity studies indicate they are present in soils across the globe, and by virtue of this

and studies of their abundance (Buckley & Schmidt, 2003) it is assumed they will have an important ecological role.

As 16S rRNA sequence information continues to accumulate it will undoubtedly further shape our perspectives of bacterial phylogeny, revealing new evolutionary lineages and phylogenetic relationships. These data will in turn be complemented by whole genome analyses and investigations of functional genes as phylogenetic markers.

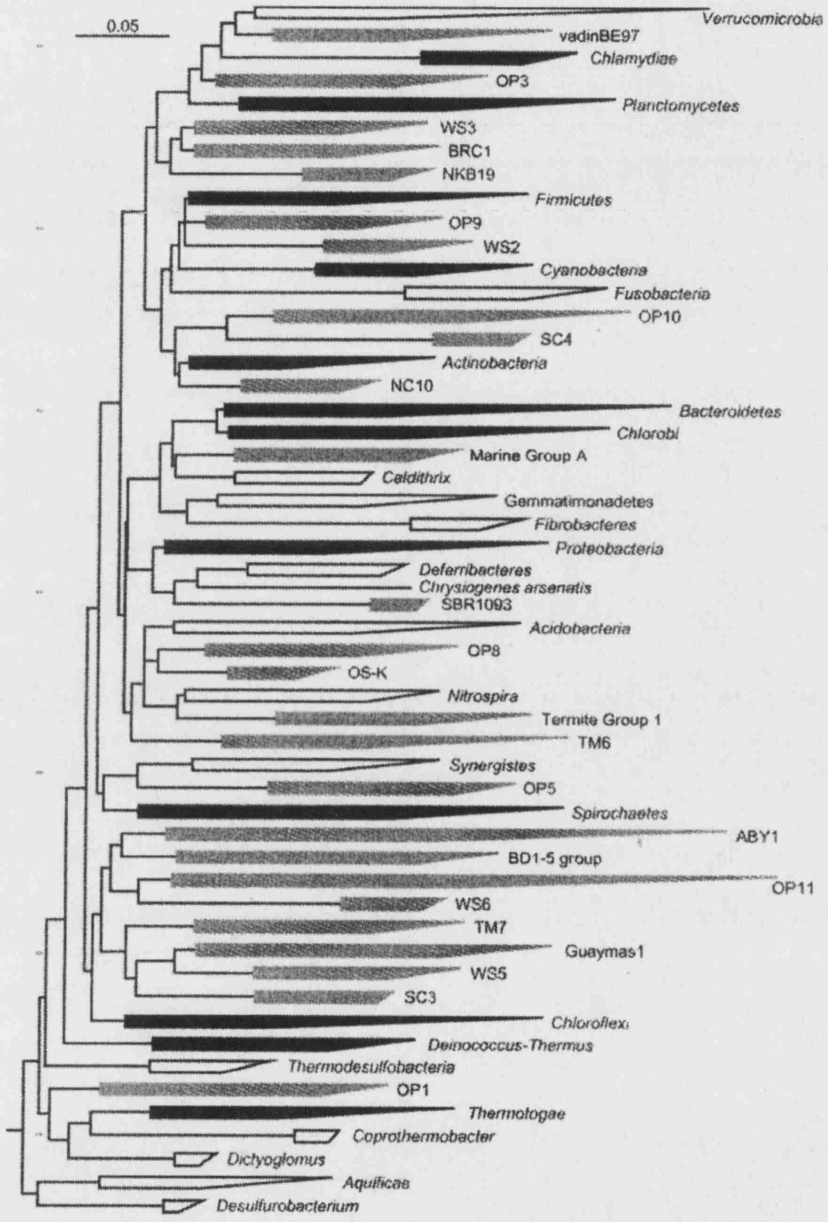


Figure 1.2: Phylogenetic tree illustrating the major phyla of the domain *Bacteria*. Taken from Rappe & Giovannoni, 2003. The branches shown in black represent the 12 phyla first described by Woese in 1987, and the white branches are the 14 phyla that were recognized since 1987. The grey branches are the candidate phyla that presently contain no cultivated representatives. The scale bar corresponds to 0.05 changes per nucleotide position.

1.5: The Antarctic Dry Valley Environment

The Miers Valley and Wright Valley lie within the McMurdo Dry Valleys, an ice-free area of 4000 km^2 positioned across McMurdo Sound from the Ross Island and the Ross Ice Shelf. The Dry Valleys remain ice-free primarily due to the passage of ice from the polar plateau being blocked by the Transantarctic Mountains. The prevailing conditions of low temperatures and low water availability within this environment, in combination with desiccating winds, diurnal freeze-thaw cycles and high seasonal UVA/ UVB radiation, render Dry Valley soils particularly unfavourable to life (Campbell & Claridge, 1987). Indeed, past microbiological investigations using cultivation techniques have indicated some of these soils to be sterile, whilst their analogy to Martian soils has prompted significant interest in the application of this environment to astrobiological studies (Campbell & Claridge, 1987). However, studies examining the Microbiology of the Antarctic Dry Valley environment using both cultivation-based and molecular techniques, have demonstrated that despite the superficial sterile appearance of this landscape, microorganisms have succeeded in colonizing a range of Dry Valley biotopes. Examples include the microbial communities identified within mats surrounding Antarctic lakes (Brambilla *et al.*, 2001; Van Trappen *et al.*, 2002; Taton *et al.*, 2003), both lake and glacial ice (Gordon *et al.*, 2000; Christner *et al.*, 2003b), and Dry Valley soils (Cameron *et al.*, 1971; Cameron *et al.*, 1972) and associated cryptoendolithic communities present within rocks (Torre *et al.*, 2003).

Annual precipitation is extremely low in the McMurdo Dry Valleys, ranging from approximately 13mm on the valley floors to a maximum of about 100 mm water equivalent in the mountains (Beyer *et al.*, 1999); considerable variation occurs both in time and location (Keys, 1980). Much of this moisture is lost to ablation due to the drying action of winds and relative humidity levels which average 50% (Clow *et al.*, 1988). Consequently, evaporation rates estimated to approach 12cm/yr in some areas (Cameron, 1972) can greatly exceed annual snowfall, creating a very arid environment where water contents of surface soils range from 0.2-5.1 % w/w (Cameron & Conrow, 1969; Horowitz *et al.*, 1972).

Winter conditions of continuous darkness, strong winds and temperatures as low as -60°C contrast starkly with the Antarctic summer, a period of 6-10 weeks in which continuous light prevails and air temperatures may reach as high as 5°C. Surface soil temperatures vary widely in the Dry Valleys, with exposure to solar radiation in seasonal diurnal cycles primarily determining maximum temperatures; the influence of altitude and wind speed along with composition of the soil itself also exert an effect (Beyer *et al.*, 1999). On occasion, temperatures may become lethal on the surface of heated rocks - Cameron & Ford (1974) observed an insolated rock in the Ross Desert to rise 42.5 degrees to 27.5°C in 3 hours. The absence of life on rock surfaces can be primarily attributed to exposure to continuous and rapid freeze-thaw cycles. Minimum soil temperatures are a factor of air temperatures and thus, soils of the highest and coldest locations in the Dry Valley environment exhibit the greatest range in surface temperatures (Beyer *et al.*, 1999).

Dry Valley soils are barren in appearance and underlain by permafrost at a depth of 0.5m; permafrost-affected soils are termed gelisols. Antarctic gelisols have formed very slowly under extreme conditions of aridity and low temperatures - mean annual temperatures within the valleys fall below -20°C (Clow *et al.*, 1988; Nienow & Friedmann, 1993; McKay *et al.*, 1998) inhibiting the chemical weathering and biological processes common to other soil types, and this is reflected in their properties (Beyer *et al.*, 1999). There is a complete absence of any visible organic matter, the main soil forming processes being oxidation and salinization. Dry Valley soils are coarse in texture, often having a surface layer of larger stones, and are extremely fragile due to the absence of structure or cohesion, a direct result of the lack of moisture and organic matter. Salt accumulations are frequently evident, either as surface efflorescence or as deposits around permanently unfrozen depressions of water, and serve to further reduce the water potential of soils. Rock weathering is the major contributor of salt and these accumulations are characteristic of soils in which evaporation exceeds leaching (Campbell & Claridge, 1987).

It is well accepted that the amount of water available is the most critical factor in determining the abundance and diversity of soil microorganisms. During the summer months water is available as meltwater on the surface of insolated rocks and as glacial

melt streams. The following additional factors have been proposed as conducive to life in this environment: a northerly aspect to enhance the receipt of solar radiation, sufficient shelter from winds to afford protection from freezing and desiccation, and the absence of inhibitory concentrations of soluble salts within the soil (Campbell & Claridge, 1987). Where ecological and climatic factors exist that are suitable for growth of microorganisms, the physical and chemical properties of the soils themselves, and the availability of nutrients most probably limits the abundance and diversity of soil microbial community (Cameron & Conrow, 1969).

1.6: Microbial Diversity in Antarctic Dry Valley Biotopes

The geographical isolation and extreme climatic conditions of continental Antarctica, together with its range of unique and pristine biotopes characterized by their minimal (if any) exposure to anthropogenic influence, have attracted considerable interest from microbiological studies. Investigations have focused on the types of microorganisms that inhabit Antarctic environments (Vishniac, 1993), their survival strategies (de los Rios *et al.*, 2003), and the ecology of Antarctic biotopes (Hughes & Lawley., 2003), and in more recent years, attention has turned to the Antarctic gene pool as a potential source for biotechnological exploitation (Wery *et al.*, 2003). Examples of environments studied include: maritime lakes and lake sediment (Bowman *et al.*, 2000; Pearce, 2003; Sjoling & Cowan, 2003), sea ice and continental shelf sediments (Bowman *et al.*, 1997; Brinkmeyer *et al.*, 2003; Bowman *et al.*, 2003), Antarctic sponges (Webster *et al.*, 2004), South Pole snow (Carpenter *et al.*, 2000) and freshwater lakes and accretion ice (Priscu *et al.*, 1999; Christner *et al.*, 2001). Microorganisms typically inhabit a variety of refugia - microenvironments that permit their survival during unfavorable conditions, where free water is available and fluctuations in temperature are to an extent, restricted. Specific to the McMurdo Dry Valleys, the refugia occupied by microorganisms include Dry Valley soils and the pores and interstitial spaces of rocks, freshwater lakes and their surrounding microbial mats, and lake and glacial ice. Studies employing both cultivation and molecular-based approaches have been conducted on microbial communities inhabiting some of these biotopes.

The ability of microorganisms to colonize the pores and spaces within rocks perhaps provides the best example of microbial exploitation of a favorable microenvironment. Cryptoendolithic communities (meaning 'hidden within rock') were first described by Friedman and Ocampo-Friedman in 1976 from the observation of green algal material growing within pore spaces between quartz grains in Beacon sandstone. These communities occupy the cracks within clear or translucent rocks such as sandstones, granites and marbles where they are protected from the desiccating surface, receiving liquid water that becomes trapped in pore holes from the melting of occasional snowfalls (Campbell & Claridge., 1987). Friedman (1982) rationalized that colonization of the endolithic zone facilitated survival in an otherwise inhospitable macroclimate without the necessary development of specific biochemical and physiological adaptive responses. Sufficient photosynthetically active radiation is available to carry out photosynthesis and the metabolites produced by phototrophs supports the growth of heterotrophic bacteria and fungi, leading to the development of a highly structured community. Two types of endolithic community have been identified based on the predominant primary producer, lichens and cyanobacteria; lichen-dominated communities are the most prevalent in the Dry Valleys.

Torre *et al.* (2003) examined the microbial diversity of cryptoendolithic communities at Battleship Promontory, a large sandstone outcrop in the Convoy range of the McMurdo Dry Valleys. SSU rRNA gene libraries were constructed from both a lichen-dominated and cyanobacterial-dominated community. Over 1,100 clones were analyzed by RFLP and grouped accordingly into 51 distinct phylotypes, of which a representative of each was sequenced. As expected, lichen and cyanobacterial rRNA gene sequences were abundant in the libraries consistent with their roles as primary producers. Additionally, abundant rRNA sequences affiliated to the *α-Proteobacteria* and *Thermus-Deinococcus* lineages were also observed in the cyanobacterium-dominated community; the authors hypothesized that these organisms were also likely to factor considerably in primary productivity. The molecular survey indicated a reduced diversity of primary producers than suggested by previous morphological studies. A single predominant fungal and cyanobacterial phylotype was apparent in each of the respective libraries – the fungal

sequence was most similar to epilithic lichens previously identified in the Dry Valley environment, whilst the cyanobacterial sequence was most closely related to sequences present in Dry Valley soils and sediments. Sixteen phylotypes were distributed amongst the bacterial class *Actinobacteria*; however the majority were not closely related to any cultivated species, and less numerous rRNA gene sequences constituted a high proportion of diversity in the libraries, indicating the survey was by no means exhaustive. No sequences affiliated to the *Archaea* were identified. Overall, this study demonstrated the existence of considerable diversity within the confines of an endolithic community that could otherwise not be accessed by microscopic and cultivation-based studies. Additionally, inferences could be made regarding the ecological role of particular organisms by virtue of their rRNA gene sequence abundance.

Comparatively little is known regarding the diversity of bacteria present in Antarctic Dry Valley soils. The majority of studies to date have employed traditional cultivation-based approaches providing a very limited perspective of the types of microorganisms inhabiting these soils. Indeed, in some of the early studies certain Antarctic samples investigated were thought to be sterile due to the failure to culture any microorganisms (Horowitz & Cameron, 1971). As for the diversity of *Archaea*, investigations have not been described in the literature. Campbell and Claridge in 1987 described life within these soils as extending to about 10 bacterial genera, two species of yeast and four species of algae, with potentially greater diversity present in algal mats and around lakes. Bacterial genera typically represented by cultured isolates included *Arthrobacter*, *Bacillus*, *Corynebacterium*, and *Micrococcus;* viable counts ranged from 10^1-10^4 microorganisms per g soil (Cameron *et al.*, 1971 & 1972). However, a recent study by Cowan *et al.* (2002) examining the ATP content of Dry Valley soils has shown them to be richer in life than previously thought. ATP analysis of desiccated surface soils was used to calculate soil microbial biomass. Values ranging from 2.6 x 10^6 to 3.9 x 10^8 cells per gram of wet weight material were obtained, indicating cell numbers to be over four orders of magnitude higher than previously suggested. It is conceivable that overlap may exist between soil and endolithic microbial communities due to the obvious similarities between the two environments. Indeed, potential colonizers of the endolithic zone may

reside in soils; however, this will not be known without a more detailed molecular assessment of diversity within these soils.

Lake ice has also been found to support growth of microbial communities in the Dry Valley environment. During the austral summer liquid water inclusions form on the surface of permanently frozen lakes. Particulate matter deposited on the surface by valley winds, results in a layer of sediment that provides an input of inorganic and organic nutrients as well as serving to seed this environment (Priscu *et al.*, 1998). Gordon *et al.* (2000) have investigated the diversity of bacteria associated with these particles by 16S rRNA gene clone library analysis. Sequences representative of the cyanobacteria, *Proteobacteria* and *Planctomycetales* were found to predominate though members of the divisions *Actinobacteria*, Green Non-Sulfur and *Acidobacterium/ Holophaga* were also identified. Further to this study, six hybridization probes designed to target the most abundant 16S rRNA gene sequences in the library were used to investigate diversity within the local terrestrial environment. The probes were hybridized to 16S rRNA gene sequences amplified from three cyanobacterial mat samples as well as to rRNA genes amplified from the lake ice community; interestingly, none of the sequences examined were found to be exclusive to the lake community. This result would indicate a number of microorganisms have adapted to life within different environments of the Dry Valleys, of which their dispersal is most likely facilitated by the strong katabatic winds that exist in the region.

Consistent with this hypothesis has been the recovery of rRNA gene sequences and bacterial isolates from an Antarctic cryoconite hole (Canada Glacier) that are most similar to those previously described in the Antarctic Dry Valley environment (Christner *et al.*, 2003a). Like lake ice, a cryoconite hole is formed by warming of particulate matter on the surface of the glacier that melts into the underlying ice. This environment affords sufficient protection to support a complex community that includes bacteria as well as eukaryotic organisms such as fungi, diatoms and rotifers. Particulates blown on to the surface of glacier from the surrounding environment most likely serve as the inoculum

and would account for similarities in microbial community structure with neighbouring Dry Valley environments.

The most significant accumulation of microbial biomass within the Dry Valley environment is observed around the peripheries of lakes and ponds. Cyanobacteria, the major phototrophs within terrestrial and freshwater biotopes, form highly structured biofilms observed as lake microbial mats covering the substrate (Taton *et al.*, 2003). At this position local melt water is freely available and the supply of nutrients is comparatively high, with material being transported to the lake by both glaciers and streams and by the air-borne deposition of particulate matter. A cultivation study has isolated a wide diversity of heterotrophic bacteria from mat samples collected from ten Antarctic lakes; isolates were affiliated to the *Proteobacteria*, the high and low percent G+C Gram-positives, and the *Cytophaga-Flavobacterium-Bacteroides* groupings (Van Trappen *et al.*, 2002). Indeed, a study of microbial diversity within a mat sample of Lake Fryxell in the McMurdo Dry Valleys has provided a stark contrast to the otherwise sterile and barren appearance of the landscape (Brambilla *et al.*, 2001). A combined approach of cultivation studies and 16S rDNA clone library analysis was used in this instance. Isolates cultivated under aerobic conditions were mostly related to *Janthinobacterium lividum*, *Flavobacterium hibernum* and *Arthrobacter flavus*, whilst those cultivated from an anaerobic enrichment belonged mainly to the *Clostridium/ Bacillus* subphylum. The molecular survey in which 325 clones were partially sequenced revealed most sequences to be phylogenetically affiliated to different sublines of the *Proteobacteria*, *Clostridium/Bacillus*, *Actinobacteria*, *Verrucomicrobia*, and CFB groupings. The majority of sequences were represented only once or twice, indicating much of the bacterial diversity still remained to be explored. Interestingly, virtually no similarity was found between diversity within the mat sample and that described in sediments of neighbouring marine salinity meromictic lakes in the Vestfold Hills, revealing them to be very different ecosystems. Uniquely, the diversity of Archaea was also assessed in this study. Analysis of 72 archaeal clones identified only two sequence types related to *Methanoculleus palmolei* of the *Euryarchaeota*, and an Antarctic clone sequence affiliated to Group 1.1a *Crenarchaeota* recovered from marine picoplankton.

Taton *et al.* (2003) focused specifically on cyanobacterial diversity within a Lake Fryxell microbial mat sample. Their objective was to determine the geographical distribution and extent to which endemism occurs amongst cyanobacteria in polar regions. Both morphological and molecular-based approaches were used to assess diversity in a natural mat sample, and also in a cultured mat community grown within a benthic gradient chamber inoculated from the same mat. Microscopic analysis identified only eight morphotypes compared to 15 phylotypes identified by clone library and DGGE analysis of 16S rDNA amplicons. Phylogenetic analysis of the rDNA sequences revealed 11 distinct lineages, 2 of which were novel, with three lineages that were exclusively Antarctic. A detailed assessment of all Antarctic and polar sequences available, gave rise to a total of 22 lineages that included the two novel lineages and seven that were exclusively Antarctic. This study clearly demonstrated both the existence of cyanobacteria endemic to Antarctica, and the occurrence of cosmopolitan phylotypes inhabiting Antarctic biotopes.

Thus, it is apparent that despite the harsh conditions of the Antarctic Dry Valley environment, a wide diversity of microorganisms has succeeded in colonizing the available sites of refugia. This has lead to the establishment of complex microbial communities identified from molecular phylogenetic studies. The extent of diversity within the soil microbial community has hitherto remained undetermined due to the absence of any molecular data.

1.7: Integrons

Analysis of rRNA gene sequences can provide information regarding microbial community structure and composition in natural environments. However, it is important to consider other genes when studying the different functional capabilities of a community. Integrons provide a mechanism by which bacteria can shuffle and recruit a variety of genes, and their typical association with mobile elements (i.e. plasmids and transposons) facilitates the spread of integrons among bacterial populations. Thus, the study of integron diversity can shed light both on the role of horizontal gene transfer in promoting bacterial fitness and adaptation, and the types of selective pressures microorganisms are exposed to.

1.7.1: Integron function

The discovery of integrons was prompted by the detailed analysis of resistance determinants encoded by plasmids and transposons. Overlapping sets of resistance genes were observed amongst plasmids and transposons that were otherwise unrelated (Hall & Vockler, 1987) whilst conversely, different combinations of resistance genes were identified at the same loci of closely related genetic elements (Ward & Grinsted, 1982; Schmidt & Klopfer-Kaul, 1984). Results of restriction mapping studies lead to suggestions that plasmids and transposons harboring different resistance genes may possess small regions of sequence homology within their resistance determinants serving as 'hot spots' for recombination (Weidemann *et al.*, 1986; Hall & Vockler, 1987). Subsequent sequencing of the resistance genes and flanking sequence lead to the identification of common architectural components now known to constitute integrons (Stokes & Hall, 1989).

Integrons provide a mechanism for horizontal gene transfer in bacteria, serving as genetic elements that permit acquisition and expression of genes contained within mobile cassette structures (Stokes & Hall, 1989; Hall & Collis, 1995). They operate as a site-specific recombination system and possess the following essential components: an integrase gene (*intI*), encoding a site-specific recombinase, a recombination site (*attI*) into which individual cassettes are inserted, and a promoter (Pc) for the expression of cassette associated genes (Recchia *et al.*, 1994; Hall & Collis, 1995). These components are located within the 5' conserved segment of the integron which is followed by a variable region consisting of the integrated gene cassettes. Class 1 integrons responsible for the widespread dissemination of antibiotic resistance amongst Gram-negative bacteria have the addition of a 3' conserved segment. This region typically contains the *qacEΔ1* conferring resistance to ethidium bromide and quaternary ammonium compounds, a *SulI* gene (encoding resistance to sulphonamides) and an ORF of unknown function, *orf5* (Stokes & Hall, 1989; Paulsen *et al.*, 1993).

Gene cassettes are among the smallest and simplest of mobile elements and include only an open reading frame (ORF) and a recombination site, *attC*, also termed the 59-base

element (59-be) located at the 3' end of the ORF (Recchia & Hall, 1995). The integrase enzyme catalyses the integration and excision of gene cassettes through the recognition of the two recombination sites, *attI* and the 59-be (Collis & Hall, 1992, b); Recchia & Hall, 1995). Insertion of gene cassettes occurs more commonly into *attI* although recombination between two 59-be structures is also permitted (Collis *et al.*, 1998). Multiple insertion/ excision events can occur, thus allowing integrons to recruit and shuffle diverse arrays of cassettes.

The 59-base elements are diverse both in sequence and in size, ranging from 57 to 141 bp in length (Recchia & Hall, 1995). However, they all share several features, that of an inverse core site (RYYYAAC) and a core site (GTTRRRY) representing the most highly conserved features of a consensus sequence encompassing the first and last 20 bases. The consensus sequences are imperfect inverted repeats and are separated by a central region which gives rise to the variation in length (Collis & Hall, 1992, b); Recchia & Hall, 1995). Moderate sequence conservation exists within these conserved features and includes eight near-invariant positions (Stokes *et al.*, 1997). The *attI* site of the integron conforms to the same consensus sequence (GTTRRRY) as that of the core site of the 59-be, with the recombination crossover point being located between the G and T residue (Hall *et al.*, 1991). Consequently, on integration of a gene cassette, the *attI* site can be regarded as a composite site, with the last six bases of the core site being derived from that of the integrated cassette. Where multiple insertion events have occurred leading to the generation of a cassette array, each of the 59-bes represent composite sites with the last six bases of the core site originating from the adjacent unit (Recchia & Hall, 1995). On excision of a gene cassette, they exist freely in a circular form but are unable to replicate (Collis & Hall, 1992, a)).

Integrons can be divided into several classes based on sequence differences in the integrase genes they encode. Class 1 integrons exhibit considerable diversity in their cassette content and geographical distribution and are the most extensively studied due to their association with multidrug-resistant strains of bacteria (Levesque *et al.*, 1995; Sallen *et al.*, 1995; Martinez-Feijo *et al.*, 1999. Similarly, integron classes 2 and 3 are also

associated with antibiotic resistance gene cassettes (Hall & Vockler, 1987; Arakawa *et al.*, 1995). The class 4 integron is located on the small chromosome of *Vibriocholerae* (Mazel *et al.*, 1998) and is termed a superintegron structure as it harbours approximately 180 cassettes, whilst the partial sequence of a fifth *intI* gene (*intI5)* has been identified in *Vibrio mimicus* (Clark *et al.*, 2000). A further three integron classes have been proposed following PCR-based recovery of novel integrase sequences from environmental samples, and are designated *intI6*, *intI7* and *intI8* (Nield *et al.*, 2001*)*. Similarly, sixteen new integrase sequences have recently been recovered from heavy metal-contaminated mine tailings (Nemergut *et al.*, 2004). Combined with identification of several chromosomal-located integrase genes as a result of genome sequencing projects (Rowe-Magnus *et al.*, 2001; Nield *et al.*, 2001), these data indicate considerable diversity exists within the integron integrase family.

Integron integrases form part of the large and diverse family of tyrosine recombinases. Members of this family possess two conserved motifs, termed boxes 1 and 2, within which are contained four very highly conserved residues, RHRY (only the histidine is not invariant but is present in nearly 95% of members) (Esposito & Scocca, 1997). Additionally, three shorter conserved regions have been identified, designated patches I, II and III, that are central to the secondary structure of this enzyme (Nunes-Duby *et al.*, 1998). A further insert of about 16 amino acids has been identified between patches II and III of class 1-4 integrases that is not found in other members of this family (Nield *et al.*, 2001). IntI integrases form a group of relative homology within the tyrosine recombinase family, with proteins from the different integron classes displaying 41-57% pairwise identity (Nunes-Duby *et al.*, 1998). All integron classes possess a common structure whereby the *attI* site is located at the 5' end of the *intI* gene, with *intI* being transcribed in the opposite direction to inserted gene cassettes. The *attI* sites of the respective integron classes are similarly variable with the core site being the only feature common to all (Collis *et al.* 1998; Partridge *et al.*, 2000).

1.7.2: The evolution of integrons

The recent discovery of a new integron within the genome of *Vibrio cholerae* has shed significant light on the origin of integrons and their associated gene cassettes (Mazel *et al.*, 1998). The structure, termed a super-integron (SI), contains 179 gene cassettes located adjacent to an integrase gene, and spans 126kb occupying approximately 4% of the 2.5 Mb genome (Heidelberg *et al.*, 2000). Key to this discovery was the identification of a series of homologous repeat sequences within the *V. cholerae* genome, that bore structural similarities to the gene cassette array associated with multi-resistant integrons (MRIs) (Barker *et al.*, 1994; Clark *et al.*, 1997; Mazel *et al.*, 1998). The *Vibrio cholerae* repeats (VCRs) comprise a family of imperfect inverted repeat sequences of 121-126 bp that include an inverse core site (ICS) and a core site (CS) region, the latter of which is identical to the gene cassette-associated consensus GTTRRRY of the 59-be. Additionally, these sequences are located in the same relative orientation as 59-be's, the majority of which are typically separated by a single ORF (Rowe-Magnus *et al.*, 1999).

The VCRs differ from the recombination sequences of their MRI counterparts in that they exhibit a high degree of relatedness, sharing approximately 74% sequence identity (Rowe-Magnus *et al.*, 1999). In contrast, the 59-be sites of antibiotic resistance genes located within MRIs are highly variable both in sequence and length. However, significantly the 59-be sites of two MRI cassettes, *CARB4* and *dfrVI*, were found to be identical to VCRs (Rowe-Magnus & Mazel, 1999). Together with the observation that a class 1 integrase enzyme could utilize SI cassettes as substrates for recombination (Mazel *et al.*, 1998), it was proposed that MRIs evolved from SIs by capturing *intI* genes and their cognate *attI* sites and incorporating them into mobile elements. Gene cassettes were then acquired from different SI sources leading to the development of a variety of MRIs. Consistent with this hypothesis was the demonstration of activity of the *V. cholerae* SI integrase *intI4* (bearing the same structural organization observed for MRIs) by deletion studies using a plasmid-borne selectable marker bordered by two VCRs (Rowe-Magnus *et al.*, 2001).

SIs appear to be widely distributed throughout the γ-proteobacterial radiation having been identified in a variety of bacterial genera that include other species of *Vibrio* (Clark *et al.*,

2000) and *Pseudomonas* species (Vaisvila *et al.*, 2001; Holmes *et al.*, 2003a & b) (See Table 1.1) (Rowe-Magnus & Mazel, 2001). All are characteristically large in size, possess an integrase gene and importantly, exhibit a high degree of homology between their endogenous *attC* sites. Integron-type integrase genes have additionally been identified in other proteobacterial genomes (Leon & Roy, 2003; Holmes *et al.*, 2003 b)). The provision of a functionally significant gene cassette pool is essential to the theory of cassette recruitment and the continued evolution of integrons. In order for SIs to serve as cassette reservoirs, the 59-be sites found within MRIs should originate from a specific SI cluster containing related *attC* sites. Indeed, of sixty-three 59-be sites associated with antibiotic resistance gene cassettes, a fifth have shown to be virtually identical to species-specific *attC* sites within SIs (Rowe-Magnus *et al.*, 2001). The shared sequence similarity observed amongst individual SI *attC* sites is a reflection of the species-specific construction of these elements. Phylogenetic analysis of VMRs (*Vibrio metschnikovii* repeats) within the *V. metschnikovii* SI demonstrates that they cluster to form a related but separate group of sequences from VCRs (Rowe-Magnus *et al.*, 1999). A total of 92 and 70 conserved nucleotide positions have been identified within the VCR and VMR families respectively, of which 50 nucleotides are common to both families. This characteristic would imply the addition of VxRs to ORFs within the *Vibrio* cell. Similarly, the family of repeat sequences associated with the SI of *Pseudomonas alcaligenes* (designated PARs) exhibit an overall sequence identity of 90%; however, they show minimal sequence conservation with VCRs and VMRs beyond the ICS and CS regions (Vaisvila *et al.*, 2001).

Whilst a characterized antibiotic-resistance gene remains to be identified within a SI structure, a number of SI cassettes have been shown to exhibit considerable homology to resistance genes; these likely represent potential progenitor cassettes. Rowe-Magnus *et al.* (2002) demonstrated that a novel ORF within the *V. cholerae* genome could be recruited by an MRI conferring phenotypic antibiotic resistance to the MRI recipient. They introduced a naturally occurring MRI, In3, located on the conjugative plasmid R388 into *V. cholerae*; a functional integrase under the control of an inducible promoter was supplied *in trans* on a second plasmid. Following induction of the integrase, plasmid R388

was transferred to *E.coli* and a variety of antibiotics employed to select for resistant transconjugants. This resulted in the identification of a chloramphenicol resistant isolate containing an MRI within which was located a CAT (chloramphenicol acetyltransferase) gene cassette, *catB9*. This cassette previously identified in the *V. cholerae* SI shared significant homology with other CATS, and the study demonstrated for the first time functionality of a class 1 integron in a host other than *E.coli* and the process of cassette recruitment from SI to MRI.

Concurrence in the relationship of different SI integrase genes with that of the evolutionary relationships of their host species indicate SIs are ancient structures that have likely contributed significantly to bacterial innovation (Rowe-Magnus & Mazel, 1999). Indeed, the degree of homology shared by different integrase classes is 45-58%, inferring integrase sequences along with their cognate *attC* sites diverged well before the onset of the antibiotic era. Unsurprisingly, SIs have also been identified in isolates from the 1800s (Mazel *et al.*, 1998). Thus, for SIs to serve as progenitors to integrons, their cassettes should confer a range of potential adaptive functions extending beyond antibiotic resistance determinants that promote their recruitment into integron structures. Such examples of SI cassettes include a psychrophilic lipase (*Moritella marina* SI), a restriction enzyme (*Xanthomonas campestris* SI) and a sulfate-binding protein (*Vibrio cholerae* SI); additionally, for many of the cassette-associated orfs identified a function has yet to be determined (Rowe-Magnus & Mazel, 2001). The current prevalence of resistance cassettes within integrons simply reflects the success of this system in proliferating adaptive functions, specifically in this case in response to the selective pressure exerted by antibiotic use.

Table 1.1: Bacterial species possessing chromosomal super-integrons and SI *intI*-like genes. Table 1.1 was compiled primarily from Rowe Magnus & Mazel, 2001 with additional information from Nemergut *et al.*, 2004.

Radiation	Strain	
γ-proteobacteria	***Vibrionaceae* & relatives:**	*Vibrio cholerae*
		V. mimicus
		V. metschnikovii
		V. parahaemolyticus
		V. harveyi
		V. hollisae
		V. anguillarium
		V. salmonicida
		V. fischeri
		V. natriegens
		V. vulnificus
		Listonella anguillarium
		L. pelagia
		Alteromonas macleodii
		Photobacterium phosphoreum
		Moritella marina
	Shewanella	*Shewanella oneidensis*
		S. putrefaciens
	Xanthomonads	*Xanthomonas campestris*
		X. oryzae
	Pseudomonads	*P. pseudoalcaligenes*
		P. alcaligenes
		P. mendocina
		P. stutzeri
	Geobacter	*Geobacter sulfurreducens*
		G. metallireducens
		Microbulbifer degradans
		Acidithiobacillus ferrooxidans
β-proteobacteria		*Nitrosomonas europaea*
Spirochaetales		*Treponema denticola*
Planctomycetales		*Gemmata obscuriglobus*

Presently, one of the greatest puzzles remaining concerns the genesis of the cassettes themselves for which a mechanism remains to be identified. The differences observed in GC content and codon usage of cassette-associated ORFs indicate they most likely originate from many different sources (Rowe-Magnus & Mazel, 1999). The extent of homology among *attC* sites specific to each SI would indicate that they are added to ORFs within the host cell. Whilst this would infer an *in vivo* process of cassette capture, generation of the ORF is unclear. Interestingly, for some of the SI ORFS examined to date, their sole homologues are of bacterial, viral and eukaryotic origin (Rowe-Magnus *et al.*, 1999). It has been suggested that VCRs could act as primers for a reverse transcriptase within the *Vibrio* cell (Hall, 1997), however, it is thought unlikely that DNA from a variety of foreign sources could be efficiently transcribed to serve as substrates for the RT. Also anomalous with this theory is the existence of several *Vibrio* cassettes that either possess their own promoter or are in the opposite orientation to the VCR.

The selective pressure imposed by the use of antibiotics in combination with a highly mobile gene acquisition system, has undoubtedly facilitated the capture of many resistance cassettes from a variety of SI sources through multiple gene transfer events. SIs themselves are highly dynamic structures, undergoing both gene acquisition and loss together with cassette rearrangement, all of which is mediated by their cognate integrase (Clark *et al.*, 2000). The generation of MRIs and their contribution to bacterial fitness as evidenced by the emergence of multidrug resistance amongst diverse Gram-negative bacteria underpins the role of integrons in promoting bacterial innovation.

1.7.3: The role of integrons in natural and clinical environments – horizontal gene transfer and cassette diversity.

Horizontal gene transfer is a fundamental process driving bacterial diversity and evolution. The processes of conjugation, transformation, transduction and transposition have enabled bacteria to shuffle and recruit genes, allowing them to respond favourably to local requirements (Ochman *et al.*, 2000). This transfer of genetic material can occur between individual DNA molecules and from one cell to another, transcending species boundaries in the process. As evidenced from whole genome sequencing projects, a large

proportion of bacterial genomes have been acquired from foreign sources (as much as 1/5 of any bacterial genome) demonstrating the important contribution of HGT to bacterial innovation (Lawrence & Ochman, 1998).

HGT has had no more of a profound impact than on the emergence of antibiotic resistance amongst clinically relevant bacteria (Rowe-Magnus & Mazel, 1999; de la Cruz & Davies, 2000; Canton *et al.*, 2003). Detailed examination of multiple resistance determinants carried by plasmids and transposons culminated in the discovery of integrons, revealing a further dimension by which genetic material could be exchanged (Hall & Vockler, 1987; Stokes & Hall, 1989). These elements serve essentially as a molecular cloning kit for bacteria and have played a significant role in the dissemination of antibiotic resistance amongst phylogenetically diverse Gram-negative bacteria. Over 60 different gene cassettes contained within integrons have been identified to date, conferring resistant to a range of antibiotics such as aminoglycosides, penicillins, cephalosporins and trimethoprim (Recchia & Hall, 1995; Leverstein-van Hall *et al.*, 2002). These include multiple distinct gene cassettes for each antibiotic family, that in some cases mediate resistance through different mechanisms. The identification of multi-resistant integrons (MRIs) containing arrays of gene cassette has further confirmed the effectiveness and versatility of integrons as a gene-capture system (Martinez & Freijo *et al.*, 1999).

The acquisition of a random assortment of foreign genes provides great scope for continued adaptation to environmental fluxes. Whilst the acquisition of a particular gene may enhance the fitness of the host organism leading to clonal amplification, much of the success of integrons can also be attributed to their close association with mobile elements. Integrons are defective for self-transposition, however, they are typically located within transposons and on conjugative plasmids facilitating their intra-and inter-species transfer (Brown *et al.*, 1996; Rice, 2002).

The recent discovery of novel environmental integrase sequences in addition to those identified within bacterial genomes and as part of SI structures have revealed integrons to be a common feature of bacterial populations (Rowe-Magnus & Mazel, 2001; Nield *et al.*,

2001 a); Nemergut *et al.*, 2004). Unsurprisingly, extensive gene cassette diversity has also been recovered from natural environments that is not simply confined to antibiotic resistance determinants (Holmes *et al.*, 2003a). Studies conducted by Stokes and colleagues on a pool of environmental gene cassettes recovered by PCR have revealed the majority to possess no database homologues (Stokes *et al.*, 2001; Holmes *et al.*, 2003a). Indeed, of 142 hypothetical proteins identified, only 17% were related to a database sequence and many of the closest matches were also hypothetical proteins. Thus, relating sequence to function for cassette-associated proteins remains one of the current challenges. Interestingly, a number of cassettes were postulated to have a role other than encoding a protein. These cassettes formed distinct families on the basis of sequence identity, though showed little conservation regarding putative orfs. The presence of sequence motifs including imperfect inverted repeats and polyA/polyT tracts indicated that the RNA transcirpts themselves may have a biological role (Holmes *et al.*, 2003a).

Similarly, a study of cassette diversity in heavy-metal-contaminated mine tailings identified 11 novel gene cassettes, 6 of which did not possess a database homologue (Nemergut *et al.*, 2004). Interestingly, a cassette encoding an orf of 143 aa related to a hydroxylaminobenzene mutase was identified. The product of this gene is involved in the catabolism of nitroaromatics that are known constituents of explosives and solvents used in gold extraction. The authors hypothesized that such a gene may enhance metabolic fitness of the host or play a role in detoxification in this particular environment. This discovery clearly demonstrates the potential adaptive functions specific to different environments encoded by gene cassettes.

The versatility of the integron platform in recognizing diverse recombination sequences of different origins, combined with a functionally significant gene cassette pool, provides great scope for the proliferation of adaptive functions and the continued evolution of bacterial genomes. Indeed, the variation identified in genomic location and host species of integrons is evidence of their success in serving as a general gene capture system for bacteria.

1.7.4: Methods for the study of integrons

The investigation of integron diversity within clinical and environmental samples can be divided into two distinct methodological approaches: those that focus on cultivation of bacteria and characterization of integron-containing isolates, and those that involve the extraction and analysis of community DNA. The former has the advantage of linking integron content to the host microorganism through phylogenetic analysis, and the location of the integron unit (i.e. chromosomal, transposon or plasmid-based) can be readily determined. Additionally, by employing selective growth media, bacterial isolates can be cultivated on the basis of phenotypic traits such as antibiotic resistance, effectively serving as an enrichment for integrons. The majority of integrons characterized from clinical environments have been identified from the cultivation of resistant isolates (Levesque *et al.*, 1995; Leverstein-van Hall *et al.*, 2001). However, as with all cultivation studies the information acquired is restricted solely to the culturable fraction of microorganisms. In contrast, analysis of community DNA extracted directly from a sample circumvents the bias posed by cultivation allowing far greater access to functional gene diversity. This approach has commonly been adopted for investigations of integron diversity in natural environments (Stokes *et al.*, 2001; Nemergut *et al.*, 2004). Whilst the entire integron and cassette gene pool can be sampled in this way, the success of this approach is highly dependent on the identification of conserved signature sequences for the design of PCR primers. As a consequence, the possibility exists that novel sequences may be missed due to their lack of complementarity with the primers. Thus unsurprisingly, a polyphasic strategy combining both cultivation and cultivation-independent approaches can be considered most informative.

Three variables should be considered in the investigation of integron diversity within samples: they are, 1) the sequence diversity exhibited by different classes of integrase enzyme, 2) the variation in cassette content, and 3) the location of the integron system i.e. host genome versus 'floating' genome. PCR-based strategies aimed at detecting new integron classes and gene cassettes in environmental DNA have exploited the conserved nature of core parts of the integron system that are necessary for their functionality in diverse genera of bacteria. Three new classes of integron have recently been proposed

following the amplification of soil DNA with primers exploiting conserved sequence at the 3' end of intI1-4 and similarity shared by 59-be sites (Nield *et al.* 2001). Similarly, a further fourteen novel integrase sequences were recovered from heavy metal contaminated mine tailings using integrase-specific primers (Nemergut *et al.* 2004). Extensive gene cassette diversity has also been detected in environmental samples using primers targeting the 59-be recombination sites (Stokes *et al.*, 2001; Holmes *et al.*, 2003a; Nemergut *et al.*, 2004).

Soil samples shown to harbour considerable gene cassette diversity have been further investigated in a cultivation-based approach, the objective to isolate and characterize integron classes present in natural environmental isolates (Holmes *et al.*, 2003a). Enrichment cultures were established and screened at weekly intervals by PCR using a combination of broad-specificity primers targeting intI or 59-be sequences. Integron-positive samples were selected for isolation attempts on solid media and individual colonies screened for integrons by PCR. This resulted in the isolation of two *P.stuzeri* strains, Q and BAM, harbouring integrons containing non-overlapping sets of genes (Holmes *et al.*, 2003b).

Similarly, a cultivation based study was used to investigate the incidence of class I integrons amongst 3000 Gram negative bacteria isolated from an estuarine environment (Rosser & Young, 1999). Bacteria were cultured on media selective for coliforms, *Pseudomonas* spp. and *Vibrio* spp. from which 1000 isolates of each type were subcultured for further study. Integron-containing isolates were initially identified by colony hybridization using an *intI* specific gene probe followed by PCR amplification of the integrase gene. Characterization of the variable region was performed by PCR and sequence analysis using primers targeting both conserved sequence within the 5CS and 3CS region and specific resistance cassettes. Additionally, the 3' conserved segment of integrons was characterized by PCR analysis and dot-blot hybridization, to assess the frequency of carriage of *sulI* and *qacE* genes in association with integrons. The incidence of class 1 integrons was found to be 3.6% and 85 integrons were selected for further investigation. Thirty-six of these were found to possess both the *qacEΔ1* and *sulI* genes in

the 3' conserved segment, whilst 38 contained a *qacE* gene but no *sulI* gene and 11 lacked both the *qacEΔ1* and *sulI* genes. Of 74 integrons for which the variable region was characterized, 40 did not contain any integrated gene cassettes. In the remaining 34 integrons, the *aadA1a* gene was found to be the most prevalent (present in 74% of the integrons) and a further 12 gene cassettes were also identified (Rosser & Young, 1999).

In a unique study conducted by Tennstedt and colleagues (2003), both molecular and cultivation-dependent techniques have been used to great effect in an investigation of integron diversity within a municipal wastewater treatment plant. This study focused on the genomic location of class 1 integrons, examining their association with antibiotic resistant plasmids. A total of 96 different resistant plasmids were isolated by a combination of exogenous plasmid isolation using a GFP-marked *Pseudomonas* sp. as a recipient, and transformation of isolated plasmid DNA obtained from bacterial isolates cultivated on media-containing antibiotics. Individual resistant plasmids and mixed-plasmid DNA preparations were analyzed for integrons by PCR using primers targeting conserved sequence within the 5CS and 3CS region. This resulted in the recovery of integrons harbouring new cassette combinations and novel gene cassettes. Additionally, a wealth of new plasmid diversity was identified providing fresh insight into their role in the dissemination of integrons and gene transfer in natural environments.

Finally, *in silico* analyses have also contributed to our current knowledge of microorganisms harbouring a chromosomal integron. Genome sequencing projects have identified *intI* genes like those associated with SIs in members of the *Gamma-*, *Delta-* and *Beta-proteobacteria* and in *Treponema denticola* (a spirochete) (Rowe-Magnus *et al.*, 2001; Nield *et al.*, 2001). This has provided valuable information regarding the bacterial species harboring SIs and chromosomal integrons and their resource as a reservoir of gene cassettes.

1.8: Aims

This project is concerned with the exploration of genome diversity in Antarctic Dry Valley soils. These soils are noteworthy for the extreme climatic conditions of the

Antarctic Dry Valley environment and their absence of exposure to anthropogenic activities. To date, no-cultivation independent investigation of prokaryote diversity in Antarctic Dry Valley soils has been reported.

The first aim of this study was to investigate the diversity of life present in Antarctic Dry Valley soils, in order to ascertain information regarding the types of microorganisms inhabiting this unique environment. A cultivation-independent approach was used here to examine the species diversity determined for both prokaryote domains (the *Bacteria* and *Archaea*) and the *Eukarya,* in soil samples collected from the Upper Wright Valley and the McMurdo Valley, two different locations in the Dry Valleys. This was studied by use of sequence analysis of phylogenetically-informative ribosomal RNA gene libraries, generated from PCR products amplified directly from DNA extracted from Dry Valley soil samples.

The second aim of this project was to determine species diversity of bacteria cultured under different growth conditions from three different locations. This enabled comparisons to be drawn between the species data from Aim 1 above. Data derived from both Aims 1 and 2 was used to assess the extent of species endemism within Antarctic Dry Valleys.

The third aim of this study was to characterize functional gene diversity using, as a model system, integrons and their associated gene cassettes. As integrons can be detected in a wide range of different bacterial species and are typically located on plasmids, they provide some insight into gene flow in environments. Here, Antarctic Dry Valley bacteria were studied using both cultivated strains and as DNA extracted directly from soil samples. These data were used to determine which adaptive functions were present within integrons in Antarctic soil bacteria. Finally, additional attempts were made to identify the genomic location of integron sequences retrieved in these bacteria.

Chapter 2
Materials & Methods

2.1: Bacterial strains and plasmids

A list of the bacterial strains and plasmids used in this study is shown in Table 2.1.

2.2: Sterilization conditions

Unless otherwise stated, all buffers, reagents, culture media and durable labware, were sterilized by autoclaving at 121°C for 20 minutes.

2.3: Sampling of Antarctic soils

Soil sampling took place during the Antarctic summer in February 2002. Samples ranging from approximately 50g to 250g of surface mineral soil were aseptically collected into autoclaved polypropylene tubes (Nalgene). In the field, samples were stored buried underground to ensure the temperature did not exceed 0°C and following transport to Scott base, samples were stored at -20°C. Subsequent transfer of samples from Scott Base to the University of Waikato, Hamilton, New Zealand, and transport from New Zealand to the U.K. (Dangerous Goods Ltd.) was carried out on dry ice. Whilst not in transit, samples were stored at all times <0°C.

A list of the samples collected in February 2002 is shown in Table 2.2. Dry Valley mineral soil samples were collected from two different locations, the Upper Wright Valley (Figure 2.1) and the Miers Valley (Figure 2.2), located within in the McMurdo Dry Valleys, a large 4000km^2 ice-free area located across McMurdo Sound from Ross Island and the Ross Ice shelf. Sampling within the Miers Valley included a vertical transect spanning an elevation of 651m from the valley floor (169m) to a high saddle on the top of the Marshall Valley (820m). Additional mineral soil samples used in this work were collected from Bratina Island, situated within the Ross Dependency region. Bratina Island is noteworthy for its large number of saline ponds of different salinity that are isolated from the marine environment by a 50m depth of sea ice. Samples from this region were

selected to provide a contrast to 'pristine' Antarctic soil samples; Bratina Island is heavily impacted by human activities and received a number of field parties throughout the 2001/2002 season, of which the permanent base camp was estimated to be occupied for a period of 50 – 80 man-days.

Sample MV11.1 was collected from desiccated surface (0-5cm) mineral soil in the Miers Valley region by D. Cowan during January 1999. Additional samples of surface mineral soil, MV8 and MV12, were also collected at this time.

Table 2.3 details the samples used for the investigation of prokaryote and eukaryote diversity and the investigation of integron sequences.

Table 2.1: Bacterial strains and plasmids used in this work.

Strain/Plasmid	Source/Reference	Characteristic
***E. coli* strain**		
TOP10	Invitrogen Corp.	Str^R
Plasmid		
pSa	Tait *et al.*, 1982	IncW, In6 (Str^R, Sp^R, Cam^R, Sul^R)
pR26	Stanisich & Ortiz, 1976	IncP1
pRN3	Lab stock kindly donated by J. M. Ward	IncN
pTOL	Williams & Worsey, 1976	IncP9
pQR445	Lab stock kindly donated by J. M. Ward	IncQ

Table 2.2: Antarctic soil samples collected during February 2002. (UWV – Upper Wright Valley; MVT – Miers Valley transect; MV – Miers Valley; Brat – Bratina Island.) ND – Not determined.

Region	Sample	GPS	Altitude(m)	Comment
Upper Wright Valley	UWV1	77° 30.693 min S 160° 40.164 min E	1319	In glacial circ: sandstone sample, thin layer of surface snow cover at time of sampling. Surface soil (1-2cm) collected.
	UWV2	77° 30.469 min S 160° 40.164 min E	1341	In glacial circ: basalt sample, thin layer of surface snow cover at time of sampling. Surface soil (1-2cm) collected.
	UWV3	77° 31.376 min S 160° 44.472 min E	~ 914	Mineral soil taken from within sheltered rock cavity of a large sandstone rock near base camp
	UWV4	77° 32.565 min S 160° 43.199 min E	937	In labyrinth structure: white crystals of salt present on surface of soil at time of sampling
	UWV5	77° 32.677 min S 160° 46.079 min E	~ 900	In labyrinth structure: moist sediment adjacent to frozen melt pool. Region comparatively sheltered from valley winds.
	UWV6	ND	~ 950	Exposed hilltop next to campsite. Dry sand with surface cover of larger stones. Surface 1cm very dry, moister soil underneath.
	UWV7	77° 31.597 min S 160° 42.787 min E	981	Water-saturated sand sample from glacial ridge path. Surface 1-2cm of soil deposited directly on glacial ice collected.
	UWV8	77° 31.652 min S 160° 42.988 min E	981	Dry soil sample from within glacial ridge path (not directly upon ice), more exposed than UWV7.

Miers Valley Transect samples	MVT1	78° 05.679 min S 163° 48.271 min E	169	Wet grounds from flood plane, soil temp. –0.2°C.
	MVT2	78°05.670 min S 163° 48.285 min E	169	Fine gravel approx. 20m from MVT1; air temp. -3.9°C, soil temp -0.5°C soil.
	MVT3	78° 05.582 min S 163° 48.324 min E	177	Assorted gravels on moraine below valley slope; air temp. -3.2°C, soil temp. -0.3°C soil
	MVT4	78° 05.541 min S 163° 48.310 min E	183	Fine gravels at base of Northern slope, soil temp. -0.4°C
	MVT5	78° 05.480 min S 163° 48.370 min E	202	Lower Northern slope, soil temp. -0.4°C.
	MVT6	78° 05.398 min S 163° 48.462 min E	234	Northern slope; soil temp. -0.2°C soil, air temp. -4.7°C.
	MVT7	78° 05.324 min S 163° 48. 520 min E	262	Northern slope, soil temp. -0.4°C.
	MVT8	78° 05.184 min S 163° 48.690 min E	333	Northern slope; soil temp. -0.2°C, air temp. -4.9°C.
	MVT9	78° 04.904 min S 163° 48.853 min E	427	Northern slope; soil temp. +0.8°C soil, air temp. -4.7°C.
	MVT10	78° 04.685 min S 163° 49.178 min E	518	Northern slope; soil temp. +1.1°C soil, air temp. -4.0°C.
	MVT11	78° 04.503 min S 163° 49.297 min E	610	Northern slope; soil temp. -0.3°C, air temp. -7.2°C.
	MVT12	78° 03.968 min S	820	Dry, fine gravels from protected rock alcove, approx. 50ft below Miers Valley and Marshall Valley saddle. Soil temp. -3.9°C, air temp. -6.9°C.

Miers Valley	MV14	78° 6.713 min S	238	Red sand sample, surface 0-3cm collected from hill above Miers Valley base camp.
	MV15	78° 5.694 min S 163° ~6 min S	225	Glacial moss sample, rich in biomass.
	MV16	78° 6.138 min S 163° 48.563 min E	~190	Desiccated surface mineral soil from valley floor; soil gave high ATP reading (DAC personal communication).
	MV17	78° 6.049 min S 163° 48.563 min E	~170	Surface 1-2cm of soil collected from beneath a dead algal mat surrounding frozen lake adjacent to base camp.
	MV18	78° 6.049 min S 163° 48.920 min E	~170	Sample taken from mud flats within valley floor.
Bratina Island	Brat1	78° 00.555 min S 165° 33.330 min E	ND	Dry surface (0-2cm) sample from steep edge of hill.
	Brat2	ND	ND	Inland surface soil sample.

Table 2.3: Samples used for the investigation of diversity of *Archaea*, *Bacteria* and *Eukarya*, and for the investigation of integron sequences.

Sample	Application	Year of Sampling
MV11.1	archaeal & bacterial 16S rRNA gene library	1999
MV16	cultivation study	2002
MV17	cultivation study	2002
UWV4	archaeal & bacterial 16S rRNA gene library cultivation study	2002
UWV6	18S rRNA gene library	2002

Samples used for the analysis of integrons:		**Year of Sampling**
Miers Valley:	MV8 & MV12*	1999
Upper Wright Valley:	UWV1-4*, UWV5-8*, UWV1 and UWV4	2002
Miers Valley Transect:	MVT3, MVT5, MVT6, MVT7, MVT8, MVT9, MVT11	2002
Miers Valley:	MV14, MV15, MV16, MV17	2002
Bratina Island:	Brat1, Brat2	2002

N.B. Asterisk (*) denotes pooled Antarctic DNA samples pooled.

Figure 2.1: Photograph of the Upper Wright Valley. The area of glacial-carved channels running along the valley floor was termed the 'labyrinth'.

Figure 2.2: Photograph of the Miers Valley.

2.4: Culture media

2.4.1: Media for growth of *E. coli*

SOC medium was used for the recovery and growth of newly transformed *E. coli* strains. Nutrient broth and nutrient agar were used for bacterial growth in liquid culture and on solid media respectively.

SOC Medium: 2% Tryptone (Oxoid)
0.5% Yeast extract (Oxoid)
10 mM NaCl
2.5 mM KCl
10 mM $MgCl_2$
10 mM $MgSO_4$
20 mM Glucose

Nutrient Broth: 25g l^{-1} Nutrient broth (Oxoid)

Nutrient Agar: 28g l^{-1} Nutrient agar (Oxoid)

2.4.2: Media for cultivation of bacteria from Antarctic soils

The following media were used for isolation of bacteria from Antarctic soil samples:

¼ Strength Nutrient Broth Agar: 6.25g l^{-1} Nutrient broth (Oxoid)
2% Agar (Oxoid)

Starch Agar: 0.5% Potato Starch (Sigma)
2% Agar

Water was added to the above reagents to achieve a volume totaling 90% of the final volume after which the broth was autoclaved. After allowing the broth to cool slightly, the following autoclaved reagents were added aseptically: 1/10 volume of M9 salts and 1/100 volume of calcium/ magnesium salts.

Czapek-Dox Agar (CZD): 25g l^{-1} CZD broth (Difco)
2% agar

(CZD broth ingredients:	Saccharose	30 g l^{-1}
	Sodium nitrate	3 g l^{-1}
	Dipotassium phosphate	1 g l^{-1}
	Magnesium sulphate	0.5 g l^{-1}
	Potassium chloride	0.5 g l^{-1}
	Ferrous sulphate	0.01 g l^{-1})

2.4.3: Addition of selective reagents

After autoclaving and allowing the media to cool to ~55°C, filter-sterilized antibiotic solution was added to the medium as required. Antibiotic final concentrations were: Ampicillin, 100μg/ml agar; kanamycin, 40μg/ml agar; streptomycin, 20μg/ml agar and gentamicin, 10μg/ml agar (supplied by SIGMA). X-Gal (Melford Laboratories Ltd) was used at a final concentration of 40μg/ml. Nystatin and cyclohexamide were used at final concentrations of 20μg/ml and 50μg/ml respectively.

2.5: Solutions

Chemicals were purchased mostly from SIGMA and BDH Laboratory Supplies, and unless otherwise stated, reagents were of general purpose grade.

- 10 X TBE: 108 g l^{-1} Tris (hydroxymethyl) methylamine
  50 g l^{-1} Orthoboric acid
  7.4 g l^{-1} EDTA.
- 1 X TAE: 40 mM Tris (hydroxymethyl) methylamine
  20 mM Acetic acid
  1 mM EDTA (pH 8.3)
- TE: 10 mM Tris-HCl pH8, 1mM EDTA pH8.0.
- M9 Salts 132 g l^{-1} $Na_2HPO_4.12H_2O$
  30 g l^{-1} KH_2PO_4
  5 g l^{-1} NaCl
  10 g l^{-1} NH_4Cl
- 100X Calcium/ Magnesium Salts: 100 mM $MgSO_4.7H_2O$, 10 mM $CaCl_2$

2.6: Direct Extraction of DNA from Antarctic Soils

2.6.1: FastPrep System

This method for extraction of DNA from soil utilizes the FastPrep System (Bio 101, Inc.) and was used for the extraction of total community DNA from sample MV11.1.

A volume of 978 µl of sodium phosphate buffer and 122 µl of MT buffer were added to a FastDNA tube containing 0.5g of soil. Tubes were shaken in a FastPrep instrument at speed 5.5 for two rounds of 30 s followed by centrifugation for 1 min at 14,000 x g. The supernatant was transferred to a clean microcentrifuge tube and 250 µl of PPS reagent added, with mixing by gentle inversion. The mixture was centrifuged at 14,000 x g for 5 min and the supernatant from each tube was transferred to a fresh 15 ml falcon tube. A volume of 1 ml of the Binding Matrix Solution was added to the supernatant and the tubes were shaken gently by hand to permit binding of the DNA to the matrix. The silica matrix was allowed to-settle for 3 min after which 1 ml of the supernatant was removed. The binding matrix was resuspended in the remaining amount of supernatant and a volume of 600 µl transferred to a SPIN™ Filter followed by centrifugation at 14,000 x g for 1 min. The flow-through in the catch tube was discarded and the remaining supernatant was added to the filter followed by centrifugation. A 500 µl volume of SEWS-M wash solution was applied to the filter and centrifuged at 14,000 x g. The flow-through was decanted and an additional centrifugation step of 14,000 x g for 2 min was used to 'dry' the matrix. The SPIN™-Filter was transferred to a fresh catch tube and allowed to air dry for 5 min after which the DNA was eluted in 50 µl of DES (DNase/ Pyrogen Free Water).

Precipitation of DNA: The DNA solutions for each sample were pooled together and 1/10 volume of 3M sodium acetate (pH 5) and 2 x volume of 100% ethanol added. The DNA was left to precipitate overnight at -20°C followed by centrifugation at 4°C for 20 min at 14,000 x g. The DNA pellet was subjected to two rounds of washing with 70% ethanol, dried under vacuum and resuspended in 20 µl of sterile water.

2.6.2: PSC-B Method

PSC-B (Phosphate, SDS, Chloroform-Bead Beater) method for DNA extraction from environmental samples is based on a procedure described by Miller *et al.*, (1999). This method was used for extraction of DNA from Antarctic soil samples collected during February 2002.

0.5 g of sample was resuspended in 300 µl of phosphate buffer (100mM NaH_2PO_4) in a bead-beater vial containing 0.5g each of 0.1mm and 3.0mm silica zirconium beads (Biospec Products, Inc). A volume of 300 µl of SDS lysis buffer (100mM NaCl, 500 mM Tris pH8.0, 10% SDS) was added and the mixture shaken gently, followed by the addition of 300 µl of chloroform-isoamyl alcohol. Vials were shaken in a Mini-bead-beater-8 (Biospec Products, Inc) for 15 s at maximum setting and then centrifuged at 14,000 x g for 10 min. The supernatant (approx. 650 µl) was transferred to a clean microcentrifuge tube and 360 µl of 7M ammonium acetate added (final concentration 2.5 M).The mixture was shaken and then centrifuged at 14,000 x g for 10min. Approximately 580 µl of supernatant was transferred to a fresh tube and 2 µl of linear acrylamide (Ambion, Inc) added followed by 315 µl isopropanol (0.54 volume).The mixture was incubated at room temperature for 15-45 min and the DNA pelleted by centrifugation at 14, 000 x g for 20 min at 4°C. The DNA was washed with 70% ethanol and allowed to air dry for 15-45 min followed by resuspension in 20-50 µl of sterile water. A negative extraction control consisting of a bead-beater vial without the addition of soil was performed on each occasion.

2.6.3: Method for extraction of high molecular weight DNA

High molecular weight DNA was extracted following the protocol described by Henne *et al.*, (1999), based on a method devised by Zhou *et al.*, (1996). This protocol was used for the extraction of DNA from samples MV8 and MV12, which were collected from the Miers Valley in February 1999 by D. A. Cowan (personal communication). High molecular weight DNA was extracted in preparation for the anticipated construction of an environmental DNA library; samples MV8 and MV12 were selected for this purpose due to the large amount of soil that was available for community DNA extraction.

Prior to DNA extraction, 250 ml polypropylene centrifuge bottles were washed thoroughly with the inclusion of an acid-wash step with concentrated hydrochloric acid and autoclaved.

To 50 g of environmental sample was added 135 ml of DNA extraction buffer (100 mM Tris-HCL pH8.0, 100 mM EDTA, 100 mM sodium phosphate, 1.5 M NaCl, 1% (wt/vol) cetyltrimethylammonium bromide) and 1 ml of proteinase K (10 mg/ml); tubes were shaken horizontally at 225 rpm for 30 min at 37°C. Next, samples were incubated with 15 ml of 20% (w/v) SDS in a 65°C water bath for 2 hours with gentle inversion every 15 minutes. Samples were centrifuged at 6,000 x g for 10 min at room temperature and the supernatant transferred to a clean centrifuge tube. DNA extraction was repeated a further two times on the remaining soil pellets by resuspending them in 45 ml of extraction buffer and 5 ml of 20% (w/v) SDS; incubation and centrifugation steps were as described before. An equal volume of chloroform-isoamyl alcohol (24:1 wt/vol) was added to the combined supernatants from each extraction step and the aqueous phase recovered by centrifugation. DNA was precipitated with the addition of 0.6 volume of isopropanol at room temperature for 1 hour. Nucleic acids were pelleted by centrifugation at 9,000 x g for 20 min, followed by washing with 70% (w/v) ethanol. DNA was resuspended in 2-4 ml of sterile water.

2.7: Purification of DNA from Soil

2.7.1: Qiagen Mini-prep plasmid purification system

DNA was extracted from Antarctic soil samples using the PSC-B method described by Miller *et al*, 1999. A preparatory 16S rRNA gene PCR was performed using DNA from each sample as template, in order to confirm the DNA was of suitable quality for PCR amplification. Samples that failed to yield a 16S rRNA product were purified using the following method:

DNA samples were adjusted to a final volume of 250 µl with the addition of sterile water. A 250µl volume of P2 solution was added to 350 µl of P3 solution, mixed by

inversion, and then added to the DNA solution, mixing with gentle inversion. Following centrifugation at 16,000 x g for 5 minutes, the supernatant was removed and applied to a spin column. Subsequent washing of the DNA was carried out according to the manufacturers instructions and the DNA was eluted in a final volume of 50-100 µl of sterile water.

2.7.2: Wizard DNA Clean-Up System.

DNA obtained from soil using the protocol for extraction of high molecular weight DNA was purified using the Wizard DNA Clean-Up System (Promega) according to the manufacturer's instructions. The DNA was eluted in 50 µl of water and then briefly dried under vacuum, after which it was resuspended in 20-50 µl of sterile water.

2.8: Quantification of DNA Extracted from Soil Samples

When measuring the amount of DNA extracted from samples, concentration was determined using the PicoGreen dsDNA quantitation reagent and kit (Molecular Probes, Eugene, OR, USA) according to the manufacturer's instructions. DNA samples were assayed in 96-well plate format using 100 µl DNA solution and 100 µl of the aqueous working solution of the PicoGreen reagent. Readings for each sample were carried out in triplicate using a Fluorocount (excitation 480nm, emission 570nm). Known concentrations of λ DNA were used to generate a standard curve of fluorescence emission intensity relative to DNA concentration and fluorescence values for samples were converted to DNA concentrations. The DNA concentration of some Antarctic DNA extractions is shown in Table 2.4.

2.9: Cultivation of Bacteria from Antarctic Soil Samples

2.9.1: Strain isolation from soil

One to two grams of soil was mixed with an equal volume of sterile water in a sterile 15ml falcon tube and vortexed for ~1 minute. Volumes of between 50 and 100µl were spread on to ¼ strength nutrient broth, CZD and starch agar plates, some of which were additionally supplemented with the antibiotics gentamicin and streptomycin. Plates were incubated at 4°C. In addition, sterile control plates for culture medium of each type used

were stored amongst spread plates to monitor and identify any potential contaminant microorganisms. Pure isolates were obtained by subculturing individual colonies two to three times on fresh culture media.

2.9.2: Storage of bacterial isolates

Bacterial growth from pure isolates was removed from the surface of the agar using a sterile loop and resuspended in the cryo-preservative solution contained within tubes. Tubes were vortexed vigorously, the preservative solution removed by pipetting and tubes stored at -80°C.

Table 2.4: Concentration of DNA in Antarctic DNA extractions

Region	Sample	DNA ng/µl	
Miers Valley		**Set1**	**Set2**
Transect	MVT1	0.51	0.44
	MVT2	5.03	4.35
	MVT3	4.57	3.39
	MVT4	0.58	0.67
	MVT5	9.43	6.02
	MVT6	24.96	18.33
	MVT7	55.85	62.88
	MVT8	16.97	12.46
	MVT9	47.85	54.02
	MVT10	3.96	3.99
	MVT11	55.83	54.88
	MVT12	15.42	12.74
Miers Valley			
	MV8	21.3	
	MV12	1.14	
	MV14	0.62	
	MV15	0.57	
	MV16	0.44-3.86	
	MV17	12.40	
Upper Wright Valley			
	UWV1	4.08	
	UWV2	0.65	
	UWV3	0.58	
	UWV4	2.51	
	UWV6	1.04	
	UWV7	0.88	
	UWV8	0.50	

Amount of DNA (approximate value) used as template for integron PCR:

Sample	ng DNA	Sample	ng DNA
Bed Dust	8	MVT3	23-32
MV8/12	50	MVT5	21-33
UWV4	12	MVT7	55-62
UWV1-4	10	MVT9	47-54
UWV5-8	5	MVT11	55
MV17	12		

2.10: PCR

2.10.1: 16S/ 18S rRNA gene Amplification

2.10.1.1: Preparation of 16S/ 18S rRNA libraries

PCR amplification was performed in a PCR Sprint thermal cycler (Thermo Hybaid). All reactions were carried out in a 50 µl total volume containing 1-10 µl template DNA, 50 pmol (each) primer, 200 µM (each) deoxynucleotide triphosphate, 1.5 mM $MgCl_2$, 1-2.5U Expand High Fidelity (Roche). The typical conditions used consisted of an initial denaturation step of 93°C for 3 min followed by 25 cycles of 94°C for 30s, 55°C for 30s and 72°C for 2.5 min. A final elongation step of 72°C for 10-30 min was included. Sample PCR reactions were carried out in triplicate and the amplicons pooled together prior to cloning.

2.10.1.2: Amplification of cloned 16S rRNA gene sequences for ARDRA.

A 0.5 µl volume of plasmid DNA was used as template for re-amplification of the cloned 16S rRNA insert using the original primers. The PCR reaction consisted of 30 pmol (each) primer, 200 µM (each) deoxynucleotide triphosphate, 1.5 mM $MgCl_2$, 1U *Taq* DNA polymerase (Roche). The PCR program was as follows: 95°C, 4 min; 25 cycles of 94°C for 30s, 55°C for 30s, and 72°C for 1 min.

2.10.1.3: 16S rRNA identification of soil isolates

DNA was obtained from cultured soil isolates using the following protocol:
Single colonies were picked using a sterile loop, resuspended in 20 µl of sterile water and vortexed briefly, followed by centrifugation to sediment the material. DNA was heat denatured in a thermocycler for 5 minutes at 95°C and the mixture centrifuged.

A 1 µl volume of the supernatant was used as template in a PCR reaction (total volume 50 µl) containing 50 pmol (each) primer, 200 µM (each) deoxynucleotide triphosphate, 1.5 mM $MgCl_2$, 5 µl Q solution (supplied with enzyme) and 1U *Taq* DNA polymerase (QIAGEN, Inc). Primer set CC/CD was used to amplify a region of ~600bp of the bacterial 16S rRNA gene. The PCR program was as described in section 2.10.1.2

2.10.1.4: Oligonucleotides

The oligonucleotide primers used for amplification of bacterial and archaeal 16S rRNA gene and eukaryal 18S rRNA gene sequences are shown in Table 2.5. Plasmids containing rRNA gene inserts were sequenced using vector-specific primers M13F and M13R and the following domain specific primers: Archaea – Arch21f, A333f: 5'-YCCAGGCCCTAYGGG-3' (Reysenbach & Pace, 1994), 785f: 5' – GGATTAGATACCCBGGTAGTC-3', Ar958r; Bacteria – 27F, 342f (Primer CC minus M13 tail): CTACGGGRSGCAGCAG-3', CD, 1073r: ACGAGCTGACGACAGCCATG-3' (Edwards *et al.*, 1989); Eukarya – EukA, 516-R (ACCAGACTTGCCCTCC) (Amann *et al.*, 1990).

2.10.2: Amplification of integron sequences

The primers used to amplify integron sequences are shown in Table 2.6. Specific primer combinations used in this work are described in Table 2.7 along with the expected size of product (where applicable) for each primer pair and the annealing temperature used. The PCR reaction was as described in section 2.10.1.3 and contained 1-7 µl of template DNA. Elongation times ranged from 1.5 min for products <1 kb, 2 min for products of 1-2 kb in size and 2.5 min for amplification of the variable region of integrons.

2.10.3: Amplification of plasmid sequences

Table 2.8 provides details of the primers used for amplification of replicon-specific plasmid sequences belonging to the incompatability (Inc) groups IncP, IncN, IncW and IncN. The PCR reaction was as described in section 2.10.1.3 and the PCR program was as follows: 95°C, 4 min; 35 cycles of 94°C for 30s, T_A°C for 30s, and 72°C for 1 min. (NB. The elongation time was increased to 1.5 min for amplification of IncW sequences.)

2.11: Agarose gel electrophoresis

2.11.1: Preparative gel electrophoresis

DNA fragments >0.5 Kb were routinely separated by gel electrophoresis in 1-1.5% (w/v) TAE agarose gels containing 0.05 µg/ml ethidium bromide as described by

Sambrook *et al* (1989). DNA bands within the gel were visualized and photographed under UV light. In some cases DNA was quantified by visual analysis on agarose gels using a known concentration of standard DNA as a reference.

2.11.2: Gel electrophoresis for ARDRA/ RFLP analysis

DNA banding patterns generated by ARDRA/ RFLP analysis and DNA fragments of <0.5 Kb were resolved by gel electrophoresis on 2-3% TBE agarose gels containing 0.05 μg/ml ethidium bromide. ARDRA gels were electrophoresed at constant voltage (80V) for approximately 4-5 hours or overnight at 15-30V.

2.12: Cloning of PCR products

2.12.1: Purification of PCR products

PCR products were purified using the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's instructions. DNA was eluted in 30-50 μl volumes of sterile water. The Qiagen MinElute system was used for purification of PCR products of low concentration with elution of the DNA in 10 μl of sterile water.

Table 2.5: Oligonucleotide primer pairs used for amplification of 16S and 18S rRNA genes.

Primer pair	Sequence 5'-3'	Target	Position	Reference
27F & Un1492R	AGTTTGATCCTGGCTCAG GGTACCTTGTTACGACTT	Bacteria	7-27[a] 1492-1510[a]	Weisburg, 1991. Delong, 1992.
CC & CD	TGTAAAACGACGGCCAGTCCAGACTCCTACGGGAGGCAGC CTT AAC TGC CCC CGG GCG TGT TC	Bacteria	334-356[a] 917-939[a]	Rudi *et al*, 1997.
Arch21f & Un1492R	TTCCGGTTGATCCTGCCGGA GGTACCTTGTTACGACTT	Archaea	7-26[a] 1492-1510[a]	Delong, 1992. Delong, 1992.
Ar4F & Ar958R	TCYGGTTGATCCTGCCRG YCCGGCGTTGAVTCCAATT	Archaea	8-25[a] 958-967[a]	Hershberger *et al*, 1996. Delong, 1992.
EukA & EukB	AACCTGGTTGATCCTGCCAGT TGATCCTTCTGCAGGTTCACCTAC	Eukarya	1-21[b] 1772-1795[b]	Medlin *et al*, 1988.

[a] *E. coli* numbering of 16S rRNA gene.
[b] *Saccharomyces cerevisiae numbering* of 18S rRNA gene.

Table 2.6: Oligonucleotides used to amplify integron-specific and Tn21 sequences.

Primer	Target	Sequence 5'-3'	Accession No.	Position	Reference
Forward primers:					
12b2	*intI1*	CGT(M)GC(R)AACGAGTG(R)CG	M95287	1444-1461	This work
intA	*intI1*	ATCATCGTCGTAGAGACGTCGG	M95287	1367-1388	Rosser & Young, 1999
Brint	*intI1*	GCAACTGGTCCAGAACCTTGAC	M95287	2238-2259	Rosser & Young, 1999
Cass1	*intI1*	TGATCCGCATGCCCGTTCCATACAG	M95287	1857-1881	Rosser & Young, 1999
HS298	*intI1*	ACRTGNGTRTADATCATNGT	M95287	1355-1374	Nield *et al.*, 2001
Int5CS	5' CS	GGTCGATGTTTGATGTTATGGAGCAG	M95287	2369-2394	This work
Int3CS-Rev	3' CS	GCACATAATTGCTCACAGCCAAACTA	M95287	5376-5401	This work
qacE-f	*qacE/ qacEΔ*	GCGAAGTAATCGCAACATCC	M95287	5508-5527	This work
sul1-f	*sulI*	GTATTGCGCCGCTCTTAGAC	M95287	6001-6021	Rosser & Young, 1999
Tn21-f	*tnpR* of Tn21	GTTGCCGTTCCGGGTTCTGGTCGAAGGT	X01298	546-573	This work
ant4	*aadA1*	CTGGCTCGAAGATACCTGCAAGAATGT	X12870	1671-1645	This work
Reverse primers:					
5CS-Rev	5' CS	CTTGCTGCTTGGATGCC	M95287	2354-2388	Levesque *et al.*, 1995
intB	*intI1*	GTCAAGGTTCTGGACCAGTTGC	M95287	2259-2238	Rosser & Young, 1999
Brint5	5' CS	CTGCTCCATAACATCAAACATCGACC	M95287	2369-2394	This work
3CS	3' CS	AAGCAGACTTGACCTGA	M95287	5418-5402	Levesque *et al.*, 1995
Int3CS	3' CS	TAGTTTGGCTGTGAGCAATTATGTGC	M95287	5401-5376	This work
HS286	59-base element	CSGCTKGARCGAMTTGTTAGVC	M95287	4134-4113	Nield *et al.*,2001
qacE-r	*qacE/ qacEΔ1*	AGCCCCATACCTACAAAGCC	M95287	5748-5729	This work
sul1-r	*sulI*	CCGACTTCAGCTTTTGAAGG	M95287	6409-6390	Rosser & Young, 1999

Table 2.7: Primer combinations used to amplify integron and Tn21 sequences, along with the expected size of PCR product (where applicable) and the annealing temperature used.

Target	Primer Pair	Expected size of product (kb)	Annealing temperature (°C)
5' CS	12b2/ 5CS-Rev	911	55
	intA/ 5CS-Rev	978	55
	Brint/ 5CS-Rev	117	55 (environmental DNA)
			60 (colony screening)
	intA/ intB	893	55
	Cass1/ 5CS-Rev	498	55
Variable region (gene cassettes)	Brint/ 3CS	Size dependent on integrated cassettes	55
	Int5CS/ Int3CS		60
	Brint/ HS286		60
	Int5CS/ HS286		57
	HS286/ HS287		55 & 57
3' CS	qacE-f/ qacE-r	241	55
	sul1f/ sul1r	408	55
	qacE-f/ sul1r	902	55
	Int3CS-Rev/ qacE-r	373	55
	Int3CS-Rev/ sul1-r	1034	55
Tn21	Tn21f/ Int5CS	~2800 bp	60
	Tn21f/ ant4	~1800 bp	60-65

Table 2.8: Primers used for amplification of replicon-specific plasmid sequences.

Inc Group	Region	Primer Sequence (5'-3')	PCR positive control	Product Size (bp)	Annealing temp. (°C)	Primer position	Accession no.	Ref.
IncN	*rep*1 *rep*2	AGTTCACCACCTACTCGCTCCG CAAGTTCTTCTGTTGGGATTCCG	pRN3	164	55	32164-32186 32328-32305	AY046276	Gotz *et al.* 1996.
IncP9	*rep*F *rep*R	CCAGCGCGGTACWTGGG GTCGGCAICTGCTTGAGCTT	TOL	399	60	1774-1790 2172-2153	AJ344068	Greated and Thomas, 1999.
IncP1	*trfA*2 1 *trfA*2 2	CGAAATTCRTRTGGGAGAAGTA CYGTTGCAATGCACCAGGTC	pR26	241	57	1334-1355 1574-1555	X00713	Gotz *et al.* 1996
IncQ	*oriV* 1 *oriV* 2	CTCCCGTACTAACTGTCACG ATCGACCGAGACAGGCCCTGC	pQR445	436	60	2328-2347 2763-2743	M28829	Gotz *et al.* 1996
IncW	*oriV*1 *oriV*2	GACCCGGAAAACCAAAAATA GTGAGGGTGAGGGTGCTATC	pSA	1140	58	1203-1184 64-83	U12441	Gotz *et al.* 1996

2.12.2: Ligation and transformation reactions

Purified PCR products were ligated into the pCR2.1 vector supplied with the TA cloning kit (Invitrogen). The procedure was carried out essentially as described in the manufacturer's instructions as outlined below.

A ratio of 2 µl of PCR product was ligated to 0.5 µl TOPO vector with the addition of 0.5 µl of salt solution and incubated at room temperature for between 5 to 30 minutes. A volume of 2-3 µl of ligation reaction was added to a vial of One Shot® Chemically competent *E. coli* and incubated on ice for 45 minutes. Cells were heat shocked at 42°C for 90 seconds, transferred immediately to ice and 250 µl of room temperature SOC medium added. Transformation reactions were incubated with shaking for 1 hour at 37°C.

The following amendments were used for transformation by electroporation: 1) a ¼ dilution of the salt solution was used in the ligation; 2) 1-2 µl of DNA was added to 50 µl of competent cells and transferred to a chilled, 0.1cm electroporation cuvette (Invitrogen). The cells were electroporated at an electrical field strength of 18.0 kV/cm.

Volumes of between 40-100 µl of cells were plated on to nutrient broth plates containing the appropriate antibiotic. For transformations requiring blue/white screening of recombinants, a volume of 40 µl of X-gal was added at a final concentration of 40 mg/ml. All plates were incubated overnight at 37°C.

2.12.3: Preparation of competent cells

Competent cells were prepared according to the following protocol:

i) 5 ml of a fresh culture of *E.coli* strain was inoculated into 100 ml of LB broth contained in a 1-L flask. The cells were incubated with shaking at 37°C until the OD reached ~0.6 units.

ii) The cells were divided into two sterile 50 ml falcon tubes and incubated on ice for 10 minutes, followed by harvesting of the cells by centrifugation at 3500 rpm for 5 minutes.

iii) After thoroughly decanting the broth, cell pellets were resuspended in 50 ml of ice cold 0.1M $MgCl_2$.

iv) Harvesting of cells by centrifugation was repeated as above and the cell pellets resuspended in 5 ml of ice cold $CaCl_2$ (100mM) containing 15% (v/v) glycerol. Cells were incubated on ice for at least 2 hours.

v) The supernatant was removed by centrifugation and the cell pellets resuspended in 1 ml of ice cold $CaCl_2$ (100mM: 15% v/v glycerol). Cells were divided into 100 µl aliquots and stored at -80°C.

2.13: Analysis of Clone Libraries

2.13.1: Isolation of plasmid DNA

Randomly selected white colonies were picked using a sterile pipette tip and inoculated into 5 ml volumes of nutrient broth containing the appropriate antibiotic. Cultures were incubated overnight with shaking (200 rpm) at 37°C. Plasmid DNA was extracted from 2-5 ml overnight cultures using the Qiagen Plasmid Mini-Prep kit according to the manufacturer's instructions. Plasmid DNA was eluted in a final volume of 50 µl of sterile water.

2.13.2: Restriction enzyme analysis of DNA

All restriction enzymes used in this work were obtained from New England Biolabs and were used in conjunction with the recommended buffer.

2.13.2.1: Identification of plasmid inserts

The following restriction digest was set up to confirm recombinant plasmids harbouring inserts of the correct size: 3 µl of plasmid DNA, 1 µl of restriction enzyme (typically *Eco*R1) and 1 µl of the appropriate 10 X restriction enzyme buffer made up to a final volume of 10 µl with sterile water. Digests were incubated at 37°C for ~2 hours.

2.13.2.2: ARDRA analysis

16S rRNA gene amplicons of correct size and purity as confirmed by preparative gel electrophoresis were subject to digestion with two tetrameric restriction enzymes, *Hin*P1

and *Msp*I. The restriction digest contained the following components: 10 µl of PCR product, 1.5 µl of restriction enzyme and 2 µl of 10 X restriction enzyme buffer made up to a final volume of 20 µl with sterile water. Digests were incubated at 37°C for 5 hours or overnight.

2.13.2.3: RFLP analysis of plasmid DNA

Plasmids containing inserts of the desired size were typed by digestion with two tetrameric restriction enzymes, typically *Hin*P1 and *Rsa*1. The restriction digest comprised the following: 2 µl of plasmid DNA, 2 µl of restriction enzyme, 2 µl of 10 X restriction enzyme buffer and 14 µl of sterile water. Digests were incubated at 37°C overnight.

2.14: DNA Sequencing

DNA sequences were determined by Oswel DNA services (Southampton) during January 2001 to May 2001. Subsequent DNA sequencing was carried out by Wolfson Institute for Biomedical Research (University College London) using a Beckman Coulter CEQ2000XL Sequencer.

2.15: Analysis of Sequence Data and Phylogenetic Reconstruction

Editing of DNA sequences and sequence assembly were carried out using the EditSeq and SeqMan programs from the DNAStar package (Windows 3.80d1 1989-1994 DNASTAR, Inc). Closest relatives of rRNA sequences were identified from database searches utilizing the BLAST (NCBI), FASTA (EBI) and Sequence Match (RDP) facilities. Reference sequences used in the phylogenetic reconstruction were obtained from the RDP and GenBank databases. Alignment of rRNA sequences was performed using the ClustalW server at the Pasteur Institute (http://bioweb.pasteur.fr/seqanal/interfaces/clustalw.html). Evolutionary distances were calculated by the method of Jukes-Cantor and phylogenetic trees were constructed by the neighbour-joining method with TREECON for Windows (van de Peer & Wachter, 1994). A bootstrap analysis consisting of 100 replicate datasets was performed. This method of phylogenetic reconstruction has been used by O'Sullivan *et al.*, (2002) and

Taton *et al.*, (2003). All 16S rRNA gene clone sequences were checked for chimera formation with the Check_Chimera software of the RDP.

Open reading frames were located within DNA sequences using the ORF finder at NCBI. Putative ORF's were analysed by PHI and PSI-BLAST analyses and multiple protein sequence alignments were achieved using CLUSTAL W.

Chapter 3

Bacterial Diversity

3.1: Introduction

The aim of the work described in this chapter was to characterise the diversity of bacteria present in Antarctic Dry Valley soils based on the sequence types identified following both *in vitro* cultivation and through molecular biological based analyses. The objectives of this work were to compare the diversity identified by both cultivation and cultivation-independent approaches, and to draw conclusions regarding the types of microorganisms inhabiting these Dry Valley soils. It was predicted that the diversity of cultivated bacterial sequences would be much lower than diversity inferred by the use of molecular biological methodologies. Similarly, the types of bacteria cultured were expected to exhibit greater similarity to microorganisms previously cultured than those present only as retrieved sequences from database searches.

3.2: Culturable Diversity

3.2.1: Background

Antarctic Dry Valley soils contain no macroscopic plant life and consequently, are low in organic matter content, the presence if which indicates both the availability of organic sources of carbon and nitrogen along with other nutrients required for growth. The lack of organic matter has been shown to limit microbial growth in Ross Desert soils when water does not (Vishniac, 1993). Studies of Dry Valley soils from the Victoria Valley have demonstrated that, whilst relatively large numbers of microorganisms can be cultivated from younger soils, the influence of the microbial population on soil composition is barely detectable. Levels of organic carbon and nitrogen were found to be extremely low even in the biologically richest of soils, at approximately 0.02-0.04% and 0.002-0.004% respectively (Cameron & Conrow, 1969; Campbell and Claridge, 1987).

Thus, due to the oligotrophic nature of Dry Valley soils, three different types of cultivation media were selected that were aimed at encouraging the growth of slower-

growing organisms as opposed to faster-growing organisms, the latter of which predominate on and contribute to the reduced diversity observed on nutrient-rich media (Hugenholtz, 2002). The growth media used in this study were as follows: one quarter-strength nutrient broth agar (1/4 NB), Czapek-Dox (CZD) agar and starch agar.

CZD broth is a base used for the cultivation of bacteria (especially actinomycetes) and fungi that are capable of utilizing an inorganic source of nitrogen, in this case provided in the form of sodium nitrate (Stergiopoulos Stergiopoulos *et al.*, 2002)2002). Whilst CZD medium contains a high amount of sucrose as the sole carbon source, the presence of nitrate can prevent the growth of more rapidly growing bacteria at the expense of slower growing organisms, which can occur when ammonium ions are available. Starch agar is a minimal medium containing organic salts and starch as the sole carbon source; hence, only microorganisms capable of degrading starch can grow. The use of polymers as growth substrates may reduce the occurrence of substrate-accelerated death observed when microorganisms from oligotrophic environments are exposed to high substrate concentrations, since the polymer must first be hydrolysed before it can be utilized (Postgate & Hunter, 1964; Koch, 1997; Sait *et el.*, 2002). Additionally, cultivation of microorganisms on the above types of media supplemented with the antibiotics gentamicin and streptomycin was also carried out. The objective of this work was to isolate antibiotic resistant bacterial species for integron screening, in order to complement the study of integron diversity in Antarctic soils discussed in Chapter 5.

Two of the samples selected for cultivation studies were designed to complement the molecular assessments of bacterial diversity in Antarctic soils. Sample MV16 is a desiccated surface soil sample collected from the valley floor of the Miers Valley and is analogous to sample MV11 from which the first bacterial 16S rRNA gene library that was constructed. (Sample MV16 was used in place of MV11.1 due to its more recent collection and limited handling, thus minimising loss of bacterial viability.) Sample UWV4 is a surface soil sample collected from within the labyrinth structure of the Upper Wright Valley (described in Chapter 2) and was used for a molecular assessment of bacterial diversity within this region. At the time of sampling, white crystals of salt

were observed on the surface of the soil. Sample MV17 is a surface soil sample collected from beneath an algal mat surrounding a small pond in the Miers Valley, and was selected for comparison of culturable diversity with MV16 and UWV4 as it represents a more nutrient-rich and sheltered environment. A single incubation temperature of 4°C was used in this study to reflect as closely as possible the temperature conditions experienced by the microbial community in its natural environment.

3.2.2: Identification of isolated bacteria

Partial sequences of the 16S rRNA gene were used to identify the cultured isolates. DNA was isolated from single colonies by heat denaturation and used as template in a PCR with primer set CC and CD. This primer set is specific for the Domain Bacteria, amplifying both Gram-positive and Gram-negative bacteria, with the forward primer, CC, having the addition of an M13-21 tail suited for automated sequencing (Rudi *et al.*, 1997). The primers amplify a region of about 600bp (nt 334-939 *E. coli* numbering) spanning variable regions V6, V7 and V8. Closest database relatives of the 16S rRNA sequences were identified by BLAST (Altschul *et al.*, 1990) and FASTA (Pearson, 1994) analyses.

Isolates were named in accordance with the medium they were cultivated on (NB: ¼ strength nutrient broth, ST: starch, CZD: Czapek-Dox), the presence (if any) of antibiotic in the media (Gm; gentamicin, Sm: streptomycin) and the sample from which they were cultivated from (17: MV17, 16: MV16, 4:UWV4).

3.2.3: Culturable diversity in the Miers Valley

3.2.3.1: Diversity on ¼ strength nutrient broth agar

Bacterial growth from sample MV17 on ¼ strength nutrient broth agar (¼ NB) was noted within one week of inoculation at 4°C. Viable counts observed ranged from 8 x 10^3 to 2 x 10^4 cells.g^{-1} wet wt for plates inoculated with 50µl and 10µl of bacterial suspension respectively. Growth on ¼ NB agar supplemented with gentamicin was slower to appear with visible colonies being recorded after eleven days. (No CFU data

available). Figure 3.1.a) shows the growth observed on the original inoculated plates, ¼ NB (50µl) and ¼ NB supplemented with gentamicin, following approximately 12 weeks incubation at 4°C.

Three different colony morphologies were initially observed on ¼ NB agar. Of these, two colony types could be successfully subcultured and a total of four isolates were selected for identification based upon colony morphology of pure isolates. Isolates were preliminarily identified from partial 16S rRNA sequences using the BLAST search program. The level of 16S rRNA identity for each isolate to a database sequence was as follows: Isolate NB 17-2 (white/ cream colony, 2-3mm diameter) shared 100% identity to a *Pseudomonas* species, isolate NB 17-6 (shiny yellow colony) shared 99% identity to an *Oxalobacter* species, isolate NB 17-3 (indistinguishable in colony morphology from NB 17-2) shared 100% identity to *Psychrobacter aquatica*, and isolate NB17-MY (matt yellow in colour) exhibited 99% identity to *Arthrobacter agilis*. The closest database relatives of these isolates are shown in Table 3.1. In contrast, only two colony types were observed and successfully subcultured on ¼ NB agar supplemented with gentamicin, both of which were highly pigmented as shown in Figure 3.1.a) ii). The yellow-coloured isolate, designated NB Gm 17-4 shown in Figure 3.1.b), shared 98% identity to a *Flavobacterium* species, whilst the orange-coloured isolate, NB Gm 17-5 and shown in Figure 3.1.c), shared 100% 16S rRNA identity to *Flavobacterium micromati* (Table 3.1). Bacterial growth was slower to appear on nutrient broth agar containing streptomycin with very small 'pin-prick' colonies forming only after 17 days of incubation at 4°C. Individual colonies were re-streaked following 30 days of incubation but failed to be subcultured, despite reducing the concentration of streptomycin from 20µg/ml to 5µg/ml. The viable count observed at this point was approximately 1 x 10^3 cells.g^{-1} wet wt.

Growth on ¼ NB agar from sample MV16 was noted after 11 days of incubation at 4°C. In excess of 300 colonies were observed on the plate inoculated with 100µl of bacterial suspension; however, all appeared to be highly similar in morphology. A greater diversity of colony morphologies was observed on the plate inoculated with 75 ul of

bacterial suspension. Viable counts observed ranged from 6 x 10^2 cfu/g to 3 x 10^4 cells.g^{-1} wet wt.

A total of four colony morphologies were initially identified and all were successfully subcultured. Five isolates were selected for identification, four of which were found to possess the following 16S rRNA identities: isolate NB 16-A shared 100% identity to a *Bacillus* species, isolate NB 16-C exhibited 99% identity to a glacial ice bacterium, isolate NB16-D shared 93% identity to a *Streptomyces* species, and isolate NB 16-EY shared 99% identity to a *Sphingobacterium* species. These isolates are pictured in Figures 3.1.d.i) and ii). Sequencing of the PCR product obtained for isolate NB 16-B (Figure 3.1.d.i) failed and was not pursued further. Isolates NB 16-C, NB 16-D, NB 16-EY could be grown in subculture within 3 days.

A single colony morphology was observed on ¼ NB agar supplemented with streptomycin but was found to be identical in morphology to isolate NB 16-EY following subculture, and thus, was not formally identified. The viable count observed on this medium was estimated to be 1.6 x 10^2 cells.g^{-1} wet wt. A number of small red colonies were observed on ¼ NB agar supplemented with gentamicin, however, these could not be subcultured.

3.2.3.2: Diversity on starch agar

The cultivation of bacteria from sample MV17 on starch agar was the slowest to appear of the three types of medium used, with small 'pin-prick' colourless colonies first observed following 17 days of incubation at 4°C. Following approximately four weeks of incubation, colonies were sufficiently large that four different morphological types could be identified. The viable count observed at this point was 1.5 x 10^4 cfu/g. Each colony type could be successfully subcultured and a total of four isolates were identified. Isolate ST 17-OC (large orange colony) was 100% identical to a *Sphingomonas* species whilst isolate ST 17-PYP (pale-yellow-pastel type) shared 99% identity to *Curtobacterium fangii*. Isolate ST 17-YC (translucent yellow) shared 99% identity to *Janthinobacterium agaricidamnosum* and isolate ST 17-2 exhibited 100% identity to a

Pseudomonas species. Details of closest database relatives are shown in Table 3.2. The *Pseudomonas* species was most numerically dominant followed by *J. agaricidamnosum*, *C. fangii*, and finally the *Sphingomonas* species.

Surprisingly, the amount of growth on starch agar supplemented with either gentamicin or streptomycin was about twice that observed for starch agar without antibiotic. A possible explanation for this phenomenon could be that the bacteria cultivated on this media were utilizing the antibiotic as an alternative carbon source. The diversity of colony morphologies cultured on starch agar supplemented with either gentamicin or streptomycin were highly similar to those identified on starch agar only and are pictured in Figure 3.1.e) Each of the four colony types described above were observed and for this reason, were not formally identified by 16S rRNA sequence analysis.

The initial attempt at culturing bacteria from MV16 on starch agar supplemented with streptomycin and gentamicin yielded confluent growth but of exceptionally small colonies that could not be differentiated; no growth was observed on starch agar alone. Culturing was repeated using a smaller volume of bacterial suspension and the starch agar was supplemented with 0.1% yeast extract, the objective of which was to increase colony size. Growth on this occasion was first noted following 11 days of incubation at 4°C on starch agar with and without gentamicin. Viable counts observed were 2.0×10^3 cells.g^{-1} wet wt for starch agar with gentamicin and 2.6×10^3 cells.g^{-1} wet wt for starch agar. Overall, five colony morphologies were observed and a total of six isolates were obtained in pure culture and identified. Figure 3.1.e) shows the bacterial growth obtained on starch agar. Two isolates cultivated on starch containing gentamicin that were yellow in colour, designated ST 16 Gm-M and ST 16-Gm-N, shared 99% identity to species of *Arthrobacter*. Isolates ST 16-7 and ST Sm 16-WP shared 92% and 99% identity respectively to *Streptomyces* species, whilst isolate ST 16-8 exhibited 99% identity to a bacterium cultivated from a municipal wastewater treatment bioreactor (Table 3.2).

3.2.3.3: Diversity on CZD agar

Bacterial growth from sample MV17 on CZD agar was first observed approximately ten days following inoculation at 4°C. Two distinct colony types were apparent, an off-white type and a pale-brown type, both of which were very small (~1mm diameter). The viable count observed on this media was 3 x 10^3 cells.g^{-1} wet wt. Each colony type could be successfully subcultured and two isolates were identified as the following (Table 3.3): CZD 17-D shared 99% identity to *Arthrobacter agilis*, whilst isolate CZD 17-4 exhibited 100% identity to *Psychrobacter aquatica*; both isolates were also cultivated on ¼ strength nutrient broth agar. Growth on CZD agar containing gentamicin was first noted following 16 days of incubation, however, the plate had to be discarded due to ice damage during incubation. No growth was observed on CZD agar supplemented with streptomycin following seven weeks of incubation and the plate was subsequently discarded due to evidence of contamination.

MV16 produced significant bacterial growth on CZD agar with at least 5 morphological types visible (Figure 3.1.f). Viable counts observed were estimated at 4 x 10^4 cells.g^{-1} wet wt for CZD agar alone, and 2-3 x 10^3 cells.g^{-1} wet wt and 3 x 10^3 cells.g^{-1} wet wt for CZD with gentamicin and streptomycin respectively. Four colony types were subcultured and two isolates identified by 16S rDNA analysis (Table 3.3). Isolate CZD 16-W shared 100% identity to *Bacillus macroides* and isolate CZD 16-Z shared 99% identity to an *Arthrobacter* species (Figure 3.1.g). A further isolate, CZD-Y, and three isolates cultivated on CZD with gentamicin could not be identified as they failed to yield a PCR product despite several attempts. Possible explanations for this include the presence of a substance within the colony debris that was inhibitory to the PCR reaction, or failure of the primers to bind and amplify the template DNA.

3.2.4: Culturable diversity in the Upper Wright

3.2.4.1: Diversity on ¼ strength nutrient broth agar

Bacterial growth on ¼ NB agar from sample UWV4 was first observed following 25 days of incubation at 4°C. The viable count observed was estimated at 6.7 x 10^3 cells.g^{-1} wet wt. Two colony morphologies predominated: an orange/ red colony, typically 1mm

in diameter, designated NB UW-RC1, and a large (3-4mm diameter) cream-coloured 'gloopy' colony, designated NB UW-GW; these isolates shared 100% identity to *Rathayibacter tritici* and 99% identity to a *Pseudomonas* species respectively. A matt yellow colony, NB UW4-3, exhibited a 100% identity to a species of *Acinetobacter* Bacterial growth cultivated from sample UWV4 on ¼ NB agar is shown in Figure 3.1.h) and details of closest relatives are given in Table 3.4. All isolates could be grown in subculture within two weeks. A single colony type was observed on ¼ NB agar containing gentamicin approximately four weeks following inoculation. The large (3-4mm diameter) 'shiny' red colony, NB Gm 4 LR, shared 99% identity to a species of *Hymenobacter* (Table 3.4); however, it could not be subcultured despite repeated attempts. A single colony type was observed on ¼ NB agar containing streptomycin but was found to be identical to a contaminant identified on the corresponding control plate, and for this reason was not investigated further.

3.2.4.2: Diversity on starch agar

No growth was observed on starch agar and starch supplemented with gentamicin following 12 weeks of incubation at 4°C. Growth was recorded on starch agar supplemented with streptomycin after 41 days of incubation whereby a single, small, light pink colony was observed but could not be subcultured. A small white colony, ST Sm UW4-SW, sharing 99% identity to a *Pseudomonas* species was later identified following further incubation at 4°C (Table 3.5).

3.2.4.3: Diversity on CZD agar

No bacterial growth on CZD agar with and without streptomycin was observed until approximately 12 weeks after inoculation; however, the growth was suspected to be the result of contamination and the plates were subsequently discarded.

A total of fourteen different isolates were identified following cultivation on ¼ NB agar, representing six isolates from sample MV17 and four each from samples MV16 and UWV4. These included two isolates cultivated on ¼ NB agar containing gentamicin from sample MV17 and a third from sample UWV4; however, the latter (NB 4Gm LR)

could not be grown in subculture. Eleven isolates were identified from cultivation on starch agar, representing four, six and one isolate from samples MV17, MV16 and UWV4 respectively. Five of these isolates were cultivated on starch agar containing antibiotic (three with streptomycin and two with gentamicin) and included four from sample MV16 and the single isolate obtained from UWV4. A total of four isolates were identified following cultivated on CZD agar, representing two each from samples MV17 and MV17. The viable count data obtained for each sample and on each type of growth medium used is summarized in Table 3.6.

Figure 3.1: Bacteria cultivated from Antarctic Dry Valley soil samples.

Figure 3.1.a: Bacteria cultivated from sample MV17 on i) ¼ NB agar, and ii) ¼ NB agar supplemented with gentamicin, following 12 weeks of incubation at 4°C.

Figure 3.1.a.i) MV17 ¼ NB

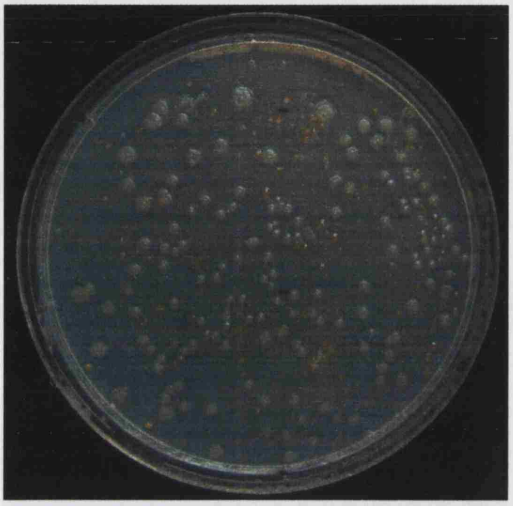


Figure 3.1.a.ii) MV17 ¼ NB + Gm

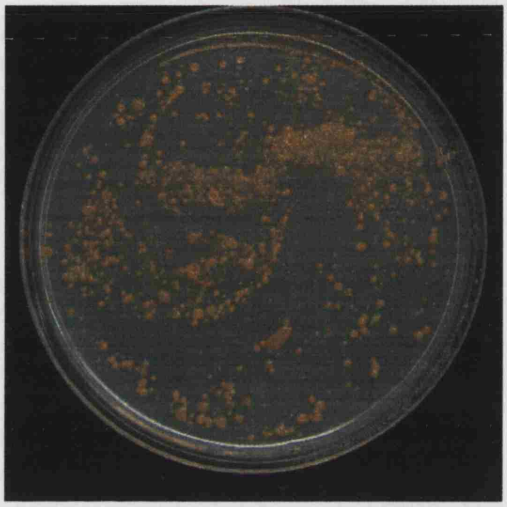


Figure 3.1.b) Isolate NB Gm 17-4

Figure 3.1.c) Isolate NB Gm 17-5

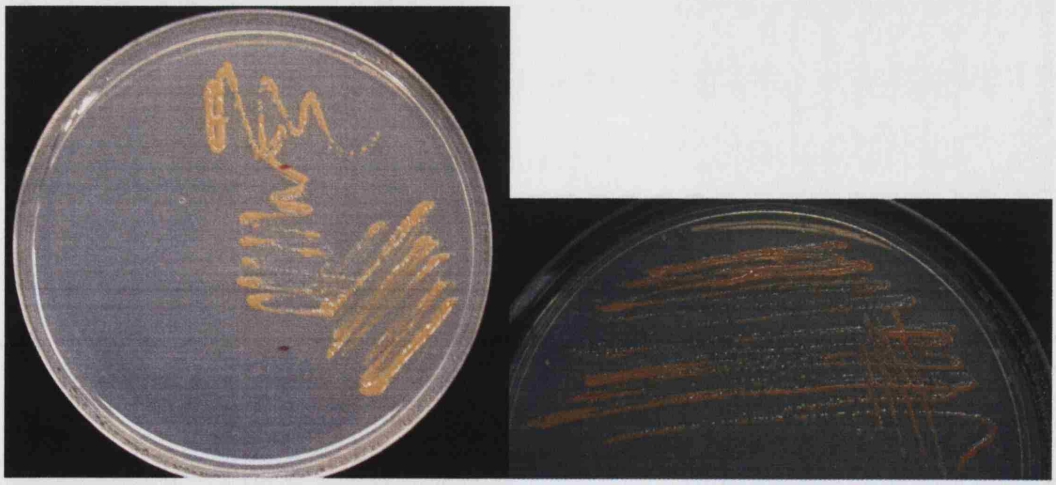


Figure 3.1.d) Bacteria cultivated from sample MV16 on ¼ NB agar: i) Clockwise from the top - isolates NB 16-B, NB16-A, NB-16D and NB 16-C; ii) Isolate NB 16-EY, a large 'glupey' yellow colony.

Figure 3.1.d. ii)

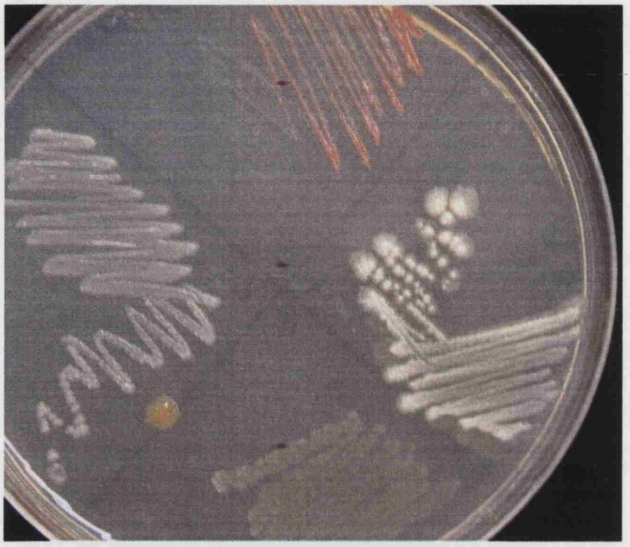


Figure 3.1.d.ii)

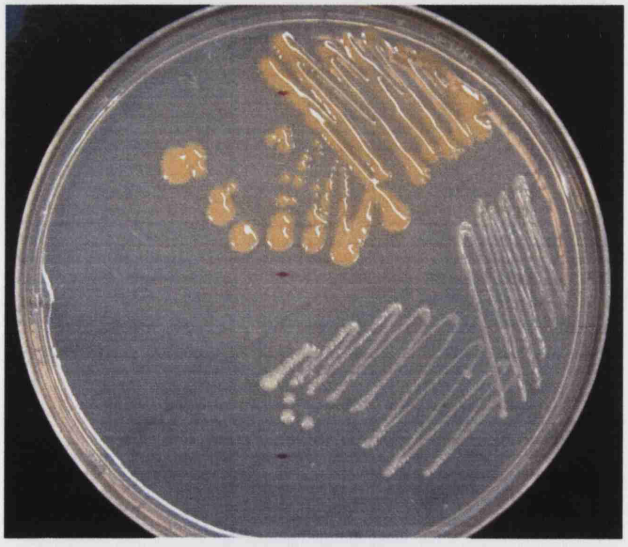
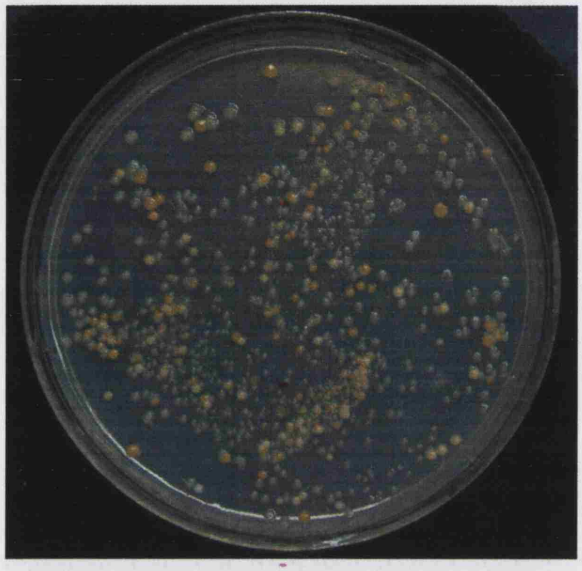


Figure 3.1.d) Bacterial cultivated from sample MV17 on i) starch agar supplemented with gentamicin, and ii) starch agar supplemented with streptomycin. The colony types observed on each medium were highly similar.

i) Starch agar with gentamicin



ii) Starch agar with streptomycin

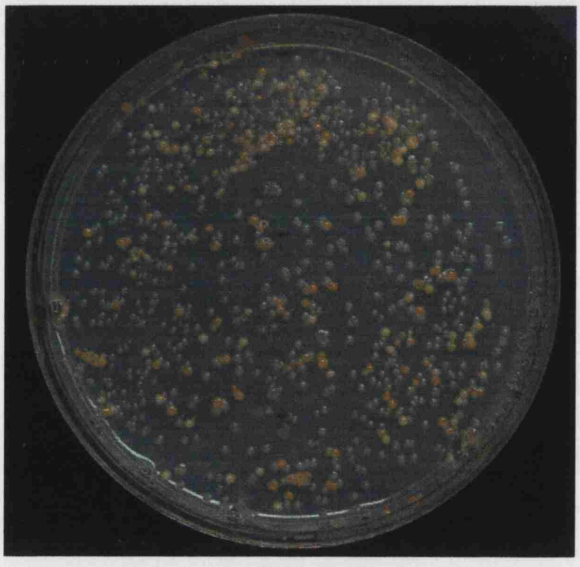


Figure 3.1.e) Bacteria cultivated from sample MV16 on starch agar.

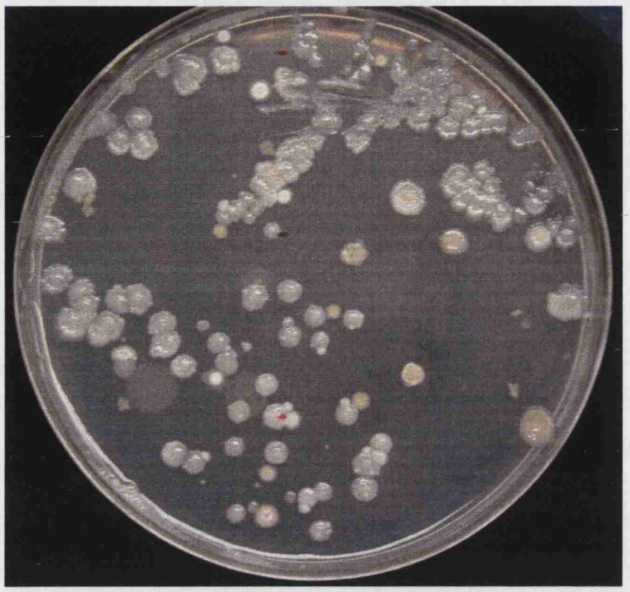


Figure 3.1.f) Bacteria cultivated from sample MV16 on CZD agar.

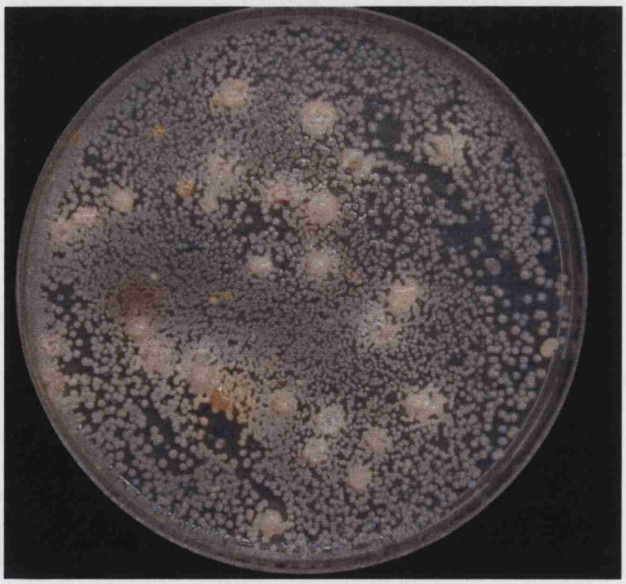


Figure 3.1.g) Isolates CZD 16-W and CZD 16-Z.

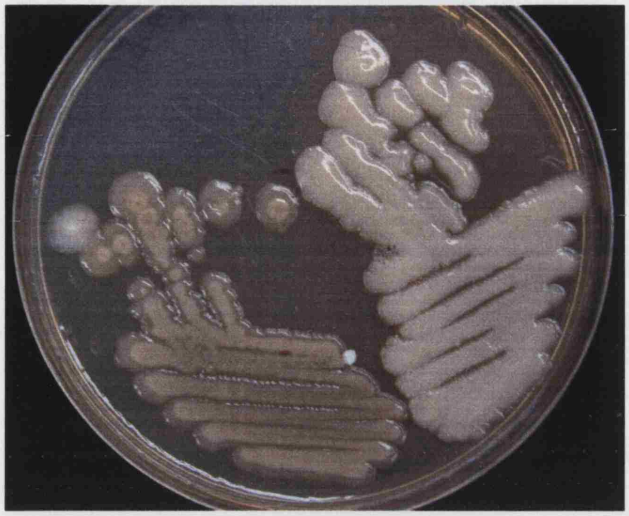


Figure 3.1.h) Bacterial growth cultivated from sample UWV4 on ¼ NB agar.

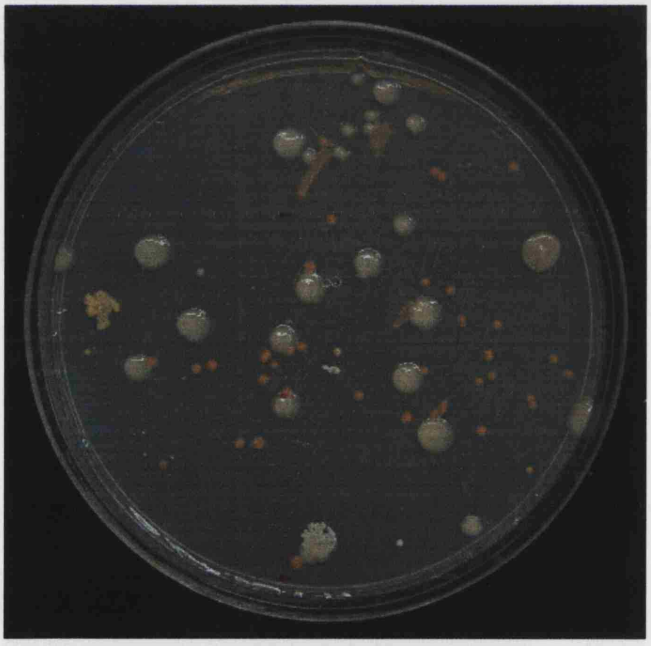


Table 3.1: Culturable diversity recovered from Miers Valley soil samples on ¼ strength nutrient broth.

Phylum	Class or subclass	Isolate name	Sequence length	% identity	Accession no. of closest relative	Closest relative
Proteobacteria	*Gamma*	NB 17-2	541	100	AY121983	*Pseudomonas sp.*
Proteobacteria	*Gamma*	NB 17-3	489	100	AJ584833	*Psychrobacter aquatica*
Proteobacteria	*Beta*	NB 17-6	440	99	AJ496038	*Oxalobacter sp.*
Actinobacteria	*Actinobacteridae*	NB 17-MY	473	99	AF134184	*Arthrobacter agilis*
Bacteroidetes	*Flavobacteria*	NB Gm 17-4	509	98	AF468432	*Flavobacterium sp.*
Bacteroidetes	*Flavobacteria*	NB Gm 17-5	474	100	AJ557888	*Flavobacterium micromati*
Firmicutes	*Bacillales*	NB 16-A	270	100	AJ316309	*Bacillus sp.*
Actinobacteria	*Actinobacteridae*	NB 16-C	495	99	AF479339	Glacial ice bacterium
Actinobacteria	*Actinobacteridae*	NB 16-D	517	93	X81574	*Streptomyces sp.*
Bacteroidetes	*Sphingobacteria*	NB 16-EY	280	99	AY167837	*Sphingobacterium sp.*

Table 3.2: Culturable diversity recovered from Miers Valley soil samples on starch agar.

Phylum	Class or subclass	Isolate name	Sequence length	% identity	Accession no. of closest relative	Closest relative
Proteobacteria	*Alpha*	ST 17-OC	448	100	AF395031	*Sphingomonas sp.*
Actinobacteria	*Actinobacteridae*	ST 17-PYP	451	99	AY273209	*Curtobacterium fangii*
Proteobacteria	*Beta*	ST 17-YC	508	99	AY167838	*Janthinobacterium agaricidamnosum*
Proteobacteria	*Gamma*	ST 17-2	532	100	AY121983	*Pseudomonas sp.*
Actinobacteria	*Actinobacteridae*	ST 16-8	395	99	AF538744	Bacterium cultivated from a municipal wastewater treatment bioreactor
Actinobacteria	*Actinobacteridae*	ST 16-7	417	92	X81574	*Streptomyces sp*
Actinobacteria	*Actinobacteridae*	ST Sm 16-WP	489	99	AF401982	*Streptomyces sp.*
Actinobacteria	*Actinobacteridae*	ST 16 Gm-M	517	99	AY176765	*Arthrobacter sp.*
Actinobacteria	*Actinobacteridae*	ST16 Gm-N	428	99	AY191847	*Arthrobacter sp*
Actinobacteria	*Actinobacteridae*	ST 16 Sm C	459	99	AJ309929	*Microbacterium aerolatum*

Table 3.3: Culturable diversity recovered from Miers Valley soil samples on CZD agar.

Phylum	Class or subclass	Isolate name	Sequence length	% identity	Accession no. of closest relative	Closest relative
Actinobacteria	*Actinobacteridae*	CZD 17-D	486	99	AF134184	*Arthrobacter agilis*
Proteobacteria	*Gamma*	CZD 17-4	210	100	AJ584833	*Psychrobacter aquatica*
Firmicutes	*Bacillales*	CZD 16-W	510	100	AJ316309	*Bacillus sp*
Actinobacteria	*Actinobacteridae*	CZD 16-Z	480	99	AY176765	*Arthrobacter sp.*

Table 3.4: Culturable diversity recovered from sample UWV4 on ¼ strength nutrient broth agar.

Phylum	Class or subclass	Isolate name	Sequence length	% identity	Accession no. of closest relative	Closest relative
Proteobacteria	*Gamma*	NB UW GW	507	99	AF513431	*Pseudomonas sp.*
Actinobacteria	*Actinobacteridae*	NB UW RC1	481	100	X77438	*Rathayibacter tritici*
Proteobacteria	*Gamma*	NB UW4-3	488	100	AY167273	*Acinetobacter sp.*
Bacteroidetes	*Sphingobacteria*	NB 4 Gm LR	500	99	AJ440980	*Hymenobacter*

Table 3.5: Culturable diversity recovered from sample UWV4 on starch agar.

Phylum	Class or subclass	Isolate name	Sequence length	% identity	Accession no. of closest relative	Closest relative
Proteobacteria	*Gamma*	ST Sm UW4-SW	470	99	AF513431	*Pseudomonas sp.*

Table 3.6: Summary of viable count data obtained following cultivation of microorganisms from samples MV16, MV17 and UWV4 on each type of growth medium used (1/4 NB: quarter-strength nutrient broth; CZD: Czapek-Dox; Gm: gentamicin; Sm: streptomycin). Values for viable counts are given as cells.g^{-1} wet.wt.

	Sample		
	MV16	**MV17**	**UWV4**
Growth Medium			
¼ NB	6×10^2-3×10^4	8×10^3-2×10^4	$6.7 \times 10^3$$6.7 \times 10^3$
¼ NB + Gm	$<2 \times 10^2$	$>6 \times 10^3$ 10^3	$<2 \times 10^2$
¼ NB + Sm	1.6×10^2	$1 \times 10^3$$1 \times 10^3$	single colony observed but suspected contaminant
Starch	2.6×10^3	$2.6 \times 10^3$$1.5 \times 10^4$	no growth
Starch + Gm	2.6×10^3	$2.0 \times 10^3$$>3 \times 10^4$	no growth
Starch + Sm	ND[a]	$>3 \times 10^4$	$<2 \times 10^2$
CZD	4×10^4	3×10^3	no growth
CZD + Gm	2-3×10^3	plate discarded due to ice damage	ND[a]
CZD + Sm	3×10^3	plate discarded due to suspected growth of contaminants	no growth

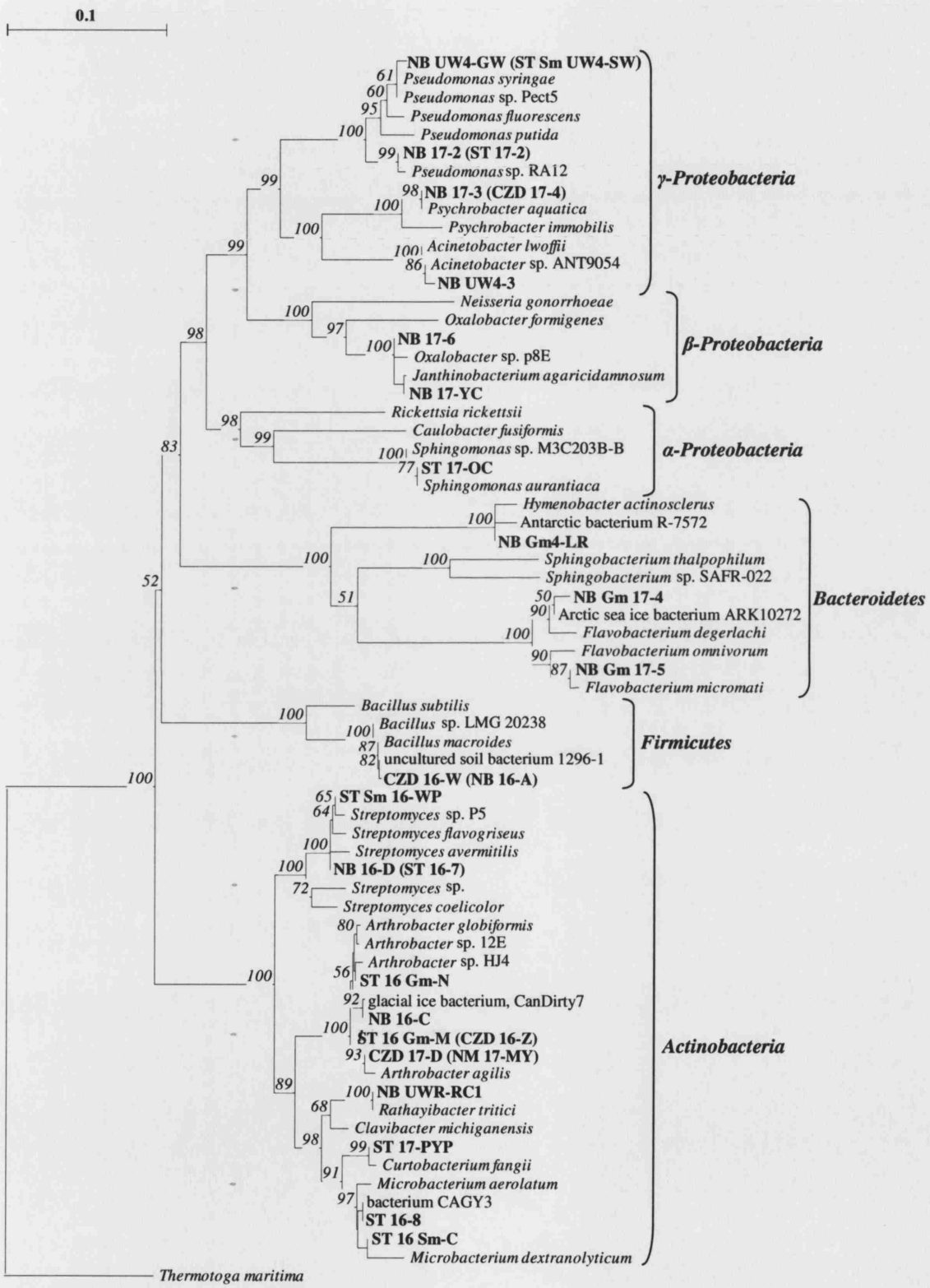
[a] ND – Not determined

3.2.5: Phylogenetic analysis of cultured isolates

The 16S rRNA sequences of 24 unique Antarctic isolates of >300 bp in length were aligned with their closest database relatives and related type strains, along with additional members of the phyla represented in the cultivation study. A total of 67 sequences were included in the alignment spanning nucleotide positions 435 to 827 of the 16S rRNA gene. The phylogenetic positioning of isolates is shown in Figure 3.2. A total of four bacterial phyla were represented: the *Actinobacteria* (eleven isolates), *Firmicutes* (two isolates), *Bacteroidetes* (four isolates), and three divisions (alpha, beta and gamma) of the *Proteobacteria* (seven isolates).

Figure 3.2: Phylogenetic affiliation of Antarctic isolates within the domain *Bacteria*. Antarctic isolates are represented in bold type. Isolates shown in parentheses represent identical isolates cultivated from the same sample but on different growth media; where two identical isolates were cultivated, the isolate for which a greater level of 16S rDNA sequence information was available was included in the phylogenetic analysis. Bootstrap values, given as a percentage of 100, are indicated for branches supported by more than 50% of trees. Scale bar: 0.1, estimated number of substitutions per nucleotide position.

Additional sequences used in the phylogenetic analysis were:
Pseudomonas syringae (AJ576247), *Pseudomonas fluorescens* (AF094726), *Pseudomonas putida* (Z76667), *Psychrobacter immobilis* (U39399), *Acinetobacter lwoffii* (Z93441), *Neisseria gonorrhoeae* (X07714), *Oxalobacter formigenes* (U49757), *Rickettsia rickettsii* (L36217), *Caulobacter fusiformis* (AB008533), *Sphingomonas aurantiaca* (AJ429238), *Hymenobacter actinosclerus* (Y17356), *Sphingobacterium thalpophilum* (D14020), *Flavobacterium dergerlachi* (AJ557886), *Flavobacterium omnivorum* (AF433174), *Bacillus subtilis* (AY162133), *Bacillus macroides* (AF157696), *Streptomyces flavogriseus* (AJ494864), *Streptomyces coelicolor* (Z76678), *Clavibacter michiganensis* (U96181), *Microbacterium dextranolyticum* (D21341).


0.1
NB UW4-GW (ST Sm UW4-SW)
Pseudomonas syringae
Pseudomonas sp. Pect5
Pseudomonas fluorescens
Pseudomonas putida
NB 17-2 (ST 17-2)
Pseudomonas sp. RA12
NB 17-3 (CZD 17-4)
Psychrobacter aquatica
Psychrobacter immobilis
Acinetobacter lwoffii
Acinetobacter sp. ANT9054
NB UW4-3
γ-Proteobacteria
Neisseria gonorrhoeae
Oxalobacter formigenes
NB 17-6
Oxalobacter sp. p8E
Janthinobacterium agaricidamnosum
NB 17-YC
β-Proteobacteria
Rickettsia rickettsii
Caulobacter fusiformis
Sphingomonas sp. M3C203B-B
ST 17-OC
Sphingomonas aurantiaca
α-Proteobacteria
Hymenobacter actinosclerus
Antarctic bacterium R-7572
NB Gm4-LR
Sphingobacterium thalpophilum
Sphingobacterium sp. SAFR-022
NB Gm 17-4
Arctic sea ice bacterium ARK10272
Flavobacterium degerlachi
Flavobacterium omnivorum
NB Gm 17-5
Flavobacterium micromati
Bacteroidetes
Bacillus subtilis
Bacillus sp. LMG 20238
Bacillus macroides
uncultured soil bacterium 1296-1
CZD 16-W (NB 16-A)
Firmicutes
ST Sm 16-WP
Streptomyces sp. P5
Streptomyces flavogriseus
Streptomyces avermitilis
NB 16-D (ST 16-7)
Streptomyces sp.
Streptomyces coelicolor
Arthrobacter globiformis
Arthrobacter sp. 12E
Arthrobacter sp. HJ4
ST 16 Gm-N
glacial ice bacterium, CanDirty7
NB 16-C
ST 16 Gm-M (CZD 16-Z)
CZD 17-D (NM 17-MY)
Arthrobacter agilis
NB UWR-RC1
Rathayibacter tritici
Clavibacter michiganensis
ST 17-PYP
Curtobacterium fangii
Microbacterium aerolatum
bacterium CAGY3
ST 16-8
ST 16 Sm-C
Microbacterium dextranolyticum
Actinobacteria
Thermotoga maritima

Isolates belonging to the phylum *Actinobacteria*

Ten isolates had phylogenetic affiliation with the phylum *Actinobacteria* and all belonged to the well-characterized subclass *Actinobacteridae*. Four isolates belonged to each of the families *Microbacteriaceae* and *Micrococcaceae* and two isolates to the family *Streptomycetaceae*; these families represent groups of commonly isolated soil bacteria.

Family *Micrococcaceae*: Isolates belonging to the genus *Arthrobacter* were cultivated on several occasions from each of the Miers Valley samples and were found to grow on all types of culture media used, including starch supplemented with gentamicin. Four different isolates were identified and were found to be most closely related to *Arthrobacter agilis* along with species of *Arthrobacter* isolated from grassland soil in China (Chen *et al.*, Unpublished), glacial ice from Canada glacier (McMurdo Dry Valleys, Antarctica) (Christner *et al.*, 2003a) and metal-contaminated lake sediment (Konstantinidis *et al.*, 2003). The level of nucleotide sequence identity amongst these isolates did not fall below 98%.

Arthrobacter species have been previously isolated from the McMurdo Dry Valleys including from soil and rock samples (Coker *et al.*, 2003) and a microbial mat from Lake Fryxell (Brambilla *et al.*, 2001; Trappen *et al.*, 2002; Reddy, 2002). In addition, they have also been cultivated from the following Antarctic environments: the sea bottom along the Antarctic ice shelf (Lonhienne *et al.*, 2001), brine samples from sea ice (Junge *et al.*, 1998) and Antarctic Krill (Turkiewicz *et al.*, 1982).

Family *Microbacteriaceae*: Three members of this family were isolated from the Miers Valleys samples and all were cultivated on starch media. The closest relative of isolate ST 17-PYP was identified as *Curtobacterium fangii*, a pathogen of wheat (Guo *et al.*, Unpublished). Isolates ST 16 Sm-C and ST 16-8 were phylogenetically affiliated to members of the genus *Microbacterium*; their closest relatives were identified as *Microbacterium aerolatum*, isolated from the air in Vienna (Zlamala *et al.*, 2002), and a bacterium isolated from a municipal wastewater treatment bioreactor (LaPara *et al.*,

Unpublished), respectively. Isolate NB UW RC-1 represents the only isolate belonging to the phylum *Actinobacteria* that was cultivated from sample UWV4 and is most closely related to *Rathayibacter tritici*.

Isolates belonging to this family have previously been cultivated from Antarctic environments and include *Cryobacterium psychrophilum* (previously named *Curtobacterium psychrophilum*) isolated from Antarctic soil to which ST 17-PYP exhibits 96% identity (Suzuki *et al.*, 1997). A bacterium designated CS117, exhibiting 99% identity to *C. psychrophilum*, has additionally been recovered from an Antarctic cryoconite hole (Christner *et al.*, 2003a) with which ST 17-PYP also shares 96% 16S rRNA sequence identity. Two further isolates affiliated to the family *Microbacteriaceae* and related to *Clavibacter michiganensis* and *Microbacterium keratinolyticum*, have also been cultivated from microbial mat samples surrounding Antarctic lakes (Van Trappen *et al.*, 2002).

Family *Streptomycetaceae*

Two isolates belonged to the family *Streptomycetaceae*, both of which were cultivated from sample MV16 on starch agar. ST Sm 16-WP shared 99% identity to a *Streptomyces* species described in a study of heavy metal resistance amongst members of this genus (unpublished). Interestingly, the second isolate NB 16-D shared only 93% rRNA identity to it closest relative, primarily due to a region of 33 bp of significant sequence variation. FASTA analysis of this region of sequence corresponding to positions 456 to 489 of the 16S rRNA gene, exhibited 93% identity in a 33 bp nucleotide overlap to the 16S rRNA gene of the psychrophilic Antarctic soil bacterium, *Cryobacterium psychrophilum* (see previous section).

Isolates belonging to the phylum *Proteobacteria*

Seven isolates were phylogenetically affiliated to the phylum *Proteobacteria*; four belonged to the class *Gammaproteobacteria*, two to the class *Betaproteobacteria* and a single isolate to the class *Alphaproteobacteria*.

***Gammaproteobacteria*:** Two *Pseudomonas* species were isolated, one each from samples MV17 and UWV4, and both of which could be cultivated on ¼ NB and starch agar. Cultivated relatives of isolate NB 17-2 included a *Pseudomonas* species isolated from a water sample of uranium mining waste (Tzvetkova & Selenska-Pobell, unpublished) and *Pseudomonas fluorescens* isolated from river water contaminated with phenolic compounds (Heinaru *et al.*, 2000). The closest relative of isolate NB UW-GW was identified as a *Pseudomonas* species cultivated from arctic sea ice (Groudieva & Antranikian, unpublished); NB UW-GW was also found to share 99% sequence identity to *Pseudomonas syringae* isolated from soil in the vicinity of Lake Zub, of the Schirmacher oasis, Antarctica (Reddy *et al.*, Unpublished). *Pseudomonas* species have been isolated from a variety of Antarctic environments including: fuel-contaminated soil at Scott base (Farrell *et al.*, 2003), microbial mat samples from Lake Ace and Lake Grace in the Vestfold hills and Lake Hoare in the McMurdo Dry Valleys (Van Trappen *et al.*, 2002), sea water of the Ross sea (Bruni *et al.*, 1999), and Antarctic sea ice (Bowman *et al.*, 1997).

Isolate NB 17-3 was affiliated to the genus *Psychrobacter* and shared 100% 16S rRNA sequence identity over the region of the gene sequenced to a bacterium, *Psychrobacter aquatica*, isolated from the McMurdo Dry Valley region (Reddy *et al.*, Unpublished (b)); an isolate with an identical 16S rRNA sequence was also cultivated on CZD agar. The closest relative of isolate NB UW4-3 was identified as a species of *Acinetobacter* present in Antarctic sea ice from the Weddel Sea (Brinkmeyer *et al.*, 2003).

***Alphaproteobacteria*:** BLAST analysis revealed isolate ST 17-0C to belong to the genus *Sphingomonas*, exhibiting 100% 16S rRNA sequence identity to species of *Sphingomonas* previously cultivated from two different Antarctic environments: an endolithic community within gypsum crust (Hughes & Lawley, 2003) and accretion ice of Lake Vostok (Christner *et al.*, 2001). The closest related type strain of isolate ST 17-OC was identified as *Sphingomonas aurantiaca* sharing 100% nucleotide identity in the region of the gene analysed. *Sphingomonas* species have additionally been identified as cultivated isolates in both a microbial mat from Lake Ace in the Vestfold hills (Van

Trappen *et al.*, 2002) and fuel-contaminated soil collected near Scott base (Farrell *et al.*, 2003), and as clone sequences present in 16S rDNA libraries surveying bacterial diversity in lake sediments of Ardley peninsula and the Schirmacher Oasis (Reddy *et al.*, Unpublished (a)).

***Betaproteobacteria*:** Two isolates belonging to the class *Oxalobacteraceae* were isolated from sample MV17. The closest relative of isolate ST 17-YC was identified as *Janthinobacterium agaricidamnosum*, an isolate recovered from a spacecraft assembly facility (Venkateswaran *et al.*, 2001). Isolate NB 17-6 was most closely related to a species of *Oxalobacter*, isolated from a packed-column bioreactor in a study of protease-producing bacteria present in Antarctic soil (Wery *et al.*, 2003); the bioreactor was inoculated using soil collected from Cape Evans on Ross Island.

Bacterial isolates belonging to the class *Oxalobacteraceae* have previously been cultivated from Antarctic samples. Isolate CD89 cultivated from an Antarctic cryoconite hole (with which ST 17-YC shares 99% identity) exhibits 97% identity to *Janthinobacterium lividum* (Christner *et al.*, 2003a). Two isolates, R-7687 and R-7614, cultivated from a microbial mat sample from Lake Hoare in the McMurdo Dry Valleys (Van Trappen *et al.*, 2002) also possess *Janthinobacterium lividum* as their closest relative sharing 99% 16S rRNA similarity; ST 17-YC exhibits 97% and 96% identity to each of these isolates respectively.

Isolates belonging to the phylum *Firmicutes*

A single isolate recovered from sample MV16 was found to be phylogenetically affiliated with the phylum *Firmicutes*. BLAST analysis revealed isolate CZD 16-W to belong to the genus *Bacillus*; an isolate with an identical 16S rRNA sequence (NB 16-A) was also cultivated from the same sample on ¼ NB agar. The Sequence Match facility of the RDP identified *Bacillus halmapalus*, an alkaliphilic bacillus, as the closest related type strain. The partial 16S rDNA sequence of CZD 16-W was also found to be identical to two bacterial isolates cultivated from comparatively low-temperature environments,

glacial ice in Bolivia and China (Christner 2002. PhD Thesis) and to several strains of *Bacillus macroides*.

Bacillus species are ubiquitous to soils across the globe and have been previously isolated from Antarctic environments including: Dry Valley air and soil samples (Cameron & Morelli, 1972), Candlemas island of the South Sandwich archipelago (Logan *et al.*, 2000), and the Ross Sea (Bruni *et al.*, 1999).

Isolates belonging to the phylum *Bacteroidetes*

Four isolates were phylogenetically affiliated to the phylum *Bacteroidetes*. Two isolates belonging to the family *Flavobacteriaceae* were cultivated from sample MV17 on ¼ NB agar containing gentamicin. The closest relatives of isolates NB Gm 17-4 and NB Gm 17-5 were identified as a *Flavobacterium* species isolated from Arctic sea ice (Brinkmeyer *et al.*, 2003), and *Flavobacterium micromati*, a psychrophilic bacterium isolated from microbial mat of an Antarctic lake (Van Trappen *et al.*, 2002), respectively. The level of shared sequence identity between the two isolates was 95%. *Flavobacterium* species are common to Antarctica having been cultivated from numerous environments including microbial mat samples from Antarctic lakes, a cryoconite hole, and sea ice (Christner *et al.*, 2003a).

Isolate NB Gm 4-LR shared 99% 16S rRNA identity in the region of the gene sequenced to a species of *Hymenobacter* isolated from a spacecraft assembly facility (La Duc *et al.*, 2003). A member of this genus, *Hymenobacter roseosalivarius,* with which NB Gm 4-LR shared 94% 16S rRNA identity, has previously been isolated from Antarctic Dry Valley soil and sandstone following inoculation of oligotrophic medium and incubation under low light intensities (Hirsch *et al.*, 1998). The closest relative of isolate NB 16-EY, a species of *Sphingobacterium*, was also identified as a bacterium isolated from a spacecraft assembly facility. In addition, NB 16-EY exhibited 97% 16S rRNA identity to *Pedobacter cryoconitis*, a bacterium recovered from a comparable low-temperature environment, alpine glacier ice (Margesin *et al.*, 2003).

3.2.6: Summary of culturable diversity

A greater diversity of bacterial species was cultivated from Miers Valley soil samples compared to soil sampled from the Upper Wright Valley. Twelve isolates cultivated on different media from samples MV16 and MV17 each were identified on the basis of their 16S rRNA gene sequence similarity to database entries. These represented eight and ten unique isolates cultured from these sites respectively. In contrast, only five bacterial species (one of which could not be subcultured) were identified following cultivation from soil sample UWV4. Moreover, the rate of bacterial growth, as assessed by the time taken to form distinct colonies, from soil sample UWV4 as compared to Miers Valley samples was slower on both ¼ NB agar and starch agar, with no growth observed on CZD agar. On the basis of these data, this tentatively infers a reduced diversity among the culturable fraction of heterotrophic bacteria in soil from the Upper Wright Valley. The inability to subculture some of the colonies may reflect those microorganisms that can only undergo a limited number of divisions on a Petri dish, afterwhich conditions are no longer conducive to their growth. Kaeberlein *et al.*, (2002) observed that for 27-30 colonies cultivated from intertidal marine sediment on solid media and selected for subculture on three independent occasions, only 14% of transfers resulted in growth.

Of the three growth media employed, a greater number of bacterial species were cultivated on ¼ NB agar and starch agar. Fourteen isolates from within four bacterial phyla were cultivated on ¼ NB agar and eleven isolates from within two phyla were cultivated on starch agar. Only four isolates within three bacterial phyla were identified following cultivation on CZD agar. In the latter case however, diversity may have been underestimated due to the failure to amplify a 16S rRNA product from four isolates cultured from soil sample MV16. Distinct bacterial colonies on samples UWV4 and MV17 were slowest to form on starch agar, whilst no discernable difference was observed in the growth rate of bacteria from sample MV16 on each of the three media types.

On the basis of 16S rRNA gene sequence similarity to all 16S rRNA gene database entries, the majority of Antarctic isolates possessed, as their closest match, a bacterial strain that had been cultured as opposed to sequences obtained through culture-independent means. The level of sequence similarity estimated throughout by using BLAST and FASTA analysis was typically high at ~98-100% over an average length of 457 bp. (Three isolates were tentatively identified from 16S rRNA sequences of <300 bp but were excluded from the subsequent phylogenetic analysis.) A consensus has been adopted from which organisms sharing less than 97% 16S rRNA identity are regarded as different species (Stackebrandt & Goebel, 1994). However, where the level of rRNA identity is greater than 97%, the resolution of the 16S rRNA gene is such that species level identification can not be made on this basis alone; a number of organisms are known to possess near identical 16S rRNA sequences but clearly constitute different species. For this reason, in this work, identification of Antarctic isolates is inferred at the level of the genus.

Isolate NB 16-D recovered from sample MV16 proved an exception in this study, sharing only 93% 16S rRNA identity in a 517 bp overlap to a *Streptomyces* species (accession number X81574). It was found that much of this difference could be explained by the presence of a region of 33bp within sequence NB 16-D, corresponding to position 456 to 489 (*E. coli* numbering) of the 16S rRNA gene, that was not shared by its database relative. Therefore, this sequence may represent a new genus within the family *Streptomycetaceae*, although this is speculative given a complete 16S rRNA gene sequence is not available.

The majority of isolates cultivated in this study belonged to bacterial genera that had previously been identified from culture-based studies in a range of Antarctic environments. Specifically, this included members of the genera *Arthrobacter*, *Flavobacterium*, *Pseudomonas*, and *Sphingomonas*. Sample MV17 was the source of greatest culturable diversity with eight bacterial genera identified compared to five and four bacterial genera for samples MV16 and UWV4 respectively. However, this was as expected given the soil sample was collected beneath a dense algal mat surrounding a

small lake in the Miers Valley and was thus, considered more conducive to bacterial growth. This particular environment is afforded some shelter by virtue of the mat, from extreme fluctuations in temperature, exposure to UV radiation and desiccating winds; additionally, it is comparatively nutrient rich and local melt water is available during the austral summer (Alger *et al.*, 1995; Brambilla *et al.*, 2001).

However, of particular note were the four actinobacterial isolates belonging to the family *Microbacteriaceae.* Cultivated relatives of these isolates (sharing >99% 16S rRNA identity) included *Rathayibacter tritici*, *Curtobacterium fangii* and species of *Microbacterium.* Furthermore, isolates affiliated to this family and most closely related to *Clavibacter michiganensis* and *Microbacterium keratinolyticum* were also recently cultivated from microbial mat samples from the Antarctic Dry Valleys (Van Trappen *et al.*, 2002). Members of the above genera appear to have a strong association with plants, either as phytopathogens or as part of the plant endophytic community. In a study conducted by Zinniel and colleagues (2002) of 853 endophytic bacteria isolated from agronomic crops and prairie plants, six endophytes selected for their ability to colonize and persist were identified as members of the genera *Cellulomonas*, *Clavibacter*, *Curtobacterium*, and *Microbacterium. R. tritic* and *C. michiganensis* are well recognized as phytopathogens of terrestrial plants and *C. fangii*, identified as the closest relative of isolate ST 17-PYP, is a pathogen of wheat (Guo *et al.*, Unpublished; Hahn *et al.*, 2003). The presence of *Microbacteriaceae* in Dry Valley soils is intriguing given the absence of macroscopic plant life in this environment, and especially since an isolate was cultivated from each of the samples investigated. Taking into account this last point along with the independent cultivation of members of this family from Antarctic Dry Valley microbial mat samples (Van Trappen *et al.*, 2002), these organisms are unlikely to represent transient contaminants. Instead, these results may indicate members of the *Microbacteriaceae* are distributed in environments that extend beyond the phytosphere.

Finally, the four closest cultivated relatives isolated in independent studies were identified for each of the Antarctic isolates described in this work. The objective here was to examine the environmental distribution of these organisms, in order to identify

both signs of endemism that may be occurring in the Antarctic environment, together with common environmental parameters that may influence the selection of these organisms. This analysis will be discussed later in Chapter 6.

3.3: Molecular Assessment of Bacterial Diversity

3.3.1: Construction and analysis of 16S rRNA gene clone libraries

The cold desert soils of the Dry Valleys in Eastern Antarctica are widely considered the harshest arid environments on Earth (Cowan *et al.*, 2002). Microbial life within these soils is primarily limited by the extreme aridity, a consequence of low precipitation and atmospheric humidity. Soil water activity is further reduced by the accumulation of salts that is typical of soils in which evaporation exceeds leaching (Campbell & Claridge, 1987). Two desiccated surface soil samples collected from different locations within the Dry Valleys were selected for the molecular assessment of bacterial diversity. Sample MV11.1 was collected from the valley floor of the Miers Valley by D. Cowan in February 1999, whilst sample UWV4 was collected from within the labyrinth structure of the Upper Wright Valley in 2001 - white crystals of salt were present on the surface of this soil at the time of sampling. Determination of the water contents of samples MV11.1 and UWV4 revealed them to be 1.008 % and 1.20 % w/w respectively, in line with typical values of 0.2-5 % w/w for Dry Valley soils (Cameron & Conrow, 1969; Horowitz *et al.*, 1972).

A 16S rRNA gene clone library was constructed for each sample from the pooled products of three replicate PCR reactions following 25 cycles of amplification with primers 27F and Un1492R. Preliminary ARDRA analysis of the MV11.1 library revealed the majority of clones to possess unique restriction patterns, thus, 16S rRNA clones were selected at random for sequencing. Partial sequences were obtained for a total of 19 clones from the MV11.1 library representing 18 bacterial phylotypes (≥97.5% 16S rRNA identity). Partial sequence analysis of 24 clones from the UWV4 library identified 13 unique phylotypes, two of which were represented by 2 and eleven clones. Closest relatives of bacterial 16S rRNA sequences were identified by BLAST and FASTA analyses and are shown in Tables 3.7 for the MV11.1 library and Table 3.8 for the UWV4 library.

Table 3.7: Closest relatives identified for MV11.1 bacterial 16S rRNA gene sequences.

Affiliation	Clone	Closest relative	Accession no.	% Identity	Sequence length	No. of clones	Comment
Betaproteobacteria	MV11-1	uncultured bacterium clone KD1-79	AY218566	97	951	1	penguin droppings sediments from Ardley Island, Antarctica
Actinobacteria	MV11-2	uncultured bacterium clone C-F-1	AF443581	96	1241	1	semi-arid soil
Planctomycetales	MV11-3	Gram-positive bacterium SOGA31	AJ244807	94	1074	1	bacterial symbiont of the cuttlefish *Sepia officinalis*
Alphaproteobacteria	MV11-6	*Rickettsia bellii*	U11014	94	1253	1	
Actinobacteria	MV11-7	*Rubrobacter radiotolerans*	U65647	90	973	1	
Alphaproteobacteria	MV11-8	uncultured bacterium clone a13104	AY102311	98	1215	1	heavy-metal contaminated soil
Actinobacteria	MV11-10	uncultured soil bacterium clone 359	AY493980	95	1291	1	soil aggregates
Actinobacteria	MV11-11	uncultured soil bacterium clone 359	AY493980	93	1291	1	soil aggregates

TM7	MV11-13	uncultured bacterium NoosaAW57	AF269013	95	908	1	sewage treatment plant
Betaproteobacteria	MV11-14	isolate CanDirty89	AF479326	99	1287	1	cryoconite hole, Canada Glacier, Antarctica
TM7	MV11-15	uncultured soil bacterium clone S1197	AF507689	94	1080	1	pinyon-juniper forest soil (Arizona)
Alphaproteobacteria	MV11-16	uncultured bacterium clone JG37-AG-18	AJ518765	98	899	1	uranium mining waste pile
Actinobacteria	MV11-19	uncultured bacterium clone ML316M-15	AF454303	95	1070	2	alkaline, hypersaline Mono Lake, California
Deltaproteobacteria	MV11-23	uncultured bacterium #0319-6G20	AF234131	93	872	1	Australian arid soil
Actinobacteria	MV11-25	uncultured bacterium clone ARKCH2Br2-66	AF468240	96	914	1	dibromomethane enrichment of Arctic sea ice
Acidobacteria	MV11-27	uncultivated soil bacterium clone S023	AF013550	98	1284	1	arid soil, south-western United States
Acidobacteria	MV11-31	clone WD254	AJ292582	93	1254	1	polychlorinated biphenyl-polluted soil
Verrucomicrobia	MV11-32	uncultured bacterium clone KD6-84	AY218772	94	1293	1	Penguin Droppings Sediments, Ardley Island (Antarctica)

Table 3.8: Closest relatives identified for UWV4 bacterial 16S rRNA gene sequences.

Affiliation	Clone	Closest relative	Accession no.	% Identity	Sequence length	No. of clones	Comment
Acidobacteria	UWV4-1	uncultured bacterium clone FBP241	AY250867	98-99	991	11	cryptoendolithic community, McMurdo Dry Valleys
Actinobacteria	UWV4-3	uncultured bacterium clone ARKCH2Br2-66	AF468240	95	946	1	Arctic pack ice
Actinobacteria	UWV4-7	uncultured *Rubrobacteridae* clone WY02C	AY150916	96	430	1	short grass steppe surface soil, Wyoming
Actinobacteria	UWV4-8	*Rubrobacter radiotolerans*	U65647	95	993	1	
Chloroflexi	UWV4-13	clone W4-UWV45	AY345496	91	985	1	lake sediment, Hawaiian Archipelago
Alphaproteobacteria	UWV4-17	isolate RSI-21	AJ252588	97	967	1	rhizosphere soil bacterium
CFB	UWV4-18	uncultured bacterium clone D132	AY274138	93	1006	1	heavy metal contaminated mine tailings
Deinococcus-Thermus	UWV4-20	uncultured *Deinococci* bacterium FBP266	AY250871	100	935	1	cryptoendolithic community, McMurdo Dry Valleys

TM7	UWV4-22	clone C129	AF507687	92%	1391	2	pinyon-juniper forest soil
CFB	UWV4-27	clone SC-I-12	AJ252615	96	960	1	agricultural soil
TM7	UWV4-33	clone SM1G12	AF445701	91	919	1	travertine depositional facies, Mammoth Hot Springs, Yellowstone Park
Actinobacteria	UWV4-34	clone FBP460	AY250884	96	958	1	cryptoendolithic community, McMurdo Dry Valleys
Alphaproteobacteria	UWV4-36	clone SC-I-56	AJ252642	99	933	1	agricultural soil

3.3.2: Phylogenetic analysis of bacterial 16S rRNA sequences

A total number of 29 bacterial 16S rRNA gene sequences ranging from 867 to 1293 bp in length were included in the phylogenetic analysis, representing 12 and 17 clone sequences from the Upper Wright and Miers Valley libraries respectively. Seven of these sequences were found to exhibit ≥98% sequence identity with their closest database relative; this level of rRNA sequence can be used to infer identification at the genus level. However, only one of these sequences exhibited 100% 16S rDNA identity (to an uncultivated bacterium) and all but one were most closely related to other environmental clone sequences. In total, only four sequences possessed a cultivated bacterium as their closest relative. Of the remaining sequences, ten were found to share ≥ 95% identity whilst twelve sequences were < 95% identical to their closest relative. Clones MV11-7 and UWV4-33 showed greatest divergence possessing only 90% and 91% 16S rDNA identity respectively to their database relatives.

Figure 3.3 illustrates the phylogenetic affiliation of Antarctic clone sequences. Clone sequences were aligned with their closest database relatives, related type strains and representatives of major phylogenetic groups; a total of 90 sequences were included in the alignment spanning nucleotide position 399 to 1001 of the 16S rRNA gene (*E. coli* numbering). Overall, nine different bacterial phyla including three subdivisions of the *Proteobacteria* were represented by Antarctic clones. The *Actinobacteria* and the *Proteobacteria* were each represented by eight sequences, followed by division TM7 (4 sequences) and the *Acidobacteria* (3 sequences). Only a single bacterial phylotype from the Miers Valley library was represented by more than one clone. Despite possessing the same database closest relative, clones MV11-10 and MV11-11 share only 93% 16S rRNA identity and thus, are both included in the phylogenetic analysis. In contrast, two phylotypes from the Upper Wright library were represented by two and eleven clone sequences, suggesting bacterial diversity may be lower in this library compared to the Miers Valley library. From this observation, it could be suggested that diversity may be lower in the Upper Wright library compared to the Miers Valley library. However, this idea is impossible to assert given that no statistical analyses were used to provide an objective comparison of diversity between the two libraries. Furthermore, both samples

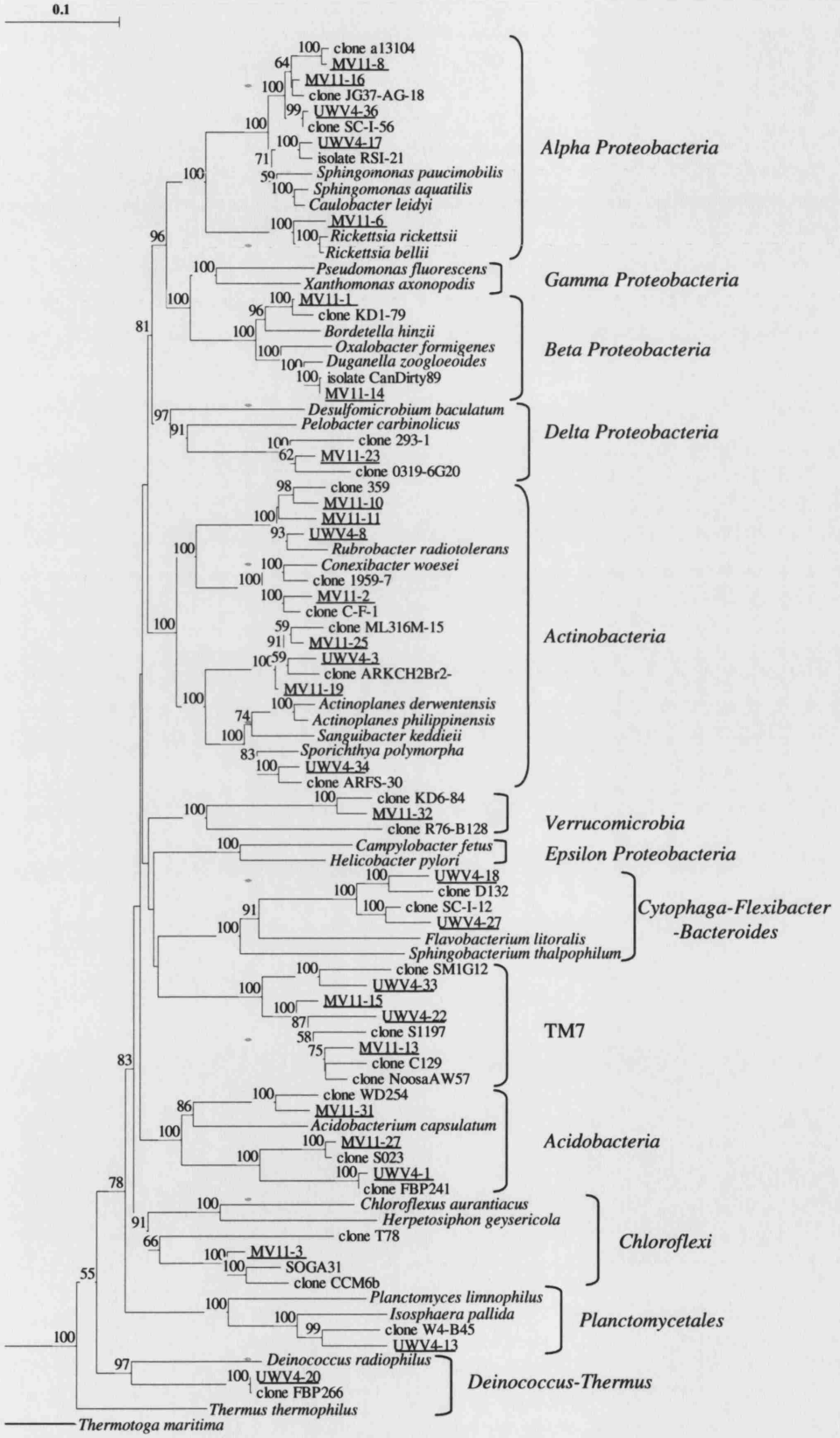

0.1
clone a13104
MV11-8
MV11-16
clone JG37-AG-18
UWV4-36
clone SC-I-56
UWV4-17
isolate RSI-21
Sphingomonas paucimobilis
Sphingomonas aquatilis
Caulobacter leidyi
MV11-6
Rickettsia rickettsii
Rickettsia bellii
Alpha Proteobacteria
Pseudomonas fluorescens
Xanthomonas axonopodis
Gamma Proteobacteria
MV11-1
clone KD1-79
Bordetella hinzii
Oxalobacter formigenes
Duganella zoogloeoides
isolate CanDirty89
MV11-14
Beta Proteobacteria
Desulfomicrobium baculatum
Pelobacter carbinolicus
clone 293-1
MV11-23
clone 0319-6G20
Delta Proteobacteria
clone 359
MV11-10
MV11-11
UWV4-8
Rubrobacter radiotolerans
Conexibacter woesei
clone 1959-7
MV11-2
clone C-F-1
clone ML316M-15
MV11-25
UWV4-3
clone ARKCH2Br2-
MV11-19
Actinoplanes derwentensis
Actinoplanes philippinensis
Sanguibacter keddieii
Sporichthya polymorpha
UWV4-34
clone ARFS-30
Actinobacteria
clone KD6-84
MV11-32
clone R76-B128
Verrucomicrobia
Campylobacter fetus
Helicobacter pylori
Epsilon Proteobacteria
UWV4-18
clone D132
clone SC-I-12
UWV4-27
Flavobacterium litoralis
Sphingobacterium thalpophilum
Cytophaga-Flexibacter
-Bacteroides
clone SM1G12
UWV4-33
MV11-15
UWV4-22
clone S1197
MV11-13
clone C129
clone NoosaAW57
TM7
clone WD254
MV11-31
Acidobacterium capsulatum
MV11-27
clone S023
UWV4-1
clone FBP241
Acidobacteria
Chloroflexus aurantiacus
Herpetosiphon geysericola
clone T78
MV11-3
SOGA31
clone CCM6b
Chloroflexi
Planctomyces limnophilus
Isosphaera pallida
clone W4-B45
UWV4-13
Planctomycetales
Deinococcus radiophilus
UWV4-20
clone FBP266
Thermus thermophilus
Deinococcus-Thermus
Thermotoga maritima

Figure 3.3: Phylogenetic affiliation of Antarctic 16S rDNA clone sequences identified from the Upper Wright (UWV) and Miers Valley (MV) libraries to members of the Domain *Bacteria*. Antarctic clone sequences are shown underlined. A total of eleven different phyla are represented by Antarctic sequences. Bootstrap values, given as a percentage of 100, are indicated for branches supported by more than 50% of trees. Scale bar: 0.1, estimated number of substitutions per nucleotide position.

were subject to different storage conditions during collection and transport, and different methods were used to extract community DNA from the samples.

The phylogenetic analysis revealed eight clones to be affiliated to the division *Actinobacteria*. Four clones belonged to the family *Rubrobacteraceae* and three possessed a clone sequence recovered from a soil environment as their closest database relative. Clone UWV4-8 exhibited 95% 16S rRNA identity to *Rubrobacter radiotolerans*, a slightly thermophilic species noted for its extreme radiation-resistance; clone MV11-7 shared 90% identity to the same relative but was not included in the phylogenetic tree. *R. radiotolerans* strain JCM 2153-T was also identified as the closest type strain for clones MV11-10 and MV11-11. The closest database relative of each of these sequences was an uncultured soil bacterium identified in soil aggregates (Kim & Crowley (a), unpublished), to which MV11-10 and MV11-11 shared 95% and 93% identity respectively. Clone MV11-2 exhibited 96% identity to an uncultured bacterium present in a semi-arid soil (Rutz & Kieft, 2004).

Figure 3.4 illustrates the phlyogenetic relationship of Antarctic clone sequences to members of the *Rubrobacteridae*. This subclass represents one of the less well characterized groups among the *Actinobacteria*, comprising sequences moderately related to the described strains *Rubrobacter radiotolerans* and *Rubrobacter xylanophilus*, which form a monophyletic group that branches deeply from the *Actinobacteria* line of descent (Holmes *et al.*, 2000). A phylogenetic reconstruction of *Rubrobacteria* soil clones conducted by Holmes *et al.*, (2000) identified three distinct groups, of which groups 1 and 2 are represented in the phylogenetic tree shown in Figure 3.4. As is evident from the tree, clones MV11-7, MV11-10, MV11-11 and UWV4-8 fall within Group 1, whilst clone MV11-2 clusters with group 2 sequences;

both groupings are supported by bootstrap values of 100. Interestingly, in the study conducted by Holmes *et al.*, (2000), group 1 Rubrobacteria were identified as the dominant group in 16S rRNA gene libraries constructed from Australian desert soils. The relative abundance of these sequences within the libraries as assessed by hybridization with an rDNA group-specific probe ranged from 2.5 to 10.2%. Closest relatives of Antarctic group 1 Rubrobacteria sequences included the desert soil clones along with sequences identified in soil aggregates (Kim & Crowley (a), Unpublished), associated with masonry and lime wall paintings (Schabereiter-Gurtner *et al.*, 2001) and identified in a lichen-dominated cryptoendolithic community of the McMurdo Dry Valleys, Antarctica (de la Torre *et al.*, 2003). Furthermore, a strain isolate of *Rubrobacter radiotolerans* isolated from thermally polluted effluent was also among the closest relatives (Carreto *et al.*, 1996). Clone MV11-2 of group 2 Rubrobacteria was most closely related to soil clones recovered from an agricultural soil in California (Valinsky *et al.*, 2002) and a thermal soil (Botero *et al.*, Unpublished), along with bacterial isolates cultivated from Australian pasture soil (Sait *et al.*, 2002).

The remaining four clones within the *Actinobacteria* formed a separate group affiliated to the order *Actinomycetales*. Clone UWV4-34 shared 96% identity to clone FBP460, an uncultured actinobacterium identified in a lichen-dominated cryptoendolithic community from the McMurdo Dry Valleys, Antarctica. In the study conducted by de la Torre *et al.*, (2003), clone FBP460 formed a distinct cluster with two additional actinobacterial clones identified from the same library. These sequences were closely related to *Geodermatophilus obscurus* and *Sporichthya polymorpha,* bacterial strains that have previously been isolated from rock-dwelling microbial communities (Eppard *et al.*, 1996). *Sporichthya polymorpha* strain IFO 12702 was also identified as the closest type strain to clone UWV4-34 sharing 93% 16S rRNA identity. Clones MV11-19, MV11-25 and UWV4-3 formed their own distinct cluster within the *Actinobacteria* and were most closely related to sequences identified in aquatic environments. Clones UWV4-3 and MV11-25 shared 95% and 96% identity respectively to an Arctic sea ice clone (Brinkmeyer & Helmke, unpublished), whilst MV11-19 shared 95% identity to a sequence recovered from Mono Lake in California, an alkaline, hypersaline soda lake

(Humayoun *et al.*, Unpublished). The same three database sequences were identified as the closest relatives for each clone and additionally included a sequence recovered from a lake in the Hawaiian Archipelago. The pairwise identity between sequences was 95 to 98%.

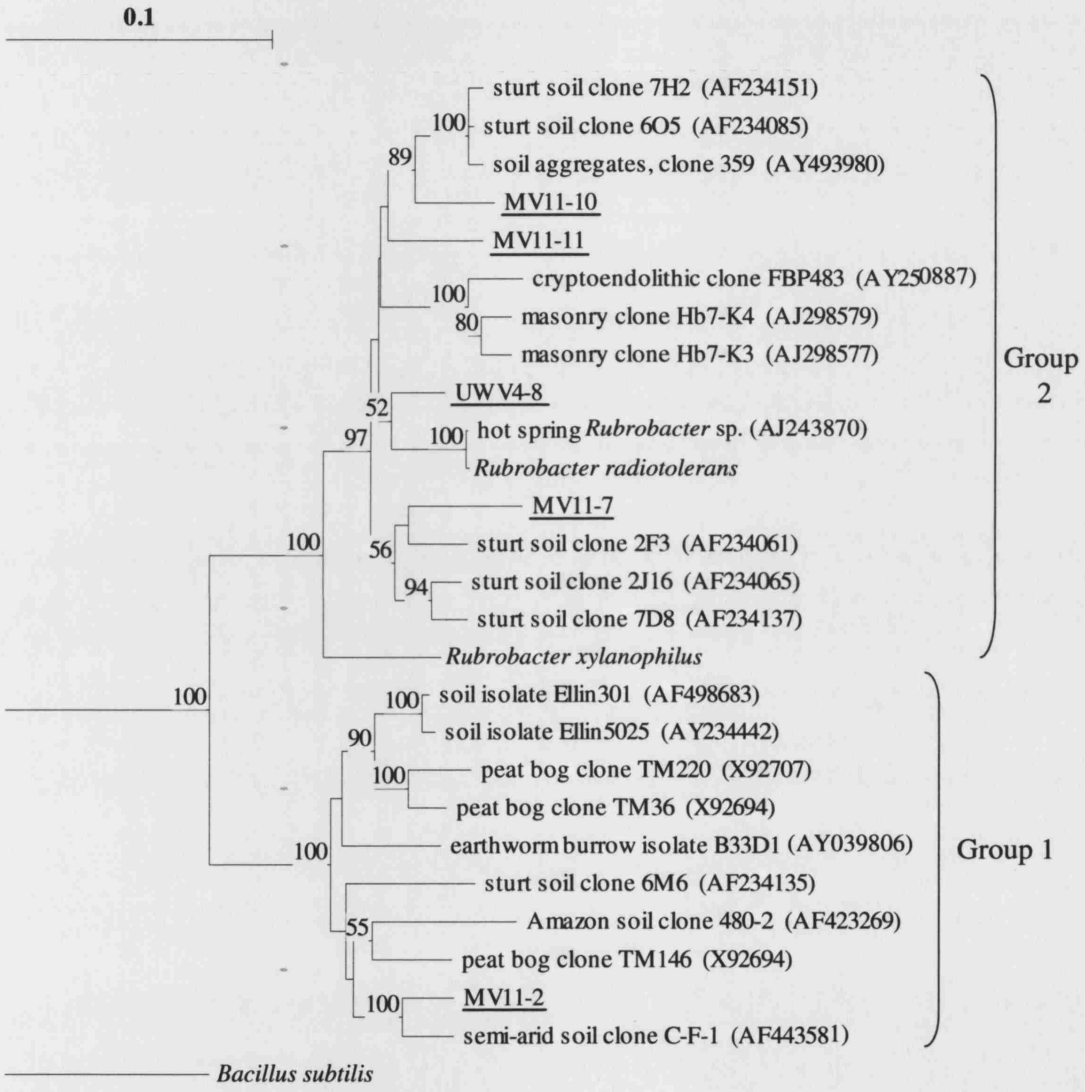


Figure 3.4: Phylogenetic affiliation of Antarctic clone sequences to members of the family *Rubrobacteridae*. Group 1 and 2 *Rubrobacter* are illustrated. Antarctic clone sequences are shown underlined. Bootstrap values, given as a percentage of 100, are indicated for branches supported by more than 50% of trees. Scale bar: 0.1, estimated number of substitutions per nucleotide position.

Three classes of the *Proteobacteria* were represented by eight Antarctic clones, five of which were affiliated to the *Alphaproteobacteria*. Four clones formed a separate cluster within the *Alphaproteobacteria* with members of the well-characterized family, *Sphingomonadaceae*. Clone UWV4-17 shared 96% identity with a bacterium isolated from rhizosphere soil (Lukow. PhD Thesis), and its closest type strain was identified as *Caulobacter leidyia* strain ATCC 15260. Clones MV11-8 and MV11-16 both shared 98% identity to sequences recovered from heavy-metal-contaminated environments (Ellis *et al.*, 2003; Geissler *et al.*, unpublished), whilst UWV4-36 exhibited 99% identity to an agricultural soil bacterium clone (Lukow, unpublished). All three clones possessed *Sphingomonas aquatilis* strain JSS-7 as their closest type strain with pairwise identities for 16S rRNA sequences ranging from 93% to 95%. Members of the *Sphingomonas* genus are commonly identified in contaminated soils by virtue of their ability to use polycyclic aromatic hydrocarbons (PAH) as a sole source of carbon and energy (Leys *et al.*, 2004). Indeed, *Sphingomonas* species have been identified in fuel contaminated soils collected near Scott base (Farrel *et al.*, 2003). The remaining clone affiliated to the *α-proteobacteria*, MV11-6, exhibited 94% identity to *Rickettsia bellii*, a tick-associated bacterium. Members of this genus are pathogenic to man and are best known as the causative agents of spotted fever and typhus (Walker & Gage, 1997). Interestingly, these organisms are obligate intracellular parasites that multiply only within a host cell and require an arthropod to serve either as a reservoir or vector (Roux & Raoult, 1995). Therefore, this may infer the presence of an arthropod such as a mite in the Miers Valley environment.

Two clones were affiliated to the *Beta-proteobacteria* and both possessed relatives identified from Antarctic environments. Clone MV11-14 shared 99% identity to a bacterium cultivated from a cryoconite hole of the Canada Glacier in the McMurdo Dry Valleys. The isolate in question, CanDirty89, was found to share 97% 16S rDNA identity to *Janthinobacterium lividum*, a freshwater bacterium of the family *Oxalobacteraceae* (Christner *et al.*, 2003a). Clone MV11-1 exhibited 97% identity to an uncultured bacterium identified in penguin dropping sediments from Ardley Island (Zhang *et al.*, Unpublished). Additionally, this clone shared 96% identity with a

Lautropia mirabilis isolate, a novel Gram-negative motile coccus isolated from the human mouth (Gerner-Smidt *et al.*, 2001). Furthermore, closest database relatives included clone sequences identified within gas hydrates in the Gulf of Mexico, a denitrifying reactor treating landfill leachate, Arctic sea ice, and benzene contaminated groundwater (Lanoil *et al.*, 2001; Etchebehere *et al.*, 2002; Brinkmeyer *et al.*, 2003; Alfreider & Vogt, unpublished).

Clone MV11-23 was the sole representative of the *Deltaproteobacteria* sharing 93% identity to an uncultured bacterium, clone 0319-6G20, present in an Australian desert soil. This clone was identified from the same library that was shown to be dominated by Rubrobacteria group I sequences (Holmes *et al.*, 2000). Additional relatives of MV11-23 included uncultured bacteria identified in Amazon soil (Borneman &Triplett, 1997; Kim & Crowley (b), unpublished), siliceous sedimentary rock (Yoshida *et al.*, Unpublished), marine sediment (Sun *et al.*, 2001) and a *Desulfosarcina* species present in Antarctic sediment (Purdy *et al.*, 2003) (although the level of identity was only 84% for the latter two sequences).

Four clone sequences, two each from the Miers Valley and Upper Wright libraries were found to be affiliated to the phylum TM7. Clone UWV4-33 exhibited 91% identity to a sequence recovered from travertine depositional facies of Mammoth Hot Springs in Yellowstone National park (Bonheyo *et al.*, Unpublished). Clones UWV4-22 and MV11-15 shared 93% and 94% identity respectively, to 16S rRNA sequences from uncultured bacteria present in pinyon-juniper forest soil in Arizona (Dunbar *et al.*, 2002). Clone MV11-13 shared 95% identity to an uncultured bacterium identified in a sewage treatment plant (Hugenholtz *et al.*, 2001). Figure 3.5 illustrates the phylogenetic positioning of Antarctic sequences relative to other members of this phylum. Clones UWV4-22, MV11-13 and MV11-15, fall within a cluster of closely related sequences, whilst clone UWV4-33 is more highly divergent. TM7 sequences are widely distributed throughout the environment having been identified in terrestrial and aquatic ecosystems (of diverse geographical locations), industrial and sludge settings and also within the

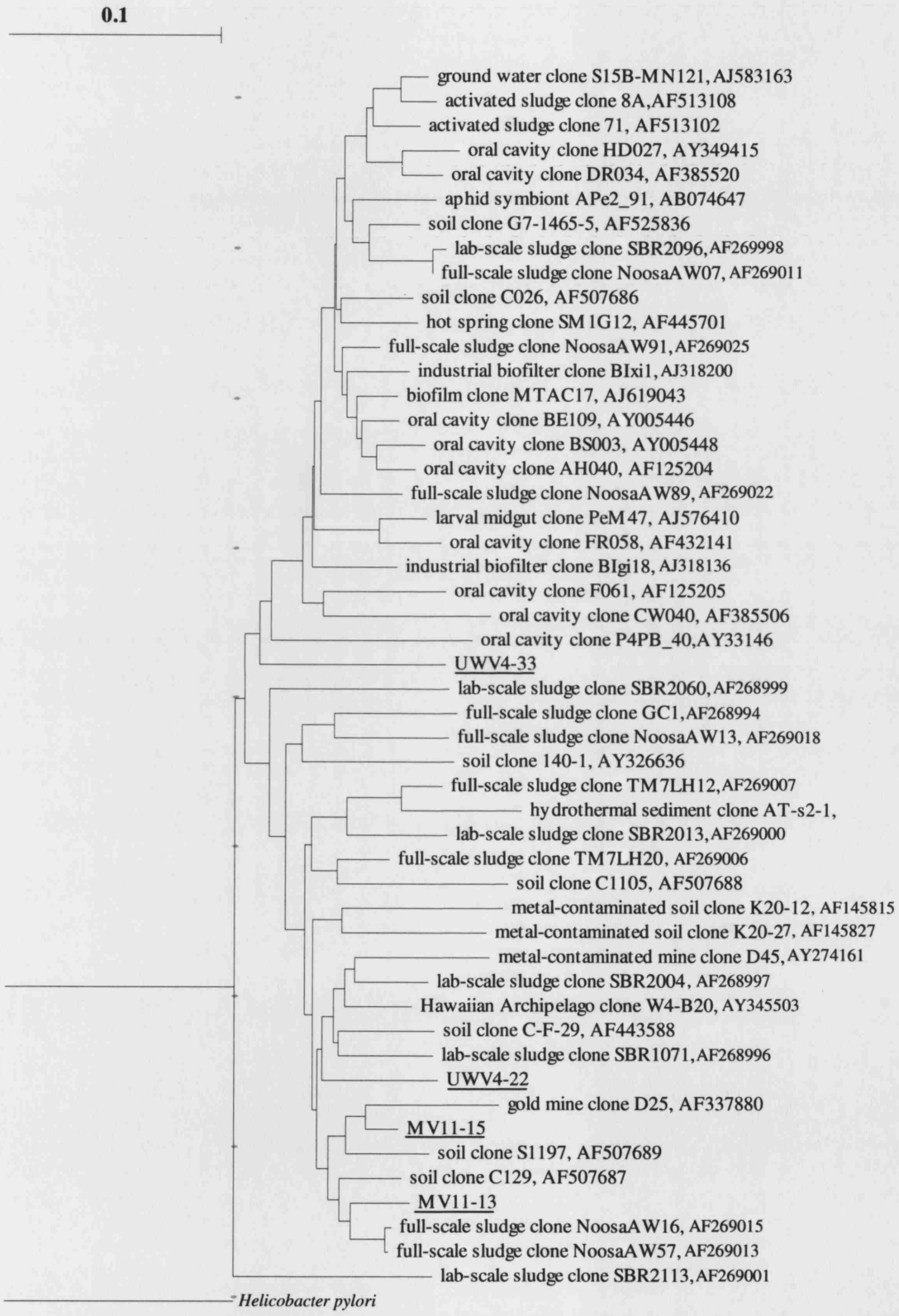

0.1
ground water clone S15B-MN121, AJ583163
activated sludge clone 8A, AF513108
activated sludge clone 71, AF513102
oral cavity clone HD027, AY349415
oral cavity clone DR034, AF385520
aphid symbiont APe2_91, AB074647
soil clone G7-1465-5, AF525836
lab-scale sludge clone SBR2096, AF269998
full-scale sludge clone NoosaAW07, AF269011
soil clone C026, AF507686
hot spring clone SM1G12, AF445701
full-scale sludge clone NoosaAW91, AF269025
industrial biofilter clone BIxi1, AJ318200
biofilm clone MTAC17, AJ619043
oral cavity clone BE109, AY005446
oral cavity clone BS003, AY005448
oral cavity clone AH040, AF125204
full-scale sludge clone NoosaAW89, AF269022
larval midgut clone PeM47, AJ576410
oral cavity clone FR058, AF432141
industrial biofilter clone BIgi18, AJ318136
oral cavity clone F061, AF125205
oral cavity clone CW040, AF385506
oral cavity clone P4PB_40, AY33146
UWV4-33
lab-scale sludge clone SBR2060, AF268999
full-scale sludge clone GC1, AF268994
full-scale sludge clone NoosaAW13, AF269018
soil clone 140-1, AY326636
full-scale sludge clone TM7LH12, AF269007
hydrothermal sediment clone AT-s2-1,
lab-scale sludge clone SBR2013, AF269000
full-scale sludge clone TM7LH20, AF269006
soil clone C1105, AF507688
metal-contaminated soil clone K20-12, AF145815
metal-contaminated soil clone K20-27, AF145827
metal-contaminated mine clone D45, AY274161
lab-scale sludge clone SBR2004, AF268997
Hawaiian Archipelago clone W4-B20, AY345503
soil clone C-F-29, AF443588
lab-scale sludge clone SBR1071, AF268996
UWV4-22
gold mine clone D25, AF337880
MV11-15
soil clone S1197, AF507689
soil clone C129, AF507687
MV11-13
full-scale sludge clone NoosaAW16, AF269015
full-scale sludge clone NoosaAW57, AF269013
lab-scale sludge clone SBR2113, AF269001
Helicobacter pylori

Figure 3.5: Phylogenetic relationship of Antarctic clone sequences (shown underlined) to members of the phylum TM7. This phylum contains no cultivated representatives. Scale bar: 0.1, estimated number of substitutions per nucleotide position.

human oral cavity (Hugenholtz *et al.*, 2001; Paster *et al.*, 2001; Dunbar *et al.*, 2002; Nemergut *et al.*, 2004).

Three clones were affiliated to the *Acidobacteria* and included clone UWV4-1, which represented the most abundant phylotype in the Upper Wright library, accounting for 11 out of the 24 clones analyzed. Clone UWV4-1 shared 98% identity to an uncultured bacterium identified in a lichen-dominated cryptoendolithic community in the McMurdo Dry Valleys (de la Torre *et al.*, 2003). Interestingly, the clone in question represented less than 1% of the library analyzed in the study by de la Torre *et al.*, (2003). Additional database relatives included clone sequences recovered from heavy-metal contaminated mine tailings, sludge and Antarctic lake ice (Nemergut *et al.*, 2004; Crocetti *et al.*, 2002; Gordon *et al.*, Unpublished). Clones MV11-27 and MV11-31 exhibited 98% identity to sequences present in arid soil and polychlorinated biphenyl-polluted soil respectively (Kuske *et al.*, 1997; Nogales *et al.*, 2001).

Figure 3.6 represents a more detailed phylogenetic analysis of sequences belonging to the division *Acidobacteria* from which eight subdivisions can be identified. Previously, this phylum was represented by only three cultivated species: *Acidobacterium capsulatum* (subdivision 1), *Holophaga foetida*, and *Geothrix fermentans* (both subdivision 8) (Hugenholtz *et al.*, 1998). However, recently Sait and colleagues (2002) were able to cultivate a further ten isolates from soil belonging to subdivisions 1 and 3, all of which were cultivated under aerobic conditions on simple media of slightly acidic pH. The phylogenetic analysis revealed clone MV11-31 to be affiliated to subdivision 3, sharing 93% 16S rRNA identity to soil isolate Ellin342. Clones MV11-27 and UWV4-1 were related to sequences within subdivision 4, which contains no cultivated representatives. The results of similar cultivation-independent studies reveal the

Acidobacteria to have a widespread distribution in nature, particularly in soils (Kuske *et al.*, 1997; McCaig *et al.*, 1999; Buckley & Schmidt, 2003).

Clone MV11-32 was the sole representative of the *Verrucomicrobia*, a phylum that was only recognized in 1995 (Ward-Rainey *et al.*, 1995). Detailed phylogenetic analysis of sequences belonging to this phylum has identified five distinct subgroups, only two of which contain cultured representatives. Subdivision 1 includes the obligate, aerobic heterotrophs of the genus *Prosthecobacter* and *Verrucomicrobium*, whilst subdivision 4 contains *Opitutus terrae* and related isolates cultivated from soil from a rice paddy field (Rappe & Giovannoni, 2003). The phylogenetic tree shown in Figure 3.7 reveals clone MV11-32 to be affiliated to subdivision 2, a group that includes sequences of soil and freshwater origin; its closest relative with which it shared 94% identity was an environmental clone sequence identified in penguin dropping sediments from Ardley Island Antarctica (Zhang *et al.*, Unpublished). Additionally, MV11-32 was closely related to soil clones of diverse geographical origins, identified in soils of the US, Australia and the Amazon (Kuske et al., 1997; Liesack & Stackebrandt, 1992; Kim & Crowley, unpublished). Subdivision 2 also includes the two *Xiphinematobacter* species that are bacterial endosymbionts of the nematode species *Xiphinema americanum*; however, these organisms have yet to be cultivated. As with the *Acidobacteria*, the *Verrucomicrobia* appear to be widely distributed throughout the environment, particularly in soils, and the phylum is composed almost entirely of environmental clone sequences (Borneman & Triplett, 1997; Kuske *et al.*, 1997; Zwart *et al.*, 1998).

Two sequences recovered from the Upper Wright library were affiliated to the *Cytophaga-Flexibacter-Bacteroides* group. Clone UWV4-27 shared 96% 16S rRNA identity to a sequence present in agricultural soil from Germany (Lukow, unpublished), whilst UWV4-18 possessed 93% identity to a clone identified in heavy metal-contaminated mine tailings from an abandoned gold mine in Colorado (Nemergut *et al.*, 2004). A single clone, UVW4-20, was affiliated to the *Deinococcus-Thermus* lineage and exhibited 100% identity to an uncultured *Deinococcus* spp. identified in a cyanobacterium-dominated cryptoendolithic community in the McMurdo Dry Valleys

(de la Torre *et al.*, 2003). In this particular study, the authors revealed the clone in question, FBP266, to form an independent group within the *Deinococcus-Thermus* group with sequences identified solely from Antarctic microbial communities. The phylum *Planctomycetales* was also represented by a single clone from the Upper Wright library, UWV4-13, sharing 90% identity to an uncultivated bacterium identified in Lake sediment in the Hawaiian Archipelago (Donachie *et al.*, Unpublished). Additional database relatives included clone sequences recovered from agricultural soil and polychlorinated biphenyl-polluted soil (Lukow, Unpublished; Nogales *et al.*, 2001), and *Nostocoida limicola*, a filamentous bacterium cultivated from activated sludge (Liu *et al.*, Unpublished). The sequence appears to be related to an *Isosphaera* sp., with the closest related type strain identified as *Isosphaera pallida* strain DSM 9630T, an aerobic, heterotrophic bacterium isolated from hot springs in North America (Giovannoni *et al.*, 1987). *Planctomycetales* is another phylum that is characterized predominantly by environmental clone sequences, including those identified from soils (Liesack & Stackebrandt, 1992; Rappe & Giovannoni, 2003).

Clone MV11-3 was phylogenetically affiliated to the *Chloroflexi*, a deep-branching lineage of the domain *Bacteria*, also known as the green non-sulfur bacteria (Rappe & Giovannoni, 2003). Its closest relative sharing 94% identity was identified as a Gram-positive bacterial symbiont of the cuttlefish *Sepia officinalis* identified in the English Channel (Grigioni *et al.*, 2000). Additional database relatives included clone sequences identified in sediments from Mammoth Cave in Kentucky (Fowler et al., Unpublished), Sanriku marine sediments (Asami & Watanabe., Unpublished), penguin dropping sediments from Antarctica (Zhang *et al.*, Unpublished), and Antarctic continental shelf sediments (Bowman & McCuaig, 2003). Like the *Verrucomicrobia* and *Acidobacteria*, this phylum contains a limited number of cultured representatives though *Chloroflexi* rRNA gene sequences are often identified in clone libraries, including those constructed from soil (Dunbar *et al.*, 2002; Furlong *et al.*, 2002; Rappe & Giovannoni, 2003).

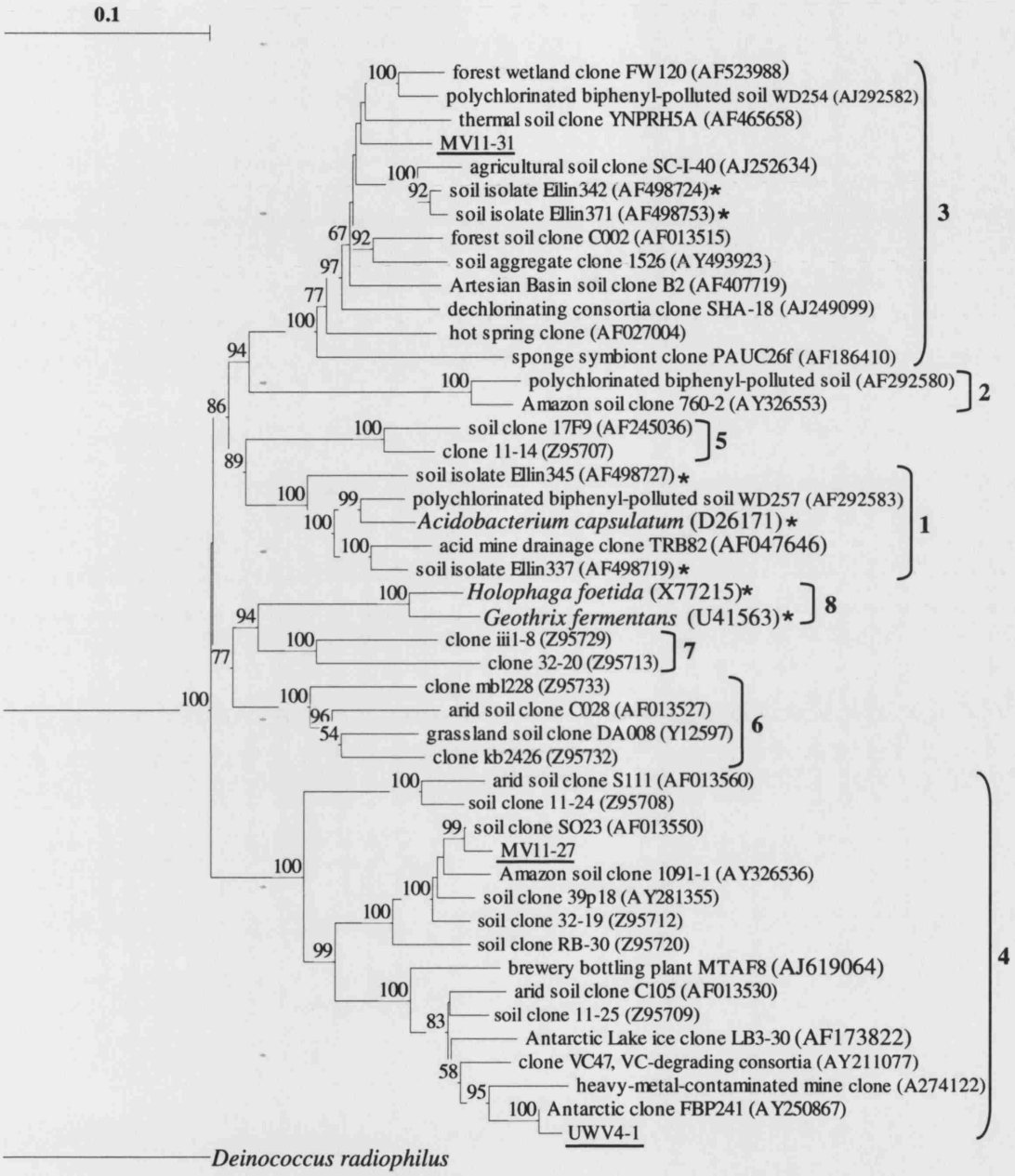


Figure 3.6: Phylogenetic affiliation of Antarctic clone sequences (shown underlined) within the division *Acidobacteria*. The eight subgroups of the *Acidobacteria* are illustrated. Clones MV11-27 and UWV4-1 are affiliated to subgroup 4 which contains not cultivated representatives; clone MV11-31 is affiliated to subgroup 3 and is related to isolates cultivated from Australian pasture soil. Asterisks denote cultivated isolates. Bootstrap values, given as a percentage of 100, are indicated for branches supported by more than 50% of trees. Scale bar: 0.1, estimated number of substitutions per nucleotide position.

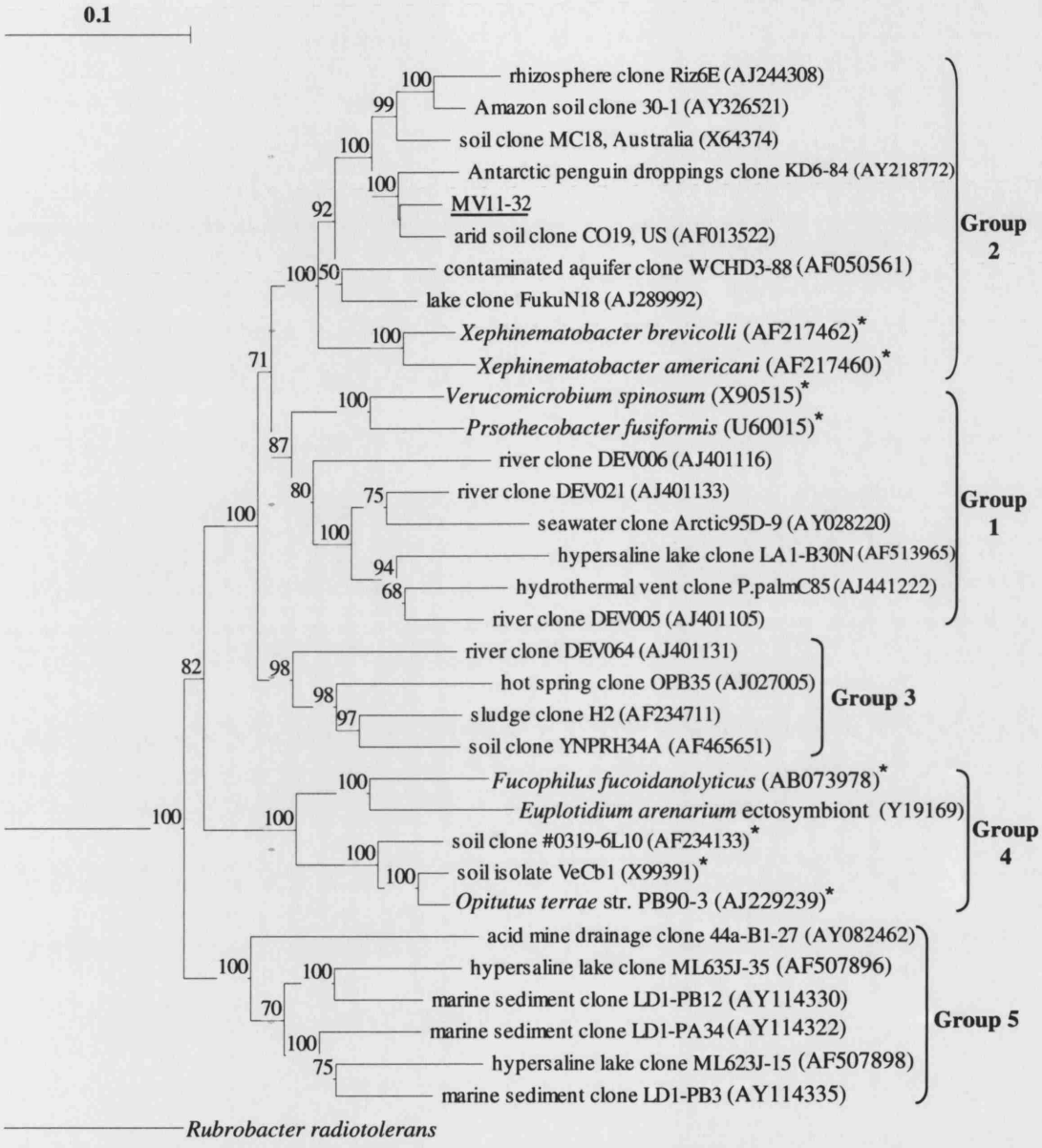


Figure 3.7: Phylogenetic positioning of Antarctic clone sequence MV11-32 (shown underlined) within the division *Veruccomicrobia*. The five subgroups of are illustrated. Asterisks denote cultivated isolates. Bootstrap values, given as a percentage of 100, are indicated for branches supported by more than 50% of trees. Scale bar: 0.1, estimated number of substitutions per nucleotide position.

3.4: Summary

Bacterial diversity within Antarctic Dry Valley soil samples collected from two different locations in the McMurdo Dry Valleys, the Miers Valley (MV) and the Upper Wright Valley (UWV), was investigated using both heterotrophic cultivation and cultivation-independent techniques.

Three samples were used in the cultivation study and included samples MV16 and UWV4, both desiccated surface mineral soils, and sample MV17 that was collected beneath a lake microbial mat in the Miers Valley. The latter was selected as it represented a comparatively more nutrient-rich and sheltered environment. Three types of growth medium were employed that were designed to limit the growth of faster-growing organisms, which predominate and contribute to the reduce diversity observed on nutrient-rich media. A single incubation temperature of 4°C was used. A total of 29 bacterial isolates was identified by 16S rRNA sequence analysis, representing 24 unique isolates distributed across 4 bacterial phyla. With the exception of a single isolate, all isolates exhibited ≥ 98% identity 16S rRNA identity to previously cultivated species.

Two samples were used for the molecular assessment of bacterial diversity and included samples MV11.1 and UWV4. Sample MV11.1, collected during February 1999, was the first soil sample to be investigated prior to collection of fresh samples in February 2002. Partial sequences were obtained for a total of 43 randomly selected bacterial 16S rRNA clones from which 30 phylotypes were identified. Twenty-nine sequences were included in the phylogenetic analysis and were distributed across nine bacterial phyla. The majority of these sequences possessed an environmental clone sequence recovered from similar culture-independent study as their closest database relative, and twelve Antarctic sequences shared <95% 16S rRNA identity to a database sequence.

In summary, the molecular survey identified a greater level of genetic diversity among bacterial species inhabiting Dry Valley soils than the diversity recovered from the heterotrophic cultivation of bacteria. Taken together these results reveal Antarctic Dry

Valley soils to support a complex microbial community, of which some members are significantly different to bacteria identified in other environments.

Chapter 4

Diversity of *Archaea* and *Eukarya*

4.1: Introduction

The aim of the work described in this chapter was to investigate diversity of other domains of life in Antarctic Dry Valley soils in addition to the bacteria described in Chapter 3. A molecular phylogenetic approach was used to characterize diversity both within the prokaryotic domain of *Archaea* and within the domain *Eukarya*. My hypothesis was that the Dry Valley environment would support a significantly lower diversity of archaeal phylotypes in contrast to the comparative molecular assessment of bacterial diversity. Additionally, archaeal clone sequences would be phylogenetically distinct from their cultured counterparts, exhibiting a higher degree of similarity to other archaeal environmental clone sequences, as evidenced by similar molecular-based studies of archaeal diversity in natural environments. The diversity of eukaryotes was also hypothesized to be low in comparison to bacterial diversity.

4.2.: Diversity of Archaea

4.2.1: Background

A molecular phylogenetic approach was used to investigate diversity of *Archaea* in two separate locations of the Antarctic Dry Valleys, the Miers Valley and the Upper Wright Valley. *Archaea*-specific PCR primers were used to construct 16S rRNA clone libraries from total community DNA extracted from samples MV11.1 (Miers Valley) and UWV4 (Upper Wright Valley). ARDRA and RFLP-typing were used for the analysis of recombinant clones from the Miers Valley and Upper Wright libraries respectively; recombinant plasmids showing differing restriction patterns to each other were randomly selected for sequencing.

4.2.2: Diversity of Archaea in the Miers Valley

4.2.2.1: Construction of 16S rRNA gene clone library

The MV11.1 archaeal 16S rRNA gene clone library was constructed using an *Archaea*-specific forward primer and a universal reverse primer, Arch21f and Un1492R

respectively. These primers correspond to positions 7 and 1510 of the 16S rRNA gene (*Escherichia coli* numbering) resulting in amplification of almost the entire sequence of the gene. The products of five replicate PCR reactions following 25 cycles of amplification were pooled together prior to cloning in order to minimize any bias introduced by the PCR. Recombinant plasmids were confirmed by digestion with *EcoR*1, after which inserts were reamplifed and ARDRA performed using the two tetrameric restriction enzymes, *Msp*1 and *Hin*P1. This allowed grouping of clones into classes containing identical restriction patterns. A representative clone(s) for each ARDRA group was selected for partial sequencing.

4.2.2.2: ARDRA analysis of archaeal clones

A total of 96 clones was analyzed by ARDRA from which 11 unique restriction patterns or ARDRA groups were identified (Table 4.1). An example of an ARDRA gel is shown in Figure 4.1. ARDRA group 1 represents the most common restriction pattern accounting for 78% of all clones analyzed. In total, six unique *Msp*1 restriction patterns were evident, excluding clone no. 31 (group 6) which had an *Msp*1 pattern that was highly similar to that of group 1. Six unique *Hin*P1 restriction patterns were also identified, excluding clone no. 36 (group 7) for which the *Hin*P1 pattern was highly similar to that of group 1. As is evident from Table 4.1, only the first three ARDRA groups exhibit both a unique *Msp*1 and *Hin*P1 restriction pattern. Subsequent new ARDRA groups were identified on the basis of possessing either one unique restriction pattern not previously identified, or alternatively, two previously characterized restriction patterns occurring together in a novel combination. Figure 4.2 illustrates a collectors curve showing the number of ARDRA groups as a function of the number of clones analyzed. The coverage values of the library, which takes into account the proportion of ARDRA patterns that occurred only once, is calculated as 0.93, inferring 93% of ARDRA diversity within the library has been identified.

Closest database relatives of archaeal clone sequences representing different ARDRA groups were identified by BLAST analysis and are represented in Table 4.2. Archaeal clones representing ARDRA groups 4, 7, 8, 9 and 10 shared the same closest database

relative, exhibiting 97-99% identity to the 16S rRNA gene sequence of a 34kb fragment of an uncultured crenarchaeote recovered from calcerous grassland soil in Germany (accession number AJ496176) (Quaiser *et al.*, 2002). Clones A100 and A31, representing ARDRA groups 5 and 6 respectively, shared 98% 16S rRNA identity with an unidentified archaeon (accession number U62819) present in a Wisconsin agricultural soil (Bintrim *et al.* 1997); clone A61 of ARDRA group 1 exhibited 99% identity to an unidentified archaeon (accession number U62811) from the same environment. The closest relative of clone A49 (ARDRA group 2) sharing 99% identity was identified as an uncultured crenarchaeote present in uranium mining waste (accession number AJ535121, unpublished).

Clone A58, representing ARDRA group 3, was found to share 93% rDNA identity to an uncultured soil bacterium present in soil from Arizona (accession number Z95720) (Ludwig *et al.*, 1997). Similarly, clone A26 of the same ARDRA group was also most closely related (92% over 400 bp) to an uncultured bacterium identified from this study (accession number AF507436). Thus, it would appear that primer Arch21f was able to bind to and amplify this specific bacterial 16S rDNA sequence in combination with Un1492R, a universal 16S rRNA primer. Clone A58 was selected for additional sequencing with the archaeal-specific primer 333f; however, this reaction failed and the sequence was not investigated further.

Seven clones representing six ARDRA groups were selected for extensive sequencing of both strands of DNA and these sequences were used in the phylogenetic reconstruction (Table 4.3)

Table 4.1: Summary of ARDRA groups. A total of 96 archaeal clones were analysed by ARDRA from which 11 groups were identified. Only clones belonging to ARDRA groups 1-3 possessed a unique *Msp*1 and *HinP*1 restriction pattern.

ARDRA Group	Clone No.	Total No. of Clones	*Msp*1 Restriction Pattern	*Hin*P1 Restriction Pattern
1	1,2,3,4,5,7,8,10,11,12,13,14,16,17,18,19,20,21,24,25, 27,30,32,33,35,37,38,39,40,41,42,43,45,46,48,50,51, 52,54,55,56,59,60,61,62,63,65,66,67,68,70,71,72,73, 75,76,77,78,79,80,82,83,85,86,87,90,92,93,94,95,96, 97,99,102,103	75	Group 1	Group 1
2	6,9,28,29,44,49	6	Group 2	Group 2
3	15,26,58	3	Group 3	Group 3
4	22	1	Group 1	Group 4
5	23, 64, 84, 100, 104	5	Group 1	Group 5
6	31	1	Group 1 'similar'	Group 5
7	36	1	Group 1	Group 1 'similar'
8	74	1	Group 1	Group 8
9	81	1	Group 9	Group 1
10	88	1	Group 10	Group 4
11	91	1	Group 11	Group 1

Figure 4.1: Example of an ARDRA gel. The *Hin*P1 and *Msp*1 restriction patterns for 15 archaeal 16S rRNA gene clones are shown below.

Figure 4.1. a) *Msp*1 digest of 15 archaeal 16S rDNA clones. Lane 1 contains a ΦX174 DNA-*Hae*III molecular weight marker (fragment sizes are shown in bp). Lanes 2, 3 and 5-16 illustrate the ARDRA group 1 *Msp*1 pattern whilst lane 4 illustrates the ARDRA group 3 *Msp*1 pattern.

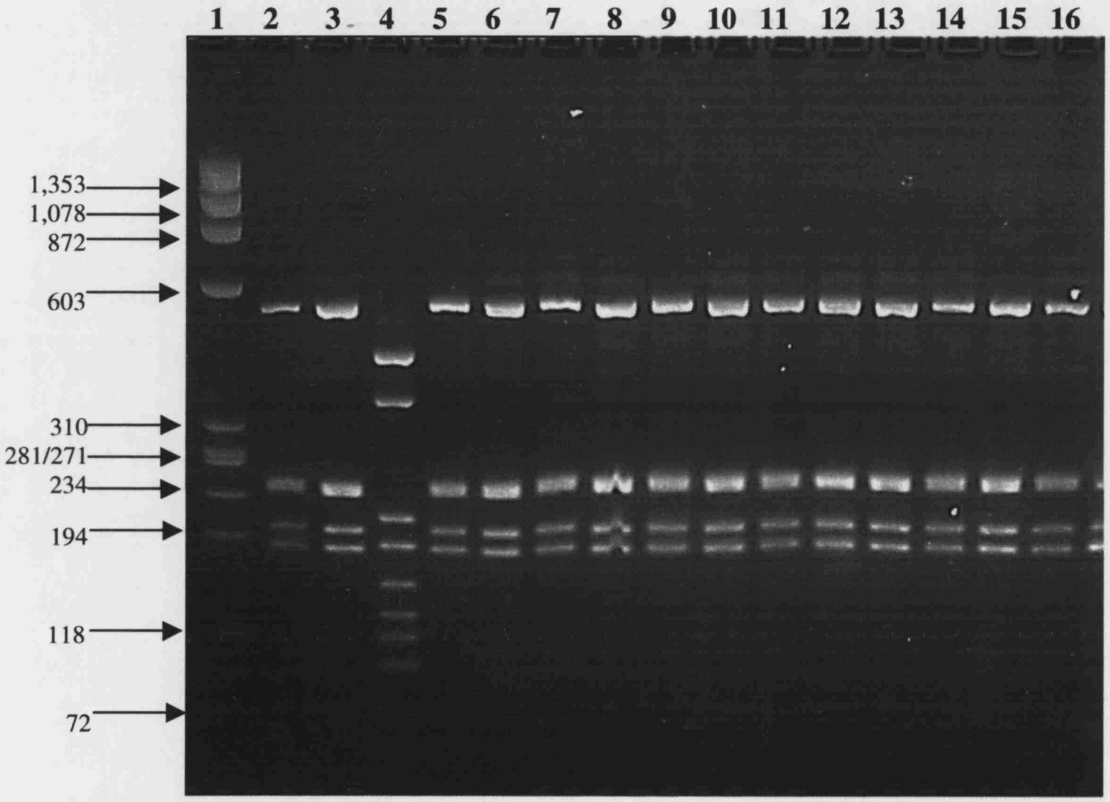
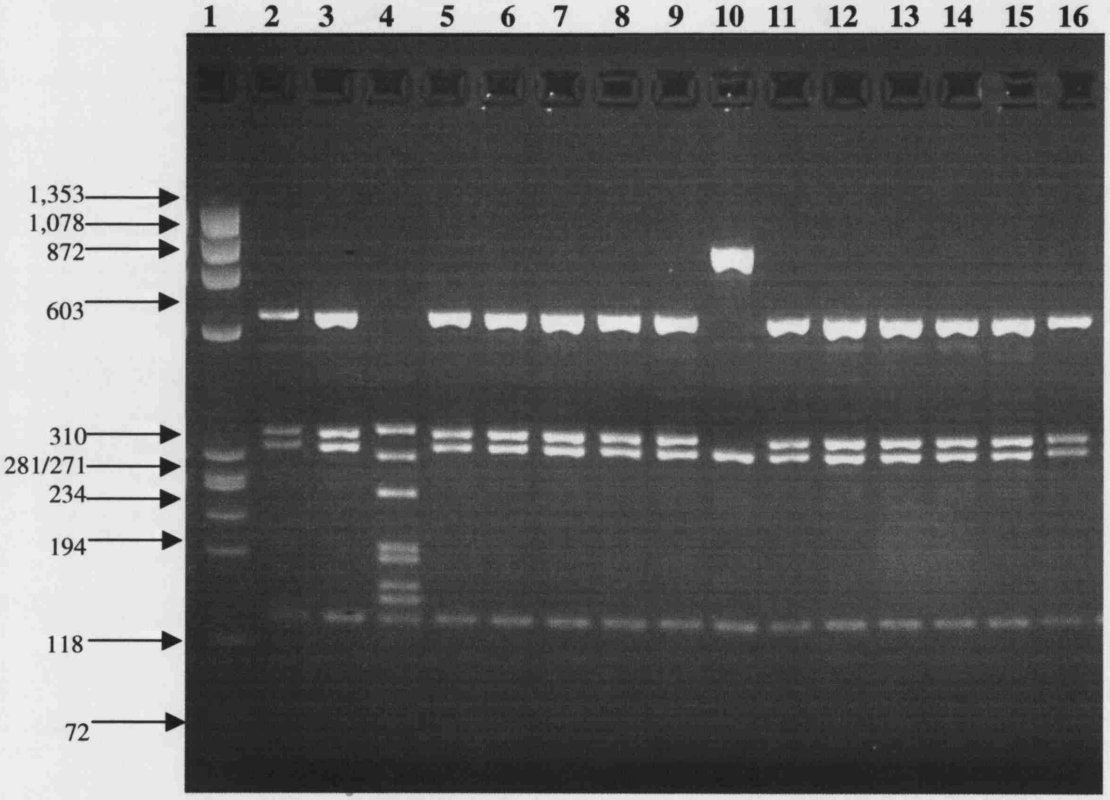


Figure 4.1. b) *Hin*P1 digest of the same 15 archaeal 16S rDNA clones. Lanes 1 contains a ΦX174 DNA-*Hae*III molecular weight marker. Lanes 2, 3, 5-9 and 11-16 illustrate the ARDRA group 1 *Hin*P1 pattern; lanes 4 and 10 illustrate ARDRA group 3 and group 5 *Hin*P1 patterns respectively. As demonstrated by these results, some ARDRA groups possess both different *Msp*1 and *Hin*P1 restriction patterns as in the case of groups 3 and 1, shown in lanes 3 and 4 of each gel photo respectively. In contrast, ARDRA groups 1 and 5 possess differing *Hin*P1 restriction patterns as shown in lanes 9 and 10 respectively of Figure 4.1.b) but share identical *Msp*1 restriction patterns (see lanes 9 and 10 of Figure 4.1.a).



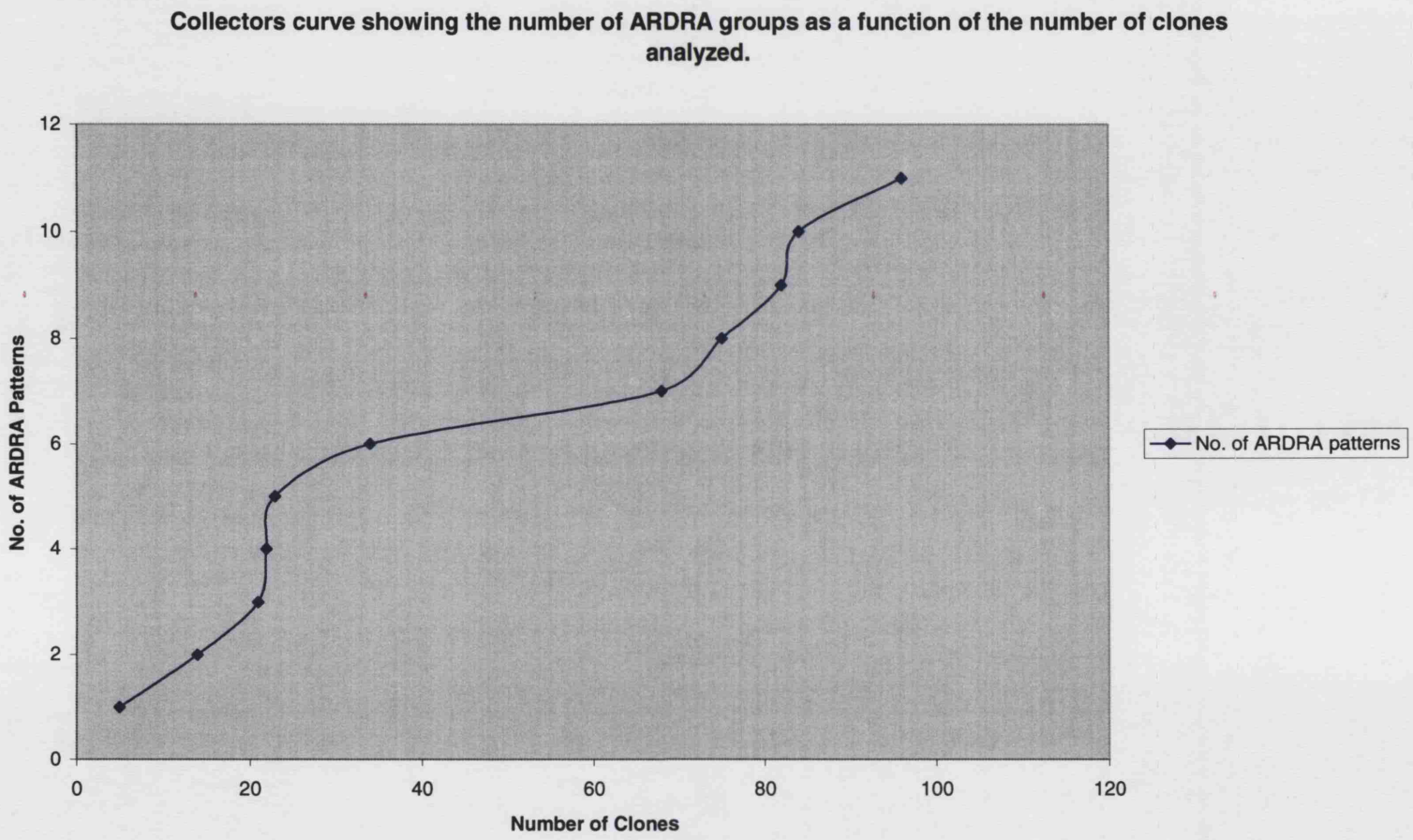


Figure 4.2: Collectors curve of ARDRA analysis of 96 clones from the MV11.1 archaeal library. Coverage = 1-(n1/N), where n1 is the number of clones that occurred only once and N is the total number of clones examined. The coverage value for ARDRA analysis of MV11.1 archaeal library is 0.93.

Table 4.2: Closest database relatives of MV11.1 archaeal sequences representing different ARDRA groups.

ARDRA Group	No. of Clones/ Group	Representative Clone	Closest Relative	Environment Description	Accession No. of Closest Relative	% Identity (Length bp)
1	75	A61	unidentified archaeon, SCA1145	agricultural soil, Wisconsin, USA	U62811	99 (1010)
2	6	A49	uncultured crenarchaeote, clone Gitt-GR-47	uranium mining waste	AJ535121	99 (1002)
3	3	A58	uncultured soil bacterium	soil, Arizona, USA	AF507752	93 (500)
4	1	A22	uncultured crenarchaeote 34 kb genomic fragment	calcerous grassland soil, Darmstadt, Germany	AJ496176	99 (974)
5	5	A100	unidentified archaeon, SCA1175	agricultural soil, Wisconsin, USA	U62819	98 (1001)
6	1	A31	unidentified archaeon, SCA1175	agricultural soil, Wisconsin, USA	U62819	98 (571)
7	1	A36	uncultured crenarchaeote 34 kb genomic fragment	calcerous grassland soil, Darmstadt, Germany	AJ496176	98 (568)
8	1	A74	uncultured crenarchaeote 34 kb genomic fragment	calcerous grassland soil, Darmstadt, Germany	AJ496176	98 (1009)
9	1	A81	uncultured crenarchaeote 34 kb genomic fragment	calcerous grassland soil, Darmstadt, Germany	AJ496176	98 (1005)
10	1	A88	uncultured crenarchaeote 34 kb genomic fragment	calcerous grassland soil, Darmstadt, Germany	AJ496176	98 (1020)

Table 4.3: MV11.1 archaeal clone sequences included in the phylogenetic analysis.

Clone Name	ARDRA Group	Sequence Length (bp)	Nearest Neighbour	% Identity	Accession Number	Ref.
A49	2	1002	uncultured crenarchaeote, Gitt-GR-47	99	AJ535121	unpublished
A61	1	1010	unidentified archaeon, SCA1145	99	U62811	Bintrim *et al.*, 1997.
A74	8	1009	uncultured crenarchaeote 34 kb genomic fragment	98	AJ496176	Quaiser *et al.*, 2002.
A81	9	1005		98		
A84	5	1020		98		
A88	10	1020		98		
A100	5	1001	unidentified archaeon, SCA1175	98	U62819	Bintrim *et al.*, 1997.

4.2.2.3: Phylogenetic analysis of archaeal clone sequences

Table 4.3 lists the archaeal clone sequences recovered from sample MV11.1 used in the phylogenetic analysis. The phylogenetic positions of clones recovered from Antarctic soil were determined relative to their closest related sequences and other crenarchaeotal clones previously described in the literature. Additional clone sequences included in the analysis and the environment in which they were found are described in Table 4.4. Also included in the phylogenetic reconstruction were representatives of the cultured *Crenarchaeota* and the archaeal kingdoms, *Euryarchaeota* and *Korearchaeota* (Table 4.5). A total of 62 sequences were used in the phylogenetic reconstruction.

Comparison of Antarctic clone sequences with existing sequences publicly available from the GenBank database, revealed 98-99% shared identity to uncultured environmental archaeal sequences retrieved from agricultural and calcerous grassland soils and uranium mining waste. Phylogenetic analysis of the sequences clearly indicates the presence of non-thermophilic *Crenarchaeota* in the Antarctic Dry Valley environment that are phylogenetically distinct from their cultured counterparts, the thermophilic *Crenarchaeota* as shown in Figure 4.3. Indeed, the analysis clearly places

Table 4.4: Crenarchaeotal clone sequences included in the phylogenetic reconstruction.

Group	Sequence Name	Environment Description	Accession No.	Reference
1.1A	SAGMA-X	waters from deep South African gold mines	AB050229	Takai *et al.*, 2001.
	GRU	freshwater lake sediment and plankton	AY278091	Ochsenreiter *et al.*, 2003.
	SB95-57	Santa Barbara Channel	U78199	Massana *et al.*, 1997.
	LMA238	Lake Michigan sediment	U87517	MacGregor *et al.*, 1997.
	Cenarchaeum symbiosum	symbiont of marine sponge, *Axinella mexicana*	U51469	Preston *et al.*, 1996.
	ARC12	freshwater ferromanganous micronodules and sediments	AF293019	Stein *et al.*, 2001.
	archaeon no. 11	Mariana Trench sediment	D87349	Kato *et al.*, 1997.
	SBAR12	coastal marine environment	M88076	Delong, 1992.
	ANTARCTIC12	Antarctic marine picoplankton	U11043	Delong, 1994.
1.1B	DAR3	Freshwater microbial mat	AY278077	Ochsenreiter *et al.*, 2003.
	SCA	Agricultural soil, Wisconsin, USA	U62811, U62814, U62815, U62817, U62819	Bintrim *et al.*, 1997.
	SAGMA	waters from deep South African gold mines	AB050221, AB050223, AB050230	Takai *et al.*, 2001.
	uncultured crenarchaeote, 34kb genomic fragment	calcerous grassland soil	AJ496176	Quaiser *et al.*, 2002.
	ROB35, ROB110	calcerous grassland soil	AY278072, AY278069	Ochsenreiter *et al.*, 2003.
	54D9	calcerous grassland soil	AY278106	Ochsenreiter *et al.*, 2003.
	TRC23	terrestrial plant roots	AF227640, (TCR23-28)	unpublished
	Gitt-GR-47	uranium mining waste	AJ535121	unpublished
	pGrfA4	freshwater sediment	U59968	
1.1C	FFSC	forest soil, Finland	AJ006920, AJ006921	Jurgens *et al.*, 1997.
	FFSB	forest soil, Finland	X96688, X96692, X96694, X96695	Jurgens *et al.*, 1997.

	S248-7	soil, mixed woodland	AY278100	Ochsenreiter *et al.*, 2003.
	GRU	freshwater lake sediment and plankton	AY278081, AY778095	Ochsenreiter *et al.*, 2003.
1.2	BBA	continental shelf sediments	AF004343, AF004344	Vetriani *et al.*, 1998.
	pGrfB286	freshwater lake sediment, Lake Griffy, USA	U59984	Hershberger *et al.*, 1996.
1.3	VAL	freshwater lake, Valkea Kotinen Lake, Finland	AJ131314, AJ131320	Jurgens *et al.*, 2000.
	JTB173	marine sediment, cold seep area, Japan Trench	AB015273	Li *et al.*, 1999.
	Arc.98	deep subsurface paleosol, Washington State, USA	AF005760	Chandler *et al.*, 1998.
Unclassified	pSL50	hot spring, Yellowstone Park	U63342	Barns *et al.*, 1994.
	pJP41	hot spring, Yellowstone Park	L25301	Barns *et al.*, 1994.
	pJP96	hot spring, Yellowstone Park	U63338	Barns *et al.*, 1994.

Table 4.5: Additional sequences included in the phylogenetic analysis representing cultured members of the *Crenarchaeota* and *Euryarchaeota* and environmental clones of the *Korearchaeota* lineage.

Kingdom	**Taxon**	**Accession Number**
Crenarchaeota	*Sulfolobus acidocaldarius*	D14876
	Thermofilum pendens	X14835
	Thermoproteus tenax	M35966
	Desulfurococcus mobilis	D36474
	Acidianus infernus	X89852
	Pyrodictium occultum	M21807
	Pyrobaculum islandicum	L07511
Euryarchaeota	*Archaeoglobus fulgidus*	Y00275
Korearchaeota	pJP27	L25852
	pJP33	L25300

the Antarctic sequences within group 1.1b of the non-thermophilic *Crenarchaeota*, also known as the 'terrestrial cluster', supported by a bootstrap value of 100%. Crenarchaeotal 16S rRNA gene sequences belonging to this group have been recovered from a range of environments and appear to have a ubiquitous distribution in soils across the globe, as represented in Table 4.6.

Figure 4.4 shows the phylogenetic position of Antarctic clone sequences (including clones A84 and A88) with respect to members of the group 1.1b *Crenarchaeota*, along with the environment from which the sequences were recovered. This phylogenetic tree was based on an alignment of group 1.1b crenarchaeotal 16S rRNA gene sequences only; *Cenarchaeum symbiosum* belonging to group 1.1a *Crenarchaeota* was used as an outgroup. The tree shown in Figure 4.4 was constructed from an alignment spanning approximately nucleotide positions 80 to 910 of the Archaea 16S rRNA compared to positions 480 to 870 for the tree shown in Figure 4.3. Thus, the former can be considered a more accurate representation of the phylogenetic position of Antarctic clones as it utilizes a greater proportion of sequence information contained within the 16S rRNA. This would account for the minor changes observed in branching order of group 1.1b sequences between the two trees.

Table 4.6: Environments in which Group 1.1b *Crenarchaeota* have been identified.

Environment	**Ref.**
agricultural soil, Wisconsin (USA)	Bintrim *et al.*, 1997
soil, Michigan (USA)	Buckley *et al.*, 1998
heavy-metal-contaminated soil (Germany)	Sandaa *et al.*, 1999
associated with terrestrial plant roots	Simon ., 2000
hindgut of a soil-feeding termite	Friedrich *et al.*, 2001
grassland, mixed woodland and sandy ecosystem (Germany)	Ochsenreiter *et al.*, 2003
soft limestone (UK), permafrost area (Russia)	"
microbial mat (Germany)	"
freshwater lake (Germany)	"
freshwater sediment, Lake Griffy (USA)	Hershberger *et al.*, 1996
subsurface goldmine waters (South Africa)	Takai *et al.* 2001

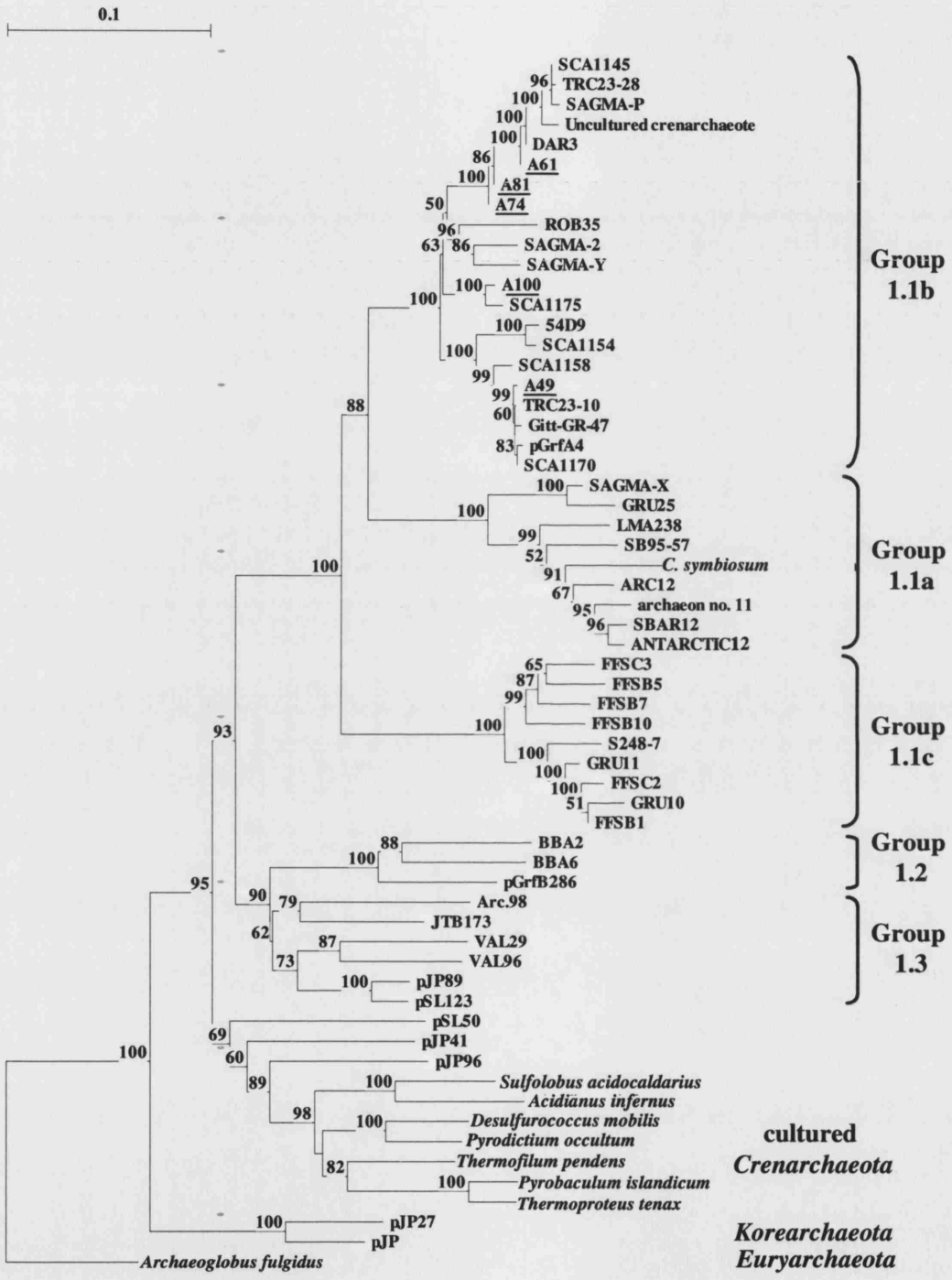


Figure 4.3: Phylogenetic position of Antarctic clone sequences within the domain Archaea. Antarctic clone sequences (shown underlined) are affiliated to Group 1.1b of the *Crenarchaeota*. Bootstrap values, given as a percentage of 100 replicate trees, are indicated for branches supported by more than 50% of the trees. Scale bar: 0.1, estimated number of substitutions per nucleotide position. Clones A84 and A88 were not included in the construction of this tree as they share the same relative as clones A74 and A81.

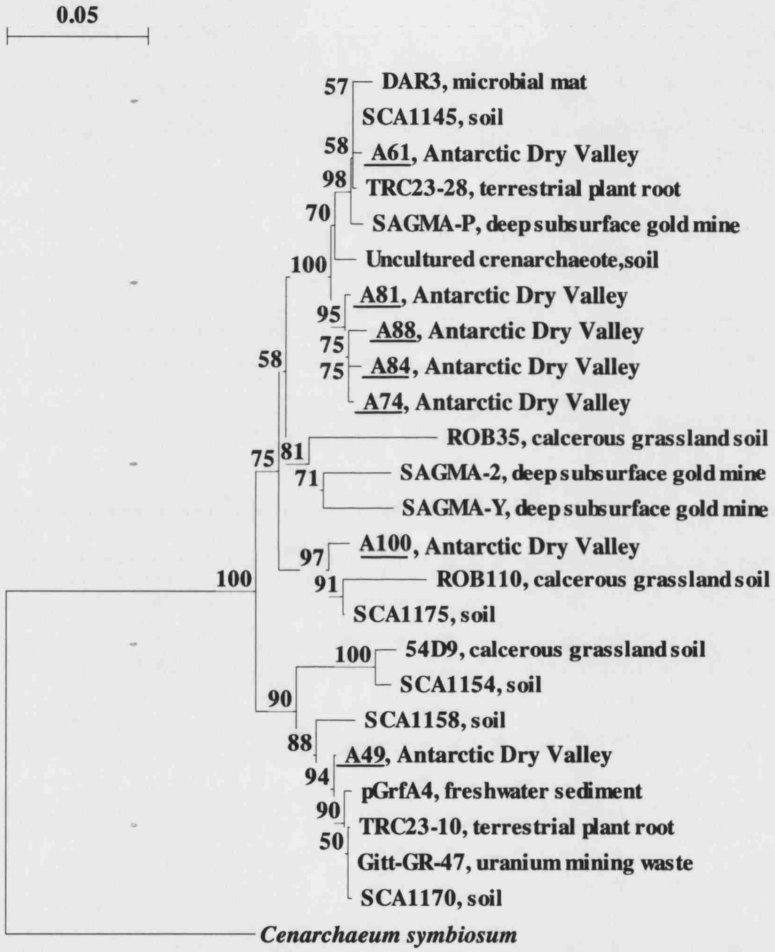


Figure 4.4: Phylogenetic relationship of Antarctic clone sequences to members of group 1.1b *Crenarchaeota*. Antarctic sequences are shown underlined. Bootstrap values, given as a percentage of 100 replicate trees, are indicated for branches supported by more than 50% of the trees. Scale bar: 0.05, estimated number of substitutions per nucleotide position.

4.2.3: Archaeal diversity in the Upper Wright Valley

4.2.3.1: Construction and analysis of 16S rDNA clone library

An archaeal 16S rDNA library was constructed from sample UWV4 using the *Archaea* specific primer pair of Ar4f and Ar958r, which amplify a region between nucleotide positions 4 and 967 of the 16S rRNA gene (*Escherichia coli* numbering). A total of 35 cycles of amplification was necessary for sample UWV4 in order to generate a visible product; the products of three replicate PCR reactions were pooled together prior to cloning. A total of 18 recombinant plasmids were analyzed by RFLP analysis using the tetrameric restriction enzyme *Msp*1, from which a single unique restriction enzyme pattern or operational taxonomic unit (OTU) was identified; two clones were selected for partial sequencing.

4.2.3.2: Phylogenetic affiliation of cloned sequences

Sequence analysis revealed clones Arch4-3 and Arch4-4 to share a pairwise identity of 99%. Their closest relative as identified by BLAST analysis was an uncultured archaeon found associated with metal-rich particles from a freshwater reservoir (Stein *et al.*, 2001); the level of 16S rRNA identity in each case was 99% over 518 bp.

The sequence of clone Arch4-4 was included in the alignment of Miers Valley archaeal clone sequences and representatives of Group 1.1b *Crenarchaeota* described earlier. Phylogenetic analysis confirms the affiliation of clone Arch4-4 to group 1.1b as shown in Figure 4.5. As is evident from the phylogenetic tree, this sequence is closely related to the group of Antarctic clone sequences recovered from the Miers Valley, comprising clones A74, A81, A84 and A88.

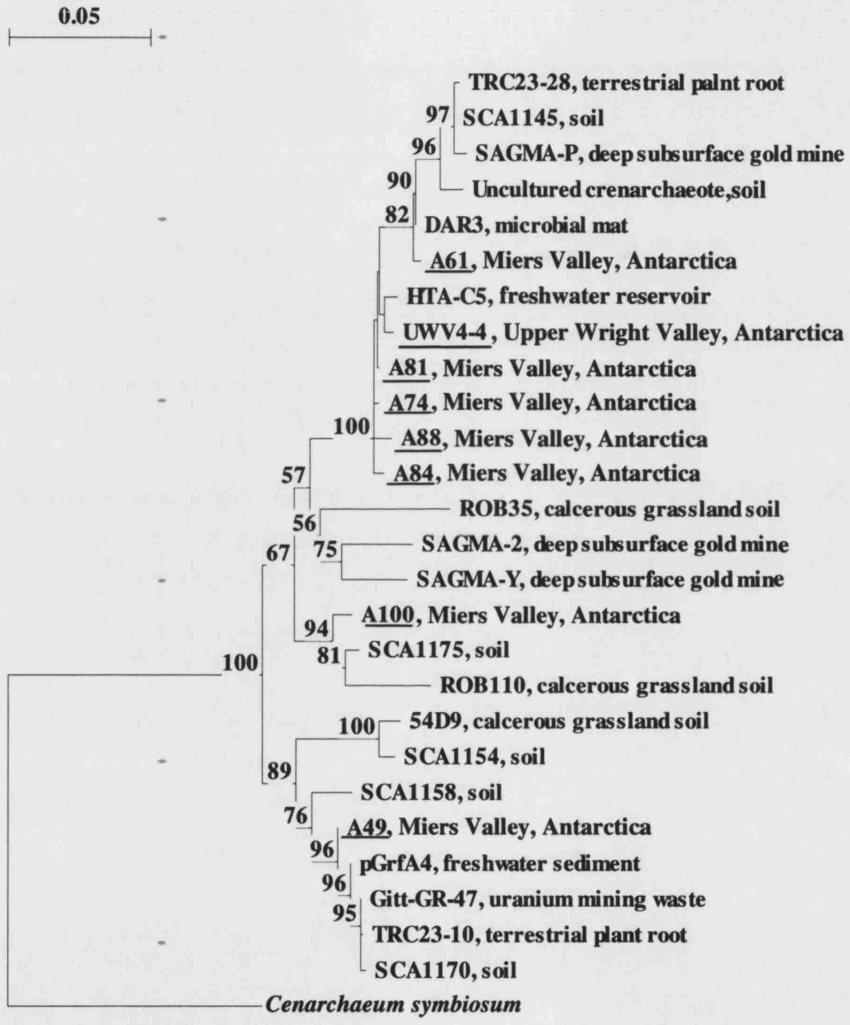


Figure 4.5: Phylogenetic relationship of archeal clone sequence UWV4-4 identified from the Upper Wright library, to Miers Valley archaeal sequences and members of group 1.1b *Crenarchaeota.* Antarctic sequences are shown underlined. Bootstrap values, given as a percentage of 100 replicate trees, are indicated for branches supported by more than 50% of the trees. Scale bar: 0.05, estimated number of substitutions per nucleotide position.

4.2.4: Validity of ARDRA

The maximum sequence diversity among Antarctic clones was 6% over approximately 1000 bp of the 16S rRNA gene, and Antarctic sequences were distributed throughout the terrestrial cluster. Three distinct phylotypes could be determined based upon <97% 16S rRNA identity among clone sequences and these are evident from the phylogenetic tree shown in Figure 4.6.

Phylotype C is represented by clone A49 belonging to ARDRA group 2. Partial sequence analysis of a further four clones from this ARDRA group (A9, A28, A29 and A44) revealed 98-99% nucleotide identity in a 535 bp overlap to clone A49. This result indicates the ARDRA analysis is consistent with the phylotype identified here, based upon sequence identity and also inferred by the phylogenetic tree.

Clone A100 of ARDRA group 5 represents phylotype B. A second clone from this ARDRA group, A64, exhibited 98% identity (over ~900 bp) to A100 and possessed the same closest database relative, clone SCA1175 (accession number U62819). However, a third clone, A84, proved an exception exhibiting 98% identity in a 1020 bp overlap to an uncultured crenarchaeote (accession number AJ496176), a database sequence associated with phylotype A. The most numerically dominant ARDRA group representing phylotype A was ARDRA group 1. The *Msp*1 restriction patterns of ARDRA groups 1 and 5 are identical, whilst the *Hin*P1 restriction patterns differ (Refer to Figure 4.1.b: Lanes 10 and 11 illustrate the *Hin*P1 restriction pattern of ARDRA groups 5 and 1 respectively; no difference is observed in the corresponding *Msp*1 restriction pattern of clones belonging to these two groups, as shown in lanes 10 and 11 of Figure 4.1.a.) This difference can be attributed to the absence of a single *Hin*P1 restriction site for clone sequences belonging to ARDRA group 5. Restriction of the ARDRA group 5 fragment of approximately 1 kb can result in two fragments of approximately 630 and 330 bp that are characteristic of the group 1 *Hin*P1 restriction pattern. As illustrated in Figure 4.7, it would appear that a single base change in the sequence of clone A84 at position 680 (position 1 corresponding to the first nucleotide of primer Arch21f) has resulted in the

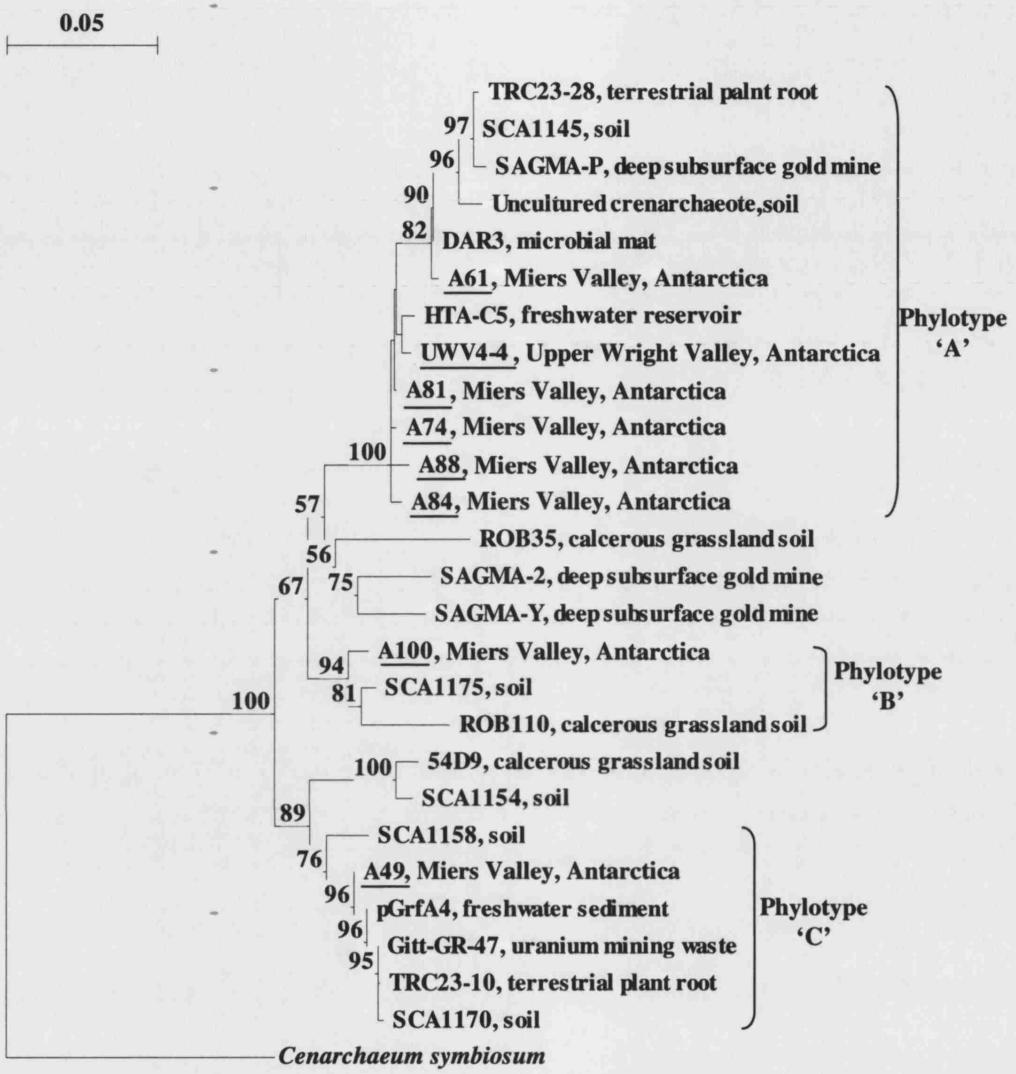


Figure 4.6: Representation of MV11.1 archaeal phylotypes. Three phylotypes are evident – designated A, B and C.

loss of the additional *Hin*P1 restriction site associated with ARDRA group 1, consequently giving rise to a group 5 restriction pattern. This would account for the anomaly in the phylogenetic positioning of clone A84 with respect to its ARDRA grouping.

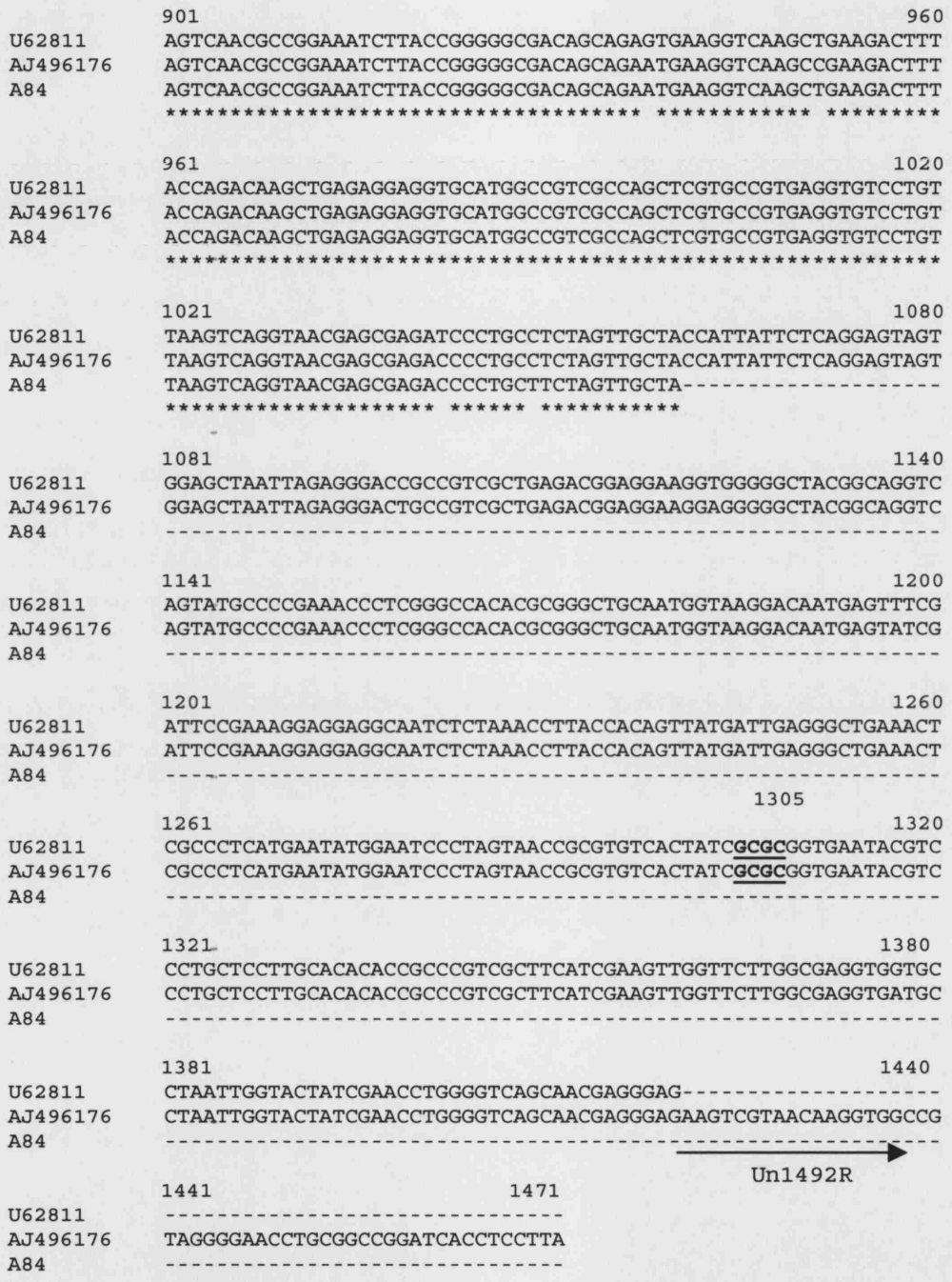
Clone A61 along with clones UWV4-4, A81, A74 and A88 represent phylotype A. A61 belongs to the most numerically dominant group, ARDRA group 1. Sequence analysis of two additional clones from this group, A33 and A79, revealed 98% identity over 680 bp to A61, inferring uniformity within the group. Clones A74, A81 and A88 formed a separate and distinct cluster from A61 within phylotype 1; all exhibited 98% identity to an uncultured crenarchaeote (accession number AJ496176) and shared 98% identity with A61. Interestingly, each of these clones are the sole representatives of ARDRA groups 8, 9 and 10 that were identified on the basis of possessing a single different restriction pattern from ARDRA group 1. Their relatedness to group 1, whilst being classified as separate ARDRA groups, is illustrated by the phylogenetic tree where they form a distinct cluster within phylotype A. Thus, it would appear that the ARDRA analysis is consistent with the phylogenetic positioning of archaeal clones belonging to phylotypes A.

In summary, the phylogenetic analysis shows phylotype B to be more closely related to phylotype A. This relationship is apparent from the ARDRA analysis as each phylotype is represented by ARDRA groups possessing identical *Msp*I restriction patterns (considering ARDRA group 1 as representative of phylotype A since it is the most numerically dominant) but different *Hin*PI patterns. In contrast, phylotype C is phylogenetically distinct from the above two and accordingly, the ARDRA analysis shows two unique restriction patterns for ARDRA group 2 representative of this phylotype. Thus, overall, ARDRA analysis of archaeal clone sequences is consistent with the phylotypes identified based upon nucleotide sequence identities and inferred from the phylogenetic tree.

Figure 4.7: *Hin*P1 restriction analysis of clone A84. This figure shows an alignment of clone A84 with its two closest database relative, an uncultured crenarchaeote (AJ496176), and an unidentified archaeon SCA1145 (U62811); all sequences cluster within phylotype A. ARDRA group 1, representing the most numerically dominant ARDRA group associated with phylotype A, possesses a *Hin*PI restriction pattern comprising four bands of approximately 630, 330, 300 and 130 bp in size (Figure 4.1.b. Lane 11). As is evident from the alignment, sequence AJ496176 (for which a complete 16S rDNA sequence is available) possesses this restriction pattern, giving rise to exact fragments of 627, 345, 323, 133 and 9 bp in size that are consistent with the restriction pattern shown in Lane 11. Clone A84 however, possesses a G-A substitution at position 680 eliminating the third *Hin*P1 restriction site and resulting in bands of 972 bp in size. This corresponds to the *Hin*P1 restriction pattern representing ARDRA group 5 shown in lane 10 of Figure 4.1.b. The dashed line represents sequence information not available and arrows denote the primer sequences.

```
           1                                                          60
U62811     ------------------GGACCCGACTGCTATCAGAGTGGGACTAAGCCATGCGAGTCA
AJ496176   AATCCGGTTGATCCTGCCGGACCCGACTGCTATCAGAGTGGGACTAAGCCATGCGAGTCA
A84        ------------------>-------------------GGACTAAGCCACGCGAGTCG
           A21f                                    ********** *******

           61                                                        120
U62811     ACATAGCAATATGTGGCATACGGCTCAGTAACACGTAGTCAACATGCCCAGGGGACGTGG
AJ496176   ACATAGCAATATGTGGCATACGGCTCAGTAACACGTAGTCAACATGCCCAGGGGACGTGG
A84        ACACAGCAATGTGTGGCATACGGCTCAGTAACACGTAGTCAACATGCCCAGGGGACGTGG
           *** ****** **************************************************

           121                                                       180
U62811     ATAACCTCGGGAAACTGAGGATAAACCGCGATAGGTCATCACTTCTGGAATGGGTAATGA
AJ496176   ATAACCTCGGGAAACTGAGGATAAACCGCGATAAGTCACTACTTCTGGAATGGGTAATGA
A84        ATAACCTCGGGAAACTGAGGATAAACCGCGATAGGTCACTACTTCTGGAATGGGTAATGA
           ********************************** ****  ********************

           181                                                       240
U62811     CTTAAATCTATATGGCCCCTGGATTGGACTGCGGCCGATCAGGCTGTTGGTGAGGTAATG
AJ496176   CTTAAATCTATATGGCCCCTGGATTGGACTGCGGCCGATCAGGCTGTTGGTGAGGTAATG
A84        CCCAAACCTACATGGCCCCTGGATTGGACTGCGGCCGATCAGGCTGTTGGTGAGGTAATG
           *  *** *** **************************************************

           241                                                       300
U62811     GCCCACCAAACCTGTAACCGGTACGGGCTCTGAGAGGAGGAGCCCGGAGATGGGCACTGA
AJ496176   GCCCACCAAACCTGTAACCGGTACGGGCTCTGAGAGGAGGAGCCCGGAGATGGGCACTGA
A84        GCCCACCAAACCTGTAACCGGTACGGGCTCTGACAGGAGGAGCCCGGAGATGGGCACTGA
           ********************************* ************************

                                    324       333
           301                                                       360
U62811     GACAAGGGCCCAGGCCCTATGGGGCGCAGCAGGCGCGAAACCTCTGCAATAGGCGAAAGC
AJ496176   GACAAGGGCCCAGGCCCTATGGGGCGCAGCAGGCGCGAAACCTCTGCAATAGGCGAAAGC
A84        GACAAGGGCCCAGGCCCTATGGGGCGCAGCAGGCGCGAAACCTCTGCAATAGGCGAAAGC
           ************************************************************
```

```
            361                                                       420
U62811      TTGACAGGGTTACTCTGAGTGATTTCCGTTAAGGAGATCTTTTGGCACCTCTAAAAATGG
AJ496176    CTGACAGGGTTACTCTGAGTGATTTCCGTTAAGGAGATCTTTTGGCACCTCTAAAAATGG
A84         TTGACAGGGTTACTCTGAGTGATTTCCGTTAAGGAGATCTTTTGGCACCTCTAAAAATGG
             ************************************************************

            421                                                       480
U62811      TGCAGAATAAGGGGTGGGCAAGTCTGGTGTCAGCCGCCGCGGTAATACCAGCACCCCGAG
AJ496176    TGCAGAATAAGGGGTGGGCAAGTCTGGTGTCAGCCGCCGCGGTAATACCAGCACCCCGAG
A84         TGCAGAATAAGGGGTGGGCAAGTCTGGTGTCAGCCGCCGCGGTAATACCAGCACCCCGAG
            ************************************************************

            481                                                       540
U62811      TGGTCGGGACGTTTATTGGGCCTAAAGCATCCGTAGCCGGTTCTACAAGTCTTCCGTTAA
AJ496176    TGGTCGGGACGTTTATTGGGCCTAAAGCATCCGTAGCCGGTTCTACAAGTCTTCCGTTAA
A84         TGGTCGGGACGTTTATTGGGCCTAAAGCATCCGTAGCCGGTTCTACAAGTCTTCCGTTAA
            ************************************************************

            541                                                       600
U62811      ATCCACCTGCTTAACAGATGGGCTGCGGAAGATACTATAGAGCTAGGAGGCGGGAGAGGC
AJ496176    ATCCACCTGCTTAACAGATGGGCTGCGGAAGATACTATGGAGCTAGGAGGCGGGAGAGGC
A84         ATCCACCTGCTTAACAGATGGGCTGCGGAAGATACTATAGAGCTAGGAGGCGGGAGAGGC
            ************************************* ********************

            601                                                       660
U62811      AAGCGGTACTCGATGGGTAGGGGTAAAATCCGTTGATCCATTGAAGACCACCAGTGGCGA
AJ496176    AAGCGGTACTCGATGGGTAGGGGTAAAATCCGTTGATCCATTGAAGACCACCAGTGGCGA
A84         AAGCGGTACTCGATGGGTAGGGGTAAAATCCGTTGATCCATTGAAGACCACCAGTGGCGA
            ************************************************************
                              678
            661                                                       720
U62811      AGGCGGCTTGCCAGAACGCGCTCGACGGTGAGGGATGAAAGCTGGGGGAGCAAACCGGAT
AJ496176    AGGCGGCTTGCCAGAACGCGCTCGACGGTGAGGGATGAAAGCTGGGGGAGCAAACCGGAT
A84         AGGCGGCTTGCCAGAACGCACTCGACGGTGAGGGATGAAAGCTGGGGGAGCAAACCGGAT
            ****************** ****************************************

            721                                                       780
U62811      TAGATACCCGGGTAGTCCCAGCTGTAAACGATGCAGACTCGGTGATGAGTTGGCTTATTG
AJ496176    TAGATACCCGGGTAGTCCCAGCTGTAAACGATGCAGACTCGGTGATGAATTGGCTTCATG
A84         TAGATACCCGGGTAGTCCCAGCTGTAAACGATGCAGACTCGGTGATGAGTTGGCTTCATG
            ************************************************* *******  **

            781                                                       840
U62811      CTAACTCAGTGCCGCAGGGAAGCCGTTAAGTTTGCCGCCTGGGGAGTACGGTCGCAAGAC
AJ496176    CCAATTCAGTGCCGCAGGGAAGCCGTTAAGTTTGCCGCCTGGGGAGTACGGTCGCAAGAC
A84         CCAACTCAGTGCCGCAGGGAAGCCGTTAAGTTTGCCGCCTGGGGAGTACGGTCGCAAGAC
            * ** ************************************************************

            841                                                       900
U62811      TGAAACTTAAAGGAATTGGCGGGGGAGCACCACAAGGGGTGAAGCCTGCGGTTCAATTGG
AJ496176    TGAAACTTAAAGGAATTGGCGGGGGAGCACCACAAGGGGTGAAGCCTGCGGTTCAATTGG
A84         TGAAACTTAAAGGAATTGGCGGGGGAGCACCTCAAGGGGTGAAGCCTGCGGTTCAATTGG
            ******************************** ****************************
```


901 960
U62811 AGTCAACGCCGGAAATCTTACCGGGGGCGACAGCAGAGTGAAGGTCAAGCTGAAGACTTT
AJ496176 AGTCAACGCCGGAAATCTTACCGGGGGCGACAGCAGAATGAAGGTCAAGCCGAAGACTTT
A84 AGTCAACGCCGGAAATCTTACCGGGGGCGACAGCAGAATGAAGGTCAAGCTGAAGACTTT
961 1020
U62811 ACCAGACAAGCTGAGAGGAGGTGCATGGCCGTCGCCAGCTCGTGCCGTGAGGTGTCCTGT
AJ496176 ACCAGACAAGCTGAGAGGAGGTGCATGGCCGTCGCCAGCTCGTGCCGTGAGGTGTCCTGT
A84 ACCAGACAAGCTGAGAGGAGGTGCATGGCCGTCGCCAGCTCGTGCCGTGAGGTGTCCTGT
1021 1080
U62811 TAAGTCAGGTAACGAGCGAGATCCCTGCCTCTAGTTGCTACCATTATTCTCAGGAGTAGT
AJ496176 TAAGTCAGGTAACGAGCGAGACCCCTGCCTCTAGTTGCTACCATTATTCTCAGGAGTAGT
A84 TAAGTCAGGTAACGAGCGAGACCCCTGCTTCTAGTTGCTA----------
1081 1140
U62811 GGAGCTAATTAGAGGGACCGCCGTCGCTGAGACGGAGGAAGGTGGGGGCTACGGCAGGTC
AJ496176 GGAGCTAATTAGAGGGACTGCCGTCGCTGAGACGGAGGAAGGAGGGGGCTACGGCAGGTC
A84 ----------
1141 1200
U62811 AGTATGCCCCGAAACCCTCGGGCCACACGCGGGCTGCAATGGTAAGGACAATGAGTTTCG
AJ496176 AGTATGCCCCGAAACCCTCGGGCCACACGCGGGCTGCAATGGTAAGGACAATGAGTATCG
A84 ----------
1201 1260
U62811 ATTCCGAAAGGAGGAGGCAATCTCTAAACCTTACCACAGTTATGATTGAGGGCTGAAACT
AJ496176 ATTCCGAAAGGAGGAGGCAATCTCTAAACCTTACCACAGTTATGATTGAGGGCTGAAACT
A84 ----------
1305
1261 1320
U62811 CGCCCTCATGAATATGGAATCCCTAGTAACCGCGTGTCACTATCGCGCGGTGAATACGTC
AJ496176 CGCCCTCATGAATATGGAATCCCTAGTAACCGCGTGTCACTATCGCGCGGTGAATACGTC
A84 ----------
1321 1380
U62811 CCTGCTCCTTGCACACACCGCCCGTCGCTTCATCGAAGTTGGTTCTTGGCGAGGTGGTGC
AJ496176 CCTGCTCCTTGCACACACCGCCCGTCGCTTCATCGAAGTTGGTTCTTGGCGAGGTGATGC
A84 ----------
1381 1440
U62811 CTAATTGGTACTATCGAACCTGGGGTCAGCAACGAGGGAG----------
AJ496176 CTAATTGGTACTATCGAACCTGGGGTCAGCAACGAGGGAGAAGTCGTAACAAGGTGGCCG
A84 ----------
Un1492R
1441 1471
U62811 ----------
AJ496176 TAGGGGAACCTGCGGCCGGATCACCTCCTTA
A84 ----------

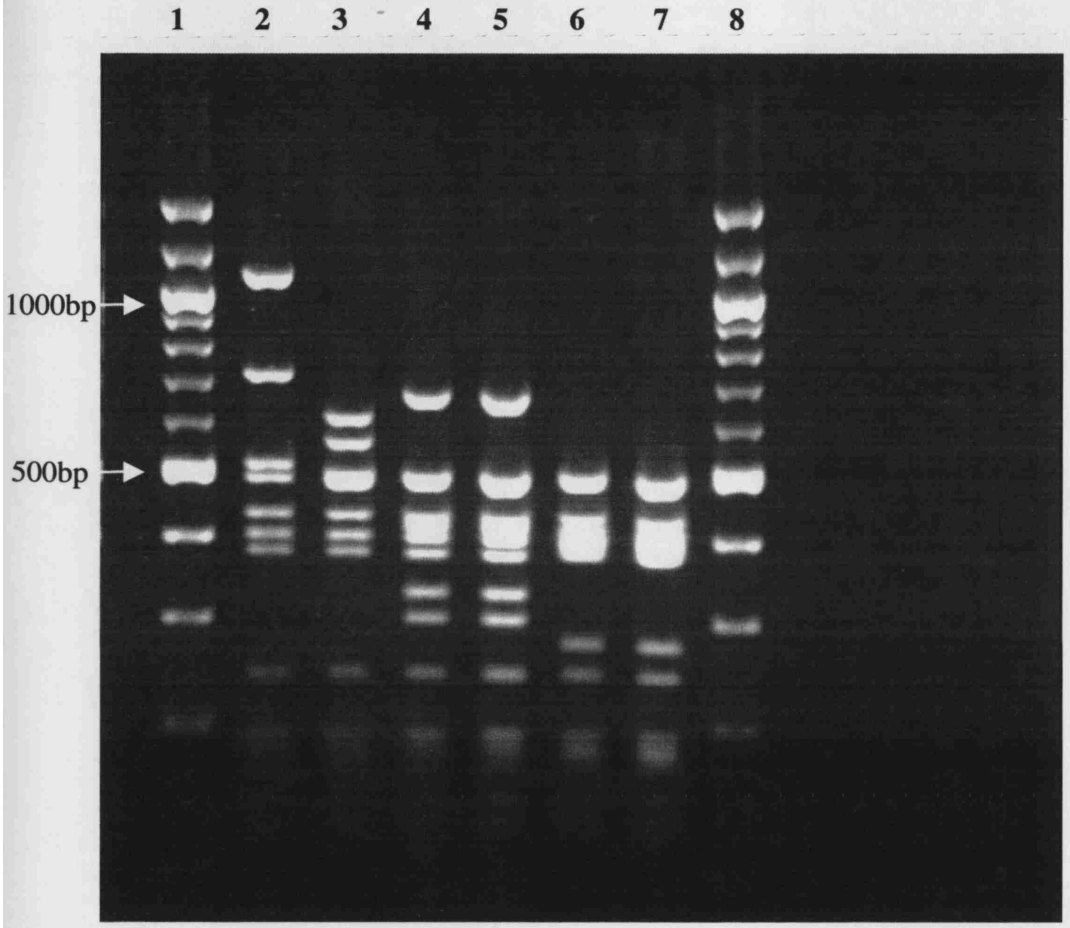
4.3: Diversity of Eukaryotes

4.3.1: Construction and analysis of a eukaryote 18S rRNA library

Eukaryote 18S rRNA gene sequences were amplified from environmental DNA using the primer set of EukA and EukB to yield a product of 1975bp in size. Three Miers Valley samples were selected for 18S rRNA gene amplification – MVT7 and MVT11 representing two of the transect samples from which significant amounts of DNA were extracted, and MV16, the sample used for the culture-based study of bacterial diversity. However, all amplifications from DNA extracted from these samples failed to yield a product after 35 cycles. Similarly, three Upper Wright Valley samples were selected: UWV1, UWV6 and UWV4, the latter being of particular interest since it was used for both a molecular and culture-based study of bacterial diversity. However, only DNA extracted from sample UWV6 yielded a PCR product, and so from this sample an 18S rRNA gene library was constructed. All DNA extractions which failed to produce an 18S PCR product were confirmed to be of suitable purity for PCR by successful amplification of bacterial 16S rRNA gene sequences.

A total of 22 clones from the UWV6 18S rRNA gene library were subject to RFLP analysis using enzyme *Msp*1. Clones were grouped manually based on the restriction patterns from which four phylotypes were identified. Figure 4.8 illustrates the restriction pattern obtained for a clone representing each phylotype. Phylotypes 2 and 3 were represented by 14 and 6 clones respectively; phylotypes 1 and 4 each comprised a single clone. The coverage value determined by RFLP analysis of the library was 0.91, inferring 91% of the total RFLP diversity within the library was determined experimentally. A clone representing each phylotype was selected for double-stranded sequencing using the primers EukA and 516R.

Figure 4.8: *Msp*1 restriction patterns identified for each eukaryal phylotype. Lanes 1 and 8 contain a 100 bp molecular weight marker. Lane 2: phylotype 1; lane 3: phylotype 4; lanes 4 and 5: phylotype 2; lanes 5 and 6: phylotype 3.



4.3.2: Phylogenetic analysis of eukaryotic clone sequences

Closest database relatives of eukaryotic clone sequences amplified from sample UWV6 were identified by BLAST analysis (Table 4.7). Clones UWV6-8 and UWV6-5 were found to exhibit 99% 18S rDNA identity to two previously cultivated species of fungi. The closest relative of UWV6-8 was identified as *Cryptococcus vishniacii*, a member of the fungal phylum *Basidiomycota*, whilst UWV6-5 was most closely related to *Taphrina padi*, an Ascomycote recovered from alder (unpublished). Clone UWV6-1 was found to exhibit only 92% 18S rRNA identity to its closest relative, an uncultured soil zygomycete present in bulk/ maize rhizosphere soil of the tropics (Gomes *et al.* 2003). The closest relative of clone UW6-12 with which it shared 97% 18S rRNA identity was *Eolimna subminuscula*, a naviculoid diatom belonging to the eukaroytic lineage, *Stramenopiles*.

Figure 4.9 illustrates the phylogenetic affiliation of Antarctic clone sequences and their closest relatives with respect to representatives of the major fungal phyla and the eukaryotic group, *Stramenopiles*. The phylogenetic tree was constructed from an alignment spanning approximately the first 500 bp of the 18S rRNA.

Table 4.7: Closest database relatives of eukaryotic clone sequences amplified from sample UWV6.

Phylotype	No. of Clones/ phylotype	Clone	Phylogenetic group	Nearest neighbour	Accession No. of nearest neighbour	% Identity (over no. of bp)
1	1	UWV6-1	Fungi; Zygomycota	Uncultured soil zygomycete	AJ515944	92 (455)
2	14	UWV6-8	Fungi; Basidiomycota	*Cryptococcus vishniacii*	AB032657	99 (469)
3	6	UWV6-5	Fungi; Ascomycota	*Taphrina padi*	AJ495833	99 (468)
4	1	UWV6-12	stramenopiles; Bacillariophyta	*Eolimna subminuscula*	AJ243064	97 (464)

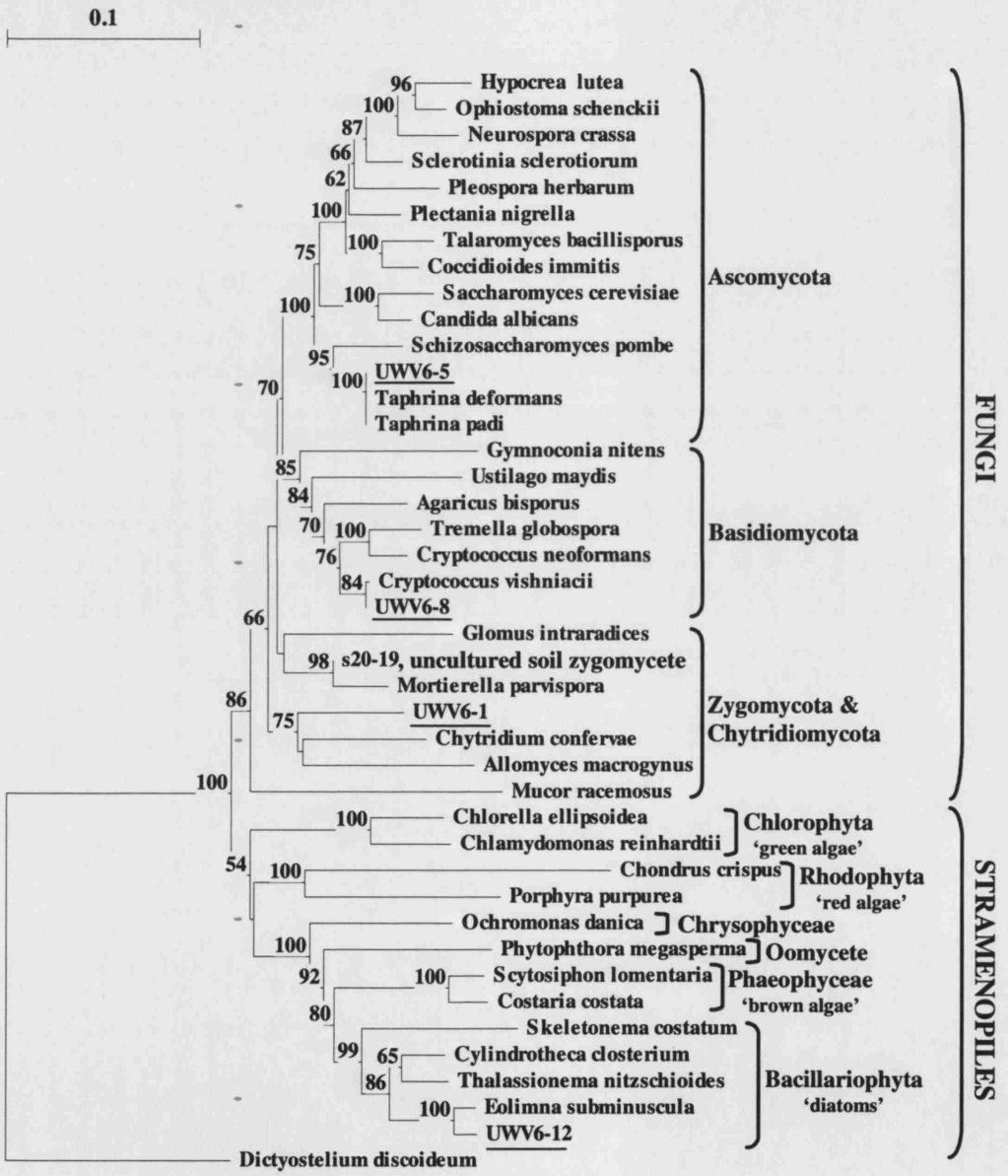


Figure 4.8: Phylogenetic affiliation of Antarctic clone sequence within the domain *Eukarya*. Antarctic sequences are shown underlined. Bootstrap values, given as a percentage of 100 replicate trees, are indicated for branches supported by more than 50% of the trees. Scale bar: 0.1, estimated number of substitutions per nucleotide position.

It is conceivable that the primer pair used in this study to amplify 18S rRNA gene sequences may have resulted in an underestimation of the diversity of eukaryotes in Dry Valley soil sample UWV6. The application of molecular methods to the investigation of eukaryotic diversity in natural environments has emerged only relatively recently (Moreira & Lopez-Garcia, 2002). Moon-van der Staay *et al.*, (2001) conducted one of the first 18S SSU rRNA-based analyses examining diversity of eukaryotes within picoplankton. The PCR primers used in this study were general eukaryotic–specific primers (versus class- or genus-specific primers) capable of recognising most protists. Their analysis of thirty-five full length 18S rRNA sequences revealed a high diversity of previously unknown picoeukaryotes belonging to important marine phyla including the Prasinophytes, Haptophytes, Dinoflagellates and Stramenopiles. At the same time Lopez Garcia *et al.*, (2001) identified an unexpectedly high diversity of eukaryotes from the analysis of 18S rRNA gene libraries constructed from oligotrophic waters of the Antarctic polar front. These two studies were subsequently followed by further publications highlighting a wealth of novel genetic diversity among marine eukaryotes (Diez *et al.*, 2001(a) and 2001(b); Lopez-Garcia *et al.*, 2003; Stoeck & Epstein, 2003; Stoeck *et al.*, 2003).

Due to the breadth of diversity observed among eukaryotic organisms (relative to prokaryotes) and the large number of novel sequences 18S rRNA sequences recently identified, the design of 18S primers of high coverage presents an formidable task. The 18S primers used in this work have previously be used by Diez et al., 2001(a) and 2001(b) and Stoeck et al., 2003(a) and 2003(b) and have been successful in amplifying sequences belonging to a range of eukaryotic taxa including the Prasinophytes, Stramenopiles, Diatoms, Alveolates, Fungi and Metazoa. However, it is feasible they may have under-sampled the diversity of eukaryotes in Dry Valley soil given the contrast of this environment to the marine environment. This may be particularly relevant to the diversity of fungal types in Dry Valley soil, where the use of fungal-specific SSU rRNA PCR primers might have identified increased fungal diversity.

4.4: Discussion of Results

Table 4.8 provides a summary of the analysis of the Miers Valley and Upper Wright Valley archaeal libraries. A direct comparison of the observed diversity of *Archaea* in Dry Valley soils from the Upper Wright Valley and the Miers Valley cannot be made due to differences in construction and analysis of the 16S rRNA clone libraries. Firstly, different PCR primers were used to amplify archaeal 16S rRNA gene sequences from both soils. Therefore, it cannot be assumed that all archaeal templates amplified with equal efficiency. Secondly, DNA template concentration was not standardized as this information was not obtained for sample MV11.1. A total of 35 cycles of amplification were required to generate a visible PCR product from sample UWV4 (in contrast to 25 cycles used for sample MV11.1). This strongly suggests that UWV4 extracts had a lower initial template concentration. PCR template concentration has been shown to affect the composition of clone libraries (von Wintzingerode *et al.* 1997) and similarly, an increase in the number of amplification cycles can result in reduced diversity within a clone library (Bonnet *et al.*, 1999).

Table 4.8: Summary of analysis of Miers Valley and Upper Wright Valley archaeal libraries.

	Miers Valley	Upper Wright Valley
No. of clones analysed	96	18
No. of different phylotypes	3	1
No. of phylotypes common to each library	1	1
Coverage of library	0.93	1.0

However, taken together these data provide strong evidence that non-thermophilic *Crenarchaeota* form part of the Antarctic Dry Valley microbial community of both regions investigated. Overall, a total of three distinct archaeal phylotypes (based upon >97% 16S rRNA identity) were identified, of which a single phylotype, phylotype A, was common to each of the libraries analysed. Phylotype A was the dominant phylotype within the Miers Valley library accounting for 85% of the clones analysed. Significantly, this was also the only phylotype identified within the Upper Wright library; this was despite differences in library construction and the analysis of only 18 clones (versus 96) by RFLP. Therefore, these results would indicate that sequences affiliated to phylotype A represent the dominant archaeal phylotype in Dry Valley soils of the two locations studied here.

Cultivation-independent studies have identified that non-thermophilic *Crenarchaeota*, designated Group 1, are an ecologically diverse group with members distributed in natural environments across the globe (reviewed by Delong, 1998 & 2003; Ochsenreiter *et al.*, 2003). Little is known regarding the physiology of the organisms as a representative of this group remains to be cultured *in vitro*. The Dry Valley archaeal clones identified in this study were affiliated specifically to Group 1.1b of the non-thermophilic *Crenarchaeota*, also known as the terrestrial cluster, a group dominated by soil clones. Abundance estimates indicate that these organisms form a stable part of soil microbial communities, and thus are likely to be of ecological significance (Buckley *et al.*, 1998; Ochsenreiter *et al.*, 2003). However, their ecological role in natural environments has yet to be elucidated.

Interestingly, of archaeal sequences affiliated to Group 1.1.b, no one particular phylotype had been found to predominate in soils across the globe. Sliwinski & Goodman (2004) examined the spatial heterogeneity of crenarchaeal phylotypes in mesophilic soil. 16S rRNA gene clone libraries were constructed from six different sampling locations and a total of 222 clones analysed by PCR-single stranded conformation polymorphism profiling (SSCP). Fourteen phylotypes were identified and sequencing and phlyogenetic analysis revealed all to be affiliated to creanarchaeotal

group 1.1b. PCR-SSCP of 30 independent soil samples collected from one location revealed different individual phylotypes to dominate undefined patches within a uniform agricultural field. Furthermore, the differences observed in SSCP profiles could not be attributed to a single specific variable (i.e. depth, soil type), leading the authors to conclude that a complex interaction of environmental and biological factors likely underlies the diversity of crenarchaeotes.

The identification of non-thermophilic *Crenarchaeota* in the terrestrial Antarctic environment has hitherto not been described. de la Torre *et al.* (2003) were unable to detect *Archaea* using archaeal-specific primers in cryptoendolithic communities of the McMurdo Dry Valleys, leading to the conclusion that *Archaea* were either absent or present at very low levels within this environment. Non-thermophilic *Crenarchaeota* affiliated to group 1.1a, the marine planktonic Archaea, have been detected in coastal waters of Antarctica and have also been identified in Antarctic sponges from the McMurdo Sound in the Ross Sea (Delong *et al.*, 1994; Murray *et al.*, 1998; Webster *et al.*, 2004).

The failure to obtain an 18S rRNA gene PCR product from five of the six samples investigated can be considered a reflection of the low numbers of eukaryotic organisms in Dry Valley soils (Vishniac, 1993). Four phylotypes were identified from the analysis of 22 18S rRNA genes amplified from DNA extracted directly from soil sample UWV6. These corresponded to three fungal species and a marine diatom. The last of these may have been deposited in Dry Valley soils through the movement of sea salt aerosols and may not represent an active member of the soil community. The strong katabatic winds that drive through the Dry Valleys represent a primary mechanism of dispersal for microorganisms (Gordon *et al.*, 2000). These winds provide a continuous input of particulate and organic matter that can be both of marine influence and also from the surrounding environment. Indeed, it would be interesting to test this hypothesis; one possible method might be to construct an 18S rRNA gene clone library from rRNA that would allow for the identification of metabolically active eukaryotes within the soil community.

Two of the fungal phylotypes exhibited 99% 18S rRNA identity to two previously cultivated fungal species, *Cryptococcus vishniacii* and *Taphrina padi*. *C. vishniacii* is a basidiomycetous yeast originally isolated from Antarctic Dry Valley soil (Vishniac & Baharaeen, 1982). Members of this genus, as well as being globally distributed (Renker *et al.*, 2004), are considered cosmopolitan to Antarctica and have been cultivated from a range of Antarctic soil samples (Vishniac, 1993). *Taphrina padi* of the Ascomycota is a fungal species isolated from alder; interestingly, members of this genus are pathogens of vascular plants. Clone UWV6-1 representing the third fungal phylotype shared 92% 18S rRNA identity to an uncultured soil zygomycete present in bulk/ maize rhizosphere soil of the tropics (Gomes *et al.* 2003). Additionally, it shared 91% identity to the cultivated species *Mortierella parvispora*, isolated from a liverwort on Adelaide Island of the Antarctic Peninsula (Hughes et al., 2003). A range of fungal species have previously been identified from Dry Valley soils; however, a direct comparison can not be made as many were characterized on the basis of morphology (Vishniac, 1993).

All fungi are chemooganotrophs (this group also lacks chemolithotrophic forms) and are ubiquitous in soils, where they play an important role in the remineralization of organic matter. Thus, fungal species present in Dry Valley soils may perform a critical function within this ecosystem, in ensuring the recycling of carbon and nutrients where these resources are of a premium. Whilst this survey was by no means exhaustive, it is perhaps interesting that fungal species present in Dry Valley cryptoendolithic communities were not detected in Dry Valley soil; similarly, no species of alga were identified. The eukaryotic community present within Dry Valley soils may bear little similarity to that inhabiting the cryptoendolithic environment. This is in contrast to observations made regarding the molecular assessment of bacterial diversity in Dry Valley soils, and would be interesting to investigate further.

Chapter 5

Integron Diversity in Antarctic Environments

5.1: Introduction

The aim of the work described in this chapter was to investigate the role of integrons in promoting gene transfer and genome diversity in a variety of Antarctic soils. Integrons have contributed significantly to the acquisition and dissemination of antibiotic resistance genes among phylogenetically diverse Gram-negative bacteria, presenting a substantial challenge to the treatment of infectious disease (Canton *et al.*, 2003). Studies of integron carriage among resistant clinical strains have demonstrated frequencies of 53% (Leverstein-van Hall *et al.*, 2002) and 75% (Levesque *et al.*, 1995) for members of the *Enterobacteriaceae* and pseudomonads. Similarly, integrons have been identified in bacterial populations from environmental samples; 3.6% of 3000 Gram-negative isolates cultivated from an estuarine environment and 12% of 95 multiresistant plasmids isolated from activated sludge were found to contain a class 1 integron. Considerable genetic diversity among integrase genes and gene cassettes has also recently been identified in environmental samples (Nield et al., 2001; Holmes et al., 2003; Nemergut et al., 2004), providing evidence that integrons function as a general gene-capture system for bacteria. Thus, the objective of this work was to examine integron diversity within microbial communities present in Antarctic soils. Three distinct regions of the integron platform were selected for investigation: 1) the integrase gene, 2) integrated gene cassettes, and 3) the 3' conserved segment (3'CS) associated with Class 1 integrons. The structure of a class 1 integron is represented in Figure 5.1.

My hypothesis was that integrons would indeed be present in Antarctic soil samples and that they would possess the same basic structure as those recovered from other environmental and clinical samples. This structure would comprise of an integrase gene, a promoter for expression of gene cassettes, a variable region containing integrated cassettes, and a degree of sequence conservation within the 3'CS. However, whilst the

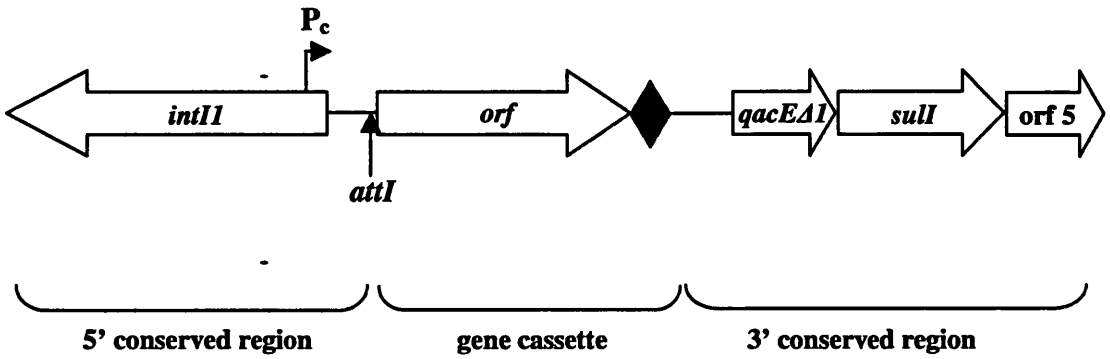


Figure 5.1: Structure of a class 1 integron containing a single gene cassette. Open arrows show direction of transcription of genes: *intI1* – integrase, *orf* - inserted gene cassette, *qacEΔ1* – encoding resistance to ethidium bromide and quaternary ammonium compounds, *sul1* – sulphonamide resistance gene, orf 5 – open reading frame of unknown function. The promoter (P_c) for expression of gene cassettes and the cassette receptor site (*attI*) are shown. The solid black diamond represents the 59-base element.

recovery of a variety of gene cassettes was expected, differences in gene cassette diversity were predicted for Antarctic soils. This would reflect the absence of selective pressures, such as exposure to synthetic antibiotics and pollutants and anthropogenic disturbance on the soil microbial communities; such selective factors are common to the majority of integron diversity studies to date. The Antarctic gene cassette pool was predicted to include cassettes similar to those previously characterized, along with cassettes encoding both novel orfs and/ or gene products conferring a selective advantage specific to the Antarctic environment i.e. UV protection. The isolation of integrons containing a multi-array of antibiotic resistance gene cassettes which have been identified in multi-drug resistant bacteria was considered unlikely.

A PCR-based strategy was used to recover both the 5' conserved segment (5'CS) of integrons that includes the integrase gene and attI site (approximately 1.4 kb in length), along with the variable region into which gene cassettes are inserted (Hall et al., 1991). Samples were also investigated for the presence of the qacE/ qacEΔ1 gene conferring resistance to ethidium bromide and quaternary ammonium compounds and the sulI gene, encoding resistance to sulphonamides. The 3'CS of class 1 integrons contains either a qacE gene, or typically in the case of multi-resistant integrons, the following

combination of three-genes: qacEΔ1, sulI, and an open reading frame of unknown function, orf5 (approximately 2 kb in length) (Paulsen et al., 1993). Additionally, attempts were made to identify the genomic location of gene cassettes based on analysis of sequences recovered.

DNA extracted from a variety of Antarctic soil samples collected from the Miers Valley (MV), the Upper Wright Valley (UWV) and Bratina Island (Brat) was investigated for the presence of integron sequences. Samples from the former two locations in the McMurdo Dry Valleys represent 'pristine' Antarctic soils, which have received minimal impact from human activities due to their geographical remoteness and harsh climatic conditions. Samples collected from Bratina Island were selected to provide a contrast to these pristine samples. Bratina Island is heavily impacted by human activities and received a number of field parties throughout the 2001/2002 season, of which the permanent base camp was estimated to be occupied for a period of 50 – 80 man-days.

Plasmid pSa of the incompatability group (Inc) W containing In6 was used as a positive PCR control (Valentine et al., 1994), along with two integron-positive environmental DNA samples, which included DNA extracted from grassland pasture soil (soil 1A) and bed dust. These samples were used to assess the efficacy of different primer combinations for the amplification of mixed template DNA.These samples were used to assess the efficacy of different primer combinations for the amplification of mixed template DNA.

5.2: Recovery of 5' Conserved Segment

As identified above, integrons have both common and variable features. This section of work describes attempts to recover the 5' conserved segment of integrons containing the integrase gene that catalyses the insertion/ excision of gene cassettes, and which also includes the promoter for expression of integrated gene cassettes

5.2.1: Recovery of integrase genes

Primer set 12b2 and 5CS-Rev were used to amplify a partial fragment of 911 bp of the integrase gene. All primers used in this work and their relative position with respect to the structure of the integron platform are represented in Figure 5.2. Primer 12b2 targets a conserved region within box 2 of class 1 and 2 integrases, whilst primer 5CS-Rev is specific to a region of the 5'CS in a position equivalent to 13 amino acids from the start codon of the integrase gene. PCR products of the expected size were clearly visible by gel electrophoresis following 35 cycles of amplification for the positive control pSa and from DNA extracted from bed dust. DNA extracted from multiple (2-4) Antarctic samples was pooled together and used as template for integrase PCR – samples pooled included Upper Wright samples 1-4 and 5-8 and Miers Valley samples 8 and 12. This permitted the quick and effective screening of multiple Antarctic samples according to geographical region, and served to compensate for low DNA yields in some samples. Following one round of amplification (35 cycles) a faint band of the expected size was visible for the pooled Antarctic samples and soil 1A on loading of 40µl (80%) of the PCR product. These products were purified from the gel and used as template for a second round of amplification in order to increase product yield, following which they were cloned into pCR 2.1 TOPO.

Integrase gene products amplified from pSa, bed dust, UWV1-4, UWV5-8, MV8/12 and soil 1A were sequenced using both vector specific primers (M13 f and r) and integrase sequencing primers IntSeq1 and IntSeq2. Sequence analysis (excluding primer sequences) revealed the cloned inserts to be virtually identical, with 99% sequence identity shared by pSa, BD (bed dust), soil 1a, UWV5-8, UWV1-4, and MV8/12 over 876 bp as shown in Figure 5.3. BLAST analysis confirmed the amplification of integrase gene products with cloned sequences exhibiting 99-100% identity to *intI1* and the corresponding region of the 5' conserved segment of a number of class 1 integrons shown in Table 5.1. The UWV4 sequence contained a total of four nucleotide substitutions, whilst two each were identified for UWV5-8 and soil 1A sequences and a single substitution for sequences from MV8/12 and bed dust. The sequence obtained from the positive control, pSa, was identical to the database relatives identified. Interestingly, the sequence retrieved from samples UWV1-4 contained a single A-G

nucleotide substitution in the start codon of *intI1* that would most likely prevent expression of the gene (Figure 5.3). In this case, due to the lack of a functional integrase

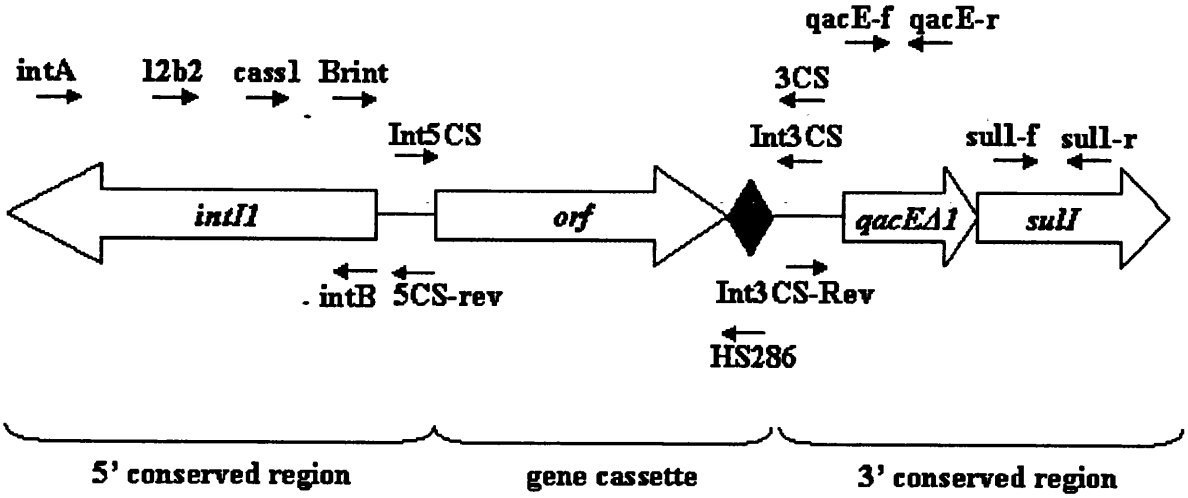


Figure 5.2: Structure of a class 1 integron containing a single gene cassette. Open arrows show direction of transcription of genes: *intI1* – integrase, orf - inserted gene cassette, *qacEΔ1* – encoding resistance to ethidium bromide and quaternary ammonium compounds, *sul1* – sulphonamide resistance gene. Solid arrows illustrate relative primer positions and direction of elongation during PCR amplification. The solid black diamond represent the 59-base element. Below are listed the primer combinations used in this study to amplify different regions of the integron, along with the expected product sizes (where applicable).

Target	Primer Pair	Expected size of product (bp)
5' conserved segment	12b2/ 5CS-Rev	911
	intA/ 5CS-Rev	978
	Brint/ 5CS-Rev	117
	intA/ intB	893
	Cass1/ 5CS-Rev	498
Variable region (gene cassette)	Brint/ 3CS	Size dependent on integrated cassettes
	Int5CS/ Int3CS	
	Brint/ HS286	
	Int5CS/ HS286	
3' conserved segment	qacE-f/ qacE-r	241
	sul1f/ sul1r	408
	qacE-f/ sul1r	902
	Int3CS-Rev/ qacE-r	373
	Int3CS-Rev/ sul1-r	1034

Figure 5.3: Alignment of 12b2 / 5CS-Rev nucleotide sequences amplified from pSa, Bed Dust (BD), soil sample 1A and pooled Antarctic samples UWV1-4, UWV5-8 and MV8/12. Accession number U37105 represents a closest database relative identified – the class 1 integron present on Tn*1404* of the R338/ R151 hybrid plasmid isolated from *P. aeruginosa*. Arrows denote the primer sequences. Sequence UWV4 has a nucleotide substitution A-G at position 854 within the start codon of the integrase gene.

```
          12b2 primer
          ──────────►
          1                                                          60
U37105    CGTCGCGAACGAGTGGCGGAGGGTGTGCGGTGTGGCGGGCTTCGTGATGCCTGCTTGTTC
pSa       CGTCGCGAACGAGTGGCGGAGGGTGTGCGGTGTGGCGGGCTTCGTGATGCCTGCTTGTTC
BD        CGTCGCAAACGAGTGGCGGAGGGTGTGCGGTGTGGCGGGCTTCGTGATGCCTGCTTGTTC
Soil 1A   CGTCGCGAACGAGTGGCGGAGGGTGTGCGGTGTGGCGGGCTTCGTGATGCCTGCTTGTTC
UWV5-8    CGTAGCGAACGAGTGGCGGAGGGTGTGCGGTGTGGCGGGCTTCGTGATGCCTGCTTGTTC
MV8/12    CGTCGCGAACGAGTGGCGGAGGGTGTGCGGTGTGGCGGGCTTCGTGATGCCTGCTTGTTC
UWV1-4    CGTCGCGAACGAGTGACGGAGGGTGTGCGGTGTGGCGGGCTCCGTGATGCCTGCTTGTTC
          *** ** ******** ************************ ******************

          61                                                         120
U37105    TACGGCACGTTTGAAGGCGCGCTGAAAGGTCTGGTCATACATGTGATGGCGACGCACGAC
pSa       TACGGCACGTTTGAAGGCGCGCTGAAAGGTCTGGTCATACATGTGATGGCGACGCACGAC
BD        TACGGCACGTTTGAAGGCGCGCTGAAAGGTCTGGTCATACATGTGATGGCGACGCACGAC
Soil 1A   TACGGCACGTTTGAAGGCGCGCTGAAAGGTCTGGTCATACATGTGATGGCGACGCACGAC
UWV5-8    TACGGCACGTTTGAAGGCGCGCTGAAAGGTCTGGTCATACATGTGATGGCGACGCACGAC
MV8/12    TACGGCACGTTTGAAGGCGCGCTGAAAGGTCTGGTCATACATGTGATGGCGACGCACGAC
UWV1-4    TACGGCACGTTTGAAGGCGCGCTGAAAGGTCTGGTCATACATGTGATGGCGACGCACGAC
          ************************************************************

          121                                                        180
U37105    ACCGCTCCGTGGATCGGTCGAATGCGTGTGCTGCGCAAAAACCCAGAACCACGGCCAGGA
pSa       ACCGCTCCGTGGATCGGTCGAATGCGTGTGCTGCGCAAAAACCCAGAACCACGGCCAGGA
BD        ACCGCTCCGTGGATCGGTCGAATGCGTGTGCTGCGCAAAAACCCAGAACCACGGCCAGGG
Soil 1A   ACCGCTCCGTGGATCGGTCGAATGCGTGTGCTGCGCAAAAACCCAGAACCACGGCCAGGA
UWV5-8    ACCGCTCCGTGGATCGGTCGAATGCGTGTGCTGCGCAAAAACCCAGAACCACGGCCAGGA
MV8/12    ACCGCTCCGTGGATCGGTCGAATGCGTGTACTGCGCAAAAACCCAGAACCACGGCCAGGA
UWV1-4    ACCGCTCCGTGGATCGGTCGAATGCGTGTGCTGCGCAAAAACCCAGAACCACGGCCAGGA
          ***************************** ****************************

          181                                                        240
U37105    ATGCCCGGCGCGCGGATACTTCCGCTCAAGGGCGTCGGGAAGCGCAACGCCGCTGCGGCC
pSa       ATGCCCGGCGCGCGGATACTTCCGCTCAAGGGCGTCGGGAAGCGCAACGCCGCTGCGGCC
BD        ATGCCCGGCGCGCGGATACTTCCGCTCAAGGGCGTCGGGAAGCGCAACGCCGCTGCGGCC
Soil 1A   ATGCCCGGCGCGCGGATACTTCCGCTCAAGGGCGTCGGGAAGCGCAACGCCGCTGCGGCC
UWV5-8    ATGCCCGGCGCGCGGATACTTCCGCTCAAGGGCGTCGGGAAGCGCAACGCCGCTGCGGCC
MV8/12    ATGCCCGGCGCGCGGATACTTCCGCTCAAGGGCGTCGGGAAGCGCAACGCCGCTGCGGCC
UWV1-4    ATGCCCGGCGCGCGGATACTTCCGCTCAAGGGCGTCGGGAAGCGCAACGCCGCTGCGGCC
          ************************************************************
```

```
241                                                                 300
U37105  CTCGGCCTGGTCCTTCAGCCACCATGCCCGTGCACGCGACAGCTGCTCGCGCAGGCTGGG
pSa     CTCGGCCTGGTCCTTCAGCCACCATGCCCGTGCACGCGACAGCTGCTCGCGCAGGCTGGG
BD      CTCGGCCTGGTCCTTCAGCCACCATGCCCGTGCACGCGACAGCTGCTCGCGCAGGCTGGG
Soil 1A CTCGGCCTGGTCCTTCAGCCACCATGCCCGCGCACGCGACAGCTGCTCGCGCAGGCTGGG
UWV5-8  CTCGGCCTGGTCCTTCAGCCACCATGCCCGTGCACGCGACAGCTGCTCGCGCAGGCTGGG
MV8/12  CTCGGCCTGGTCCTTCAGCCACCATGCCCGTGCACGCGACAGCTGCTCGCGCAGGCTGGG
UWV1-4  CTCGGCCTGGTCCTTCAGCCACCATGCCCGTGCACGCGACAGCTGCTCGCGCAGGCTGGG
        ****************************** ******************************

        301                                                         360
U37105  TGCCAAGCTCTCGGGTAACATCAAGGCCCGATCCTTGGAGCCCTTGCCCTCCCGCACGAT
pSa     TGCCAAGCTCTCGGGTAACATCAAGGCCCGATCCTTGGAGCCCTTGCCCTCCCGCACGAT
BD      TGCCAAGCTCTCGGGTAACATCAAGGCCCGATCCTTGGAGCCCTTGCCCTCCCGCACGAT
Soil 1A TGCCAAGCTCTCGGGTAACATCAAGGCCCGATCCTTGGAGCCCTTGCCCTCCCGCACGAT
UWV5-8  TGCCAAGCTCTCGGGTAACATCAAGGCCCGATCCTTGGAGCCCTTGCCCTCCCGCACGAT
MV8/12  TGCCAAGCTCTCGGGTAACATCAAGGCCCGATCCTTGGAGCCCTTGCCCTCCCGCACGAT
UWV1-4  TGCCAAGCTCTCGGGTAACATCAAGGCCCGATCCTTGGAGCCCTTGCCCTCCCGCACGAT
        ************************************************************

        361                                                         420
U37105  GATCGTGCCGTGATCGAAATCCAGATCCTTGACCCGCAGTTGCAAACCCTCACTGATCCG
pSa     GATCGTGCCGTGATCGAAATCCAGATCCTTGACCCGCAGTTGCAAACCCTCACTGATCCG
BD      GATCGTGCCGTGATCGAAATCCAGATCCTTGACCCGCAGTTGCAAACCCTCACTGATCCG
Soil 1A GATCGTGCCGTGATCGAAATCCAGATCCTTGACCCGCAGTTGCAAACCCTCACTGATCCG
UWV5-8  GATCGTGCCGTGATCGAAATCCAGATCCTTGACCCGCAGTTGCAAACCCTCACTGATCCG
MV8/12  GATCGTGCCGTGATCGAAATCCAGATCCTTGACCCGCAGTTGCAAACCCTCACTGATCCG
UWV1-4  GATCGTGCCGTGATCGAAATCCAGATCCTTGACCCGCAGTTGCAAACCCTCACTGATCCG
        ************************************************************

        421                                                         480
U37105  CATGCCCGTTCCATACAGAAGCTGGGCGAACAAACGATGCTCGCCTTCCAGAAAACCGAG
pSa     CATGCCCGTTCCATACAGAAGCTGGGCGAACAAACGATGCTCGCCTTCCAGAAAACCGAG
BD      CATGCCCGTTCCATACAGAAGCTGGGCGAACAAACGATGCTCGCCTTCCAGAAAACCGAG
Soil 1A CATGCCCGTTCCATACAGAAGCTGGGCGAACAAACGATGCTCGCCTTCCAGAAAACCGAG
UWV5-8  CATGCCCGTTCCATACAGAAGCTGGGCGAGCAAACGATGCTCGCCTTCCAGAAAACCGAG
MV8/12  CATGCCCGTTCCATACAGAAGCTGGGCGAACAAACGATGCTCGCCTTCCAGAAAACCGAG
UWV1-4  CATGCCCGTTCCATACAGAAGCTGGGCGAACAAACGATGCTCGCCTTCCAGAAAACCGAG
        ****************************** ******************************

        481                                                         540
U37105  GATGCGAACCACTTCATCCGGGGTCAGCACCACCGGCAAGCGCCGCGACGGCCGAGGTCT
pSa     GATGCGAACCACTTCATCCGGGGTCAGCACCACCGGCAAGCGCCGCGACGGCCGAGGTCT
BD      GATGCGAACCACTTCATCCGGGGTCAGCACCACCGGCAAGCGCCGCGACGGCCGAGGTCT
Soil 1A GATGCGAACCACTTCATCCGGGGTCAGCACCACCGGCAAGCGCCGCGACGGCCGAGGTCT
UWV5-8  GATGCGAACCACTTCATCCGGGGTCAGCACCACCGGCAAGCGCCGCGACGGCCGAGGTCT
MV8/12  GATGCGAACCACTTCATCCGGGGTCAGCACCACCGGCAAGCGCCGCGACGGCCGAGGTCT
UWV1-4  GATGCGAACCACTTCATCCGGGGTCAGCACCACCGGCAAGCGCCGCGACGGCCGAGGTCT
        ************************************************************
```

```
541                                                                 600
U37105  TCCGATCTCCTGAAGCCAGGGCAGATCCGTGCACAGCACCTTGCCGTAGAAGAACAGCAA
pSa     TCCGATCTCCTGAAGCCAGGGCAGATCCGTGCACAGCACCTTGCCGTAGAAGAACAGCAA
BD      TCCGATCTCCTGAAGCCAGGGCAGATCCGTGCACAGCACCTTGCCGTAGAAGAACAGCAA
Soil 1A TCCGATCTCCTGAAGCCAGGGCAGATCCGTGCACAGCACCTTGCCGTAGAAGAACAGCAA
UWV5-8  TCCGATCTCCTGAAGCCAGGGCAGATCCGTGCACAGCACCTTGCCGTAGAAGAACAGCAA
MV8/12  TCCGATCTCCTGAAGCCAGGGCAGATCCGTGCACAGCACCTTGCCGTAGAAGAACAGCAA
UWV1-4  TCCGATCTCCTGAAGCCAGGGCAGATCCGTACACAGCACCTTGCCGTAGAAGAACAGCAA
        ****************************** ******************************

        601                                                         660
U73105  GGCCGCCAATGCCTGACGATGCGTGGAGACCGAAACCTTGCGCTCGTTCGCCAGCCAGGA
pSa     GGCCGCCAATGCCTGACGATGCGTGGAGACCGAAACCTTGCGCTCGTTCGCCAGCCAGGA
BD      GGCCGCCAATGCCTGACGATGCGTGGAGACCGAAACCTTGCGCTCGTTCGCCAGCCAGGA
Soil 1A GGCCGCCAATGCCTGACGATGCGTGGAGACCGAAACCTTGCGCTCGTTCGCCAGCCAGGA
UWV5-8  GGCCGCCAATGCCTGACGATGCGTGGAGACCGAAACTTTGCGCTCGTTCGCCAGCCAGGA
MV8/12  GGCCGCCAATGCCTGACGATGCGTGGAGACCGAAACCTTGCGCTCGTTCGCCAGCCAGGA
UWV1-4  GGCCGCCAATGCCTGACGATGCGTGGAGACCGAAACCTTGCGCTCGTTCGCCAGCCAGGA
        ********************************** ************************

        661                                                         720
U73105  CAGAAATGCCTCGACTTCGCTGCTGCCCAAGGTTGCCGGGTGACGCACACCGTGGAAACG
pSa     CAGAAATGCCTCGACTTCGCTGCTGCCCAAGGTTGCCGGGTGACGCACACCGTGGAAACG
BD      CAGAAATGCCTCGACTTCGCTGCTGCCCAAGGTTGCCGGGTGACGCACACCGTGGAAACG
Soil 1A CAGAAATGCCTCGACTTCGCTGCTGCCCAAGGTTGCCGGGTGACGCACACCGTGGAAACG
UWV5-8  CAGAAATGCCTCGACTTCGCTGCTGCCCAAGGTTGCCGGGTGACGCACACCGTGGAAACG
MV8/12  CAGAAATGCCTCGACTTCGCTGCTGCCCAAGGTTGCCGGGTGACGCACACCGTGGAAACG
UWV1-4  CAGAAATGCCTCGACTTCGCTGCTGCCCAAGGCTGCCGGGTGACGCACACCGTGGAAACG
        ******************************** ****************************

        721                                                         780
U73105  GATGAAGGCACGAACCCAGTTGACATAAGCCTGTTCGGTTCGTAAACTGTAATGCAAGTA
pSa     GATGAAGGCACGAACCCAGTTGACATAAGCCTGTTCGGTTCGTAAACTGTAATGCAAGTA
BD      GATGAAGGCACGAACCCAGTTGACATAAGCCTGTTCGGTTCGTAAACTGTAATGCAAGTA
Soil 1A GATGAAGGCACGAACCCAGTTGACATAAGCCTGTTCGGTTCGTAAACTGTAATGCAAGTA
UWV5-8  GATGAAGGCACGAACCCAGTTGACATAAGCCTGTTCGGTTCGTAAACTGTAATGCAAGTA
MV8/12  GATGAAGGCACGAACCCAGTTGACATAAGCCTGTTCGGTTCGTAAACTGTAATGCAAGTA
UWV1-4  GATGAAGGCACGAACCCAGTTGACATAAGCCTGTTCGGTTCGTAAACTGTAATGCAAGTA
        ************************************************************

        781                                                         840
U73105  GCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGC
pSa     GCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGC
BD      GCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGC
Soil 1A GCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCAGTGGTAACGGCGC
UWV5-8  GCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGC
MV8/12  GCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGC
UWV1-4  GCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGC
        ********************************************** **************
```

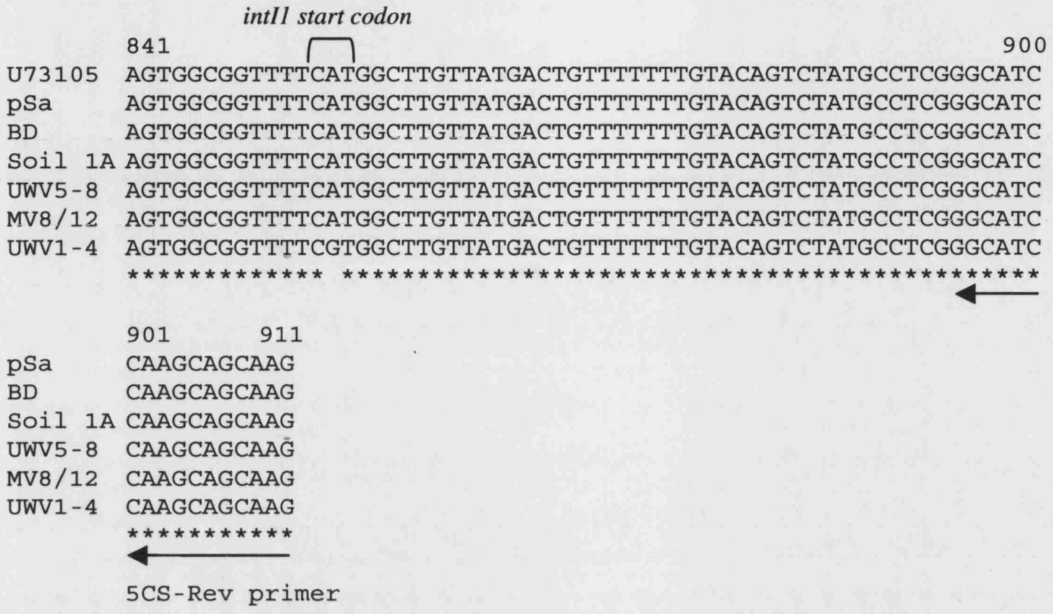


Table 5.1: Closest database relatives identified for 12b2/ 5CS-Rev sequences obtained from Antarctic samples and soil 1A, bed dust and pSa. All clone sequences were most closely related to the same eleven database relatives. Details are given (where available) of the plasmid harboring the *intI1* gene, the name of the integron (In) and any additional mobile elements (ME) associated with the integron (i.e. IS- insertion sequences; Tn – transposon).

Accession No.	Plasmid	Additional ME	In	Host Organism
AF550679	p1658/97	IS*26*		*Escherichia coli*
AY339625	p541	IS*26*	In-e541	*Escherichia coli*
U37105	R388-R151	Tn*1404*		*Pseudomonas aeruginosa*
U12441	R388		In3	*Escherichia coli*
X04555	pDGO100			
AJ311891	pS21		InS21	*Salmonella enterica*
X72585	R751	Tn*402*		*Klebsiella aerogenes*
L06418	pDGO100		In7	*Escherichia coli*
U67194	R751			*Enterobacter aerogenes*
AF263520			In58	*Pseudomonas aeruginosa*
AY123253	pRMH760			*Klebsiella pneumoniae*

enzyme, integration and excision of gene cassettes from this integron would only occur in the presence of an integrase supplied *in trans*.

A further 15 Antarctic DNA samples (MVT5-8 & 10-12, MV14-17, UWV1 & 8, Brat 1 & 2) were investigated for the presence of integrase genes but failed to yield a visible product following a threshold of 35 cycles of amplification. Due to the unexpectedly high level of integrase sequence identity shared by pSa, bed dust and soil 1A sequences with those recovered from Antarctic soils, the potential amplification of a contaminant was considered. The integrase gene, *intI1*, is associated with a number of different class 1 integrons present on various mobile elements, which have correspondingly been isolated from a wide range of host organisms; it is highly conserved exhibiting 99-100% nucleotide sequence identity (Partridge *et al.*, 2001). However, by virtue of their geographical isolation and the presence of different selective pressures, integrase genes recovered from Antarctic environments were expected to show a greater degree of sequence divergence. The integrase gene, intI1, is associated with a number of different class 1 integrons present on various mobile elements, which have correspondingly been isolated from a wide range of host organisms; it is highly conserved exhibiting 99-100% nucleotide sequence identity (Partridge et al., 2001). However, by virtue of their geographical isolation and the presence of different selective pressures, integrase genes recovered from Antarctic environments were expected to show a greater degree of sequence divergence. Whilst the negative control for each round of PCR gave no discernable product, introduction of a contaminant during the purification of first round PCR products prior to subsequent reamplification could not be ruled out. This was of concern given the failure to recover an integrase gene product from other Antarctic DNA samples.

A product of 819 bp was recovered from sample MVT9 (designated MVT9.11) following two rounds of amplification with primer pair 12b2 / 5CS-Rev. A second round of amplification was performed in this instance to increase product yield. Sequence

analysis identified an orf of 152 aa. As shown in Figure 5.4, this orf exhibited 58% identity and 72% similarity to the carboxy terminus of a transposase (405 aa) identified within *Thermoanaerobacter tengcongensis* (phylum *Firmicutes*) The orf spanned positions 1-459 of sequence MVT9.11, position 1 corresponding to the 5' end of primer 12b2. The remaining 343 bp of the sequence (positions 460-802) excluding the 5CS-Rev primer, was found to exhibit no homology to integron-related sequences or to possess any integron-related features. Therefore, it was concluded that this sequence was not part of an integron structure and was most likely the result of a non-specific annealing event.

In an attempt to confirm the genuine presence of *intI1* in the above Antarctic DNA samples, amplification of the integrase gene was repeated using an alternative integrase-specific primer, intA, that targets the 3' end of *intI1*. Primer set intA and 5CS-Rev amplify a region of 978 bp of the 5' conserved segment of class I integrons that includes all but the last 22 amino acids of the carboxy termini of the integrase gene (Figure 5.2). Products of the expected size after 35 cycles of amplification were clearly visible for pSa, bed dust and soil 1A although product yield for the latter was significantly lower. Recovery of an integrase gene was confirmed by partial sequence analysis of the cloned bed dust product which exhibited 100% (over 500 bp) nucleotide sequence identity to the *intI1* gene of class 1 integrons; this sequence also shared 100% identity to the 12b2/5CS-Rev PCR product previously amplified from bed dust. Pooled Antarctic samples UWV1-4, UWV5-8 and MV8/12 failed to yield a product with this primer combination, despite purification of the initial PCR reaction through a spin column for use as template in a second round of amplification. (Re-amplification of the original PCR reaction was carried out in case there was insufficient product yield to be visualized under UV light.)

Partial sequence analysis of a second intA/5CS-Rev product of approximately 700bp in size recovered from bed dust (designated BD2.2) revealed 99% identity in a 505 bp overlap to a region of the *E. coli* K12 genome (accession number D90706). An orf of 165 aa was identified exhibiting 99% identity to the conserved hypothetical protein,

YleA (accession number NP_752677), as shown in Figure 5.5. The YleA protein of 474 aa in length is identical to the product of the *miaB* gene, encoding a putative tRNA-thiotransferase in *Salmonella enterica* subsp. *enterica* serovar *typhi*. Analysis of the nucleotide sequence of the *yleA* gene revealed positions 77-85 to be identical to the 3' end of the 5CS-Rev primer, whilst positions 734-756 were of sufficient similarity to permit binding of the intA primer, thus resulting in amplification of a 688 bp product.

Figure 5.4: Sequence MVT9.11

Figure 5.4: a) Annotation of nucleotide sequence MVT9.11 amplified with primer pair 12b2 / 5CS-Rev from sample MVT9 and deduced amino acid sequence of partial orf.

12b2 primer →

```
1    CGTAGCGAACGAGTGGCGGAAGTGATCGTCTCCGAGATTGGTGTCGAGATGAGTTGTTTTCCTTCAGAC   69
     R  S  E  R  V  A  E  V  I  V  S  E  I  G  V  E  M  S  C  F  P  S  D

70   GGACACCTTGCGAGTTGGGCTGGCGTGTGTCCCGGCAGTCATGAATCAGCGGGCAAGCGCAAAAGCGGC  138
     G  H  L  A  S  W  A  G  V  C  P  G  S  H  E  S  A  G  K  R  K  S  G

139  AAGACAACAAAAGGCAATGTCTATGTACGAACGGCGTTAGTCCAAGCGGCGTGGGCGGCGACACACACC  207
     K  T  T  K  G  N  V  Y  V  R  T  A  L  V  Q  A  A  W  A  A  T  H  T

208  AAAGACACATACCTCGCGGCGCAGTACCGGCGGTTGGTCAAACACAAAGGCAAACAGCGGGCGTTAGTC  276
     K  D  T  Y  L  A  A  Q  Y  R  R  L  V  K  H  K  G  K  Q  R  A  L  V

277  GCCGTCGGACACAGCATCTTGGTCATCATCTACCATGTCCTGAAGCACAAGGTTAGTTATCATGAGTTA  345
     A  V  G  H  S  I  L  V  I  I  Y  H  V  L  K  H  K  V  S  Y  H  E  L

346  GGTGGCGGCTACTTTGACCGTCAGAATGTTGAGCGCCAACGCCAGATGTTTATTCGTAAACTTGAAGGG  414
     G  G  G  Y  F  D  R  Q  N  V  E  R  Q  R  Q  M  F  I  R  K  L  E  G

425  CTTGGACTCAAGGTCACAGTTGAGGAACTGCCCGTCGCGGCTTAGTCCCAACCAATCTGTTTTCATAAG  483
     L  G  L  K  V  T  V  E  E  L  P  V  A  A  .

484  AGTAGTCCATGCCGTGGACCTGATAGCCGCGCTTCAGTAACCAGTTGACATCATTGATACTGCCGCCTC  552

553  CGGCATCGACGCGGATGACCGTCCGGCGGCGTTTGCTCTGATCCAAGTTGAGAGTCTGTTCGGCGGCGG  621

622  TGATCAACTCCGGTAAGGTCGTGTTGAGTTGCCGGTTGCCCGGATAGAGTTGATCGACCACGACCTCGC  690

691  CATACTGCGCGGCACTTACCGCCCTAACTGCCGCCTCTGCGCCGCCTCATAGTGGAAATAGCCAGGGGT  759

760  GGCGCGCTCGGCCTTCTTGCCCAGGGCATCCCGCTCAGATCGGCATCCAAGCAGCAAG              891
```

← 5CS-Rev primer

Figure 5.4:b) Alignment of the partial orf identified from sequence MVT9.11 with the transposase of *Thermoanaerobacter tengcongensis* (accession number NP_622741). The alignment spans 142 amino acids of the carboxy terminus of the transposase. (N.B. tsp: transposase.)

Identities = 84/143 (58%), Positives = 104/143 (72%)

```
MVT9:   4 RVAEVIVSEIGVEMSCFPSDGHLASWAGVCPGSHESAGKRKSGKTTKGNVYVRTALVQAA 63
          R AE I++EIG  M  FPS  HL SWAGV PG++ESAGKRKSG+T KGN  +R+ LV+AA
tsp: 260 RTAEHIIAEIGTNMDQFPSAAHLCSWAGVAPGNNESAGKRKSGRTRKGNEKLRSVLVEAA 319

MVT9: 64 WAATHTKDTYLAAQYRRLVKHKGKQRALVAVGHSILVIIYHVLKHKVSYHELGGGYFDRQ 123
           AA HTKDTYL+AQY R+   +G  RA VAV HSIL I+Y++LK K  Y+ELG  Y++ +
tsp: 320 RAAAHTKDTYLSAQYHRIAARRGANRAAVAVAHSILTIVYYLLKRKERYNELGVNYYEER 379

MVT9:124 NVERQRQMFIRKLEGLGLKVTVE 146
```

```
            E   +  I+KLE LGLKVTVE
tsp: 380 KKEAIVKQSIKKLEALGLKVTVE 402
```

Figure 5.5: Sequence BD2.2

Figure 5.5: a) Potential binding sites of primers 12b2 and 5CS-Rev to the nucleotide sequence of the gene encoding the hypothetical protein YleA (accession number NP_752677) (partial sequence shown).

```
                                                                    CTT
 1  ATGACCAAAAAACTCCATATTAAAACCTGGGGCTGTCAGATGAACGAGTACGATTCATCGAAGATGGCCG  70
   5CS-Rev primer
    GCTGCTTGGATGCC
71  ATCTGCTGGATGCCACCCACGGCTATCAACTGACCGACGTGGCGGAAGAAGCGGATGTGCTGCTGCTGAA  140

141 CACCTGCTCAATCCGCGAGAAGGCTCAGGAAAAAGTCTTCCATCAGTTGGGTCGCTGGAAACTGTTAAAA  210

211 GAGAAGAATCCAGACCTGATTATCGGCGTCGGCGGCTGCGTGGCATCGCAAGAAGGCGAGCACATTCGCC  280

281 AGCGCGCCCACTATGTCGATATTATTTTTGGGCCGCAAACGCTGCACCGCCTGCCGGAGATGATCAACTC  350

351 CGTGCGCGGCGACCGCAGCCCGGTTGTAGATATCAGCTTCCCGGAAATCGAGAAGTTTGACCGTCTGCCG  420

421 GAACCGCGCGCCGAAGGGCCGACCGCGTTTGTCTCCATCATGGAAGGCTGCAATAAATATTGCACCTACT  490

491 GCGTAGTGCCTTACACCCGTGGTGAAGAGGTAAGCCGTCCGTCCGACGATATTCTGTTTGAGATTGCCCA  560

561 GCTTGCGGCTCAGGGCGTGCGTGAAGTTAACCTGCTCGGCCAGAACGTGAACGCCTGGCGCGGTGAGAAC  630

631 TACGACGGCACCACCGGATCGTTTGCCGATCTGCTGCGTCTGGTTGCGGCGATCGACGGGATCGATCGTA  700
                                    12b2 Primer
                                    CCGACG-TCTCTACGACGATGAT
701 TTCGCTTTACCACCAGCCATCCGATCGAATTCACCGACGATATCATCGAAGTGTACCGCGACACGCCGGA  770

771 GCTGGTGAGCTTCCTGCATCTGCCGGTACAGAGCGGTTCCGATCGCATTCTGAACCTGATGGGACGTACC  840

841 CATACGGCGCTGGAGTACAAAGCGATCATCCGTAAACTGCGTGCGGCGCGCCCGGATATTCAGATCAGCT  910
```

Figure 5.5:b) Alignment of partial orf identified from sequence BD2.2 with the hypothetical protein YleA.

Identities = 164/165 (99%), Positives = 164/165 (99%)

```
BD2.2: 1  LDATHGYQLTDVAEEADVLLLNTCSIREKAQEKVFHQLGRWKLLKEKNPDLIIGVGGCVA 60
          LDATHGYQLTDVAEEADVLLLNTCSIREKAQEKVFHQLGRWKLLKEKNPDLIIGVGGCVA
YleA: 26  LDATHGYQLTDVAEEADVLLLNTCSIREKAQEKVFHQLGRWKLLKEKNPDLIIGVGGCVA 85

BD2.2:61  SQEGEHIRQRAHYVDIIFGPQTLHRLPEMTNSVRGDRSPVVDISFPEIEKFDRLPEPRAE 120
          SQEGEHIRQRAHYVDIIFGPQTLHRLPEM NSVRGDRSPVVDISFPEIEKFDRLPEPRAE
YleA: 86  SQEGEHIRQRAHYVDIIFGPQTLHRLPEMINSVRGDRSPVVDISFPEIEKFDRLPEPRAE 145

BD2.2:121 GPTAFVSIMEGCNKYCTYCVVPYTRGEEVSRPSDDILFEIAQLAA 165
          GPTAFVSIMEGCNKYCTYCVVPYTRGEEVSRPSDDILFEIAQLAA
```

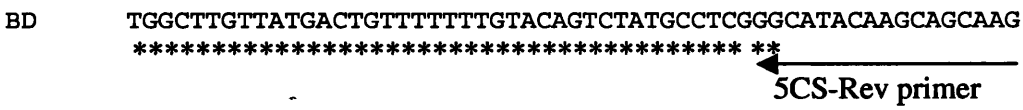
```
YleA: 146 GPTAFVSIMEGCNKYCTYCVVPYTRGEEVSRPSDDILFEIAQLAA 190
```

5.2.2: Development of integron-screening PCR

The inconsistency in PCR amplification of the integrase gene coupled with the low product yield from soil DNA after one round of amplification, raised concerns of the potential for amplification of a contaminant. This lead to the development of an alternative integron-screening PCR comprising the primer pair Brint/5CS-Rev, allowing the recovery of 117 bp fragment of the 5' conserved segment (Figure 5.2). The Brint primer is specific for the sequence at the 5' end of *IntI1* and permits identification of the first 20 amino acids of the gene. Products were recovered after one round of amplification for pSa, bed dust and Antarctic samples Brat2, UWV4 and MV17. Sequence analysis confirmed amplification of the 5'CS with sequences exhibiting 100% identity to the corresponding region of class 1 integrons. The nucleotide sequence of this region of the 5'CS is highly conserved among class 1 integrons to the same degree as *intI1*. Figure 5.6 shows that allis highly conserved among class 1 integrons to the same degree as *intI1*. Figure 5.6 shows that all sequences were identical (excluding primer sequences) with the exception of the sequence retrieved from UWV4, which possessed a single substitution, G to T, at the first nucleotide position directly following the 5CS-Rev primer.

Figure 5.6: Alignment of Brint/5CS-Rev nucleotide sequences amplified from the positive control, pSa, bed dust DNA (BD) and Antarctic DNA samples UWV4, Brat2 and MV17. Direct sequencing of the pSa and bed dust PCR products was performed, accounting for their reduced length. PCR products recovered from Antarctic samples were cloned prior to sequencing.

```
          Brint primer
          ──────────►
UWV4      GCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCA
Brat2     GCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCA
MV17      GCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCA
pSa       ----------=--------------------GGTGGTAACGGCGCAGTGGCGGTTTTCA
BD        ------------------------------GGTGGTAACGGCGCAGTGGCGGTTTTCA
                                          ****************************

UWV4      TGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCCTCTGGCATCCAAGCAGCAAG
Brat2     TGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCCTCGGGCATCCAAGCAGCAAG
MV17      TGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCCTCGGGCATCCAAGCAGCAAG
pSa       TGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCCTCGGG--------------
```



A product of 288 bp, designated UWV1 B5-2, was also recovered with primer pair Brint/5CS-Rev from sample UWV1. Sequence analysis revealed it to share 78% identity in a 251 nucleotide overlap to a region of the *Pseudomonas putida* genome, corresponding to a hypothetical protein. A partial orf of 54 aa was subsequently identified exhibiting 88% identity and 94% similarity at the amino acid level to the hypothetical protein, as shown in Figure 5.7. Additionally, this orf also exhibited 88% identity (98% similarity) to an unknown environmental sequence identified in an environmental DNA library constructed from DNA extracted from seawater of the Sargasso Sea (Venter *et al.*, 2004). However, the DNA sequence upstream of the orf bore no resemblance to either database relative identified.

As a reproducible PCR product was recovered from DNA extracted from sample MV17 on three separate occasions using the integron-screening PCR, this sample was selected for renewed attempts to amplify integrase genes. The following three primer combinations were used: 1) intA & intB, 2) Cass1 & 5CS-Rev and 3) 12b2 & 5CS-Rev (Figure 5.2). This strategy utilized three alternative integrase-specific forward primers in combination with a reverse primer (intB or 5CS-Rev), the target sequence of which was confirmed to be present in sample MV17 as evidenced by a positive result with the Brint/5CS-Rev screening PCR. (NB. Primer Brint is the reverse equivalent of intB.)

Amplification of MV17 DNA with the primer pair Cass 1/5CS-Rev failed to yield any products, despite recovery of a product of the correct size from pSa. A single product was amplified with primer pair 12b2/5CS-Rev; however, as shown in Figure 5.8 this was significantly shorter that the expected size of 911 bp. FASTA analysis of the 671 bp sequence identified a level of 68% identity in a 385 nucleotide overlap to a region of the *Streptomyces avermitilis* genome (accession number AP005048); this corresponded to a partial overlap with the *fadE2* gene, encoding a putative acyl-CoA dehydrogenase.

The expected size of product amplified with primer pair intA / intB was 893 bp. Amplification of pSa produced a single band of the correct size whilst amplification of MV17 DNA gave rise to multiple bands as shown in Figure 5.8. Bands A and B were

Figure 5.7: Sequence UWV1-B5

a) Partial orf identified from sequence UWV1-B5 amplified from sample UWV1 with primer pair Brint/5CS-Rev.

```
      Brint primer
      ──────────►
 1    GCAACTGGTCCAGAACCTTGACGCGCGCAACCAATCACTCGGGGGAGTGGAAACTCAAACTGTAGTTCAC  70

 71   GGCCTTGTTGCCTGGGTCGAGAATGTCCAGGGCGATATGGATCGGGGTCTGGGACGGCATCTCGGTCTGG 140
                        M  S  R  A  I  W  I  G  V  W  D  G  I  S  V  W

141   CCGGCCAACTCGCCGCTGAGGTATTCGGCGGGTTTGAAGCGCCGGCTGGCGATCATCCGGCCATTAAGGT 210
      P  A  N  S  P  L  R  Y  S  A  G  L  K  R  R  L  A  I  I  R  P  L  R  S

211   CGGCAAAGCGCAGCTCCAGCAGCGGGAACGGCTGGGAGAACGTCGCACGGTTATAGATGATGGCATCCAA 280
        A  K  R  S  S  S  S  G  N  G  W  E  N  V  A  R  L  .
                                                          ◄──────────

281   GCAGCAAG                                                               288

      ◄──────
      5CS-Rev primer
```

Figure 5.7: b) Alignment of the partial orf identified from sequence UWV1-B5 with its closest database relatives identified by PSI-Blast.

Accession number EAK34636: unknown environmental sequence from the Sargasso Sea.

Identities = 48/53 (90%), Positives = 52/53 (98%)

```
UWV1-B5:   1   MIASRRFKPAEYLSGELAGQTEMPSQTPIHIALDILDPGNKAVNYSLSFHSPE 53
               +IASRRFKPAEYLSGELAG +EMPSQTPIHI+LDILDPGN+AVNYSLSFHSPE
EAK34636: 367 LIASRRFKPAEYLSGELAGVSEMPSQTPIHISLDILDPGNRAVNYSLSFHSPE 419
```

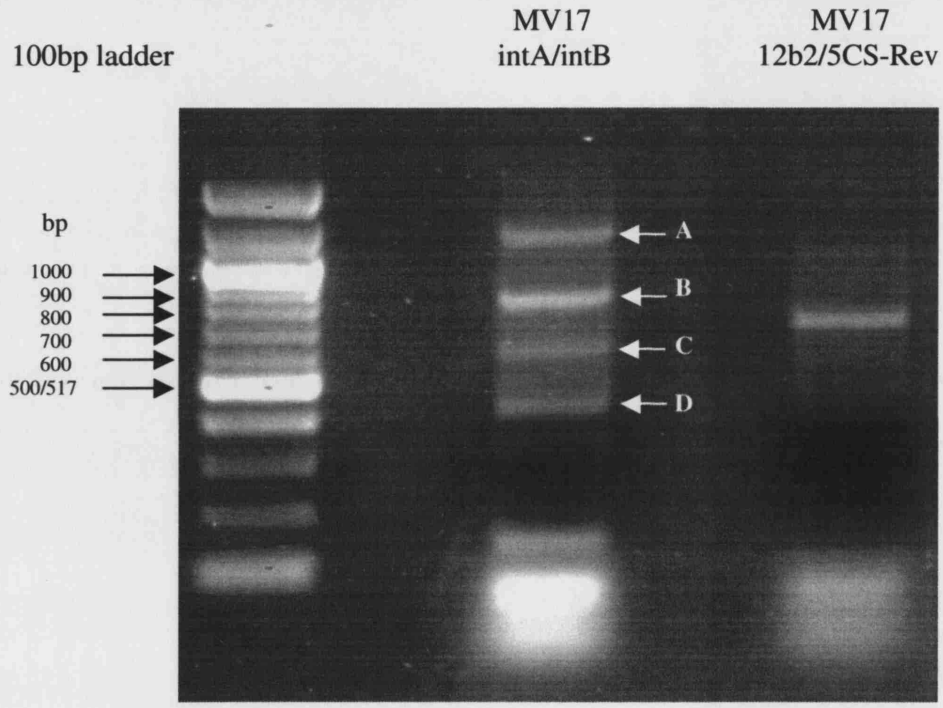
Accession number NP_746924: conserved hypothetical protein (*Pseudomonas putida* KT2440)

Identities = 47/53 (88%), Positives = 50/53 (94%)

```
UWV1-B5:      1 MIASRRFKPAEYLSGELAGQTEMPSQTPIHIALDILDPGNKAVNYSLSFHSPE 53
                +IASRRFKP+EYLSGELAG+ EMPSQTPIHIALDILDPG KAVNYSLSF SPE
NP_746924: 399 LIASRRFKPSEYLSGELAGRGEMPSQTPIHIALDILDPGPKAVNYSLSFRSPE 451
```

cloned for further analysis whilst C and D were discounted as they were considered too small to be genuine products. Two clones containing 'band A' inserts were sequenced, MV17-int4 (1150 bp) and MV17-int5 (1160 bp); however neither were found to be integrase sequences. A complete orf of 224 aa was identified from sequence MV17-int4 commencing 20 nucleotides downstream of the 3' end of primer intB. The orf identified here matched, at the amino acid level, a conserved hypothetical protein of the plant pathogen *Pseudomonas syringae* pv. *tomato* (sequence accession number NP_791317.1) to the level of 50% identity and 64% similarity over a region of 220 amino acids. Within this 220 amino acid region, a putative conserved domain of 140 amino acids was identified that corresponded to that identified for a protein, DUF533, of unknown function. Members of this family (pfam 04391.5) have been identified within the *Proteobacteria* and may be secreted or integral membrane proteins. A partial orf of 386 amino acids was identified from MV17-int5 spanning the entire length of the sequence, although it bore little similarity to any database sequence. The closest database match identified by PSI-Blast was a maturase-related protein from *Serratia maracescens* (sequence accession number AAL51020). In this case, the level of sequence identity and similarity was 33% and 52% respectively over a region of 98 amino acids. Interestingly, a conserved putative domain was identified corresponding to a reverse transcriptase enzyme that is indicative of a mobile element such as retrotransposon or retrovirus. However, given that the orf was incomplete and a substantial proportion of the orf failed to resemble a database sequence, its significance is unknown. Analysis of recombinant clones from the 'band B' library revealed that none contained inserts of the correct size. As this was one of the final pieces of work carried out, time constraints meant that it could not be pursued further. This was unfortunate given the band was of the expected size (~900 bp) and a genuine 12b2 / 5CS-Rev PCR product could be reproducibly obtained from sample MV17.

Figure 5.8: PCR products generated from amplification of MV17 DNA with integrase-specific primer pairs 1) intA/intB, and 2)12b2/5CS-Rev. Bands 'A' and 'B' amplified with intA/intB were selected for cloning and sequence analysis; band 'B' corresponds to the product of expected size of 893 bp. Bands 'C' and 'D' were considered too small to be genuine and were not analyzed further. A single product of ~700 bp was amplified from sample MV17 with primers 12b2 and 5CS-Rev; the expected size of product in this case was 911 bp.



5.3: Amplification of the Variable Region

The primer pair of Brint/3CS was originally selected for amplification of the variable region of integrons, which contains the integrated gene cassettes. The reverse primer 3CS targets conserved sequence in the 3' conserved segment (Figure 5.2). A band of approximately 1.1 kb was recovered from bed dust DNA following a single round of PCR amplification. Sequence analysis of a single clone (BD3) containing an insert of 1107 bp revealed an orf of 262 aa. The orf identified here, matched at the amino acid level, a streptomycin adenyltransferase to the degree of 85% identity and 93% similarity as shown in Figure 5.9. This protein is encoded by the *aadA1* gene present on the plasmid R100 (accession number BAA78801) and confers resistance to the antibiotics streptomycin and spectinomycin. Nucleotide positions 1-61 encoding the first 20 amino acids of the integrase in the reverse orientation were found to be 100% identical to *intI1*. Six clones from this library were analyzed by restriction digest analysis using three tetranucleotide restriction enzymes – *Hin*P1, *Msp*1 and *Mbo*1. A second clone possessing a small variation in *Hin*P1 restriction pattern was selected for sequencing but was found to share 99% nucleotide sequence identity to BD3.

A PCR product of 1108 bp was also obtained from soil sample 1A using primer pair Brint/3CS. Sequence analysis of the clone in question, DP-A3, revealed revealed an orf of 262 aa that that possessed 87% identity and 92% similarity at the amino acid level to the same database relative as BD3. The level of nucleotide sequence identity between DPA3 and BD3 was 99%. The same gene cassette was also recovered from soil 1A, designated 1AR, using an alternative reverse primer, HS286, which targets the 59-be at the 3' end of the cassette (Figure 5.2). Excluding the HS286 primer sequence due to its degeneracy, 1AR shared 99% nucleotide identity with DP-A3. Figure 5.10 represents an alignment of the amino acid sequence of the orf identified for each clone – BD3, DP-A3 and 1AR; a total of five residues exhibit an amino acid substitution.

Amplification of the pooled Antarctic DNA samples from which 12b2/5CS-Rev integrase products were cloned failed to yield any products. However, a product of approximately 1 kb was amplified and cloned from sample UWV4. Sequence analysis of

Figure 5.9: Sequence BD3

a) Annotation of the nucleotide sequence of BD3 amplified from bed dust DNA with primer pair Brint/3CS. Solid line arrows denote the primer sequences, dashed-line arrow represents the direction of transcription of the integrase gene. The deduced amino acid sequence of *intI1* and the *aadA1* gene are shown. The attachment site, *attI,* into which the gene cassette is inserted is highlighted in bold type and the start codon of the *aadA1* gene is underlined in bold type. The DNA sequence underlined at the 3'end of the aadA1 gene represents the 59-base element.

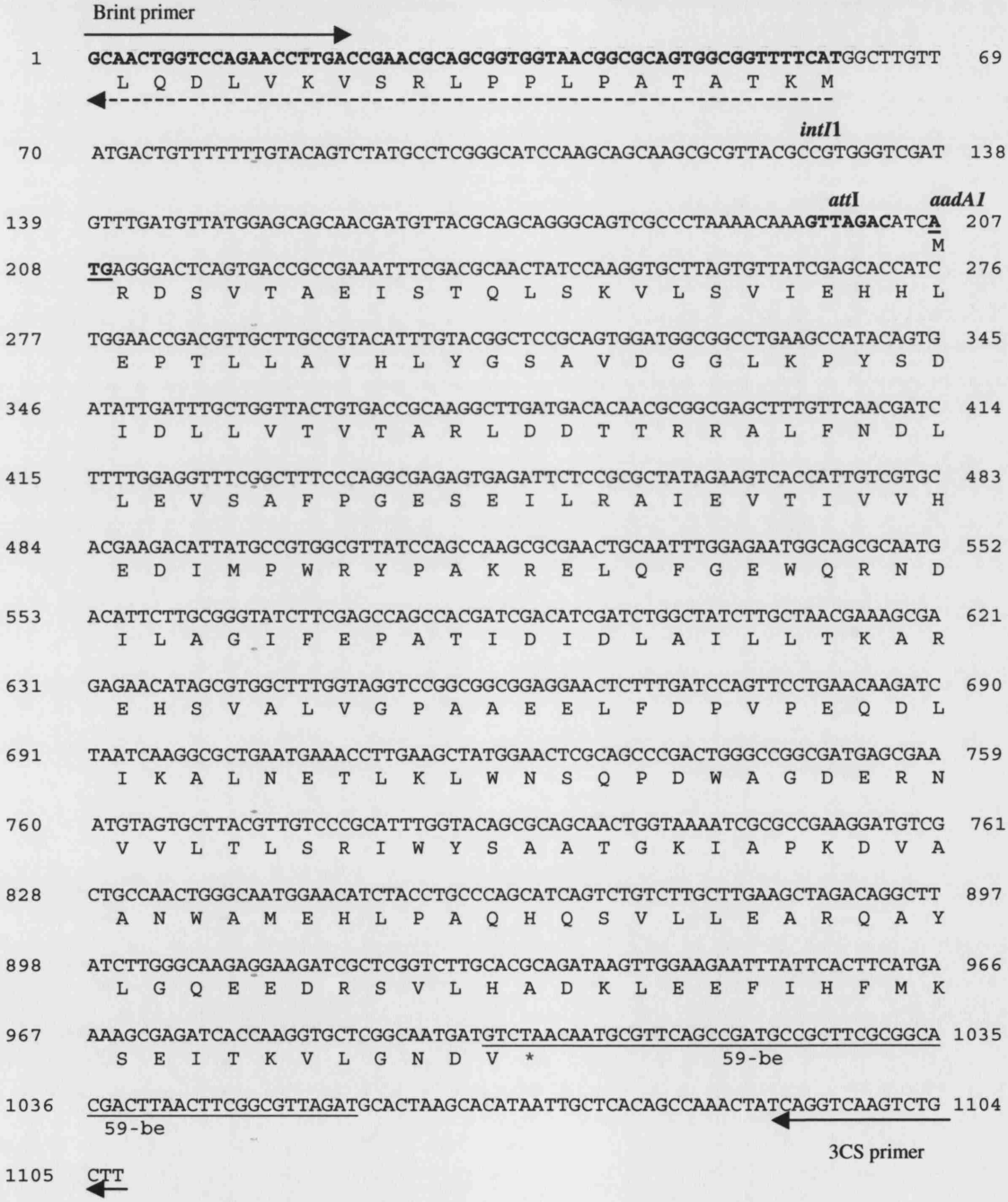


Figure 5.9: b) Alignment of the amino acid sequence of the orf identified from BD3 with the product of the *aadA1* gene of plasmid R100, (accession number BAA78801).

Identities = 225/262 (85%), Positives = 244/262 (93%)

```
BD3:     1  MRDSVTAEISTQLSKVLSVIEHHLEPTLLAVHLYGSAVDGGLKPYSDIDLLVTVTARLDD   60
            MR++V AE+STQLS+V+ VIE HLEPTLLAVHLYGSAVDGGLKP+SDIDLLVTVT RLD+
aadA1: 74  MREAVIAEVSTQLSEVVGVIERHLEPTLLAVHLYGSAVDGGLKPHSDIDLLVTVTVRLDE  133

BD3:    61  TTRRALFNDLLEVSAFPGESEILRAIEVTIVVHEDIMPWRYPAKRELQFGEWQRNDILAG  120
            TTRRAL NDLLE SA PGESEILRA+EVTIVVH+DI+PWRYPAKRELQFGEWQRNDILAG
aadA1:134  TTRRALINDLLETSASPGESEILRAVEVTIVVHDDIIPWRYPAKRELQFGEWQRNDILAG  193

BD3:   121  IFEPATIDIDLAILLTKAREHSVALVGPAAEELFDPVPEQDLIKALNETLKLWNSQPDWA  180
            IFEPATIDIDLAILLTKAREHSVALVGPAAEELFDPVPEQDL +ALNETL LWNS PDWA
aadA1:194  IFEPATIDIDLAILLTKAREHSVALVGPAAEELFDPVPEQDLFEALNETLTLWNSPPDWA  253

BD3:   181  GDERNVVLTLSRIWYSAATGKIAPKDVAANWAMEHLPAQHQSVLLEARQAYLGQEEDRSV  240
            GDERNVVLTLSRIWYSA TGKIAPKDVAA+WAME LPAQ+Q V+LEARQAYLGQEEDR
aadA1:254  GDERNVVLTLSRIWYSAVTGKIAPKDVAADWAMERLPAQYQPVILEARQAYLGQEEDRLA  313

BD3:   241  LHADKLEEFIHFMKSEITKVLG 262
              AD+LEEF+H++K EITKV+G
aadA1:314  SRADQLEEFVHYVKGEITKVVG 335
```

Figure 5.10: Alignment of the amino acid sequence of *aadA1* gene cassettes conferring resistance to streptomycin and spectinomycin amplified from bed dust (BD3) and soil 1A (1AR and DP-A3).

```
           1                                                          60
BD3        MRDSVTAEISTQLSKVLSVIEHHLEPTLLAVHLYGSAVDGGLKPYSDIDLLVTVTARLDD
1AR        MRDSVTAEISTQLSKVLSVIEHHLEPTLLAVHLYGSAVDGGLKPYSDIDLLVTVTARLDD
DP-A3      MRDSVTAEISTQLSKVLSVIEHHLEPTLLAVHLYGSAVDGGLKPYSDIDLLVTVTARLDD
           ************************************************************

           61                                                         120
BD3        TTRRALFNDLLEVSAFPGESEILRAIEVTIVVHEDIMPWRYPAKRELQFGEWQRNDILAG
1AR        TTRRALFNDPLEVPAFPGESEILRAIEVTIVVHEDIMPWRYPAKRELQFGEWQRNDILAG
DP-A3      TTRRALFNDLLEASAFPGESEILRAIEVTIVVHEDTMPWRYPVKRELQFGEWQRNDILAG
           ********* **..******************** ******.*****************

           121                                                        180
BD3        IFEPATIDIDLAILLTKAREHSVALVGPAAEELFDPVPEQDLIKALNETLKLWNSQPDWA
1AR        IFEPATIDIDLAILLTKAREHSVALVGPAAEELFDPVPEQDLIKALNETLKLWNSQPDWA
DP-A3      IFEPATIDIDLAILLTKAREHSVALVGPAAEELFDPVPEQDLIEALNETLKLWNSQPDWA
           **********************************************:***************

           181                                                        240
BD3        GDERNVVLTLSRIWYSAATGKIAPKDVAANWAMEHLPAQHQSVLLEARQAYLGQEEDRSV
1AR        GDERNVVLTLSRIWYSAATGKIAPKDVAANWAMEHLPAQHQSVLLEARQAYLGQEEDRSV
DP-A3      GDERNVVLTLSRIWYSAATGKIAPKDVAANWAMEHLPAQHQSVLLEARQAYLGQEEDRSV
           ************************************************************

           241                262
BD3        LHADKLEEFIHFMKSEITKVLG
1AR        LRADKLEEFIHFMKSEITKVLG
DP-A3      LRADKLEEFIHFMKSEITKVLG
           *:********************
```

the 1146 bp insert, designated UWV4-B1, revealed an orf of 281 amino acids. The orf identified shared 98% identity (98% similarity) at the amino acid level to an aminoglycoside adenyltransferase as shown in Figure 5.11. This protein is the product of the *aadA6* gene located within In51 (sequence accession number AAD54217), a class 1 integron of *Pseudomonas*, and confers high level resistance to the antibiotics streptomycin and spectinomycin (Naas *et al.*, 1999). A second clone from this library, UWV4-B4, was selected for sequencing on the basis of restriction enzyme analysis. Unfortunately, the cloned insert was not sequenced in its entirety (due to loss of the plasmid sample); however, a total of 933 bp of sequence information was obtained. Two orfs of 81 and 138 amino acids in length were identified from the 5' and 3' portions of the sequence respectively. Each orf was found to exhibit 98% identity (98% similarity) at the amino acid level to the same database relative as UWV4-B1 (AAD54217). The level of nucleotide identity of the 5' and 3' portion of the UWV4-B4 sequence to UWV4-B1 was 99% and 98% respectively. An alignment of the partial sequence of the UWV4-B4 gene cassette with that of UWV4-B1 is shown in Figure 5.111. c).

Amplification with primer pair Brint/3CS did result in some spurious products that upon sequencing were found to be unrelated to integrons. Two products were amplified from soil 1A, DP-A5 (512 bp) and DP-A8 (597 bp). A partial orf of 168 amino acids running the entire length of DP-A5 as shown in Figure 5.12:a) was found to share 37% identity and 56% similarity at the amino acid level to an unknown environmental sequence cloned from DNA extracted from the Sargasso Sea (accession number EAJ14596) (Figure 5.12:b). The closest database relative identified with an assigned function was a putative exopolysaccharide biosynthesis protein from *Vibrio cholerae* (accession number NP_230584), to which the DP-A5 orf exhibited 31% amino acid identity and 52% similarity (Figure 5.12:c). Two consecutive putative orfs were identified within the DP-A8 sequence, represented in Figure 5.13.a). The first complete orf of 81 amino acids shared 55% identity and 75% similarity at the amino acid level to an orf from an environmental sequence retrieved from the Sargasso Sea (accession number EAG92777) as shown in Figure 5.13.b.i) Similarly, this orf also shared 53% identity and 73% similarity to a conserved hypothetical protein of *Synechococcus* (accession number

NP_896671) (Figure 5.13.b.ii). A conserved domain was identified in this orf corresponding to the Antitoxin protein of the toxin-antitoxin stability system involved in cell division and chromosome partitioning. The second orf comprising 81 amino acids and for which no obvious start codon was apparent, shared at the amino acid level, 62% identity (85% similarity) to an environmental sequence from the Sargasso Sea (accession number EAA81077) as shown in Figure 5.13.c.i) and 71% identity (83% similarity) to a conserved hypothetical protein of *Nitrosomonas europaea* (accession number NP_841401) (Figure 5.13.c.ii).

A product unrelated to an integron was also amplified from soil 1A with the primer pair Brint/HS286. Amplification of non-integron sequences was of increased risk in this instance due to the degeneracy of the HS286 primer. The product of 596 bp contained a partial orf of 73 amino acids commencing 354 nucleotides downstream of the Brint primer. The orf identified here shared 76% identity and 87% similarity at the amino acid level to the LexA protein, an SOS response regulator of *Mesorhizobium loti* (accession number NP_102395) of 245 amino acids in length.

Due to the limited success in recovery of integron sequences from Antarctic soil samples and the amplification of a series of spurious products, an alternative primer pair was designed. The primers, Int5CS and Int3CS, targeting conserved sequence within the 5' and 3' conserved segments respectively, were designed to be longer in length than the primers previously used, in an attempt to improve the efficacy and specificity of the PCR. Primer Int5CS is located 15 bp downstream of the 3' end of the 5CS-Rev primer that was used as part of the integron screening PCR. Int5CS is specific to a region of highly conserved sequence located between the 5' end of the integrase gene and the *attI* site preceding an integrated cassette. Primer Int3CS targets conserved sequence directly adjacent to the 3' end of the 3CS primer sequence. It was hoped that primers targeting other regions of sequence within the 5'CS and 3'CS region than those used previously would result in the amplification of additional integron sequences.

Figure 5.11: Sequence UWV4-B1

Figure 5.11: a) Annotation of UWV4-B1 nucleotide sequence showing the deduced amino acid sequence of *intI1* and *aadA6*. Solid line arrows denote primer sequences, dashed line arrow represents direction of transcription of *intI1*, *attI* – attachment site for insertion of gene cassette. The start codon of the *aadA6* gene is underlined in bold type and the 59-base element at the 3' end of the sequence is underlined.

Brint primer

intI1

GCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTT

L Q D L V K V S R L P P L P A T A T K M

ATGACTGTTTTTTTGTACAGTCTATGCCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGGTCGAT

attI *aadA6*

GTTTGATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAGCAAA**GTTAGAC**ATC**A**

M

TGAGTAACGCAGTACCCGCCGAGATTTTGGTACAGCTATCACTGGCTCTCAACGCCATCGAGCGTCATC

S N A V P A E I L V Q L S L A L N A I E R H L

TGGAATCAACGTTGCTGGCCGTGCATTTGTACGGCTCTGCACTGGACGGTGGCCTGAAGCCATACAGTG

E S T L L A V H L Y G S A L D G G L K P Y S D

ATATTGATTTGCTGGTTACTGTGGCTGCACGGCTCGATGAGACTGTCCGACAAGCCCTGGTCGTAGATC

I D L L V T V A A R L D E T V R Q A L V V D L

TCTTGGAAATTTCTGCCTCCCCTGGCCAAAGTGAGGCTCTCCGCGCCTTGGAAGTTACCACCGTCGTGC

L E I S A S P G Q S E A L R A L E V T T V V H

ATGGTGATGTTGTCCCTTGGCGTTATCCGGCCAGACGGGAACTGCAATTCGGGGAGTGGCAGCGTAAGG

G D V V P W R Y P A R R E L Q F G E W Q R K D

ACATTCTTGCGGGCATCTTCGAGCCCGCCACAACCGATGTTGATCTGGCTATTCTGCTAACTAAAGTAA

I L A G I F E P A T T D V D L A I L L T K V R

GGCAGCATAGCCTTGCATTGGCAGGTTCGGCCGCAGAGGATTTCTTTAACCCAGTTCCGGAAGGCGATC

Q H S L A L A G S A A E D F F N P V P E G D L

TATTCAAGGCATTGAGCGACACTCTGAAACTATGGAATTCGCAGCCGGATTGGGTAGGCGATGAGCGGA

F K A L S D T L K L W N S Q P D W V G D E R N

ATGTAGTGCTTACCTTGTCTCGCATTTGGTACAGCGCAGCAACCGGCAAGATCGCACCGAAGGATATCG

V V L T L S R I W Y S A A T G K I A P K D I V

TTGCCAACTGGGCAATGGAGCGTCTGCCAGATCAACATAAGCCCGTACTGCTTGAAGCCCGGCAGGCTT

A N W A M E R L P D Q H K P V L L E A R Q A Y

ATCTTGGACAAGGAGAAGATTGCTTGGCCTCACGCGCGGATCAGTTGGCGGCGTTCGCTCACTTCGTGA

L G Q G E D C L A S R A D Q L A A F A H F V K

AACATGAAGCCACTAAATTGCTTAGTGCCATGCCAGTGATGTCTAACAATTCATTCAAGCCGACGCCGC

H E A T K L L S A M P V M S N N S F K P T P L

TTCGCGGCGCGGCTTAATTCAGGCGTTAGATGCACTAAGCACATAATTGCTCACAGCCAAACTATCAGG

R G A A *

TCAAGTCTGCTT

3CS primer

Figure 5.11: b) Alignment of the amino acid sequence of the orf identified from UWV4-B1 with product of the aadA6 gene identified on In51 of *Pseudomonas aeruginosa* (accession number AAD54217).

Identities = 277/281 (98%), Positives = 277/281 (98%)

```
UWV4-B1:   1  MSNAVPAEILVQLSLALNAIERHLESTLLAVHLYGSALDGGLKPYSDIDLLVTVAARLDE 60
              MSNAVPAEI VQLSLALNAIERHLESTLLAVHLYGSALDGGLKPYSDIDLLVTVAARLDE
aadA6:     1  MSNAVPAEISVQLSLALNAIERHLESTLLAVHLYGSALDGGLKPYSDIDLLVTVAARLDE 60

UWV4-B1:  61  TVRQALVVDLLEISASPGQSEALRALEVTTVVHGDVVPWRYPARRELQFGEWQRKDILAG 120
              TVRQALVVDLLEISASPGQSEALRALEVT VVHGDVVPWRYPARRELQFGEWQRKDILAG
aadA6:    61  TVRQALVVDLLEISASPGQSEALRALEVTIVVHGDVVPWRYPARRELQFGEWQRKDILAG 120

UWV4-B1: 121  IFEPATTDVDLAILLTKVRQHSLALAGSAAEDFFNPVPEGDLFKALSDTLKLWNSQPDWV 180
              IFEPATTDVDLAILLTKVRQHSLALAGSAAEDFFNPVPEGDLFKALSDTLKLWNSQPDW
aadA6:    121  IFEPATTDVDLAILLTKVRQHSLALAGSAAEDFFNPVPEGDLFKALSDTLKLWNSQPDWE 180

UWV4-B1: 181  GDERNVVLTLSRIWYSAATGKIAPKDIVANWAMERLPDQHKPVLLEARQAYLGQGEDCLA 240
              GDERNVVLTLSRIWYSAATGKIAPKDIVANWAMERLPDQHKPVLLEARQAYLGQGEDCLA
aadA6:    181  GDERNVVLTLSRIWYSAATGKIAPKDIVANWAMERLPDQHKPVLLEARQAYLGQGEDCLA 240

UWV4-B1: 241  SRADQLAAFAHFVKHEATKLLSAMPVMSNNSFKPTPLRGAA 281
              SRADQLAAF HFVKHEATKLLSAMPVMSNNSFKPTPLRGAA
aadA6:    241   SRADQLAAFVHFVKHEATKLLSAMPVMSNNSFKPTPLRGAA 281
```

Figure 5.11: c) Alignment of partial orf identified from sequence UWV4-B4 with the complete orf of sequence UWV4-B1. The dashed line represents unavailable sequence for UWV4-B4.

```
UWV4-B1    1 MSNAVPAEILVQLSLALNAIERHLESTLLAVHLYGSALDGGLKPYSDIDLLVTVAARLDE  60
UWV4-B4      MSNAVPAGISVQLSLALNAIERHLESTLLAVHLYGSALDGGLKPYSDIDLLVTVAARLDE
             ******* * **************************************************

UWV4-B1   61 TVRQALVVDLLEISASPGQSEALRALEVTTVVHGDVVPWRYPARRELQFGEWQRKDILAG 120
UWV4-B4      TVRQALVVDLLEISASPGQSE-----------------------------------
             ********************

UWV4-B1  121 IFEPATTDVDLAILLTKVRQHSLALAGSAAEDFFNPVPEGDLFKALSDTLKLWNSQPDWV 180
UWV4-B4      ----------------------ALAGSAAEDFFNPVPEGDLFKALSDTLKLWNSQPDWE
                                    **********************************

UWV4-B1  181 GDERNVVLTLSRIWYSAATGKIAPKDIVANWAMERLPDQHKPVLLEARQAYLGQGEDCLA 240
UWV4-B4      GDERNVVLTLSRIWYSAATGKIAPKDIVANWAMGCLPDQHKPVLLEARQAYLGQGEDCLA
             ********************************  ************************

UWV4-B1  241 SRADQLAAFAHFVKHEATKLLSAMPVMSNNSFKPTPLRGAATRGAA 286
UWV4-B4      SRADQLAAFVHFVKHEATKLLSAMPVMSNNSFKPTPLRGAA-----
             *********.********************************
```

Figure 5.12: Sequence DP-A5

Figure 5.12: a) Annotated sequence of DP-A5 amplified from soil 1a with primer pair Brint/3CS showing the amino acid sequence of the partial orf identified. Arrows denote primer sequences.

```
      Brint primer
      ------------------------>
  1   GCAACTGGTCCAGAACCTTGACCAAGTGCAGCCAGTGGTGCTGAAGTGGTGGCAAAAATGGCTGCCAAA   69
          L  V  Q  N  L  D  Q  V  Q  P  V  V  L  K  W  W  Q  K  W  L  P  K

 70   AGAATGGCTTCCTGTTCCCAGCAAGCCCGACCCACGAGCGCAGCGCGAAGGCTTGATCTCCGGTATTCT  138
       E  W  L  P  V  P  S  K  P  D  P  R  A  Q  R  E  G  L  I  S  G  I  L

139   GGGCGGCATCCAGGTGCAGCCAATTCGTTCCAGTCAGCTAGCGCGCATTTCCTACGTGGATCGTGATCC  207
       G  G  I  Q  V  Q  P  I  R  S  S  Q  L  A  R  I  S  Y  V  D  R  D  P

208   GCAACAAGCCGCCGATTTCGCTAACGCCGCCGCGAACGCCTATATCGTTGAGACGCTCGAAGGCCGTCT  276
       Q  Q  A  A  D  F  A  N  A  A  A  N  A  Y  I  V  E  T  L  E  G  R  L

277   GCAAATGGTCAGTTTGGGCTCGCAGTGGTTAGCCAAGCGACTTGATGGACTGCGCAAGAAAGTCGACGA  345
       Q  M  V  S  L  G  S  Q  W  L  A  K  R  L  D  G  L  R  K  K  V  D  D

346   TTCAGAGCGCGCGTTGCAAGCGTACAAAGAGCGCGAGAACATTGTCGACGTTAAAACTTCGCAATCTTT  414
       S  E  R  A  L  Q  A  Y  K  E  R  E  N  I  V  D  V  K  T  S  Q  S  L

415   GGAAGTCCAGCAAGTCAACGTGGCATCCTCACAGTTGACCGAAGCGCGCTCCGAGAGGATGCGCATGCA  483
       E  V  Q  Q  V  N  V  A  S  S  Q  L  T  E  A  R  S  E  R  M  R  M  Q

484   ATCAATCTACGATCAGGTCAAGTCTGCTT
       S  I  Y  D  Q  V  K  S  A
                <------------------
                3CS Primer
```

Figure 5.12.b) Alignment of DP-A5 orf with the partial orf from an unknown environmental sequence, accession number EAJ14596.

Identities = 60/158 (37%), Positives = 90/158 (56%)

```
DP-A5:       7  QVQPVVLKWWQKWLPKEWLPVPSKPDPRAQ-----REGLISGILGGIQVQPIRSSQLARI 61
                +V+P ++   +  LP  +LP P      A+     RE LI+     + + PIR +QL  I
EAJ14596:   64  EVKPGLMTRLRDLLP--FLPTPESELTEAELAHKAREDLITEFQQRLTISPIRKTQLVNI 121

DP-A5:      62  SYVDRDPQQAADFANAAANAYIVETLEGRLQMVSLGSQWLAKRLDGLRKKVDDSERALQA 121
                S+    PQ AAD ANA A+AYI   LE +L +    + WL  RL  LR ++DDSE  LQA
EAJ14596:  122  SFQTYSPQLAADVANAIADAYIESQLEAKLGITQKAASWLGGRLGELRAQLDDSEARLQA 181

DP-A5:     122 YKERENIVDVKTSQSLEVQQVNVASSQLTEARSERMRM 159
               Y+E+E ++DV+  + L  +++   S  LTEARS + ++
EAJ14596:  182 YREKEGLIDVEGVRGLGAKELERLSEALTEARSRKAQV 219
```

Figure 5.12.c) Alignment of DP-A5 orf with putative exopolysaccharide biosynthesis protein of *Vibrio cholerae* (737 amino acids in length), accession number NP_230584.

Identities = 49/157 (31%), Positives = 83/157 (52%)

```
DP-A5:      23  EWLP---------VPSKPDPRAQREG--LISGILGGIQVQPIRSSQLARISYVDRDPQQA 71
                EW+P           + PD  A R+   ++     G+++ PIR +QL  I Y   DP+ A
NP_230584: 128  EWMPFLQGFKQEETVADPDLEAYRQNRLILYKFKRGMEISPIRKTQLVNIYYTAGDPKLA 187

DP-A5:      72  ADFANAAAANAYIVETLEGRLQMVSLGSQWLAKRLDGLRKKVDDSERALQAYKERENIVDV 131
                A  AN  A  Y+   LE +L++    + WL  R++ LR ++ +SE  LQA+ + E +VDV
NP_230584: 188  AKIANEIARVYMDSHLEAKLEVELKANTWLNTRMEELRTQLRESEAKLQAFLQAEGLVDV 247

DP-A5:     132  KTSQSLEVQQVNVASSQLTEARSERMRMQSIYDQVKS 168
                +  + L  Q++   +SQ+ +AR  R+  +++Y    S
NP_230584: 248  QGVEGLATQELEELTSQMNKARDRRVAAETLYQVANS 284
```

Figure 5.13: Sequence DP-A8

Figure 5.13: a) Annotated sequence of DP-A8 amplified from soil 1a with primer pair Brint/3CS. Arrows denote primer sequences. The amino acid sequence of orf1 is shown in regular type and orf2 is represented in bold type.

```
      3CS Primer
      ──────────►
  1   AAGCAGACTTGACCTGACGAAATCATACGTTGTGTGGAGCTAAAACGTACCGTTTGACGTACGTACGCT   69

 70   ATGACGTACAATATCGTCAATAACTCGTGGAGGTGTGTCATGACTACGATTTCGGCCAGTGAGGCCCGT  138
      M  T  Y  N  I  V  N  N  S  W  R  C  V  M  T  T  I  S  A  S  E  A  R
      orf1
139   GCCAACCTGTATCGTCCGATCGATGAGGCGGCATCCTCTCACCAGCCACTGCTTATCTCGGGCAAGCGT  207
      A  N  L  Y  R  P  I  D  E  A  A  S  S  H  Q  P  L  L  I  S  G  K  R

277   AACAAAGCTGTGCTGATCTCCGAGGAGGATTGGACGGCCATTCAGGAGACTCTTTTCCTTCTGTCCGTA  276
      N  K  A  V  L  I  S  E  E  D  W  T  A  I  Q  E  T  L  F  L  L  S  V

277   CCTGGAATGCGTGAGTCTATCCGGGAAGGGATGGATTCCCCTGTTGATGAGTGTAACGAGGATCTGGAT  345
      P  G  M  R  E  S  I  R  E  G  M  D  S  P  V  D  E  C  N  E  D  L  D
                                                                        L
                                                                      orf2
346   TGGTGAGTTGGCGACTCGTATATACAAAACAAGCGCGAAAAGACGCGAAGAAGCTGGCCTCAAGTGGTC  414
      W  *
        V  S  W  R  L  V  Y  T  K  Q  A  R  K  D  A  K  K  L  A  S  S  G  L  483

415   TGAAGCCGAGAGCCCAGGAATTGTTGGATATACTCGCCGAAAATCCTTTCCAGAAGCCACCTCCTTTCG  552
        K  P  R  A  Q  E  L  L  D  I  L  A  E  N  P  F  Q  K  P  P  P  F  E

553   AGAAGCTGGTGGGGGATCTTTCGGGCGCCTACTCGCGGCGGATCAACATTCAACGTTGGTTGGTGTATC  621
        K  L  V  G  D  L  S  G  A  Y  S  R  R  I  N  I  Q  R  W  L  V  Y  Q

622   AGGTGCTGGAGGAAGGGCGAGTCGTCAAGGTTCTGGACCAGTTGC                          667
        V  L  E  E  G  R  V  V  K  V  L  D  Q  L
                        ◄──────────────────────
                                Brint primer
```

Figure 5.13.b.i) Alignment of DP-A8 orf1 with unknown environmental sequence, accession number EAG92777.

Identities = 44/80 (55%), Positives = 60/80 (75%)

```
DP-A8:    14 MTTISASEARANLYRPIDEAASSHQPLLISGKRNKAVLISEEDWTAIQETLFLLSVPGMR 73
             M +IS +EAR  L+  +DE A SH P+ I GKR+ AVL+SE+DW AIQETL+L +VPGMR
EAG92777: 1 MASISVTEARKRLFALVDEVAESHTPVEIHGKRSNAVLVSEDDWRAIQETLYLTAVPGMR 60

DP-A8:    74 ESIREGMDSPVDECNEDLDW 93
             ESI +GM +P+ + +E+  W
EAG92777:61 ESIVDGMATPIGDLSEEPGW 80
```

b.ii) Alignment of DP-A8 orf1 with conserved hypothetical protein of *Synechococcus* sp. WH 8102, accession number NP_896671.

Identities = 43/80 (53%), Positives = 59/80

```
DP-A8:    14 MTTISASEARANLYRPIDEAASSHQPLLISGKRNKAVLISEEDWTAIQETLFLLSVPGMR 73
             M +IS +EAR  L+  +DE A SH P+ I GKR+ AVL+SE+DW AIQETL+L ++PGMR
NP_896671: 1 MASISVTEARKRLFALVDEVADSHTPVEIHGKRSNAVLVSEDDWRAIQETLYLAAIPGMR 60

DP-A8:    74 ESIREGMDSPVDECNEDLDW 93
             ESI +GM +P  + +E+  W
NP_896671:61 ESIVDGMATPTSDLSEEPGW 80
```

Figure 5.13.c.i) Alignment of DP-A8 orf2 with unknown environmental sequence, accession number EAA81077.

Identities = 51/81 (62%), Positives = 69/81 (85%)

```
DP-A8:     1  LVSWRLVYTKQARKDAKKLASSGLKPRAQELLDILAENPFQKPPPFEKLVGDLSGAYSRR 60
              +V+++L YTKQA+KDAKK+ SSGLKP+ QELL+IL  +P+Q  PP+E+L+GDLSGAYSRR
EAA81077:  1  MVNYKLAYTKQAQKDAKKITSSGLKPKVQELLEILENDPYQSYPPYERLIGDLSGAYSRR 60

DP-A8:    61 INIQRWLVYQVLEEGRVVKVL 81
             INIQ  LVYQ+LE  + +K++
EAA81077: 61 INIQHRLVYQILESEKTIKII 81
```

c.ii) Alignment of DP-A8 orf2 with conserved hypothetical protein of *Nitrosomonas europaea*, accession number NP_841401.

Identities = 56/78 (71%), Positives = 65/78 (83%)

```
DP-A8:     4 WRLVYTKQARKDAKKLASSGLKPRAQELLDILAENPFQKPPPFEKLVGDLSGAYSRRINI 63
             W L YT QA+KDAKKLASSGLK +A+ELL ++  NP+Q PPP+EKLVGDL+GA SRRINI
NP_841401: 2 WELRYTHQAQKDAKKLASSGLKDKAEELLAVVRNNPYQTPPPYEKLVGDLAGACSRRINI 61

DP-A8:    64 QRWLVYQVLEEGRVVKVL 81
             Q  LVYQVLE  R+VKVL
NP_841401:62 QHRLVYQVLERERIVKVL 79
```

Products of approximately 1 kb in size were recovered from DNA extracted from bed dust and Antarctic samples MVT7 and MVT11 (transect samples), MV17 and Brat2 following a single round of amplification. All products were cloned and a representative clone from each library was selected for sequencing. Clones were designated as follows: 53 MVT7-2, 53 MVT11-1, 53 MV17-A, 53 Brat2-1/2, 53 BD6/9. All clones were found to possess identical inserts of 959 bp with sequences sharing >99% nucleotide sequence identity. BLAST analysis of this sequence identified the closest database relative as the cloning vector pHRBar-6 (accession number AY283058), a broad-host-range conditional lethal plasmid, sharing 90% nucleotide identity in the region of 264 amino acids corresponding to the *aadA* gene. A single orf of 286 aa was identified for all clone sequences that was found to share 86% identity and 93% similarity at the amino acid level to a streptomycin adenlytransferase of the plasmid R100 (accession number BAA78801) A pairwise alignment of the orf identified from sequence MVT11-1 and the *aadA* gene of plasmid R100 is shown in Figure 5.13.a). RFLP analysis using two tetranucleotide restriction enzymes was performed on eight clones from each library in order to determine if all clones possessed the same insert. All clones analyzed from the MVT7, MVT11, and MV17 libraries were identical. Two RFLP types (highly similar) were identified among the Brat2 clones; however, sequence analysis of a clone belonging to the second RFLP type (53Brat 2-2) revealed only six nucleotide substitutions. This resulted in a total of five amino acid substitutions compared to the Brat2-1 orf but did not alter the closest database relative or level of amino acid sequence identity. A second Bed Dust clone (53BD9) was also selected for sequencing based upon RFLP typing. The sequence of 961 bp shared 99% nucleotide identity to a region corresponding to the *aadA2* gene of the plasmid pCG4 of *Corynebacterium glutamicum* (accession number AAF9763) (Nesvera *et al.*, 1998). An orf of 197 aa was identified sharing 98% identity and 99% similarity at the amino acid level to the product of the *aadA2* gene, an aminoglycoside adenyltransferase, present on a class I integron of *Vibrio cholerae* (Figure 5.14.b) (Dalsgaard *et al.*, 2000).

Figure 5.14: orf's identified from Int5CS/Int3CS sequences
Figure 5.14:a) Alignment of orf identified from Int5CS/ Int3CS gene cassette sequences amplified from Bed Dust, MVT7, MVT11, MV17 and Brat2 DNA samples with the *aadA* gene of plasmid R100. Identical sequences were recovered from each sample and the MVT11 sequence is represented in the alignment.

Identities = 244/283 (86%), Positives = 264/283 (93%)

```
MVT11: 1   MLWSSNDVTQQGSRPKTKLDIMRDSVTAEISTQLSKVLSVIEHHLEPTLLAVHLYGSAVD 60
           MLWSSNDVTQQGSRPKTKL+IMR++V AE+STQLS+V+ VIE HLEPTLLAVHLYGSAVD
R100:   53 MLWSSNDVTQQGSRPKTKLNIMREAVIAEVSTQLSEVVGVIERHLEPTLLAVHLYGSAVD 112

MVT11: 61  GGLKPYSDIDLLVTVTARLDDTTRRALFNDLLEVSAFPGESEILRAIEVTIVVHEDIMPW 120
           GGLKP+SDIDLLVTVT RLD+TTRRAL NDLLE SA PGESEILRA+EVTIVVH+DI+PW
R100: 113  GGLKPHSDIDLLVTVTVRLDETTRRALINDLLETSASPGESEILRAVEVTIVVHDDIIPW 172

MVT11:121  RXPAKRELQFGEWQRNDILAGIFEPATIDIDLAILLTKAREHSVALVGPAAEELFDPVPE 180
           R PAKRELQFGEWQRNDILAGIFEPATIDIDLAILLTKAREHSVALVGPAAEELFDPVPE
R100: 173  RYPAKRELQFGEWQRNDILAGIFEPATIDIDLAILLTKAREHSVALVGPAAEELFDPVPE 232

MVT11:181  QDLIKALNETLKLWNSQPDWAGDERNVVLTLSRIWYSAATGKIAPKDVAANWAMEHLPAQ 240
           QDL +ALNETL LWNS PDWAGDERNVVLTLSRIWYSA TGKIAPKDVAA+WAME LPAQ
R100: 233  QDLFEALNETLTLWNSPPDWAGDERNVVLTLSRIWYSAVTGKIAPKDVAADWAMERLPAQ 292

MVT11:241  HQSVLLEARQAYLGQEEDRSVLHADKLEEFIHFMKSEITKVLG 283
           +Q V+LEARQAYLGQEEDR    AD+LEEF+H++K EITKV+G
R100: 293  YQPVILEARQAYLGQEEDRLASRADQLEEFVHYVKGEITKVVG 335
```

Figure 5.14.b) Alignment of orf identified from sequence 53BD9 with the product of the *aadA2* gene, present on a class I integron of *Vibrio cholerae* (accession number AAF97630).

Identities = 195/197 (98%), Positives = 196/197 (99%)

```
53BD9: 1   MRVAVTIEISNQLSEVLSVIERHLESTLLAVHLYGSAVDGGLKPYSDIDLLVTVAVKLDE 60
           MRVAVTIEISNQLSEVLSVIERHLESTLLAVHLYGSAVDGGLKPYSDIDLLVTVAVKLDE
aadA2: 1   MRVAVTIEISNQLSEVLSVIERHLESTLLAVHLYGSAVDGGLKPYSDIDLLVTVAVKLDE 60

53BD9: 61  TTRRALLNDLMEASAFPGESETLRAIEVTLVVHDDIIPWRYPAKRELQFGEWQRNDILAG 120
           TTRRALLNDLMEASAFPGESETLRAIEVTLVVHDDIIPWRYPAKRELQFGEWQRNDILAG
aadA2: 61  TTRRALLNDLMEASAFPGESETLRAIEVTLVVHDDIIPWRYPAKRELQFGEWQRNDILAG 120

53BD9: 121 IFEPAMIDIDLAILLTKAREHSVALVGPAAEEFFDPVPEQDLFEALRETLKLWNSQPDWA 180
           IFEPAMIDIDLAILLTKAREHSVALVGPAAEEFFDPVPEQDLFEALRETLKLWNSQPDWA
aadA2: 121 IFEPAMIDIDLAILLTKAREHSVALVGPAAEEFFDPVPEQDLFEALRETLKLWNSQPDWA 180

53BD9: 181 GDERNVVLTLSRIWYRS 197
           GDERNVVLTLSRIWY +
aadA2: 181 GDERNVVLTLSRIWYSA 197
```

Additional products of an identical size were recovered from Antarctic samples UWV1, MVT5 and MV16 using primer pair Int5CS and Int3CS, although the decision was taken not to clone these due to extensive characterization of the sequences described above. A further three Antarctic DNA samples (MVT3, MVT9 and MV8/12) were amplified with primers Int5CS and Int3CS but failed to yield products. In an attempt to increase the diversity of integron products recovered, primer Int5CS was used in combination with the reverse primer HS286 that targets the 59-base element flanking gene cassettes. Bed dust DNA was used as an initial target for PCR amplification with Int5CS / HS286 as it was considered a positive environmental control for gene cassettes. A single product of identical size to the Int5CS/Int3CS product was recovered; however, for this reason the product was not characterized further. Given that amplification with primer HS286 of the Int5CS/Int3CS cassette sequence previously characterized, would result in a product 32 nucleotides shorter in length, the Int5CS/HS286 product was not considered to be a different gene cassette.

An additional combination of primers targeting the 59-base elements was also investigated. Primer pair HS286 and HS287 permit the recovery of gene cassettes present in a multicassette array. Amplification of a product is dependent on a gene cassette being bordered by two 59-bes – the reverse primer, HS286, anneals to the 59-be of a specific gene cassette whilst the forward primer, HS287, targets the 59-be site of a preceding cassette. Bed dust DNA was first used as template for this PCR, employing High Fidelity Taq Polymerase and three different annealing temperatures – 55, 57 and 59 °C. This resulted in the amplification of three faint and diffuse bands ranging from approximately 300-500 bp in size. These products were not considered suitable for cloning due to the lack of specificity of the PCR in combination with previous problems encountered with the amplification of spurious products. A second PCR was performed reproducing the original conditions described by Stokes *et al.*, (2001), although this result was found to be less satisfactory with considerably greater non-specific annealing. DNA extracted from Antarctic samples MVT5, MVT11 and MV17 (from which Int5CS/ Int3CS products were recovered) was also used as template for amplification with primers HS286 and HS287. It was predicted that, given the expected lower diversity in

mixed templates of environmental DNA extracted from Antarctic samples compared to bed dust, the risk of non-specific annealing of the primers would be reduced. Optimal annealing temperatures of 55 and 57°C were used during the PCR but failed to yield any amplification products.

5.4: Amplification of 3' Conserved Segment

The 3'CS of class 1 integrons comprises either a *qacE* gene, conferring resistance to quaternary ammonium compounds and disinfectants, or the following combination of three genes: *qacEΔ1*, *sulI*, encoding resistance to sulphonamides, and *orf5*, the function of which is unknown (Stokes & Hall, 1989; Paulsen *et al.*, 1993; Radstrom *et al.*, 1994). The product of the *qacEΔ1* gene is a partially functional deletion derivative of *qacE*, in which the C-terminal 16 amino acids have been deleted. This deletion event is postulated to have occurred through the insertion of a DNA sequence carrying both the *sulI* gene and orf5 between codons 94 and 95 of the *qacE* gene (Paulsen *et al.*, 1993).

The *qacE* gene was first identified on plasmid R751, a broad host range plasmid isolated from *Klebsiella aerogenes*, as part of an integron bearing a complete *qacE* gene cassette and no *sulI* resistance gene (Paulsen *et al.*, 1993). The *qacE* gene has also subsequently been identified within the 3'CS of integrons isolated from Gram-negative bacteria cultivated from an estuarine environment and clinical isolates of *Pseudomonas aeruginosa* (Rosser & Young, 1999; Kazama *et al.*, 1998). The 3'CS structure comprising of the *qacE* gene only is proposed to represent the ancestral integron (Bissonnette & Roy, 1992). Evolution of the 3'CS containing *qacEΔ1*, *sulI* and orf5, that is frequently found associated with multi-resistant integrons isolated from clinical environments, occurred through the insertion of the *sulI*-orf5 DNA sequence into the 3' end of *qacE*. The latter genetic combination appears to predominate; however, this most likely reflects a bias towards studies of integron carriage among multi-drug resistant bacteria.

A number of Antarctic DNA samples were screened for the presence of the *qacE/ qacEΔ1* gene and *sulI* genes by PCR employing a threshold of 35 cycles of

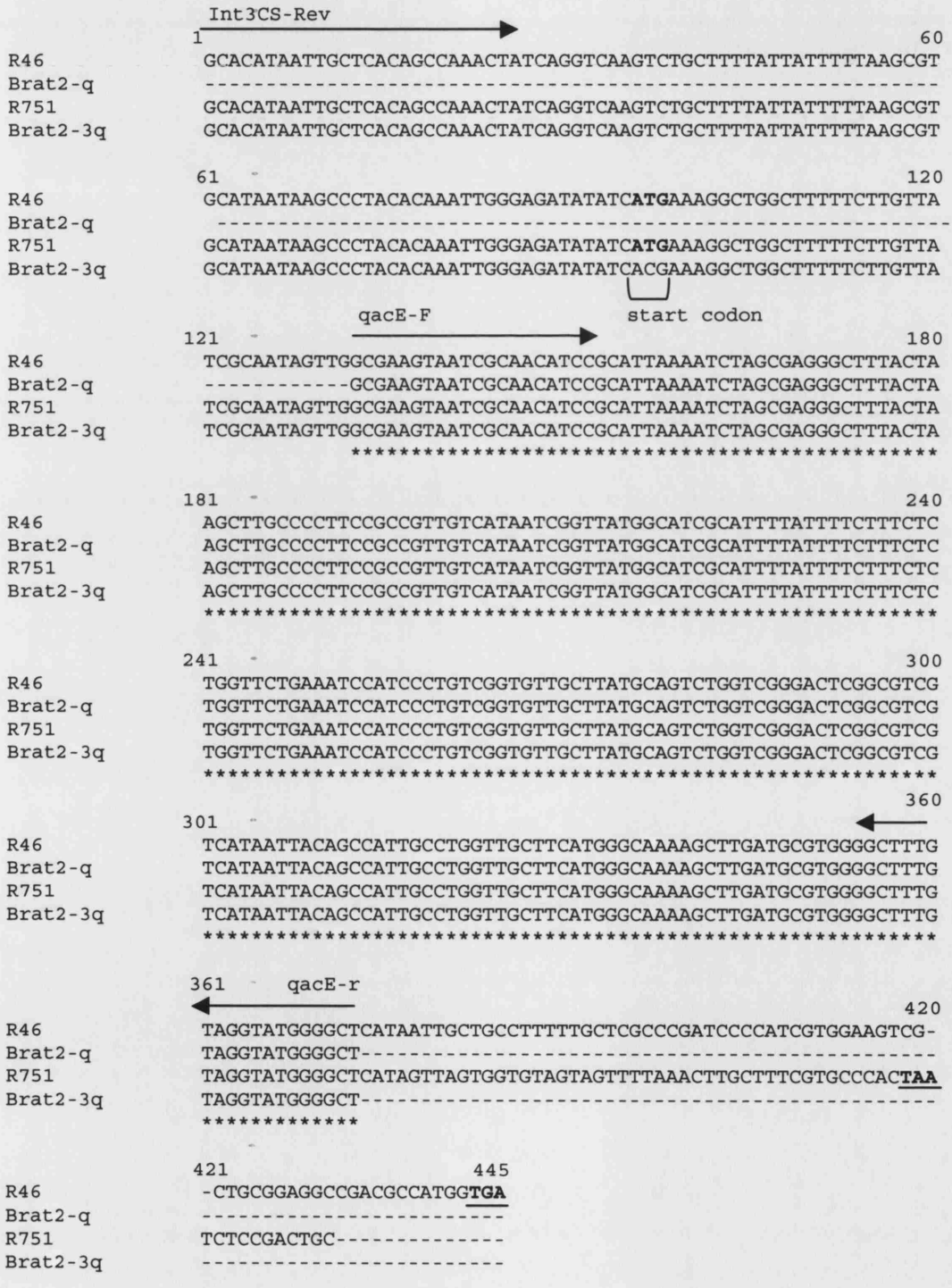
amplification. The objective was to determine if this part of the conserved integron structure typically associated with class 1 integrons, formed part of integrons present in the Antarctic environment.

5.4.1: Presence of the *qacE/qacEΔ1* gene

The presence of the *qacE/ qacEΔ1* gene in environmental DNA samples was assessed by PCR. The primers qacE-f and qacE-r were designed to amplify a partial fragment of 241 bp common to both the *qacE* and *qacE Δ1* genes. Products were amplified from pSa, bed dust DNA and Antarctic samples UWV1 and Brat2 following a threshold of 35 cycles of amplification. Analysis of cloned sequences revealed all to be identical with the exception of a single base change at position 135 of the UWV1 sequence (that did not alter the amino acid sequence), exhibiting 100% nucleotide identity to the *qacE/ qacEΔ1* gene. A total of seven Antarctic DNA samples were negative for the presence of *qacE/ qacEΔ1* and included: UWV4, MV8, MV16, MV17, MVT3 and 5, and pooled samples UWV1-4.

A second PCR reaction was performed on Antarctic DNA samples UWV1 and Brat2 in order to determine if the sequence amplified formed part of the 3'CS of an integron. Primer Int3CS-Rev (Int3CS in the reverse orientation) extending downstream from the variable region of the integron, was used in combination with the qacE-r primer to amplify a product of expected size 373 bp. PCR products of the correct size (373 bp) were recovered from pSa, bed dust and both Antarctic samples, suggesting in each case a *qacE/ qacEΔ1* gene was linked to the 3' CS of an integron. Sequence analysis of the product obtained from Brat2 confirmed it to be genuine, exhibiting 99% nucleotide identity to the corresponding region of a number of class 1 integrons. An alignment of this sequence along with the *qacE* gene of plasmid R751 and the *qacEΔ1* gene located on plasmid R46 is shown in Figure 5.15. Interestingly, the single nucleotide difference present in the Brat2 sequence was found to occur in the start codon of the *qacE/ qacEΔ1* gene. The T-C substitution identified would most likely render this gene non-functional if it is genuine and not a PCR or sequencing error.

Figure 5.15: An alignment of the sequences amplified from Antarctic sample Brat2 with primers i) qacE-f and qacE-r (Brat2-q), and ii) IntCS-Rev and qacE-r (Brat2-3q), with the *qacE* gene of plasmid R751 (accession number NC_001735) and the *qacEΔ1* gene of plasmid R46 (accession number NP_511227). Arrows denote the primer sequences. The start codon is highlighted in bold type and the termination codon is shown in bold type underlined. The Brat2-3q sequence possesses a T-C substitution in the start codon that would most likely prevent expression of this gene. The *qacE* and *qacEΔ1* sequences diverge immediately following the 5'of the qacE-r primer. (The dashed line represents sequence not available.)



5.4.2: Presence of the *sulI* gene

The primer pair of sulI and sulI.rev (Rosser & Young, 1999) was used to amplify a partial fragment of 408 bp of the *sulI* gene. The *sulI* gene is 840 bp in length and encodes a dihydropteroate synthase of 279 aa conferring resistance to sulphonamides (Stokes & Hall, 1989). Amplicons of the correct size and intensity were obtained from pSa, bed dust, soil 1A and pooled Antarctic samples MV8/12 following a single round of PCR amplification, the latter three of which were selected for cloning. Products were also obtained from pooled Antarctic samples UWV1-4 and UWV5-8 of considerably lower yield and the UWV5-8 product was reamplified prior to cloning. Double-stranded DNA sequencing of clones BD-1, 1A-2, MV8/12-1 and UW58-4 was performed. Analysis of sequences obtained from bed dust DNA and Antarctic samples MV8/12 revealed 100% nucleotide identity to the *sulI* gene. Sequences recovered from soil 1A and samples UWV5-8 each contained a single nucleotide difference corresponding to positions 627 and 369 of the *sulI* gene respectively. This resulted in a shared identity to *sulI* of 99% at the nucleotide level and a single amino acid substitution for each sequence.

Products of the correct size were also amplified from Antarctic samples MV17, UWV4, Brat2 and MV16. Direct sequencing of the PCR products obtained from Brat2 and MV17 revealed each to share a 100% nucleotide identity (over 336 bp) with the *sulI* gene. A total of five Miers Valley transect samples along with sample UWV1 from which a *qacE/ qacEΔ1* sequence was obtained, were also screened for the presence of the *sulI* gene but appeared to be negative under the threshold conditions described.

Samples that were *sulI* positive were further investigated for linkage of the *sulI* gene to the 3'CS of an integron. This was carried out by PCR using the primer combination Int3CS-Rev/sulI-r that amplify a product of expected size 1034 bp. Antarctic DNA samples Brat2, MV16, MV17, UWV4, and UWV1-4 were used as templates for amplification but failed to yield products at the threshold set for detection. Therefore, this suggested that the *sulI* genes identified in these samples were not located within the 3'CS of an integron. (Amplification of the plasmid positive control, pSa, did result in a

product of the correct size.) A product of approximately 300bp was amplified from bed dust DNA; however this was not cloned as it was concluded to be too small to represent a *sulI* gene only located within the 3'CS, in the absence of an upstream *qacEΔ1* gene. In a separate PCR investigating linkage of *sulI* to *qacEΔ1*, a product of the expected size (902 bp) was amplified from bed dust DNA using the primer pair, qacE-f and sul1-r. This primer combination permits the recovery of 103 amino acids of the C-terminal of the *qacEΔ1* protein and 199 amino acids of the hydroxyl terminal of the *sulI* protein. Sequence analysis of the cloned product revealed 100% nucleotide identity to the corresponding region of numerous class I integrons. One explanation for this anomaly is that primers Int3CS-Rev and SulI-r annealed to and preferentially amplified a smaller sequence present in bed dust DNA as opposed to the genuine 902 bp sequence, despite demonstrating linkage of 1) *sulI* to *qacEΔ1* and 2) *qacEΔ1* to the 3'CS of an integron. Antarctic DNA sample Brat2, which was the only Antarctic sample positive for both *qacE/qacEΔ1* and *sulI*, was also used as template for amplification with qacE-f and sul1-r. No amplification products were visible providing further support to the contention that the *sulI* gene in this sample was not part of an integron structure.

5.4: Analysis 59-Base Element Sequences

Gene cassettes typically are comprised of a single gene and an integrase-specific recombination site, known as the 59-base element (Stokes & Hall, 1989). The 59-base elements exhibit considerable variation both in sequence and in length, ranging from 57 to 141 bp (Recchia & Hall, 1995, Stokes *et al.*, 1997). A total of five different gene cassettes were recovered from DNA extracted from bed dust, pasture soil and Antarctic Dry Valley soil and their 59-be sequences were identified for comparative analysis. Figure 5.16.a. represents an alignment of these sequences.

All cassettes were found to possess a recombination site of 60 bp in length with the exception of BD3, which was 59 bp. The two most highly conserved features of the 59-be were clearly identifiable in all cases: i) a 7 bp core site conforming to the consensus GTTRRRY located in the right-hand region (at the 3' end) of the 59-be, ii) an inverse core site with the consensus RYYYAAC at the left hand end (the 5' end) of the 59-be. In

each case, the core site and inverse core site were identical possessing the sequences GTTAGAT and GTCTAAC, respectively (Figure 5.16.b). In addition, the eight invariant residues were present in each 59-be. Four 59-be sequence types were evident: 1) DP-A3 and all Int5CS/Int3CS gene cassettes (excluding 53BD-9), 2) BD3, 3) UVW4-B1 and UWV4-B4, 4) 53BD-9. Table 5.2 shows a comparison of closest relatives for 59-be sites and their associated gene cassettes identified from database searches.

Figure 5.16: Analysis of 59-base element sequences

Figure 5.16: a) Alignment of 59-base element sequences from recovered gene cassettes. Sequences UWV4-B1, UWV4-B4, BD3 and DPA3 were amplified with primer pair Brint/ 3CS; the remaining sequences designated '53' were amplified with primer pair Int5CS/ Int3CS. Antarctic sequences include UWV4, Brat2, MV17, MVT11, and MVT7; BD – bed dust, DPA3 – pasture soil 1a. The eight invariant residues characteristic of 59-be's are highlighted in bold type. 59-be groupings are shown in parentheses.

```
UWV4-B1     GTCTAACAATTCATTCAAGCCGACGCCGCTTCGCGGCGCGGCTTAATTCAGGCGTTAGAT  (3)
UWV4-B4     GTCTAACAATTCATTCAAGCCGACGCCGCTTCGCGGCGCGGCTTAATTCAGGCGTTAGAT  (3)
BD3         GTCTAACAATGCGTTC-AGCCGATGCCGCTTCGCGGCACGACTTAACTTCGGCGTTAGAT  (2)
DPA3        GTCTAACAATGCGTTCAAGCCGATGCCGCTTCGCGGCACGACTTAACTTCGGCGTTAGAT  (1)
53BD6       GTCTAACAATGCGTTCAAGCCGATGCCGCTTCGCGGCACGACTTAACTTCGGCGTTAGAT  (1)
53Brat2-1   GTCTAACAATGCGTTCAAGCCGATGCCGCTTCGCGGCACGACTTAACTTCGGCGTTAGAT  (1)
53Brat2-2   GTCTAACAATGCGTTCAAGCCGATGCCGCTTCGCGGCACGACTTAACTTCGGCGTTAGAT  (1)
53MV17-A    GTCTAACAATGCGTTCAAGCCGATGCCGCTTCGCGGCACGACTTAACTTCGGCGTTAGAT  (1)
53MVT11-1   GTCTAACAATGCGTTCAAGCCGATGCCGCTTCGCGGCACGACTTAACTTCGGCGTTAGAT  (1)
53MVT7-2    GTCTAACAATGCGTTCAAGCCGATGCCGCTTCGCGGCACGACTTAACTTCGGCGTTAGAT  (1)
53BD9       GTCTAACAATTCGTTCAAGCCGACCGCGCTACGCGCGGCGGCTTAACTCCGGCGTTAGAT  (4)
            ********** * ****  * **   **** ****   ** ***** *  **********
```

Figure 5.16.b) Annotation of 53Brat2-1 59-be site. Asterisks mark the eight invariant residues.

```
                core site                                          inverse core site
                ---------->                                        <----------
53Brat2-1   GTCTAACAATGCGTTCAAGCCGATGCCGCTTCGCGGCACGACTTAACTTCGGCGTTAGAT
               ***          *           *                          ***
```

The recovery of gene cassettes harboring an *aadA1* gene from bed dust and Antarctic DNA samples, which were most closely related to the *aadA1* gene present on Tn*21* (and elsewhere) of plasmid R100, lead to the hypothesis that these cassettes may be located on Tn*21*. To test this hypothesis, two separate PCR reactions were designed to examine for linkage of the gene cassettes to Tn*21*. A forward primer designated Tn21f, specific to the *tnpR* gene encoding a transposition resolvase of Tn*21*, was used in combination with two reverse primers: i) ant4, complementary to the *aadA1* gene; ii) Brint5 (primer Int5CS in the reverse orientation). Plasmid R26 was used in this instance as the positive PCR control. R26 possesses an identical profile for resistance pattern and restriction enzyme analysis as plasmid R1033 (J.M Ward, Personal Communication). Plasmid R1033 harbours the transposon Tn*1696*, a derivative of Tn*21*, that has the addition of an *aacC1* gene encoding an acetyltransferase preceding the *aadA* gene (Wohlleben *et al.*, 1989). Thus, amplification of pR26 with primers Tn21f and ant4 results in recovery of part of the *tnpR* gene (encoding a resolvase), the *tnpM* gene (encoding a putative transposition regulator), *intI1* (encoding the integron integrase), the *aacCI* gene cassette and part of the aadA gene cassette; the expected size of product is approximately 2800 bp (Liebert *et al*, 1999). Amplification of pR26 with primers Tn*21* and Brint5 permits recovery of a smaller fragment of approximately 1800 bp that does not include the two resistance gene cassettes described above.

PCR templates investigated were DNA extracted from bed dust, MVT7, MVT11, MV16 and Brat2 all of which yielded *aadA1* gene cassettes with primers Int5CS and Int3CS. However, no amplification products were visible with either combination of PCR primers following 35 cycles of amplification and it was concluded that the *aadA1* gene cassette identified in these samples did not appear to be located on Tn*21*.

Table 5.2: Summary of analysis of gene cassette sequences.

Gene Cassette	59-be group	orf identified	Closest relative of orf	% identity (% similarity)	length of 59-be	closest relative of 59-be	% identity (bp overlap)
DPA3	1	aadA1	*aadA* gene, plasmid R100 (NP_052897)	87 (92)	60	Tn21 recombination site	89 (58)
53BD6 53Brat2-1 53Brat2-2 53MV17-A 53MV17-D 53MVT11-1 53MVT7-2	1	aadA1	*aadA* gene, plasmid R100 (NP_052897)	86 (93)	60	Tn21 recombination site	89 (58)
BD3	2	aadA1	*aadA* gene, plasmid R100 (NP_052897)	85 (93)	59	Tn21 recombination site	87 (58)
UWV4-B1 UWV4-B4	3	aadA6	*aadA6* gene, In51, *Pseudomonas aeruginosa* (AF140629)	98 (98)	60	59-be of aadA6 gene cassette from *Pseudomonas aeruginosa* class I integron (AY460181, AF140629)	100 (58)
53BD9	4	aadA2	aadA2, class 1 integron of *Vibrio cholerae* (AAF97630)	98 (99)	60	59-be of aadA2 gene cassette present in numerous class I integrons	100 (60)

5.5. Prevalence of Plasmids in Antarctic soils

As linkage of the *aadA* gene cassette recovered from Dry Valley soils to transposon Tn*21* could not be demonstrated, the possibility that this cassette might reside on a plasmid was explored. PCR amplification of different replicon-specific DNA regions of plasmids belonging to the incompatibility (Inc) groups IncP, IncN, IncW and IncQ was performed on DNA extracted from Dry Valley soils (Gotz et al., 1996; Greated & Thomas, 1999). The objective of this research was to assess whether these specific markers for these plasmid groups were present in the Dry Valley environment, and as such, might serve as vehicles for the carriage and dissemination of an integron harbouring the *aadA* gene cassette. The incompatibility groups selected for investigation were chosen primarily for their broad host range and the availability of replicon-specific PCR primers. Members of the IncP, IncN and IncW groups are all conjugative plasmids, whilst IncQ plasmids can be mobilized when transfer functions are supplied *in trans* by a conjugative helper plasmid.

Table 5.2 lists the primers used in this study; all primer sets were successful in amplifying a product of the expected size from a designated positive control of plasmid DNA. A summary of the results obtained from amplification of different Antarctic DNA samples with each of the primer sets is shown in Table 5.3. No IncN, IncW and IncP1 amplification products of the expected size were generated for any of the samples tested. Amplification with IncP9-specific primers resulted in the recovery of a 465 bp sequence from samples UWV1 and Brat2. However, sequence analysis revealed no similarity to the IncP9 *rep* gene; instead, an orf of 90 amino acids was identified exhibiting 50% identity and 71% similarity at the amino acid level to a hypothetical protein of *Nostoc punctiforme* (ZP_00111895) of 92 amino acids in length. Products of approximately the expected size were obtained from four Antarctic DNA samples following amplification with IncQ-specific primers as shown in Figure 5.17. Due to the presence of multiple bands, the products were to be further analyzed by Southern blot hybridization in order to confirm amplification of the IncQ *oriV* gene. However, unfortunately this work was not carried out due to time constraints. Therefore, it is impossible to determine whether

Table 5.3: Details of primers used for amplification of replicon-specific plasmid sequences.

Inc Group	Region	Primer Sequence (5'-3')	PCR positive control	Product Size (bp)	Annealing temp. (°C)	Primer position	Accession no.	Ref.
IncN	*rep*1 *rep*2	AGTTCACCACCTACTCGCTCCG CAAGTTCTTCTGTTGGGATTCCG	pRN3	164	55	32164-32186 32328-32305	AY046276	Gotz *et al.* 1996.
IncP9	*rep*F *rep*R	CCAGCGCGGTACWTGGG GTCGGCAICTGCTTGAGCTT	TOL	399	60	1774-1790 2172-2153	AJ344068	Greated and Thomas, 1999.
IncP1	*trfA2* 1 *trfA2* 2	CGAAATTCRTRTGGGAGAAGTA CYGTTGCAATGCACCAGGTC	pR26	241	57	1334-1355 1574-1555	X00713	Gotz *et al.* 1996
IncQ	*oriV* 1 *oriV* 2	CTCCCGTACTAACTGTCACG ATCGACCGAGACAGGCCCTGC	pQR445	436	60	2328-2347 2763-2743	M28829	Gotz *et al.* 1996
IncW	*oriV*1 *oriV*2	GACCCGGAAAACCAAAAATA GTGAGGGTGAGGGTGCTATC	pSa	1140	58	1203-1184 64-83	U12441	Gotz *et al.* 1996

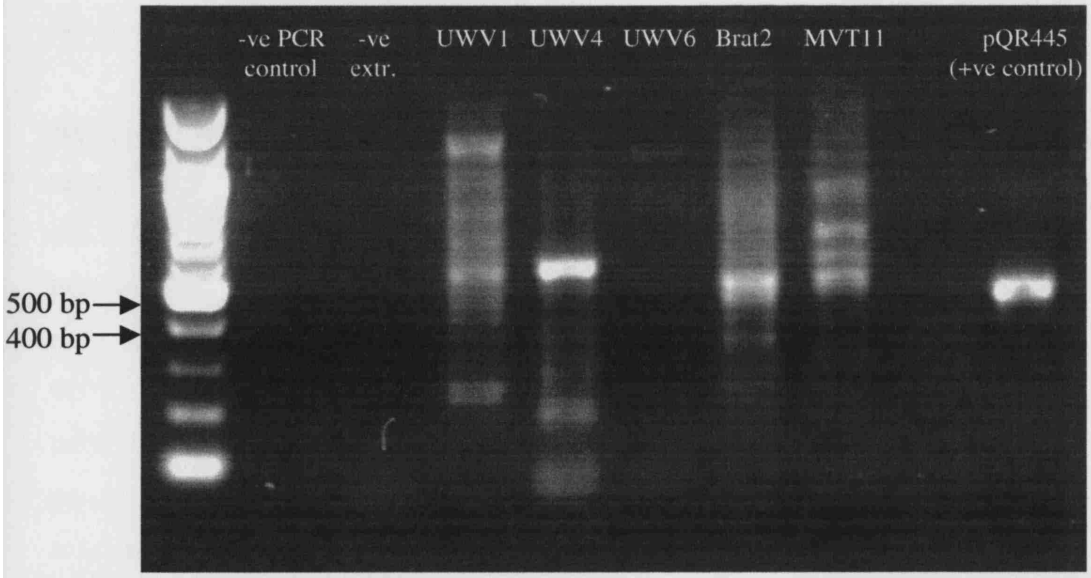
Table 5.4: Summary of results of plasmid screening by PCR of Antarctic DNA samples.

Plasmid incompatbility group	Brat2	MVT5	MVT11	MV16	UWV1	UWV4	UWV6
IncN	-	-	-	-	-	-	-
IncP1	-	-	-	-	-	NT	-
IncP9	-	-	-	-	-	-	-
IncQ	+ (?)	NT	+ (?)	NT	+ (?)	+ (?)	-
IncW	-	-	-	-	-	NT	-

NT: Not tested.
+ (?): PCR product of expected size generated but not confirmed by sequence analysis.

these products were indeed genuine, particularly in light of the amplification of spurious products with IncP9-specific primers.

Figure 5.17: Result of PCR amplification of Antarctic DNA samples with IncQ-specific primers. The expected product size is 436 bp as represented by pQR445, the positive control of plasmid DNA. Products of identical size to the positive control were amplified from samples Brat2 and MVT11, whilst sample UWV1 and UWV4 gave rise to products of slightly larger size; sample UWV6 was negative. The presence of multiple PCR products makes it impossible to state with certainty whether these products were genuine. (-ve extr: negative extraction control)



5.6: Integron Screening of Bacterial Isolates

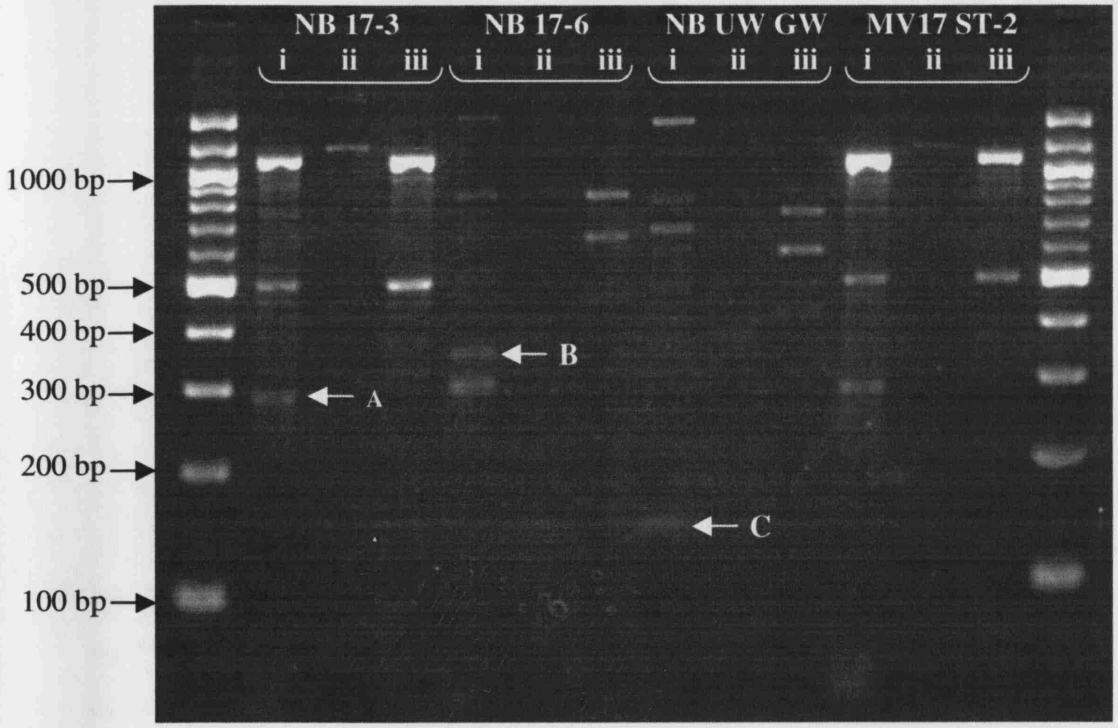
Bacterial isolates cultivated from Antarctic samples (Section 3.2) were investigated for the presence of integrons by PCR. Integrons residing on mobile elements such as plasmids or transposons or located within the bacterial chromosome, were screened for using primers Brint and 5CS-Rev. These primers were successful in the amplification of a 117 bp fragment of the 5'CS from community DNA extracted from Antarctic samples (Section 5.2.2). DNA extracted by heat-denaturation from single colonies was used as template in the PCR; all templates were confirmed to be of suitable quality by an accompanying 16S rRNA gene PCR. A total of 46 individual colonies were screened, representing 21, 19, and 6 colonies cultivated from samples MV16, MV17 and UWV4 respectively. Sixteen colonies were grown on media containing antibiotic. The 29 bacterial isolates identified in Section 3.2 were among those screened, and included a further 17 colonies randomly selected on the basis of varying morphology.

All isolates failed to yield a product of the correct size. A product of 138 bp was amplified from isolate NB UW GW, a *Pseudomonas* species; however, sequence analysis did not identify significant similarity to any database sequence. The amplification of products of larger than the expected size was commonplace and a number of isolates gave rise to multiple products, despite employing a relatively high annealing temperature of 60°C. In these cases, the original colony PCR was repeated along with amplification using both the Brint primer and 5CS-Rev primer individually. Sequence analysis of products previously amplified with integron-specific primers had revealed a number to result from amplification with a single primer, particularly in the case of primer 5CS-Rev. Thus, any bands resulting from amplification with primer pair Brint/5CS-Rev that were also present following amplification with the individual primers were discounted as one-primer PCR products.

Three products of particular frequency and intensity that were present only following amplification with both Brint and 5CS-Rev, as shown in Figure 5.18, were selected for further investigation. Sequences of 288 bp amplified from isolate NB 17-2, a *Pseudomonas* sp. and NB17-3, *Psychrobacter aquatica*, were identical to the sequence

recovered with the same primers from DNA extracted from sample UWV1; this corresponded to a region of the *Pseudomonas putida* genome (see section 5.2.2). Isolate NB 17-6, an *Oxalobacter* species, gave rise to a product of 345 bp. Sequence analysis identified a partial orf of 65 commencing from position 145 (position 1 corresponding to the first nucleotide of the 5CS-rev primer). The orf identified here shared 50% identity and 75% similarity at the amino acid level to a putative orf of *Nitrosomonas europaea* (accession number NP_841494).

Figure 5.18: Integron screening of four Antarctic isolates – NB 17-3, NB 17-6, NB UW GW and MV17 ST-2. DNA extracted from a single colony of each isolate was amplified with the following primers: i) Brint/ 5CS-Rev, ii) Brint only, and iii) 5CS-Rev only. Products present within lane 'i' and also identified within either of lanes 'ii' or 'iii' for each isolate were discounted as 'one-primer' PCR products e.g. The products of approximately 500 and 100 bp in size present in lane i for isolate NB 17-3 are also present in lane iii; these represent products amplified with primer 5CS-Rev only. Brint/ 5CS-Rev products selected for further investigation are highlighted by arrows and include: A – 288 bp product amplified from NB 17-3, B – 345 bp product amplified from NB 17-6, and C – 138 bp product amplified from NB UW GW.



5.5: Summary

DNA extracted from a range of Antarctic soils was investigated for the presence of class 1 integrons and their associated gene cassettes using a PCR-based strategy. A threshold limit of 35 cycles of amplification to generate a visible product was observed in this work. The plasmid pSa containing In6 was used as a positive control for integron PCRs, and DNA extracted from pasture soil and bed dust served as positive controls of environmental DNA.

A summary of the products obtained in this study using primers specific to the 5'CS of integrons that includes the integrase gene, is shown in Table 5.5. Integrase sequences exhibiting ≥99% nucleotide identity to *intI1* were amplified from pooled Antarctic samples UWV1-4, UWV5-8 and MV8/12, along with DNA extracted from pasture soil and bed dust. This level of sequence identity is observed among integron integrase genes identified from geographically widespread locations. A further fifteen Antarctic DNA samples were negative for the presence of *intI1*. Sequences of 117 bp, amplified as part of the integron screening PCR, were recovered from Antarctic samples UWV4, MV17 and Brat2, and from bed dust DNA. These sequences were identical to the corresponding region of the 5'CS of class 1 integrons, which similarly possesses the same level of sequence conservation as *intI1*.

Amplification of the variable region of integrons resulted in the recovery of three different gene cassettes in this study, all of which encoded aminoglycoside adenyltransferase enzymes. These enzymes confer resistance to the antibiotics streptomycin and spectinomycin. The products obtained using primers specific to the variable region are summarized in Table 5.6. Two types of gene cassette were detected in Antarctic soil samples. Firstly, an *aadA* gene sharing 86% identity (93% similarity) at the amino acid level to the *aadA1* cassette located within In2 of Tn*21* on plasmid R100 was identified in three Dry Valley soil samples - MV17, MVT7 and MVT11, and soil from Bratina Island (Brat2). PCR products of identical size were obtained from a further 3 out of 6 Antarctic samples investigated but were not characterized further. Minor sequence variants of the same gene cassette were additionally recovered from pasture soil and bed

dust DNA. The second Antarctic gene cassette, amplified from sample UWV4, exhibited 98% identity (98% similarity) to the *aadA6* gene of In51, a class 1 integron identified in *Pseudomonas aeruginosa*. Finally, a third cassette recovered from DNA extracted from bed dust, shared 98% identity (99% similarity) to the *aadA2* gene of a class 1 integron from *Vibrio cholerae*.

Antarctic soil samples were also investigated for the presence of the *qacE/qacEΔ1* gene and the *sulI* gene, which form part of the 3'CS typically associated with class1 integrons identified from clinical resistant bacterial strains. A summary of these results is shown in Table 5.7. Two Antarctic soil samples out of nine screened, UWV1 and Brat2, were positive for the presence of the *qacE/qacEΔ1* gene (the PCR primers used in this work did not discriminate between *qacE* or the truncated version, *qacEΔ1*.) Furthermore, linkage of the *qacE/qacEΔ1* gene to the 3'CS of an integron was demonstrated for each of these samples. *sulI* sequences were identified from four Antarctic samples (UWV5-8, MV8-12, Brat2 and MV17), and a further three samples (UWV1-4, UWV4 and MV16) gave rise to PCR products of the expected size following amplification with *sulI*-specific primers; six samples were *sulI*-negative. Linkage of the *sulI* gene to the 3'CS of an integron could not be demonstrated for five of the *sulI*-positive Antarctic samples investigated. Brat2 represented the only sample that was both *qacE/qacEΔ1* positive and *sulI* positive but no evidence of the linkage of these genes was obtained. All *qacE/qacEΔ1* and *sulI* sequences analysed in this study exhibited ≥99% nucleotide identity to corresponding database sequences.

Linkage of the *aadA* gene-cassette to Tn*21* could not be demonstrated for four of the Antarctic samples investigated from which *aadA1*-type gene cassettes were identified. Furthermore, no replicon-specific PCR products of plasmid incompatability (Inc) groups IncN, IncP and IncW were detected in 6-7 Antarctic samples tested. Products of approximately the expected size were amplified from four Antarctic samples using IncQ-specific primers; however, confirmation of these sequences was not obtained and thus, this result in inconclusive. A total of 46 individual colonies cultivated from Antarctic soil samples and screened for the presence of integrons were negative.

Table 5.5: Products obtained from amplification of DNA samples with primers specific to the 5'CS of integrons.

Clone Name	Sample	Origin of Sample	Primer Pair	Size of product (bp)	orf Identified (aa)	Homology
Int UWV14	UWV1-4	Dry Valley	12b2/ 5CS-Rev	911		99-100% nucleotide identity to *intI1* and corresponding region of 5'CS of class 1 integrons
Int UWV58	UWV5-8	Dry Valley	12b2/ 5CS-Rev	911		
Int MV812	MV8/12	Dry Valley	12b2/ 5CS-Rev	911		
Int 1AL2	1A	Pasture soil	12b2/ 5CS-Rev	911		
Int BD1	BD	Bed Dust	12b2/ 5CS-Rev	911		
MVT9.11	MVT9	Dry Valley	12b2/ 5CS-Rev	819	152	58% identity and 72% similarity to transposase of *T. tengcongensis* (NP_622741)
IntA BD	BD	Bed Dust	IntA/ 5CS-Rev	~1000		100% (500 bp) to *intI1*
IntA-BD2.2	BD	Bed Dust	IntA/ 5CS-Rev	~700	165	99% identity to YleA protein (NP_752677)
UWV4-B5	UWV4	Dry Valley	Brint/ 5CS-Rev	117		100% nucleotide identity to 5'CS of class 1 integrons 100% (85 bp)
Brat2-B5	Brat2	Bratina Island	Brint/ 5CS-Rev	117		
MV17-B5	MV17	Dry Valley	Brint/ 5CS-Rev	117		
BD-B5pp*	BD	Bed Dust	Brint/ 5CS-Rev	117		
UWV1 B5-2	UWV1	Dry Valley	Brint/ 5CS-Rev	288	54	88% identity and 94% similarity to a hypothetical protein of *P. putida* (NP_746924)
Cass-MV17	MV17	Dry Valley	Cass1/ 5CS-Rev	671		68% nucleotide identity (385 bp) to *Streptomyces avermitilis* genome (AP005048)
MV17-int4	MV17	Dry Valley	intA/ intB	1150	224	50% identity and 64% similarity (220 aa) to hypothetical protein of *P. syringae* (NP_791317)
MV17-int5	MV17	Dry Valley	intA/ intB	1160	386	33% identity and 52% similarity (98 aa) to maturase-related protein *S. maracescens* (AAL51020)

* indicates PCR product directly sequenced

Table 5.6: Products obtained from DNA samples following amplification of the variable region of integrons.

Clone Name	Sample	Origin of Sample	Primer Pair	Size of product (bp)	orf Identified (aa)	Homology
BD3	BD	Bed Dust	Brint/ 3CS	1107	262	85% identity (93% similarity) aadA protein
DP-A3	Soil 1A	Pasture Soil	Brint/ 3CS	1108	262	87% identity (92% similarity) of plasmid
1AR	Soil 1A	Pasture Soil	Brint/ HS286	~1000	262	87% identity (93% similarity) R100 (BAA78801)
UWV4-B1	UWV4	Dry Valley	Brint/ 3CS	1116	281	98% identity/ similarity to aadA6 protein of (AAD54217) from *Pseudomonas* sp.
UWV4-B4	UWV4	Dry Valley	Brint/ 3CS	~1100	81+138 In51	
DP-A5	Soil 1A	Pasture Soil	Brint/ 3CS	512	168	31% identity (52% similarity) to putative exopolysaccharide biosynthesis protein of *V. cholerae*
DP-A8	Soil 1A	Pasture Soil	Brint/ 3CS	597	93 & 81	53% identity (73% similarity) to hypothetical protein of *Synechococcus* (NP_896671) 71% identity (83% similarity) to a hypothetical protein of *N. europea* (NP_841401)
	Soil 1A	Pasture Soil	Brint/ HS286	596	73	75% identity and 87% similarity to the LexA protein of *M. loti* (NP_102395)
53 MVT7-2	MVT7	Dry Valley	Int5CS/ Int3CS	959	264	86%identity and 93% similarity to aadA protein of plasmid R100
53 MVT11-1	MVT11	Dry Valley	Int5CS/ Int3CS	959	264	
53 MV17-A	MV17	Dry Valley	Int5CS/ Int3CS	959	264	
53 Brat2-1	Brat2	Bratina Island	Int5CS/ Int3CS	959	264	
53 Brat2-2	Brat2	Bratina Island	Int5CS/ Int3CS	959	264	
53 BD6	BD	Bed Dust	Int5CS/ Int3CS	959	264	
53 BD9	BD	Bed Dust	Int5CS/ Int3CS	961	197	98% identity and 99% similarity to the aadA2 protein of *V. cholerae*

Table 5.7: Products obtained from DNA samples following amplification of the 3'CS of integrons.

Clone Name	Sample	Origin of Sample	Primer Pair	Size of product (bp)	Homology
	UWV1	Dry Valley	qacE-f/ qacE-r	241	99% } nucleotide identity to *qacE/ qacEΔ1*
	Brat2	Bratina Island	qacE-f/ qacE-r	241	100% } nucleotide identity to *qacE/ qacEΔ1*
	BD	Bed Dust	qacE-f/ qacE-r	241	100% } nucleotide identity to *qacE/ qacEΔ1*
	Brat2	Bratina Island	Int3CS-Rev/qacE-r	373	99% nucleotide identity to corresponding region of 3'CS and *qacEΔ1* of numerous class 1 integrons
MV8/12-1	MV8/12	Dry Valley	sulI/ sulI.rev	408	100% } nucleotide identity to *sulI*
UW58-4	UWV5-8	Dry Valley	sulI/ sulI.rev	408	99% } nucleotide identity to *sulI*
1A-2	Soil 1A	Pasture Soil	sulI/ sulI.rev	408	99% } nucleotide identity to *sulI*
BD-1	BD	Bed Dust	sulI/ sulI.rev	408	100% } nucleotide identity to *sulI*
Brat2-pp*	Brat2	Bratina Island	sulI/ sulI.rev	~400	100% (336 bp) } *sulI*
MV17-pp*	MV17	Dry Valley	sulI/ sulI.rev	~400	100% (336 bp) } *sulI*
BD-QS	BD	Bed Dust	qacE-f/ sulI.rev	902	100% nucleotide identity to *qacEΔ1/ sulI* of class 1 integrons

* indicates PCR product directly sequenced

Given the difficulty encountered in the recovery of integrase gene sequences and the limited diversity of gene cassettes that were obtained from Antarctic soil samples, these results would indicate that integrons are relatively rare in the Antarctic environments investigated in this study. It is conceivable that Antarctic integrase gene sequences may have diverged sufficiently that the primers used in this work were not able to adequately amplify all sequences. Alternatively, the Antarctic gene cassettes that were identified may not in all cases be located adjacent to an integrase gene as expected (i.e. a transposase may instead be located upstream of the gene cassette). Furthermore, integrase gene sequences may be present at numbers beyond the limits of detection employed in this study. This factor may be especially relevant given the lower numbers of microorganisms present in Antarctic Dry Valley soils compared to temperate soils, and is indeed reflected in the relatively low amounts of DNA that were extracted from Dry Valley soil samples, notably those from the Upper Wright Valley. Nonetheless, the recovery of class 1 integrase gene sequences from Antarctic samples sharing 100% nucleotide identity to database sequences is consistent with the sequence conservation and widespread geographical distribution of this gene.

Local environmental conditions (i.e. nutrient status) are known to affect horizontal gene transfer (HGT) and HGT is thought to coincide with periods of pronounced metabolic activity among soil microbial communities (van Elsas & Bailey, 2002). Thus, the limited diversity of integron sequences recovered from Antarctic soils may be influenced by a relatively low level of metabolic activity among these soil microbial communities.

The Antarctic Dry Valley environment may be considered one that is relatively stable given that it is not subject to environmental perturbations due to anthropogenic activities. In this case, there may be less of requirement for gene transfer systems that proliferate adaptive functions in response to environmental change. Given the extreme nature of the Antarctic Dry Valley environment, it is expected that microorganisms inhabiting these soils have evolved the necessary metabolic and physiological capabilities that enable them to survive. Therefore, the role of gene transfer systems such

as integrons in Antarctic ecosystems may be reduced due to the limited selection pressures that exist for other functional attributes. However, the *aadA* gene cassettes recovered from Antarctic soils are considered to be genuine given the prevalence and stable maintenance of these cassettes in integrons from diverse environments including soils. This point will be discussed in more detail in Chapter 6.

Chapter 6
Discussion

6.1: Discussion Overview

The research presented in Chapters 3, 4 and 5 describes an investigation of genome diversity in Antarctic Dry Valley soils. As part of this work, the diversity of life within these soils belonging to the domains Archaea, Bacteria and Eukarya was investigated. Furthermore, Antarctic soils were examined for the presence of class 1 integrons and their associated gene cassettes. In this next section, the bacterial diversity identified within Dry Valley soils using both cultivation and cultivation-independent techniques is discussed. Furthermore, these results are examined with respect to the notion of endemism among bacteria in the Antarctic environment. Finally, the results and implication of the investigation of integron diversity are discussed

6.2: Culturable Bacterial Diversity

Figure 6.1 is a phylogenetic tree detailing the environmental description of the four closest cultivated relatives identified by database searches for each Antarctic isolate. Only 16S rRNA sequences of cultivated bacteria for which an environmental description was available were selected, and where possible, the four closest relatives isolated in separate studies were chosen. The purpose of this task was to examine the environmental distribution of the types of bacteria cultivated from Antarctic soils: i) in order to determine the extent (if any) of endemism in the Antarctic environment, and ii) to identify common environmental parameters underlying the distribution of certain types of bacteria in natural environments.

Of the 22 Antarctic isolates cultivated in this study and represented in Figure 6.1, 12 possessed a bacterium of Antarctic origin as one of their four closest database relatives. In each case, the level of 16S rRNA gene sequence identity was ≥98%. Database relatives of Antarctic origin included isolates identified in the following environments: soil samples collected from the McMurdo Dry Valleys, Ross Island, Scott base, and the vicinity of Lake Zub in the Schirmacher oasis (Reddy *et al.*, Unpublished (a & b)); ice

samples collected from Lake Vostok, a cryoconite hole from Canada Glacier in the McMurdo Dry Valleys, Antarctic pack ice and sea ice brine (Christner *et al.*, 2001 & 2003; Brinkmeyer *et al.*, 2003; Junge *et al.*, 1998); microbial mat samples surrounding lakes within the McMurdo Dry Valleys and Vestfold Hills (Brambilla *et al.*, 2001; Van Trappen *et al.*, 2002); sea water from Terra Nova Bay in the Ross Sea ,and an endolithic community within gypsum crusts (Hughes & Lawley, 2003; Michaud *et al.*, 2004). Thus, it seems likely that a number of bacterial genera are common to Antarctic biotopes, in particular, those affiliated to the phyla *Proteobacteria* and *Bacteroidetes.* (From hereon, reference with respect to the diversity of bacteria cultivated from different environments is made at genus level identification. Whilst it is accepted that considerable phenotypic variation can occur within members of a genus, the purpose here was to identify environments possessing a similar community composition among the culturable fraction of bacteria as that identified for Dry Valley soils in this work.)

In a study conducted by Christner *et al.*, (2003a) of bacterial diversity in an Antarctic cryconite hole (Canada Glacier, McMurdo Dry Valleys), many of the isolated bacteria and rRNA sequences obtained using culture-independent methods were also highly similar to those previously described in Antarctic Dry Valley environments, notably lake ice and microbial mat samples. A comparison of isolates cultivated from the cryoconite hole and Antarctic dry valley soil samples reveal several genera of bacteria in common, namely *Pseudomonas*, *Janthinobacterium*, *Flavobacterium*, and *Arthrobacter.* This could be explained by inoculation of the cryoconite hole by air-borne deposition of particulate matter blown from the surrounding environment. Similarly, constituents of the lake ice microbial community have been identified in neighbouring terrestrial environments (Gordon *et al.*, 2000). Strong winds within the McMurdo Dry Valleys provide the primary mechanism for dispersal of microorganisms, accounting for the overlap observed in bacterial diversity among different Dry Valley biotopes (Gordon *et al.*, 2000). Furthermore, movement of sea salt aerosols from Antarctic coastal regions to the South Pole is known to occur within about two days (Carpenter *et al.*, 2000).

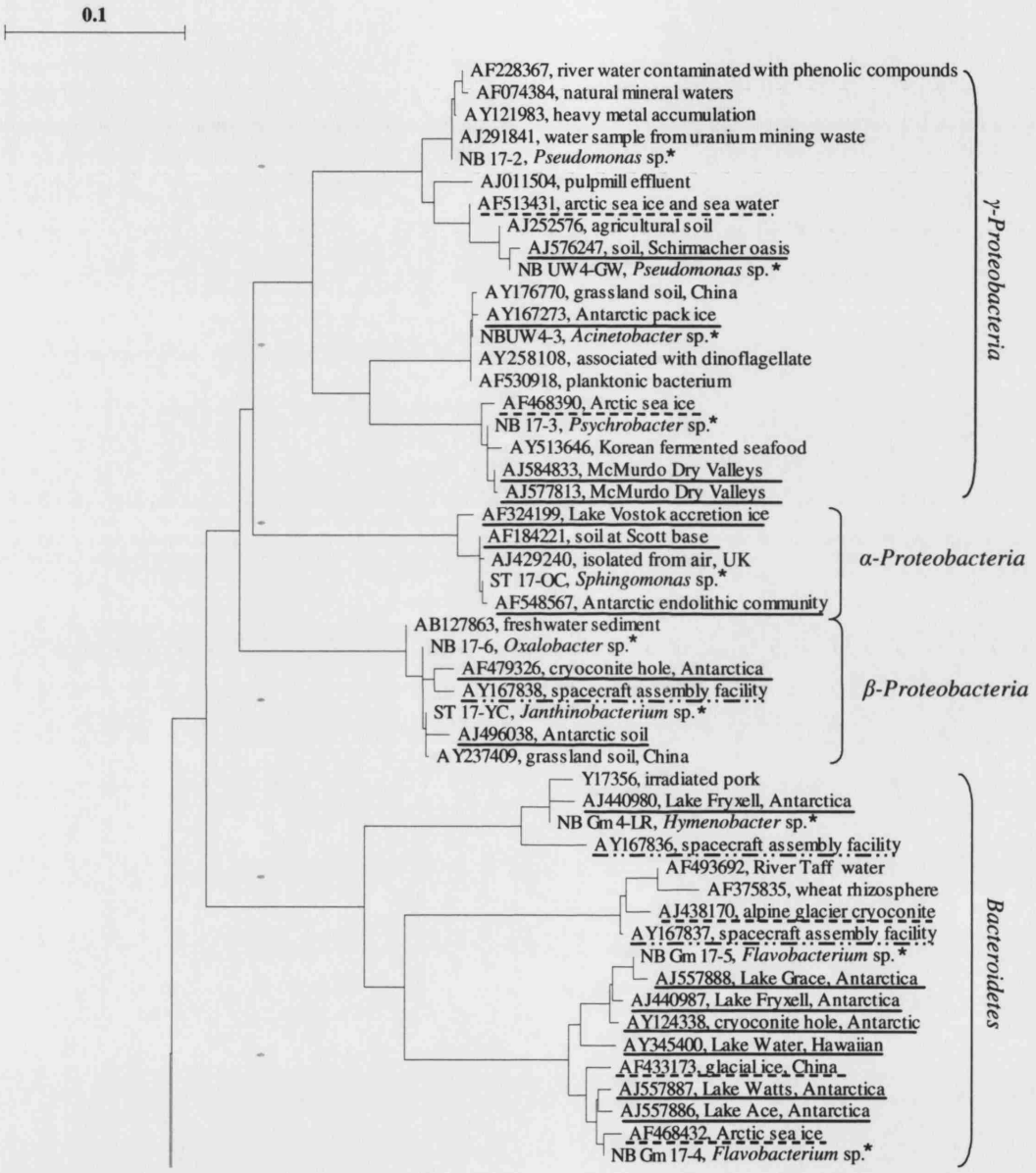
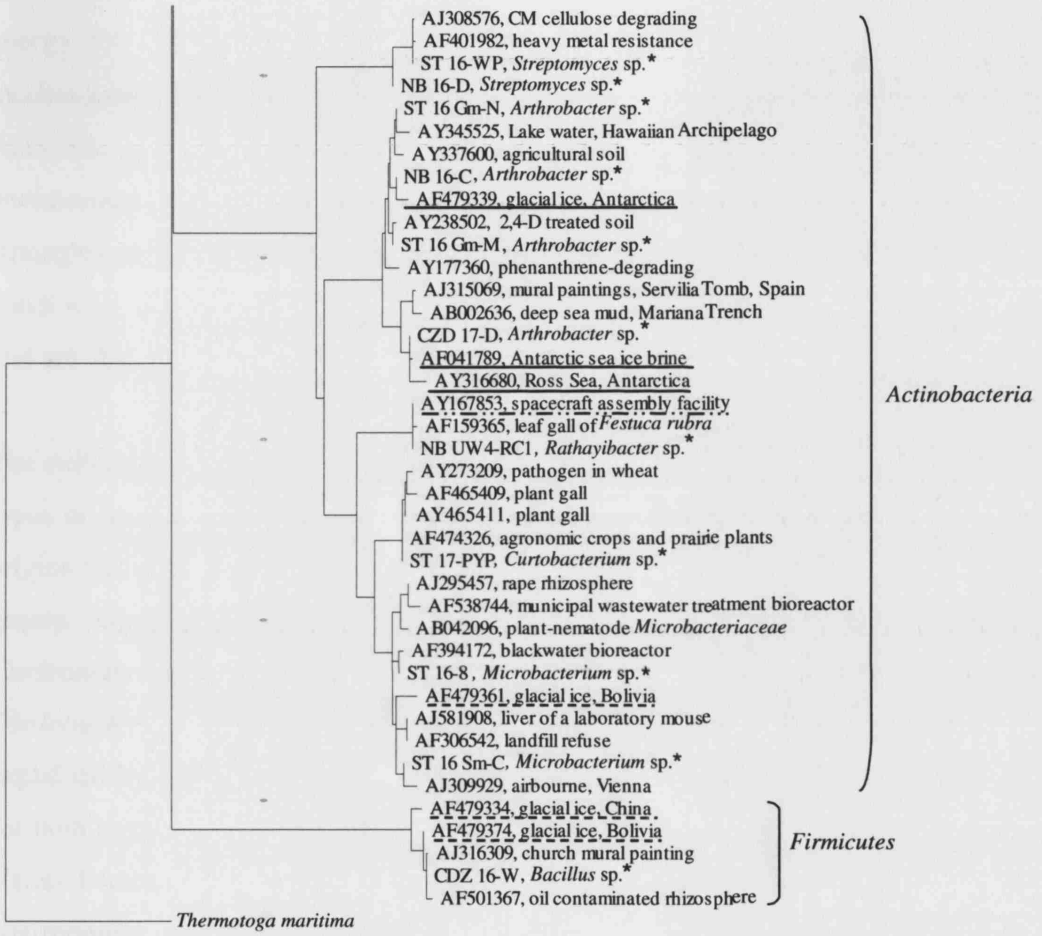


Figure 6.1: Phylogenetic tree detailing the closest database relatives of Antarctic isolates along with their environmental description. Antarctic isolates are marked with an asterisk. Cultivated bacteria of Antarctic origin (including isolates obtained in this work) are underlined; bacteria cultivated from ice of non-Antarctic origin are indicated by the dashed line, and bacteria cultivated from a spacecraft assembly facility are indicated by the dash-dot line. Tree is continued over page.


AJ308576, CM cellulose degrading
AF401982, heavy metal resistance
ST 16-WP, Streptomyces sp.*
NB 16-D, Streptomyces sp.*
ST 16 Gm-N, Arthrobacter sp.*
AY345525, Lake water, Hawaiian Archipelago
AY337600, agricultural soil
NB 16-C, Arthrobacter sp.*
AF479339, glacial ice, Antarctica
AY238502, 2,4-D treated soil
ST 16 Gm-M, Arthrobacter sp.*
AY177360, phenanthrene-degrading
AJ315069, mural paintings, Servilia Tomb, Spain
AB002636, deep sea mud, Mariana Trench
CZD 17-D, Arthrobacter sp.*
AF041789, Antarctic sea ice brine
AY316680, Ross Sea, Antarctica
AY167853, spacecraft assembly facility
AF159365, leaf gall of Festuca rubra
NB UW4-RC1, Rathayibacter sp.*
AY273209, pathogen in wheat
AF465409, plant gall
AY465411, plant gall
AF474326, agronomic crops and prairie plants
ST 17-PYP, Curtobacterium sp.*
AJ295457, rape rhizosphere
AF538744, municipal wastewater treatment bioreactor
AB042096, plant-nematode Microbacteriaceae
AF394172, blackwater bioreactor
ST 16-8, Microbacterium sp.*
AF479361, glacial ice, Bolivia
AJ581908, liver of a laboratory mouse
AF306542, landfill refuse
ST 16 Sm-C, Microbacterium sp.*
AJ309929, airbourne, Vienna
AF479334, glacial ice, China
AF479374, glacial ice, Bolivia
AJ316309, church mural painting
CDZ 16-W, Bacillus sp.*
AF501367, oil contaminated rhizosphere
Thermotoga maritima
Actinobacteria
Firmicutes

These aerosols form cloud condensation nuclei that can serve as vehicles for the transport of microbes and organic matter. Indeed, the carriage of particulate organic matter and marine microbes in snow fall has been proposed as a potential source of energy for bacteria in South Pole snow (Carpenter *et al.*, 2000). Thus, there are mechanisms that enable the widespread dispersal of microorganisms over the entire Antarctic continent. The presence of genera that appear ubiquitous in the Antarctic environment is a reflection of those organisms that have developed the necessary strategies to survive and carry out metabolic activities under the extreme Antarctic conditions; these organisms inhabit a variety of microbial refugia (as described above) and are able to colonize new ecological niches opportunistically.

The culturable diversity identified in Dry Valley soils also bears marked similarity to the types of bacteria isolated from glacial ice and sea ice samples of diverse geographical origins (Bowman *et al.*, 1997; Christner *et al.*, 2003a; Miteva *et al.*, 2004). Bacterial genera common to both environments include *Arthrobacter*, *Microbacterium*, *Curtobacterium*, *Pseudomonas*, *Sphingomonas*, *Acinetobacter*, *Psychrobacter*, *Oxalobacter*, *Flavobacterium*, *Streptomyces* and *Bacillus* species. This may be explained by the use in this study of a single incubation temperature of 4°C that selects for both psychrophilic and psychrotolerant microorganisms; a total of 15 of the 22 Dry Valley bacterial isolates possess a relative identified in a comparative low temperature environment. However, as well as low temperatures, microbial communities inhabiting ice must similarly contend with desiccation and oligotrophic conditions. *Sphingomonas* species appear particularly well-adapted to these conditions, having been isolated from Lake Vostok accretion ice that has been devoid of surface input for over 420,000 years (Christner *et al.*, 2001), and the Guliya ice cap of the Qinhan-Tibetan plateau in Western China that is over 750,000 years old (Christner *et al.*, 2003b).

Interestingly, four Dry Valley isolates possessed bacteria cultivated from a spacecraft assembly facility among their closest relatives. Once again, a number of features are common to this environment and the Antarctic Dry Valleys. The assembly of space craft components is carried out in an ultra clean environment that can be considered an

extreme environment created by the filtered air circulation, desiccation, and low nutrient conditions (Venkateswaran *et al.*, 2001; La Duc *et al.*, 2003). This presents a similarly challenging environment to microbial growth as does a polar desert.

As discussed in Chapter 3 (section 3.2.6), there appears to be little correlation among actinobacterial isolates affiliated to the family *Microbacteriaceae* and the Antarctic Dry Valley environment. Members of this family are commonly identified through their association with plants, either as phytopathogens or as part of the plant endophytic community (Zinniel *et al.*, 2002), a relationship that is evident among the relatives of Antarctic *Microbacteriaceae* isolates (Figure 6.1). However, recent studies have indicated these bacterial types may have a wider distribution in natural environments. Psychrophilic actinobacterial isolates sharing highest levels of sequence similarity to *Rathayibacter tritici* and *Clavibacter michiganensis* have been cultivated from permanently cold ground waters in Finland (Mannisto *et al.*, 2000). The isolates, *Subtercola boreus* and *Subtercola frigoramans*, represent a new genus within the family *Microbacteriaceae.* Growth of these organisms was observed down to -2°C with a temperature optimum of 15-17°C; no growth was observed at temperatures over 30°C.

Additionally, nine ultramicro-size (<0.1 μm^3) strains isolated from freshwater habitats in Europe and Asia (Hahn *et al.*, 2003) also possess *R. tritici* and *C. michiganensis* as their closest described relatives. In this instance, cultivation of isolates was carried out at 15°C. Phylogenetic analysis of 16S rRNA gene sequences revealed the isolates to form a separate cluster with clone sequences recovered from freshwater environments. Furthermore, isolates possessing identical 16S rRNA sequences were cultivated from different continents and exhibited different phenotypic properties. Thus, it would appear that members of the *Microbacteriaceae* have adapted to growth at low temperatures and are likely to be extensively distributed throughout these environments, both in soil and freshwater systems. Their apparent predominant association with plants and the rhizosphere may reflect a sampling bias of cultivation studies, given that members of this family are important plant pathogens.

Finally, isolate NB16-D cultivated from sample MV16 and sharing 93% 16S rRNA identity to its closest database relative, a *Streptomyces* species, possessed the lowest similarity score of all Antarctic isolates. BLAST analysis of the 33 bp section of sequence responsible for this difference revealed it to be identical with the exception of two nucleotide substitutions to the 16S rRNA gene of *Cryobacterium psychrophilum*. *Cryobacterium psychrophilum* is an obligate psychrophilic bacterium isolated from Antarctic soil and affiliated to the family *Microbacteriaceae*; it possesses a growth optimum between 9 and 12°C and does not grow above 18°C (Suzuki *et al.*, 1997).

Interestingly, *C. psychrophilum* possesses an insertion of 14 bp within its 16S rRNA gene that is similar to that identified in only a few members of the *Microbacteriaceae*. These include *Rhodoglobus vestalii* (AJ459101) isolated from a cyanobacterial mat sample collected from a pond near Bratina Island in Antarctica (Sheridan *et al.*, 2003); *Leifsonia rubeus* (AJ438585) and *Leifsonia aureus* (AJ438586) isolated from a pond located in the Wright Valley, Antarctica (Reddy *et al.*, 2003); *Subtercola frigormans* isolated from permanently cold ground water in Finland (Mannisto *et al.*, 2000); isolate QSSC9-20 (AF170755) cultivated from a sublithic community in Antarctic quartz stone (Smith *et al.*, Unpublished), and a clone sequence WCHB1-18 identified in a chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation (Dojka *et al.*, 1998). In each case, the insertion corresponds to the same region of the 16S rRNA gene and is also present in the sequence of Antarctic isolate NB 16-D. Sheridan *et al.*, (2003) postulate that this insertion is unlikely to be a PCR artefact given that the inserted nucleotides can form a number of possible hairpin structures that would stabilize the insertion. Whilst the functional relevance of this insertion is unclear, its presence in isolates cultivated from low temperature environments is particularly intriguing. Furthermore, the identification of a similar insertion in isolate NB16-D of the family *Streptomycetaceae*, raises questions about how such an insertion might have evolved.

6.3: Molecular Assessment of Bacterial Diversity

Total community DNA extracted from two Antarctic Dry Valley soil samples collected from the Upper Wright Valley and Miers Valley was used for direct amplification of 16S rRNA genes and clone library construction. Twenty-four clones were analyzed from the Upper Wright library representing 13 unique phylotypes and 7 bacterial phyla; analysis of 19 clones from the Miers Valley library identified 18 phylotypes within 8 bacterial phyla. Overall, a total of 12 bacterial phyla were represented by 42 partial 16S rRNA gene clone sequences of minimum length 872 bp analyzed in this work. Of these sequences, 22 exhibited <97% 16S rRNA identity to database sequences, inferring a wealth of as yet uncharacterized diversity. This would indicate that, despite possible preconceptions based on the extreme conditions within the Antarctic Dry Valleys, these soils support a complex microbial community. Moreover, these bacterial types most likely represent considerable diversity in terms of physiological and metabolic capabilities.

The data obtained in this study were compared to previous studies from which similar strains or sequences from culture-independent research had been identified. This was done for two reasons; firstly, to provide evidence on the level of endemism within Antarctic Dry Valley soils and secondly, to elucidate possible physiological characteristics that these species may share. The former question will be more fully addressed in section 6.4. However, the initial focus here is on the physiological properties that may well shape the growth of bacteria in Antarctic environments.

The database relatives of Antarctic clone sequences were examined in order to gain some insight into common environmental parameters that may explain the distribution of these organisms. In combination with the phylogenetic affiliation of Antarctic sequences, this allowed tentative inferences to be made regarding their potential physiological characteristics. From this, hypotheses could be devised to form the basis of future investigations examining microbial community function in Dry Valley soils. However, these inferences were made with caution as it is accepted that the physiological properties of an organism can not be deduced from the phylogenetic

position of its 16S rRNA when divergence from a cultivated relative is high (Brambilla *et al.*, 2001). It is noted that where the level of 16S rRNA sequence identity of an Antarctic clone to a cultivated bacterium is ≥98%, identification can be made with confidence at genus level. However, due to the limited resolution of the 16S rRNA gene, identification at species level is not permitted on the basis of this information only. A level of 94-98% 16S rRNA identity equates to variation within a genus, whilst <94% 16S rRNA identity indicates distinct genera. Relatives of Antarctic clone sequences affiliated to the i) the *Proteobacteria*, and ii) the *Actinobacteria* and CFB lineages are represented in the phylogenetic trees shown in Figures 6.2 and 6.3 respectively.

Five of the sequences identified in this culture-independent study belonged within the subclass *Rubrobacteridae*, one of five major lineages of the *Actinobacteria*. Only two species from this subclass have been cultivated *in vitro* namely, *Rubrobacter radiotolerans* and *Rubrobacter xylanophilus*; both organisms are moderate thermophiles possessing temperature optima of 48 and 60°C respectively. These organisms were first isolated from Japanese hot springs and thermally polluted effluent but have also recently been isolated from hot springs in Portugal (Carreto *et al.*, 1996; Ferreira *et al.*, 1999). Furthermore, renewed cultivation attempts have obtained isolates from within this subclass from Australian pasture soil and an earthworm burrow (Janssen *et al.*, 2002; Sait *et al.*, 2002; Furlong *et al.*, 2002). Subsequent studies of other soil bacterial communities have shown that the *Rubrobacteridae* are geographically widespread, with three subgroups proposed based upon 16S rRNA sequence analysis (Holmes *et al.*, 2000). The phylogenetic analysis of Antarctic clone sequences placed four clones within subgroup 1 that includes the thermophilic species, and a single clone within subgroup 2 that contains the soil isolates. The four group 1 clones were most closely related to *R. radiotolerans* and *Rubrobacter*-like organisms identified in soil aggregates, arid Australian soil, an Antarctic cryptoendolithic community, and masonry and lime wall paintings (Holmes *et al.*, 2000; de la Torre *et al.*, 2003; Schabereiter-Gurtner *et al.*, 2001). Desiccation is a theme common to both the last three environments and Antarctic Dry Valley soils. Similarly, the group 2 Antarctic clone was most closely related to a

sequence retrieved from a semi-arid soil. Thus, the distribution of these organisms in natural environments may in part, be linked to their ability to withstand desiccation.

Rubrobacter radiotolerans is an extremely radiation-resistant bacterium, exhibiting higher resistance than *Deinococcus radiodurans* (Asgarani *et al.*, 2000). The capacity of *D. radiodurans* to withstand the lethal effects of ionizing radiation is a function of its highly effective DNA repair system, which is proposed to have evolved in response to the DNA damage caused by desiccation (Mattimore & Battista, 1996). High levels of ionizing radiation do not occur in natural environments, thus, no selective pressure exists for such a phenotype. However, single strand DNA breaks and cross-linking of DNA to protein that are induced by desiccation are a general phenomenon, and present a principle constraint to the long term survival of microorganisms (Dose *et al.*, 1995). Thus, the DNA repair mechanism of *D. radiodurans* is thought to be the product of evolution of the ancestral *Deinococcus* species in a very arid environment.

Consequently, similar mechanisms of DNA repair likely exist in *R. radiotolerans* that protect against desiccation and also confer the radiation-resistant phenotype. Consistent with this hypothesis, Asgarani *et al.*, (2000) have identified a DNA repair enzyme of *R. radiotolerans* that corrects DNA lesions induced by ionizing radiation. It is therefore conceivable that *Rubrobacter*-like organisms possess DNA repair mechanisms that enable them to survive in arid environments. Furthermore, as with *D. radiodurans*, both *R. radiotolerans* and *R. xylanophilus* are pigmented organisms due to the presence of carotenoids typical of halophilic bacteria (Saito *et al.*, 1994). Carotenoids serve as antioxidants and have been demonstrated to have a protective function against oxidative DNA damaging reagents, ionizing radiation and UV light (Shahmohammadi *et al.*, 1998). Interestingly, the sequences identified from masonry and lime wall paintings that were among the closest relatives of Antarctic *Rubrobacter* sequences, were specifically associated with the rosy discolouration of these surfaces. In this study conducted by Schabereiter-Gurtner *et al.*, (2001), the bacterial community of three red-pigmented biofilms was investigated by 16S rRNA gene-dependent phylogeny using clone libraries and DGGE analysis. Of the 16S rRNA sequences analysed, 70% were most closely

related to the genus *Rubrobacter*, and specifically in one sample, only *Rubrobacter*-like sequences were detected. No *Rubrobacter*-like organisms were detected in non-rosy biofilms. This would imply the presence of carotenoids in *Rubrobacter* inhabiting the masonry biofilm and also, potentially within Antarctic *Rubrobacter* given their relationship to the masonry clones and the two cultivated species of *Rubrobacter*.

These properties of resistance to aridity and UV would certainly be advantageous for organisms inhabiting Dry Valley environments. Indeed, it would be interesting to determine whether the distribution of this subclass is correlated with soil aridity, by correlating the presence of the subclass along a soil transect of increasing water content within the Dry Valleys. This could be achieved by Real Time PCR analysis of *Rubrobacteria* spp. sequences using DNA extracted directly from transect samples and amplification with group-specific PCR primers. Additionally, this work could be extended to the analysis of 16S rRNA templates, to quantify the metabolically active members within this subclass and in turn correlate to water content.

A number of the clone sequences described here possessed relatives of Antarctic origin, incorporating eight of the bacterial phyla represented. Among those related to constituents of cryptoendolithic communities present in the McMurdo Dry Valleys and investigated by de la Torre *et al.*, (2003), were the *Rubrobacteria*-like sequences described above, clone UVW4-34 also of the *Actinobacteria*, and clone UWV4-20, an uncultured deinococcus. *Deinococcus* species have additionally been identified in Dry Valley soil and South Pole snow and as discussed earlier, this is thought in part to reflect their ability to withstand desiccation (Shravage *et al.*, Unpublished; Carpenter *et al.*, 2000). Interestingly, *Deinococcus*-like sequences were found to be abundant within the cyanobacterium-dominated cryptoendolithic community and the authors postulated these organisms may have a role in primary productivity due to their high abundance (de la Torre *et al.*, 2003). Additionally, clone UWV4-1 of the *Acidobacteria* lineage was most similar to a member of a lichen-dominated cryptoendolithic community in the McMurdo Dry Valleys. This phylotype accounted for 11 of the 24 clones analysed from the Upper Wright library, suggesting that this organism may be numerically dominant in the soil

studied. Clone UWV4-1 along with a second clone were affiliated to group 4 of the *Acidobacteria*, a group that contains currently no cultured representatives. Consequently, little is known regarding the physiology of these organisms. However, culture-independent studies have indicated that they are widespread in soils and are therefore likely to be of ecological significance (Kuske *et al.*, 1997; Nogales *et al.*, 1999; Buckley & Schmidt, 2003; Liles *et al.*, 2003).

The identification of bacteria inhabiting soils of the McMurdo Dry Valleys that are closely related to members of cryptoendolithic communities present in this region is a reflection of the similarity between these two environments, where rock serves as the substrate for microbial colonization (Nienow & Friedmann, 1993). Dry Valley soils are likely to represent the major reservoir of microorganisms that are able to colonize the cracks and fissures within rocks opportunistically (Hirsch *et al.*, 1985). It is therefore possible that the bacterial types common to both communities may represent primary colonizers of the endolithic zone.

Three sequences obtained in this study, affiliated to the *Proteobacteria*, were of particular note. The first sequence, MV11-6 shared 94% 16S rRNA sequence identity over 1253 bp to *Rickettsia bellii* within the *Alphaproteobacteria*. Interestingly, this species is tick associated, and members of this genus are obligate intracellular parasites that are typically maintained in arthropod vectors (Roux & Raoult, 1995; Walker & Gage, 1997). The level of sequence divergence indicates clone MV11-6 may represent a new genus within the family *Rickettsiaceae*, which also includes the well-known genera *Ehrlichia* and *Wolbachia*. Again, these organisms can only reproduce intracellularly and in particular, *Wolbachia* are extremely common in insects where they are responsible for a range of effects on the host organism (Charlat *et al.*, 2003). Thus, the identification of this sequence might suggest the presence of an arthropod, such as a mite, in the Miers Valley. Terrestrial arthropods do occur in Antarctic environments and have developed the necessary strategies to survive sub-zero temperatures and desiccating conditions. Based upon the information available, these include the free-living oribatid mite, *Alaskozetes antarcticus*, which is widely distributed throughout maritime Antarctic, and

two species of collembolan (Worland & Block, 2003). Collembola are primitive wingless insects that primarily inhabit soil and litter and are best known by the common name of springtails, due to their forked 'tail' that enables them to move by a springing action. Such species include *Cryptopygus antarcticus*, often the dominant micro-arthropod in moss communities in maritime and Antarctic habitats (Worland & Block, 2003), and *Gomphiocephalus hodgsoni*, which has been shown to be endemic to Southern Victoria Land (Stevens & Hogg, 2003). All species feed on fungi, algae and organic detritus, of which a variety of fungal species have been cultivated from Ross Desert Soils and are also present in cryptoendolithic communities (Vishniac, 1993; de la Torre *et al.*, 2003). Furthermore, macroscopic growth of moss and lichens were observed in the Miers Valley during the 2002 field expedition, on rocks located within a glacial melt-stream directly adjacent to the glacier. The collembola *Gomphiocephalus hodgsoni* has been identified in samples taken from Taylor Valley (Stevens & Hogg, 2003), a site neighbouring the Miers Valley, and as such may be considered the strongest candidate host species of the rickettsial species identified here. The amplification of this sequence from DNA extracted from sample MV11.1 is significant given the considerable back-drop of soil microbial diversity and further suggests the presence of an arthropod in the soil sampled.

Clone MV11-1, affiliated to the *Betaproteobacteria*, was most similar (97% identity) to a sequence identified in penguin dropping sediments from Ardley Island, Antarctica (Zhang *et al.*, Unpublished). Furthermore, MV11-1 shared 96% 16S rRNA gene identity to *Lautropia mirabilis*, an opportunistic pathogen of humans first isolated from the oral cavity (Gerner-Smidt *et al.*, 1994). *L. mirabilis* is a heterotrophic, facultative anaerobe capable of nitrate and nitrite reduction and was similarly identified as the closest relative of an uncultured bacterium associated within gas hydrates in the Gulf of Mexico (Lanoil *et al.*, 2001). The material with which this uncultured bacterium was associated – gas hydrate – is an ice like crystalline solid composed of molecules of natural gas (typically methane) surrounded by a cage of water molecules. This material is formed under conditions of high pressure and low temperature, and is found in both oceanic sediments along with permanently frozen soils and glacial ices of high latitudes (Kvenvolden,

1999). As it is a primary site where the anaerobic oxidation of methane (AOM) occurs, gas hydrates have been the subject of a number of microbial diversity studies (Bidle *et al.*, 1999; Lanoil *et al.*, 2001; Fry *et al.*, 2001). Two phylogenetically distinct groups of archaea (ANME-1 and ANME-2 belonging to the *Euryarchaeota*) perform AOM and are frequently observed with sulfate-reducing bacteria within gas hydrates, forming an association based upon transfer of reduced compounds (Valentine, 2002). The result is an intricate microbial consortia where primary productivity is based upon consumption of methane and hydrogen sulfide (the waste product of AOM) (Boetius *et al.*, 2000; Lanoil *et al.*, 2001; Nauhaus *et al.*, 2002).

However, the presence of gas hydrates within Miers Valley soils is unlikely given this region is not of high altitude. Furthermore, the investigation of archaeal diversity within Dry Valley soils did not identify any members of the *Euryarchaeota*. This suggests that there may be either few or no archaea present that are capable of AOM. However, clone MV11-1 may represent a bacterium capable of anaerobic respiration using nitrate as an alternative electron acceptor to oxygen. This notion is based upon the similarity of MV11-1 to *L. mirabilis* (96% 16S rRNA identity) and the environmental context of its closest relatives; gas hydrates represent an anaerobic environment and penguin dropping sediments likely support anaerobic growth of microorganisms. Significantly, nitrate is reported to be a major anion in Ross Desert soils as the marine influence wanes and is largely of atmospheric origin (Vishniac, 1993). Thus, the possibility exists that nitrate may represent a suitable alternative electron acceptor for anaerobic and facultative anaerobic organisms in Dry Valley soils.

Nitrate reduction is carried out by phylogenetically diverse bacteria and includes members of the *alpha-*, *beta-*, and *gammaproteobacteria*, Gram-positive bacteria and the *Archaea* (Philippot, 2002). Recently, diversity among nitrate-reducing communities in soils was examined by PCR amplification of the *narG* gene (Philippot *et al.*, 2002). The *narG* gene encodes the α subunit of the membrane bound nitrate reductase that incorporates the catalytic site of nitrate reduction; as such, this gene can be used as a functional gene marker due to its universal presence in members of this community. In

the study conducted by Philippot *et al.* (2002), clone libraries were constructed from PCR-amplified *narG* sequences obtained from planted and unplanted soils; RFLP analysis of 472 clones randomly selected from 6 libraries identified 118 unique restriction patterns. Sequencing of clones revealed the predicted amino acid identity to range from 62 to 99.5%, indicating substantial functional gene diversity among nitrate reducers. Similarly, Gregory *et al.*, (2003) examined diversity of *narG* gene sequences among nitrate-reducing isolates cultivated from freshwater sediment and amplified from total DNA extractions. This investigation yielded considerable species diversity among the culturable fraction of nitrate reducers, and demonstrated a degree of spatial organization in terms of the species detected and *narG* sequence diversity relative to sediment depth. Therefore, *narG* could be used as a functional gene marker to explore diversity among nitrate reducers in Antarctic Dry Valley soils, targeting both the culturable population and *narG* gene sequence diversity within total DNA extractions. This would provide considerable insight into the role and significance of nitrate reduction in Antarctic Dry Valley soils and the results could be contrasted with environments previously investigated.

Interestingly, three sequences affiliated to the *Actinobacteria* were most closely related to clone sequences obtained from aquatic environments. Warnecke and coworkers (2004) recently analysed the phylogenetic relationship of 63 actinobacterial sequences obtained from freshwater environments with over 300 sequences obtained from a variety of soil habitats. Four distinct clusters were evident, designated acI to acIV, from which lineage acIV was found to contain the Lake Schohsee clone S4 (AJ575506) identified as a database relative of clone UWV4-3. From this it can be inferred that the Antarctic sequences are affiliated to group acIV. Members of this cluster appear to occupy a diverse array of habitats that include soils, marine water and sediments, and activated sludge (Warnecke *et al.*, 2004). Furthermore, sequences have been identified in lake sediment and microbial mat samples from Antarctica (Bowman *et al.*, 2000; Brambilla *et al.*, Unpublished). However, given that there are no cultivated isolates among this group or data regarding their abundance and activity in natural environments, nothing is known about the physiology of these organisms.

Similarly, little information is known about the physiologies corresponding to the closest database relatives of the remaining clones that were identified in this study; namely clones affiliated to the lineages TM7, *Veruccomicrobia*, *Planctomycetales* and *Chloroflexi*. A total of four clones were affiliated to division TM7, which at present contains no cultivated representatives, and all exhibited ≤95% 16S rRNA identity to their closest relatives. Sequence divergence was also comparatively high (<94% identity to database relatives) among representatives of the other three phyla. This emphasizes the importance of a polyphasic approach to research, investigating both functional and phylogenetic diversity among communities and changes in specific environmental parameters, in order to determine the role these organisms play in Antarctic and other environments.

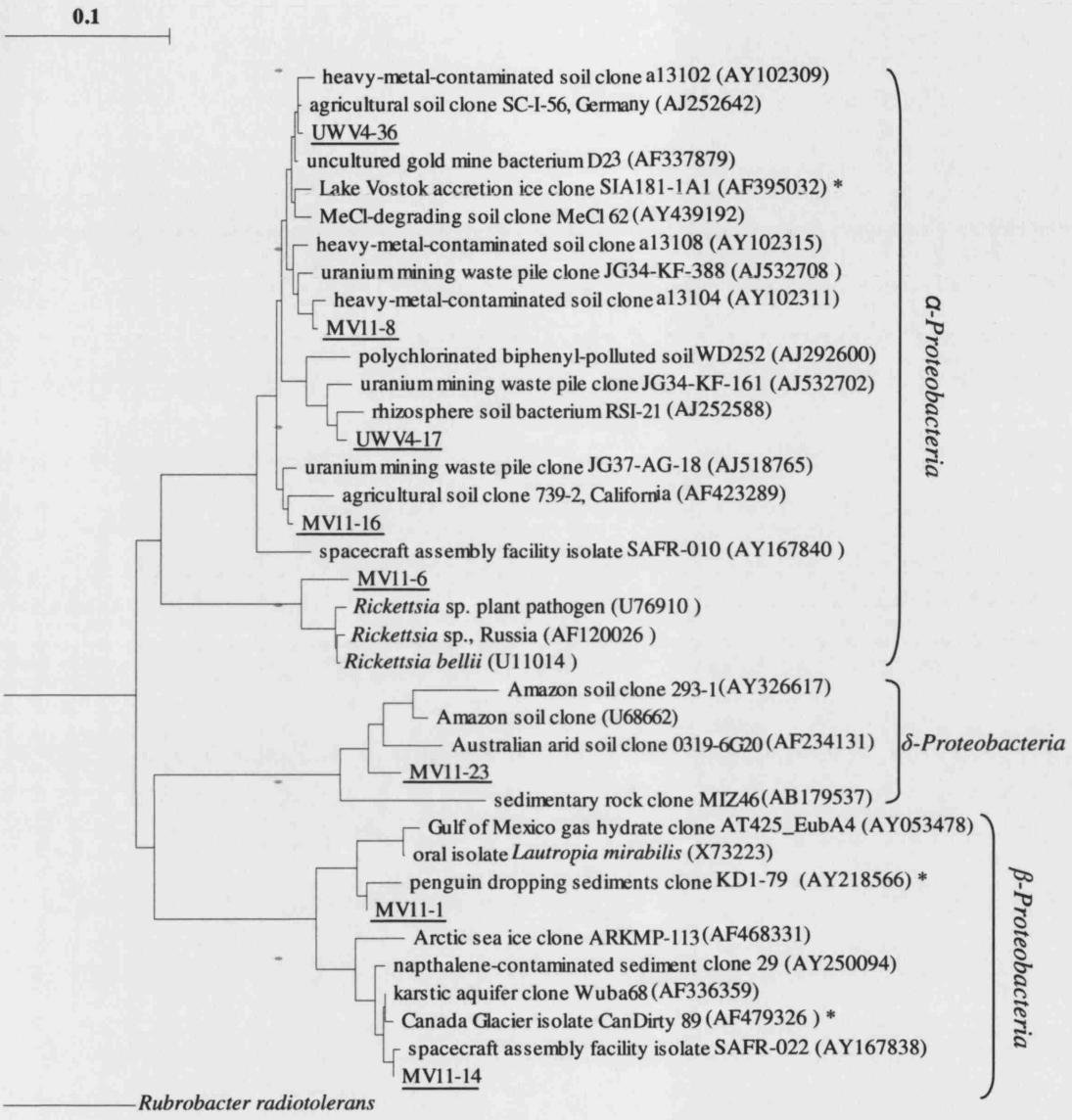


Figure 6.2: Phylogenetic tree illustrating the closest database relatives identified for Antarctic clone sequences (shown underlined) affiliated to the *Proteobacteria*. Database relatives of Antarctic origin are marked with an asterisk. Scale bar: 0.1, estimated number of substitutions per nucleotide position.

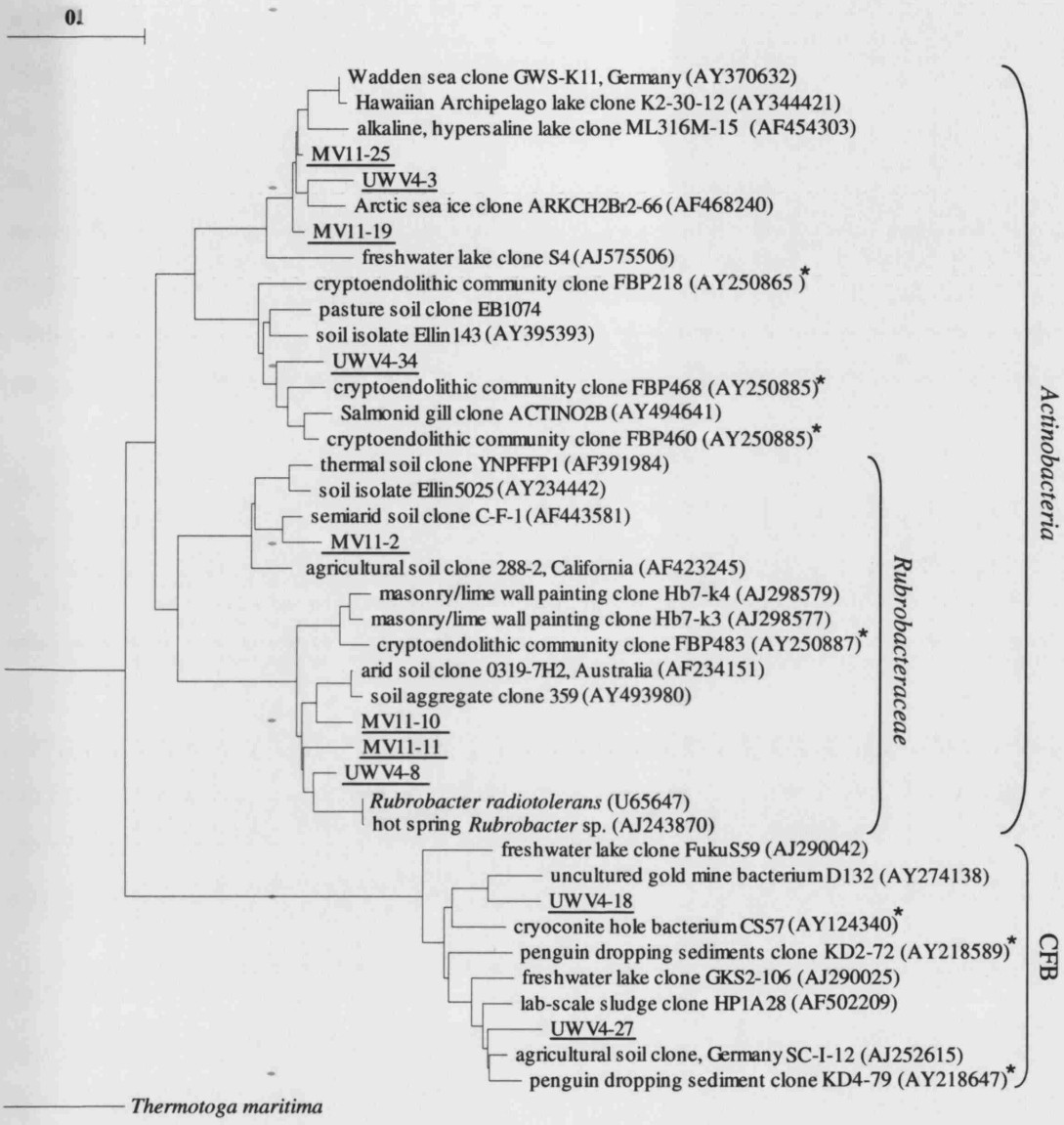


Figure 6.3: Phylogenetic tree illustrating the closest database relatives identified for Antarctic clone sequences (shown underlined) affiliated to the lineages *Actinobacteria* and CFB. Database relatives of Antarctic origin are marked with an asterisk. Scale bar: 0.1, estimated number of substitutions per nucleotide position.

6.4: Are Bacteria Endemic to Antarctica?

The term endemism refers to species unique to a certain region, and follows biogeographical constraints that impact on the evolution of organisms. With respect to bacteria, the geographical isolation of species and their evolution over time can result in genotypic differences; endemism occurs when a particular genotype is identified in a single region only. This concept requires that the mixing of genotypes only occurs within a region and not between regions, a factor that is dictated by the available mechanisms of dispersal.

When considering the case for endemism, a distinction must be made between a species that is transient and one that is resident. For example, a transient organism may be identified in the region of interest but will be not be truly endemic. In contrast, an endemic organism should theoretically be sampled at any time, that is, taking into account the species population size at the time of sampling. This distinction is important, and when investigating endemism there are clear advantages to testing for species presence on a temporal scale. However, this was not possible in this present study. Thus, the data presented here have been interpreted as evidence to support or reject the potential for species endemism in the Antarctic environment.

This study focused on the Antarctic Dry Valleys of South Victoria Land, a region of 4,800 sq km. Only 2% of Antarctica is ice-free and the Dry Valleys account for a mere 1.7% of the ice-free region. This biotope is extreme even by Antarctic standards, characterized by aridity, oligotrophy, extreme fluctuations in temperature and the strong katabatic winds that sweep through the valleys. For this reason, it can be argued that the combination of these conditions might make it more likely that species detected in Dry Valley soils are endemic.

Over half of the bacterial isolates cultivated in this study from Antarctic Dry Valley soil possessed a bacterium of Antarctic origin (Dry Valley or otherwise) as one of their four closest relatives. Moreover, a number of genera identified here were represented by bacterial species previously cultivated from different Antarctic biotopes. However, given

that only partial 16S rRNA gene sequences were analyzed, and all Antarctic isolates shared the same level of sequence identity to cultivated bacteria of non-Antarctic origin, this provides little evidence for endemism. Instead, it would imply a range of bacterial genera are suitably adapted to life in Antarctic environments.

A number of parallels could be drawn with the Antarctic Dry Valleys, from environments that shared similarities in terms of the culturable fraction of bacterial species. From this it could be inferred that a broad scale selection of certain functional attributes operated as a product of these common environmental parameters. Thus, the bacterial species inhabiting for example, ice or a spacecraft assembly facility and the Antarctic Dry Valleys, may represent organisms best adapted to growth under such conditions. This could be investigated by testing for the presence of these 'indicator' species using genus- or species-specific primers in similar environments of geographically distinct locations.

In the molecular assessment of bacterial diversity, six of the Dry Valley clones included in the phylogenetic analysis also possessed a clone sequence or cultivated bacterium of Antarctic origin as their closest relative. However, none of these were most closely related exclusively to Antarctic sequences or isolates. Furthermore, for three of the Dry Valley soil clones, the level of 16S rRNA identity was <98% strongly suggesting these uncultured bacteria represent new species. In particular, for clone MV11-32 that was most closely related to a sequence identified in penguin dropping sediments from Ardley Island, Antarctica, the sequence divergence was 6%. This clone was affiliated to the *Verrucomicrobia*, a phylum that was only recently recognized (Ward-Rainey *et al.*, 1995). Therefore, in this instance it would be premature to infer endemism given that our knowledge of the distribution of this phylum is still limited to an extent. The same can be considered of sequences exhibiting comparatively high divergence from their database relatives, as it is highly probable that future molecular phylogenetic studies investigating microbial diversity both in natural and other environments will identify further related sequence types.

Endemism may exist in the Antarctic environment but presently, there is insufficient sequence data available regarding the diversity of Antarctic microbial communities to infer endemism using the 16S rRNA gene. A cluster of 16S rRNA gene sequences that are exclusively Antarctic would imply endemism. Furthermore, this phylogenetic marker may not be considered a good indicator of endemism due to its lack of resolution, since the rRNA operon is highly conserved. Indeed, Hahn *et al.*, (2003) reported the isolation of two strains affiliated to the *Actinobacteria* from different freshwater ecosystems of separate continents, which possessed identical 16S rRNA gene sequences. However, both strains were found to exhibit different ecophysiological and phenotypic traits.

Clone UWV4-20 of the *Deinococcus* lineage may perhaps represent the strongest candidate in terms of evidence of endemism in the Antarctic environment. This clone shared 100% 16S rDNA identity to an uncultured *Deinococcus* sp. identified in a cryptoendolithic community from the McMurdo Dry Valleys. Furthermore, the second database relative sharing 98.5% identity was an uncultured bacterium present in hydrocarbon-contaminated soil from Ross Island, Antarctica. Therefore, it would be interesting to examine the distribution of members of this genus in more detail and using a phylogenetic marker of increased resolution. The study of genotypic diversity within a smaller phylogenetic range may be considered a more directed and easier method of detecting endemism.

For example, Cho and Tiedje (2000) examined the extent of endemicity among 248 fluorescent *Pseudomonas* strains isolated from soil collected from five different continents. In thAT study, three methods of increasing resolution were used to identify genotypes: ARDRA of 16S rRNA gene amplicons, 16-23S rRNA ITS-RFLP, and BOX-PCR. No endemicity was detected using ARDRA, whilst a weak level of endemicity was observed from ITS-RFLP analysis (the ITS serves as a more variable marker as it less subject to selection pressure than the 16S rRNA gene). In contrast, the BOX-PCR revealed strong endemicity among strains, and individual genotypes were found to be specific to a certain site and not distributed between sites or regions. Consequently, the 16S rRNA sequence data analyzed here for cultivated isolates and uncultured bacteria

identified in Dry Valleys soils, is perhaps unlikely to detect evidence of endemism. Indeed, sequences identified from the molecular survey of diversity appear to represent bacterial types that are widespread in their geographical distribution.

In a study conducted by Taton *et al.*, (2003), evidence of endemism in the Antarctic environment was observed using 16S rRNA sequence data; however, the group of organisms subject to investigation was comparatively broader than the *Pseudomonas* strains described above. Taton and coworkers investigated the diversity of cyanobacteria in microbial mats of Lake Fryxell in the McMurdo Dry Valleys, Antarctica. The phylogenetic analysis of 49 partial 16S rRNA sequences obtained in this study with 125 previously published sequences of cyanobacteria of Antarctic and polar origin, revealed seven clusters that were exclusively Antarctic. In this instance, the identification of possible endemism was undoubtedly aided by the significant amount of 16S rRNA sequence data available for cyanobacteria. However, both cyanobacterial groups that were cosmopolitan in their distribution and those that appeared endemic to Antarctic were identified in this study.

The results of the cultivation and cultivation-independent study of bacterial diversity presented here, indicate Antarctic Dry Valley soils to support a complex community of bacterial species. Given that a relatively small number of 16S rRNA gene sequences were analysed in this study, together with the fact that many of these sequences were encountered only once, means that it is difficult to draw any firm conclusions about the significance of these sequences in the Dry Valley environment. This study finds little evidence of endemism based upon the analysis of partial 16S rRNA sequences. The high degree of similarity among Dry Valley soil isolates to previously cultivated Antarctic bacteria can be considered a reflection of the limited number of species that can be cultivated *in vitro*. These organisms appear to represent bacterial genera suitably adapted to growth in the Antarctic and similar environments. Overall, these data in combination with the 16S rRNA gene clone sequences analysed here and the results of other diversity studies, strongly support the idea that these bacterial groups have a widespread global distribution. This is interesting given the extreme conditions within the Dry Valleys.

Endemism may indeed occur within certain taxa inhabiting Antarctic environments, for which *Deinococcus* is a candidate, and this would be interesting to investigate further using more sensitive methods. In conclusion, Dry Valley soils harbour considerable novelty in terms of bacterial species and thus, warrant further investigation.

6.5: Integron Diversity

6.5.1: Integrons in the Antarctic Environment

Chapter five describes the investigation of integron diversity in bacteria within Antarctic soils, with respect to the three distinct regions of the integron platform: the 5'-CS that includes the integrase gene, the variable region containing integrated gene cassettes, and the 3'-CS typical of class 1 integrons. It was hypothesized both that integrons in the Antarctic environment would possess the same basic structure as those recovered from bacteria within clinical and environmental samples of other geographical regions, and that the Antarctic cassette pool would include cassettes similar to those previously characterized along with those encoding novel orfs and/or gene products.

Partial integrase sequences were recovered from the pooled DNA of Antarctic samples UWV1-4, UWV5-8 and MV8-12, along with DNA extracted from bed dust and pasture soil. All sequences exhibited 99-100% identity to *intI1* of class 1 integrons. Additionally, three Antarctic samples - UWV4, MV17 and Brat2, were positive for the integron screening PCR, in which a 117 bp region of the 5'-CS was amplified. Once again, these sequences and the sequence recovered from bed dust DNA exhibited 100% identity to the corresponding region of class 1 integrons. The 5'-CS of class 1 integrons includes the integrase gene (*intI1*), a promoter for expression of integrated gene cassettes (P_c), and the cassette integration site (*attI*). This region is highly conserved with >99% nucleotide sequence identity shared among a wide range of different integrons that vary both in their genomic context and host species, and that are of diverse geographical origins (Liebert *et al.*, 1999; Partridge *et al.*, 2001). Indeed, to illustrate this point, integrons bearing integrase genes of ≥99% nucleotide identity include those present on the resistant plasmids R100 and R46, and plasmids isolated from a wastewater treatment plant in Germany and a fish pathogen of Norwegian salmon (Sorum *et al.*, 2003), along

with integrons of clinical resistant strains isolated in Europe (Lauretti *et al.*, 1999; Poirel *et al.*, 2000; Sabate *et al.*, 2002), South Africa (Elisha & Steyn, 1994), Japan (Kazama et al.,1995) and China (Siu et al., 2000). Sequence differences that do occur, appear most frequently within the -35 and -10 regions of the P_c promoter located within the integrase coding region, and account for variation observed in promoter strength (Levesque *et al.*, 1995; Martinez-Freijo *et al.* 1999). Specifically, the integrase sequences amplified from Antarctic soils all included the strong promoter - $TTGACAN_{17}TAAACT$. An insertion of three guanosine molecules approximately 100 bp downstream of P_c can create a second weak promoter, providing a further initiation point for transcription. This insertion is observed among In2 of Tn*21* and close relatives (Levesque *et al.*, 1995). Class 1 integrons were identified in a number of the first multidrug-resistant plasmids isolated, including plasmid NR1 (R100) isolated in Japan in the late 1950s, and plasmids R46, R751 and R388 isolated from resistant strains in Europe during the 1960s and 1970s (Hall & Vockler, 1987; Radstrom *et al.*, 1994). Thus, given their diverse genetic backdrop, it is likely that integrons were present in bacteria in a range of environmental and geographical settings prior to the introduction of antibiotics for the treatment of disease.

Two different gene cassettes were identified in Antarctic soils, both of which encode aminoglycoside-modifying enzymes. An *aadA* gene cassette was amplified from Antarctic DNA samples MVT7, MVT11, MV17 and Brat2 that was most closely related to the *aadA1* gene cassette of In2, a class 1 integron located on Tn*21* of plasmid R100 (Stokes & Hall, 1989). Plasmid R100 was first isolated from *Shigella flexneri* in Japan during the late 1950s (Nakaya *et al.*, 1960) and carries transposon Tn*21*. Transposon Tn*21* contains the mercury resistance (*mer*) operon and is widely distributed in nature, both in mercury-polluted and pristine environments (Pearson *et al.*, 1996; Osborn *et al.* 1997). The same gene cassette was also recovered on two separate occasions (using different primer pairs) from both DNA extracted from bed dust and pasture soil. The second gene cassette, amplified from Antarctic sample UWV4, comprised an *aadA6* gene most similar to the gene cassette located within a class 1 integron, In51, isolated from *Pseudomonas aeruginosa* (Naas *et al.*, 1999). In this case, *P. aeruginosa* was the

cause of a urinary tract infection in a patient hospitalized in France, although the patient was suspected to have imported this strain from Thailand following hospitalization there. A further third cassette containing an *aadA2* gene was recovered from bed dust DNA that was most similar to the *aadA2* gene present on a class 1 integron of *Vibrio cholerae*. Once again, this particular isolate of *V. cholerae* originated from Thailand (Dalsgaard *et al.*, 2000).

Attempts to recover an increased diversity of gene cassettes from both bed dust and Antarctic DNA samples using different primer combinations were unsuccessful. Typically, products amplified with primer pairs used in this study (Brint/3CS, Brint/HS286, Int5CS/Int3CS and Int5CS/HS286) were all ~1,000 bp in size. Similar results have been observed in studies examining the cassette content of integrons among the *Enterobaceriaceae*. In a study conducted by Martinez-Freijo and colleagues (1999), in which the variable region of class 1 integrons of clinical isolates obtained from different European hospitals were characterized, they observed the most commonly inserted DNA regions to comprise of 800, 1,000 and 1,500 bp. Of the nine isolates for which a 1,000 bp insert was obtained, the *aadA* gene cassette was identified in each case. Furthermore, White *et al.* (2002) examined antimicrobial resistance among 50 *E. coli* O111 isolates of human and animal origin collected between 1976 and 1999. Twenty of the 23 integron-containing isolates gave rise to a 1 kb amplicon following amplification of the variable region, all of which comprised an *aadA* gene. Similarly, 8 out of 9 Shiga toxin-producing *E. coli* isolates isolated from cattle resulted in a 1 kb amplicon that was found to contain the *aadA* gene (Zhao *et al.*, 2001).

Thus, these results would indicate the presence of class 1 integrons and their associated gene cassettes in Antarctic soils. The sequence conservation identified for 5'-CS sequences of Antarctic origin is consistent with that observed for class 1 integrons of different geographical origins. This supports the notion that class 1 integrons function as a ubiquitous gene-transfer system for Gram-negative bacteria in both clinical and natural environments. Additionally, the recovery of identical gene cassettes from Antarctic soils, bed dust and pasture soil provides evidence that endemism does not occur in the

Antarctic environment with respect to the integron platform or the associated gene cassettes. Given that only products of 1,000 bp were obtained from Antarctic DNA samples following amplification of the variable region of integrons, this would infer the sole presence of the *aadA* gene cassette in these samples (as detected by this method). The absence of products following amplification with primers HS286/ HS287 that target the 59-be sites, would also indicate that arrays of cassettes are not present in these samples. This PCR strategy is only successful if one cassette is preceded by another, as amplification depends on a cassette being bordered by two 59-be sites. Thus, where multiple cassettes are located within an integron, all can potentially be amplified with the exception of the first cassette in the array. The absence of HS286/ HS287 products in this study contrasts with previous investigations of cassette diversity in soil and natural environments, in which a variety of highly novel gene cassettes have been identified using this strategy (Stokes *et al.*, 2001; Holmes *et al.*, 2003a; Nemergut *et al.*, 2004).

6.5.2: The *aadA* family of gene cassettes

The gene cassettes identified in this study encode aminoglycoside-3"-adenyltransferases (AAD), enzymes which confer resistance to the antibiotics streptomycin and spectinomycin. They form part of large and diverse group of aminoglycoside-modifying enzymes that catalyze the covalent modification of aminoglycoside antibiotics (White & Rawlinson, 2001). These antibiotics represent a large group of secondary metabolites whose primary target is the ribosome, to which they bind to and inhibit the biochemical steps involved in translation of mRNA, thus affecting protein synthesis. Over 50 different aminoglycoside-modifying enzymes have been identified and these can be classified into three groups: N-acetyltransferases (AAC) use acetyl-coenzyme A as a donor molecule and affect amino functions, whilst O-nucleotidyl transferases (ANT or AAD) and O-phosphotransferases (APH) both use ATP as a donor molecule and affect hydroxyl functions (Davies & Wright, 1997; Wright, 1999). Specifically, AAD enzymes confer resistant to streptomycin and spectinomycin by adenylation of the 3" -hydroxyl position and the 9-hydroxyl position of the antibiotic respectively.

Presently, the *aadA* family of genes comprises eleven different types based upon sequence differences in the amino acid sequence of the proteins encoded. Table 6.1 illustrates the number of nucleotide entries identified within GenBank for each member of the of *aadA* family, along with information regarding the range of host species from which each cassette has been isolated and the relevant genomic context (where applicable). Based upon the number of nucleotide entries in GenBank, the *aadA1* gene cassette would appear to be detected more frequently followed by the *aadA2* gene cassette, together accounting for 74% of the *aadA* entries. Relatives of both of these cassettes were identified in this study. Six of the different cassette types are represented by less than 5 nucleotide entries in GenBank.

An alignment of the AadA protein encoded by each type of *aadA* gene cassette is shown in Figure 6.2. Included is the orf identified from the MVT11-1 cassette that is representative of the *aadA1*-type cassettes recovered in this study, and hereby designated *aadAX*. The *aadA10* cassette has been excluded from this alignment as it may not represent a complete cassette (Partridge *et al.*, 2002).

A total of 104 residues are conserved in all sequences and are represented as a consensus sequence in Figure 6.4. Pairwise identities of the AadA proteins range from 53% to 97%. The AadAX protein exhibits 85% amino acid identity to AadA1, followed by 83% identity to both AadA3 and AadA8, and 82% identity to AadA2. The AadA3 and AadA8 proteins are considered hybrids of AadA1 and AadA2, sharing 87% and 90% identity respectively to AadA1, and 97% and 93% respectively to AadA2 (Partridge *et al.*, 2002). Whereas these proteins are most closely related to AadA2, the AadAX protein shares greater similarity to AadA1. Due to the extent of sequence divergence between AadAX and AadA1, this would indicate the AadAX protein represents a new and additional member of the AadA family.

To date, surveys of integron diversity have consistently identified the *aadA* gene cassette to be the most commonly found resistance gene inserted into the variable region of integrons. This has been widely demonstrated for clinical isolates of *Enterobacteriaceae*

(Levesque *et al.*, 1995; Martinez-Freijo *et al.*, 1999). Additionally, the *aadA* gene cassette has been identified in the following integron-containing bacteria: 28 out of 109 Gram-negative isolates cultivated from an estuarine environment (Rosser & Young, 1999); 23 of 24 bacterial isolates cultivated from cattle faecal samples (Barlow *et al.*, 2004), and 22 of 23 *E. coli* O111 isolates of human and animal origin (White *et al.*, 2002). Furthermore, of 21 gene cassette arrays identified on resistance plasmids isolated from activated sludge and amplified from total plasmid DNA preparations, 13 contained an *aadA* gene (Tennstedt *et al.*, 2003).

Significantly, in multiple cassette-containing integrons, the *aadA* gene is virtually always observed at the 3' end of the variable region. Given that cassettes preferentially recombine into the *attI* site of an integron (Collis & Hall, 1995), this phenomenon would indicate that the *aadA* gene was the first cassette to be integrated. Contrary to the belief that integrons are highly mosaic structures permitting frequent genetic exchange, these observations infer they are relatively stable structures. Indeed, it has been proposed that transfer of the entire integron may occur more frequently than cassette integration/ excision events (Martinez-Freijo *et al.*, 1999), and the predominance of the *aadA* gene cassette may reflect a common evolutionary origin.

Considering the majority of integron studies to date have focused on the cassette content of class 1 integrons and have shown the *aadA* gene cassette to be the most frequent and stably integrated resistance cassette, it is perhaps not surprising that an *aadA* cassette was identified in Antarctic soils. All the primer combinations used in this work to amplify the variable region of integrons were designed based upon class 1 integron sequences. Given that Antarctic Dry Valley soils represent a 'pristine' environment, the absence of both a variety and multiple arrays of resistance gene cassettes so frequently observed in integrons of clinical origin, might have been expected. Whilst extensive cassette diversity has been identified in natural environments, these studies have employed a different strategy. PCR primers used in these instances have targeted the 59-be sites of cassettes, thus operating independently of integron class (Stokes *et al.*, 2001; Holmes *et al.*, 2003a). When this strategy was applied to Antarctic soils, no PCR

Table 6.1: The *aadA* family of gene cassettes. The number of nucleotide entries in GenBank for each type of *aadA* gene cassette is shown. (Values were obtained by entering the gene name in Entrez.) Details of the range of host species from which class 1 integrons containing the specified *aadA* gene cassette are shown, along with information regarding the genomic context of the integron where available.

aadA gene	No. of entries in GenBank	Host species	Genomic context
aadA1	71	*Proteus mirabilis*	
		Escherichia coli	p541, pIP1100, pAK33; R751
		Bordetella parapertussis	genome sequence
		Vibrio sp.	
		In86, *Acinetobacter baumannii*	
		Pseudomonas aeruginosa	Tn1403
		Kluyvera ascorbata	
		Shigella flexneri	
		Pasteurella multocida	pJR2
		Klebsiella pneumoniae	pLQ1000
		Uncultured bacterium	plasmid-encoded class I integron
			Tn21 of plasmid R100
			R46
			Tn27
aadA2	55	In4, *Pseudomonas aeruginosa*	Tn1696 R1033
		& *Pseudomonas* sp.	Tn1404
		Salmonella sp.	
		In36 *Escherichia coli*	pHSH1
		Campylobacter jejuni	
		Aeromonas sp.	
		Citrobacter sp.	
		Corynebacterium glutamicum	pCG4
		Vibrio sp.	
		Klebsiella pneumoniae	
		Serratia marcescens	

aadA gene	No. of entries in GenBank	Host species	Genomic context
aadA3	1	In21, *Escherichia coli*	Tn2424 of plasmid NR79
aadA4	10	*Escherichia coli*	Tn21 relative of plasmid pTJ100
		Acinetobacter baumannii	
		Klebsiella pneumoniae	
		Bordetella bronchiseptica	genome sequence
		Bordetella parapertussis	genome sequence
aadA5	15	*Escherichia coli*	
		Salmonella sp.	
		Acinetobacter baumannii	
		Klebsiella pneumoniae	
		Uncultured bacterium	plasmid-encoded class I integron
aadA6	2	In51, *Pseudomonas aeruginosa*	
aadA7	8	In163, *Pseudomonas aeruginosa*	
		Salmonella sp.	
		Vibrio sp.	
		Escherichia coli	
aadA8	2	*Klebsiella pneumoniae*	
		Uncultured bacterium	plasmid-encoded class I integron
aadA9	4	*Corynebacterium glutamicum*	pCG4
		Corynebacterium glutamicum	pTET3
aadA10	1	*Pseudomonas aeruginosa*	Tn1404 of hybrid plasmid R388-R151
aadA11	1	*Escherichia coli*	

Figure 6.4: Alignment of AadA proteins of different classes. AadAX represents the protein encoded by the *aadA* gene recovered in this study. The consensus sequence is shown in bold type.

```
AadA2       ----MTIEISNQLSEVLSVIERHLESTLLAVHLYGSAVDGGLKPYSDIDLLVTVAVKLDE
AadA3       MRVAVTIEISNQLSEVLSVIERHLESTLLAVHLYGSAVDGGLKPYSDIDLLVTVAVKLDE
AadA8       MRVAVTIEISNQLSEVLSVIERHLESTLLAVHLYGSAVDGGLKPYSDIDLLVTVAVKLDE
AadA11      MREAVTAEISTQLSEVLSVIERHLESTLLAVHLYGSAVDGGLKPYSDIDLLVTVTVRLNE
AadA1       MREVVIAEVSTQLSEVVGVIERHLEPTLLAVHLYGSAVDGGLKPHSDIDLLVTVTVRLDE
AadAX       MRDSVTAEISTQLSKVLSVIEHHLEPTLLAVHLYGSAVDGGLKPYSDIDLLVTVTARLDD
AadA6       MSNAVPAEISVQLSLALNAIERHLESTLLAVHLYGSALDGGLKPYSDIDLLVTVAARLDE
AadA7       MSEKVPAEISVQLSQALNVIGRHLESTLLAVHLYGSALDGGLKPYSDIDLLVAVAAPLND
AadA4       MGEFFPAQISEQLSHARGVIERHLAATLDTIHLFGSALDGGLKPDSNIDLLVTVSAAPND
AadA5       MGEFFPAQVFKQLSHARAVIERHLAATLDTIHLFGSAIDGGLKPDSDIDLFVTVSAAPND
AadA9       MSNSIHTGISRQLSQARDVIKRHLASTLKAIHLYGSAIDGGLKPYSDIDLLVTVDARLDE
                .   :  *** .  .* :** .** ::**:***:****** *:***:*:* .  ::

            ----------QLS-----I--HL--TL---HL-GSA-DGGLKP-S-IDL-V-V------

AadA2       TTRRALLNDLMEASAFPGESETLRAIEVTLVVHDDIIPWRYPAKRELQFGEWQRNDILAG
AadA3       TTRRALLNDLMEASAFPGESETLRAIEVTLVVHDDIIPWRYPAKRELQFGEWQRNDILAG
AadA8       TTRRALLNDLMEASAFPGESETLRAIEVTLVVHDDIIPWRYPAKRELQFGEWQRNDILAG
AadA11      ATRQALLNDLLEASAFPGESETLRAIEVTIVVHDDIIPWRYPAKRELQFGEWQRNDILAG
AadA1       TTRRALINDLLETSASPGESEILRAVEVTIVVHDDIIPWRYPAKRELQFGEWQRNDILAG
AadAX       TTRRALFNDLLEVSAFPGESEILRAIEVTIVVHEDIMPWRYPAKRELQFGEWQRNDILAG
AadA6       TVRQALVVDLLEISASPGQSEALRALEVTIVVHGDVVPWRYPARRELQFGEWQRKDILAG
AadA7       AVRQALLVDLLEVSASPGQNKALRALEVTIVVHSDIVPWRYPARRELQFGEWQRKDILAG
AadA4       SLRQALMLDLLKVSSPPGNGGPWRPLEVTVVARSEVVPWRYPARRGLQFGEWLRHDILSG
AadA5       SLRQALMLDLLKVSSPPGDGGTWRPLELTVVARSEVVPWRYPARRELQFGEWLRHDILSG
AadA9       ATRRSLMLDFLNISAPPCESSILRPLEVTVVACNEVVPWRYPARRELQFGEWLREDILEG
            : *::*. *::: *: * :.   *.:*:*:*.  :::******:* ****** *.*** *

            --R--L--D----S--P------R--E-T-V------PWRYPA-R-LQFGEW-R-DIL-G

AadA2       IFEPAMIDIDLAILLTKAREHSVALVGPAAEEFFDPVPEQDLFEALRETLKLWNSQPDWA
AadA3       IFEPAMIDIDLAILLTKAREHSVALVGPAAEEFFDPVPEQDLFEALRETLKLWNSQPDWA
AadA8       IFEPAMIDIDLAILLTKAREHSVALVGPAAEEFFDPVPEQDLFEALRETLKLWNSQPDWA
AadA11      ISEPAMIDVDLAILLTKAREHSVALVGPAAEELFDPVPEQDLFEALNETLKLWNSQPDWA
AadA1       IFEPATIDIDLAILLTKAREHSVALVGPAAEELFDPVPEQDLFEALNETLTLWNSPPDWA
AadAX       IFEPATIDIDLAILLTKAREHSVALVGPAAEELFDPVPEQDLIKALNETLKLWNSQPDWA
AadA6       IFEPATTDVDLAILLTKVRQHSLALAGSAAEDFFNPVPEGDLFKALSDTLKLWNSQPDWE
AadA7       IFEPATTDSDLAILLTKAKQHSIVLAGSAAKDLFSSVPESDLFKALADTLKLWNSPPDWA
AadA4       TFEPAVLDHDLAILLTKARQHSLALLGPSAVTFFEPVPNEHFSKALFDTIAQWNSESDWK
AadA5       TFEPAVLDHDLAILLTKARQHSLALLGPSAATFFEPVPKEHFSKALFDTIAQWNAESDWK
AadA9       VFEPAALDADLAILITKARQHSIALVGPVAQKVFMPVPEHDFLQVLSDTLKLWNTHEDWE
              ***  * *****:**.::**:.* *. *  .* .**: .: :.* :*:  **:  **

            --EPA--D-DLAIL-TK---HS--L-G--A---F--VP-------L--T---WN---DW-
```

```
AadA2       GDERNVVLTLSRIWYSAITGKIAPKDVAADWAIKRLPAQYQPVLLEAKQAYLGQKEDHLA
AadA3       GDERNVVLTLSRIWYSAITGKIAPKDVAADWAIKRLPAQYQPVLLEAKQAYLGQKEDHLA
AadA 8      GDERNVVLTLSRIWYSVVTGKIAPKDVAADWAMERLPAQYQPVILEARQAYLGQEEDRLA
AadA11      GDERNVVLTLSRIWYTEVTGKIVPKDVAADWAMERLPAQHQPVLLEARQAYLGQKEDRLA
AadA 1      GDERNVVLTLSRIWYSAVTGKIAPKDVAADWAMERLPAQYQPVILEARQAYLGQEEDRLA
AadAX       GDERNVVLTLSRIWYSAATGKIAPKDVAANWAMEHLPAQHQSVLLEARQAYLGQEEDRSV
AadA6       GDERNVVLTLSRIWYSAATGKIAPKDIVANWAMERLPDQHKPVLLEARQAYLGQGEDCLA
AadA7       GDERNVVLTLSRIWYTAATGKIAPKDVAATWAMARLPAQHQPILLNAKRAYLGQEEDYLP
AadA4       GDERNVVLALARIWYSASTGLIAPKDVAAAWVSERLPAEHRPIICKARAAYLGSEDDDLA
AadA 5      GDERNVVLALARIWYSASTGLIAPKDVAAAWVSERLPAEHRPLICKARAAYLGSEDDDLA
AadA9       NEERNIVLTLARIWYSTETGGIVPKDVAAEWVLERLPAEHKPILVEARQAYLGLCKDSLA
            .:***:**:*:****:  ** *.***:.* *.  :** :::.:: :*: ****  .*

            --ERN-VL-L-RIWY---TG-I-PKD--A-W----LP---------A--AYLG---D---

AadA2       SRADHLEEFIRFVKGEIIKSVGKK-------------------------
AadA3       SRADHLEEFIHYVKGEITKVVGK--------------------------
AadA8       SRADQLEEFVHYVKDEITKVVGKSEKVPAEISVQLSQALNVIGRHLEST
AadA11      SRADQLEEFVHFVKGEITKVVGKMSKTKLDALS----------------
AadA1       SRADQLEEFVHYVKGEITKVVGK--------------------------
AadAX       LHADKLEEFIHFMKSEITKVLGNDV------------------------
AadA6       SRADQLAAFVHFVKHEATKLLSAMPVMSNNSFKPTPLRGAA--------
AadA7       ARADQVAALIKFVKYEAVKLLGASQATMMSKTKDAST------------
AadA4        MRVEETAAFVRYAKATIERILR---------------------------
AadA5        MRVEETAAFVRYAKATIERILR---------------------------
AadA9        LRADETSAFIGYAKSAVADLLEKRKSQTSHICDGAKNV-----------
              :.:.   :: : *      :

             ------------K-----------------------------------
```

amplicons were obtained, thus inferring low cassette diversity in the samples investigated.

6.5.3: Functional significance of the *aadA* gene cassette

Taking into account the prevalence of the *aadA* family of gene cassettes, together with their stable and early recruitment within class 1 integrons, the recovery of *aadA* gene cassettes from Antarctic soils is consistent with the hypothesis that members of this family show a widespread distribution. Thus, it further suggests that the product of this gene should confer an adaptive function enhancing the fitness of the host cell.

The antibiotic streptomycin is produced by *Streptomyces griseus* of the order *Actinomycetales*. Members of this genus are Gram positive, filamentous bacteria that occur ubiquitously in soils and are responsible for the production of a diverse range of antibiotics. Streptomycin is a broad spectrum antibiotic active against Gram-negative bacteria. Therefore, acquisition of the *aadA* gene and the ability to express it in the

presence of antibiotic by Gram-negative bacteria would confer a significant advantage allowing resistant organisms to out-compete non-resistant phenotypes. The presence of *S. griseus* in Antarctic Dry Valley soils is conceivable given *Streptomycetes* are common soil borne organisms and two species were cultivated from Dry Valley soil in this study. Thus, the stable recruitment of the *aadA* gene cassette within integrons, which represent a primary mechanism of gene transfer among Gram-negative bacteria, can perhaps be considered the strongest evidence of the fitness benefits conferred by the gene.

6.5.4: Genomic context of the *aadA* cassette

The dissemination of antibiotic resistance genes among the *Enterobacteriaceae* and Pseudomonads can be largely attributed to the transfer of broad-host-range plasmids between species (Liebert *et al.*, 1999; Canton *et al.*, 2003). In combination with their association with integrons and transposons, which have facilitated the stockpiling of varying combinations of resistance genes, a mosaic of different resistance elements have been observed. Identification of the closest relative of the *aadAX* gene cassette recovered from Antarctic soils, as the *aadA1* cassette located within In2 of Tn*21* led to the hypothesis that this cassette may be part of Tn*21*. Transposon Tn*21* was first identified on plasmid R100, a self-transmissible, multiple antibiotic resistance plasmid isolated from *Shigella flexneri* (Liebert *et al.*, 1999). Tn*21* carries genes required for its own transposition, along with the mercury resistance operon and In2, a class 1 integron incorporating the *aadA1*, *qacEΔ1* and *sulI* antibiotic resistance determinants. Due to its mobility and the adaptive functions conferred on the host cell, Tn*21* occurs widely both in clinical and environmental isolates of gram-negative bacteria. However, PCR analysis of four Antarctic DNA samples from which the *aadAX* gene cassette was recovered could not demonstrate linkage of the gene cassette to Tn*21*. Furthermore, as the combined presence of the *qacEΔ1* and *sulI* genes in Antarctic soils and their linkage to the 3'-CS was not observed, this provided further support to the contention that the *aadA1* gene in this case, was not present on Tn*21* or related transposons.

A model of evolution of In2 has been proposed by Brown *et al.*, (1996) based upon the finding that class 1 integrons In2, In0 and In5, are transposon derivatives defective for

self-transposition. In each of these integrons, an insertion sequence, IS1326, appears to have caused the partial deletion of a former transposition module (Brown *et al.*, 1996). The integrons are postulated to have evolved from a transposon similar to Tn*402*; Tn*402*, a class 1 integron comprises two integrated gene cassettes, *dfrB3* and orfD, followed by the complete *qacE* and does not include a *sulI* gene. Additionally, it contains four orfs downstream of *qacE* that are homologous to genes identified as necessary for transposition of transposon Tn5053. The model proposes that replacement of *dfrB3* and orfD gene cassettes with the *aadA1* gene cassette, and insertion of the *sulI-orf5* DNA sequence into the 3'end of the *qacE* gene, led to the creation of In2. In2 was thus, a highly evolved structure prior to its insertion into Tn*21* (Liebert *et al.*, 1999). Therefore, it is conceivable that the *aadAX* gene cassette identified in this study may represent an ancestral-form of In2 that was not integrated into Tn*21* and subsequently diverged in sequence. This would account for the absence of an identifiable *qacEΔ1-sulI* unit in Antarctic samples. Interestingly, Tn*402* located on plasmid R751 was among the closest database relatives identified for the partial integrase sequences amplified from Antarctic soils.

Antarctic DNA samples were additionally investigated for the presence of plasmids belonging to different incompatability (Inc) groups, which might potentially serve as vehicles for the transfer of integrons in the Antarctic environment. Plasmids were selected for investigation based upon their broad host range in Gram-negative bacteria. However, no amplification of replicon-specific DNA regions of incompatability groups IncP1, IncP9, IncN and IncW was detected in Antarctic samples. Of particular note, the plasmid R751 harbouring Tn402 belongs to the IncP1 group. In a previous study examining the distribution of plasmids among environmental bacteria, IncP plasmids were identified in a variety of soils and manure, whilst plasmids of the IncN and IncW groups were generally observed in manure only (Gotz *et al.*, 1996). Amplicons of approximately the expected size were recovered from four Antarctic DNA samples following amplification with IncQ replicon-specific primers. IncQ plasmids are relatively small in size and can be mobilized when transfer functions are supplied *in trans* by a self-transmissable plasmid (Rawlings & Tietze, 2001). By virtue of these

properties, IncQ plasmids are highly promiscuous and have been identified in a variety of environments including soil (Gotz *et al.*, 1996; Smalla *et al.*, 2000). However, due to failure to confirm these PCR products as genuine IncQ-specific sequences, the presence of IncQ plasmids in the Antarctic environment remains inconclusive.

Thus, in conclusion, whilst the genomic context of the *aadAX* gene cassette remains unknown it can be said that integron structures are present in Antarctic Dry Valley soils. The high sequence similarity values obtained for integrase sequences from either Antarctic Dry Valley soils or other, non-Antarctic habitats suggests these integrons share common ancestry. This is supported by the identification of the *aadAX* gene cassette in both Antarctic and non-Antarctic environments and is consistent with the hypothesis that integrons function as a general gene capture system for bacteria.

References

Alger,S.A., Spaulding,S.A., Shupe,G.H., and McKnight,D.M. (1995) McMurdo LTER: Species composition and spatial distribution of algal mats in Green Creek, Taylor Valley, Antarctica. *Antarctic Journal of the US*: 289-291.

Altschul,S.F., Gish,W., Miller,W., Myers,E.W., and Lipman,D.J. (1990) Basic local alignment search tool. *J Mol.Biol.* **215**: 403-410.

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

Amann,R.I., Krumholz,L., and Stahl,D.A. (1990) Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J Bacteriol.* **172**: 762-770.

Amann,R.I., Ludwig,W., and Schleifer,K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol.Rev.* **59**: 143-169.

Arakawa,Y., Murakami,M., Suzuki,K., Ito,H., Wacharotayankun,R., Ohsuka,S., Kato,N., and Ohta,M. (1995) A novel integron-like element carrying the metallo-beta-lactamase gene blaIMP *Antimicrob Agents Chemother.* **39**: 1612-1615.

Asgarani,E., Terato,H., Asagoshi,K., Shahmohammadi,H.R., Ohyama,Y., Saito,T., Yamamoto,O., and Ide,H. (2000) Purification and characterization of a novel DNA repair enzyme from the extremely radioresistant bacterium *Rubrobacter radiotolerans. J Radiat.Res.(Tokyo)* **41**: 19-34.

Axelsson,M., Davison,W., Forster,M.E., and Farrell,A.P. (1992) Cardiovascular responses of the red-blooded antarctic fishes *Pagothenia bernacchii* and *P. borchgrevinki J Exp.Biol.* **167**: 179-201.

Baharaeen,S., Vishniac,H.S. (1982) A fixation method for visualization of yeast ultrastructure in the electron microscope. *Mycopathologia.* **77**: 19-22.

Barker,A., Clark,C.A., and Manning,P.A. (1994) Identification of VCR, a repeated sequence associated with a locus encoding a hemagglutinin in *Vibrio cholerae* O1. *J Bacteriol.* **176**: 5450-5458.

Barlow,R.S., Pemberton,J.M., Desmarchelier,P.M., and Gobius,K.S. (2004) Isolation and characterization of integron-containing bacteria without antibiotic selection. *Antimicrob Agents Chemother.* **48**: 838-842.

Barns,S.M., Fundyga,R.E., Jeffries,M.W., and Pace,N.R. (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc.Natl.Acad.Sci.U.S.A.* **91**: 1609-1613.

Beyer,L., Bockheim,J.G., Campbell,I.B., and Claridge,G.G.C. (1999) Genesis, properties and sensitivity of Antarctic Gelisols. *Antarctic Science.* **11**: 387-398.

Bidle,K.A., Kastner,M., and Bartlett,D.H. (1999) A phylogenetic analysis of microbial communities associated with methane hydrate containing marine fluids and sediments in the Cascadia margin (ODP site 892B). *FEMS Microbiol Lett.* **177**: 101-108.

Bintrim,S.B., Donohue,T.J., Handelsman,J., Roberts,G.P., and Goodman,R.M. (1997) Molecular phylogeny of Archaea from soil *Proc.Natl.Acad.Sci.U.S.A.* **94**: 277-282.

Bissonnette,L., Roy,P.H. (1992) Characterization of In0 of *Pseudomonas aeruginosa* plasmid pVS1, an ancestor of integrons of multiresistance plasmids and transposons of gram-negative bacteria. *J Bacteriol.* **174**: 1248-1257.

Boetius,A., Ravenschlag,K., Schubert,C.J., Rickert,D., Widdel,F., Gieseke,A., Amann,R., Jorgensen,B.B., Witte,U., and Pfannkuche,O. (2000) A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature.* **407**: 623-626.

Bonnet,R., Suau,A., Dore,J., Gibson,G.R., and Collins,M.D. (2002) Differences in rDNA libraries of faecal bacteria derived from 10- and 25-cycle PCRs. *Int.J Syst.Evol.Microbiol.* **52**: 757-763.

Borneman,J., Skroch,P.W., O'Sullivan,K.M., Palus,J.A., Rumjanek,N.G., Jansen,J.L., Nienhuis,J., and Triplett,E.W. (1996) Molecular microbial diversity of an agricultural soil in Wisconsin. *Appl.Environ.Microbiol.* **62**: 1935-1943.

Borneman,J., Triplett,E.W. (1997) Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Appl.Environ.Microbiol.* **63**: 2647-2653.

Borneman,J. (1999) Culture-independent identification of microorganisms that respond to specified stimuli. *Appl.Environ.Microbiol.* **65**: 3398-3400.

Bowman,J.P., McCammon,S.A., Brown,M.V., Nichols,D.S., and McMeekin,T.A. (1997) Diversity and association of psychrophilic bacteria in Antarctic sea ice. *Appl.Environ.Microbiol.* **63** : 3068-3078.

Bowman,J.P., Rea,S.M., McCammon,S.A., and McMeekin,T.A. (2000) Diversity and community structure within anoxic sediment from marine salinity meromictic lakes and a coastal meromictic marine basin, Vestfold Hilds, Eastern Antarctica. *Environ.Microbiol.* **2**: 227-237.

Bowman,J.P., McCuaig,R.D. (2003) Biodiversity, community structural shifts, and biogeography of prokaryotes within Antarctic continental shelf sediment. *Appl.Environ.Microbiol.* **69** : 2463-2483.

Bowman,J.P., McCammon,S.A., Gibson,J.A., Robertson,L., and Nichols,P.D. (2003) Prokaryotic metabolic activity and community structure in Antarctic continental shelf sediments. *Appl.Environ.Microbiol.* **69**: 2448-2462.

Brambilla,E., Hippe,H., Hagelstein,A., Tindall,B.J., and Stackebrandt,E. (2001) 16S rDNA diversity of cultured and uncultured prokaryotes of a mat sample from Lake Fryxell, McMurdo Dry Valleys, Antarctica. *Extremophiles.* **5**: 23-33.

Brinkmeyer,R., Knittel,K., Jurgens,J., Weyland,H., Amann,R., and Helmke,E. (2003) Diversity and structure of bacterial communities in Arctic versus Antarctic pack ice. *Appl.Environ.Microbiol.* **69**: 6610-6619.

Brown,H.J., Stokes,H.W., and Hall,R.M. (1996) The integrons In0, In2, and In5 are defective transposon derivatives. *J Bacteriol.* **178**: 4429-4437.

Bruni,V., Gugliandolo,C., Maugeri,T., and Allegra,A. (1999) Psychrotrophic bacteria from a coastal station in the Ross sea (Terra Nova Bay, Antarctica). *New Microbiol.* **22**: 357-363.

Buckley,D.H., Graber,J.R., and Schmidt,T.M. (1998) Phylogenetic analysis of nonthermophilic members of the kingdom crenarchaeota and their diversity and abundance in soils. *Appl.Environ.Microbiol.* **64**: 4333-4339.

Buckley,D.H., Schmidt,T.M. (2001) Environmental factors influencing the distribution of rRNA from Verrucomicrobia in soil. *FEMS Microbiol.Ecol.* **35**: 105-112.

Buckley,D.H., Schmidt,T.M. (2003) Diversity and dynamics of microbial communities in soils from agro-ecosystems *Environ.Microbiol.* **5**: 441-452.

Cameron,R.E., Conrow,H.P. (1969) Soil mositure, relative humidity, and microbial abundance in Dry Valleys of Southern Victoria Land. *Antarctic Journal of the US.* **4**: 23-28.

Cameron,R.E., Conrow,H.P., Gensel,D.R., Lacy,G.H., and Morelli,F.A. (1971) Surface distribution of microorganisms in Antarctic Dry Valley soils: a Martian analogue. *Antarctic Journal of the US*: 211-213.

Cameron,R.E., Ford,A,B. (1974) Bseline analyses of soils from the Pensacola Mountains. *Antarctic Journal of the US* **9**: 116-119.

Cameron,R.E., Morelli,F.A., and Johnson,R.M. (1972) Bacterial species in soil and air of the Antarctic Continent. *Antarctic Journal*: 187-189.

Cameron,R.E. (1972) Microbiological and ecological investigations in Victoria Valley, Southern Victoria Land, Antarctica. *Antarctic Research Series* **20**: 195-220.

Campbell,I.B., Claridge,G.G.C. (1987) *The Biology of Antarctic Soils. In Antarctica: Soils, Weathering Processes and Environment.* Campbell,I.B., Claridge,G.G.C. (eds)., pp. 73-96.

Canton,R., Coque,T.M., and Baquero,F. (2003) Multi-resistant Gram-negative bacilli: from epidemics to endemics. *Curr.Opin.Infect.Dis.* **16**: 315-325.

Carpenter,E.J., Lin,S., and Capone,D.G. (2000) Bacterial activity in South Pole snow. *Appl.Environ.Microbiol.* **66**: 4514-4517.

Carreto,L., Moore,E., Nobre,M.F., Wait,R., Riley,P.W., Sharp,R.J., and da Costa,M.S. (1996) *Rubrobacter xylanophilus* sp.nov., a new thermophilic species isolated from a thermally polluted effluent. *Int.J Syst.Bacteriol.* **46**: 460-465.

Chandler,D.P., Brockman,F.J., Bailey,T.J., and Fredrickson,J.K. (1998) Phylogenetic Diversity of Archaea and Bacteria in a Deep Subsurface Paleosol. *Microb.Ecol.* **36**: 37-50.

Charlat,S., Hurst,G.D., and Mercot,H. (2003) Evolutionary consequences of Wolbachia infections. *Trends Genet.* **19**: 217-223.

Chin,K.J., Hahn,D., Hengstmann,U., Liesack,W., and Janssen,P.H. (1999) Characterization and identification of numerically abundant culturable bacteria from the anoxic bulk soil of rice paddy microcosms. *Appl.Environ.Microbiol.* **65**: 5042-5049.

Cho,J.C., Tiedje,J.M. (2000) Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. *Appl.Environ.Microbiol.* **66**: 5448-5456.

Christner,B.C., Mosley-Thompson,E., Thompson,L.G., and Reeve,J.N. (2001) Isolation of bacteria and 16S rDNAs from Lake Vostok accretion ice. *Environ.Microbiol.* **3**: 570-577.

Christner,B.C. Recovery of bacteria from glacial and subglacial environments. Thesis (2002) Ohio State University, 484 W. 12th Avenue, Columbus,OH, USA.

Christner,B.C., Kvitko,B.H., and Reeve,J.N. (2003a) Molecular identification of bacteria and Eukarya inhabiting an Antarctic cryoconite hole. *Extremophiles.* **7**: 177-183.

Christner,B.C., Mosley-Thompson,E., Thompson,L.G., and Reeve,J.N. (2003b) Bacterial recovery from ancient glacial ice. *Environ.Microbiol.* **5**: 433-436.

Clark,C.A., Purins,L., Kaewrakon,P., and Manning,P.A. (1997) VCR repetitive sequence elements in the *Vibrio cholerae* chromosome constitute a mega-integron. *Mol.Microbiol.* **26**: 1137-1138.

Clark,C.A., Purins,L., Kaewrakon,P., Focareta,T., and Manning,P.A. (2000) The *Vibrio cholerae* O1 chromosomal integron. *Microbiology* **146 (Pt 10)**: 2605-2612.

Clow,G.D., McKay,C.P., Simmons,G.M., Jr., and Wharton,R.A., Jr. (1988) Climatological observations and predicted sublimation rates at Lake Hoare, Antarctica *J Clim.* **1**: 715-728.

Coker,J.A., Sheridan,P.P., Loveland-Curtze,J., Gutshall,K.R., Auman,A.J., and Brenchley,J.E. (2003) Biochemical characterization of a beta-galactosidase with a low temperature optimum obtained from an Antarctic *arthrobacter* isolate. *J Bacteriol.* **185**: 5473-5482.

Collis,C.M., Hall,R.M. (1992a) Gene cassettes from the insert region of integrons are excised as covalently closed circles. *Mol.Microbiol.* **6**: 2875-2885.

Collis,C.M., Hall,R.M. (1992b) Site-specific deletion and rearrangement of integron insert genes catalyzed by the integron DNA integrase. *J Bacteriol.* **174**: 1574-1585.

Collis,C.M., Hall,R.M. (1995) Expression of antibiotic resistance genes in the integrated cassettes of integrons. *Antimicrob Agents Chemother* **39**: 155-162.

Collis,C.M., Kim,M.J., Stokes,H.W., and Hall,R.M. (1998) Binding of the purified integron DNA integrase Intl1 to integron- and cassette-associated recombination sites. *Mol.Microbiol.* **29**: 477-490.

Cowan,D.A., Russell,N.J., Mamais,A., and Sheppard,D.M. (2002) Antarctic Dry Valley mineral soils contain unexpectedly high levels of microbial biomass. *Extremophiles.* **6**: 431-436.

Crocetti,G.R., Banfield,J.F., Keller,J., Bond,P.L., and Blackall,L.L. (2002) Glycogen-accumulating organisms in laboratory-scale and full-scale wastewater treatment processes. *Microbiology.* **148**: 3353-3364.

de la Cruz.F., Davies,J. (2000) Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol.* **8**: 128-133.

Dalsgaard,A., Forslund,A., Serichantalergs,O., and Sandvang,D. (2000) Distribution and content of class 1 integrons in different *Vibrio cholerae* O-serotype strains isolated in Thailand. *Antimicrob Agents Chemother* **44**: 1315-1321.

Davies,J., Wright,G.D. (1997) Bacterial resistance to aminoglycoside antibiotics. *Trends Microbiol.* **5**: 234-240.

DeLong,D.C. (1996) Defining biodiversity. *Wildlife Soc Bull* **24**: 738-749.

Delong,E.F. (1992) Archaea in coastal marine environments. *Proc.Natl.Acad.Sci.U.S.A* **89**: 5685-5689.

Delong,E.F., Wu,K.Y., Prezelin,B.B., and Jovine,R.V. (1994) High abundance of Archaea in Antarctic marine picoplankton. *Nature* **371**: 695-697.

Delong,E.F. (1998) Everything in moderation: archaea as 'non-extremophiles'. *Curr.Opin.Genet.Dev.* **8** : 649-654.

Delong,E.F. (2003) Oceans of *Archaea. ASM News* **69**: 503-511.

Derakshani,M., Lukow,T., and Liesack,W. (2001) Novel bacterial lineages at the (sub)division level as detected by signature nucleotide-targeted recovery of 16S rRNA genes from bulk soil and rice roots of flooded rice microcosms. *Appl.Environ.Microbiol.* **67**: 623-631.

Diez,B., Pedros-Alio,C., and Massana,R. (2001a) Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Appl.Environ.Microbiol* **67**: 2932-2941.

Diez,B., Pedros-Alio,C., Marsh,T.L., and Massana,R. (2001b) Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques. *Appl.Environ.Microbiol* **67**: 2942-2951.

Dojka,M.A., Hugenholtz,P., Haack,S.K., and Pace,N.R. (1998) Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Appl.Environ.Microbiol.* **64**: 3869-3877.

Dose,K., Stridde,C., Dillmann,R., Risi,S., and Bieger-Dose,A. (1995) Biochemical constraints for survival under Martian conditions. *Adv.Space Res.* **15**: 203-210.

Dunbar,J., Barns,S.M., Ticknor,L.O., and Kuske,C.R. (2002) Empirical and theoretical bacterial diversity in four Arizona soils. *Appl.Environ.Microbiol.* **68**: 3035-3045.

Edwards,U., Rogall,T., Blocker,H., Emde,M., and Bottger,E.C. (1989) Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res.* **17**: 7843-7853.

Elisha,B.G., Steyn,L.M. (1994) High level kanamycin resistance associated with the hyperproduction of AAC(3)II and a generalised reduction in the accumulation of aminoglycosides in *Acinetobacter* spp. *J Antimicrob Chemother* **34**: 457-464.

Ellis,R.J., Morgan,P., Weightman,A.J., and Fry,J.C. (2003) Cultivation-dependent and -independent approaches for determining bacterial diversity in heavy-metal-contaminated soil. *Appl.Environ.Microbiol.* **69**: 3223-3230.

Esposito,D., Scocca,J.J. (1997) The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res.* **25**: 3605-3614.

Etchebehere,C., Errazquin,M.I., Cabezas,A., Planzzola,M.J., Mallo,M., Lombardi,P., Ottonello,G., Borzaccŏni,L., and Muxi,L. (2002) Sludge bed development in denitrifying reactors using different inocula-performance and microbiological aspects. *Water Sci.Technol.* **45**: 365-370.

Farrell,R.L., Rhodes,P.L., and Aislabie,J. (2003) Toluene-degrading Antarctic Pseudomonas strains from fuel-contaminated soil. *Biochem.Biophys.Res.Commun.* **312**: 235-240.

Felske,A., Engelen,B., Nubel,U., and Backhaus,H. (1996) Direct ribosome isolation from soil to extract bacterial rRNA for community analysis. *Appl.Environ.Microbiol.* **62**: 4162-4167.

Felske,A., Rheims,H., Wolterink,A., Stackebrandt,E., and Akkermans,A.D. (1997) Ribosome analysis reveals prominent activity of an uncultured member of the class Actinobacteria in grassland soils. *Microbiology* **143 (Pt 9)**: 2983-2989.

Ferreira,A.C., Nobre,M.F., Moore,E., Rainey,F.A., Battista,J.R., and da Costa,M.S. (1999) Characterization and radiation resistance of new isolates of *Rubrobacter radiotolerans* and *Rubrobacter xylanophilus. Extremophiles.* **3**: 235-238.

Friedman,E.I. (1982) Endolithic microorganisms in the Antarctic desert. *Science* **215**: 1045-1053.

Friedmann,E.I., Ocampo,R. (1976) Endolithic blue-green algae in the dry valleys: primary producers in the Antarctic desert ecosystem. *Science* **193**: 1247-1249.

Friedrich,M.W., Schmitt-Wagner,D., Lueders,T., and Brune,A. (2001) Axial differences in community structure of Crenarchaeota and Euryarchaeota in the highly compartmentalized gut of the soil-feeding termite *Cubitermes orthognathus.* *Appl.Environ.Microbiol.* **67**: 4880-4890.

Furlong,M.A., Singleton,D.R., Coleman,D.C., and Whitman,W.B. (2002) Molecular and culture-based analyses of prokaryotic communities from an agricultural soil and the burrows and casts of the earthworm *Lumbricus rubellus*. *Appl.Environ.Microbiol.* **68**: 1265-1279.

Gerner-Smidt,P., Keiser-Nielsen,H., Dorsch,M., Stackebrandt,E., Ursing,J., Blom,J., Christensen,A.C., Christensen,J.J., Frederiksen,W., Hoffmann,S., and . (1994) *Lautropia mirabilis* gen. nov., sp. nov., a gram-negative motile coccus with unusual morphology isolated from the human mouth. *Microbiology* **140 (Pt 7)**: 1787-1797.

Giovannoni,S.J., Godchaux,W., III, Schabtach,E., and Castenholz,R.W. (1987) Cell wall and lipid composition of *Isosphaera pallida*, a budding eubacterium from hot springs *J Bacteriol.* **169**: 2702-2707.

Goebel,B.M., Stackebrandt,E. (1994) Cultural and phylogenetic analysis of mixed microbial populations found in natural and commercial bioleaching environments. *Appl.Environ.Microbiol.* **60**: 1614-1621.

Gomes,N.C., Fagbola,O., Costa,R., Rumjanek,N.G., Buchner,A., Mendona-Hagler,L., and Smalla,K. (2003) Dynamics of fungal communities in bulk and maize rhizosphere soil in the tropics. *Appl.Environ.Microbiol.* **69**: 3758-3766.

Gordon,D.A., Priscu,J., and Giovannoni,S. (2000) Origin and Phylogeny of Microbes Living in Permanent Antarctic Lake Ice. *Microb.Ecol.* **39**: 197-202.

Gotz,A., Pukall,R., Smit,E., Tietze,E., Prager,R., Tschape,H., van Elsas,J.D., and Smalla,K. (1996) Detection and characterization of broad-host-range plasmids in environmental bacteria by PCR. *Appl.Environ.Microbiol.* **62**: 2621-2628.

Greated,A., Thomas,C.M. (1999) A pair of PCR primers for Incp-9 plasmids. *Microbiology.* **145 (Pt 11)** : 3003-3004.

Gregory,L.G., Bond,P.L., Richardson,D.J., and Spiro,S. (2003) Characterization of a nitrate-respiring bacterial community using the nitrate reductase gene (narG) as a functional marker. *Microbiology.* **149**: 229-237.

Grigioni,S., Boucher-Rodoni,R., Demarta,A., Tonolla,M., and Peduzzi,R. (2000) Phylogenetic characterisation of bacterial symbionts in the accessory nidamental glands of the Sepioid *Sepia officinalis*. *Mar Biol.* **136**: 217-222.

Hahn,M.W., Lunsdorf,H., Wu,Q., Schauer,M., Hofle,M.G., Boenigk,J., and Stadler,P. (2003) Isolation of novel ultramicrobacteria classified as actinobacteria from five freshwater habitats in Europe and Asia. *Appl.Environ.Microbiol.* **69**: 1442-1451.

Hall,R.M., Vockler,C. (1987) The region of the IncN plasmid R46 coding for resistance to beta-lactam antibiotics, streptomycin/spectinomycin and sulphonamides is closely related to antibiotic resistance segments found in IncW plasmids and in Tn21-like transposons. *Nucleic Acids Res.* **15**: 7491-7501.

Hall,R.M., Brookes,D.E., and Stokes,H.W. (1991) Site-specific insertion of genes into integrons: role of the 59-base element and determination of the recombination cross-over point. *Mol.Microbiol.* **5**: 1941-1959.

Hall,R.M., Collis,C.M. (1995) Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol.Microbiol.* **15**: 593-600.

Hall, R.M. Mobile gene cassettes and integrons: moving antibiotic resistance genes in gram-negative bacteria, Ciba Foundation Symposium 207 (1997) 192-202; discussion 202-205.

Head,I.M., Saunders,J.R., and Pickup,R.W. (1998) Microbial Evolution, Diversity, and Ecology: A Decade of Ribosomal RNA Analysis of Uncultivated Microorganisms. *Microb.Ecol.* **35**: 1-21.

Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L., Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D., Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I., Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D., Nierman,W.C., and White,O. (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature.* **406**: 477-483.

Heinaru,E., Truu,J., Stottmeister,U., and Heinaru,A. (2000) Three types of phenol and p-cresol catabolism in phenol- and p-cresol-degrading bacteria isolated from river water continuously polluted with phenolic compounds. *FEMS Microbiol.Ecol.* **31**: 195-205.

Henne,A., Daniel,R., Schmitz,R.A., and Gottschalk,G. (1999) Construction of environmental DNA libraries in Escherichia coli and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl.Environ.Microbiol.* **65**: 3901-3907.

Henne,A., Schmitz,R.A., Bomeke,M., Gottschalk,G., and Daniel,R. (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl.Environ.Microbiol.* **66**: 3113-3116.

Hershberger,K.L., Barns,S.M., Reysenbach,A.L., Dawson,S.C., and Pace,N.R. (1996) Wide diversity of Crenarchaeota. *Nature.* **384**: 420.

Hill,G.T., Mitkowski,L., Aldrich-Wolfe,L., Emele,L.R., Jurkonie,D.D., Ficke,A., Maldonado-Ramirez,S., Lynch,S.T., and Nelson,E.B. (2000) Methods for assessing the composition and diversity of soil microbial communities. *Appl.Soil Ecol.* **15**: 25-36.

Hirsch,P., Gallikowski,C.C., and Friedman,E.I. (1985) Microorganisms in soil samples from Linnaeus Terrace southern Victoria Land: preliminary observations. *Antarctic Journal of the US* **20**: 183-186.

Hirsch,P., Ludwig,W., Hethke,C., Sittig,M., Hoffmann,B., and Gallikowski,C.A. (1998) *Hymenobacter roseosalivarius* gen. nov., sp. nov. from continental Antartica soils and sandstone: bacteria of the *Cytophaga/Flavobacterium/Bacteroides* line of phylogenetic descent. *Syst.Appl.Microbiol.* **21**: 374-383.

Holmes,A.J., Bowyer,J., Holley,M.P., O'Donoghue,M., Montgomery,M., and Gillings,M.R. (2000) Diverse, yet-to-be-cultured members of the Rubrobacter subdivision of the *Actinobacteria* are widespread in Australian arid soils. *FEMS Microbiol.Ecol.* **33**: 111-120.

Holmes,A.J., Gillings,M.R., Nield,B.S., Mabbutt,B.C., Nevalainen,K.M., and Stokes,H.W. (2003a) The gene cassette metagenome is a basic resource for bacterial genome evolution. *Environ.Microbiol.* **5**: 383-394.

Holmes,A.J., Holley,M.P., Mahon,A., Nield,B., Gillings,M., and Stokes,H.W. (2003b) Recombination activity of a distinctive integron-gene cassette system associated with *Pseudomonas stutzeri* populations in soil. *J Bacteriol.* **185**: 918-928.

Horowitz, N.H., and R. E. Cameron. 1971. Microbiology of the Dry Valleys of Antarctica. (Abstract.) Committee on Space Research. International Council of Scientific Unions. 14th Plenary Meeting, Seattle, Washington, June 17-July 2. Referred to Cameron *et al.*, 1971.

Horowitz,N.H., Cameron,R.E., and Hubbard,J.S. (1972) Microbiology of the Dry Valleys of Antarctica. *Science* **176**: 242-245.

Hugenholtz,P., Goebel,B.M., and Pace,N.R. (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol.* **180**: 4765-4774.

Hugenholtz,P., Tyson,G.W., Webb,R.I., Wagner,A.M., and Blackall,L.L. (2001) Investigation of candidate division TM7, a recently recognized major lineage of the domain Bacteria with no known pure-culture representatives. *Appl.Environ.Microbiol.* **67**: 411-419.

Hugenholtz,P. (2002) Exploring prokaryotic diversity in the genomic era. *Genome Biol.* **3**: REVIEWS0003.

Hughes,K.A., Lawley,B. (2003) A novel Antarctic microbial endolithic community within gypsum crusts. *Environ.Microbiol.* **5**: 555-565.

Hughes,K.A., Lawley,B., and Newsham,K.K. (2003) Solar UV-B radiation inhibits the growth of Antarctic terrestrial fungi. *Appl.Environ.Microbiol.* **69**: 1488-1491.

Janssen,P.H., Schuhmann,A., Morschel,E., and Rainey,F.A. (1997) Novel anaerobic ultramicrobacteria belonging to the Verrucomicrobiales lineage of bacterial descent isolated by dilution culture from anoxic rice paddy soil. *Appl.Environ.Microbiol.* **63**: 1382-1388.

Janssen,P.H., Yates,P.S., Grinton,B.E., Taylor,P.M., and Sait,M. (2002) Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions *Acidobacteria*, *Actinobacteria*, *Proteobacteria*, and *Verrucomicrobia*. *Appl.Environ.Microbiol.* **68**: 2391-2396.

Junge,K., Gosink,J.J., Hoppe,H.G., and Staley,J.T. (1998) *Arthrobacter*, *Brachybacterium* and *Planococcus* isolates identified from antarctic sea ice brine. Description of *Planococcus mcmeekinii*, sp. nov. *Syst.Appl.Microbiol.* **21**: 306-314.

Jurgens,G., Lindstrom,K., and Saano,A. (1997) Novel group within the kingdom Crenarchaeota from boreal forest soil. *Appl.Environ.Microbiol.* **63**: 803-805.

Jurgens,G., Glockner,F., Amann,R., Saano,A., Montonen,L., Likolammi,M., and Munster,U. (2000) Identification of novel Archaea in bacterioplankton of a boreal forest lake by phylogenetic analysis and fluorescent in situ hybridization(1). *FEMS Microbiol.Ecol.* **34**: 45-56.

Kaeberlein,T., Lewis,K., and Epstein,S.S. (2002) Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* **296**: 1127-1129.

Kato,C., Li,L., Tamaoka,J., and Horikoshi,K. (1997) Molecular analyses of the sediment of the 11,000-m deep Mariana Trench. *Extremophiles.* **1**: 117-123.

Kazama,H., Kizu,K., Iwasaki,M., Hamashima,H., Sasatsu,M., and Arai,T. (1995) Isolation and structure of a new integron that includes a streptomycin resistance gene from the R plasmid of *Pseudomonas aeruginosa*. *FEMS Microbiol.Lett.* **134**: 137-141.

Kazama,H., Hamashima,H., Sasatsu,M., and Arai,T. (1998) Distribution of the antiseptic-resistance genes qacE and qacE delta 1 in gram-negative bacteria. *FEMS Microbiol.Lett.* **159**: 173-178.

Keys, J.R. 1980. Air Temperature, Wind, Precipitation, and Atmospheric Humidity in the McMurdo Region. Department of Geology Publication 17 (Antarctic Data Series 9). Victoria University of Wellington, New Zealand. 52 rpp.

Koch,A.L. (1997) Microbial physiology and ecology of slow growth. *Microbiol.Mol.Biol.Rev.* **61**: 305-318.

Konstantinidis,K.T., Isaacs,N., Fett,J., Simpson,S., Long,D.T., and Marsh,T.L. (2003) Microbial diversity and resistance to copper in metal-contaminated lake sediment. *Microb.Ecol.* **45**: 191-202.

Kuske,C.R., Barns,S.M., and Busch,J.D. (1997) Diverse uncultivated bacterial groups from soils of the arid southwestern United States that are present in many geographic regions. *Appl.Environ.Microbiol.* **63**: 3614-3621.

Kvenvolden,K.A. (1999) Potential effects of gas hydrate on human welfare. *Proc.Natl.Acad.Sci.U.S.A* **96**: 3420-3426.

La Duc,M.T., Nicholson,W., Kern,R., and Venkateswaran,K. (2003) Microbial characterization of the Mars Odyssey spacecraft and its encapsulation facility. *Environ.Microbiol.* **5**: 977-985.

Lanoil,B.D., Sassen,R., La Duc,M.T., Sweet,S.T., and Nealson,K.H. (2001) Bacteria and Archaea physically associated with Gulf of Mexico gas hydrates. *Appl.Environ.Microbiol.* **67**: 5143-5153.

Lauretti,L., Riccio,M.L., Mazzariol,A., Cornaglia,G., Amicosante,G., Fontana,R., and Rossolini,G.M. (1999) Cloning and characterization of blaVIM, a new integron-borne metallo-beta-lactamase gene from a *Pseudomonas aeruginosa* clinical isolate. *Antimicrob Agents Chemother.* **43**: 1584-1590.

Lawrence,J.G., Ochman,H. (1998) Molecular archaeology of the *Escherichia coli* genome. *Proc.Natl.Acad.Sci.U.S.A.* **95**: 9413-9417.

Leon,G., Roy,P.H. (2003) Excision and integration of cassettes by an integron integrase of *Nitrosomonas europaea. J Bacteriol.* **185**: 2036-2041.

Leverstein-van Hall,M.A., Box,A.T., Blok,H.E., Paauw,A., Fluit,A.C., and Verhoef,J. (2002) Evidence of extensive interspecies transfer of integron-mediated antimicrobial resistance genes among multidrug-resistant *Enterobacteriaceae* in a clinical setting. *J Infect.Dis.* **186**: 49-56.

Levesque,C., Piche,L., Larose,C., and Roy,P.H. (1995) PCR mapping of integrons reveals several novel combinations of resistance genes. *Antimicrob Agents Chemother.* **39**: 185-191.

Leys,N.M., Ryngaert,A., Bastiaens,L., Verstraete,W., Top,E.M., and Springael,D. (2004) Occurrence and phylogenetic diversity of Sphingomonas strains in soils contaminated with polycyclic aromatic hydrocarbons. *Appl.Environ.Microbiol.* **70**: 1944-1955.

Li,L., Kato,C., and Horikoshi,K. (1999). Microbial Diversity in Sediments Collected from the Deepest Cold-Seep Area, the Japan Trench. **1**: 391-400.

Liebert,C.A., Hall,R.M., and Summers,A.O. (1999) Transposon Tn21, flagship of the floating genome. *Microbiol.Mol.Biol.Rev.* **63**: 507-522.

Liesack,W., Stackebrandt,E. (1992) Occurrence of novel groups of the domain Bacteria as revealed by analysis of genetic material isolated from an Australian terrestrial environment. *J Bacteriol.* **174**: 5072-5078.

Liles,M.R., Manske,B.F., Bintrim,S.B., Handelsman,J., and Goodman,R.M. (2003) A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Appl.Environ.Microbiol.* **69**: 2684-2691.

Lio,P., Goldman,N. (1998) Models of molecular evolution and phylogeny. *Genome Res.* **8**: 1233-1244.

Logan,N.A., Lebbe,L., Hoste,B., Goris,J., Forsyth,G., Heyndrickx,M., Murray,B.L., Syme,N., Wynn-Williams,D.D., and De Vos,P. (2000) Aerobic endospore-forming bacteria from geothermal environments in northern Victoria Land, Antarctica, and Candlemas Island, South Sandwich archipelago, with the proposal of *Bacillus fumarioli* sp. nov. *Int.J Syst.Evol.Microbiol.* **50 Pt 5**: 1741-1753.

Lonhienne,T., Mavromatis,K., Vorgias,C.E., Buchon,L., Gerday,C., and Bouriotis,V. (2001) Cloning, sequences, and characterization of two chitinase genes from the Antarctic *Arthrobacter* sp. strain TAD20: isolation and partial characterization of the enzymes. *J Bacteriol.* **183**: 1773-1779.

Lopez-Garcia,P., Philippe,H., Gail, F., Moreira, D. (2003) Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proc.Natl.Acad.Sci.U.S.A* **2**: 697-702.

Lopez-Garcia,P, Rodriguez-Valera. F., Pedros-Alio. C., Moreira D. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature. **409** :603-7.

Ludwig,W., Bauer,S.H., Bauer,M., Held,I., Kirchhof,G., Schulze,R., Huber,I., Spring,S., Hartmann,A., and Schleifer,K.H. (1997) Detection and in situ identification of representatives of a widely distributed new bacterial phylum. *FEMS Microbiol.Lett.* **153**: 181-190.

MacGregor,B.J., Moser,D.P., Alm,E.W., Nealson,K.H., and Stahl,D.A. (1997) Crenarchaeota in Lake Michigan sediment. *Appl.Environ.Microbiol.* **63**: 1178-1181.

Manefield,M., Whiteley,A.S., Griffiths,R.I., and Bailey,M.J. (2002) RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. *Appl.Environ.Microbiol.* **68**: 5367-5373.

Mannisto,M.K., Schumann,P., Rainey,F.A., Kampfer,P., Tsitko,I., Tiirola,M.A., and Salkinoja-Salonen,M.S. (2000) *Subtercola boreus* gen. nov., sp. nov. and *Subtercola frigoramans* sp. nov., two new psychrophilic actinobacteria isolated from boreal groundwater. *Int.J Syst.Evol.Microbiol.* **50 Pt 5**: 1731-1739.

Marchesi,J.R., Weightman,A.J., Cragg,B.A., Parkes,R.J., and Fry,J.C. (2001) Methanogen and bacterial diversity and distribution in deep gas hydrate sediments from the Cascadia Margin as revealed by 16S rRNA molecular analysis. *FEMS Microbiol Ecol.* **34**: 221-228.

Margesin,R., Sproer,C., Schumann,P., and Schinner,F. (2003) *Pedobacter cryoconitis* sp. nov., a facultative psychrophile from alpine glacier cryoconite. *Int.J Syst.Evol.Microbiol.* **53**: 1291-1296.

Martinez-Freijo,P., Fluit,A.C., Schmitz,F.J., Verhoef,J., and Jones,M.E. (1999) Many class I integrons comprise distinct stable structures occurring in different species of *Enterobacteriaceae* isolated from widespread geographic regions in Europe. *Antimicrob Agents Chemother.* **43**: 686-689.

Massana,R., Murray,A.E., Preston,C.M., and Delong,E.F. (1997) Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Appl.Environ.Microbiol.* **63**: 50-56.

Mattimore,V., Battista,J.R. (1996) Radioresistance of *Deinococcus radiodurans*: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. *J Bacteriol.* **178**: 633-637.

Mazel,D., Dychinco,B., Webb,V.A., and Davies,J. (1998) A distinctive class of integron in the *Vibrio cholerae* genome. *Science* **280**: 605-608.

McCaig,A.E., Glover,L.A., and Prosser,J.I. (1999) Molecular analysis of bacterial community structure and diversity in unimproved and improved upland grass pastures. *Appl.Environ.Microbiol.* **65**: 1721-1730.

McKay,C., Mellon,M.T., and Friedmann,E.I. (1998) Soil temperatures and stability of ice-cemented ground in the McMurdo Dry Valleys, Antarctica. *Antarct.Sci.* **10**: 31-38.

Medlin,L., Elwood,H.J., Stickel,S., and Sogin,M.L. (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **71**: 491-499.

Michaud,L., Di Cello,F., Brilli,M., Fani,R., Lo,G.A., and Bruni,V. (2004) Biodiversity of cultivable psychrotrophic marine bacteria isolated from Terra Nova Bay (Ross Sea, Antarctica). *FEMS Microbiol.Lett.* **230**: 63-71.

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

Miteva,V.I., Sheridan,P.P., and Brenchley,J.E. (2004) Phylogenetic and physiological diversity of microorganisms isolated from a deep greenland glacier ice core. *Appl.Environ.Microbiol.* **70**: 202-213.

Moon-van der Staay SY, De Wachter,R., and Vaulot,D. (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409**: 607-610.

Moreira,D., Lopez-Garcia,P. (2002) The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends Microbiol* **10**: 31-38.

Munson,M.A., Nedwell,D.B., and Embley,T.M. (1997) Phylogenetic diversity of Archaea in sediment samples from a coastal salt marsh. *Appl.Environ.Microbiol.* **63**: 4729-4733.

Murray,A.E., Preston,C.M., Massana,R., Taylor,L.T., Blakis,A., Wu,K., and Delong,E.F. (1998) Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island, Antarctica. *Appl.Environ.Microbiol.* **64**: 2585-2595.

Muyzer,G., de Waal,E.C., and Uitterlinden,A.G. (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl.Environ.Microbiol.* **59**: 695-700.

Naas,T., Poirel,L., and Nordmann,P. (1999) Molecular characterisation of In51, a class 1 integron containing a novel aminoglycoside adenylyltransferase gene cassette, *aadA6*, in *Pseudomonas aeruginosa*. *Biochim.Biophys.Acta* **1489**: 445-451.

NAKAYA,R., NAKAMURA,A., and MURATA,Y. (1960) Resistance transfer agents in *Shigella*. *Biochem Biophys.Res.Commun.* **3**: 654-659.

Nauhaus,K., Boetius,A., Kruger,M., and Widdel,F. (2002) *In vitro* demonstration of anaerobic oxidation of methane coupled to sulphate reduction in sediment from a marine gas hydrate area. *Environ.Microbiol.* **4**: 296-305.

Nemergut,D.R., Martin,A.P., and Schmidt,S.K. (2004) Integron diversity in heavy-metal-contaminated mine tailings and inferences about integron evolution. *Appl.Environ.Microbiol.* **70**: 1160-1168.

Nesvera,J., Hochmannova,J., and Patek,M. (1998) An integron of class 1 is present on the plasmid pCG4 from gram-positive bacterium *Corynebacterium glutamicum*. *FEMS Microbiol.Lett.* **169**: 391-395.

Nield,B.S., Holmes,A.J., Gillings,M.R., Recchia,G.D., Mabbutt,B.C., Nevalainen,K.M., and Stokes,H.W. (2001) Recovery of new integron classes from environmental DNA. *FEMS Microbiol.Lett.* **195**: 59-65.

Nienow,J.A., Friedman,E.I. (1993) Terrestrial lithophytic (rock) communities. In *Antarctic Microbiology*. Friedman,E.I. (ed). New York: Wiley-Liss, pp. 343-412.

Nogales,B., Moore,E.R., Abraham,W.R., and Timmis,K.N. (1999) Identification of the metabolically active members of a bacterial community in a polychlorinated biphenyl-polluted moorland soil. *Environ.Microbiol.* **1**: 199-212.

Nogales,B., Moore,E.R., Llobet-Brossa,E., Rossello-Mora,R., Amann,R., and Timmis,K.N. (2001) Combined use of 16S ribosomal DNA and 16S rRNA to study the bacterial community of polychlorinated biphenyl-polluted soil. *Appl.Environ.Microbiol.* **67**: 1874-1884.

Nunes-Duby,S.E., Kwon,H.J., Tirumalai,R.S., Ellenberger,T., and Landy,A. (1998) Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.* **26**: 391-406.

O'Farrell,K.A., Janssen,P.H. (1999) Detection of *Verrucomicrobia* in a pasture soil by PCR-mediated amplification of 16S rRNA genes. *Appl.Environ.Microbiol.* **65**: 4280-4284.

O'Sullivan,L.A., Weightman,A.J., and Fry,J.C. (2002) New degenerate *Cytophaga-Flexibacter-Bacteroides*-specific 16S ribosomal DNA-targeted oligonucleotide probes reveal high bacterial diversity in River Taff epilithon. *Appl.Environ.Microbiol.* **68**: 201-210.

Ochman,H., Lawrence,J.G., and Groisman,E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature.* **405**: 299-304.

Ochsenreiter,T., Selezi,D., Quaiser,A., Bonch-Osmolovskaya,L., and Schleper,C. (2003) Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environ.Microbiol.* **5**: 787-797.

Osborn,A.M., Bruce,K.D., Strike,P., and Ritchie,D.A. (1997) Distribution, diversity and evolution of the bacterial mercury resistance (mer) operon. *FEMS Microbiol.Rev.* **19**: 239-262.

Pace,N.R. (1997) A molecular view of microbial diversity and the biosphere. *Science* **276**: 734-740.

Partridge,S.R., Recchia,G.D., Scaramuzzi,C., Collis,C.M., Stokes,H.W., and Hall,R.M. (2000) Definition of the attI1 site of class 1 integrons. *Microbiology.* **146 (Pt 11)**: 2855-2864.

Partridge,S.R., Recchia,G.D., Stokes,H.W., and Hall,R.M. (2001) Family of class 1 integrons related to In4 from Tn1696. *Antimicrob Agents Chemother.* **45**: 3014-3020.

Partridge,S.R., Collis,C.M., and Hall,R.M. (2002) Class 1 integron containing a new gene cassette, *aadA10*, associated with Tn1404 from R151. *Antimicrob Agents Chemother.* **46**: 2400-2408.

Paster,B.J., Boches,S.K., Galvin,J.L., Ericson,R.E., Lau,C.N., Levanos,V.A., Sahasrabudhe,A., and Dewhirst,F.E. (2001) Bacterial diversity in human subgingival plaque. *J Bacteriol.* **183**: 3770-3783.

Paulsen,I.T., Littlejohn,T.G., Radstrom,P., Sundstrom,L., Skold,O., Swedberg,G., and Skurray,R.A. (1993) The 3' conserved segment of integrons contains a gene associated with multidrug resistance to antiseptics and disinfectants. *Antimicrob Agents Chemother.* **37**: 761-768.

Pearce,D.A. (2003) Bacterioplankton community structure in a maritime antarctic oligotrophic lake during a period of holomixis, as determined by denaturing gradient gel electrophoresis (DGGE) and fluorescence in situ hybridization (FISH). *Microb.Ecol.* **46**: 92-105.

Pearson,W.R. (1994) Using the FASTA program to search protein and DNA sequence databases. *Methods Mol.Biol.* **24**: 307-331.

Philippot,L., Piutti,S., Martin-Laurent,F., Hallet,S., and Germon,J.C. (2002) Molecular analysis of the nitrate-reducing community from unplanted and maize-planted soils. *Appl.Environ.Microbiol.* **68**: 6121-6128.

Philippot,L. (2002) Denitrfying genes in bacteria and Archaeal genomes. *Biochim.Biophys.Acta.* **1577**: 355-376.

Phillips,A., Janies,D., and Wheeler,W. (2000) Multiple sequence alignment in phylogenetic analysis. *Mol.Phylogenet.Evol.* **16**: 317-330.

Poirel,L., Le,T., I, Naas,T., Karim,A., and Nordmann,P. (2000) Biochemical sequence analyses of GES-1, a novel class A extended-spectrum beta-lactamase, and the class 1 integron In52 from *Klebsiella pneumoniae*. *Antimicrob Agents Chemother.* **44**: 622-632.

Postgate, J.R., Hunter, J.R. (1964) Accelerated death of Aerobacter aerogenes starved in the presence of growth-limiting substrates. *J Gen.Microbiol.* **34**: 459-473.

Preston,C.M., Wu,K.Y., Molinski,T.F., and Delong,E.F. (1996) A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proc.Natl.Acad.Sci.U.S.A.* **93**: 6241-6246.

Priscu,J.C., Fritsen,C.H., Adams,E.E., Giovannoni,S.J., Paerl,H.W., McKay,C.P., Doran,P.T., Gordon,D.A., Lanoil,B.D., and Pinckney,J.L. (1998) Perennial Antarctic lake ice: an oasis for life in a polar desert. *Science* **280**: 2095-2098.

Priscu,J.C., Adams,E.E., Lyons,W.B., Voytek,M.A., Mogk,D.W., Brown,R.L., McKay,C.P., Takacs,C.D., Welch,K.A., Wolf,C.F., Kirshtein,J.D., and Avci,R. (1999) Geomicrobiology of subglacial ice above Lake Vostok, Antarctica. *Science* **286**: 2141-2144.

Purdy,K.J., Nedwell,D.B., and Embley,T.M. (2003) Analysis of the sulfate-reducing bacterial and methanogenic archaeal populations in contrasting Antarctic sediments. *Appl.Environ.Microbiol.* **69**: 3181-3191.

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

Radstrom,P., Skold,O., Swedberg,G., Flensburg,J., Roy,P.H., and Sundstrom,L. (1994) Transposon Tn5090 of plasmid R751, which carries an integron, is related to Tn7, Mu, and the retroelements. *J Bacteriol.* **176**: 3257-3268.

Ranjard,L., Poly,F., and Nazaret,S. (2000) Monitoring complex bacterial communities using culture-independent techniques: application to soil environment. *Res.Microbiol.* **151**: 167-177.

Rappe,M.S., Giovannoni,S.J. (2003) The uncultured microbial majority. *Annu.Rev.Microbiol.* **57**: 369-394.

Rawlings,D.E., Tietze,E. (2001) Comparative biology of IncQ and IncQ-like plasmids. *Microbiol.Mol.Biol.Rev.* **65**: 481-96, table.

Recchia,G.D., Stokes,H.W., and Hall,R.M. (1994) Characterisation of specific and secondary recombination sites recognised by the integron DNA integrase. *Nucleic Acids Res.* **22**: 2071-2078.

Recchia,G.D., Hall,R.M. (1995) Gene cassettes: a new class of mobile element. *Microbiology* **141 (Pt 12)**: 3015-3027.

Reddy,G.S., Prakash,J.S., Matsumoto,G.I., Stackebrandt,E., and Shivaji,S. (2002) *Arthrobacter roseus* sp. nov., a psychrophilic bacterium isolated from an antarctic cyanobacterial mat sample. *Int.J Syst.Evol.Microbiol.* **52**: 1017-1021.

Reddy,G.S., Prakash,J.S., Srinivas,R., Matsumoto,G.I., and Shivaji,S. (2003) *Leifsonia rubra* sp. nov. and *Leifsonia aurea* sp. nov., psychrophiles from a pond in Antarctica. *Int.J Syst.Evol.Microbiol.* **53**: 977-984.

Renker,C., Blanke,V., Borstler,B., Heinrichs,J., and Buscot,F. (2004) Diversity of *Cryptococcus* and *Dioszegia* yeasts (Basidiomycota) inhabiting arbuscular mycorrhizal roots or spores. *FEMS Yeast Res.* **4**: 597-603.

Reysenbach,A.L., Wickham,G.S., and Pace,N.R. (1994) Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. *Appl.Environ.Microbiol.* **60**: 2113-2119.

Reysenbach,A.L., Ehringer,M., and Hershberger,K. (2000) Microbial diversity at 83 degrees C in Calcite Springs, Yellowstone National Park: another environment where the Aquificales and "Korarchaeota" coexist. *Extremophiles.* **4**: 61-67.

Rice,L.B. (2002) Association of different mobile elements to generate novel integrative elements. *Cell Mol.Life Sci.* **59**: 2023-2032.

de los Rios.A., Wierzchos,J., Sancho,L.G., and Ascaso,C. (2003) Acid microenvironments in microbial biofilms of antarctic endolithic microecosystems. *Environ.Microbiol.* **5**: 231-237.

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

Rosado,A.S., Duarte,G.F., Seldin,L., and van Elsas,J.D. (1998) Genetic diversity of nifH gene sequences in paenibacillus azotofixans strains and soil samples analyzed by denaturing gradient gel electrophoresis of PCR-amplified gene fragments. *Appl.Environ.Microbiol.* **64**: 2770-2779.

Rosser,S.J., Young,H.K. (1999) Identification and characterization of class 1 integrons in bacteria from an aquatic environment. *J Antimicrob Chemother* **44**: 11-18.

Roux,V., Raoult,D. (1995) Phylogenetic analysis of the genus Rickettsia by 16S rDNA sequencing. *Res.Microbiol.* **146**: 385-396.

Rowe-Magnus,D.A., Guerout,A.M., and Mazel,D. (1999) Super-integrons *Res.Microbiol.* **150**: 641-651.

Rowe-Magnus,D.A., Mazel,D. (1999) Resistance gene capture. *Curr.Opin.Microbiol.* **2**: 483-488.

Rowe-Magnus,D.A., Mazel,D. (2001) Integrons: natural tools for bacterial genome evolution. *Curr.Opin.Microbiol.* **4**: 565-569.

Rowe-Magnus,D.A., Guerout,A.M., Ploncard,P., Dychinco,B., Davies,J., and Mazel,D. (2001) The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. *Proc.Natl.Acad.Sci.U.S.A* **98**: 652-657.

Rowe-Magnus,D.A., Guerout,A.M., and Mazel,D. (2002) Bacterial resistance evolution by recruitment of super-integron gene cassettes. *Mol.Microbiol.* **43**: 1657-1669.

Rowe-Magnus,D.A., Guerout,A.M., Biskri,L., Bouige,P., and Mazel,D. (2003) Comparative analysis of superintegrons: engineering extensive genetic diversity in the *Vibrionaceae*. *Genome Res.* **13**: 428-442.

Rudi,K., Skulberg,O.M., Larsen,F., and Jakobsen,K.S. (1997) Strain characterization and classification of oxyphotobacteria in clone cultures on the basis of 16S rRNA sequences from the variable regions V6, V7, and V8. *Appl.Environ.Microbiol.* **63**: 2593-2599.

Rutz,B.A., Kieft,T.L. (2004) Phylogenetic Characterization of Dwarf Archaea and Bacteria from a Semiarid Soil. *Soil Biol.Biochem* **36**: 825-833.

Sabate,M., Navarro,F., Miro,E., Campoy,S., Mirelis,B., Barbe,J., and Prats,G. (2002) Novel complex sul1-type integron in Escherichia coli carrying bla(CTX-M-9). *Antimicrob Agents Chemother.* **46**: 2656-2661.

Sait,M., Hugenholtz,P., and Janssen,P.H. (2002) Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. *Environ.Microbiol.* **4**: 654-666.

Saito,T., Terato,H., and Yamamoto,O. (1994) Pigments of *Rubrobacter radiotolerans*. *Arch.Microbiol.* **162**: 414-421.

Saitou,N., Nei,M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol.Biol.Evol.* **4**: 406-425.

Sallen,B., Rajoharison,A., Desvarenne,S., and Mabilat,C. (1995) Molecular epidemiology of integron-associated antibiotic resistance genes in clinical isolates of enterobacteriaceae. *Microb.Drug Resist.* **1**: 195-202.

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

Sandaa,R.A., Enger,O., and Torsvik,V. (1999) Abundance and diversity of Archaea in heavy-metal-contaminated soils. *Appl.Environ.Microbiol.* **65**: 3293-3297.

Schabereiter-Gurtner,C., Pinar,G., Vybiral,D., Lubitz,W., and Rolleke,S. (2001) *Rubrobacter*-related bacteria associated with rosy discolouration of masonry and lime wall paintings. *Arch.Microbiol.* **176**: 347-354.

Schleper,C., Holben,W., and Klenk,H.P. (1997) Recovery of crenarchaeotal ribosomal DNA sequences from freshwater-lake sediments. *Appl.Environ.Microbiol.* **63**: 321-323.

Schmidt,F., Klopfer-Kaul,I. (1984) Evolutionary relationship between Tn21-like elements and pBP201, a plasmid from *Klebsiella pneumoniae* mediating resistance to gentamicin and eight other drugs. *Mol.Gen.Genet.* **197**: 109-119.

Shahmohammadi,H.R., Asgarani,E., Terato,H., Saito,T., Ohyama,Y., Gekko,K., Yamamoto,O., and Ide,H. (1998) Protective roles of bacterioruberin and intracellular KCl in the resistance of Halobacterium salinarium against DNA-damaging agents. *J Radiat.Res.(Tokyo)* **39**: 251-262.

Sheridan,P.P., Loveland-Curtze,J., Miteva,V.I., and Brenchley,J.E. (2003) Rhodoglobus vestalii gen. nov., sp. nov., a novel psychrophilic organism isolated from an Antarctic Dry Valley lake. *Int.J Syst.Evol.Microbiol.* **53**: 985-994.

Simon,H.M., Dodsworth,J.A., and Goodman,R.M. (2000) Crenarchaeota colonize terrestrial plant roots. *Environ.Microbiol.* **2**: 495-505.

Siu,L.K., Lo,J.Y., Yuen,K.Y., Chau,P.Y., Ng,M.H., and Ho,P.L. (2000) beta-lactamases in *Shigella flexneri* isolates from Hong Kong and Shanghai and a novel OXA-1-like beta-lactamase, OXA-30. *Antimicrob Agents Chemother* **44**: 2034-2038.

Sjoling,S., Cowan,D.A. (2003) High 16S rDNA bacterial diversity in glacial meltwater lake sediment, Bratina Island, Antarctica. *Extremophiles.* **7**: 275-282.

Sliwinski,M.K., Goodman,R.M. (2004) Spatial heterogeneity of Crenarchaeal assemblages within mesophilic soil ecosystems as revealed by PCR-single-stranded conformation polymorphism profiling. *Appl.Environ.Microbiol.* **70**: 1811-1820.

Smalla,K., Krogerrecklenfort,E., Heuer,H., Dejonghe,W., Top,E., Osborn,M., Niewint,J., Tebbe,C., Barr,M., Bailey,M., Greated,A., Thomas,C., Turner,S., Young,P., Nikolakopoulou,D., Karagouni,A., Wolters,A., van Elsas,J.D., Dronen,K., Sandaa,R., Borin,S., Brabhu,J., Grohmann,E., and Sobecky,P. (2000) PCR-based detection of mobile genetic elements in total community DNA. *Microbiology* **146 (Pt 6)**: 1256-1257.

Sorum,H., L'Abee-Lund,T.M., Solberg,A., and Wold,A. (2003) Integron-containing IncU R plasmids pRAS1 and pAr-32 from the fish pathogen *Aeromonas salmonicida. Antimicrob Agents Chemother* **47**: 1285-1290.

Stackebrandt, E., and B. M. Goebel. (1994) Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. **44**: 846-849.

Stanisich,V.A., Ortiz,J.M. (1976) Similarities between plasmids of the P-incompatibility group derived from different bacterial genera. *J Gen.Microbiol.* **94**: 281-289.

Stein,L.Y., La Duc,M.T., Grundl,T.J., and Nealson,K.H. (2001) Bacterial and archaeal populations associated with freshwater ferromanganous micronodules and sediments. *Environ.Microbiol.* **3**: 10-18.

Stergiopoulos,I., Gielkens,M.M., Goodall,S.D., Venema,K., and De Waard,M.A. (2002) Molecular cloning and characterisation of three new ATP-binding cassette transporter genes from the wheat pathogen *Mycosphaerella graminicola. Gene* **289**: 141-149.

Stevens,M.I., Hogg,I.D. (2003) Long-term isolation and recent range expansion from glacial refugia revealed for the endemic springtail *Gomphiocephalus hodgsoni* from Victoria Land, Antarctica. *Mol.Ecol.* **12**: 2357-2369.

Stoeck,T., Taylor,G.T., and Epstein,S.S. (2003) Novel eukaryotes from the permanently anoxic Cariaco Basin (Caribbean Sea). *Appl.Environ.Microbiol* **69**: 5656-5663.

Stoeck,T., Epstein,S. (2003) Novel eukaryotic lineages inferred from small-subunit rRNA analyses of oxygen-depleted marine environments. *Appl.Environ.Microbiol* **69**: 2657-2663.

Stokes,H.W., Hall,R.M. (1989) A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol.Microbiol.* **3**: 1669-1683.

Stokes,H.W., O'Gorman,D.B., Recchia,G.D., Parsekhian,M., and Hall,R.M. (1997) Structure and function of 59-base element recombination sites associated with mobile gene cassettes. *Mol.Microbiol.* **26**: 731-745.

Stokes,H.W., Holmes,A.J., Nield,B.S., Holley,M.P., Nevalainen,K.M., Mabbutt,B.C., and Gillings,M.R. (2001) Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Appl.Environ.Microbiol.* **67**: 5240-5246.

Stokes,H.W., Hall,R.M. (2004) A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol.Microbiol.* **3**: 1669-1683.

Sun,B., Cole,J.R., and Tiedje,J.M. (2001) *Desulfomonile limimaris* sp. nov., an anaerobic dehalogenating bacterium from marine sediments. *Int.J.Syst.Evol.Microbiol.* **51**: 365-371.

Suzuki,K., Sasaki,J., Uramoto,M., Nakase,T., and Komagata,K. (1997) *Cryobacterium psychrophilum* gen. nov., sp. nov., nom. rev., comb. nov., an obligately psychrophilic actinomycete to accommodate "*Curtobacterium psychrophilum*" Inoue and Komagata 1976. *Int.J Syst.Bacteriol.* **47**: 474-478.

Suzuki,M.T., Taylor,L.T., and Delong,E.F. (2000) Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. *Appl.Environ.Microbiol.* **66**: 4605-4614.

Tait,R.C., Lundquist,R.C., and Kado,C.I. (1982) Genetic map of the crown gall suppressive IncW plasmid pSa. *Mol.Gen.Genet.* **186**: 10-15.

Takai,K., Moser,D.P., DeFlaun,M., Onstott,T.C., and Fredrickson,J.K. (2001) Archaeal diversity in waters from deep South African gold mines. *Appl.Environ.Microbiol.* **67**: 5750-5760.

Taton,A., Grubisic,S., Brambilla,E., De Wit,R., and Wilmotte,A. (2003) Cyanobacterial diversity in natural and artificial microbial mats of Lake Fryxell (McMurdo Dry Valleys, Antarctica): a morphological and molecular approach. *Appl.Environ.Microbiol.* **69**: 5157-5169.

Tennstedt,T., Szczepanowski,R., Braun,S., Puhler,A., and Schluter,A. (2003) Occurrence of integron-associated resistance gene cassettes located on antibiotic resistance plasmids isolated from a wastewater treatment plant. *FEMS Microbiol.Ecol.* **45**: 239-252.

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

Tiedje,J.M., Asuming-Brempong,S., Nusslein,K., Marsh,T.L., and Flynn,S.J. (1999) Opening the black box of soil microbial diversity. *Appl.Soil Ecol.* **13**: 109-122.

de la Torre., Jr., Goebel,B.M., Friedmann,E.I., and Pace,N.R. (2003) Microbial diversity of cryptoendolithic communities from the McMurdo Dry Valleys, Antarctica. *Appl.Environ.Microbiol.* **69**: 3858-3867.

Torsvik,V., Goksoyr,J., and Daae,F.L. (1990) High diversity in DNA of soil bacteria. *Appl.Environ.Microbiol.* **56**: 782-787.

Torsvik,V., Daae,F.L., Sandaa,R.A., and Ovreas,L. (1998) Novel techniques for analysing microbial diversity in natural and perturbed environments. *J Biotechnol.* **64**: 53-62.

Torsvik,V., Ovreas,L. (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr.Opin.Microbiol.* **5**: 240-245.

Turkiewicz,M., Galas,E., and Kalinowska,H. (1982) Microflora of Antarctic Krill (*Euphausia superba*). *Acta Microbiol.Pol.* **31**: 175-184.

Vaisvila,R., Morgan,R.D., Posfai,J., and Raleigh,E.A. (2001) Discovery and distribution of super-integrons among pseudomonads. *Mol.Microbiol.* **42**: 587-601.

Valentine,C.R., Heinrich,M.J., Chissoe,S.L., and Roe,B.A. (1994) DNA sequence of direct repeats of the sulI gene of plasmid pSa. *Plasmid* **32**: 222-227.

Valentine,D.L. (2002) Biogeochemistry and microbial ecology of methane oxidation in anoxic environments: a review. *Antonie Van Leeuwenhoek* **81**: 271-282.

Valinsky,L., Della,V.G., Scupham,A.J., Alvey,S., Figueroa,A., Yin,B., Hartin,R.J., Chrobak,M., Crowley,D.E., Jiang,T., and Borneman,J. (2002) Analysis of bacterial community composition by oligonucleotide fingerprinting of rRNA genes. *Appl.Environ.Microbiol.* **68**: 3243-3250.

Van de,P.Y., De Wachter,R. (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput.Appl.Biosci.* **10**: 569-570.

Van de,P.Y., Caers,A., De Rijk,P., and De Wachter,R. (1998) Database on the structure of small ribosomal subunit RNA. *Nucleic Acids Res.* **26**: 179-182.

van Elsas,J.D., Bailey,M.J. (2002) The ecology of transfer of mobile genetic elements. *FEMS Microbiol Ecol.* **42**: 187-197.

Van Trappen,S., Mergaert,J., Van Eygen,S., Dawyndt,P., Cnockaert,M.C., and Swings,J. (2002) Diversity of 746 heterotrophic bacteria isolated from microbial mats from ten Antarctic lakes. *Syst.Appl.Microbiol.* **25**: 603-610.

Venkateswaran,K., Satomi,M., Chung,S., Kern,R., Koukol,R., Basic,C., and White,D. (2001) Molecular microbial diversity of a spacecraft assembly facility. *Syst.Appl.Microbiol.* **24**: 311-320.

Venter,J.C., Remington,K., Heidelberg,J.F., Halpern,A.L., Rusch,D., Eisen,J.A., Wu,D., Paulsen,I., Nelson,K.E., Nelson,W., Fouts,D.E., Levy,S., Knap,A.H., Lomas,M.W., Nealson,K., White,O., Peterson,J., Hoffman,J., Parsons,R., Baden-Tillson,H., Pfannkoch,C., Rogers,Y.H., and Smith,H.O. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66-74.

Vetriani,C., Reysenbach,A.L., and Dore,J. (1998) Recovery and phylogenetic analysis of archaeal rRNA sequences from continental shelf sediments. *FEMS Microbiol.Lett.* **161**: 83-88.

Vishniac,H.S. (1993) *The Microbiology of Antarctic Soils, in Antarctic Microbiology.* Friedman,E.I. (ed). New York: Wiley-Liss, Inc., pp. 297-341.

von Wintzingerode,F., Gobel,U.B., and Stackebrandt,E. (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol.Rev.* **21**: 213-229.

Wagner,R. (1994) The regulation of ribosomal RNA synthesis and bacterial cell growth. *Arch.Microbiol.* **161**: 100-109.

Walker,D.H., Gage,K.L. (1997) *Rickettsia, orienta, ehrlicha* and *coxiella.* In *Medical Microbiology.* Greenwood,D., Slack,R., and Peutherer,J. (eds). New York: Churchill Livingstone, pp. 371-380.

Ward-Rainey,N., Rainey,F.A., Schlesner,H., and Stackebrandt,E. (1995) Assignment of hitherto unidentified 16S rDNA species to a main line of descent within the Domain *Bacteria. Microbiology* **141**: 3247-3250.

Ward,J.M., Grinsted,J. (1982) Physical and genetic analysis of the Inc-W group plasmids R388, Sa, and R7K. *Plasmid* **7**: 239-250.

Warnecke,F., Amann,R., and Pernthaler,J. (2004) Actinobacterial 16S rRNA genes from freshwater habitats cluster in four distinct lineages. *Environ.Microbiol.* **6**: 242-253.

Weber,S., Stubner,S., and Conrad,R. (2001) Bacterial populations colonizing and degrading rice straw in anoxic paddy soil. *Appl.Environ.Microbiol.* **67**: 1318-1327.

Webster,N.S., Negri,A.P., Munro,M.M., and Battershill,C.N. (2004) Diverse microbial communities inhabit Antarctic sponges. *Environ.Microbiol.* **6**: 288-300.

Wiedemann B, Meyer JF, Zuhlsdorf MT. Insertions of resistance genes into Tn21-like transposons. J Antimicrob Chemother. 1986 Oct;18 Suppl C:85-92.

Weisburg,W.G., Barns,S.M., Pelletier,D.A., and Lane,D.J. (1991) 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol.* **173**: 697-703.

Wellington,E.M., Berry,A., and Krsek,M. (2003) Resolving functional diversity in relation to microbial community structure in soil: exploiting genomics and stable isotope probing. *Curr.Opin.Microbiol.* **6**: 295-301.

Wery,N., Gerike,U., Sharman,A., Chaudhuri,J.B., Hough,D.W., and Danson,M.J. (2003) Use of a packed-column bioreactor for isolation of diverse protease-producing bacteria from antarctic soil. *Appl.Environ.Microbiol.* **69**: 1457-1464.

Whelan,S., Lio,P., and Goldman,N. (2001) Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends Genet.* **17**: 262-272.

White,C., Sayer,J.A., and Gadd,G.M. (1997) Microbial solubilization and immobilization of toxic metals: key biogeochemical processes for treatment of contamination. *FEMS Microbiol.Rev.* **20**: 503-516.

White,D.G., Zhao,S., McDermott,P.F., Ayers,S., Gaines,S., Friedman,S., Wagner,D.D., Meng,J., Needle,D., Davis,M., and DebRoy,C. (2002) Characterization of antimicrobial resistance among *Escherichia coli* O111 isolates of animal and human origin. *Microb.Drug Resist.* **8**: 139-146.

White,P.A., Rawlinson,W.D. (2001) Current status of the aadA and dfr gene cassette families. *J Antimicrob Chemother.* **47**: 495-496.

Wiedemann,B., Meyer,J.F., and Zuhlsdorf,M.T. (1986) Insertions of resistance genes into Tn21-like transposons. *J Antimicrob Chemother*: 85-92.

Williams,P.A., Worsey,M.J. (1976) Ubiquity of plasmids in coding for toluene and xylene metabolism in soil bacteria: evidence for the existence of new TOL plasmids. *J Bacteriol.* **125**: 818-828.

Woese,C.R., Fox,G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc.Natl.Acad.Sci.U.S.A* **74**: 5088-5090.

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

Woese,C.R., Kandler,O., and Wheelis,M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc.Natl.Acad.Sci.U.S.A* **87**: 4576-4579.

Wohlleben,W., Arnold,W., Bissonnette,L., Pelletier,A., Tanguay,A., Roy,P.H., Gamboa,G.C., Barry,G.F., Aubert,E., Davies,J., and . (1989) On the evolution of Tn21-like multiresistance transposons: sequence analysis of the gene (aacC1) for gentamicin acetyltransferase-3-I(AAC(3)-I), another member of the Tn21-based expression cassette *Mol.Gen.Genet.* **217**: 202-208.

Worland,M.R., Block,W. (2003) Desiccation stress at sub-zero temperatures in polar terrestrial arthropods. *J Insect Physiol* **49**: 193-203.

Wright,G.D. (1999) Aminoglycoside-modifying enzymes. *Curr.Opin.Microbiol.* **2**: 499-503.

Zengler,K., Toledo,G., Rappe,M., Elkins,J., Mathur,E.J., Short,J.M., and Keller,M. (2002) Cultivating the uncultured. *Proc.Natl.Acad.Sci.U.S.A* **99**: 15681-15686.

Zhao,S., White,D.G., Ge,B., Ayers,S., Friedman,S., English,L., Wagner,D., Gaines,S., and Meng,J. (2001) Identification and characterization of integron-mediated antibiotic resistance among Shiga toxin-producing *Escherichia coli* isolates. *Appl.Environ.Microbiol.* **67**: 1558-1564.

Zhou,J., Bruns,M.A., and Tiedje,J.M. (1996) DNA recovery from soils of diverse composition. *Appl.Environ.Microbiol.* **62**: 316-322.

Zinniel,D.K., Lambrecht,P., Harris,N.B., Feng,Z., Kuczmarski,D., Higley,P., Ishimaru,C.A., Arunakumari,A., Barletta,R.G., and Vidaver,A.K. (2002) Isolation and characterization of endophytic colonizing bacteria from agronomic crops and prairie plants. *Appl.Environ.Microbiol.* **68**: 2198-2208.

Zlamala,C., Schumann,P., Kampfer,P., Valens,M., Rossello-Mora,R., Lubitz,W., and Busse,H.J. (2002) *Microbacterium aerolatum* sp. nov., isolated from the air in the 'Virgilkapelle' in Vienna. *Int.J Syst.Evol.Microbiol.* **52**: 1229-1234.

Zwart,G., Hiorns,W.D., Methe,B.A., van Agterveld,M.P., Huismans,R., Nold,S.C., Zehr,J.P., and Laanbroek,H.J. (1998) Nearly identical 16S rRNA sequences recovered from lakes in North America and Europe indicate the existence of clades of globally distributed freshwater bacteria. *Syst.Appl.Microbiol.* **21**: 546-556.

Appendix

References to unpublished work. GenBank accession numbers are shown in parentheses.

Alfreider,A. and Vogt,C. (AY214204) Analysis of bacterial populations in benzene-contaminated groundwater.

Asami,H. and Watanabe,K. (AB116399) Phylogenetic diversity of prokaryotes in Sanriku marine sediments.

Bonheyo,G.T., Fouke,B.W., Frias-Lopez,J. and Sanzenbacher,B. (AF445701) Microbial 16S rRNA diversity within travertine depositional facies from 73 degrees to 25 degreec C at Angel Terrace, Mammoth Hot Springs, Yellowstone National Park, USA.

Botero,L.M., Burr,M.D., Willits,D., Elkins,J.G., Inskeep,W.P. and McDermott,T.R. (AF391984) Prokaryote diversity in an extreme thermal soil.

Brambilla,E., Tindall,B., Hippe,H. and Stackebrandt,E. (AJ287660) Cellular and molecular diversity of mat sample from lake Fryxell, Antarctica.

Brinkmeyer,R. and Helmke,E. (AF468240) Evidence for methylotrophic processes in Arctic pack ice.

Chen,X., Wang,J., Nan,Z., Ren,J. and Yang,C. (AY176765)

Donachie,S.P., Hou,S., Lee,K.S., Riley,C.W., Pikina,A., Liu,J., Kempe,S., Gregory,T.S., Bossuyt,A., Boerema,J., Malahoff,A. and Alam,M. (AY345496) Microbial Communities in the Hawaiian Archipelago: A Microbial Diversity Hotspot.

Fowler,R., Roberson,E., Groves,C. and Sahi,S. (AY221067) Survey of bacteria from Mammoth Cave, KY by molecular analysis of 16S rRNA genes.

Geissler,A., Tzvetkova,T., Flemming,K. and Selenska-Pobell,S. (AJ518765) Comparison of natural bacterial communities found in uranium mining waste piles and mill tailings.

Gordon,D.A., Priscu,J.C. and Giovannoni,S.J. (AF173822) Origin and phylogeny of Microbes Living in Permanent Antarctic Lake Ice.

Groudieva,T. and Antranikian,G. (AF513431) Diversity of bacteria isolated from Arctic sea ice and seawater and their cold-active hydrolytic enzymes.

Guo,J.-H., Guo,Y.-H., Zhang,X.-M. and Ge,Y.-Y. (AY273209) *Clavibacter fangii* sp. nov., a new speices of pathogen in wheat.

Humayoun,S.B., Bano,N., Lecleir,G. and Hollibaugh,J.T. (AF454303) Composition of bacterial assemblages from alkaline, hypersaline Mono Lake, California.

Kim,J.-S. and Crowley,D.E. (a) (AY493980) Fingerprinting bacterial communities in soil aggregates.

Kim,J.-S. and Crowley,D.E. (b) (p114: AY326617; p117: AY326522) Bacterial diversity of Amazon soil by oligonucleotide fingerprinting of rRNA genes.

LaPara,T.M. (AF538744) Cultivation dependent and cultivation independent community analysis of a municipal wastewater treatment bioreactor.

Liu,J.R., McKenzie,C., Seviour,E.M., Webb,R., Blackall,L.L., Saint,C. and Seviour,R.J. (AF244750) The phylogeny of the bulking filamentous bacterium '*Nostocoida limicola*' III from activated sludge.

Lukow,T. (p113: AJ252642; p117: AJ252615; p118: AJ252623) Vergleichende Charakterisierung der bakteriellen Rhizosphaerengemeinschaften transgener versus nicht-transgener Kartoffelpflanzen.

Reddy,G.S.N., Ravi Prasad,A., Razia,Kutty., Kathrin,R. and Shivaji,S. (a) (AJ576092-AJ576100) Bacterial diversity of a soil sample from Schirmacher Oasis, Antarctica.

Reddy,G.S.N., Schumann,P., Stackebrandt,E., Matsumoto,G.I. and Shivaji,S. (b) (AJ584833) Identification of *Psychrobacter psychrophila* sp. nov., *Psychrobacter vallis* sp. nov. and *Psychrobacter aquatica* sp. nov. from McMurdo dry valley region, Antarctica.

Shravage *et al.* Unpublished data. p209 Refer to de la Torre *et al.* 2003.

Smith,M.C., Bowman;J.P. and Line,M.A. (AF170755) Antarctic quartz stone sublithic communities.

Tzvetkova,I. and Selenska-Pobell,S. (AJ291841) Bacterial diversity in water samples of Monticello mill tailings site.

Yoshida,H., Yamamoto,K., Murakami,Y., Hoshii,D., Nishikawa,Y., Miyoshi,T., Naganuma,T., Milodowski,A.E. and Metcalfe,R. (AB179537) Biogenic redox front formation: microbial consortium involved in reduction and oxidation of iron in siliceous sedimentary rock.

Zhang,R., Zeng,R., Zhao,J. and Lin,N. (p117: AY218772; p118: AY218649; p113 & p227: AY218566) 16S rDNA Diversity of Bacteria in Penguin Droppings Sediments from Ardley Island, Antarctica.