Detection of Microbial Taxa in Complex Communities: impacts of relative abundance, gene transfer and persistence of target DNA

David William Cleary BSc

A Thesis submitted for the Degree of Doctor of Philosophy

University of Warwick, the School of Life Sciences

April 2012

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisors, Professor Elizabeth M. H. Wellington, Dr William H. Gaze and Dr Alistair H. Bishop, for their help, support and enthusiasm.

Huge thanks must go to the members of the (extended) Molecular Biology team (in all its previous guises) for their friendship, support and understanding. In particular Dr Carol Stone, Dr Martin Pearce and Dr Lois Blackman deserve a special mention for starting me on this journey. For assistance in the laboratory and at the computational biology 'coal-face', particularly in times of great need, I would also like to thank Bry Lingard, Amber Murch, Dr Phil Rachwal and Morgan Ely. For bioinformatic support and much 'hand-holding' my gratitude is extended to collaborators at the University of Exeter, Dr Konrad Paszkiewicz and Dr David Studholme. For assistance with experiments involving soil DNA extractions and high-containment pathogens I am grateful to Caroline Dewey and Dr Helen L. Rose respectively.

Lastly, and most importantly, I thank my two boys (and as a fellow PhD student, for proofreading and technical discussion over the dinner table) my wife for all their love, patience and understanding during a period of stress and distraction.

Declaration

I hereby declare that I, David Cleary, conducted the work described in this thesis under the supervision of Professor Elizabeth M. H. Wellington, University of Warwick, Dr William H. Gaze, Peninsula College of Medicine and Dentistry, Universities of Exeter and Plymouth and Dr Alistair H. Bishop, DSTL Porton Down; with the exception of those instances where the contribution of others has been specifically acknowledged. None of the information contained herein has been used in any previous application for a degree. All sources of information have been specifically acknowledged by means of reference.

David W. Cleary

Summary

The aim of this study was to understand the ecological interactions of *B. pseudomallei*, a highly pathogenic potential biological warfare agent, in both soils and bio-aerosols. This study examined the impact of DNA persistence on its detection in environmental samples. The potential for members of the Burkholderiaceae to use persistent extracellular DNA in the process of natural transformation was also determined. Lastly, a comparative analysis of methods using 16S rRNA analysis and metagenomics was also undertaken to evaluate their utility for identification of pathogens in complex samples.

Soil microcosms were used to determine the impacts of soil mineralogy on DNA persistence using real-time PCR. This study showed that the presence of clay minerals with contrasting adsorption affinities for DNA resulted in significantly different rates of DNA decay in soil.

The capacity for *Burkholderia* sp. to undergo natural transformation was determined through the use of both chromosomal allelic rescue and non-homologous plasmid uptake strategies. Transformation was suggested as a result of observations of non-homologous plasmid uptake in *B. multivorans* 13010. However further work is required in this area.

The performance of 16S rRNA sequencing using assays targeting the V1-3 and V4-6 variable regions was assessed using soil samples taken from one location that had undergone contrasting application strategies of veterinary antibiotics. This approach was shown to be able to discern subtle shifts in the relative abundances of certain microbial taxa in response to antibiotic application. Outcomes here support previous observations regarding the increased prevalence of mobile genetic elements (integrons) that harbour antibiotic resistance determinants.

For bio-aerosols, samples were acquired from regions of known endemicity of *B. pseudomallei* (N. Australia) and both 16S rRNA and metagenomic analysis methods were used to determine the fluctuations in microbial diversity and the prevalence of this pathogenic organism. The fluctuations of bio-aerosol microbial diversity and in particular *B. pseudomallei*, within wet and dry seasons of Northern Australia were identifiable using a combination of 16S rRNA and total community DNA metagenomic analyses. The importance of robust bioinformatic analysis is highlighted.

List of Abbreviations

ATCC	American type culture collection
bp	base pairs
Bcc	Burkholderia cepacia complex
BSA	bovine serum albumen
BWA	biological warfare agent
°C	degrees centigrade
CF	cystic fibrosis
cfu	colony forming units
ds	double stranded
dH_2O	distilled water
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	equimolar mix of deoxyribonucleoside triphosphates dATP, dCTP, dGTP and dTTP
EDTA	ethylenediaminetetra-acetic acid
gDNA	genomic DNA
GMO	genetically modified organism

h	hour
HGT	horizontal gene transfer
HWWC	horizontal wet-walled cyclone
IPTG	isopropyl β-D-1-thiogalactopyranoside
IS	insertion sequences
Kbp	kilobase pair
Mbp	megabase pair
MID	multiplex identifier
MDA	multiple displacement amplification
MG-RAST	metagenomic-rapid annotation by subsystem technology
MIC	minimum inhibitory concentration
min	minutes
mRNA	messenger ribonucleic acid
NCTC	national collection of type cultures
NGT	natural genetic transformation
OTU	operational taxonomic unit
PBS	phosphate buffered saline
PCA	principal component analysis
PCIAA	phenol:chloroform:isoamyl alcohol

PCR	polymerase chain reaction
QIIME	quantitative insights into microbial ecology
RFLP	restriction fragment length polymorphism
RDP	ribosomal database project
RPM	revolutions per minute
rRNA	ribosomal ribonucleic acid
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulphate
sec	seconds
SOC	super optimal media with catabolite repression
ss	single stranded
TAE	tris acetic acid EDTA
tcDNA	total community DNA
Tris	tris-(hydroxymethyl)-methylamine
UPGMA	unweighted-pair group method using average
UV	ultra-violet
v/v	volume per volume
vol	volume

w/v	weight per volume
WGA	whole genome amplification

Chapter 1

Introduction

1.1 **Introduction**

Gene transfer through natural genetic transformation (NGT) among bacterial communities in terrestrial environments has garnered much interest. This is principally related to the potential dissemination and uptake of recombinant plant DNA into the surrounding microflora (Nielsen *et al*., 1998; van den Eede *et al*., 2004). Similar considerations concerning the release of pathogenic bacteria into the environment have received less attention.

Bioterrorism or biological warfare is the intentional use of micro-organisms or toxins to cause death or disease in humans, animals or plants (Carter, 1992; Ashford, 2003). In this context there is the potential for such releases of pathogenic organisms to contribute to the horizontal gene pool in the environment and thus potentially result in the transfer of virulence determinants to environmental microbiota. Conversely, the acquisition of genetic material by the pathogen may be a risk that could result an increase in pathogenicity e.g. antibiotic resistance genes. There is considerable observed variability in viable persistence of bacterial biological warfare agents (BWAs). Three potential BWAs, as defined by the Centre for Disease Control (CDC), and termed 'Select Agents', are *Bacillus anthracis*, *Yersinia pestis* and *Burkholderia pseudomallei* (Rotz *et al.*, 2002). *B. anthracis*, a spore forming Gram positive, is well known for its potential to survive in soil for many years which contrasts with *Y. pestis* where persistence is measured in days or weeks (Manchee *et al.*, 1981; Eisen *et al.*, 2008). The persistence of *B. pseudomallei* in comparison is very much dependent upon soil type, osmolarity and pH. It has been demonstrated, for example, that waterlogged, heavy clay soils support persistence much better than

dry, sandy soils (Inglis *et al.*, 2006). When using techniques that focus on the detection of nucleic acid signatures rather than culturable organisms, it should be remembered that the presence of non-viable organisms, or simply DNA from dead cells, will also be included in any analysis. Therefore, given the unlikely prolonged survival of a released pathogen the questions of persistence and availability of genetic material within the environment are important factors of pathogen identification and concurrently, gene exchange. In this context the following events should be considered: 1 – the persistence of genetic material from the pathogen in the environment, and 2 – the horizontal transfer of genes to endogenous closely-related species or, where conditions favour persistence of viable organisms such as for *B. pseudomallei*, acquisition by the released pathogen of genetic elements that may increase pathogenicity. The *Burkholderia* genus therefore contains candidate bacterial pathogens for study in these contexts. As well as the two species of interest for their potential as BWAs: *B. mallei* and *B. pseudomallei*, the genus also contains opportunistic pathogens belonging to the *Burkholderia cepacia* complex (Bcc).

1.2 **The Burkholderiaceae**

This highly diverse genus, first described by William Burkholder in 1950 but not classified as Burkholderia until 1992 (Yabuuchi *et al.*, 1992), contains over 30 distinct and phenotypically versatile species (Mahenthiralingam *et al.*, 2008). It includes the pathogenic *B. pseudomallei* and *B. mallei* lineages and the opportunistic pathogens of the *B. cepacia* complex. The majority are soil saprophytes or plant pathogens with the notable exception of *B. mallei* which is an obligate pathogen of equines (Cheng and Currie, 2005; Ramette *et al.*, 2005; Waag and DeShazer, 2004; Vandamme *et al.*, 2007).

On average, members of this genus have genome sizes of 7.5 Mbp and possess two large chromosomal replicons. The presence of rRNA operons and housekeeping genes on the smaller chromosome suggests this is a stable genomic architecture (Mahenthiralingam and Drevinek, 2007). Evolution through the lateral acquisition of foreign DNA also appears to have been a prominent feature of this genus with ~10% of the genomes examined to date comprising genomic islands (GI) with a large number of plasmid-like genes, insertion sequence (IS) elements and bacteriophages (Mahenthiralingam and Drevinek, 2007). However, to date only *B. thailandensis* and some strains of *B. pseudomallei* have been shown to undergo transformation (Thongdee *et al.*, 2008). The number and diversity of IS elements, which are capable of gene activation, are also a feature of this genus with some, such as IS*1356*, *1363* and an IS*402*/*1356* hybrid linked to clinically important strains e.g. the epidemic ET12 and PHDC lineages of the *Burkholderia cepacia* complex (Bcc) (Tyler *et al.*, 1996; Liu *et al.*, 2003). Given the reduced number of IS elements in *B. pseudomallei* and *B. thailandensis* in comparison to *B. mallei*, it is reasoned that IS elements have played a reduced role in their evolution (Kim *et al.*, 2005). The evolution of *B. mallei*, in contrast, has involved deletion and rearrangements, a consequence of the higher number of IS elements, which has resulted in its restricted host habitat and functioning as an obligate pathogen (Losada *et al.*, 2010).

1.2.1 ***Burkholderia pseudomallei, B. mallei*** **and** ***B. thailandensis***

B. pseudomallei is the causative agent of melioidosis, a tropical disease of both humans and animals that is endemic in Northern Australia and South East Asia with sporadic occurrences in many other countries as shown in Figure 1.1

(Cheng and Currie, 2005). It should be noted that melioidosis is not evenly distributed within these endemic areas and occurs in hot-spots (Dance, 2000). It is a facultative intracellular bacterium that usually infects individuals with underlying risk factors such as diabetes mellitus, alcoholism and renal complications (Ulrich *et al.*, 2006). Inhalational infection presents as a more severe disease with pneumonia (Cheng and Currie, 2005). This occurs during monsoon seasons and is thought to arise as a result of severe weather aerosolising surface dwelling organisms. Infection commonly starts either in the lungs for inhalational exposure or as focal infections at the sites of skin abrasions, whereupon it can spread to any organ of the infected host (Chaowagul, 2000). Clinically, disease can present asymptomatically, as focal infections or pneumonia. The later presentation can be chronic or proceed rapidly to fatal septicaemia within 48 hrs from the onset of symptoms (Currie, 2000; White, 2003). Virulence factors of *B. pseudomallei* include products secreted by the general secretory pathway, Type III secretion systems, flagella, lipopolysaccharide, capsule, quorum sensing and Type IV pili (DeShazer *et al.*, 1997, 1998, 1999; Atkins *et al.*, 2002; Rainbow *et al.*, 2002; Stevens *et al.*, 2002; Chua *et al.*, 2003; Ulrich *et al.*, 2003; Essex-Lopresti *et al.*, 2005; Reckseidler-Zenteno *et al.*, 2005). As will be discussed later, in the context of gene transfer through transformation, Type IV pili have been shown to be important components of DNA uptake mechanisms in Gram negative bacteria.

Gene mutations, deletions and acquisitions appear to have been the major drivers of *B. pseudomallei* evolution with less emphasis on rearrangements and bacteriophage-mediated recombination (Yu *et al.*, 2006). In particular, the

genome of *B. pseudomallei* K96243 appears to have evolved through horizontal capture of GIs resulting in the acquisition of some of the virulence-related genes that are absent in *B. thailandensis* (Kim *et al.*, 2005). These include the replacement of a distinct set of capsular polysaccharide synthesis genes within the polysaccharide cluster and additionally the acquisition of a *Yersinia*-like fimbrial (YLF) cluster with consequent replacement of a flagellum biosynthesis cluster (Kim *et al.*, 2005; Yu *et al*., 2006). Through comparisons of whole genome sequences, Tuanyok *et al*., (2007) were able to demonstrate that the YLF is not a universal feature of *B. pseudomallei*. It appears to be mutually exclusive to the *B. thailandensis*-like flagella and chemotaxis gene clusters (BTFC) and consequently can be used to separate *B. pseudomallei* strains into two groups. These groups show contrasting geographical distribution: BTFC strains are common in Australia but rare elsewhere and YLF strains predominate in Thailand. YLF strains are also more common as clinical isolates although BTFC strains can cause disease (Tuanyok *et al.*, 2007).

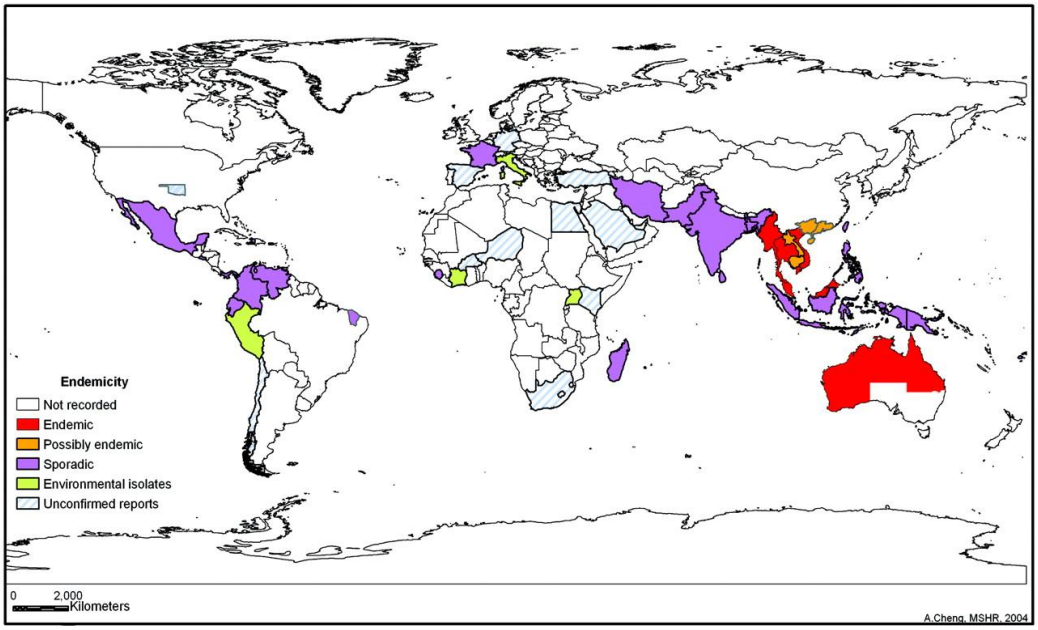

Figure 1.1 Worldwide Distribution of Melioidosis (Taken from Cheng and Currie, 2005). Clear endemic regions can be seen in SE Asia and Australia.

Until its reclassification in 1998, *B. thailandensis* was considered a subtype of *B. pseudomallei* (Brett *et al.*, 1998). It shares a number of phenotypic characteristics with *B. pseudomallei* including habitat and genomic structure but is not pathogenic to immunocompetent individuals (Yu *et al.*, 2006) and can be distinguished by its ability to assimilate arabinose (Smith *et al.*, 1997). An interesting feature of *B. thailandensis* is its absence from Northern Australia where *B. pseudomallei* is endemic (Cheng and Currie, 2005). Through examinations of 16S rRNA gene sequences from each of *B. pseudomallei*, *B. mallei* and *B. thailandensis*, Yu *et al.*, (2006) demonstrated a close genetic relationship between *B. pseudomallei* and its derivative *B. mallei*. *B. thailandensis* was found to occupy a separate branch of the phylogenetic tree with a divergence time of 47 million years before present (Yu *et al.*, 2006). Like *B. pseudomallei*, numerous genomic islands exhibiting either atypical G+C content, stretches of bacteriophage-related genes and phage-like integrases have

been identified (Yu *et al.*, 2006). Whilst these differ in content to the genomic islands found in *B. pseudomallei*, and variation is seen between *B. thailandensis* strains, the locations appear to be conserved, suggesting that these might represent integration hot-spots for laterally transferred sequences (Yu *et al.*, 2006).

In contrast to both *B. pseudomallei* and *B. thailandensis*, *B. mallei* is an obligate pathogen of equines. It is responsible for the disease 'glanders', which is occasionally transmitted to humans causing either acute or chronic infections depending on the route of exposure (Waag and DeShazer, 2004). It was first described by Hippocrates in 425 BC and given the name '*melis'* by Aristotle in 350 BC. Disease presentation in humans is indistinguishable from melioidosis and without correct treatment is invariably fatal (Ulrich *et al.*, 2006). The similarities in genomic architecture, supportive of the hypothesis that *B. mallei* evolved from an ancestral strain of *B. pseudomallei*, was confirmed when strains ATCC 23344 and K96243 were sequenced in 2004. In both species the second, smaller chromosome harboured a lower proportion of genes encoding core cell function combined with more accessory functions (Holden *et al*., 2004; Nierman *et al.*, 2004). Virulence factors of *B. mallei* include the capsular polysaccharide, Type III secretion systems, quorum sensing network, a two-component transcriptional regulatory system (VirAG) and Type IV secretion systems (DeSchazer *et al.*, 2001; Ulrich *et al.*, 2004; Nierman *et al.*, 2004; Ribot and Ulrich, 2006; Schell *et al.*, 2007). VirAG in particular appears to be part of a regulatory cascade for controlling ~60 genes of known, likely or putative virulence determination as demonstrated through transcriptome comparison of *B.*

mallei with *virAG* expression mutants (Schell *et al.*, 2007). In terms of lateral gene exchange, subtractive hybridisation comparisons of *B. mallei* strain C5 with the type strain ATCC 23344 yielded sequences suggestive of conjugative transfer processes (Fushan *et al.*, 2005). These included sequences with similarity to TrbI, a component of the mating pair formation system belonging to the VirB10 family of Type IV secretion systems, and plasmid-encoded genes (Fushan *et al.*, 2005). Although the incidence of cases in the Western world is virtually zero (following modernisation of transportation systems thereby reducing the use of horses) sporadic cases are still reported in Asia, the Middle East, South America and Africa (Ulrich *et al.*, 2006).

There is documented use of *B. mallei* as a biological weapon. It was exploited by confederate forces during the American civil war, Germany during WWI and the former USSR (Waag and DeShazer, 2004). Similarly, research into the use of *B. pseudomallei* as a biological warfare agent (BWA) is thought to have taken place in the former USSR, the USA and possibly Egypt (Cheng and Currie, 2005). As a result, it is widely believed that both *B. pseudomallei* and *B. mallei* are candidate organisms for use as BWAs and as such have been classified as category B select agents by the Centre for Disease Control (CDC) (Lehavi, 2001, Rotz 2002, Voskul, 2003). The use of *B. thailandensis* as a model for *B. pseudomallei* has become popular given the close genetic relationship and the absence of requirements for high-containment facilities for its handling (Yu *et al.*, 2006)

1.2.2 ***Burkholderia cepacia*** **Complex**

The *Burkholderia cepacia* complex (Bcc) is a closely related group of bacteria within the *Burkholderia* genus (Figure 1.2) (Spilker *et al.*, 2009). Current molecular typing of the Bcc uses multi-locus sequence typing (MLST) analysis of seven housekeeping genes (ATP synthase β chain, *atpD*; glutamate synthase large subunit, *gltB*; DNA gyrase B, *gyrB*; recombinase A, *recA*; GTP-binding protein, *lepA*; acetoacetyl-CoA reductase, *phaC* and tryptophan synthase, *trpB*) and assigns each strain a unique clonal sequence type (Spilker *et al.*, 2009). The complex consists of opportunistic pathogenic species that can be divided into at least seventeen genomovars of which some are very recent additions. These are *B. cepacia*, *B. multivorans*, *B. cenocepacia*, *B. stabilis*, *B. vietnamiensis*, *B. dolosa*, *B. ambifaria*, *B. anthina*, *B. pyrocinia B. ubonensis*, *B. latens* sp. nov., *B. lata*, *B. arboris*, *B. diffusa*, *B. seminalis*, *B. metallica* and *B. contaminans* (Vermis *et al.*, 2002; Ramette *et al.*, 2005; Vanlaere *et al.*, 2008 and Vanlaere *et al.*, 2009).

Genome size and complexity have been determined for a number of strains including *B. cenocepacia* J2315, *B. cenocepacia* H12424, *B. xenovorans* LB400 and the environmental isolate *Burkholderia* sp. 383 (Mahenthiralingam and Drevinek, 2007). There are commonly 2 – 4 large chromosomes varying in size from 4 – 9 Mb in total with at least one or more plasmids. Within genomovars, strains possess diverse genotypic and metabolic potential exhibiting antibacterial and antifungal properties and pathogenicity in both plants and animals (Tsang, 2004).

Members of the Bcc have been isolated from a broad range of habitats including soil, plant and animal surfaces, the rhizosphere and water (Vandamme *et al.*, 2007). This diversity in genetic and metabolic potential has seen the use of the Bcc as biological control agents, promoters of plant growth and bioremediation (Bevivino *et al.*, 1998; Holmes *et al.*, 1998). An example of this is *B. vietnamiensis* strain G4, of interest due to its potential for degradation of complex aromatic pollutants such as trichloroethylene, through the presence of the catabolic plasmid, pTOM (O'Sullivan *et al.*, 2007). However the widespread use has been limited owing to the concerns regarding the potential risk to human health (Butler, 1995). Their notoriety as opportunistic pathogens of humans has arisen through the common implication with 'cepacia syndrome' in Cystic Fibrosis (CF) sufferers where infection is characterised by a rapid decline in pulmonary function (Mahenthiralingam *et al.*, 2002). Investigations into the epidemiology of Bcc infections in CF populations demonstrated that *B. cenocepacia* dominated with *B. multivorans* as the next most common cause of infection (Lipuma, 2001; Kidd *et al.*, 2003; Reik *et al.*, 2005). In contrast to infection with *Pseudomonas aeruginosa*, where median life expectancy is 27.8 years, infection with Bcc results in a substantial drop to 15.6 years (Speert *et al.*, 2001). A further confounding factor is the high level of intrinsic resistance to antimicrobial therapies; although this can vary depending on the isolate examined (Nzula *et al.*, 2002). In addition, the demonstration that new strains of Bcc are continuously isolated from CF patients and that clinical isolates have been shown to be indistinguishable from environmental strains, suggests an ongoing acquisition from the environment (Lipuma *et al.*, 2002). The remaining members of the Bcc account for less than 10% of infections although *B. cepacia*,

B. stabilis, *B. vietnamiensis* and *B. dolosa* are more commonly encountered (Reik, 2005).

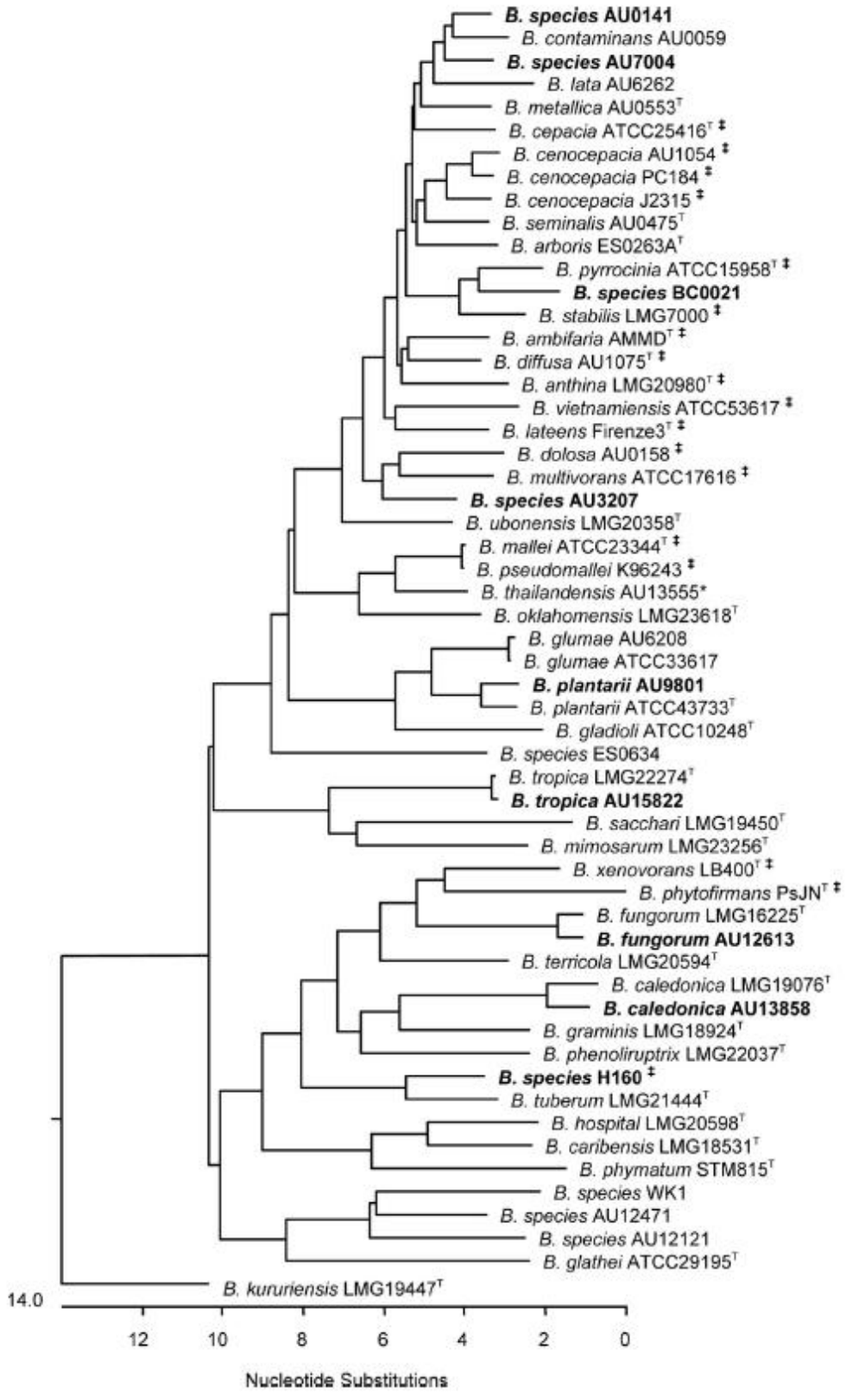

Figure 1.2: Phylogenetic Structure of the genus Burkholderia based on sequence analysis of MLST targets. (Taken from Spilker *et al*., 2009)

1.3 **Natural Genetic Transformation of *Burkholderia* sp.**

Transformation involves the uptake, integration and stable inheritance of cell free DNA by bacterial cells (Dröge *et al*. 1998; Thomas and Nielsen 2005). This phenomenon has been identified in over 60 bacterial and archeal species (Johnsborg *et al.*, 2007). At present, it is unknown if any member of the Bcc are competent for transformation. Although, as has been discussed, available sequence data suggest that these large variable genomes do have a propensity for gene exchange (Lessie *et al.*, 1996; Summer *et al.*, 2004) with the architecture likely to allow for high levels of homologous and illegitimate recombination (Holmes *et al.*, 1998). Importantly, Thongdee *et al.*, (2008) have recently shown the capacity for members of the highly pathogenic lineages to undergo natural transformation. Here they demonstrated PCR fragment-mediated allelic replacement in both *B. thailandensis* (strains E264 and S95019) and *B. pseudomallei* (1026a and b) using 800 bp flanking homologous sequences under conditions that had previously been shown to promote transformation in *Ralstonia* sp. Interestingly they were unable to transform *B. pseudomallei* strain K96243. It should be noted that given the presence of insertion sequences common to both *B. cepacia* and *B. pseudomallei* (Mack and Titball, 1998) the possibility of virulence-related gene transfer between these organisms cannot be excluded (Holmes *et al.*, 1998).

As many *Burkholderia* sp. are soil saprophytes (Mahenthiralingham *et al.*, 2008) then this transformation, if present, would need to be considered within this environment. Understanding this process would therefore take into consideration many of the important temporospatial factors that modulate transformation.

Transformations of various species within soil microcosms or in the presence of particular constituents of soil have previously been demonstrated. For example, Gallori *et al*., (1994) used chromosomal and covalently-closed-circular plasmid DNA, adsorbed onto montmorillonite clay and inoculated into non-sterile soil, to monitor the persistence of DNA transformation potential. Here they demonstrated that clay-bound DNA retained its ability to transform *B. subtilis* up to 15 days post-inoculation. In contrast, Nielsen *et al.*, (2000) assessed the persistence of DNA transformation potential in non-sterile soil by inoculating with cell lysates of *Acinetobacter* sp. strain BD413, *P. fluorescens* or *B. cepacia*. In this case, using an allelic rescue system in recipient *Acinetobacter* sp. strain BD413, they found that only homologous system rescue was observed and that this was only achievable up to 8 hours after inoculation of the lysate. Sikorski *et al.* (1998), used histidine-requiring derivatives of *P. stutzeri* JM300 to examine the transformation frequency in non-sterile sandy loam microcosms. They observed that the presence of nucleases indigenous to the microcosm did not impede the transformation process although a gradual decline was observed during prolonged incubation with 3 % activity after 3 days incubation (Sikorski *et al.*, 1998). Ultimately, transformation relies, concurrently, on the persistence of DNA in the environment as well as availability for competent organisms. In this regard, DNA is continuously released from a variety of sources such as decomposing cells, disrupted cells / viral particles and through excretion from living cells. The latter phenomenon has been demonstrated in a number of genera of bacteria including *Pseudomonas* (Thomas and Nielsen, 2005). The persistence of extracellular DNA in an environment long enough to allow for

transformation processes is an important underlying factor, and one that can vary dramatically depending on the environmental conditions present.

1.4 **DNA Persistence**

Inoculation of bacteria into soil, where there is little or no selectivity or advantage, is likely to result in a rapid decline in population size (van Veen *et al.*, 1997). This phenomenon is known as microbiostasis and extends to the introduction of typical soil bacteria (Ho and Ko, 1985). The factors governing this effect can be divided into biotic and abiotic. The former consists of predation by protozoa, competition and the presence of root material which acts as a protective, nutrient rich niche. Abiotic factors include the presence of clay minerals, water tension (where high levels prior to inoculation prevents migration of bacteria into protective pores), organic carbon, inorganic nutrients (N, P) pH and temperature (Gray 1975). Disregarding the notable exception of *B. anthracis* spores (Sinclair *et al.*, 2008), bacteria used in the scenario of biological warfare would more than likely decline in population size rapidly following introduction to the environment. Consequently, it is the long term persistence and biological activity of released nucleic acids that is of interest here.

Nucleic acids are ubiquitous in the environment (Trevors, 1996). The necessity to understand the release and persistence of DNA, extracellular in particular, in the environment is well documented for gene transfer and microbial ecology studies (Levy-Booth *et al.*, 2007; Nielsen *et al.*, 2007). Attempts have been made to link extracellular DNA into a continuously replenished cycle where it has three proposed fates:

1) Persistence through adsorption to soil components,

2) Nuclease degradation or

3) Transformation of competent organisms (Levy-Booth *et al.*, 2007).

In addition to the persistence of nucleic acids, through their analysis, contributing to the study of microbial biodiversity, it is also an important consideration when risk assessments are conducted that cover the potential introductions of non-native species e.g. genetically modified organisms (GMOs) (Nielsen *et al.*, 2007). The availability of extracellular DNA derived from these species within the horizontal gene pool ultimately determines the contribution of these elements to the evolutionary potential of indigenous populations (Nielsen *et al.*, 2007). Furthermore, it is not hard to imagine that DNA persistence may play a key role in efficacy of microbial forensics, attribution and disease surveillance. Additionally, the development of techniques that discriminate between nucleic acids derived from viable and non-viable organisms will aid in these types of investigations. These include those that target messenger RNA (mRNA) such as nucleic acid sequence based amplification (NASBA) (Deiman *et al.*, 2002) or those that utilise propidium monoazide which penetrates only membrane-damaged cells, intercalating with their DNA and consequently preventing amplification from non-viable organisms (Nocker *et al.*, 2007).

1.5 **Metagenomics**

Understanding of microbial ecology and, more specifically here, the ecology of pathogenic species within natural environments, can be achieved by examining the microbial community structure and functioning of such microbiomes. This

approach ultimately yields information related to the impact of the introduction of pathogenic species upon microbial community structures; an important source of information in the context of gene transfer and acquisition of mobile genetic elements.

Metagenomics is defined as the culture-independent genomic analysis of all micro-organisms of a particular environment (Handelsman *et al.*, 1998). The use of molecular techniques bypasses the requirement for the isolation, laboratory cultivation and observation of individual members of that community. This may include protein and metabolite-based investigations (termed meta-proteomics and meta-metabolomics. respectively). However here gene- or gene-product-based analyses will be focussed on, hereafter referred to as metagenomics, or meta-transcriptomics where mRNA is the subject of investigation. The latter approach gives an additional advantage over metagenomics in that it enables both the elucidation of community structure as well as insights into its functional capacity.

Metagenomic studies can broadly be grouped into two types depending on the desired goal. Following isolation of target nucleic acids from a particular environment, these are:

1. Shotgun sequencing of environmental DNA, whereby total gene content is discerned by random sequence analysis.
2. Amplicon sequencing from total community DNA using phylogenetic markers for particular microbial groups (Kowalchuk *et al.*, 2007).

Investigations used to rely upon technologies which were costly and laborious - notably based on the amplification and subsequent cloning of 16S rRNA targets. More recently development of high-throughput sequencing technologies such as those that involve sequencing-by-synthesis have altered the scope of metagenomic studies (Hall, 2007). Their application beyond the realm of *de novo* sequencing of individual bacterial genomes into microbial ecology, phylogenetics, transcriptomics and identification of uncharacterised infectious agents is well documented (Gilbert *et al.*, 2008; Palacios *et al.*, 2008; Urich *et al.*, 2008; Lauber *et al.*, 2009). These technologies provide several advantages over traditional sequencing methods. These include the increased data output per run with reduced cost to the researcher and the lack of a requirement for laborious clone library generation. There are a number of platform technologies that enable unprecedented sequence interrogation. Those such as the Illumina® HiSeq and MiSeq generate many millions of sequences at read lengths of ~150 bp. These utilise sequencing-by-synthesis whereby a reversible, fluorescently-labelled dNTP terminator is used to allow imaging of incorporations of single bases into growing DNA strands. In contrast, pyrosequencing using the 454 Sequencing™ platforms offers the advantage for metagenomic studies of longer read lengths (~700bp) that, in comparison to similar technologies, are the closest to that achieved through Sanger sequencing. This is particularly useful in the application of this technology to 16S rRNA-based phylogenetic studies where read length increases the confidence of phylogenetic classification.

1.5.1 **Pyrosequencing**

Pyrosequencing is a sequencing-by-synthesis technology that utilizes enzyme-coupled reactions and bioluminescence. The concept is based on detection of pyrophosphate (PPi) released during DNA amplification and was first proposed by Pål Nyrén in 1985 (Nyrén and Lundin, 1985). The reaction begins with nucleic acid polymerization which releases PPi. This is achieved by employing a modified Klenow polymerase enabling simultaneous primer extension and PPi release. This Klenow polymerase does not possess 3'-5' exonuclease activity thereby alleviating issues regarding unsynchronized DNA polymerisation (Ahmadian *et al.*, 2006). The released PPi is subsequently converted to ATP by ATP sulfurylase, which provides the energy for luciferase to oxidise luciferin. This reaction produces photons at a wavelength of 560nm. These are detected by a charge-coupled device (CCD) camera with the consequent generation of a pyrogram (Ronaghi, 2001). As the identity of each nucleotide added is known, the sequence of the template DNA can be determined by analysis of the pyrogram. The intensity of the signal received by the CCD camera is proportional to the number of nucleotides incorporated in the growing sequence. Therefore, the larger the run of identical nucleotides the greater the signal produced (Wicker *et al.*, 2006). The complete reaction from nucleotide polymerisation to production of light takes approximately 3-4 seconds (Ronaghi, 2001). Since the original idea, several improvements have been made to increase the efficiency of the reaction. The level of background luminescence was reduced by the addition of deoxyadenosine α-thiotriphosphate (dATPαS) as a substitute for deoxyadenosine triphosphate (dATP). DNA polymerase recognizes dATPαS and will incorporate it into the growing DNA strand. However, unlike

dATP, luciferase does not recognise dATPαS. Therefore, the addition of dATPαS increases the signal to noise ratio, as luciferase will not oxidase luciferin in the presence of dATPαS (Ronaghi *et al.*, 1996). The second improvement was the introduction of a nucleotide-degrading enzyme, apyrase, creating a four-enzyme system. This enabled nucleotides to be added without washing the template (Ronaghi *et al.*, 1998). The introduction of small quantities of the nucleotide-degrading enzyme allowed unincorporated dNTPs, dATPαS and ATP to be removed after each reaction cycle. The most recent addition is single-stranded binding protein. These proteins have positive effects on the ability of pyrosequencing to translate long and difficult sequences. The proteins achieve this by increasing the efficiency of the reaction, reducing mispriming and improving signal intensity (Ronaghi, 2000).

1.5.2 **Phylogenetic Markers and Application of High-throughput Sequencing for Analyses of Microbial Community Structure**

Phylogenetic marker-based analysis of environmental prokaryotic community structure has been extensively undertaken using the small subunit of the ribosomal RNA gene, more commonly referred to as 16S ribosomal RNA (rRNA) (Weisburg *et al.*, 1991). Prior to the introduction of high-throughput sequencing technologies these were done using cloned fragments of 16S rRNA regions, isolated using broad-specificity PCR primers, which were then sequenced using Sanger sequencing (Großkopf *et al.*, 1998; van Waasbergen *et al.*, 2000). Phylogenetic reconstruction of a sampled community is then based upon evolutionary distances between orthologous sequences or similarities to database entries identified through BLAST, FASTA or Bayesian classifiers.

These identify operational taxonomic units (OTU) that correspond to species of organisms (Sogin *et al.*, 2006).

Markers for microbial community analysis are required to meet certain criteria. They must be common to all bacteria, consist of variable and conserved regions and can function as an evolutionary clock (Dahllöff *et al.*, 2000). Whilst 16S rRNA analysis may be regarded as the gold standard for these studies, it suffers from both intra-species heterogeneity where sequence variation exists even within a single cell, and intra-species heterogeneity that can affect accurate community profiling (Dahlloff *et al.*, 2000). To overcome this issue, single copy genes have thus been explored. One such gene is *rpoB* and codes for the β-subunit of RNA polymerase. This subunit performs the majority of the catalytic functioning of the RNA polymerase responsible for synthesising mRNA, rRNA and tRNA in bacteria (Adékambi, 2008). Use of this target, in particular the hypervariable region between positions 2300 and 3300, has been shown to overcome the issue of heterogeneity in PCR-DGGE comparisons with 16S rRNA (Dahllöff *et al.*, 2000). Furthermore it can provide accurate and reproducible profiles for species identification, simple mixtures and complex microbial communities. In some instances it can clarify species boundaries where 16S rRNA profiling has failed (Adékambi, 2008). A further alternative for species reconstruction of complex microbiomes is *gyrB* which encodes the β-subunit of DNA gyrase. Like *rpoB* this gene is ubiquitous in bacteria and has thus been exploited for its utility in microbial community profiling. In one example, Yin *et al.*, (2008) used RFLP analysis of 16S rRNA and *gyrB* amplicons to investigate the microbial community dynamics and impact of local geochemical composition

in acid mine drainage sites. The authors were able to demonstrate not only that both markers suggested the dominance of the same bacterial classes but that *gyrB* gave a higher diversity index for the three sites investigated (Yin *et al.*, 2008). Currently the issue for both these alternative targets is the smaller database sizes when compared to those available for 16S rRNA. Until these are available to the same level as the Ribosomal Database Project, for example, it is unlikely that their use in high-throughput sequencing approaches will increase.

Certain pyrosequencing approaches have enabled a substantial increase in the depth of 16S rRNA coverage achievable from a given sample. However at present the read length is still not sufficient to cover the full length gene. As a result many approaches target specific variable regions of the 16S rRNA gene to elucidate metagenomic data from environmental samples. Sogin *et al.*, (2006) used PCR amplification of the V6 region to examine microbial diversity in oceanic samples. Through comparison of the sequences to a database of 40,000 unique V6 regions the authors were able to examine the complexity of microbial species richness within each sample using rarefaction analysis. With this reduction in achievable fragment length it has been necessary to compare the outputs of such deep sequencing studies with those achieved using near full length sequence reads and Sanger sequencing. For example, research has been undertaken to elucidate the impact of shorter read lengths achieved and the impact on estimated species richness and assignment of phylogeny (Youssef *et al.*, 2009). These studies have helped pinpoint specific regions of the 16S rRNA gene that provide greater phylogenetic resolutions. Additionally they have determined that the shorter read length is in fact not a limitation when estimating

species richness. Youssef *et al.*, (2009) compared species richness estimates and operational taxonomic units (OTU) from 1132 16S rRNA clones from a soil sample with simulated shorter read outputs. They determined that selection of target variable regions impacts on these outputs with some over-estimating and others under-estimating diversity. Amplicons that cover V4, V5+V6 and V6+V7 were found to produce comparable results to full length analyses. However, these regions were selected on the basis of determining species richness through clustering at 97 % sequence similarity, and therefore in the context of phylogentic resolution alternative regions may be more applicable. Investigating fragment lengths of conserved regions and the impact on phylogenetic classification, Liu *et al.*, (2007) demonstrated fragments of 250 bp around the V1+V2, V3 and V4 region provided the best targets. In a similar study Wang *et al.*, (2007) demonstrated similar utility through the use of V2 and V4 regions. Ultimately, the choice of variable region target depends entirely on the goal of the study. For pathogen detection use of 16S rRNA variable regions has been demonstrated using both short (~100 bp) and long (~600 bp) pyrosequencing approaches (Jordan *et al.*, 2005; Luna *et al.*, 2007; Dowd *et al.*, 2008; Price *et al.*, 2009). However, selection should be guided by those that give the greatest resolution and the use of additional targets is recommended, particularly for closely related species (Ruppitsch *et al.*, 2007). This was demonstrated by Chakravorty *et al.*, (2007) by comparison of sequence similarity dendrograms generated through the analysis of variable regions one to eight in 110 bacterial species that included many pathogens. Here the authors showed that selection of V2 and V3 enabled the greatest level of differentiation to genus level (Chakravorty *et al.*, 2007).

1.5.3 **Application of Whole Genome Sequencing to Pathogen Detection**

An advantage of utilising whole genome shotgun sequencing approaches for pathogen detection is that it negates the requirement for a pre-defined panel of assays that use either antibody or PCR technology. As a consequence, sequencing-based identification of emerging pathogens is possible. Recently, Nakamura *et al.* (2008) highlighted this approach for the identification of the causative agent of diarrhoeal illness where culture methods and reverse transcriptase PCR analysis failed to identify the causative agent. Comparative analysis of sequences generated showed that, of the 96,941 sequences obtained for the illness sample, 156 aligned to those from *Campylobacter jejuni* with absence from the recovery samples (Nakamura *et al.*, 2008). Adopting a similar approach, Palacios *et al.* (2008) identified a new Old World Arenavirus in kidney and liver tissues from a cluster outbreak in transplant patients. Three patients, having received organs from a single donor, developed febrile illness and died 4-6 weeks post-transplantation. Culture techniques, PCR assays for 21 viral candidates as well as toxoplasma, *Mycobacterium tuberculosis*, *Mycoplasma pneumoniae* with viral and panmicrobial oligonucleotide microarray analysis revealed no candidate pathogens (Palacios *et al.*, 2008). Un-biased sequencing yielded data that, when converted into deduced protein sequences and analyzed with BLASTX, aligned with 14 fragments of an Old World Arenavirus closest to lymphocytic choriomeningitis viruses (LCMV) (Palacios *et al.*, 2008). Whether similar approaches can be applied to the detection of known or unknown pathogens in more complex environmental sample matrices remains to be elucidated.

1.5.4 **Emerging Approaches for Metagenomic Studies - Metatranscriptomics**

As the depth of sequence coverage that is now achievable for a single sample has increased substantially whole genome sequencing has, in some cases, replaced 16S rRNA profiling for the characterisation of environmental samples. However, it is now suggested that total RNA profiling will supersede both these approaches. Here indications of phylogenetic structure, from determining actively transcribed rRNA, and functional gene content, by mRNA analysis, are achieved simultaneously. Ultimately this yields information relating to the adaptation and evolution of communities that would not be achieved through DNA-based metagenomic approaches (Cubillos-Ruiz *et al.*, 2010). It is clear that given the complexity and heterogeneity of soil environments, DNA shotgun sequencing alone will not capture its diversity (Gabor *et al.*, 2007). As an example, only 1% of 150,000 sequences that totalled 100 Mbp which were derived from a farm soil could be assembled into contigs (Tringe *et al.*, 2005). In demonstration of alternative approaches, Urich *et al.*, (2008) investigated a sandy soil ecosystem using a 'double-RNA' approach whereby the total RNA pool was transcribed, sequenced and then binned for searches using databases of both small and large subunit rRNA and MEGAN for taxonomic and functional characterisation. In this instance the authors did not attempt to enrich for any particular domain of life. The power of this approach is demonstrated where the authors, from a single sample, acquired over 250,000 reads of which nearly 200,000 were annotated as rRNA sequences (giving taxonomic identifications) with the remainder of mRNA origin. It is also possible to enrich for mRNA sequences present in an environmental sample and then combine these data with those derived from DNA-based metagenomics, as demonstrated by Gilbert *et al.*,

(2008) in their study of marine mesocosm responses to ocean acidification. In this way a greater level of enrichment for mRNA can be achieved whilst still permitting phylogenetic reconstruction. Additionally, suppressive subtractive hybridisation (SSH) may also be applied to cDNA to increase the probability of finding rarer transcripts or metgenomic DNA fragments (Rebrikov *et al.*, 2004).

Figure 1.3 shows the general steps in metagenomic studies that may be applied to a sample for identification of an unknown pathogen. The wider application of these approaches to the detection and identification of pathogenic micro-organisms in environmental sample types remains an issue. In these contexts it needs to be determined which approach will yield suitably sensitive and specific identification of pathogens within a complex biological background. Moreover, where total community nucleic acids content is limited, the use of whole genome amplification strategies to increase analysable material has been employed (Gilbert *et al.*, 2008). This is particularly relevant if the detection of low levels of pathogenic organisms is a challenge.

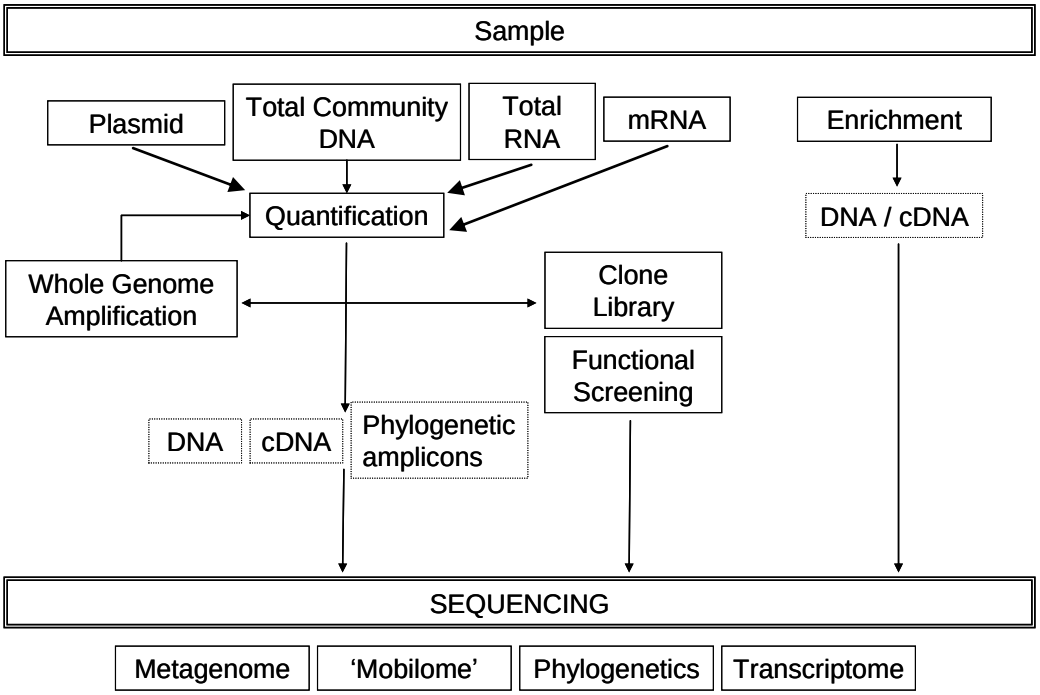

Figure 1.3: Schematic highlighting common steps in metagenomics strategies.

1.6 **Project Aims**

This study aims to address a number of facets of microbial detection and adaptation in complex communities. These include the availability and biological activity of nucleic acids from the Burkholderiaceae in soil and the analysis of such complex communities by high-throughput sequencing methods. Specific aims are detailed within each chapter however the overall aims of this project can be summarised as follows:

1. Determine how soil type can influence the persistence of nucleic acids.

2. Determine if members of the *Burkholderia cepacia* complex (Bcc) undergo gene acquisition from extracellular DNA through natural genetic transformation.

3. Investigate the impact, in terms of accurate microbial community profiling, of the choice of 16S rRNA variable gene for high-throughput sequencing analysis. These will be achieved through the comparative analysis of the changes in the community profile of soils that have been exposed to long-term antibiotic treatment using various 16S rRNA assays.

4. Where the identification of pathogens is the desired goal of microbial community profiling of environmental samples it is necessary to compare methods that may facilitate this. Here, the suitability of shotgun metagenomic and 16S rRNA community profiling approaches using high-throughput sequencing were compared for the detection of bacterial pathogens within complex bio-aerosol communities.

Chapter 2

Materials and Methods

2.1 **Clay Minerals**

2.1.1 **Generating Homo-ionic Minerals**

Bentonite, kaolinite or silicon dioxide (Sigma-Aldrich, UK) were made homo-ionic according to the method of Khanna and Stotzky (1992). Briefly, the mineral (0.5 g) was washed in 1M $CaCl_2$ / $MgCl_2$ or NaCl solution (5 ml) then centrifuged at 16,000 g at 4 ^oC for 10 min. The supernatant was removed and the wash repeated three times. After the final wash the mineral suspension was divided into 1 ml aliquots. The suspension was then centrifuged at 16,000 g at 4 ^oC for 10 min. The supernatant was removed leaving the remaining pellet containing approximately 100 mg homo-ionic mineral.

2.1.2 **DNA Adsorption to Homo-ionic Minerals**

Briefly, soil mineral (0.5 g) was re-suspended in 5 ml of a 1M solution of $MgCl_2$, $CaCl_2$, or NaCl. Following three centrifugations (16,000 g at 4 ^oC, 10 min) and washing steps the homo-ionic mineral solution was aliquoted into 1 ml volumes, each containing approximately 100 mg of material. Volumes of 0.1 to 100 µl containing ~0.1 to ~10 mg of mineral were used for adsorption studies. Minerals were incubated with Calf Thymus DNA (Sigma-Aldrich, UK) at a concentration of 500 ng μl^{-1}. Quantitative analysis of the reduction in DNA concentration (used to determine the level remaining, and thus adsorbed to the mineral) was performed using 260/280 nm adsorption on a NanoDrop® spectrophotometer (NanoDrop® Technologies, UK).

2.2 **Media and Reagents**

2.2.1 **Microbiological Media**

A number of microbiological growth media were used in this study. These are listed in Table 2.1 below.

Table 2.1: Microbiological Growth Media

Media	Constituents
SOB	0.5% (w/v) yeast extract, 2% (w/v) tryptone, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$.
Nutrient Agar	Oxoid, pH 7.4 ± 0.2
Tryptic Soy Agar	Oxoid, pH 7.3 ± 0.2
Tryptic Soy Broth	Oxoid, pH 7.3 ± 0.2
Luria-Bertoni Agar	Tryptone 10 g, Yeast Extract 5 g, NaCl 10 g, Bacto Agar 15 g pH 7.4 ± 0.2
Luria-Bertoni Broth	Tryptone 10 g, Yeast Extract 5 g, NaCl 10 g, pH 7.4 ± 0.2
Defined Minimal Medium (DMM)	Defined Minimal Medium (DMM): 1/4 strength M63 medium supplemented with glucose 2g L^{-1}, glycerol 4ml L^{-1}, 1mM MgSO4.7H20, thiamine 1µg mL^{-1} and leucine, isoleucine, valine, tryptophan, glutamic acid, and glutamine at 40µg ml^{-1} each.

2.2.2 **Antibiotics**

Table 2.2: Antibiotics Used in this Study

Antibiotic	Stock Solution	Final Concentration
Trimethoprim[1]	100 mg ml^{-1} solution in dimethylsulfoxide	1.5 mg ml^{-1}
Kanamycin[1]	100 mg ml^{-1} in dH_2O	4 – 258 µg ml^{-1}
Ampicillin[2]	50 mg ml^{-1} in dH_2O	50 µg ml^{-1}
Gentamicin[1]	50 mg ml^{-1} in dH_2O	15 µg ml^{-1}

[1] **Sigma-Aldrich**
[2] **Roche Applied Sciences**

2.2.3 **Buffers and Solutions**

Table 2.3: Buffers used in this study

Buffer / Solutions	Constituents
Crombach Buffer	Tris-HCl (33 mM), EDTA (1 mM)
Soil Lysis Buffer	100 mM Tris-HCl, 100 mM sodium-EDTA, 1.5 M NaCl
Tatrazine Loading Buffer	50 % v/v glycerol, 0.1 g Tartrazine
Klenow Filling Buffer	500 mM Tris-HCl pH 7.5, 100 mM $MgCl_2$, 10 mM DTT, 500 µg ml^{-1} BSA
S1 Nuclease Buffer	30 mM sodium acetate, pH 4.5, 50 mM NaCl, 1 mM $ZnCl_2$
50 x TAE[1]	Tris Base (242 g), Glacial Acetic Acid (57.1 ml), 0.5 M EDTA, pH 8.0 (100 ml)
DNaseI 'Stop Solution'	EDTA (10 mM) in 20 mM NaCl
Gel Loading Buffer	Sucrose (60 %), EDTA (100 mM), Bromophenol Blue (0.25 % w/v)

2.3 **Bacterial Strains**

2.3.1 ***Bacillus anthracis*** **Ames Spores**

Stocks of *B. anthracis* Ames spores were held at Dstl at ACDP containment level 3 (CL3). Stocks, enumerated at 1 x 10^{10} spores ml^{-1} by plate count, were used here.

2.3.2 ***Bacillus globigii*** **NCTC 10073**

Vegetative cells of *B. globigii* NCTC 10073 were produced from glycerol stocks held at Dstl and stored at -80 ^{o}C. To culture, 10 µl of glycerol stock were streaked onto LB agar plates and incubated overnight in a static 37 ^{o}C incubator (Jencons, UK). A single colony was then used to inoculate 100 ml of LB broth in a 250 ml vented conical flask (VWR, UK). Cultures were propagated in a Stuart shaking incubator at 200 rpm (Jencons, UK) at 37 ^{o}C, overnight. The cell density was calculated by plate counts where 100 µl of a $\log_{10}$ dilution series were plated in triplicate on LB Agar plates. These were incubated at 37 ^{o}C overnight prior to counting of colonies.

2.3.3 ***Burkholderia cepacia*** **Complex**

The *Burkholderia cepacia* Complex strain panel used was kindly supplied by Prof. John Govan (Edinburgh University). Details of the strains in this panel are given in Table 2.4.

Table 2.4: *Burkholderia cepacia* Complex strains used in this study

Species	Strain Name	Genomovar	Strain Collection	Isolation
B. cepacia	25416	I	ATCC	Environmental
B. cepacia	17759	I	ATCC	Environmental
B. cepacia	C2970	I	-	Clinical
B. cepacia	C3159	I	-	Clinical
B. cepacia	C1963	I	-	Clinical
B. cepacia	J2540	I	-	Environmental
B. cepacia	9091	I	NCIMB	-
B. cepacia	25608	I	ATCC	-
B. cepacia	C1964	I	BCCM / LMG	Environmental
B. multivorans	13010	II	BCCM / LMG	Clinical
B. multivorans	17616	II	ATCC	Environmental
B. multivorans	7732	II	-	-
B. multivorans	7897	II	-	-
B. cenocepacia	18863	IIIA	BCCM / LMG	Clinical
B. cenocepacia	J2956	III	BCCM / LMG	Clinical
B. cenocepacia	C1394	IIIB	-	Clinical
B. cenocepacia	J415	IIIB	-	Clinical
B. cenocepacia	C2836	III	-	Clinical
B. stabilis	ERL347	IV	-	Clinical
B. stabilis	14086	IV	BCCM / LMG	Environmental
B. stabilis	14294	IV	BCCM / LMG	Clinical
B. stabilis	7639	IV	-	-
B. stabilis	J493	IV	-	Environmental
B. stabilis	C3172	IV	BCCM / LMG	Clinical
B. stabilis	8088	IV	-	-
B. stabilis	J1743	IV	-	Environmental
B. stabilis	J1763	IV	-	-
B. vietnamiensis	10929	V	BCCM / LMG	Environmental
B. vietnamiensis	18836	V	BCCM / LMG	Clinical
B. vietnamiensis	549	V	-	-
B. vietnamiensis	638	V	-	-
B. vietnamiensis	ERL126	V	-	Clinical
B. vietnamiensis	C2978	V	BCCM / LMG	Clinical
B. vietnamiensis	159	V	-	-
B. vietnamiensis	C3175	V	BCCM / LMG	Clinical
B. ambifaria	BAA-244	VII	ATCC	Environmental
B. anthina	J2540	VIII	-	Environmental
B. pyrocinia	15958	IX	ATCC	Environmental

2.3.4 ***Burkholderia pseudomallei* K96243**

B. pseudomallei K96243 inocula were produced from glycerol stocks held at Dstl and stored at -80 ^{o}C. A sterile loop was used to transfer ~10 µl of the glycerol stock onto LB agar plates. These were subsequently incubated overnight in a static 37 ^{o}C Binder incubator (VWR, UK). A single colony was then used to

inoculate 100 ml of LB broth in a vented conical flask (VWR, UK). Cells were grown in a shaking incubator (Eppendorf, UK) at 200 rpm and 37 ^{o}C, overnight. Enumeration of CFU ml^{-1} was undertaken by plate count where 100 µl of a $\log_{10}$ dilution series were plated in triplicate on LB Agar plates. These were incubated at 37 ^{o}C overnight prior to counting of colonies.

2.3.5 **Determination of Antibiotic Minimum Inhibitory Concentration**

Where applicable, the antibiotic susceptibility of bacterial isolates was determined by broth dilution method. Stock solutions of antibiotics were prepared according to Table 2.2. These were diluted in LB broth in an exponential series from 4 µg ml^{-1} up to 256 µg ml^{-1}. The determination of Log phase cultures for inoculation was made by optical density measurements of 1 ml in 2.5 ml Sterilin disposable plastic spectrophotometer cuvettes (Fisher Scientific, UK) at 600 nm in a HEλIOS γ spectrophotometer (Thermo Scientific, UK). Log phase cultures were then adjusted to an O.D$_{600}$ of 1.0 using LB Broth. Finally, 20 ml of LB broth with antibiotic at the above concentrations were inoculated with 500 µl culture and incubated overnight at 37 ^{o}C with shaking (100 rpm). The minimum inhibitory concentration was determined by triplicate O.D$_{600nm}$ measurements of overnight cultures.

2.3.6 **Storage of Strains**

A single bacterial colony was streaked onto an LB agar plate and incubated at 37 °C, overnight. A sterile inoculating loop was used to take a selection of colonies from this plate and emulsify them in 1 ml of 30 % (v/v) glycerol in a cryotube. These were then stored at –80 °C.

2.4 **Nucleic Acid Extractions and Manipulations**

2.4.1 **Genomic DNA Extraction**

Unless stated otherwise, genomic DNA was extracted using a QIAamp DNA mini-prep kit or the Gentra Purgene Yeast/Bact. DNA Isolation kit (both Qiagen, UK) according to the manufacturer's instructions.

2.4.2 **DNA Extraction from Soil**

A number of methods were used for the extraction of total DNA from soil. These are summarised below. All chemicals were purchased from Sigma-Aldrich, UK unless stated otherwise.

2.4.2.1 MoBio UltraClean® Soil DNA Isolation Kit

This commercial kit enables the extraction of total DNA from ~0.25 g soil. The kit was used according the manufacturer's guidelines. Soil samples were added to the provided 'bead-solution' tube with 60 µl of solution S1. Following vortexing, 200 µl of inhibitor removal solution (IRS) was added. Samples underwent bead-beating using a Vortex-Genie 2 (Cambio Ltd, UK) at maximum speed for 10 min. After centrifugation at 10,000 g for 30 sec the supernatant was transferred to a sterile 2 ml tube. Solution S2 (250 µl) was then added for the precipitation of proteins. This was then incubated at 4 ^{o}C for 5 min. Following centrifugation at 10,000 g for 1 min the supernatant was transferred to a 2 ml tube containing 900 µl of solution S3. Approximately 700 µl were then loaded onto a silica spin column and centrifuged at 10,000 g for 1 min. This was repeated until the entire sample had been passed through the column. The column was washed with 300 µl of solution S4 prior to a dry centrifugation to

remove residual S4. DNA was eluted from the column by the addition of 50 µl of solution S5 and centrifugation at 10,000 g for 30 sec.

The alternative lysis method from this kit for higher molecular weight DNA recovery was also used. Here, mechanical lysis through bead beating was replaced by incubation at 70 ^{o}C for 5 min after the addition of S1 and IRS solutions. Following the initial 5 min incubation the sample was vortexed and the heat lysis step repeated. The method then follows that stated above.

2.4.2.2 MoBio PowerSoil® Soil DNA Isolation Kit

This commercial kit enables the extraction of total DNA from ~0.25 g soil. The kit was used according the manufacturer's guidelines. Soil samples were added to the provided 'bead-solution' tube with 60 µl of solution C1. Samples underwent bead-beating using a Vortex-Genie 2 (Cambio Ltd, UK) at maximum speed for 10 min. After centrifugation at 10,000 g for 30 sec, the supernatants were transferred to a sterile 2 ml tube (provided) to which 250 µl of solution C2 were then added. This was incubated at 4 ^{o}C for 5 min. Following centrifugation at 10,000 g for 1 min the supernatant was transferred to a 2 ml tube containing C3 solution (200 µl). After incubation at 4 ^{o}C for 5 min and subsequent centrifugation at 10,000 g for 1 min, the supernatant was transferred to a 2 ml tube containing 1.2 ml of solution C4. Approximately 700 µl were then loaded onto a silica spin column (provided) and centrifuged at 10,000 g for 1 min. This was repeated until the entire sample had been passed through the column. The column was washed with 300 µL of solution C5 prior to a dry centrifugation to remove residual C5. DNA was eluted from the column by the addition of 50 µl of solution C6 and centrifugation at 10,000 g for 30 sec.

2.4.2.3 Modified UltraClean® Soil DNA Isolation Kit (Andrews *et al.*, 2004)

This method follows that detailed above with the addition of an ethanol precipitation step following DNA elution from the silica spin column. Firstly, an equal volume of ammonium acetate (5M) and 2 volumes of cold absolute ethanol were added. The sample was then incubated at -20 ^{o}C for 20 min. Samples were centrifuged for 15 min at 13,600 g. The supernatant was decanted and 1 ml of 70 % (vol/vol) ethanol was used to wash the DNA pellet. The sample was centrifuged as before and, following removal of the supernatant, the pellet was air-dried at room temperature before resuspension in 50 µl of TE Buffer. To this was added 1 ml of solution S3 from the UltraClean® kit which was then processed through provided spin columns as detailed above.

2.4.2.4 Phenol / Chloroform (Krsek and Wellington, 1999)

Total DNA was extracted from 0.3 g of soil as follows. Soil was resuspended in 400 µl Crombach buffer (Table 2.3) with 0.25 g glass beads (0.10 - 0.11 mm diameter) (Sigma-Aldrich, UK) and shaken vigorously on a Vortex-Genie 2 (Cambio Ltd, UK) for 5 min. This was followed by the addition of 100 µl of 5 mg ml^{-1} Lysozyme (Roche Applied Sciences, Germany) at 5mg ml^{-1} whereupon the sample was incubated for 1 hour at 37 ^{o}C on an Eppendorf thermomixer (Fisher Scientific, UK). Sodium dodecyl sulphate (SDS) (Roche Applied Sciences, Germany) was then added to a final concentration of 1 %. The sample was then incubated for a further 30 min at 65 ^{o}C. Following centrifugation at 4,000 g and removal of the supernatant, a precipitation was performed on ice using potassium acetate (8M). The sample was then centrifuged for 20 min at 13,800 g at 4 ^{o}C. The supernatant then underwent a further precipitation overnight at 4 ^{o}C in 50 % polyethylene glycol (PEG) (wt/vol) with 1/10 vol. NaCl (5M). Following centrifugation a phenol: chloroform: isoamyl alcohol

(PCIAA) (25:24:1) extraction was performed on the supernatant using an equal volume. Following thorough vortexing the mixture was centrifuged at 12,000 g for 20 min at room temperature. Aqueous phases from two repeats of the PCIAA extraction were then pooled. DNA was precipitated overnight with 0.9 vol. isopropanol at 4 ^{o}C. Following centrifugation the pellet was resuspended in 0.9 vol. 5 mM spearmine-HCl and incubated for 2 h. The sample was then centrifuged and the pellet resuspended in TE Buffer.

2.4.2.5 Epicentre SoilMaster™ DNA Extraction Kit

Total DNA from <100 mg of soil was extracted according to the manufacturer's guidelines. Briefly, 250 µl of soil DNA extraction buffer and 2 µl Proteinase K were first added to the soil in a 1.5 ml sterile eppendorf tube (Fisher, UK). This was then incubated at 37 ^{o}C for 10 min, with shaking at 400 rpm, on a thermomixer (Eppendorf, UK). To this were added 50 µl of soil lysis buffer. The sample was briefly vortexed before incubation at 65 ^{o}C for 10 min in a Grant thermostatic water bath (Grant Instruments Ltd, UK). Following centrifugation at 2000 g for 2 min, the supernatant was removed into a sterile 1.5 ml eppendorf tube. To this was added 1/3 (vol/vol) of a protein precipitation reagent followed by incubation on ice for 8 min. Precipitated material was removed by centrifugation at 13,000 g and the supernatant transferred to a spin column containing a resin to remove enzyme inhibitors. The sample was passed through the resin by centrifugation at 2000 g for 1 min. DNA in the flow-through was then precipitated by adding 6 µl of DNA precipitation reagent followed by incubation at room temperature for 5 min. Following centrifugation at 13, 000 g for 5 min the supernatant was removed and the DNA pellet washed in 500 µl of

pellet wash solution; this was repeated twice. The pellet was resuspended in 50 µl TE Buffer.

2.4.2.6 Phenol / Chloroform-based Extraction (Yeates *et al.*, 1998)

The method was adapted from Yeates *et al.*, (1998) to enable extraction from 0.3 g of soil. Lysis was performed in 400 µl soil lysis buffer (Table 2.3) with 0.25 g glass beads (0.10 – 0.11 mm diameter, Sigma-Aldrich, UK) with vigorous shaking on a Vortex-Genie 2 (Cambio Ltd, UK) for 2 min. Lysis was achieved by the addition of 25 µl 20 % SDS (Roche Applied Sciences, Germany) and incubation at 65 ^{o}C for one hour on an Eppendorf thermomixer (Fisher Scientific, UK) with shaking at 500 rpm. Following centrifugation at 6,000 g for 10 min, ½ vol. PEG (30 %) in 1.6 M NaCl was added to the supernatant. This was incubated at room temperature for 2 h. Following incubation the sample was then centrifuged and the supernatant discarded. The pellet was resuspended in 1 ml of 10 mM Tris-HCl, 1 mM sodium-EDTA. Potassium acetate (7.5 M) was then added to a final concentration of 0.5 M and incubated on ice for 5 min. This was centrifuged for 30 min at 12,000 g at 4 ^{o}C. A PCIAA extraction was performed on the supernatant. Following thorough vortexing the mixture was centrifuged at 12,000 g for 20 min at room temperature. Aqueous phases from two repeats of the PCIAA extraction were then pooled. This was followed by an isopropanol precipitation (0.6 vol. for 2 h at room temperature). Following centrifugation at 12,000 g at 4 ^{o}C for 30 min the supernatant was discarded and the pellet resuspended in TE Buffer.

2.4.2.7 MoBio PowerMax® Soil DNA Isolation Kit

This kit enables the extraction of total DNA from >10 g of soil. Firstly, 15 ml of 'PowerBead' solution were added to the 'PowerBead' tube. Up to 10 g of soil

sample were then added to these tubes and vortexed for 1 min using a Vortex-Genie® 2. Following this, 1.2 ml of solution C1 was added and the samples vortexed again for 10 min. Tubes were centrifuged at 2,500 g for 3 min. Supernatants were subsequently transferred to a fresh collection tube. To this was added 5 ml of Solution C2. The tubes were inverted twice to mix and then incubated at 4 ^{o}C for 10 min. After incubation, samples were centrifuged at 2,500 g for 4 min at room temperature. The supernatant was then transferred to a clean collection tube containing 4 ml of Solution C3**.** After briefly vortexing the samples were incubated at 4 ^{o}C for 10 min. Following centrifugation at 2,500 g for 4 minutes at room temperature the supernatant was transferred to a clean collection tube containing 30 ml of solution C4. The sample was then passed through a silica spin filter by centrifugation at 2,500 g for 2 minutes at room temperature. The spin filter was washed using 10 ml of solution C5 by centrifugation at 2,500 g for 3 min at room temperature. The spin column was then centrifuged for 5 min at 2,500 g for 5 min and placed into a new collection tube. DNA was eluted from the column with 5 ml of solution C6. This was centrifuged at 2,500 g for 3 min at room temperature.

2.4.3 **DNA Extraction from Collected-Aerosol Samples**

Aerosols were collected using high-efficiency, wetted walled cyclones (HWWC) with samples provided in 10 mM HEPES / 0.01 % Tween-80 buffer. Sub-samples (50 ml) were filtered using 25 mm Durapore® PVDF 0.22 µm filters in Swinnex® syringe filter holders (Millipore, UK). Filters were curled inwards to expose the 'contaminated' surface to 1 ml Instagene matrix (Bio-Rad Laboratories Ltd, UK) in a screw cap 1.5 ml microfuge tube. Each sample was vigorously vortexed to encourage displacement of bacteria and particulate matter from the filter surface

and subsequently heated in an Eppendorf thermomixer with shaking at 1400 rpm for 15 min. The tubes were immediately placed on ice for 5 min before centrifugation at 14,000 rpm at 4 °C. Supernatants, containing crude DNA extracts, were transferred to a new 1.5 ml tube and stored at -20 ^{o}C.

2.4.4 **Plasmid DNA Extraction**

Small- and large-scale preparations of plasmid DNA were generated using the QIAprep Spin Miniprep Kit and HiSpeed Plasmid Maxi Kit, respectively (Qiagen, UK). All preparations were carried out as per manufacturer's instructions.

2.4.5 **Agarose Gel Electrophoresis**

Between 10 µl and 25 µl of samples (amplicon or DNA extracts) were loaded onto agarose gels with a final concentration of 1 X Tartrazine loading buffer. Gel density percentages were based on w/v of agarose (Type II, Sigma) in 1 X TAE buffer (Bio-Rad Laboratories Ltd, UK) and contained a final concentration of 1.5 mg ml^{-1} ethidium bromide (Bio-Rad Laboratories, UK) or 1.5 mg crystal violet (Sigma-Aldrich, UK). Molecular weight standards XIV (100 bp ladder), III or IV (all obtained from Roche Applied Sciences, Germany) were loaded into adjacent lanes. Each gel was run in 1 x TAE buffer at 90 volts until the amplicons were separated (30 to 60 min). Alternatively, E-gels® (Invitrogen, Life Technologies, UK) were used to visualise amplicons. E-gels® were used according to the manufacturer's instructions. Agarose gels were visualised using a Chemi HR410 BioSpectrum® Imaging System (Ultraviolet Products Ltd, UK).

2.4.6 **Nucleic Acid Extraction from Agarose Gels**

Gel slices were excised using a scalpel blade whilst the DNA band was illuminated by UV light on a transilluminator (Ultraviolet Products Ltd, UK) and transferred to a 1.5 ml Eppendorf tube. Gel slices were then processed using the QIAquick Gel Extraction kit (Qiagen, UK) according to manufacturer's instructions.

2.4.7 **Microcon® Purification**

Purification of nucleic acids using Microcon® YM-100 centrifugal filter units (Millipore, UK) was performed according to the manufacturer's instructions.

2.4.8 **Ethanol Precipitation**

Unless otherwise stated ethanol precipitation was performed as follows. Firstly, the volume of the DNA solution to be precipitated was measured. Sodium acetate (3 M) pH 5.2 (Sigma-Aldrich, UK) was then added (1/10 vol.) followed by 2.5 volumes of cold absolute ethanol (Sigma-Aldrich, UK) and the mixture vortexed and incubated at -20 ^{o}C for $\geq$ 20 min. The sample was centrifuged at 13,600 g for 15 min. The supernatant was then decanted and 1 ml of 70% (vol/vol) ethanol (Sigma-Aldrich, UK) was used to wash the DNA pellet. The sample underwent a further centrifugation as above and following decanting of the supernatant the pellet was air-dried overnight at room temperature. Finally the pellet was resuspended in 50 µl TE Buffer.

2.4.9 **Quantification of Nucleic Acids**

2.4.9.1 UV Densitometry

DNA samples were analysed on the Nanodrop® (Labsystems, UK), which measured the absorbance at 260 nm and 280 nm of 1 µl of the DNA solution and calculated the DNA concentration accurately, provided it was between 1 and 1000 ng μl^{-1}. It also compared the ratio of 260/280. A value of 1.8 to 2 was indicative of a good quality DNA preparation with little or no protein contamination.

2.4.9.2 Picogreen/Ribogreen Quantification of DNA

Double- and single-stranded DNA quantification was carried out using the Qubit® fluorometer using Picogreen® or Ribogreen® Quant-iT™ intercalaters (Invitrogen®, Life Technologies®, UK), according to the manufacturer's instructions.

2.4.9.3 Agilent 2100 Bioanalyser DNA Analysis

Where stated, DNA extracts were analysed on an Agilent 2100 Bioanalyser (Agilent, UK) for fragment size distribution and concentration according to the manufacturer's instructions. Briefly, a gel-dye mix was prepared by adding 15 µl of High sensitivity dye to a vial of gel matrix. Gel-dye matrix (9 µl) was added to the High-sensitivity DNA chip. Following chip priming, 5 µl of marker were added to each well and 5 µl of sample or water were added to each unmarked lane. The DNA ladder (1 µl) was added to the yellow well and the chip was vortexed at 2400 rpm for 1 min. The chip was then loaded into the Bioanalyser and run for 30 min.

2.4.10 **Whole Genome Amplification**

Whole genome amplification of genomic or environmental DNA was done using the Qiagen® REPLI-g midi kit (Qiagen, UK). The kit was used according to manufacturer's instructions.

2.4.11 **Removal of Multiple Displacement Amplification-generated, Hyperbranched DNA Structures**

For the removal of hyperbranched structures resulting from multiple displacement amplification (MDA), purified DNA underwent a three-step debranching strategy as outlined by Zhang *et al.*, (2006). Purified amplicon was incubated with 8 U μl^{-1} RepliPHI™ Phi29 DNA polymerase (Epicentre®, UK), 1mM dNTP (Roche Diagnostics, Germany) and 1 x RepliPHI reaction buffer in 50 µl at 30 ^{o}C for 2 h, 65 ^{o}C for 3 min. This was followed by digestion with 1 U μl^{-1} S1 nuclease (NEB, UK) in 200 µl 1 x buffer at 37 ^{o}C for 30 min. Debranched DNA was then extracted by phenol/chloroform. An equal volume of PCIAA (25:24:1) was added to the DNA containing reaction mixture and vortexed gently. The mixture was then centrifuged at 2,000 rpm for 5 min whereupon the aqueous phase was transferred into a fresh microfuge tube. DNA was then precipitated by ethanol precipitation.

2.4.12 **Nucleic Acid Manipulations**

2.4.12.1 Restriction Endonuclease Digestion of DNA

DNA was digested with restriction enzymes using the buffers and conditions as specified in the manufacturer's (NEB, USA or Roche Diagnostics, Germany) instructions. Digestions were typically carried out in a volume of between 20 and 100 µl.

2.4.12.2 Klenow Treatment

Fill-in reactions of sticky-end fragments were performed as follows. Typically 1 U of Klenow enzyme (Roche Diagnostics, Germany) was incubated with >1 µg DNA, 1 mM of each desired dNTP and 2 µl of 10 x Klenow filling buffer in 20 µl volume adjusted with dH_2O. Samples were then incubated at 37 ^oC for 15 min in an Eppendorf Thermomixer Compact (Fisher Scientific, UK).

2.4.12.3 Alkaline Phosphatase Treatment

DNA was dephosphorylated using shrimp Alkaline Phosphatase (SAP) according to the manufacturer's guidelines (Roche Diagnostics, Germany). DNA (50 ng) was incubated with 3 U SAP in 0.9 µl 10 x dephosphorylation buffer. This was followed by incubation at 37 ^oC for 10 min (60 min for blunt-end cloning). SAP was then deactivated by incubation at 65 ^oC for 15 min.

2.4.13 **Plasmids**

A number of plasmids were used in this study for the construction on knock-out mutants of the *Burkholderia cepacia* Complex; these are listed in Table 2.5.

Table 2.5: Plasmids used in this study

Plasmid	Features	Reference
pKT230	Km^r Sm^r	Bagdasarian *et al.*, (1981)
pDT1558	Te^r Amp^r	Walter and Taylor (1989)
pPW601a	Amp^r	White et al., (1996)
pJQ200	Gm^r SacB	Quandt and Hynes (1993)
pGA1	Amp^r Km^r	This study
pHKT2	Tm^r GFP	Tomlin *et al.*, (2004)
pTrcHis-TOPO®	Amp^r	Invitrogen®
pTrcHis-Km	Amp^r Km^r	This study

2.4.14 **Ligations with T4 DNA Ligase**

Ligations with T4 DNA Ligase were performed as per the manufacturer's guidelines (Roche Diagnostics, Germany) in volumes of 30 µl. T4 DNA ligase (1 U) was incubated with <1 µg of DNA in 1 x ligation buffer (66 mM Tris-HCl; 5 mM $MgCl_2$; 1 mM dithiothreitol; 1 mM ATP; pH 7.5) at 4 ^oC (blunt ends) or 15-25 ^oC (sticky ends) for 16 h.

2.4.15 **Transformation of Competent HB101 *E. coli* (Promega Ltd, USA)**

Between 1 and 50 ng of DNA were added to 100 µl of cells and incubated on ice for 10 min. Cells were then heat-shocked at 42 ^oC for 45 sec and incubated on ice for 2 min. SOC broth (900 µl) was then added and the mixture incubated at 37 ^oC for 1 h. Aliquots of the undiluted and a 1:10 dilution (prepared using SOC Broth) was then spread onto appropriate selective media agar plates.

2.4.16 **One Shot® TOP10 *E. coli***

Between 1 and 5 µl of DNA (10 pg to 100 ng) was transferred into a vial of thawed Invitrogen™ One Shot® TOP10 competent *E. coli* cells (Life Technologies, UK) and mixed gently. Cells were incubated on ice for 30 min and subsequently heat-shocked for 30 sec at 42 °C without shaking. There followed an incubation on ice for 2 min prior to the addition of 250 µl of pre-warmed SOC medium. This was then incubated at 37 ^{o}C for 1 h at 225 rpm in a Stuart shaking incubator (Jencons, UK). The transformation mixture, < 200 µl, was plated on LB agar plates containing the appropriate antibiotic plus IPTG and X-gal. The plates were then incubated at 37 °C overnight.

2.4.17 **Filter Transformations**

Overnight cultures (20 µl), centrifuged and resuspended in 0.7 % NaCl, were spotted onto 0.22 µm nitrocellulose membranes (Millipore, UK) and placed onto pre-warmed LB agar plates. Plasmid DNA (0.5 µg) was spotted in the same positions and a control spot where no cells had been placed. Plates were incubated at 37 ^{o}C, overnight in a static incubator (Jencons, UK). Following incubation, filters were placed in 50 ml tubes and washed in 1 ml buffer (0.9 % NaCl, 15 mM $MgCl_2$, 1 mM $CaCl_2$ and 100 µg DNaseI). Washes were incubated at 37 ^{o}C for 10 min from which 100 µl was plated onto appropriate selective media.

2.4.18 **Polymerase Chain Reaction (PCR)**

Specific primer details and amplicon sizes are given in Table 2.6 with reagent concentrations and reaction conditions in Table 2.7. Standard end-point PCR was done in volumes of 50 µl in thin-walled 0.2 ml PCR tubes (Bio-Rad

Laboratories Ltd, UK). Amplifications were done in a GeneAmp® PCR System 9700 (Applied Biosystems, UK). The basic PCR protocol consisted of an initial 5 min denaturation step at 94 °C, followed by > 40 cycles of denaturation (94°C) for 1 min, primer annealing (variable temperature) for 30 – 60 sec, and an extension step (72 °C) for 30 sec – 4.5 min. A final single extension step of 72 °C for 10 min was also included. Where high fidelity amplification was required, the thermostable, proofreading (3' – 5' exonuclease activity) Expand High Fidelity PCR System (Roche Applied Science, Germany) was used in place of *Taq* DNA polymerase.

Real-time PCR reactions were performed in volumes of 20 or 25 µl using with LightCycler™ (Roche Applied Sciences, Germany), SmartCylcer™ (Cepheid, France) or Rotor-Gene™ (Qiagen, UK) instruments. The basic PCR protocol consisted of a 30 sec denaturation at 94 ^{o}C, followed by >40 cycles of denaturation (94 ^{o}C) for 5 sec and combined primer annealing and extension (variable temperature) for 30 - 60 sec. Unless stated otherwise TaqMan® reporting chemistry utilised 3'- carboxyfluorescein (FAM) and 5' Black Hole Quencher-1 (BHQ-1) dyes (ATDBio Ltd, UK).

2.4.19 **PCR Optimisation**

2.4.19.1 Primers

Primer optimisation was performed using a matrix of concentrations at 50, 300 and 900 nM equating to 1.25, 7.5 and 22.5 µM final concentrations in a 25 µl PCR reaction. PCR reactions consisted of 2 µl template DNA at 1 ng μl^{-1}, 1 µl each primer, 2 µl SYBR Green I (Roche Diagnostics, Germany), 2 µl 10 x SYBR

Green I mastermix (Roche Diagnostics, Germany), 0.8 µl $MgCl_2$ (25 mM) and 10.75 µl molecular grade dH_2O. Optimal primer concentration pairings were established by comparing cycle threshold (*ct*) values and end-point fluorescence on a LightCycler (Roche Diagnostics, Germany) or RotorGene™ real-time PCR instrument (Qiagen, UK).

2.4.19.2 Magnesium and dNTP Concentration

Magnesium optimisation was performed by altering the added volume of a 25 mM $MgCl_2$ stock solution (Roche Diagnostics, Germany). Volumes, and resulting concentration in a 20 µl PCR reaction were 0.8 µl (1 mM), 1.2 µl (2 mM) 1.6 µl (3 mM), 2.4 µl (4 mM) and 3.2 µl (5 mM). The concentration of dNTPs was optimised in parallel using a 200 µM stock solution (Roche Diagnostics, Germany). Concentrations in 20 µl PCR reactions were adjusted to 200 µM, 250 µM, 300 µM, 350 µM and 400 µM. Optimal concentrations of $MgCl_2$ and dNTPs were established by comparing cycle threshold (*ct*) values and end-point fluorescence on a LightCycler (Roche Diagnostics, Germany) or RotorGene™ real-time PCR instrument (Qiagen, UK).

2.4.19.3 TaqMan® Real-time PCR Probe

The optimisation of TaqMan® real-time PCR probes was done by evaluating concentrations of 50 nM, 100 nM, 150 nM, 200 nM and 250 nM. This gave final concentrations of 1.25, 2.5, 3.75, 5 and 6.25 µM in 25 µl PCR reactions. PCR reactions consisted of 2 µl template DNA (1 ng μl^{-1}), 1 µl each primer, 1 µl TaqMan® probe, 2 µl 10 x PCR mastermix (500 mM Tris-HCl, pH 8.8, 4 mM $MgCl_2$, 0.2 mM each dNTP, 0.04 units ul^{-1} Taq polymerase, 250 ng ul^{-1} BSA and 8 % glycerol) and 19 µl molecular biology grade dH_2O. Optimal primer

concentration pairings were established by comparing cycle threshold (ct) values and end-point fluorescence on a LightCycler (Roche Diagnostics, Germany) or RotorGene™ real-time PCR instrument (Qiagen, UK).

2.4.19.4 Taq DNA Polymerase

The concentration of *Taq* DNA Polymerase (Roche Diagnostics, Germany) was adjusted to include the following units per PCR reaction: 0.8, 1, 1.25, 2.5 and 4U in 20 µl PCR reactions with the optimised conditions as determined above. The optimal concentration was established by comparing cycle threshold (*ct*) values and end-point fluorescence on a LightCycler (Roche Diagnostics, Germany) or RotorGene™ real-time PCR instrument (Qiagen, UK).

Primer	Sequence (5' – 3')	Features	Amplicon (bp)	Target	Chapter	Reference
KmF KmR-SphI	TCTAGATATGAGCCATATTCAACGGGA AAAAACTCATCGAGCATCAAATGGCATGC	XbaI SphI	823	Km' pKT230 (kanamycin resistance gene)	4	This Study
1-2RRP3 (1-2FFP2	CCAACCTGAACTATTCGG TGCGGATCGTGGCGC		228	*zmpA* (zinc metalloprotease gene from Bcc)	4	Gingues *et al.* (2005)
ZMPA2F ZMPA2R	TGAAGTTCTACGCGAACAACCC GAYAGCAGRTAGAAGAAGCGGT	Y = C/T R = G/A	208	*zmpA* (zinc metalloprotease gene from Bcc)	4	This Study
Zhom1F Zhom1R Zhom2F Zhom2R	CTCGACTGCTGCCCATCAC CAAAAGCGCTGAGACTCGC TCTGGTACCGGACGCTGAC TTCGGATAGCTCGAGTTGGC		50	*zmpA*(5' and 3' regions of zinc metalloprotease gene from Bcc)	4	This Study
Z2F Z2R	ACAAGGCACTGCAACTGATCC GCTCGAGTTGGCGTTCAGG		492	*zmpA* (zinc metalloprotease gene from Bcc)	4	This Study
TelF TelR	TCTAGAACGAGGGATAGAAAGTTTAGC CCGCTTATGGCTCTGCCGTACG	XbaI SphI	2951	*kilAtelAB* pDT1558 (potassium tellurite resistance cassette)	4	This Study
GC-EcoR12 GC-AcI1	GAATTCGAATTCTTCGGATAGCTCGAGTTGGC AACGTTAACGTTGGCGCGCCGATATCCTCGAC	EcoR1 AcI1	1041	*zmpA* k/o construct pGA1	4	This Study
KmC 6F KmC R	TTTTGCGCCGACATCATAACG GGTACCGCATGCAAAAACTCATCGAGCATCA	Acc651	1205	k/o construct pTrcHis-Km-77	4	This Study
GfpF GfpR	ATGTGCTGATGCGATACCGA GACTAGGATGGACTCCATCAT		700	Gfp pHKT2	4	This study
BAPA2F BAPA2R BAPA-P	ACTAGTGAAGTACATGGAAATGCAGAA TGCGACCGTACTTGAATTCG CATGCGTCGTTCTTTGATATTGGTGGGAG		100	*Pag* (protective antigen of *B. anthracis*)	3	This study

CAPS2F CAPS2R CAPS-P	CATGGCCGAGAAAATCATGA CTGAATACGCGAGGTCGATTC CCATTTCGGTGATCGCTTCGCC		65	*wcbM* (capsular polysaccharide synthesis gene from *B. pseudomallei/mallei*)	3	This study
RA-F2 RA-R2 RA-Prb	GCTCGATCAAGAAGAACGACG GGCGGCGACACCTTGTT TCGGCAACGAAACCCGCGTG		76	*recA* (recombinase A gene from *B. cepacia*)	3	This Study
PerfF PerfR PerfT1	AGTGCAAGTGTTAATCGTTATCA TGGACAGATCATTTTCTAAGAT CATCAACTAAAGTCTACGCTTGGG		247	*Plc* (phospholipase C gene from *C. perfringenz)*	3	Chatwell, N (DSTL)
BGCSP-F BGCSP-R BGCSP-P	AAGGTTTCGGGCTCATCGA TTGAAGCCTTCGCCTTGAAT AGGTCAAGACGATGTATTCGTTCATTTCTCTG		79	*cspB* (cold shock protein gene of *B. globigii*)	3	Campbell, G (DSTL)
F1 R1	GATTTGTAAGGCGCTTATTTGTGCTACG CATGTAGTCATCTGTTCCAGCATCTTTTTC		1059	*Plc* (phospholipase C gene from *C. perfringenz)*	3	This Study
16SV1-3454F (Based on 8F) 16SV1-3454R (Based on 534R)	**CGTATCGCCTCCCTCGCGCCATCAG**AGAGTT TGATCCTGGCTCAG **CTATGCGCCTTGCCAGCCCGCTCAG**TIACCGI IICTICTGGCAC	*E. coli* position 8 to 533	575	16S rRNA (partial) gene (V1-3)	5/6	Baker *et al.*, (2003)
16SV4-6454F1 (Based on 1099F) 16SV4-6454R (Based on 1114R)	**CGTATCGCCTCCCTCGCGCCATCAG**KAGGKI GCRAGCGTTRWYCG **CTATGCGCCTTGCCAGCCCGCTCAG**CCCAAC ATYTCACRACACGA	*E. coli* position 539 to 1111	622	16S rRNA (partial) gene (V4-6)	5	This Study

Table 2.6: Primers Used in this Study. Degenerate alphabet: R (A or G, D (A, G or T), N (A, C, G or T), Y (C or T), S (C or G), W (A or T). Underlined regions indicate restriction enzyme sites.

Table 7.7 Amplification Conditions for PCR Primers Used in this Study. S: Standard end-point PCR.

Primer Pair	$MgCl_2$ (mM) & Primer (pmoles)	PCR Program Type	Annealing Temp.	Cycles	Notes
TelF TelR	4 10	S	64 °C	35	
KmF KmR-Sph1	4 10	S	66 °C	35	
GC-AscI1 GC-EcoR12	4 10	S	60 °C	35	
KmC 6F KmC R	4 10	S	55 °C	35	
Gfp	2 10	S	53 °C	35	
BAPA2F BAPA2R BAPA2P	4 22.5 (F) 7.5(R) 2 (P)	TaqMan®	60 °C	40	
CAPS2F CAPS2R CAPSP	5 7.5 (F) 22.5 (R) 3.75 (P)	TaqMan®	66 °C	40	1M Betaine
BGCSP-F BGCSP-R BGCSP-P	4 22.5 (F) 7.5 (R) 2 (P)	TaqMan®	60 °C	40	
RA-F2 RA-R2 RA-P	5 22.5 (F) 7.5 (R) 0.1(P)	TaqMan®	62 °C	40	
PerfF PerfR PerfT	4 5	TaqMan®	55 °C	40	
F1 R1	1.5 10	Long-PCR	63 °C	35	
ZMPA2F ZMPA2R	4 10	S	59 °C	35	
ZMPA F (1-2RRP3) ZMPA R (1-2FFP2)	4 10	S	56 °C	30	
Zhom1F Zhom1R Zhom2F Zhom2R	4 10	TaqMan®	60 °C	40	
Z2F Z2R	1.5 10	S	51 °C	35	
16SV1-3454F 16SV1-3454R	10 1.5	S	42 °C	35	
16SV4-6454F 16SV4-6454R	10 1.25	S	40 °C	35	

2.5 **454™ Sequencing**

All processes related to 454™ sequencing, including shotgun and amplicon library preparations, emulsion-PCR and sequencing were carried out according to the manufacturer's instructions. Briefly, for shotgun sequencing, 500 ng of genomic material was fragmented through nebulisation to yield a fragment libaray with an average size of 700 bp. Following small fragment removal using Agencourt® AMPure® XP size selection beads (Beckman Coulter Inc, UK) the fragment ends were end-repaired prior to the ligation of 454™ adapters. Libraries were then quantified using a Qubit® fluorometer using Picogreen® Quant-iT™ intercalaters (Invitrogen®, Life Technologies®, UK). Libraries were then captured onto beads and isolated in emulsion PCR (emPCR) droplets using a Qiagen® Tissue-Lyser. Following emPCR on a Perkin Elmer 9700 PCR machine, the amplified, bead-bound libraries were recoved from the emulsion, quantified using a Z2 Beckman Coulter Counter® (Beckman Coulter Inc, UK) and loaded onto the PicoTitrePlate (PTP) for sequencing. For amplicon sequencing approaches the procedure only differed in so much that the libraries did not undergo nebulisation. Instead, following quantification, libraries were used directly in emPCR and sequencing.

Multiplex Identifier Adaptors (MIDs), taken from Roche Applied Sciences technical bulletin TCB-10010, and used for bar-coding of PCR amplicons for 454 Sequencing™ (as shown in Figure 2.1) are listed in Table 2.8.

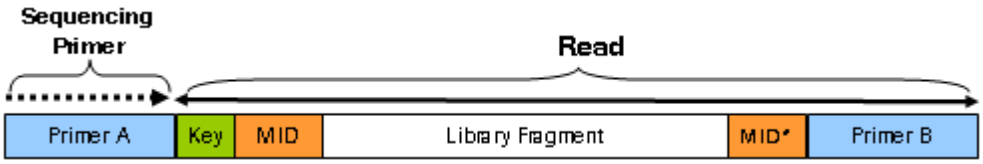

Figure 2.1: Depiction of the location of Multiplex Identifier Adaptors (MIDs) in relation to amplicons generated for 454 Sequencing™. (Image reproduced from 454™ technical bulletin TCB-10010 [Roche Applied Sciences, Germany]).

Table 2.8: Multiplex Identifier Adaptor (MID) Sequences used in generation of 454 Sequencing™ amplicon libraries

MID	Sequence (5' – 3')	MID	Sequence (5' – 3')
1	ACGAGTGCGT	2	ACGCTCGACA
3	AGACGCACTC	4	AGCACTGTAG
5	ATCAGACACG	6	ATATCGCGAG
7	CGTGTCTCTA	8	CTCGCGTGTC
10	TCTCTATGCG	11	TGATACGTCT
13	CATAGTAGTG	14	CGAGAGATAC
15	ATACGACGTA	16	TCACGTACTA
17	CGTCTAGTAC	18	TCTACGTAGC
19	TGTACTACTC	20	ACGACTACAG
21	CGTAGACTAG	22	TACGAGTATG

Chapter Three

DNA Persistence in Soil

3.1 **Introduction**

For the reliable detection of pathogenic microorganisms in soil, which may have been used in the contect of biological warfare / bioterrorism, the persistence of nucleic acid signatures is of paramount importance. An understanding as to the dynamics of this persistence is required in order to discriminate between the naturally occurring background and that which would indicate a nefarious release.

As has been shown, (Lorenz and Wackernagel, 1987; Ogram *et al.*, 1988; Khanna and Stotzky, 1992; Alvarez *et al.*, 1998; Crecchio and Stotzky 1998; Cai *et al.*, 2006), DNA persistence in terrestrial environments is affected by many factors including molecular conformation of the nucleic acids, the use of DNA as a nutrient source by microorganisms, the presence of degradative enzymes and the adsorption of DNA onto various mineral components. The recovery of nucleic acid material from archeaological / paleontological samples suggests that, under the right conditions, DNA can persist for extremely extended periods of time (DeSalle *et al.*, 1992). Recently, Allentoft *et al.*, (2012) calculated the half-life of mitochondrial DNA in bone to be 521 years.

Prolonged persistence of DNA released from dead or metabolically active microbes in soil was thought to be restricted by rapid degradation by soil nucleases which have been shown to be mainly of bacterial origin (Lorenz and Wackernagel, 1994; Blum *et al.*, 1997). However, extensive studies have shown protection afforded by clay minerals and other soil colloids to bound nucleic acids. For example, nucleic acids adsorb to clays, sand, humic acids and bulk

soil (Lorenz and Wackernagel, 1987; Ogram *et al.*, 1988; Khanna and Stotzky, 1992; Crecchio and Stotzky 1998). Bound nucleic acids are protected from nuclease degradation remaining available for both PCR amplification and as material for transformation of competent cells (Khanna and Stotzky 1992; Alvarez *et al.*, 1998; Cai *et al.*, 2006). For example, Crecchio and Stotzky (1998) demonstrated that total inactivation of humic-bound DNA required 100 times the concentration of DNase I than that required for unbound DNA. The impact of variations in clay mineralogy were highlighted by Cai *et al.*, (2006) who demonstrated that DNA adsorbed to montmorillonite, fine clays and inorganic clays were more protected from DNase I degradation than inorganic and coarse clays. The nature of the resistance to degradation may be a function of DNA adsorption modifying the electron distribution of the adsorbed DNA, thus making it more resistant to hydrolysis, or by physical separation of the DNA from nucleases (Khanna *et al.*, 1998; Demanèche *et al.*, 2001). Interestingly, some clay minerals, such as kaolinite and montmorillonite, appear to prevent PCR amplification of adsorbed DNA when DNA-clay complexes are used directly in the amplification process (Alvarez *et al.*, 1998; Peng *et al.*, 2007). In the case of clay minerals, binding is affected by the surrounding pH where protonation of certain chemical groups on the DNA molecule increase its affinity for the negatively charged surface of the mineral (Khanna and Stozky, 1992). Adsorption of DNA to sand is also affected by pH, salt concentration and salt valency (Lorenz and Wackernagel, 1987).

A limitation of these studies, from the viewpoint to determining extracellular DNA persistence in soil, is that they have, for the most part, concentrated on the

adsorption of nucleic acids to purified mineral components or isolated soil colloids. Where adsorption was investigated for bulk soil (Ogram *et al.*, 1988) no attempt was made to extrapolate to persistence. Where inferences have been made to the impact of these factors in non-sterile soil systems, nucleic acids are often introduced into the soil as mineral conjugates and thus do not accurately represent the initial dynamics of these interactions. In addition, no study to date has investigated the impact of contrasting soil mineralogy on DNA persistence with a view to investigating the true impact of *in-situ* colloid-based adsorption.

3.1.1 **Release and Persistence of Bacterium-derived DNA**

The release of DNA from bacterial cells can occur either as a consequence of cell death (Palmen and Hellingwerf, 1994; Palmen and Hellingwerf, 1997) or through active secretion. This latter phenomenon is well established with examples from *Acinetobacter*, *Alcaligenes*, *Azotobacter*, *Bacillus*, *Flavobacterium*, *Micrococcus*, *Neisseria* and *Pseudomonas* (Lorenz *et al.*, 1991; Lorenz and Wackernagel, 1994; Paget and Simonet, 1994; Yin and Stotzky, 1997; Dillard and Seifert, 2001; Matsui *et al.*, 2003; Hamilton *et al.*, 2005; Thomas and Nielsen, 2005). Whilst it is recognised that the latter genera secrete DNA *in vitro*, the importance in environmental settings remains to be determined (Nielsen *et al.*, 2007); although the positive contribution in terms of acting as a structural element in the formation of biofilms has been reported (Whitchurch *et al.*, 2002). In the context of cell death, autolysis leads to the release of numerous cytoplasmic contents including DNA. The quantity is reliant upon the conditions of growth preceding autolysis and whether it is triggered by abrupt changes in environmental conditions e.g. introduction of antibiotics (Freidlander, 1975). Nevertheless, DNA persistence is required for transformation and thus has been

the focus of much study, particularly in relation to the issue of genetically modified organisms (GMO). Examples of DNA persistence studies are given in Table 3.1. These studies demonstrate the large variation in observed stability of DNA in soil environments. However it should be noted that the majority of these studies used microcosms and did not tackle the issue of open-environment degradation kinetics (Nielsen *et al.,* 2007). Investigations into DNA persistence alone do not give the required information to determine the likely potential of its contribution to lateral gene exchange events. One must also determine the biological activity of the DNA within the environment. As an example Sikorski *et al.,* (1998) demonstrated that although plasmid-mediated transformants of *P. stuzeri* JM300 were detectable after five days incubation of pPSI1 in non-sterile sandy loam microcosms, the plasmid DNA had lost approximately 88 % of its transformation activity within the first 24 h.

Inocula	Microcosm	Detection Method	Persistence	Reference
Plasmid pUC8-ISP	Loamy Sand Clay Silty Clay	PCR and transformation of *E. coli*	PCR: 0.2% remaining at day 60 Transformation activity: 0.011% at day 60 PCR: 0.05% remaining at day 60 Transformation activity: 0.004% at day 40 PCR: 0.01% remaining at day 60 Transformation activity: 0.002% at day 10	Romanowski *et al.*, 1993
E. coli K12 CM1170 Plasmid	Non-sterile desiccated vs. non-desiccated	Dot-blot / Southern hybridisations and transformation of *E. coli* DH5a	Increased persistence in desiccated soils. Plasmids remained detectable and biologically active for 17 days after disappearance of *E. coli* counts	Brim *et al.*, 1994
Bacillus subtilis DNA	Sandy loam	Transformation	15 days	Gallori *et al.*, 1994
Acinetobacter sp. DNA	Silt loam, loamy sand	Transformation	30 minutes	Nielsen *et al.*, 1997
Pseudomonas stutzeri DNA	Loamy sand	Transformation	3 days	Sikorski et al., 1998
Cell lysates of Acinetobacter sp., Pseudomonas fluorescens, Burkholderia cepacia	Silt loam	Transformation of *Acinetobacter* sp. strain BD413	4-8 hours	Nielsen *et al.*, 2000
Killed Gaeumannomyces graminis var. tritici	Sodosol Soil Organic Matter Sand (solodised solanetz)	PCR	0.2% remaining at day 4 Undetectable DNA levels 2 weeks post-inoculation 8% remaining at day 4	Herdina *et al.*, 2004
Genomic DNA of baculovirus *Choristoneura fumiferana*	Forest litter	PCR	90 days	(England *et al.*, 2004)

Table 3.1: Studies of DNA Persistence in Soil. Listed above are eight examples of soil microcosm studies.

3.1.2 **Extraction of DNA from Soil**

An underpinning requirement for the accurate characterisation of DNA persistence in soil is the availability of appropriate extraction methods. Broadly, the efficient extraction of DNA from environmental samples is regarded as of paramount importance in the application of molecular methods to the study of microbial communities and/or pathogen detection (Rudi and Jakobsen, 2006). A selected method must enable the isolation of sufficient yields of DNA, in some instances at high molecular weight, with the simultaneous removal of potential inhibitors of downstream processes. Given the heterogeneity of microbial communities, as well as the samples from which they may be derived (e.g. soil, manure or sediment) there exists a plethora of methods within the literature from which to select. These can be broadly categorised as to whether the extraction occurs *ex* or *in situ*. In other words, cells are either first separated from the sample prior to lysis (*ex situ* or indirect extraction) or separated and then lysed within the matrix (*in situ* or direct extraction). The benefit of the former is that it enables the extraction of high molecular weight DNA of predominantly prokaryotic origin (Holben, 1994). In contrast *in situ* methods can give DNA of comparable purity but with an overall higher yield (Steffan *et al.*, 1988). Additionally, recent efforts to characterise the impact of indirect vs. direct extraction methods on accessing the soil metagenome (estimated at 10^{15} bp) through sequencing has revealed little difference between the methods for metabolic function and species detection (Delmont *et al.*, 2011).

Extraction methods may consist of physical (bead-beating) or chemical (SDS, lysozyme, or Proteinase K) lysis or indeed a combination of both. This is

followed by a purification process with the aim of concentrating the extracted DNA whilst removing inhibitors of potential downstream processes e.g. PCR, such as humic acids, the major constituent of soil organic matter (Tsai and Olson, 1992; Tebbe and Vhajen, 1993; Paul and Clark, 1996; Dong *et al.*, 2006). Previous efforts to elucidate the impact of the various stages of DNA extraction are well documented (Kresk and Wellington, 1999; Delmont *et al.*, 2011). Delmont *et al*., (2011), in particular, highlighted the challenges of obtaining a single extraction method for capturing soil diversity. The authors showed that only through the application of a combination of gradient methods, lysis stringency protocols and molecular weight fractionation was it possible to obtain a more genetically diverse representation of soil communities: an improvement of 80% in comparison to the best single method (Delmont *et al.*, 2011). It is clear then that alterations at each step will impact on purity, yield and representation of the microbial community from which the DNA is being obtained. On this basis, methods may be developed to broadly capture as much diversity as possible or to target specific applications, such as the detection of specific pathogens (Dineen *et al.*, 2010). There are many examples from the latter category where methods have been evaluated for the purpose of detecting BWAs where rapidity, cost and ease of use have been investigated alongside sensitivity and yield (Dineen *et al.*, 2010; Fitzpatrick *et al.*, 2010; Gulledge *et al.*, 2010).

3.2 **Aims**

The overall aim of this section is to determine if soil types with characteristics amenable to adsorption will enable prolonged DNA persistence. Specifically:

- determine the levels of DNA adsorption to soil mineral components.
- determine the protection of adsorbed DNA from nuclease degradation.
- establish suitable soil DNA extraction methods for PCR analysis of BWAs in soil.
- use real-time PCR analysis to determine the persistence of free DNA, DNA from dead intact or lysed bacterial cells in soil.

The hypothesis here is that clay-based soils will result in greater levels of DNA persistence and thus enable extended periods of time within which pathogenic organisms can be identified from the soil matrix.

3.3 **Impact of Mineral Concentration on Adsorption of DNA**

3.3.1 **Comparison of Minerals**

Comparative adsorption levels for kaolinite, bentonite and SiO_2 (homo-ionic to Ca^{2+}; Section 2.1.1) were determined using Calf Thymus DNA according to the method of Khanna and Stozky (1992). Clay minerals were chosen based on their contrasting physical properties, such as cation exchange capacity, that are known to impact on the levels of DNA adsorption (Cai *et al.*, 2006). Quantitative analysis of the reduction in DNA concentration using 260/280 nm adsorption was done according to Section 2.1.2. Results are shown in Figures 3.1 – 3.4.

For all three minerals tested, maximum adsorption was observed quickly (within the first minute for SiO_2 and 30 min for each clay mineral) with little increase in the overall DNA levels adsorbed over the remainder of the time-course. Adsorption levels were in the order of Bentonite > Kaolinite > SiO_2. Measurable DNA levels in supernatants were completely diminished for Ca^{2+}-Bentonite by 30 min at the maximum clay quantity of 10 mg. In contrast, Ca^{2+}-Kaolinite adsorbed approximately 60 % of the total DNA by 30 min for the same clay quantity. Finally, Ca^{2+}-SiO_2 DNA adsorption reached approximately 30 % within the first minute. No clear trend for the control sample of SiO_2 without the presence of a cation was apparent. Observed *p* values, calculated by ANOVA using MiniTab v.14, were <0.05 indicating that mineral concentration and time were statistically significant variables impacting on DNA adsorption. Adjusted R^2 values were used to describe the goodness-of-fit for each clay concentration vs. time models. These were 99.36 % (Ca^{2+}-Bentonite), 97.58 % (Ca^{2+}-Kaolinite), 82.02 % (SiO_2 no cation) and 87.79 % (Ca^{2+}-SiO_2). It is clear that

variability within the observed adsorption levels for SiO_2 was greater than for the two clay minerals examined. However this did not affect the overall statistically significant trend of increased DNA adsorption over time with increasing mineral concentration. Whilst there was significant difference in the adsorption levels for each concentration of SiO_2 over time, there was no clear trend.

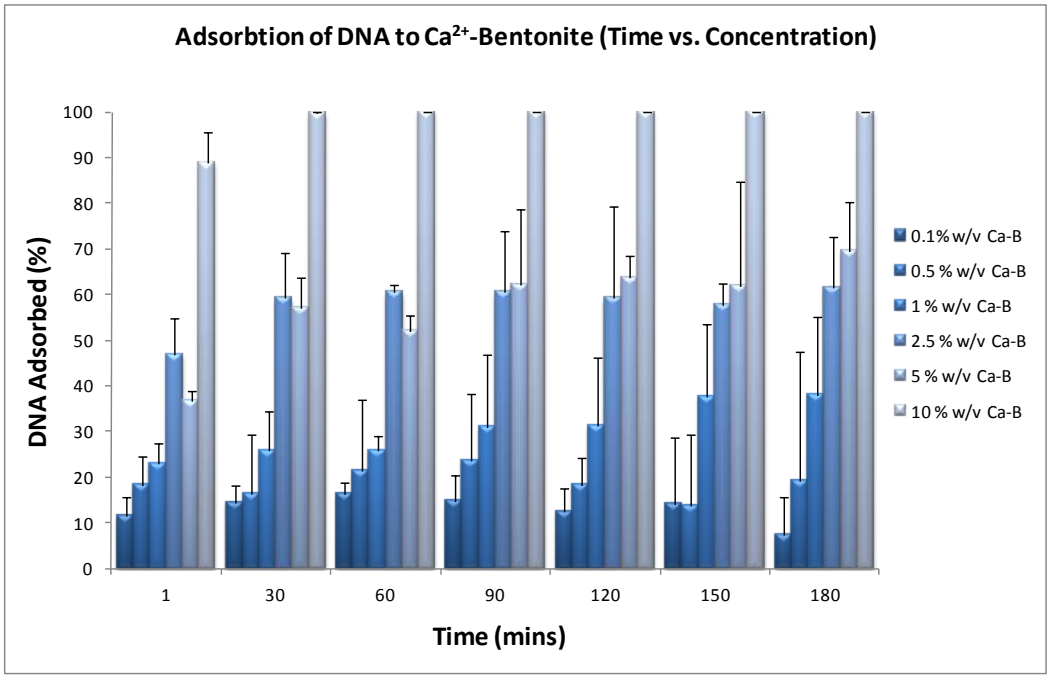

Figure 3.1: Effect of concentration of homo-ionic bentonite and time on DNA adsorption. The level of adsorption of DNA (~5 µg) was determined for increasing concentrations (0.1 % to 10 % w/v) of homo-ionic bentonite over a 180 min time course. Error bars show standard deviation of three replicates.

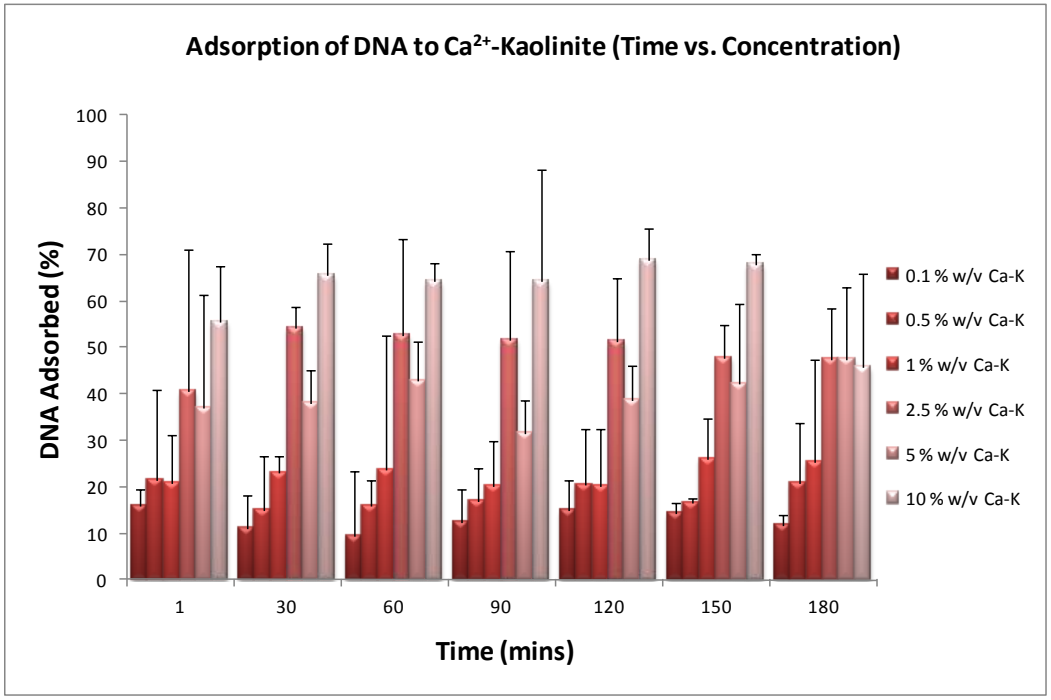

Figure 3.2: Effect of concentration of homo-ionic kaolinite and time on DNA adsorption. The level of adsorption of DNA (~5 µg) was determined for increasing concentrations (0.1 % to 10 % w/v) of homo-ionic kaolinite over a 180 min time course. Error bars show standard deviation of three replicates.

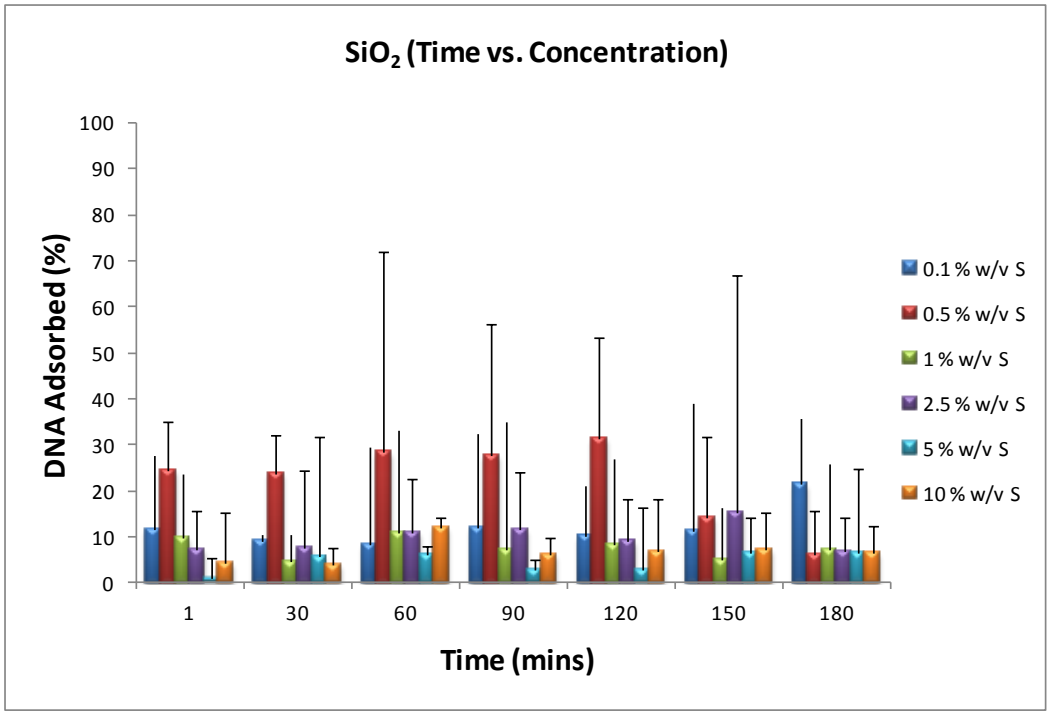

Figure 3.3: Effect of concentration and time on DNA adsorption to SiO_2. The level of adsorption of DNA (~5 µg) was determined for increasing concentrations (0.1 % to 10 % w/v) of SiO_2 over a 180 min time course. Error bars show standard deviation of three replicates.

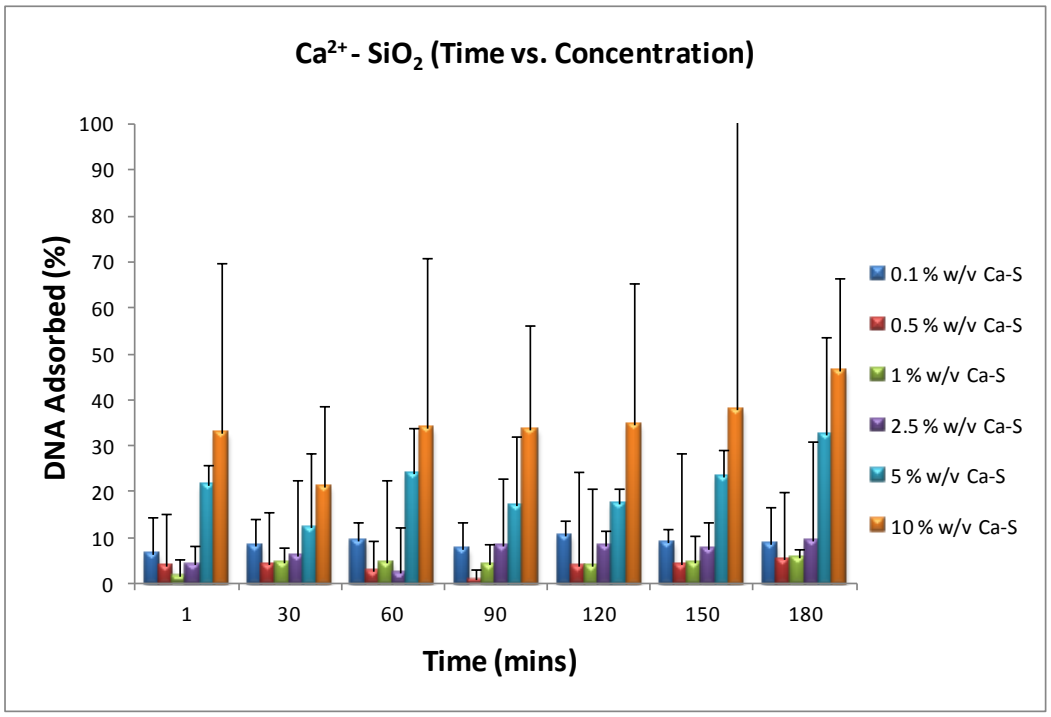

Figure 3.4: Effect of concentration and time on DNA adsorption to homo-ionic SiO_2. The level of adsorption of DNA (~5 µg) was determined for increasing concentrations (0.1 % to 10 % w/v) of homo-ionic silicon dioxide over a 180 min time course. Error bars show standard deviation of three replicates.

3.3.2 **Effect of Cation on DNA Adsorption**

The effect of altering the available cation was examined for its effect on DNA adsorption. Here, 5 mg of mineral (Bentonite, Kaolinite or SiO_2) was made homo-ionic to either Ca^{2+}, Na^{+} or Mg^{2+} according to the method outlined in Section 2.1.2. DNA adsorption was determined across a 180 min time course as previously outlined. Results are shown in Figures 3.5-3.7.

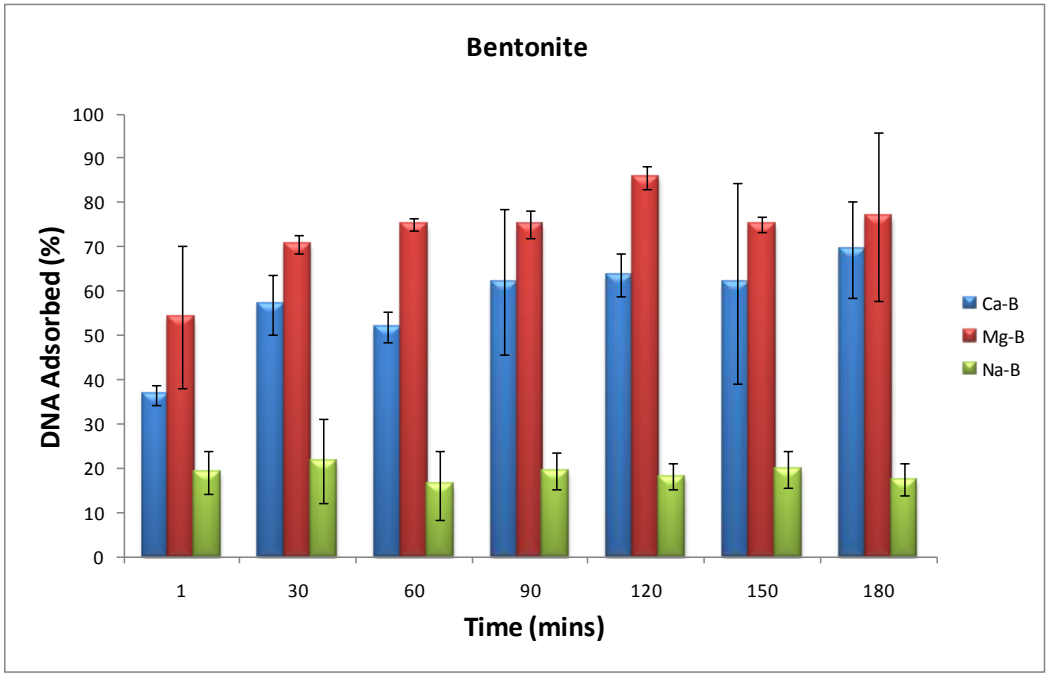

Figure 3.5: Effect of available cation on levels of DNA adsorption to bentonite. Levels of DNA adsorption were determined in the presence of Ca^{2+}, Mg^{2+} and Na^{+}. Adsorption was determined for 5 % w/v of clay mineral over a 180 min time course. Error bars show standard deviation of three replicates.

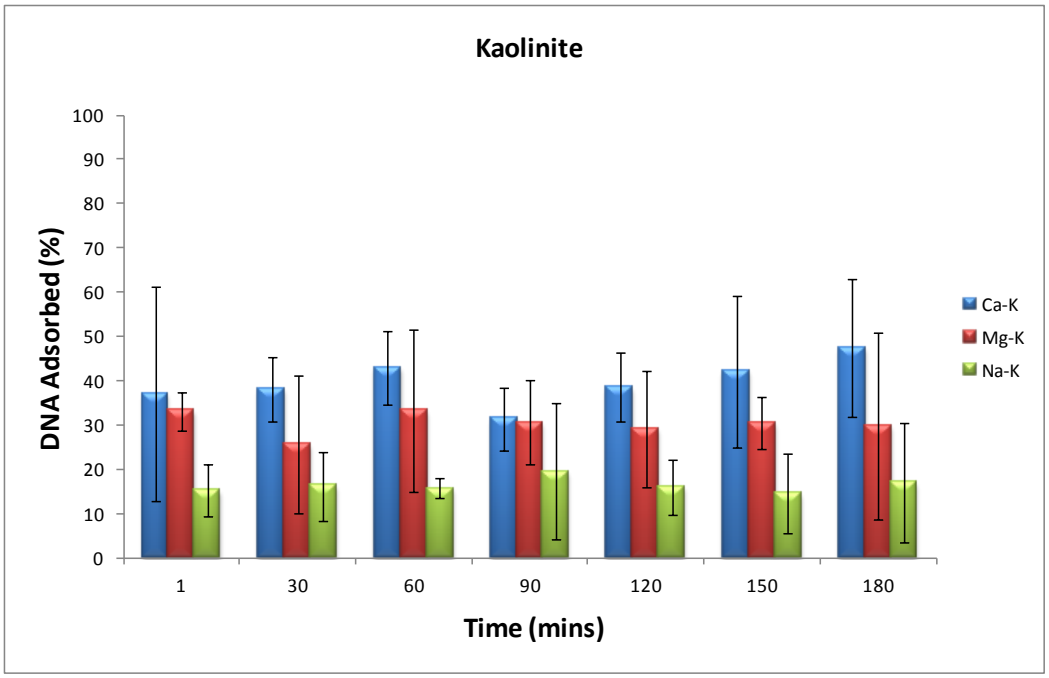

Figure 3.6: Effect of available cation on levels of DNA adsorption to kaolinite. Levels of DNA adsorption were determined in the presence of Ca^{2+}, Mg^{2+} and Na^{+}. Adsorption was determined for 5 % w/v of clay mineral over a 180 min time course. Error bars show standard deviation of three replicates.

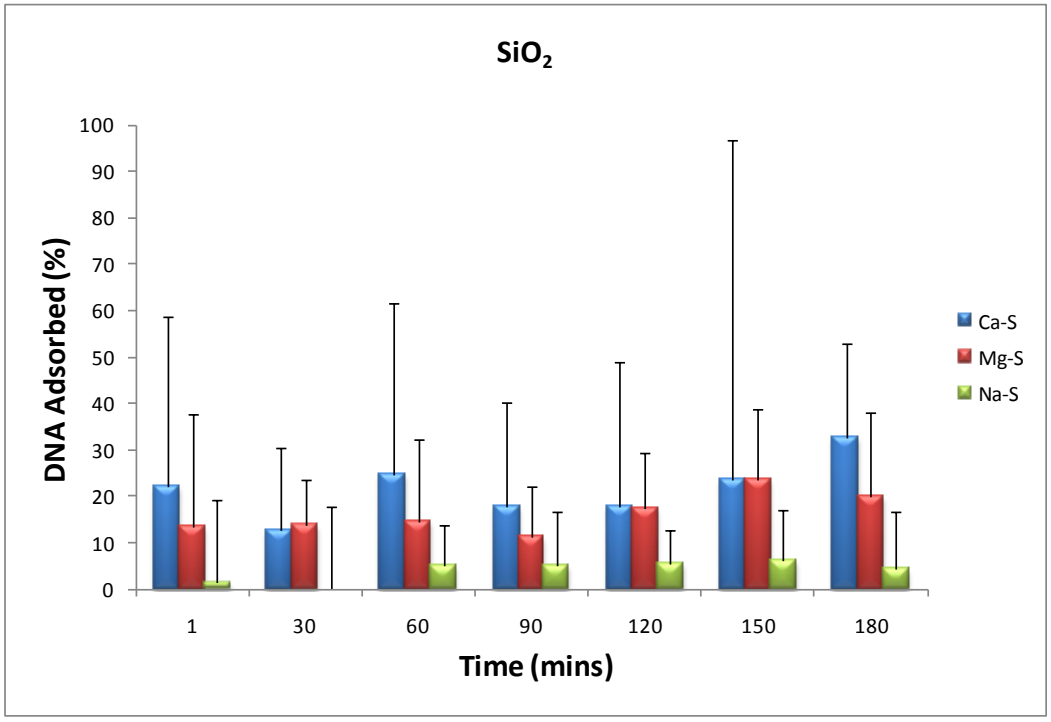

Figure 3.7: Effect of available cation on levels of DNA adsorption to silicon dioxide. Levels of DNA adsorption were determined in the presence of Ca^{2+}, Mg^{2+} and Na^{+}. Adsorption was determined for 5 % w/v of clay mineral over a 180 min time course. Error bars show standard deviation of three replicates.

For Bentonite, Ca^{2+} and Mg^{2+} enabled substantially increased levels of adsorption (~60 % and 70 % respectively) in comparison to Na^{+} where observed levels were approximately 20 % (Figure 3.5). Whilst Na^{+} still generated lower levels of adsorption for Kaolinite in comparison to Ca^{2+} and Mg^{2+} (Figure 3.6), there was a reversal with Ca^{2+} allowing higher levels of DNA adsorption. A similar trend to Kaolinite was observed for SiO_2 however the variability observed was higher for Ca^{2+} than Mg^{2+} or Na^{+}.

3.3.3 **Effect of pH on DNA Adsorption**

The effects of pH on DNA adsorption are shown in Figure 3.8. Here, 25 μg of Calf Thymus DNA (Sigma-Aldrich, UK) was incubated with 0.5 % w/v Kaolinite, Bentonite or SiO_2 for 180 min in DNA Buffer (10 mM Tris-HCl, 0.1 mM EDTA, 4 mM NaCl) adjusted to pH 2, 4, 6, 8, 10 or 12. Following

incubation at 24 ^{o}C for 180 min the mineral suspensions were centrifuged and DNA adsorption was determined by analysing suspension supernatants using 260/280 nm adsorption as described previously.

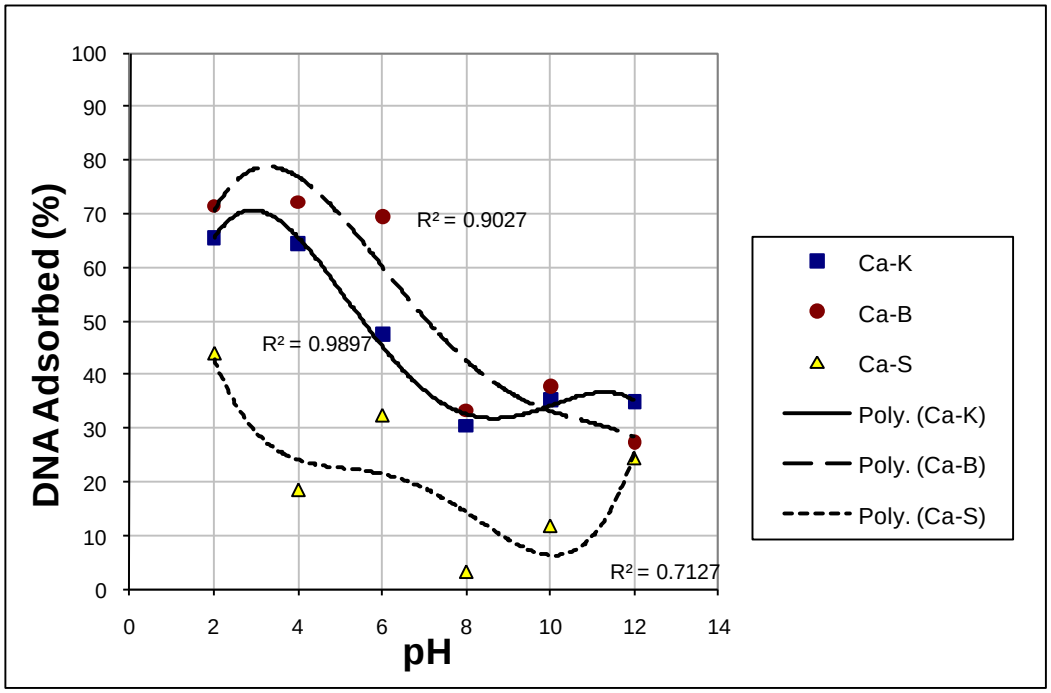

Figure 3.8: Effect of pH on DNA adsorption to homo-ionic Bentonite, Kaolinite and SiO_2. Homo-ionic mineral (0.5 % w/v) was exposed to 25 µg of calf thymus DNA for 180 min in DNA Buffer (10mM Tris-HCl, 0.1mM EDTA, 4mM NaCl pH 6.5) adjusted to pH 2, 4, 6, 8, 10 or 12. Polynomial trendlines with R^2 value for each data set are shown.

For Ca^{2+}-Bentonite and Ca^{2+}-Kaolinte a clear trend of increased DNA adsorption at lower pHs was observed (Figure 3.8). No linear trend for SiO_2 was observed with changing pH. However an R^2 of 0.7127 was achieved using polynomial trend-line analysis.

3.3.4 **Release of Adsorbed DNA**

Kaolinite or Bentonite (10 mg) homo-ionic to Ca^{2+} were incubated with 100 µl Calf Thymus DNA (final concentration of 458.3 ng μl^{-1}). The samples were incubated at 24 ^{o}C with shaking for 180 min followed by centrifugation at 23,000

g for 15 min. The supernatant was removed and 50 µl dH_2O added to re-suspend the pellets. The samples were then incubated at 24 ^{o}C with shaking for 5 or 10 min. At both time points the samples were centrifuged and the supernatants analysed by 260/280 absorbance (Section 2.4.9.1). Results are shown in Figure 3.9.

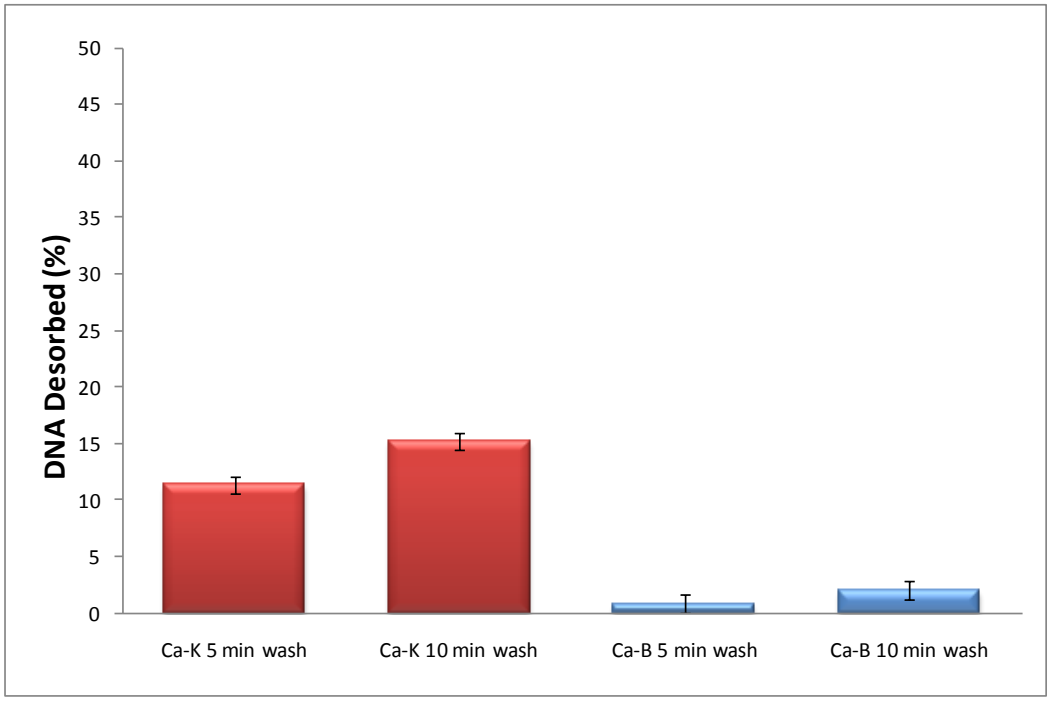

Figure 3.9: Release of DNA after adsorption to homo-ionic bentonite and kaolinte. Homo-ionic (Ca^{2+}) clay mineral (10 % w/v were incubated with 45 µg calf thymus DNA for 180 min. Post-incubation washes of 5 or 10 min were performed in dH_2O followed by centrifugation to remove clay mineral and adsorbed DNA. Levels of released DNA were determined by 260/280 nm analysis of the post-wash supernatant. Error bars show standard deviation of three replicates.

As seen in Figure 3.9, approximately 25 % of adsorbed DNA was released from Kaolinite after washing for 10 min. In comparison release of adsorbed DNA from Bentonite was lower with <5 % released after 10 min.

3.3.5 **Protection of Mineral-adsorbed DNA from DNaseI Mediated Degradation**

Two approaches, both utilising PCR amplification of specific sequences, were used to determine the protection afforded to DNA by adsorption to minerals.

The first approach used a long-PCR assay targeting the phospholipase C (*plC*) gene of *Clostridium perfringens*. The second approach used a real-time TaqMan® PCR assay (kindly donated by N Chatwell, DSTL, UK) that targeted an internal region of *plC*. For the design of the long-PCR assay the *plC* sequence (1,365 bp) of *C. perfringens* (Accession number: M24904.1) was first obtained from GenBank. Primer Express v1.0 (Applied-Biosystems, UK) was then used to design primers that targeted the near full-length gene.

The sensitivity of these two assays were determined using a dilution series of *C. perfringens* genomic DNA (Sigma-Aldrich, UK). Primers and PCR conditions are detailed in Tables 2.6 and 2.7, Section 2.4.9. PCR products were analysed on a 1 % E-gel (Invitrogen, UK). Results are shown in Figure 3.10. Detection of *C. perfringens* DNA was possible to 1 pg μL^{-1}. With a genome size of 3,031,430 bp (Shimizu *et al.*, 2002) this equates to ~500 genome equivalents.

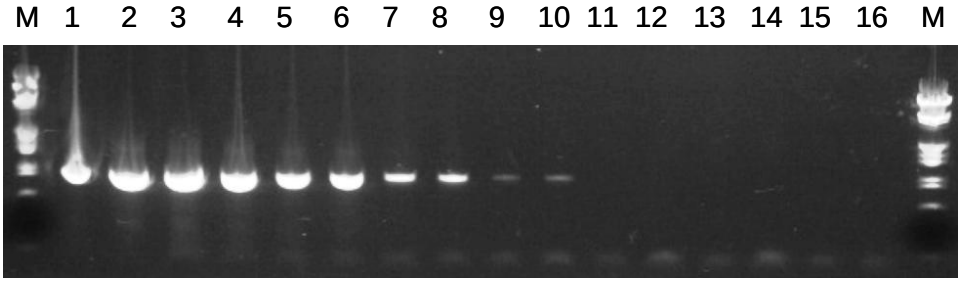

Figure 3.10: 2% Agarose gel demonstrating sensitivity of *plc* long-PCR assay. Lane identities are as follows: 1 and 2 10 ng; 3 and 4 1 ng; 5 and 6 100 pg; 7 and 8 10 pg; 9 and 10 1 pg; 11 and 12 100 fg; 13 and 14 10 fg; 15 and 16 no template control (ntc); M DNA molecular weight marker III.

The standard curve for the real-time TaqMan® PCR assay, and used for these analyses, is shown in Figure 3.11. The sensitivity of the assay was 1 pg μl^{-1}.

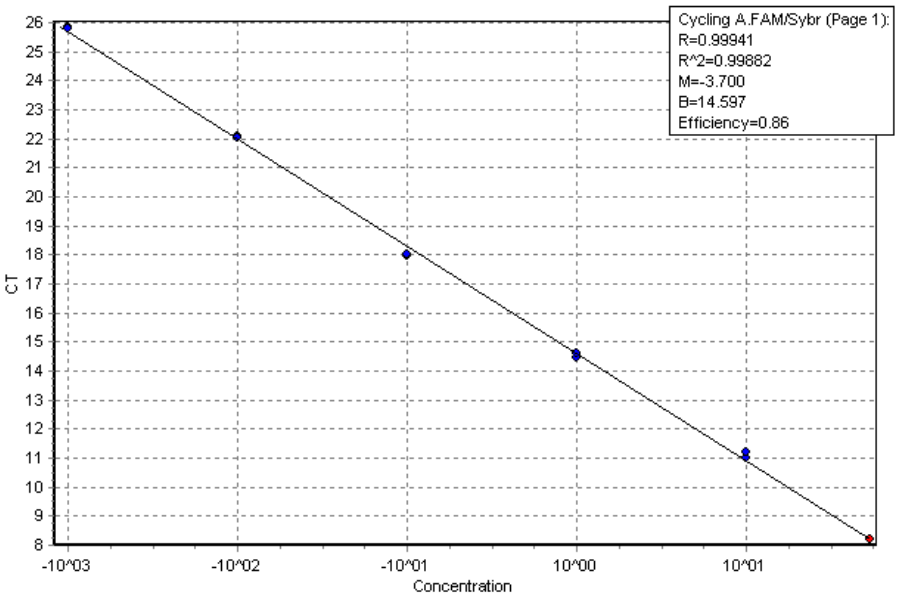

Figure 3.11: *Clostridium perfringens plc* real-time PCR Standard Curve. PCR reactions using 10 ng, 1 ng, 100 pg, 10 pg and 1 pg *C. perfringens* genomic DNA were performed in triplicate on a Rotor-Gene™ thermal cycler.

3.3.6 **Exposure of Mineral-adsorbed DNA to DNase I**

Mineral (0.1 % w/v) made homo-ionic to Ca^{2+} or treated with dH_2O, were incubated with 2 μg of *C. perfringens* genomic DNA for 1 h at 350 rpm on a Thermomixer (Eppendorf, UK) and then exposed to DNase I according to Section 2.1.3. At each time point (1, 10 and 20 min incubation with DNase I) 50 μl of 10 mM EDTA in 20 mM NaCl was added to stop the reaction. Control samples included a DNA only sample, DNA incubated with DNase I for 20 min in the abscence of mineral and two controls for the mineral with DNA in the abscence of DNase I. Samples were centrifuged, supernatants were removed and the pellets re-suspended in 150 μl of DNase I 'Stop Solution' (Table 2.3). Both the supernatant and pellet samples were analysed using real-time TaqMan® PCR analysis for the semi-quantification of amplifiable *C. perfringens* DNA. Results are shown in Figures 3.12 – 3.16 and show marked variability in both the levels

of protection from nuclease degradation and also the availability of adsorbed DNA for amplification.

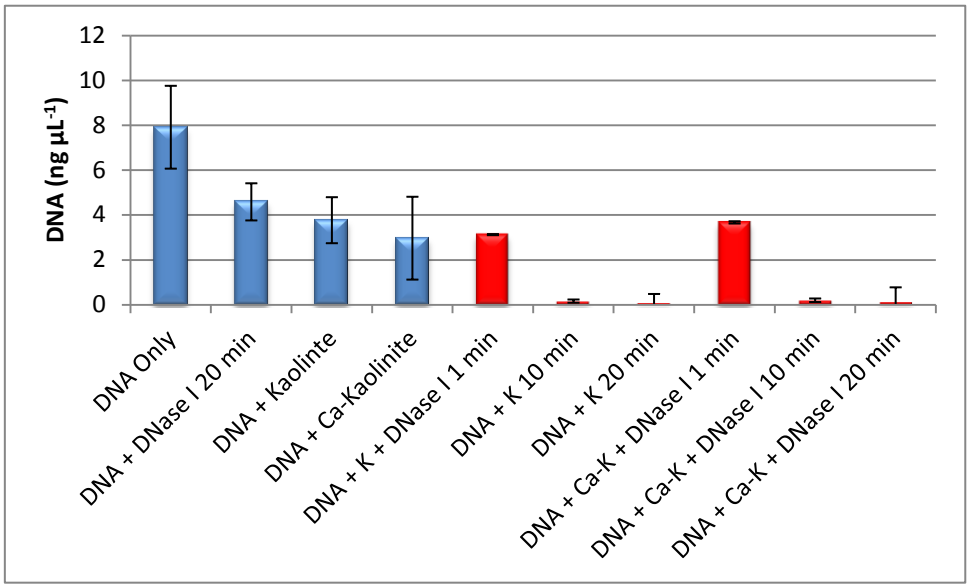

Figure 3.12: Results showing real-time TaqMan® PCR detection of *C. perfringens* genomic DNA in supernatants of Kaolinitic-DNA mixtures. Mineral concentrations (control and homo-ionic) were 0.1 % w/v for initial DNA adsorptions. Blue bars represent control samples with red being those of Kaolinitic-DNA mixes exposed to DNase I over a 20 min time course. Error bars show standard deviation of four PCR replicates.

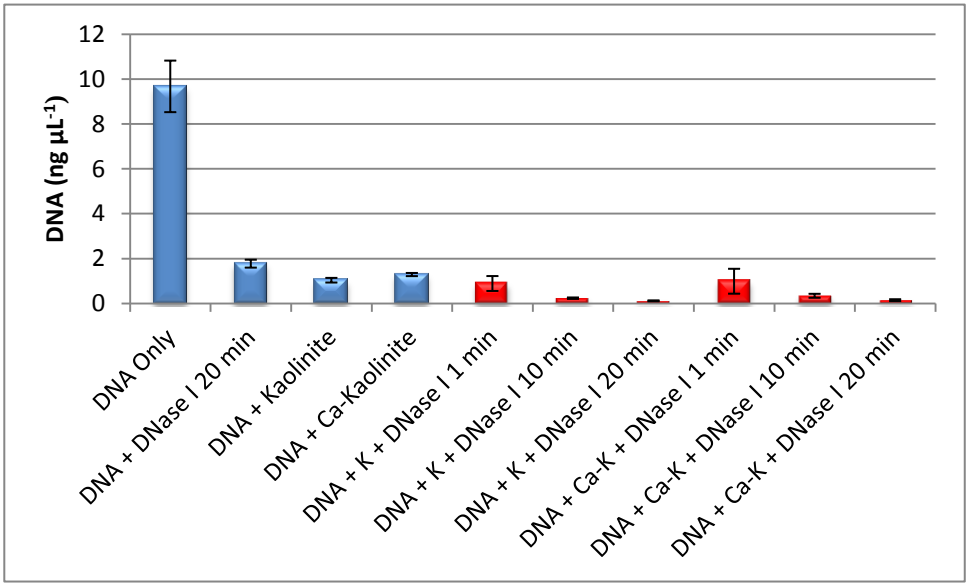

Figure 3.13: Results showing real-time TaqMan® PCR detection of *C. perfringens* genomic DNA in resuspended pellets of Kaolinitic-DNA mixtures. Mineral concentrations (control and homo-ionic) were 0.1 % w/v for initial DNA adsorptions. bars represent control samples with red being those of Kaolinitic-DNA mixes exposed to DNase I over a 20 min time course. Error bars show standard deviation of four PCR replicates.

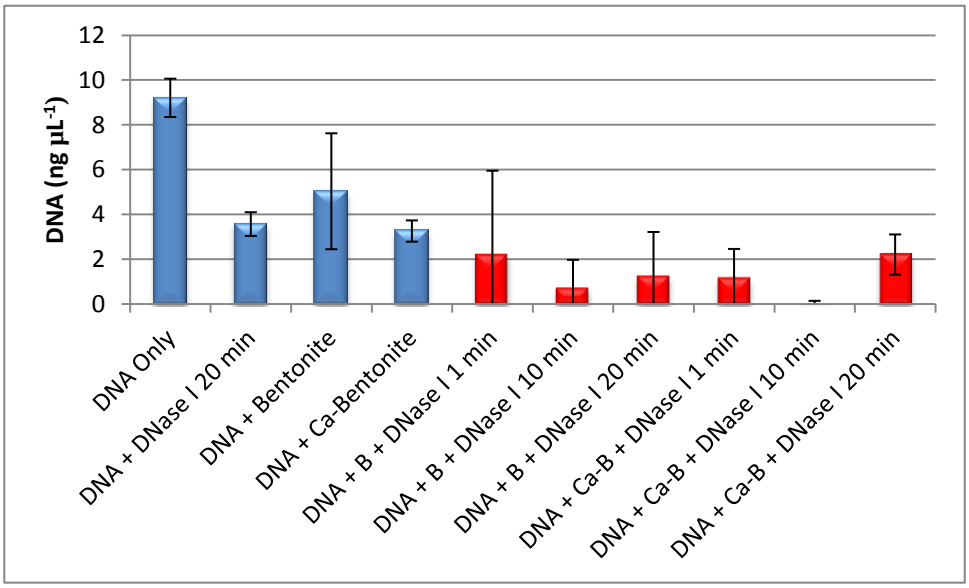

Figure 3.14: Results showing real-time TaqMan® PCR detection of *C. perfringens* genomic DNA in supernatants of Bentonite-DNA mixtures. Mineral concentrations (control and homo-ionic) were 0.1 % w/v for initial DNA adsorptions. Blue bars represent control samples with red being those of Bentonite-DNA mixes exposed to DNase I over a 20 min time course. Error bars show standard deviation of four PCR replicates.

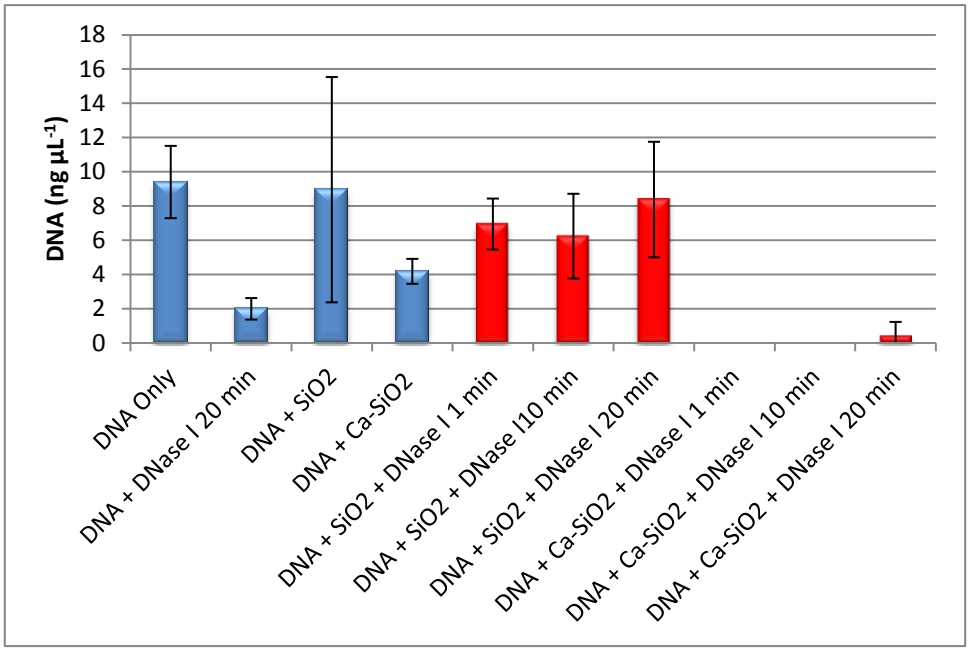

Figure 3.15: Results showing real-time TaqMan® PCR detection of *C. perfringens* genomic DNA in supernatants of SiO_2-DNA mixtures. Mineral concentrations (control and homo-ionic) were 0.1 % w/v for initial DNA adsorptions. Blue bars represent control samples with red being those of SiO_2-DNA mixes exposed to DNase I over a 20 min time course. Error bars show standard deviation of four PCR replicates.

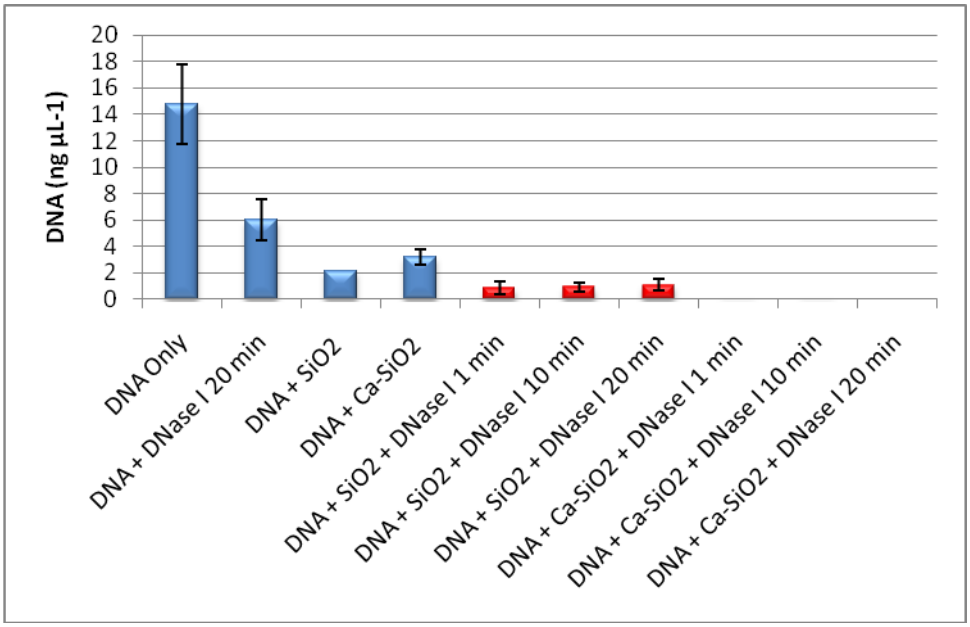

Figure 3.16: Results showing real-time TaqMan® PCR detection of *C. perfringens* genomic DNA in resuspended pellets of SiO_2-DNA mixtures. Mineral concentrations (control and homo-ionic) were 0.1 % w/v for initial DNA adsorptions. Blue bars represent control samples with red being those of SiO_2-DNA mixes exposed to DNase I over a 20 min time course. Error bars show standard deviation of four PCR replicates.

For Kaolinitic supernatants, degradation of DNA in the presence of DNase I was observed for both Ca^{2+} adjusted and no cation controls with starting concentrations of 2.96 and 3.7 ng μl^{-1} and final concentrations of 128.8 pg μl^{-1} and 87 pg ul^{-1} respectively. Observed levels of DNA degradation in resuspended Kaolinitic pellets (Figure 3.13) were from ~ 1 ng μl^{-1} to final concentrations of 112 and 146 ng μl^{-1} after 20 min incubation for Kaolinite and Ca-Kaolinite, respectively. Figure 3.14 shows DNase I degradation of DNA in the presence of Bentonite. DNA concentrations in the supernatants were reduced from 5 and 3.2 ng μl^{-1} to 1.2 and 2.2 ng μl^{-1} for Bentonite and Ca-Bentonite, respectively, after 20 min incubation. No amplification of *C. perfringens* DNA in resuspended Bentonite pellets was observed (data not shown). DNase I mediated DNA degradation in the presence of SiO_2 is shown in Figures 3.15 and 3.16. No degradation of DNA in the supernatants was observed for SiO_2 where the

concentration after 20 min exposure (8.37 ng μl^{-1}) was comparable to starting levels (8.95 ng μl^{-1}). In comparison, DNA was completely degraded in the presence of Ca-SiO_2. Starting DNA concentrations in SiO_2 pellet resuspensions (Figure 3.16) were 2.14 and 3.2 ng μl^{-1} for SiO_2 and Ca-SiO_2 respectively. After 20 min incubation, levels in SiO_2 supernatants were ~ 1ng μl^{-1}. For Ca-SiO_2, after 10 min, the level of DNA had been reduced to 6 x 10^{-5} ng μl^{-1}. These data suggests that DNA is most protected from DNase I degradation in the presence of bentonite (homo-ionic and dH_2O control) and SiO_2 (dH_2O treated). The absence of amplification from either the pellet or supernatant kaolinite samples suggests a low level of nuclease protection.

Figure 3.17 shows the amplification of a 1 Kbp portion of *plC* from *C. perfringens* following exposure of a clay-DNA complex to DNase I through a 20 min time course. Samples were treated and prepared as described above for real-time PCR analysis except that here amplification products were examined on a 2 % agarose gel in 1 x TAE prepared as described in Section 2.4.5.

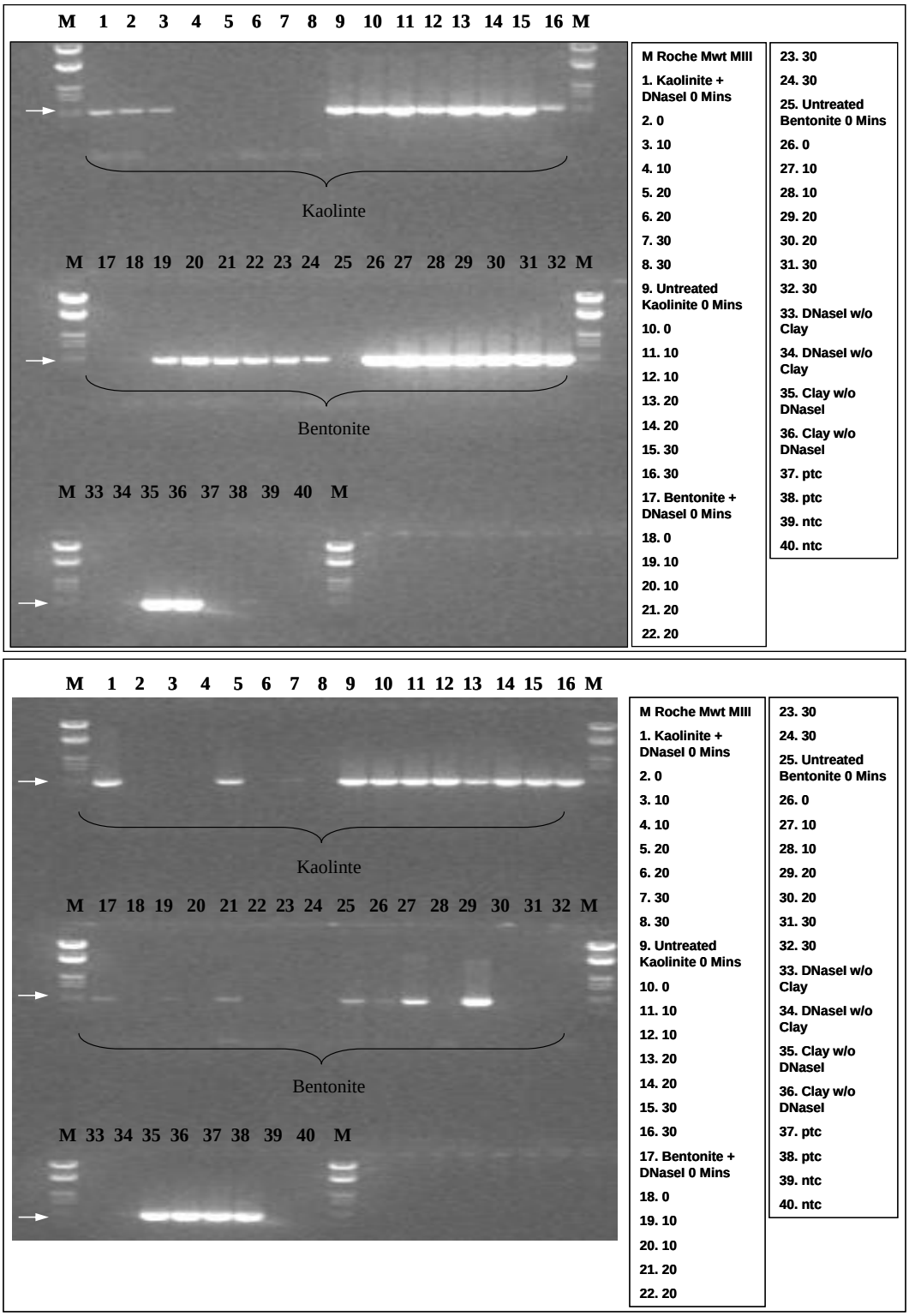

Figure 3.17: Amplification of *C. perfringens plc* in the Presence of Clay Minerals and DNase I. Duplicate end point PCR reactions were performed on supernatants (top) and resuspended pellets (bottom) obtained from Bentonite/Kaolinte-DNA-DNase I mixtures through a 20 min time course. Amplification of an ~1 Kbp product of the *plc* gene from *C. perfringens* is denoted by the white arrow. M - Roche molecular size marker III.

Results shown in Figure 3.17 indicate that, for supernatants, amplification of *plC*

was not possible for Kaolinite after 10 min incubation but remained for Bentonite throughout the time course. Amplification from resuspended pellets of Bentonite was sporadic in both DNase I treated and controls. Amplification for Kaolinitic-DNA complexes in resuspended pellets was almost completely removed in comparison to control samples.

3.4 **Soil DNA Extraction**

3.4.1 **Introduction**

As has been discussed previously, the ability to extract DNA from soil is a requirement for the accurate characterisation of DNA persistence in soil. For the determination of DNA persistence from highly pathogenic organisms, released in the context of biological warfare / bioterrorism, such methods must not only fulfil requirements of sensitivity but also ease of use (for use by military or first responders) and rapidity (Whitehouse and Hottel, 2007). Consequently a number of *in situ* extraction methods, including a number of those that are commercially available, were assessed here for sensitivity of detection and the capacity to process diverse soil types.

3.4.2 **Soil Types**

A number of soil types were selected based on specific challenges associated with nucleic acid extraction or dynamics of DNA persistence. These were peat loam, sandy loam, two clay loam soils with contrasting clay mineralogy and sand. Air-dried, 2 mm sieved soils were supplied by LandLook Ltd (Midlands, UK). Sand was acquired from a local retailer. Analysis of the soil samples was undertaken at The Macaulay Land Use Research Institute (Aberdeen, UK).

Results of this analysis and general characteristics of the soils are shown in Table 3.2.

Table 3.2: Characteristics of Soil Types used in Soil DNA Extraction Method Evaluation and DNA Persistence Studies. Soils were sourced by LandLook Ltd (UK) and analysed at The Macaulay Land Use Research Institute (Aberdeen, UK). ND – not determined.

Soil / Soil Characteristics	148/07/1 Denchworth (Clay Loam)	148/07/2 Teigngrace (Clay Loam)	148/07/3 Bromsgrove (Sandy Loam)	148/07/4 Adventurers (Peat Loam)
Dry Matter Content (% m/m)	92.2	97.7	98.4	71.0
Water Content (% m/m)	8.5	2.4	1.6	40.8
0.0633mm – 2mm (%)	19.88	20.71	ND	ND
0.002mm – 0.063mm (%)	24.08	48.5	ND	ND
<0.002mm (%)	56.04	30.8	ND	ND
pH (1:5) in H_2O	8.0	5.0	6.7	6.1
pH (1:5) in 1 mol L^{-1} KCl	6.6	4.0	5.4	5.7
pH (1:5) in 1 mol L^{-1} $CaCl_2$	7.2	4.0	5.7	5.7
Cation Exchange Capacity (cmol+/kg)	9.7	6.5	ND	ND
Clay	Smectitic	Kaolinitic	ND	ND
Total Organic Carbon (%)	0.75	0.68	1.24	
Total Carbon (%)	ND	ND	ND	24.79
Loss on ignition (%)	ND	ND	ND	48.2

3.4.3 **Evaluation of Soil DNA Extraction Methods**

Seven soil DNA extraction methods were evaluated against three soil types chosen to represent a clay loam (Teigngrace), peat loam (Adventurers) and sand, for the efficient extraction of nucleic acids for subsequent real-time PCR analysis. DNA extractions on triplicate 0.3 g soil samples seeded with $1x10^7$, $1x10^6$, $1x10^5$ or $1x10^4$ CFU vegetative *Bacillus atrophaeus* NCTC 10073 were performed. Methods, selected to represent variations of *in situ* DNA extraction protocols, are summarised in Table 3.3 with full details in Section 2.4.2.

Table 3.3: Brief Details of Selected Soil DNA Extraction Methods

Method	Description
MoBio UltraClean® Soil DNA Isolation Kit	Chemical and physical lysis; silica membrane capture
MoBio UltraClean® (Method for Maximum Yield)	Chemical and heat lysis; silica membrane capture
MoBio Powersoil® Soil DNA Isolation Kit	Chemical and physical lysis; silica membrane capture
Modified Ultraclean® (Andrews *et al.*, 2004)	Standard protocol with additional ethanol precipitation step
Phenol / Chloroform (Kresk and Wellington, 1999)	Physical and chemical lysis; phenol / chloroform extraction
Epicentre SoilMaster™ DNA Isolation Kit	Chemical and heat lysis; ethanol precipitation
Phenol / Chloroform (Yeates et al., 1998)	Physical, chemical and heat lysis; phenol / chloroform extraction

Analysis was performed by quintuplicate real-time TaqMan® PCR assay targeting the cold shock protein (*csp*) of *B. atrophaeus* and kindly supplied by Dr G Campbell (Dstl, UK). PCR was performed in 25 µl reaction volumes on a Smartcycler™ (Cepheid, France) PCR machine. Full assay and cycling conditions can be found in Section 2.4.19. Results are shown in Table 3.4. There was marked variability in the performance of each method against each soil type tested. All the methods allowed *B. atrophaeus* detection in Kaolinitic soil, although with variation in sensitivity. Methods based on phenol: chloroform extraction performed poorly in comparison to commercially available methods. Only the MoBio UltraClean® Soil DNA Isolation kit was able to allow detection to 1 x 10^4 CFU, the lowest concentration tested. The most effective method for soils with high humic acid content (peat loam) was the MoBio Powersoil® kit. Here, detection of vegetative *B. atrophaeus* was possible over the full range tested to 1 x 10^4 CFU. Real-time PCR cycle threshold (*ct*) comparisons were used to assess levels of target DNA in soil extracts. The *ct* is the cycle number at which fluorescence signals from a real-time PCR reaction

exceeds an automatically derived baseline. This value can then be used to quantitatively compare samples. A lower *ct* is indicative of a higher concentration of target DNA in the PCR reaction. The *Ct* value at the lowest concentration was 37.29 (±3.06) for this kit, and represented the lowest *ct* achieved across all the methods tested. The UltraClean® kit with an additional ethanol precipitation step, introduced for further purification of eluted DNA, also enabled detection at the lowest concentration with a *ct* of 41.08 (±4.13). For sand, the PowerSoil® kit performed better than the other methods tested. Here, *B. atrophaeus* was detected down to 1 x 10^5 CFU with a *ct* of 39.1 (±3.82).

Soil Type	Soil DNA Extraction Methods							*Bacillus atrophaeus* concentration CFU $0.3g^{-1}$ Soil
	MoBio UltraClean®	MoBio UltraClean® (Max)	MoBio PowerSoil®	Modified UltraClean® (Andrews *et al.*, 2004)	Phenol / Chloroform (Kresk and Wellington, 1999)	Epicentre SoilMaster™	Phenol / Chloroform (Yeates *et al.*, 1998)	
Kaolinitic Clay Soil	25.43 (±0.54)	26.6 (±1.3)	23.33 (±0.76)	ND	34.13 (±3.97)	29.18 (±2.13)	30.45 (±1.67)	$1x10^7$
	28.55 (±0.39)	30.54 (±0.71)	27.88 (±0.65)	ND	-	32.78 (±1.12)	35.40 (±3.12)	$1x10^6$
	31.61 (±0.79)	-	35.71 (±3.15)	ND	-	-	-	$1x10^5$
	34.98 (±0.56)	-	-	ND	-	-	-	$1x10^4$
Peat Loam	-	-	34.96 (±3.22)	34.19 (±0.47)	42.05 (±0.62)	-	-	$1x10^7$
	-	-	36.15 (±3.03)	35.17 (±1.13)	-	-	-	$1x10^6$
	-	-	35.63 (±2.02)	36.48 (±0.7)	-	-	-	$1x10^5$
	-	-	37.29 (±3.06)	41.08 (±4.13)	-	-	-	$1x10^4$
Sand	43.57 (±2.80)	-	24.87 (±0.63)	ND	-	30.09 (±1.74)	-	$1x10^7$
	-	-	29.16 (±1.06)	ND	-	34.15 (±2.95)	-	$1x10^6$
	-	-	39.1 (±3.82)	ND	-	-	-	$1x10^5$
	-	-	-	ND	-	-	-	$1x10^4$

Table 3.4: Comparison of Soil DNA Extraction Methods for Three Soil Types. Cycle threshold (*ct*) values are shown with SE of 15 replicates in parentheses. (-) no amplification observed; ND – not determined.

3.4.4 **Sensitivity of PowerSoil® Soil DNA Extraction Method with *Burkholderia pseudomallei* and *Bacillus anthracis* Spores**

The sensitivity of the MoBio Powersoil® kit was assessed using one soil type (Kaolinitic clay loam) and bacterial species of interest owing to their potential use as BWAs. *Burkholderia pseudomallei* K96243 and *Bacillus anthracis* Ames were selected as representatives of Gram negative and Gram positive spore forming BWAs respectively.

3.4.4.1 Real-time PCR Assay Development for *B. anthracis* and *B. pseudomallei*

3.4.4.1.1 B. anthracis

Bacillus anthracis, the causative agent of anthrax, is a Gram positive, capsulated bacillus that forms spores. The closest relatives to *B. anthracis* are strains of *B. thuringiensis* and *B. cereus.* Key differences in the virulence of *B. anthracis* in comparison to its close relatives are encoded in the two main plasmids, pX01 (encoding the protective antigen, lethal and oedema factors) and pX02 (encoding the capsule). Capabilities for the accurate detection of *B. anthracis* have been of interest owing to its recognised potential as a BWA and, particularly, subsequent to the US anthrax letters of October 2001 (Hoffmaster *et al.*, 2002).

3.4.4.1.1.1 Real-time TaqMan® PCR Design and Optimisation

The protective antigen (PA) is the central component of a three-part toxin complex secreted by *B. anthracis* (Turnbull, 1990). As a membrane-inserting heptamer, it facilitates the translocation of oedema and lethal factor into the cytosol (Milne and Collier, 1993, Milne *et al.*, 1994). As such it represents a

useful target for the design of assays that enable the detection of potentially virulent strains of *B. anthracis*.

Ten sequences for *B. anthracis* PA were acquired from GenBank with a further sequence from the Comprehensive Microbial Resource (CMR) at the J. Craig Venter Institute (formerly the Institute for Genomic Research – TIGR). Accession numbers for the sequences are listed in Table 3.5. Sequences were aligned using MegAlign ClustalW (DNASTAR Lasergene 6). A consensus sequence was exported into Primer Express v1.0 (Applied-Biosystems, UK) which was then used to design TaqMan® PCR primers and probes.

Table 3.5: *B. anthracis* Protective Antigen Sequences used in Design of Real-Time TaqMan® Assay. NK – not known.

Isolate	Accession Number
B. anthracis IT-Carb3-6254	AJ413936.1
B. anthracis	AF306782
B. anthracis BA1024	AF306783
B. anthracis BA1035	AF306780
B. anthracis IT-Carb1-6241	AJ413937
B. anthracis A2012	AE011190
B. anthracis 33	AF306781
B. anthracis	AF306778
B. anthracis	AF268967
B. anthracis	AF005404
B. anthracis 28	AF306779
B. anthracis Krugger B	NK

3.4.4.1.1.2 Sensitivity and Cross-reactivity

The performance, in terms of detection of *B. anthracis* strains and cross-reactivity with closely related strains, of assay BAPA2 was assessed firstly against an inclusivity panel consisting of 14 $pX01^{+}/pX02^{+}$ *B. anthracis*, and secondly an exclusivity panel of 28 related *B. cereus* group members and other *Bacillus* sp. Genomic DNA at 1 ng μl^{-1} was used in triplicate 25 µl PCR reactions. PCR reactions were performed on a Cepheid SmartCycler® (Cepheid,

France) using the cycling conditions outlined in Section 2.4.19. Results are shown in Tables 3.6 and 3.7.

Table 3.6: Inclusivity Panel Assessment of *B. anthracis* BAPA1 Assay. Triplicate PCRs were performed on a Cepheid SmartCycler® (Cepheid, France) using 1 ng μl^{-1} genomic DNA from 14 strains of *B. anthracis*; (-) indicates no amplification with (+) indicating amplification of strain.

B. anthracis LSU34	+
B. anthracis LSU102	+
B. anthracis LSU149	+
B. anthracis LSU188	+
B. anthracis LSU248	+
B. anthracis LSU256	+
B. anthracis LSU264	+
B. anthracis LSU267	+
B. anthracis 293	+
B. anthracis 328	+
B. anthracis 379	+
B. anthracis 442	+
B. anthracis 465	+
B. anthracis AMES	+

Table 3.7: Exclusivity Panel Assessment of *B. anthracis* BAPA1 Assay. Triplicate PCRs were performed on a Cepheid SmartCycler® (Cepheid, France) using 1 ng μl^{-1} genomic DNA from 28 strains of closely-related *Bacillus* sp.; (-) indicates no amplification with (+) indicating amplification of strain.

B. thuringiensis Al-Hakam	-
B. thuringiensis HD-571	-
B. thuringiensis var. Kurkstaki	-
B. thuringiensis 97-27	-
B. cereus 34	-
B. cereus D17	-
B. cereus F1-15	-
B. cereus NCTC10320	-
B. cereus ATCC10876	-
B. cereus NCTC11143	-
B. cereus NCTC11145	-
B. cereus NCTC4312	-
B. cereus NCTC6474	-
B. cereus NCTC7464	-
B. cereus NCTC8035	-
B. cereus NCTC9939	-
B. cereus NCTC9945	-
B. cereus NCTC9946	-
B. cereus S2-8	-
B. atrophaeus NCTC10073	-
B. atrophaeus CEB 93/029	-
B. atrophaeus ATCC9372	-
B. circulans NCTC2510	-
B. coagulans NCTC10334	-
B. lichenforms NCTC10341	-
B. mycoides NCTC9680	-
B. pasteuri NCTC4822	-
B. brevis NCTC2611	-

Sensitivity was assessed with a dilution series of genomic DNA from *B. anthracis* AMES. Results are shown in Figure 3.18. The assay enabled detection to 0.1 pg of genomic DNA per PCR reaction.

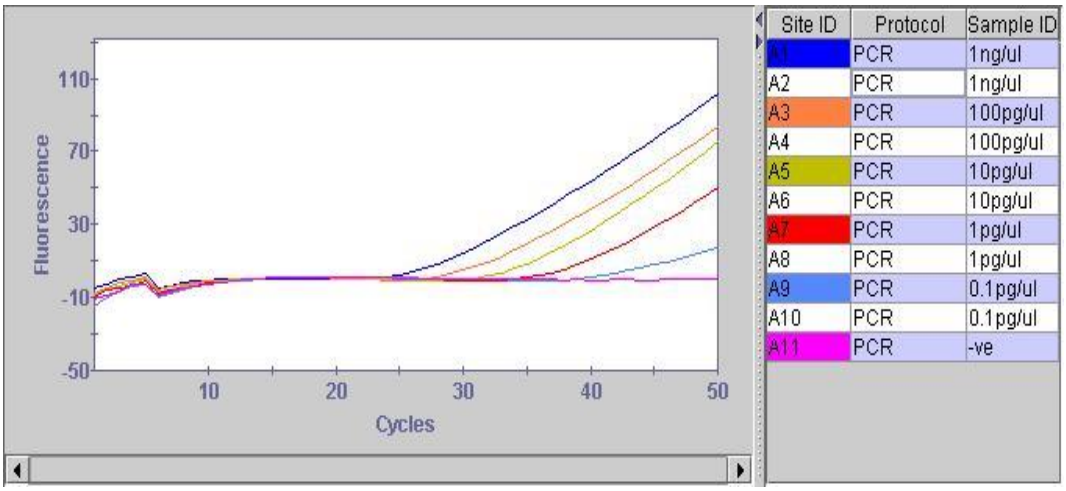

Figure 3.18: Assessment of the sensitivity of a real-time PCR assay targeting the protective antigen of *B. anthracis*. Duplicate PCR reactions for a dilution series of genomic DNA from *B. anthracis* AMES (1 ng μl^{-1} to 0.1pg μl^{-1}) were performed on a Cepheid SmartCylcer® PCR machine. Amplification was observed at 0.1 pg μl^{-1}.

3.4.4.1.2 B. pseudomallei

B. pseudomallei is the causative agent of melioidosis, a tropical disease of both humans and animals that is endemic in Northern Australia and South East Asia with sporadic occurrences in many other countries (Cheng and Currie, 2005). Its use, as is the case with *B. mallei* as a biological weapon, is well documented (Waag and DeShazer, 2004; Cheng and Currie, 2005). As a result both have been classified as category B select agents by the Centre for Disease Control (CDC) (Larsen and Johnson, 2009).

3.4.4.1.2.1 Real-time TaqMan® PCR Design and Optimisation

It has previously been shown that both *B. pseudomallei* and *B. mallei* contain polysaccharide gene clusters which are important virulence determinants (DeShazer *et al.*, 2001). Importantly, from the view of avoiding amplification of the closely related, non-pathogenic soil saprophyte *B. thailandensis*, this cluster is truncated in this organism and therefore some of the genes lack homologues in

this organism. Consequently one gene, *wcbM* (encoding d-glycero-d-manno-heptose 1-phosphate) was chosen as a target for the design of a real-time PCR assay capable of specifically detecting both *B. pseudomallei* and *B. mallei.*

Eleven sequences for *B. pseudomallei* and *B. mallei wcbM* were acquired from GenBank. Accession numbers for the sequences are listed in Table 3.8. Sequences were aligned using MegAlign ClustalW (DNASTAR Lasergene 6). A consensus sequence was exported into Primer Express v1.0 (Applied-Biosystems, UK) which was then used to design TaqMan® PCR primers and probes.

Table 3.8: *Burkholderia pseudomallei* and *B. mallei wcbM* Sequences used in Design of Real-Time TaqMan® Assay.

Isolate	Accession Number
Burkholderia pseudomallei 1106a	CP000572.1
Burkholderia pseudomallei 1710b	CP000124.1
Burkholderia pseudomallei MSHR346	CP001408.1
Burkholderia pseudomallei	AF228583.1
Burkholderia mallei NCTC 10247	CP000548.1
Burkholderia pseudomallei 668	CP000570.1
Burkholderia mallei NCTC 10229	CP000546.1
Burkholderia mallei SAVP1	CP000526.1
Burkholderia mallei ATCC 23344	CP000010.1
Burkholderia pseudomallei K96243	BX571965.1
Burkholderia mallei	AF285636.1
Burkholderia pseudomallei 1106a	CP000572.1

3.4.4.1.2.2 Sensitivity and Cross-reactivity

The performance, in terms of detection of *B. pseudomallei* / *B. mallei* strains and cross-reactivity with closely related organisms, of assay CAPS2 (Table 2.6) (which targets *wcbM*) was assessed firstly against an inclusivity panel consisting of 22 *B. pseudomallei* / *B. mallei* strains, and secondly an exclusivity panel of 29 related *Burkholderia* sp.. Genomic DNA at 1 ng μl^{-1} was used in triplicate 25 µl PCR reactions. PCR reactions were performed on a Cepheid SmartCycler®

(Cepheid, France) using the cycling conditions outlined in Section 2.4.10.1. Results are shown in Tables 3.9 and 3.10.

Table 3.9: Inclusivity Panel Assessment of *B. pseudomallei* / *B. mallei* CAPS2 Assay. Triplicate PCRs were performed on a Cepheid SmartCycler® (Cepheid, France) using 1 ng μl^{-1} genomic DNA from 22 strains of *B. pseudomallei* / *B. mallei*; (-) indicates no amplification with (+) indicating amplification of strain.

B. mallei ATCC 23344	+
B. mallei NCTC 10230	+
B. mallei NCTC 10245	+
B. mallei NCTC 10260	+
B. mallei NCTC 120	+
B. mallei NCTC 10247	+
B. mallei NCTC 10248	+
B. pseudomallei DSTO 20	+
B. pseudomallei DSTO 21	+
B. pseudomallei DSTO 22	+
B. pseudomallei DSTO 23	+
B. pseudomallei DSTO 24	+
B. pseudomallei DSTO T9	+
B. pseudomallei DSTO T14	+
B. pseudomallei DSTO T17	+
B. pseudomallei DSTO T18	+
B. pseudomallei DSTO T43	+
B. pseudomallei DSTO T75	+
B. pseudomallei DSTO T83	+
B. pseudomallei DSTO T87	+
B. pseudomallei DSTO T88	+
B. pseudomallei DSTO T106	+

Table 3.10: Exclusivity Panel Assessment of *B. pseudomallei* / *B. mallei* CAPS2 Assay. Triplicate PCRs were performed on a Cepheid SmartCycler® (Cepheid, France) using 1 ng μl^{-1} genomic DNA from 29 strains of *B. pseudomallei* / *B. mallei*; (-) indicates no amplification with (+) indicating amplification of strain.

B. thailandensis 251	-
B. thailandensis E264	-
B. thailandensis E82	-
B. thailandensis E27	-
B. thailandensis 257	-
B. thailandensis 253	-
B. thailandensis 215	-
B. cepacia 32540	-
B. cepacia 17754	-
B. cepacia ATCC17759	-
B. cepacia ATCC25416	-
B. cepacia 9091	-
B. cepacia C3159	-
B. cepacia C1964	-
B. cenocepacia J2596	-
B. multivorans 13010	-
B. multivorans 7897	-
B. vietnamiensis C1375	-
B. vietnamiensis 159	-
B. vietnamiensis NCPPB 4256	-
B. stabilis 8088	-
B. stabilis 14086	-
B. stabilis C1372	-
B. ambifaria ATCC BAA-2440-S	-
B. plantari NCPPB 3590	-
B. glumae NCPPB 2981	-
B. gladiolii NCPPB 2151	-
B. caryophylli NCPPB 1891	-
B. andropogonis NCPPB 934	-

Sensitivity was assessed using a dilution series of genomic DNA from *B. pseudomallei* K96243 and *B. mallei* ATCC 23344. Results are shown in Figure 3.18. Reliable detection was possible at 100 fg μl^{-1} concentration.

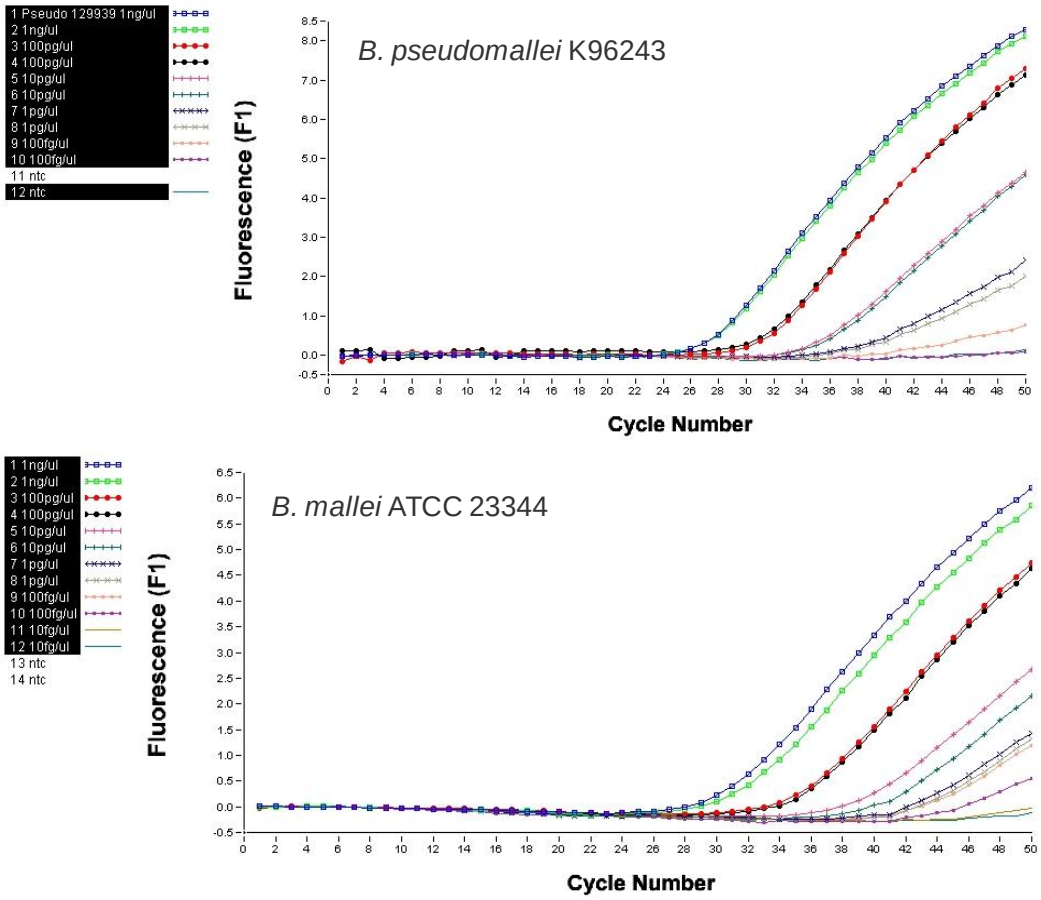

Figure 3.19: Assessment of the sensitivity of a real-time PCR assay targeting *Burkholderia pseudomallei* and *B. mallei wcbM*. Duplicate PCR reactions for a dilution series of genomic DNA (1 ng μl^{-1} to 0.01pg μl^{-1}) from *B. pseudomallei* (top) and *B. mallei* (bottom) were performed on a Roche LightCylcer™ PCR machine. In both cases amplification was observed for 100 fg μl^{-1}.

3.4.4.2 Evaluation of DNA Extraction Method Sensitivity

Here, triplicate 0.3 g kaolinitic clay loam soil samples were seeded with a concentration gradient of $1x10^7$ 0.3 g^{-1} down to $1x10^0$ CFU of *B. pseudomallei* K96243 or spores of *B. anthracis* AMES. Quintuplicate real-time PCR reactions were then performed. Results are shown in Table 3.11. For both inoculants, detection was possible to 1 x 10^5 CFU 0.3 g^{-1}.

Table 3.11: Sensitivity of PowerSoil™ DNA Extraction Protocol from Clay loam Soil Inoculated with *B. anthracis* Spores or *B. pseudomallei*. Ct values are shown with SE of 15 replicates in parentheses. (-) no amplification observed

Inoculums	Concentration (CFU) 0.3 g^{-1} Soil							
	$1x10^7$	$1x10^6$	$1x10^5$	$1x10^4$	$1x10^3$	$1x10^2$	$1x10^1$	$1x10^0$
B. anthracis AMES (Spores)	27.1 (±0.2)	30.6 (±0.2)	35.3 (±0.2)	-	-	-	-	-
B. pseudomallei K96243	23.1 (±2.6)	30.1 (±1)	32.6 (±0.2)	-	-	-	-	-

3.5 **Persistence of DNA in Soil Microcosms**

DNA persistence was compared across three soil types; two contrasting clay loams (Kaolinitic and Smectitic) and Sandy loam soil. Inocula used were either inactivated but intact bacterial cells, inactivated and lysed bacterial cells or extracellular DNA. These were chosen to represent the potential levels of DNA exposure to degradation that may be present in soil. *Burkholderia cepacia* ATCC 17759 was used for the bacterial cell inocula. Genomic DNA from *Clostridium perfringens* was used as extracellular nucleic acids. In both cases, real-time TaqMan® PCR analysis was used to determine DNA decay over a time course of 128 days at ambient temperatures (~19 ^{o}C).

3.5.1 **Development of a Real-time PCR Assay Targeting *recA* of the *Burkholderia cepacia* complex**

RecA is commonly used as a real-time PCR target for the detection and differentiation of *Burkholderia cepacia* complex strains (Payne *et al.*, 2005). Nineteen sequences for *recA* from members of the *Burkholderia cepacia*

complex were acquired from GenBank. Accession numbers for the sequences are listed in Table 3.12. Sequences were aligned using MegAlign ClustalW (DNASTAR Lasergene 6). A consensus sequence was exported into Primer Express v1.0 (Applied-Biosystems, UK) which was then used to design TaqMan® PCR primers and probes.

Table 3.12: *Burkholderia cepacia* complex *recA* Sequences used in Design of Real-Time TaqMan® Assay.

Isolate	Accession Number
Burkholderia cenocepacia J2315	AM747720.1
Burkholderia cepacia C5424	AF143781.1
Burkholderia cepacia C4455	AF143782.1
Burkholderia cepacia K56-2	AF143779.1
Burkholderia sp. Y20	EF426458.1
Burkholderia sp. Y10	EF426457.1
Burkholderia sp. M279	DQ989505.1
Burkholderia sp. M229	DQ989504.1
Burkholderia cenocepacia LMG16656	AY951880.1
Burkholderia cepacia ATCC25416	AF143786.1
Burkholderia cepacia ATCC17759	AF143788.1
Burkholderia pyrrocinia	AF143794.1
Burkholderia cenocepacia HI2424	CP000458.1
Burkholderia cenocepacia AU 1054	CP000378.1
Burkholderia cepacia genomovar VI	AF323971.1
Burkholderia sp. 383	CP000151.1
Burkholderia cepacia J2540	AF456053.1
Burkholderia pyrrocinia JK-SH007	GQ169786.1
Burkholderia multivorans HI2308	AF143777.1

3.5.1.1 Sensitivity of *recA* Real-time PCR Assay

Sensitivity was assessed using a dilution series of genomic DNA from *B. cepacia* ATCC 17759. Results are shown in Figure 3.20. Amplification was observed to 10fg μl^{-1}.

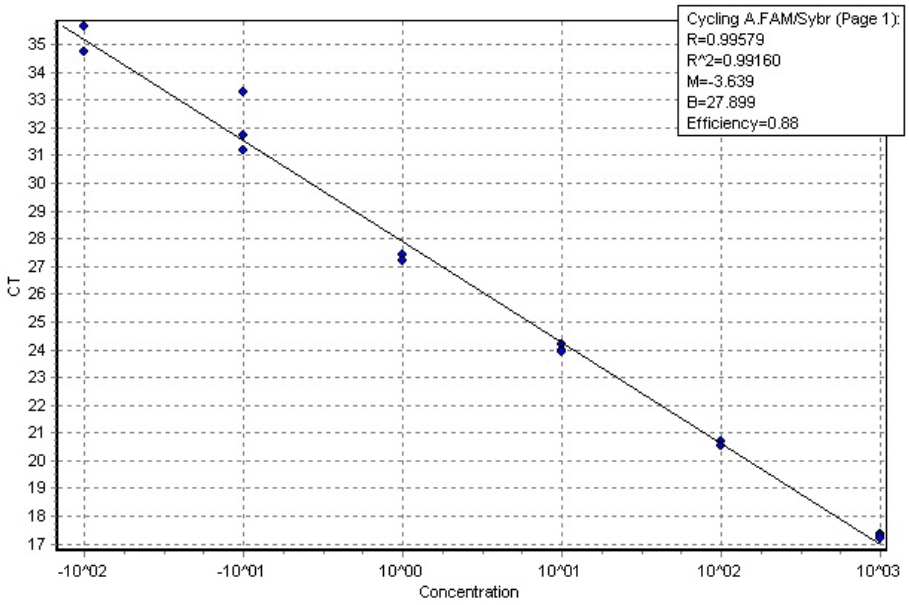

Figure 3.20: *Burkholderia cepacia* complex *recA* real-time PCR Standard Curve. PCR reactions using 1 ng, 100 pg, 10 pg, 1 pg, 100 fg and 10fg *B. cepacia* ATCC 17759 genomic DNA were performed in triplicate on a Rotor-Gene™ thermal cycler.

3.5.2 Detection of *Burkholderia cepacia* ATCC 17759 in Soil Microcosms

Soil microcosms, consisting of 0.3 g of Kaolinitic clay loam, Smectitic clay loam or sandy loam soil prepared in triplicate, were destructively sampled using the MoBio PowerSoil® kit at time points 0, 4, 8, 16, 32, 64 and 128 days following initial inoculation with 1 x 10^7 CFU of *B. cepacia* ATCC 17759. At each time point DNA was extracted using the PowerSoil® DNA Isolation kit according to the manufacturers' instructions. Real-time PCR was performed using the conditions outlined in Section 2.4.19. Results are shown in Figure 3.21. PCR amplification of *recA* was possible in all three soil types up to and including day 128.

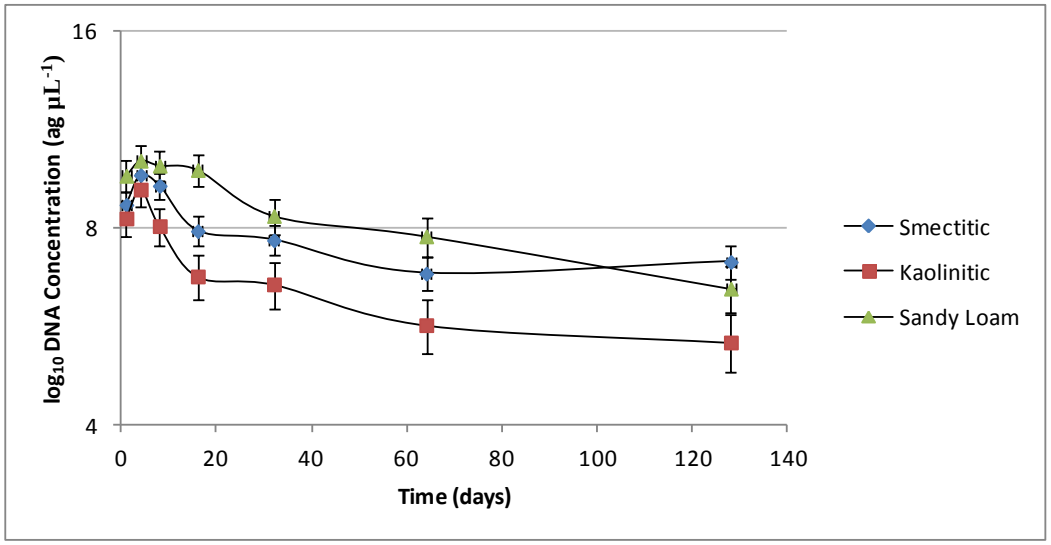

Figure 3.21: Detection of *B. cepacia* ATCC 17759 in Soil Microcosms. Replicate 0.3 g soil microcosms (n = 3) of Kaolinitic clay loam, Smectitic clay loam and Sandy loam were seeded with 1x10^7 CFU of *B. cepacia* ATCC 17759. DNA extracts (2 µL) obtained at each time point were subjected to triplicate quantitative PCR analysis. Error bars show standard deviation of nine PCR reactions.

3.5.3 **Persistence of DNA from Lysed Bacterial Cells**

Soil microcosms, consisting of 0.3 g of Kaolinitic clay loam, Smectitic clay loam or sandy loam soil prepared in triplicate, were destructively sampled using the MoBio PowerSoil® kit at time points 0, 4, 8, 16, 32, 64 and 128 days following initial inoculation with 1 x 10^7 CFU of lysed *B. cepacia* ATCC 17759. Lysis was performed by incubation of 1 ml cell suspension aliquots in PBS at 99 ^{o}C in a water-bath for 10 min. Seeded microcosms were incubated at 19 ^{o}C prior to sampling and extraction. To determine lysis efficacy, 100 µl of the cell suspension was subsequently plated on LB Agar plates and incubated at 37 ^{o}C in a static incubator for 72 h. At each time point DNA was extracted using the PowerSoil® DNA Isolation kit according to the manufacturers' instructions. Real-time PCR was performed using the conditions outlined in Section 2.4.19. Results are shown in Figures 3.22 – 3.24. PCR amplification of *recA* was possible in all three soil types up to and including day 128.

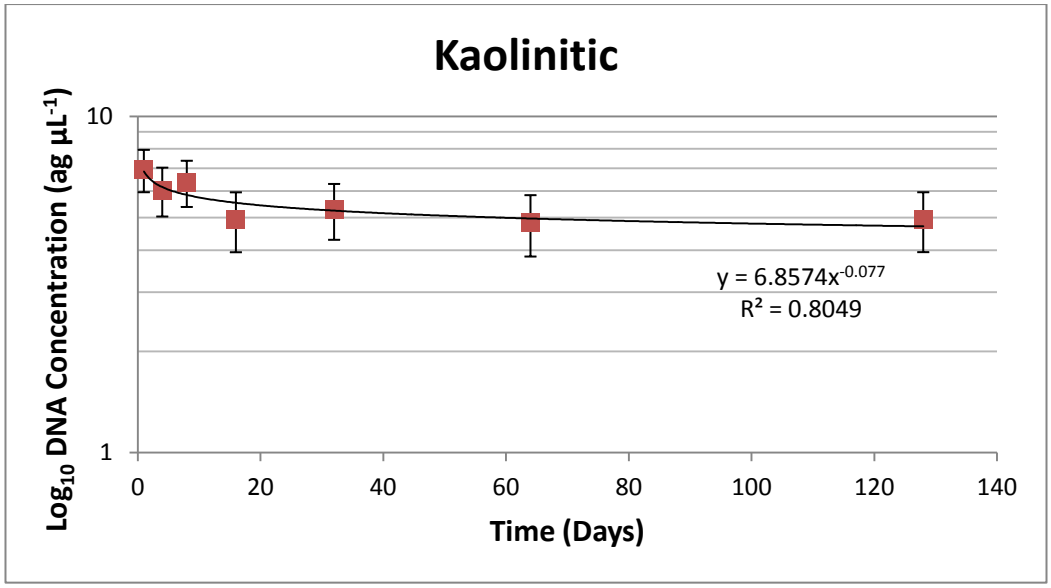

Figure 3.22: Real-time PCR Detection of DNA from Lysed *B. cepacia* ATCC 17759 Cells in Kaolinitic Clay Loam Soil. Replicate 0.3 g soil microcosms (n = 3) of Kaolinitic clay loam were seeded with $1x10^7$ CFU of lysed *B. cepacia* ATCC 17759. DNA extracts (2 µl) obtained at each time point (0, 4, 8, 16, 32, 64 and 128 days post-inoculation) were subjected to triplicate quantitative real-time PCR analysis. Error bars show standard deviation of nine PCR reactions.

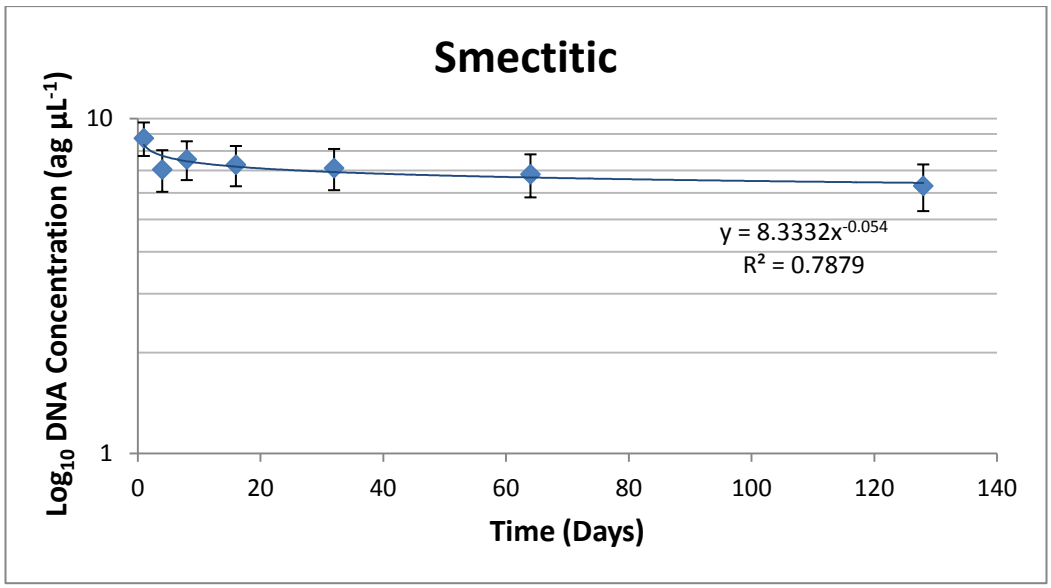

Figure 3.23: Real-time PCR Detection of DNA from Lysed *B. cepacia* ATCC 17759 Cells in Smectitic Clay Loam Soil. Replicate 0.3 g soil microcosms (n = 3) of Smectitic clay loam were seeded with $1x10^7$ CFU of lysed *B. cepacia* ATCC 17759. DNA extracts (2 µl) obtained at each time point (0, 4, 8, 16, 32, 64 and 128 days post-inoculation) were subjected to triplicate quantitative real-time PCR analysis. Error bars show standard deviation of nine PCR reactions.

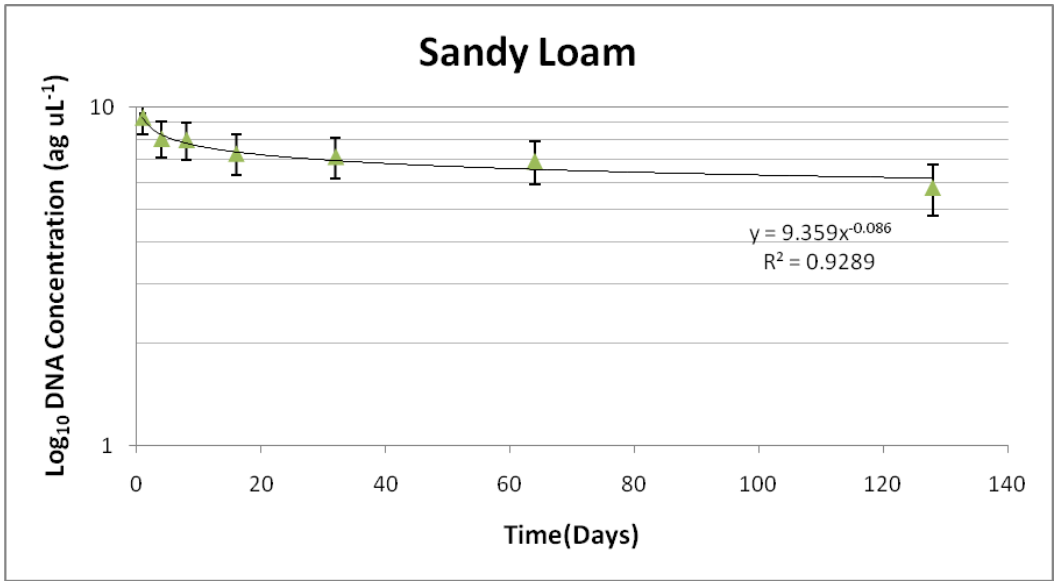

Figure 3.24: Real-time PCR Detection of DNA from Lysed *B. cepacia* ATCC 17759 Cells in Sandy Loam Soil. Replicate 0.3 g soil microcosms (n = 3) of Sandy loam were seeded with $1x10^7$ CFU of lysed *B. cepacia* ATCC 17759. DNA extracts (2 µl) obtained at each time point (0, 4, 8, 16, 32, 64 and 128 days post-inoculation) were subjected to triplicate quantitative real-time PCR analysis. Error bars show standard deviation of nine PCR reactions.

To determine the statistical significance of soil type on DNA persistence in the three soils tested, linear regression analysis and ANOVA were performed in Excel. Here, $\log_{10}$ values for DNA concentration were plotted against $\log_{10}$ time values as shown in Figure 3.25. ANOVA outputs and *p*-values are shown in Table 3.13.

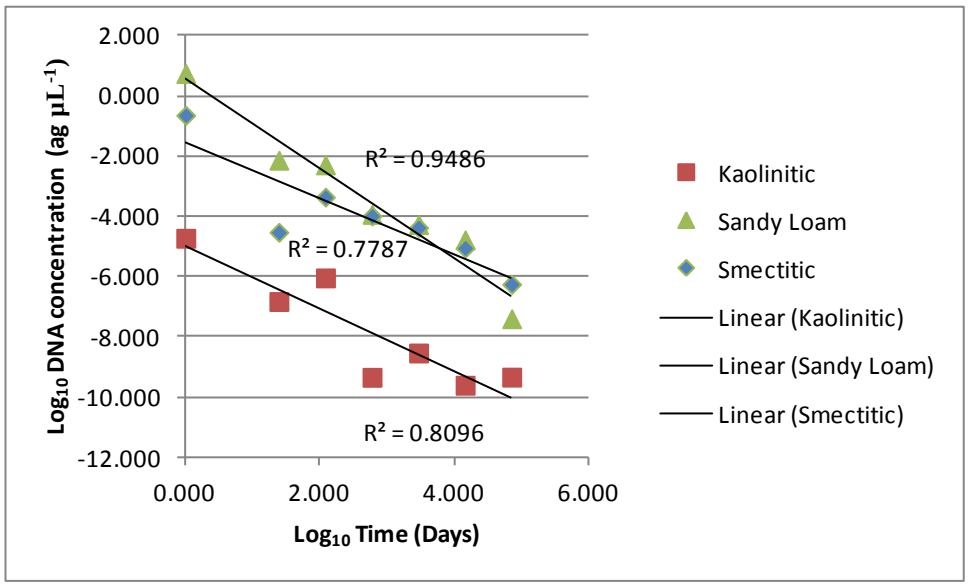

Figure 3.25: Linear Regression Analysis of DNA Decay for Kaolinitic, Smectitic Clay loam and Sandy Loam Soil Microcosms Seeded with 1 x 10^7 CFU Lysed *B. cepacia* ATCC 17759 Cells.

Table 3.13: ANOVA Outputs for Linear Regression Analysis of Soil Type vs. Time. *p*-values are shown for comparing decay curves of Kaolinitic and Smectitic Clay Loam Soils to Sandy Loam Soil.

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	146.3433986	29.26868	42.908766	2.31444E-08
Residual	15	10.23171347	0.682114		
Total	20	156.5751121			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.533704645	0.623363031	0.85617	0.4053738	-0.794962197	1.862371487
log time	-1.48453652	0.201816747	-7.35586	2.383E-06	-1.914698736	-1.054374314
soil indicator K	-5.54617953	0.881568452	-6.29126	1.446E-05	-7.425198203	-3.667160867
soil indicator Sm	-2.0966498	0.881568452	-2.37832	0.0311145	-3.975668469	-0.217631133
log time x K	0.4528269	0.28541198	1.586573	0.1334613	-0.155514333	1.061168133
log time x Sm	0.563289719	0.28541198	1.973602	0.0671413	-0.045051514	1.171630952

The *p*-values shown in Table 3.13 indicate that the decay curve for Kaolinitic clay loam soil is not significantly different to that of Sandy loam (*p*-value = 0.1334613). Smectitic clay loam however does appears to exhibit a decay curve

that is significantly slower when compared to that of Sandy loam (*p*-value = 0.0671413) and thus also Kaolinitic clay.

3.5.4 **Persistence of DNA from Dead-Intact Bacterial Cells**

Soil microcosms, consisting of 0.3 g of Kaolinitic clay loam, Smectitic clay loam or sandy loam soil prepared in triplicate, were destructively sampled at time points 0, 4, 8, 16, 32, 64 and 128 days following initial inoculation with 1 x 10^7 CFU of UV-killed *B. cepacia* ATCC 17759. UV inactivation was performed by exposure of 1 ml cell suspension aliquots in PBS to a 254 nm light source for two min in a UV cross-linker (UVTEC). Seeded microcosms were incubated at 19 ^{o}C prior to sampling and extraction. To determine UV inactivation efficacy, 100 µL of the cell suspension was subsequently plated on LB Agar plates and incubated at 37 ^{o}C in a static incubator for 72 h. At each time point DNA was extracted using the PowerSoil™ DNA Isolation kit according to the manufacturers' instructions. Real-time PCR was performed using the conditions outlined in 2.4.19 and described above. Results are shown in Figures 3.26 – 3.30.

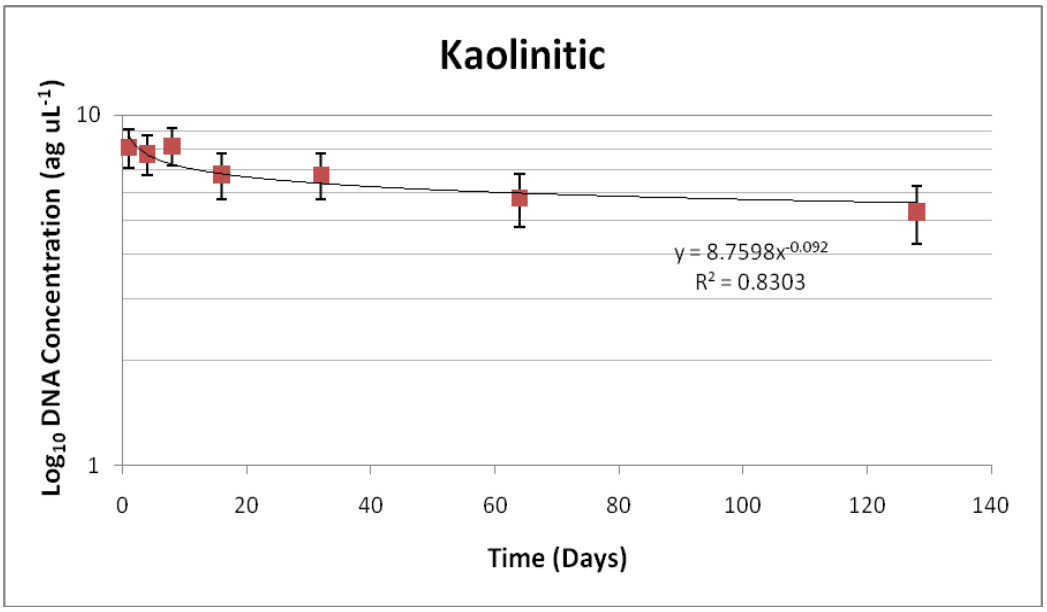

Figure 3.26: Real-time PCR Detection of DNA from UV-killed *B. cepacia* ATCC 17759 Cells in Kaolinitic Clay Loam Soil. Replicate 0.3 g soil microcosms (n = 3) of Kaolinitic clay loam were seeded with $1x10^7$ CFU of UV-killed *B. cepacia* ATCC 17759. DNA extracts (2 µL) obtained at each time point (0, 4, 8, 16, 32, 64 and 128 days post-inoculation) were subjected to triplicate quantitative real-time PCR analysis. Error bars show standard deviation of nine PCR reactions.

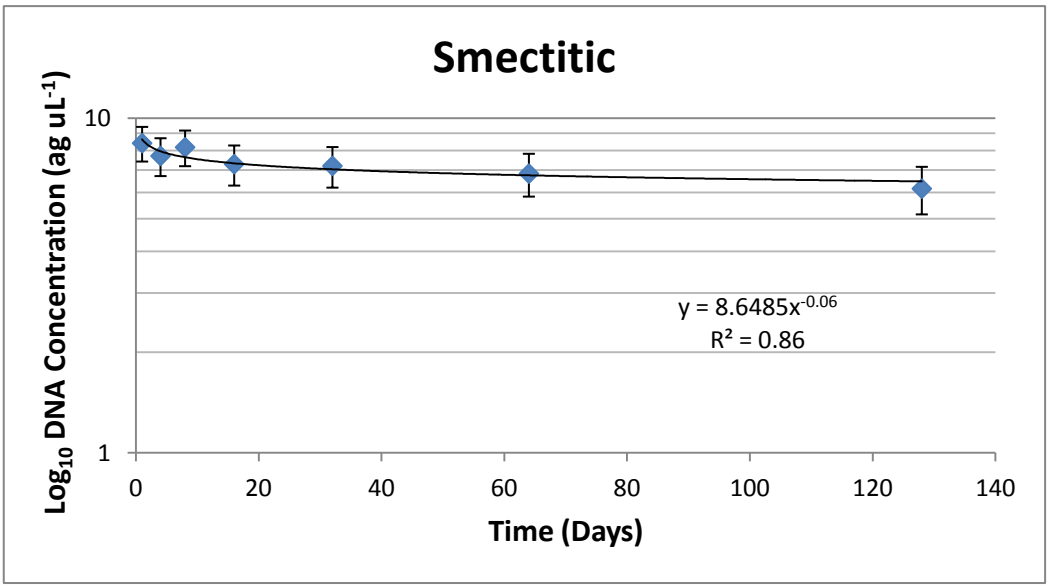

Figure 3.27: Real-time PCR Detection of DNA from UV-killed *B. cepacia* ATCC 17759 Cells in Smectitic Clay Loam Soil. Replicate 0.3 g soil microcosms (n = 3) of Smectitic clay loam were seeded with $1x10^7$ CFU of UV-killed *B. cepacia* ATCC 17759. DNA extracts (2 µL) obtained at each time point (0, 4, 8, 16, 32, 64 and 128 days post-inoculation) were subjected to triplicate quantitative real-time PCR analysis. Error bars show standard deviation of nine PCR reactions.

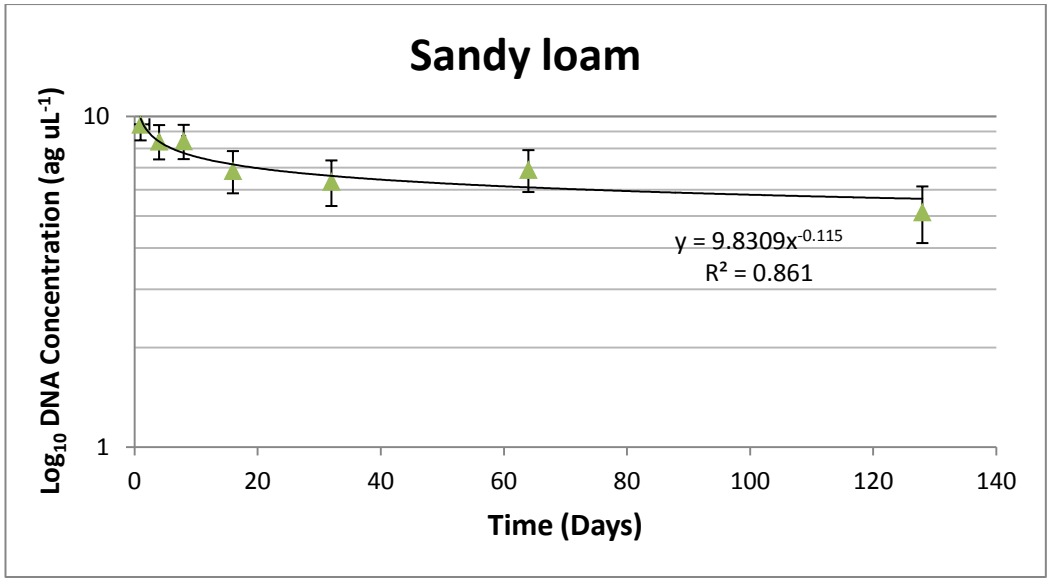

Figure 3.28: Real-time PCR Detection of DNA from UV-killed *B. cepacia* ATCC 17759 Cells in Sandy Loam Soil. Replicate 0.3 g soil microcosms (n = 3) of Sandy loam were seeded with $1x10^7$ CFU of UV-killed *B. cepacia* ATCC 17759. DNA extracts (2 µL) obtained at each time point (0, 4, 8, 16, 32, 64 and 128 days post-inoculation) were subjected to triplicate quantitative real-time PCR analysis. Error bars show standard deviation of nine PCR reactions.

PCR amplification of *recA* was possible for all three soil types up to and including day 128. Using linear regression and ANOVA analysis methods outlined above the decay curves for Kaolinitic clay loam and Sandy loam soils were not statistically different (*p*-value = 0.188489). The decay observed for Smectitic clay loam soil was shown to be significantly slower than that observed for Sandy loam with a *p*-value of 0.024149.

3.5.5 **Persistence of Extracellular DNA**

Soil microcosms, consisting of 0.3 g of Kaolinitic clay loam, Smectitic clay loam or sandy loam soil prepared in triplicate, were destructively sampled at time points 0, 4, 8, 16, 32, 64 and 128 days following initial inoculation with 200 ng genomic DNA (Sigma-Aldrich, UK) representing approximately 10^8 genome equivalents of *C. perfringens*. At each time point DNA was extracted using the PowerSoil® DNA Isolation kit according to the manufacturer's instructions.

Real-time PCR was performed using the conditions outlined above and in 2.4.19. Results are shown in Figure 3.29.

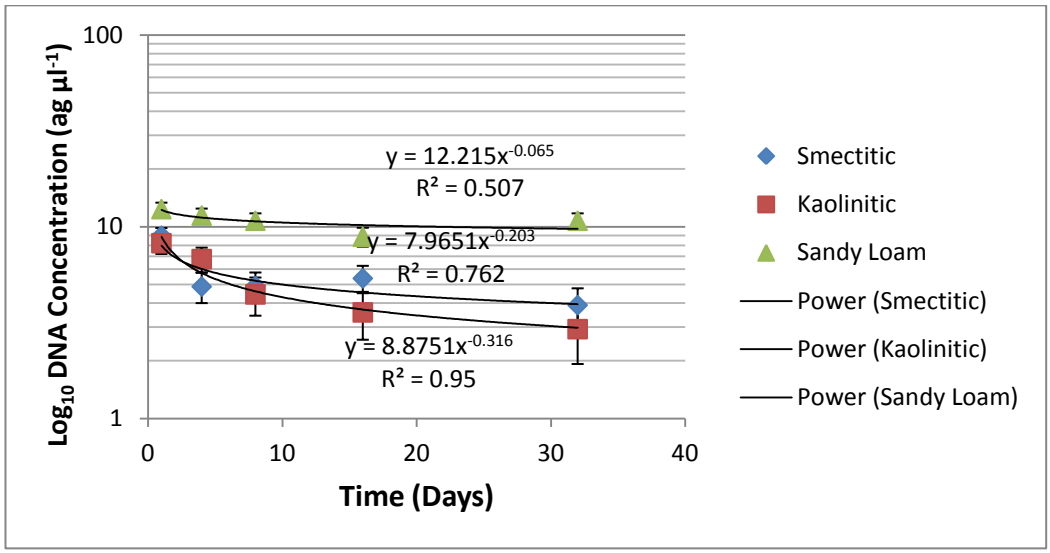

Figure 3.29: Real-time PCR Detection of genomic DNA of *C. perfringens* in Kaolinitic Clay Loam, Smectitic Clay Loam and Sandy Loam Soil. Replicate 0.3 g soil microcosms (n = 3) of each soil were seeded with 10^8 genome equivalents (100 µl volume) of genomic DNA from *C. Perfringens*. DNA extracts (2 µL) obtained at each time point (0, 4, 8, 16, 32, 64 and 128 days post-inoculation) were subjected to triplicate quantitative real-time PCR analysis. Error bars show standard deviation of nine PCR reactions.

Amplification of DNA was achieved up to day 32 of the time course for both clay loam soils. No amplification was observed at day 64. Amplification of DNA from Sandy loam soil was possible up to and including day 64 where the time course was terminated. For visual clarity to aid comparative analysis, Log_{10} DNA concentration is only plotted up to day 32 for this soil type (Figure 3.29). Whilst the decay curve for Sandy loam soil was shown to be statistically significant to both Kaolinitic and Smectitic clay loam soil (*p*-value = 0.018666 and 0.084013 respectively), comparison of the two clay loams did not reveal a statistically significant difference (*p*-value = 0.465583).

3.5.6 **DNA Half-life in Soil Microcosms**

Half-lives (t½) and remaining DNA levels (%) were calculated for each of the decay curves shown in Figures 3.22 – 3.29. These are shown in Table 3.14.

Table 3.14: Half-lives (t½) and remaining DNA levels reported for each soil type (smectitic clay loam, kaolinitic clay loam and sandy loam) for each inoculum investigated. ND – not determined.

Inocula	Soil Type	Half-life (t½) Days	Remaining DNA Levels (%) at Expt. End
Lysed B. cepacia	Smectitic Clay Loam	27.8	0.367
	Kaolinitic Clay Loam	41.4	1.000
	Sandy Loam	11.03	0.029
Dead-intact *B. cepacia*	Smectitic Clay Loam	30.2	0.544
	Kaolinitic Clay Loam	23.5	0.157
	Sandy Loam	7.3	0.004
Extracellular DNA	Smectitic Clay Loam	6.1	0.0008
	Kaolinitic Clay Loam	6.5	0.0005
	Sandy Loam	ND	3.024

3.6 **Discussion**

Initial observations of DNA adsorption to minerals, representative of those within the soil types that would ultimately be used in DNA persistence studies, were undertaken. In addition to kaolinite and bentonite, silicon dioxide was chosen, over quartz (the commonly found form in soil) as a direct comparative for the two clays as these were both used in highly purified form. Ideally the results presented here would have been confirmed using minerals directly extracted from the soils that were ultimately used. Regardless, the contrast between the high variations observed for the silicon dioxide compared to the clay minerals used here suggests this approach was flawed. There was good agreement for both adsorptions over time and responses to available cation and pH with previous studies (Khanna and Stotzky, 1992; Pietramellara *et al.*, 1997; Cai *et al.*, 2006). Results here indicated that swelling-clays with high cation exchange capacities such as bentonite adsorb greater quantities of DNA at higher affinity than non-swelling types such as kaolinite. This contrast had previously been shown to be 4 -7 times greater by Cai *et al.*, (2006). Protection from DNase I degradation was also in keeping with previous studies with bentonite providing greater levels of resistance than kaolinite. There had been contrasting evidence as to the potential for montmorillonite to inhibit amplification of DNA (Vettori *et al.*, 1996; Alvarez *et al.*, 1998). It is clear from this study that the presence of bentonite, a structurally similar clay mineral to montmorillonite, does indeed have an inhibitory effect on amplification of both small (~100 bp) and large (~1 Kbp) targets, most likely a result of physical separation of polymerase and DNA template within the structure of the clay mineral.

The approach used here to evaluate persistence of DNA of bacterial origin in contrasting soil types included direct DNA extraction of total DNA and real-time PCR-based detection of *B. cepacia* and *C. perfringens*. Prior to the evaluation of DNA persistence dynamics a number of methods were evaluated for the direct extraction of total DNA from soil. These were assessed using real-time PCR detection of vegetative cells of *B. atrophaeus*, spores of *B. anthracis* and the Gram negative *B. pseudomallei*. Of the methods evaluated, only the MoBio PowerSoil® kit enabled reliable DNA extraction and detection of *B. atrophaeus* vegetative cells within the three diverse soil types tested (clay loam, peat loam and sand). As a consequence this method was evaluated with two organisms of interest due to their potential use as biological warfare agents (*B. anthracis* and *B. pseudomallei*). The detection levels of *B. anthracis* spores presented here were comparable to that recently reported for other commercial extraction kits (Gulledge *et al.*, 2010). Here, the authors were able to demonstrate detection of 10^5 spores using the MoBio UltraClean® kit. This method however did not enable comparable detection levels across the soils tested. This is in keeping with the data presented here and mitigated by the selection of the MoBio PowerSoil® kit which proved superior for more challenging soils such as peat loam. Gulledge *et al.*, (2010) also examined the effects of pre-treatment and enrichment on detection sensitivities. The pre-treatment, involving the suspension of the soil in 5 ml of 0.1 % sodium pyrophosphate tetrabasic ($Na_4P_2O_7$), 1 mmol l^{-1} EDTA, 10 $mmol^{-1}$ Tris-Cl, increased the detection limit by 1 log, but not consistently across all the kits and the three soil types tested. The enrichment process, where soil was suspended in 25 ml of a selective medium (polymyxin B-lysozyme-EDTA-thallium acetate broth (PLET) supplemented

with 38 µg ml^{-1} sulfamethoxazole and 2 µg ml^{-1} trimethoprim, followed by overnight culture proved more beneficial where detection limits for all the soils tested increased to 10^2 and sporadically 10^1 spores. However, in this case there is an obvious trade-off between sensitivity, capacity of the method to cope with diverse soil types and speed of analysis. The latter is particularly pertinent where detection of pathogens in the context of bio-terrorism is the major objective and speed of initial analysis is potentially a higher priority. In this case, whilst not achieving an increase in overall sensitivity, the PowerSoil® method shown here does potentially enable consistent detection from diverse soil types. A further study by Dineen *et al.*, (2010) using spores of *Bacillus cereus* highlighted the FastDNA® SPIN kit as the most effective for sand, sandy clay and sandy loam soils. Interestingly, the PowerSoil® kit did not enable detection below 10^9 spores g^{-1} for the clay soil. This is in contrast to results presented here where detection levels were 4 logs higher at 10^5. It is possible that the type of clay in the *B. cereus* study, although not detailed, may have resulted in the sequestering of spores decreasing the efficiency of spore lysis. For *B. pseudomallei*, the detection of only 10^5 CFU was unexpected. It was thought that extraction from a Gram negative organism would be at a greater level than that of a Gram positive spore-former such as *B. anthracis* as a result of increased lysis efficiency. However recent studies, although only evaluating methods on sandy loam, demonstrated detection limits of between 10^4 and 10^5 genome equivalents per gram of soil for both adapted and commercial kits (Trung *et al.*, 2011). Clearly, any comparison of performance is restricted by lack of comparisons of assay and use of similar soil type. An additional limitation of the study presented here was the choice of only one soil type.

As a consequence of the observations regarding DNA adsorption to clay and other soil minerals presented here, ultimately the hypothesis was whether soil types with mineralogy that might enable increased adsorption would enable prolonged DNA survival. Results here show DNA persistence using either intact or lysed cells of *B. cepacia*, regardless of soil type, up to at least 128 days post-inoculation at ambient (~19 ^{o}C) temperatures. This long-term persistence has been observed previously with PCR-based detection achieved from three months to at least two years (Widmer *et al.*, 1996, 1997; Paget *et al.*, 1998; Gebhard and Smalla, 1999). Examining the persistence of recombinant neomycin phosphotransferase II (rNPT-II) at two sites of genetically modified plant cultivation, Widmer *et al.*, (1997) detected rNPT-II up to 77 days for tobacco leaf litter buried in fibreglass bags in soil, and 137 days for leaf, stem tissues and tubers left on the surface at a second site. Mirroring the approach used here, Pontiroli *et al.*, (2009) demonstrated the persistence of purified plant DNA in 10 g microcosms for 210 weeks. Final quantitative analysis determined the remaining DNA represented approximately 0.002 % of the initial inoculums (Pontiroli *et al.*, 2009). The authors noted that the greatest decrease (10,000-fold) was observed within the first two weeks (Pontiroli *et al.*, 2009). This is in agreement with results shown here where DNA persistence appeared to stabilise after sixteen days incubation in all soil types evaluated. DNA decay curve analysis, half-lives and remaining DNA levels for each soil type indicate that both clay soils enabled increased levels of DNA persistence from lysed and intact *B. cepacia* cells when compared to sandy loam. This is suggestive of a positive effect for the presence of clay for prolonged DNA survival possibly as a result of

adsorption and protection from nuclease degradation. Although linear regression and ANOVA revealed statistically significant differences between the decay curves of kaolinitic clay loam and smectitic clay loam soil it was not possible to determine if this difference revealed a greater propensity for one clay soil type to enable increased persistence over the other. Nielsen *et al.*, (2000) had previously determined that cellular debris may enable increased persistence of DNA of bacterial origin in soil. This is supported by the observation that extracellular DNA exhibited the shortest half-life and lowest remaining quantity (6.1 and 6.5 days with 0.0008 % and 0.0005 % remaining for smectitic and kaolinitic clay soil respectively) in comparison to soils inoculated with intact or lysed cells. Additionally, at least for smectitic clay loam, soils seeded with intact *B. cepacia* cells appeared to generate shorter half-lives for DNA decay than lysed cells.

The complexity of soil environments represents a methodological challenge when simulating DNA release *in situ* and subsequent persistence. Thus certain limitations to this study must be noted. Firstly, perturbation of the microcosms, either physically or through rewetting cycles was not undertaken which may have impacted the duration of persistence. Additionally, factors such as microcosm scale, temperature fluctuations and the presence of a rhizosphere could have been implemented and/or varied. Finally, in order to verify the suggested positive impact of clay mineral presence on DNA persistence presented here, one could use a non-clay soil as the base microcosm that could then be supplemented with various forms and types of clay mineral in a controlled and quantifiable fashion. The impact of the biotic elements of the soil environment should also be considered. Here, the diversity and activity of

microbial communities will also determine the duration of DNA persistence and the potential for adaptation of introduced pathogenic species to the prevailing environmental conditions.

In summary, this chapter aimed to determine the impact of various abiotic factors on the fate of extracellular DNA in soil. The purpose of this was to examine how these could impact on the utility of real-time PCR for the detection of BWAs in this environmental matrix; recognising that long-term, the persistence of the pathogens nucleic acid is more likely than the organism itself. The influence of both clay type and surrounding pH on adsorption of DNA was shown. Clay type was shown to influence both the quantity of DNA adsorbed as well as the resultant level of protection from nuclease degradation. Here bentonite clay mineral was shown to provide the greatest protection, although, in contrast to kaolinite, it was also shown to prevent amplification of adsorbed DNA. The impact of these clay minerals when present in soil was then assessed for their affects on DNA persistence. The soil types chosen were first used in a comparative evaluation of soil DNA extraction methods – an important consideration, as the efficient extraction of nucleic acids from this complex matrix is necessary for the use of real-time PCR. Once a suitable method was chosen, each soil type was compared for the duration of persistence for inoculated nucleic acid material (cell-associated and extracellular). Here, soil types with clay minerals were shown to enable longer DNA persistence compared to non-clay soils although it was not possible to discern differences between those with different clay type.

Chapter 4

Natural Genetic Transformation in the Burkholderia cepacia Complex

4.1 **Introduction**

As demonstrated in Chapter 3, DNA from bacteria, closely related to potential BWA, can persist in soil for extended periods of time. Aside from the benefits that this gives in terms of being able to detect a released BWA by real-time PCR, there are also potentially negative implications. These concern the horizontal transfer of genetic material from highly pathogenic microorganisms. As discussed in Chapter 1, from the perspective of horizontal gene transfer (HGT), there are two possibilities that arise from the introduction of pathogenic microorganisms into an existing microbial community. These are 1) the transfer of genetic material from the released BWA to existing microbial community members, and/or 2) the acquisition of genetic elements by the pathogen. As the focus in Chapter 3 has been on the the persistence of DNA rather than the pathogen itself then the HGT process most applicable, as it requires extracellular DNA, is that of natural transformation.

The transfer of genetic information between distantly or even unrelated organisms during evolution has been inferred from nucleotide sequence comparisons (Dröge *et al.*, 1998). Analysis of 88 prokaryotic genomes revealed that laterally transferred genes account for between 0 and 22% of the total gene content depending on the genus examined (Garcia-Vallve *et al.*, 2003). The genetic diversity created by these events enables organisms to adapt to adverse environmental conditions or facilitate the evasion of host defence mechanisms (Robertson and Meyer, 1992). As it relates specifically to the persistence and availability of extracellular DNA, only transformation and not conjugative

transfer or transduction was the method of horizontal gene transfer investigated here.

Transformation has received a great deal of attention owing to the frequent use of bacterial antibiotic resistance markers in the construction of transgenic plants (Nielsen *et al.*, 1998). The most common marker is the Tn5 derived *npt*II gene and has been used in numerous plant species (Flavell *et al.*, 1992). The risk of antibiotic resistance transfer to naturally competent soil dwelling bacteria has thus warranted intensive research. Of particular note is the fact that plant-derived DNA persists in soil environments for long periods of time (Gebhard and Smalla, 1999; Paget *et al.*, 1998; Widmer *et al.*, 1997) and that allelic rescue in competent bacteria using transgenic plant DNA has been described (Iwaki and Arakawa, 2005) even within soil microcosms (Nielsen *et al.*, 2000). It is considered, if natural transfer from plant material to indigenous microbial populations does occur, that without the presence of homologous sequences to facilitate uptake / recombination then this would be a rare event (Dröge *et al.*, 1998). However, concerns regarding the validity of experimental observations of horizontal gene transfer into soil microbial populations to date have been raised (Nielsen and Townsend, 2004). These centre on the inadequate sampling strategies to detect rare transformation events, the lack of effort focused on the long-term impact of transfers and the conditions that would facilitate their persistence (Nielsen and Townsend, 2004).

Transformation involves the uptake, integration and stable inheritance of cell free DNA by bacterial cells and enables the transfer of DNA across genus boundaries

(Dröge *et al*. 1998; Thomas and Nielsen, 2005). Competence, the bacterial physiological state required for transformation, has been identified in over 60 bacterial and archeal species from the Phyla Eukaryarchaeota, Deinococcus-Thermus, Cyanobacteria, Cholorobi, alpha, beta, epsilon and gamma subdivisions of the Proteobacteria, the Firmicutes and Actinobacteria (Johnsborg *et al.*, 2007). Importantly, many human pathogens are naturally transformable including *Haemophilus influenzae*, *Neisseria gonorrhoeae*, *Vibrio cholerae*, *Helicobacter pylori* and *Burkholderia pseudomallei* (Mathis and Scocca, 1982; Meibom *et al.*, 2005; Thongdee *et al.*, 2008). Taken together with the observed homology between competence systems of Gram positive and negative bacteria, it is suggested that transformation is a phylogenetically ancient adaptation (Johnsborg *et al.*, 2007). However, the sporadic distribution of competence-related genes among organisms capable of transformation and those that are not suggests a facet that is frequently lost (Johnsborg *et al.*, 2007). Aside from the presence of transformable DNA three events must occur in the process of transformation:

1. Development of competence,
2. Capture and uptake of free DNA and
3. Integration and expression of the gene(s) into the host chromosome or the integration / recircularisation of translocated plasmid DNA.

In most species, competence can be described as a transient state that arises as a response to environmental stimuli that result in the up-regulation of genes that enable gene acquisition (Lorenz and Wackernagel, 1994). With reference to

growth phase, some organisms have been shown to become competent during logarithmic growth phase, such as species belonging to the genera *Acinetobacter* and *Neisseria*, others at the onset of stationary phase (*Bacillus subtilis*) and others that develop this physiological response for only short periods (*Streptococcus pneumoniae*) (Palmen *et al.*, 1994). Moreover, the proportion of a bacterial culture that will develop a competent state can vary from virtually 0% to 100% depending on the organism in question (Thomas and Nielsen 2005).

4.1.1 **Transformation in Gram Negative Bacteria**

As is the case for Gram positives, there has been extensive study of a select number of species from the beta and gamma subdivisions of the Proteobacteria which have been used to generate a model for uptake in Gram negatives (Johnsborg *et al.*, 2007). These are *Neiserria meningitidis*, *N. gonorrhoeae* and *Haemophilus influenzae*. DNA uptake in the majority of cases occurs via Type IV pili. Especially for the Neisseriaceae (and the Pasturellaceae of the gamma subdivision), cell receptors are however restricted to DNA from closely related organisms (Mathis and Scocca, 1982).

Whilst there is currently no indication of the propensity for the *Burkholderia cepacia* Complex (Bcc) to undergo transformation, Thongdee *et al*., (2008) demonstrated natural transformation within the human pathogenic lineages of the Burkholderiaceae. Prior to this, the closest relatives shown to utilise this transfer mechanism belonged to the genus *Ralstonia* (a member of the Burkholderiaceae) and *Pseudomonas*. In particular, the phytopathogenic *R. solanacearum* and the ubiquitous soil dwelling saprophyte *P. stutzeri* (Bertolla *et al.*, 1997; Lorenz and Sikorski, 2000).

R. solanacearum GMI1000 has been shown to be competent for transformation *in vitro* as well as during plant infection processes (Bertolla *et al.*, 1997; Bertolla and Simonet, 1999). Bertolla *et al*. (1997) demonstrated that competence was a physiological response of the cells that occurred during exponential growth phases and was not a result of an extracellular competence factor. They also determined that a minimum of 50 bp of homologous sequence was required for recombination. More recently, Mercier *et al*., (2007) showed that a sequence divergence of 2 % was sufficient to prevent homologous recombination and allelic rescue; indicating that this species may possess considerable barriers to the uptake and integration of heterologous sequences. A mutator phenotype (Δ*mutS*) achieved a sexual isolation value 100-fold lower than the wild-type (304.9 to 3.1) indicating the importance of the mismatch repair system in preventing this acquisition. In examining the extent to which a mutator phenotype would allow for interspecific DNA transformation, Mercier *et al.*, (2007) further demonstrated that transformation frequencies were routinely 10-fold higher than the wild-type strain even when exposed to sequence divergence of up to 20 %. As it is likely that SOS responses (the error-prone DNA repair global response that is activated in response to DNA damage (Michel, 2005)) would be up-regulated during plant infection cycles then the potential consequence of a hyper-*rec* phenotype has been postulated to have beneficial consequences which allow transient phenotypes that permit non-homologous DNA acquisition.

The ability of *P. stutzeri* strains to undergo natural transformation, even within soil extract media has been known for some time (Lorenz and Wackernagel, 1990, 1991). It was shown that the period of competence development in liquid cultures of *P. stutzeri* occurs as a sharp peak in late exponential phase that is rapidly lost in stationary and absent in overnight cultures (Lorenz and Wackernagel, 1990). *P. stutzeri* JM300 utilises a restriction system to discriminate against foreign DNA (Lorenz *et al.,* 1998). In terms of physiological responses and structural components, *P. stutzeri* requires the development of competent state, Type IV pili and ComA (Meier *et al.,* 2002). The pili of *P. stutzeri* mediate the uptake of DNA into the periplasmic space. At this point the DNA becomes resistant to DNase I. Mutant *P. stutzeri* strains carrying inactivated *pilA* (coding for the structural sub-unit of Type IV pili) and *pilC* (coding for an accessory protein for pilus biosynthesis and thus unable to form pili) were shown by Meier *et al.*, (2002) to be transformation deficient. Interestingly, a *pilT* mutant (Tf59) that had excessive but non functional pili was also non-transformable. It has been suggested that PilT, which is involved in the depolymerisation of pili and hence retraction as part of the twitching motility phenotype, physically draws bound DNA into the periplasm (Graupner *et al.*, 2001; Wolfgang *et al.*, 1998). Although non-transformable, *pilA* of *P. aeruginosa* have been shown to restore competence in a mutant strain of *P. stutzeri* where *pilA* had been inactivated through insertion of a gentamicin resistance cassette (van Schaik *et al.,* 2005). It should be noted however that not all Gram negatives rely on the same type IV pili mechanisms for DNA uptake. In *N. meningitidis* for example, class I but not class II type IV pili were shown to be required for competence (Sun *et al.,* 2005). ComA, which has characteristics

similar to an integral polytopic membrane protein, is thought to form an inner membrane pore through which single stranded DNA molecules pass (Meier *et al.*, 2002). In this example, ComA is facilitated by ExbB, a two membrane domain protein (Graupner and Wakernagel, 2001).

4.1.2 **Aims**

Given the transformation potential of *B. pseudomallei* and *B. thailandensis* it is possible that members of the Bcc can undergo gene transfer through this mechanism. Targeting the Bcc to try and identify this process serves two purposes. Firstly, as the Bcc are closely related to the highly pathogenic Burkholderiaceae, then barriers to transformation between these lineages are potentially weak or non-existent. In this regard it could be possible for Bcc to acquire genes from *B. pseudomallei* for example. Secondly, if the Bcc do demonstrate this process they could provide insight into the evolution and functioning of transformation in this genus. Lastly, members of the Bcc are also pathogenic in their own right, albeit to a susceptible human subpopulation, and therefore understanding possible routes for acquiring new genetic elements is important from this perspective.

The specific aims were as follows:

- Develop a selectable, defined gene knock-out construct for mutagenesis.
- Assess the propensity for transformation through homologous sequence mediated integration using genomic DNA from the above mutants.
- Investigate non-homologous sequence uptake and integration.

4.2 *Burkholderia cepacia* complex Mutant Generation Strategies

To investigate the propensity for members of the Bcc to undergo transformation an allelic marker rescue strategy was devised. Here, a single-gene chromosomal knock-out mutant would be exposed to gDNA from the wild-type strain. If allelic rescue was observed, through restoration of phenotype and loss of selectable marker, then transformation would have occurred. *ZmpA*, encoding a zinc metalloprotease, was the target gene for mutagenesis and subsequent studies of transformation events in the Bcc. This metalloprotease was originally described in *B. cenocepacia* Pc715j (Mckevitt, 1989) and has been shown to be important for virulence in *B. cenocepacia* K56-2 (Kooi *et al.*, 2009). Presence of *zmpA* was previously demonstrated in *B. cepacia*, *B. cenocepacia*, *B. stabilis*, *B. ambifaria*, and *B. pyrocinia*, but not in *B. multivorans*, *B. vietnamiensis*, *B. dolosa* or *B. anthina* and correlated to detectable extracellular proteolytic activity (Gingues *et al.*, 2005). Two strategies for the generation of *zmpA* knock-out mutants were investigated. These are shown in Figure 4.1. For clarity each strategy is colour-coded dark-blue or red which in turn are used to identify results applicable to each.

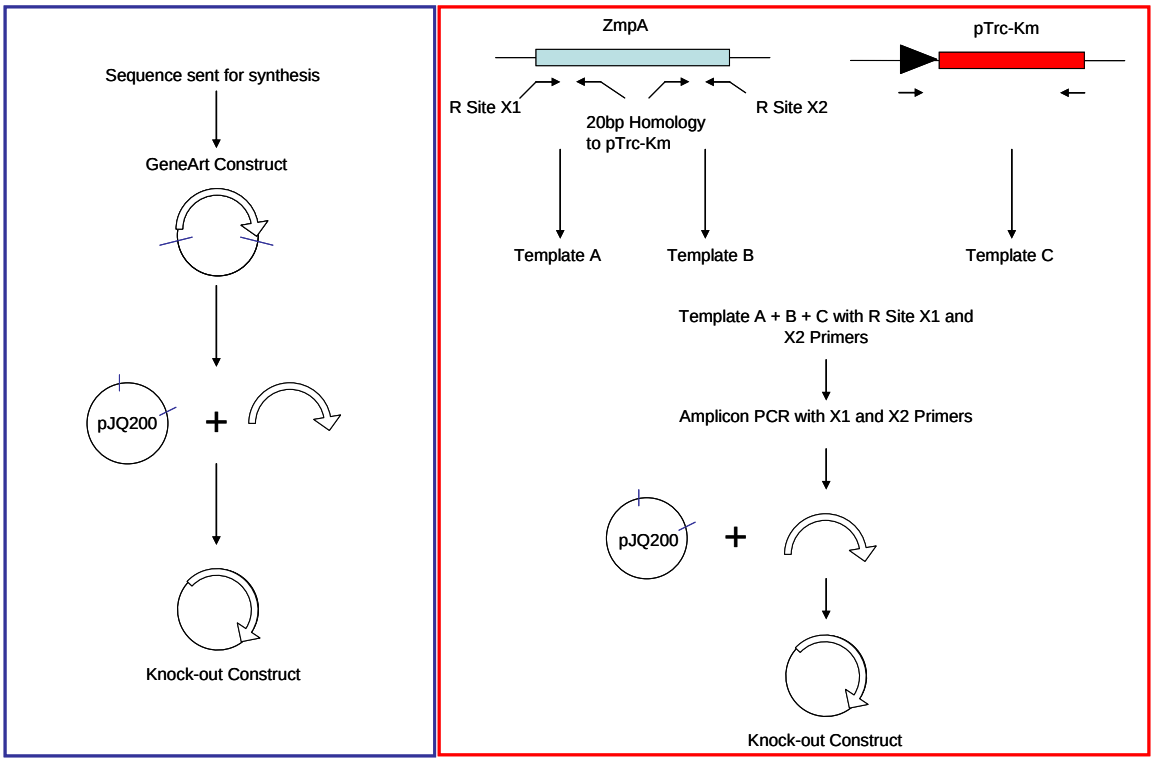

Figure 4.1: Strategies for Creating ZmpA Knock-out Constructs on the Suicide Vector pJQ200 for use in the Bcc

4.2.1 *ZmpA* in the *Burkholderia cepacia* Complex

The presence of *zmpA* in Bcc strains was determined using two primer pairs: 1-2RRP3 and 1-2FFP2 (Gingues *et al.*, 2005) and ZmpA2F / 2R (Section 2.4.19). Primers ZmpA2F and 2R were designed using twelve *zmpA* sequences from GenBank. Accession numbers for the sequences are listed in Table 4.1.

Table 4.1: *Burkholderia cepacia* complex *zmpA* sequences

Burkholderia cepacia complex Strain	Accession Number
B. cenocepacia strain LMG 18832	DQ069250.1
B. cenocepacia HI2424 chromosome 3	CP000460.1
B. cenocepacia AU 1054	CP000378.1
B. cenocepacia strain LMG 16654	DQ069249.1
B. cenocepacia MC0-3	CP000960.1
B. cenocepacia J2315	AM747722.1
B. cepacia	AY143552.1
B. cenocepacia strain LMG 18827	DQ069248.1
B. cenocepacia strain K56-2	DQ069247.1
B. ambifaria AMMD	CP000441.1
B. ambifaria MC40-6	CP001026.1
Burkholderia sp. 383	CP000150.1
B. cenocepacia strain LMG 18832	DQ069250.1

Sequences were aligned using MegAlign ClustalW (DNASTAR LaserGene 8). A consensus sequence was exported into Primer Express v1.0 (Applied-Biosystems, UK) which was then used to design PCR primers. Amplification of *zmpA* was determined using genomic DNA extracted from overnight LB broth cultures according to the method detailed in Sections 2.3.3 and 2.4.1. Primers and PCR conditions are detailed in Tables 2.6 and 2.7. PCR products were analysed on a 2 % agarose gel (Section 2.4.5). Results are summarised in Table 4.2. Although there was sporadic amplification between the two assays overall there was agreement with the expected species that appeared to contain *zmpA*.

Table 4.2: Amplification of *zmpA* in *Burkholderia* sp. Triplicate PCRs were performed on a GeneAmp® PCR System 9700 (Applied Biosystems, UK) using 1 ng µl^{-1} genomic DNA from 15 strains of *Burkholderia* sp.; (-) indicates no amplification with (+) indicating amplification of strain.

Burkholderia sp.	1-2RRP3/ 1-2FFP2	ZmpA2F/ ZmpA2R
B. cepacia J2540	-	-
B. cepacia C1964	+	+
B. cepacia C1359	+	-
B. cepacia 17759	+	+
B. cepacia 9091	-	+
B. cenocepacia J2596	+	+
B. multivorans 7897	-	-
B. multivorans 13010	-	-
B. stabilis 14086	-	+
B. stabilis C3172	+	+
B. stabilis 8088	-	+
B. vietnamiensis 159	-	-
B. vietnamiensis C1375	ND	-
B. mallei 10248	-	-
B. pseudomallei 23344	-	-

4.2.2 **Amplification of 5' and 3' Regions of *zmpA* for Knock-out Construct**

To enable homologous recombination between the knock-out cassette located on the suicide vector pJQ200 and *zmpA*, two regions (50 bp) of homology to the 5' and 3' regions the target gene were amplified for subsequent use in the PCR extension strategy (Figure 4.1). Figure 4.2 shows the target regions of *zmpA* for amplification. Primer design was performed on a MegAlign consensus sequence generated through ClustalW alignment (DNASTAR LaserGene 8) using Primer Express software v1.0 (Applied-Biosystems, UK). Amplification was performed using the profile detailed in Section 2.4.19 on a Rotor-Gene™ real-time PCR instrument (Qiagen, UK). Results for *B. cepacia* ATCC 17759 are shown in

Figure 4.3. Amplicons were purified using Microcon® YM-100 columns according to the manufacturer's instructions (Section 2.4.7).

Figure 4.2: Amplification strategy for 5' and 3' regions of *zmpA* knock-out construct. Bases highlighted in yellow represent areas with inter-species sequence heterogeneity.

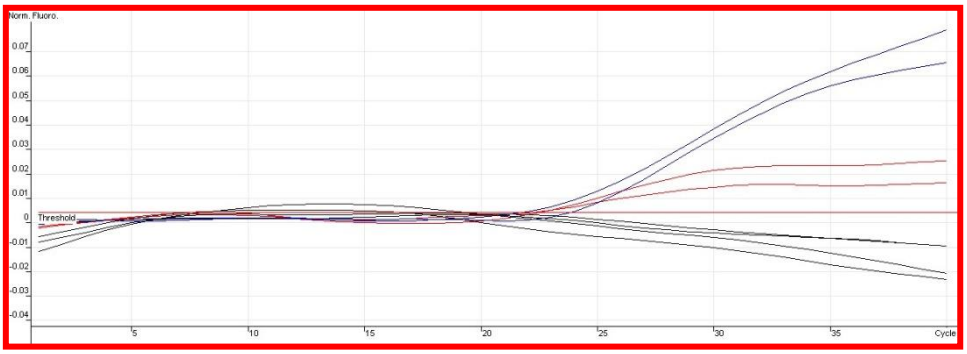

Figure 4.3: Amplification of 5' and 3' sequences of zmpA from *B. cepacia* ATCC 17759. Duplicate PCR reactions using genomic DNA (1 ng μl^{-1}) were performed on a RotorGene (Qiagen, UK) real-time PCR machine. Trace identities are as follows: blue trace – 5' region; red trace – 3' region; black trace – no template control.

4.2.3 **Amplification of Selective Markers**

For use as the selectable marker in the knock-out cassette both antibiotic (kanamycin) or heavy metal selection (potassium tellurite) were evaluated in parallel. Selection using tellurite had previously shown utility for *Pseudomonas* sp. (Sanchez-Romero *et al.*, 1998). The sensitivity of *B. cepacia* ATCC 17759 was first determined according to the method in Section 2.3.5. Minimum inhibitory concentration (determined using $O.D_{600}$) for each is shown in Figure 4.4.

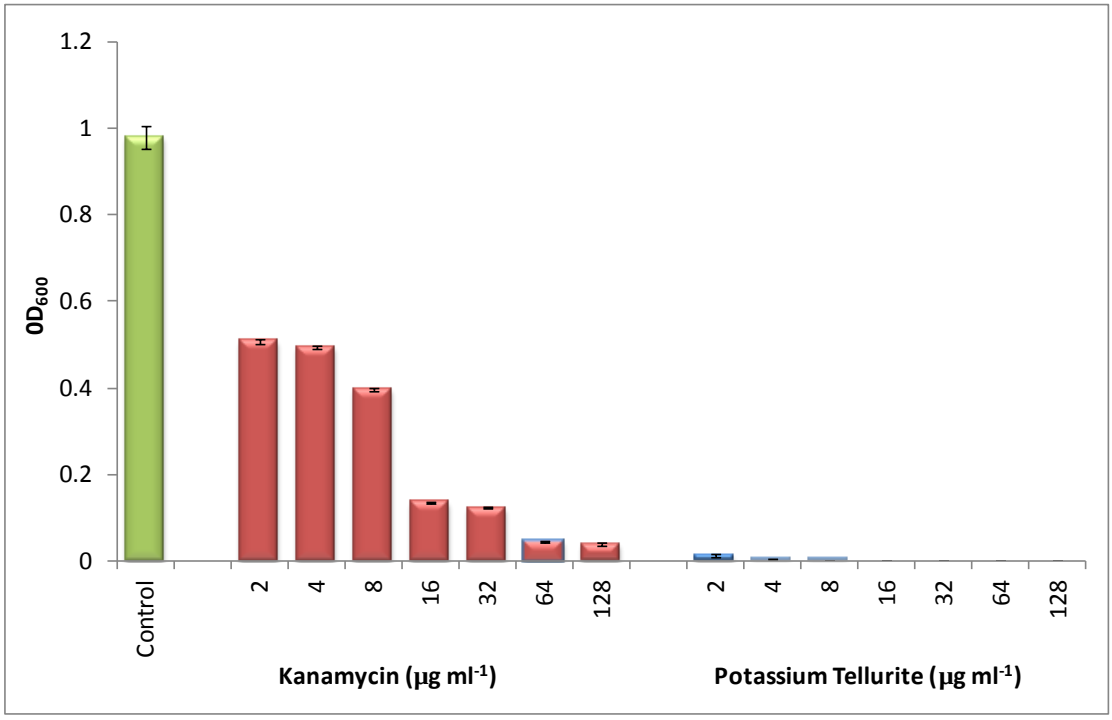

Figure 4.4: Minimum inhibitory concentration (determined using O.D$_{600}$) of kanamycin and potassium tellurite for *B. cepacia* ATCC 17759. Error bars show standard deviation of replicates (n = 3).

Potassium tellurite was seen to inhibit growth of *B. cepacia* ATCC 17759 to a greater level than kanamycin with almost no growth observed at the lowest concentration used (2µg ml^{-1}). In contrast growth was observed even at 128 µg ml^{-1} of kanamycin although there was a clear bactericidal effect with increasing concentration.

4.2.4 **Amplification of Kanamycin Resistance Gene from pKT230**

Amplification of Km^r for inclusion in the PCR ligation strategy (Figure 4.1) was from the incQ plasmid pKT230 (Km^r, Sm^r), kindly supplied by Prof. Petra Oyston (Dstl, UK). Plasmid DNA was transformed into HB101 *E. coli* Competent Cells (Section 2.4.12.6) and cultured overnight at 37 ^{o}C in LB broth containing 25 µg ml^{-1} kanamycin. Extraction of plasmid DNA was done as outlined in Section 2.4.4. Confirmation of pKT230 isolation was achieved using

*Bst*N1 (NEB, UK) restriction digestion (Section 2.4.12.1). Digests were visualised on a 2 % agarose gel (Figure 4.5). The restriction digest banding pattern was predicted using the restriction enzyme database (REBASE) (Roberts *et al.*, 2005).

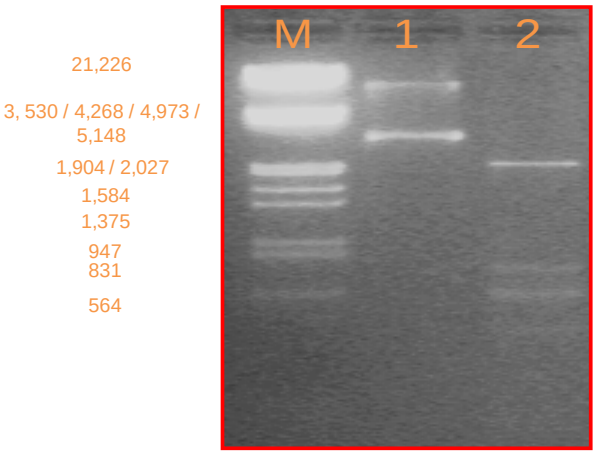

Figure 4.5: Agarose gel (2 %) showing intact (lane 1) and *Bst*N1 digestion (lane 2) of pKT230. M – Molecular weight marker III.

PCR Primers KmF and KmR-SphI, designed using Primer Express v. 1.0 and using standard PCR conditions (Section 2.4.19), were used to amplify the kanamycin resistance gene of pKT230. Results are shown in Figure 4.6. Correctly sized amplicons were excised and purified by gel extraction (Section 2.4.6).

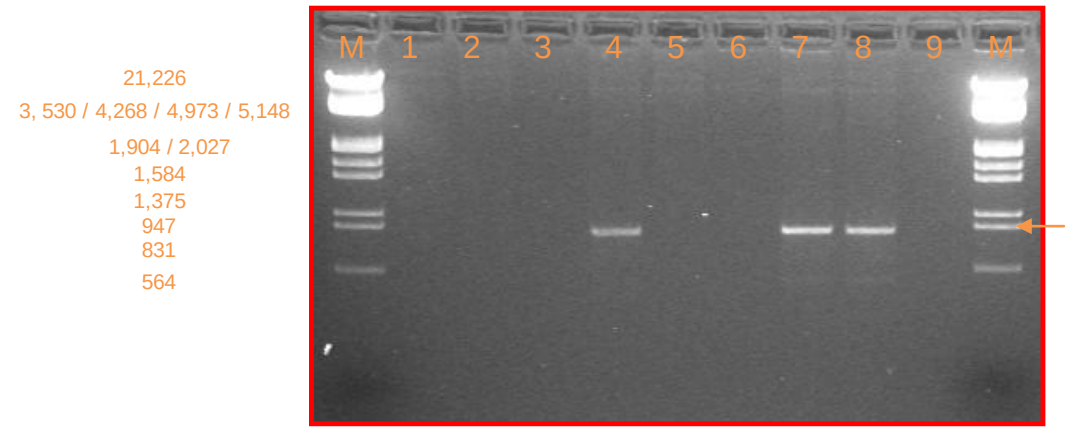

Figure 4.6: Agarose gel (1.5 %) showing optimisation of Km^r amplification from pKT230 using 2, 3 and 4 mM $MgCl_2$. Lane identities are as follows: M molecular weight marker III (Roche Applied Sciences, UK); 1 and 2 - 2mM $MgCl_2$; 4 and 5 - 3mM $MgCl_2$; 7 and 8 - 4mM $MgCl_2$; 3, 6 and 9 no template controls.

4.2.4.1 Amplification of Tellurite Resistance Operon from pDT1558

Amplification of Tel^r operon for inclusion in the PCR ligation strategy (Figure 4.1) was from plasmid pDT1558, kindly provided by Prof. Diane E. Taylor (University of Alberta, Canada). Primers TelF and TelR (Section 2.4.19), were used to amplify the ~3.0 Kbp potassium tellurite resistance cassette (*kilA*/*telAB*) of pDT1558. Results are shown in Figure 4.7. Correctly sized amplicons were excised and purified (Section 2.4.6).

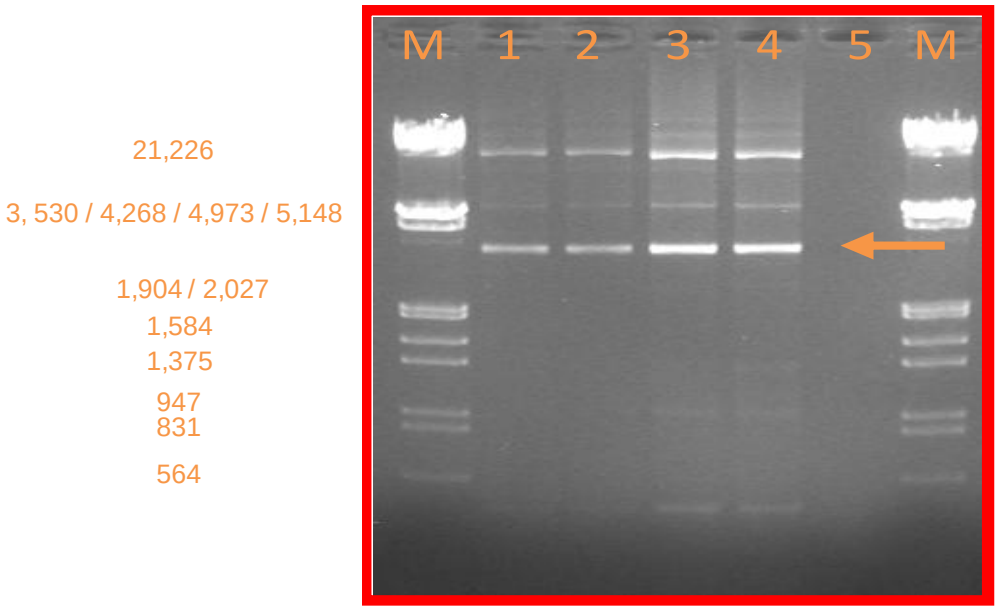

Figure 4.7: Agarose gel (2 %) showing amplification of *kilA*/*telAB* from pDT1558. Lane identities are as follows M molecular weight marker III; 1 to 4 – 1ng μl^{-1} pDT1558 plasmid DNA; 5 - no template control.

4.2.4.2 Generation of Ptac-Km^r and Tel^r Constructs

Selectable markers (Km^r or Tel^r) were cloned into plasmid pPW601a placing expression under the control of the strong hybrid promoter Ptac, which is composed of the -35 region of the trp promoter and the -10 region of the lacUV5 promoter/operator. Plasmid pPW601a (Figure 4.8) was kindly provided by Dr Peter White (Dstl, UK). Both the Km^r (of pKT230) and Tel^r (of pDT1558) were amplified with primers that introduced 5' *Xba*I and 3' *Sph*1 restriction sites for compatibility with cloning into pPW601a whereby the resistance determinants are under the control of the Ptac promoter. Plasmid pPW601a DNA was transformed into heat-shock competent *E. coli* JM109 and cultured overnight at 37 ^oC in LB broth containing 25µg ml^{-1} Ampicillin. Extraction of plasmid DNA was done as outlined in Section 2.4.4.

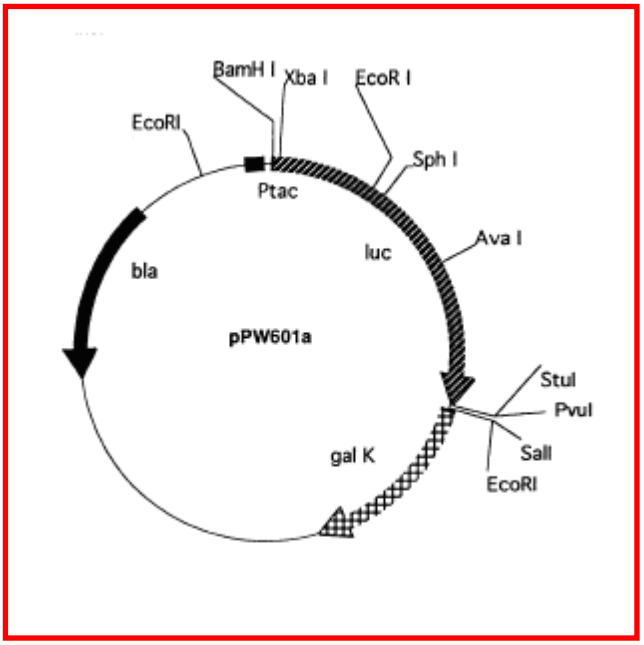

Figure 4.8: Plasmid pPW601a. Image reproduced with permission (White *et al.*, 1996).

Restriction digests with *Xba*1 and *Sph*1 were performed according to manufacturer's instructions (Section 2.4.12.1). Digestion of pPW601a revealed an unexpected fragment distribution (Figure 4.9). Sequence analysis using SeqBuilder (DNASTAR LaserGene 8) revealed that Sph1 was not a unique restriction site but occurred twice. The fragment of the expected size was excised and purified (Section 2.4.6).

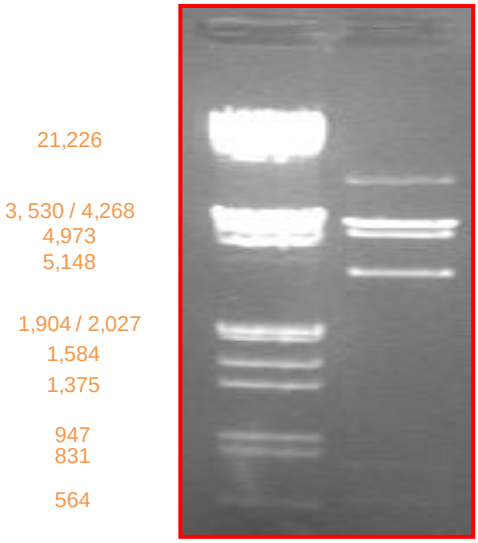

Figure 4.9: *Xba*1 / *Sph*1 Restriction Digest of pPW601a. M - Molecular weight marker III. Expected double digest fragment indicated by arrow.

Following dephosphorylation (Section 2.4.12.2) digested Km^r and Tel^r were ligated into pPW601a using T4 DNA ligase (Section 2.4.12.5). Transformations using HB101 *E. coli* Competent Cells (Section 2.4.12.6) were screened using LB agar plates supplemented with 25 µg ml^{-1} Ampicillin. No transformants were observed for Km^r ligations. Eight colonies were observed for Tel^r ligations. Putative transformants were sub-cultured onto LB agar plates supplemented with 25 µg ml^{-1} Ampicillin. Following overnight incubation at 37 ^oC, colony PCR was performed using primers TelF and TelR using amplification conditions outlined in Section 2.4.19. No amplification was observed.

4.2.5 **Custom Synthesis of Ptac-Km^r and Ptac-Tel^r Constructs**

In parallel to the PCR ligation strategy, the *zmpA* knock-out cassette was also sent to GeneArt® for synthesis. Only the construct containing the Km^r was synthesised. This was supplied in plasmid pGA1 (Figure 4.10).

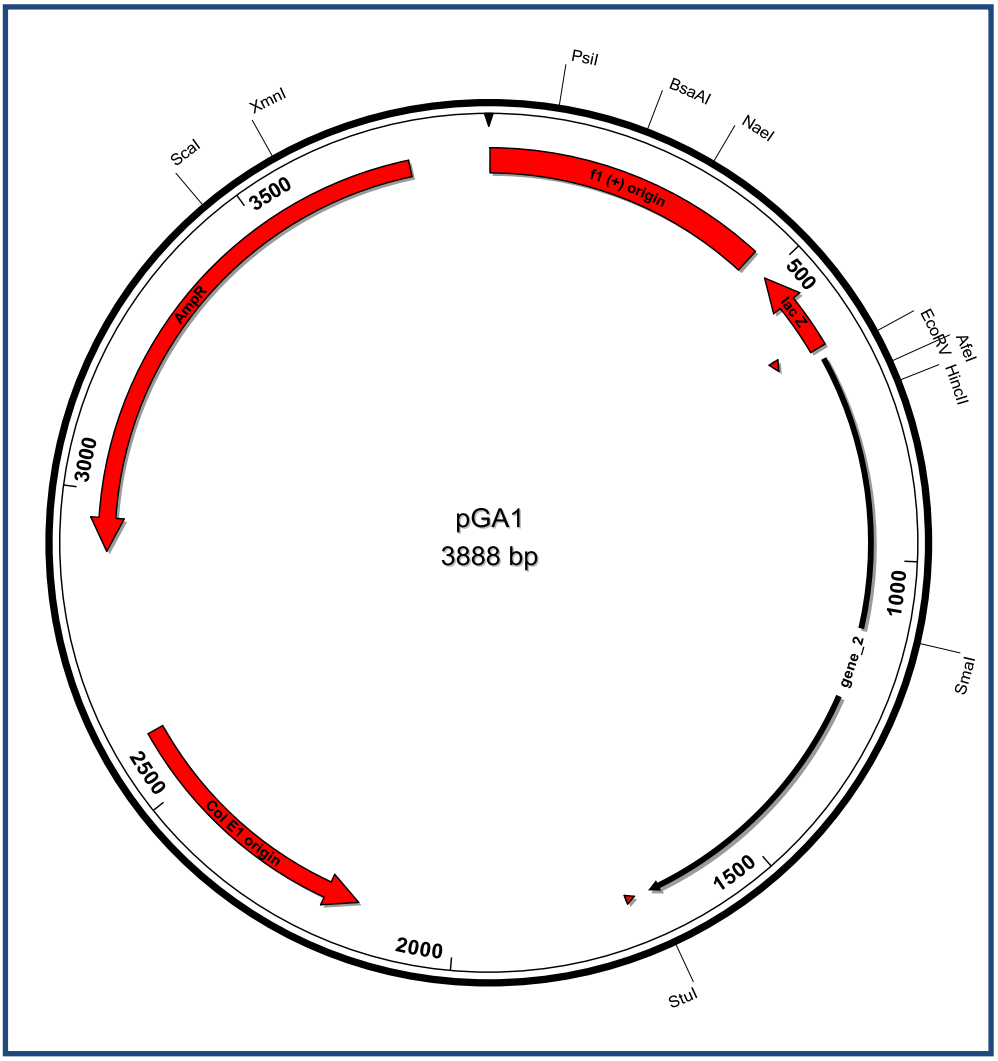

Figure 4.10: Circular Map of Plasmid pGA1

Plasmid pGA1 was transformed into HB101 *E. coli* Competent Cells (Section 2.4.12.6) and cultured overnight at 37 ^{o}C in LB broth containing 50 µg ml^{-1} ampicillin. Extraction of plasmid DNA was done as outlined in Section 2.4.4. Confirmation of pGA1 isolation was done using *Bst*N1 (NEB, UK) restriction digestion (Section 2.4.12.1). Digests were visualised on a 2 % agarose gel (Figure 4.11). The restriction digest banding pattern was predicted using the restriction enzyme database (REBASE) (Roberts *et al.*, 2005).

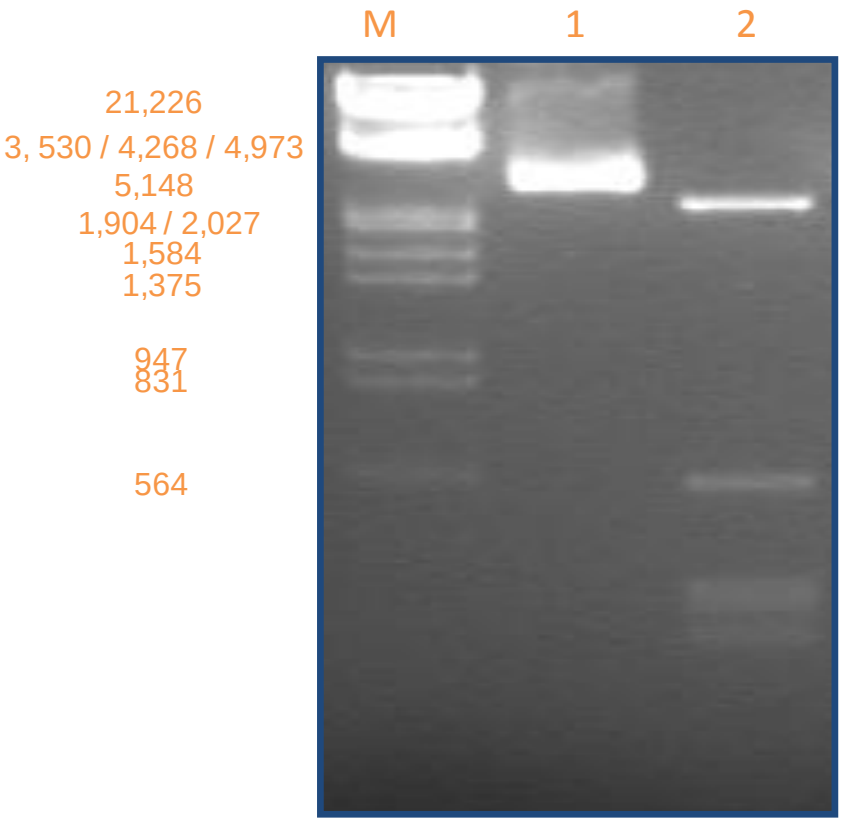

Figure 4.11: Agarose Gel (2 %) showing *Bst*N1 digest of pGA1. Lane identities are as follows: 1 - intact pGA1; 2 - *Bst*NI digested pGA1; M – Molecular Weight Marker III.

Expression of kanamycin resistance in pGA1 was checked using *E. coli* HB101 competent cells. Resistance was observed at 10 µg ml^{-1} kanamycin on LB agar plates with 40 µl of 100 mM Isopropyl-β-D-thio-galactoside (IPTG).

4.2.6 **Amplification of Construct from pGA1**

In order to ligate the *zmpA* kanamycin resistance construct of pGA1 into the suicicide vector pJQ200, primers GC-*Ac*1I and GC-*Eco*RI, designed using Primer Express v. 1.0 and using PCR conditions outlined in Section 2.4.19, were used to amplify the cassette. PCR products were analysed on a 2 % agarose gel (Section 2.4.5). Results are shown in Figure 4.12. Correctly sized amplicons were excised and purified (Section 2.4.6).

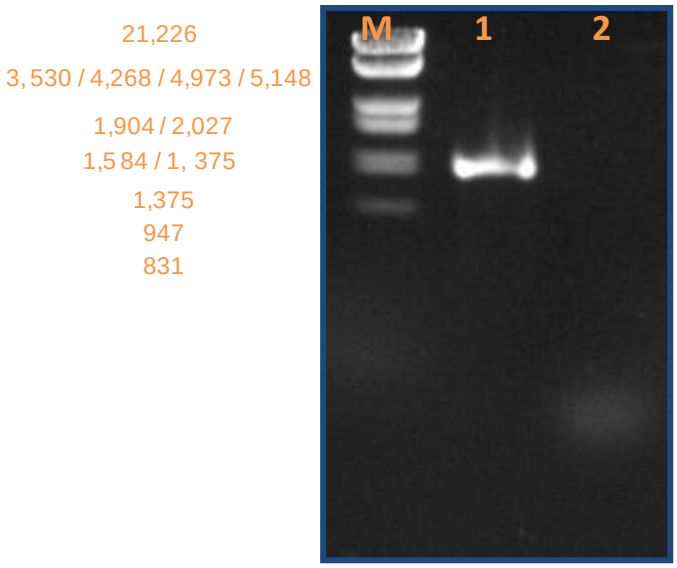

Figure 4.12: Agarose gel (2 %) showing amplification of *zmpA*-kanamycin resistance cassette of pGA1. M – Molecular weight marker III

Following gel extraction, the kanamycin cassette was double-digested with *Ac*I1 and *Eco*R1, dephosphorylated and ligated into the suicide vector pJQ200 (ATCC 77482 - Figure 4.13) (Section 2.4.12.3). Plasmid pJQ200 was first transformed into HB101 *E. coli* Competent Cells (Section 2.4.15) and cultured overnight at 37 ^{o}C in LB broth containing 15 µg ml^{-1} gentamicin. Extraction of plasmid DNA was done as detailed in Section 2.4.4. The site of insertion for the *zmpA-kanamycin* resistance cassette was *sacB*, which when functional confers sensitivity to sucrose in Gram negative bacteria (Gay *et al.*, 1985). The plasmid was double-digested with *Ac*I1 and *Eco*R1 and dephosphorylated (Section 2.4.12.3). Digests were visualised on an agarose gel (2 %) as shown in Figure 4.14.

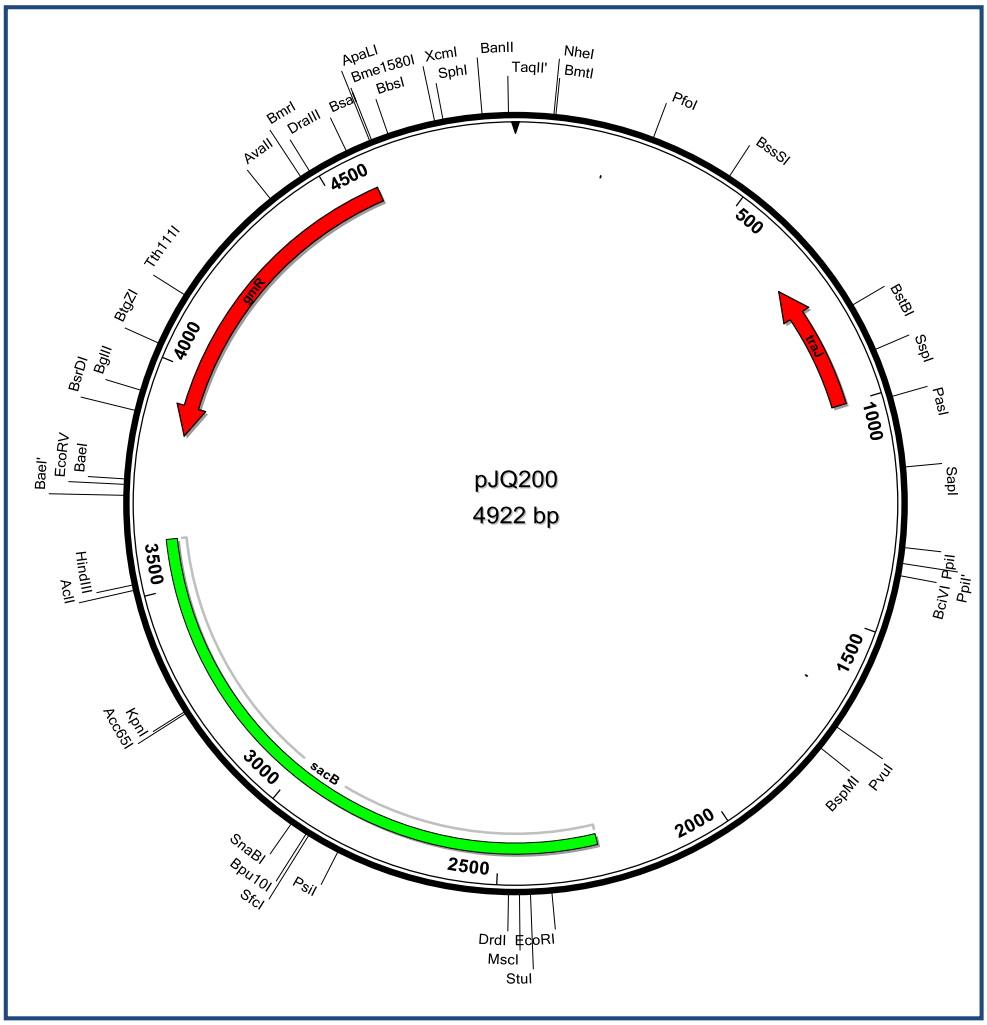

Figure 4.13: Circular Map of Plasmid pJQ200. Unique restriction sites are shown.

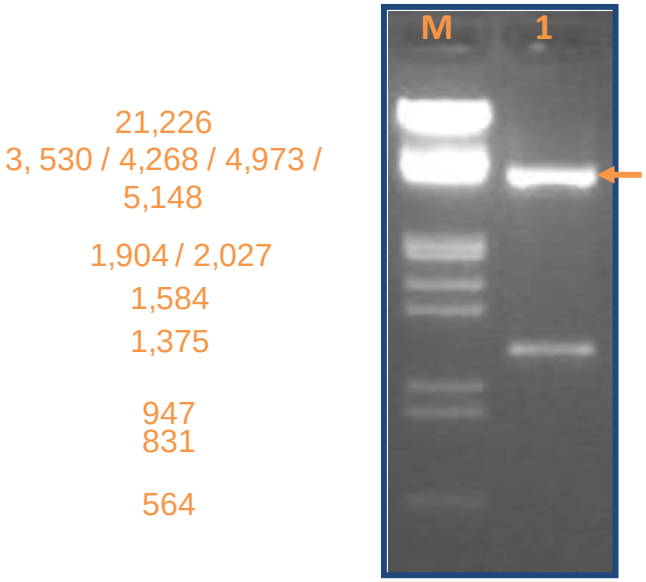

Figure 4.14: Agarose Gel (2 %) showing *Ac*I1 and *Eco*R1 digest of pJQ200. Lane identities are as follows: *Ac*I1 and *Eco*R1 digested pJQ200; M – Molecular Weight Marker III.

Bands representing the larger fragment, where *sacB* had been removed, were excised and purified as detailed in Section 2.4.6. Overnight ligations using T4 DNA ligase were done using *Ac*I1/*Eco*R1 digested construct from pGA1 and pJQ200 (Section 2.4.12.5). Transformations using HB101 *E. coli* Competent Cells (Section 2.4.12.6) were screened using LB agar plates supplemented with 15 µg ml^{-1} Gentamicin. No colonies were observed following overnight incubation at 37 ^{o}C.

4.2.7 **Ligation of *zmpA*-Km^{r} Construct from pGA1 into pTrcHIS-TOPO®**

Following observations by Fehlner-Gardiner and Valvano, (2002) that Ptac promoters were not functional in *B. vietnamiensis*, the kanamycin resistance gene from pKT230 was ligated into the TA cloning vector pTrcHis-TOPO® (Section

2.4.12.6). Here expression of kanamycin resistance was under the control of the Ptrc promoter. Cloning was done according to the manufacturer's instructions. Ligations were transformed into One-shot® TOP10 *E. coli* as detailed in Section 2.4.12.7. Transformants were selected for on LB agar plates supplemented with ampicillin (50 µg ml^{-1}) and glucose (0.5 %). Following overnight incubation the resultant 80 colonies were subcultured onto LB agar plates as above and screened through colony PCR using primers KmF and KmR-*Sph*1 targeting the kanamycin resistance gene. Two colonies were shown to harbour Km^r and showed resistance to kanamycin (25 µg ml^{-1}). Amplification of Ptrc-Km^r was done using primers KmC6F and KmCR and cycling conditions detailed in Section 2.4.19. PCR products were analysed on a 2 % agarose gel as shown in Figure 4.15. Primer KmC6F introduced an *Acc*651 restriction site at the 5' end of Ptrc-Km^r for subsequent ligations into pGA1. The removal of 5' single base (A) overhangs introduced by *Taq* DNA polymerase was achieved by treatment using DNA polymerase I (Klenow) large fragment (Section 2.4.12.2).

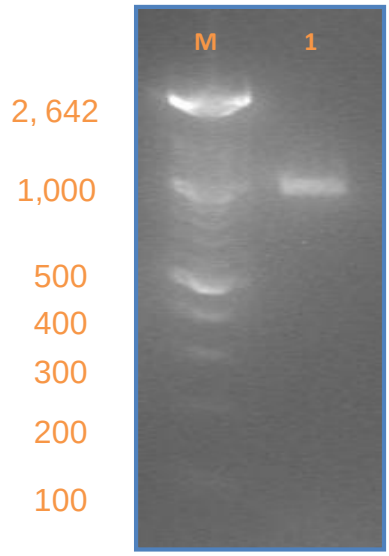

Figure 4.15: Agarose gel (2 %) showing amplification Ptrc-Km^r cassette of pTrc-Km^r. M – Molecular weight marker XIV

The Ptac-Km^r portion of the Km^r cassette was excised from pGA1 through digestion with *Acc*651 and *Hinc*II (Figure 4.16). The larger fragment was excised and purified (Section 2.4.6). Both Ptrc-Km^r and the pGA1 backbone were dephosphorylated (Section 2.4.12.2) prior to ligations using T4 DNA ligation (Section 2.4.12.5).

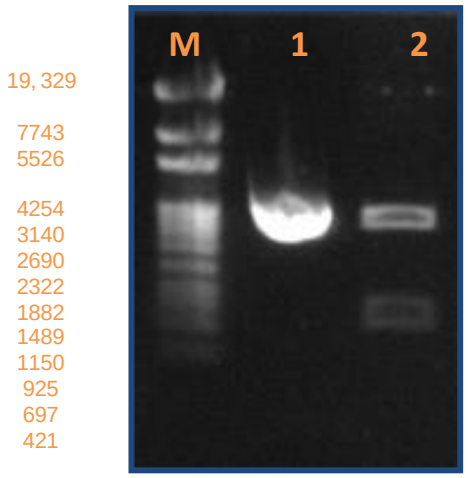

Figure 4.16: Agarose gel (2 %) showing *Acc*65I / *Hinc*II Digest of pGA1. Lane identities are as follows: 1 un-digested pGA1; 2 - double digested pGA1. M - DNA Molecular Weight Marker IV

Ligation reactions were transformed into HB101 *E. coli* Competent Cells (Section 2.4.12.6). Transformations were screened using were screened both using LB agar plates supplemented with 50 µg ml^{-1} ampicillin and agar plates with ampicillin and 25 µg ml^{-1} kanamycin. No colonies were observed. Owing to time constraints within this study, no further efforts were made to construct a selectable *zmpA* knock-out cassette.

4.3 **Sequence-independent (Non-homologous) Transformation**

To determine the potential for Bcc members to undergo transformation in the absence of homologous sequences, filter transformations were done using the *mob*$^{+}$/*tra*$^{-}$ plasmid pHKT2 following the method of Ray and Nielsen, (2005) (Section 2.4.25). This plasmid was kindly donated by Dr Howard Ceri, University of Calgary, Canada. The plasmid had previously been used in the study of Bcc biofilms and therefore its expression and stability was well characterised (Tomlin *et al.*, 2004). Additionally, the utility of trimethoprim as a selectable marker was apparent given the relatively low levels of resistance observed throughout members of the Bcc (Nzula *et al.*, 2002).

Triplicate filter transformations were done with *B. stabilis* 14086, *B. multivorans* 13010 and *B. cepacia* 17759. These strains were selected based on their sensitivity to trimethoprim with MICs of 2, 2, and 4 µg ml^{-1} respectively (Nzula *et al.*, 2002). Cells were cultured in defined minimal media (Section 2.2.1). Control transformations where no plasmid DNA was present were also done at the MIC for each strain. Results are shown in Table 4.3. No colonies were observed on LB agar plates containing trimethoprim concentrations above the MICs for *B. stabilis* or *B. cepacia*. For *B. multivorans* 13010, sixty-three colonies were observed at a trimethoprim concentration of 8 µg ml^{-1} suggesting that pHKT2 had been taken up and expressed within this organism.

Table 4.3: Increased resistance to trimethoprim in *B. stabilis* LMG 14086, *B. multivorans* 13010 and *B. cepacia* ATCC 17759 following filter transformations with pHKT2. ND – not determined; Tm – trimethoprim

Bcc Strain (MIC for Tm)	Transformation 1	Transformation 2	Transformation 3	Control	Tm (μg ml^{-1})
B. stabilis LMG 14086 (MIC 2μg ml^{-1})	Confluent	Confluent	Confluent	ND	1
	0	1	0	4	2
	0	0	0	ND	4
B. multivorans 13010 (MIC 2μg ml^{-1})	Confluent	Confluent	Confluent	ND	1
	0	0	25	0	2
	0	0	63	ND	4
B. cepacia ATCC 17759 (MIC 4μg ml^{-1})	Confluent	Confluent	Confluent	ND	2
	0	0	0	0	4
	1	0	0	ND	8

Each of the sixty-three colonies from LB agar plates supplemented with 8 μg ml^{-1} trimethoprim were sub-cultured and then subjected to colony PCR using primers GfpF and GfpR using cycling conditions outlined in Section 2.4.26. This assay was designed to amplify the green fluorescent protein gene of pHKT2. Amplicons were visualised using an E-gel (2 %) as detailed in Section 2.4.12 as seen in Figure 4.17.

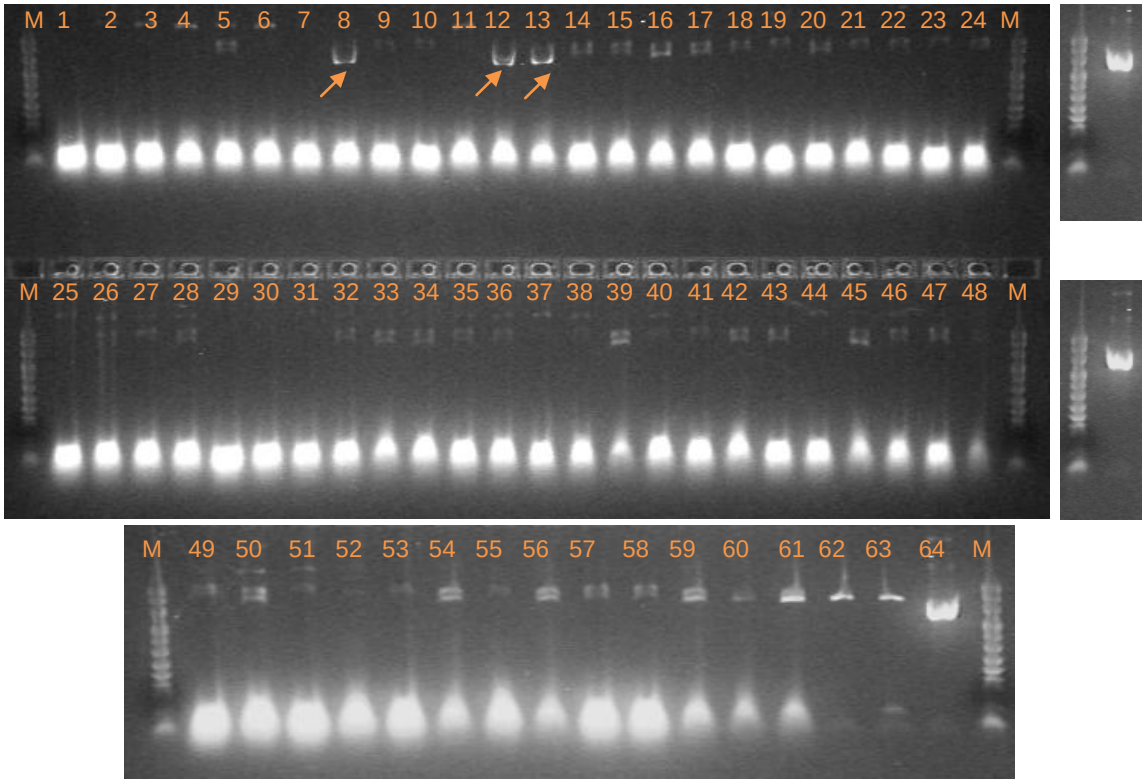

Figure 4.17: E-gel (2 %) of colony PCR amplification of GFP in *B. multivorans* 13010 putative pHKT2 transformants. Lane identities are as follows: 1 to 63 - results of GFP PCR amplification reactions performed on single colony, putative *B. multivorans* mutants cultured on LB agar supplemented with 8 µg ml^{-1} trimethoprim; 64 - PCR positive control using 10 ng μl^{-1} pHKT2 M – Molecular Weight Marker VIII. Amplicons of the correct size are indicated by the orange arrows.

Of the sixty-three colonies that were subjected to PCR for the presence of the green fluorescent protein gene three were putatively positive for the presence of that gene. No further evaluations were done on these mutants as a consequence of study time constraints.

4.4 Discussion

As it requires the presence of extracellular DNA, here the focus of understanding horizontal gene transfer within the Bcc was on the process of transformation. This mechanism is widely distributed amongst bacteria (Johnsborg *et al.*, 2007) however as discussed only the closely related strains of *B. pseudomallei* and *B. thailandensis* have been shown to possess the capacity (Thongdee *et al.*, 2008). To determine if members of the Bcc were capable of the uptake, integration and expression of extracellular DNA two experimental approaches were adopted - the generation of defined knock-outs within a zinc mettaloprotease (*zmpA*) with subsequent allelic rescue and uptake of non-homologous plasmid pHKT2.

Zinc mettaloprotease was an attractive target for mutagenesis owing to its presence in *B. cenocepacia* as the most common aetiological agent in cystic fibrosis Bcc infections (Speert *et al.*, 2002). Additionally *zmpA* was well conserved amongst some of the other genomovars although here spurious results for its amplification were observed. Regardless of this only one strain, *B. cepacia* J2540 failed to produce amplification of *zmpA* when tested with either PCR assay. Furthermore, the regulation of *zmpA* through both CepIR and CciIR quorum sensing systems has been well characterised (Kooi *et al.*, 2006) and would not likely be involved in any hypothesised mechanisms of DNA uptake. The chosen method for generating gene-specific mutant i.e. allelic exchange on a compatible suicide vector, was one that had been used previously for the generation of chromosomal knock-outs in *Yersinia pestis* (Derbise *et al.*, 2003). Although a construct was generated, the observations regarding the poor levels of expression resulting from the use of Ptac promoters in *B. vietnamiensis* (Fehlner-

Gardiner and Valvano, 2002) necessitated the adoption of an alternative promoter system. Although several components for the construct were generated, including expressed kanamycin resistance under the control of Ptrc, final synthesis was not possible within the timeframes of this study. Whilst still a useful approach, there have been developments in the use of the Phage λ-Red proteins for the generation of knock-out mutants in those strains of *B. pseudomallei* and *B. thailandensis* that had previously been shown to be naturally transformable (Thongdee *et al.*, 2008; Kang *et al.*, 2011). This system had, until recently, been limited to the γ-proteobacteria but now represents a useful tool for the Burkholderiaceae. The system involves the amplification of a selectable cassette e.g. one encoding antibiotic resistance, with one region of homology to the target gene (>30 bp in most cases) generated through primer extension. A suicide vector, such as pJQ200 used in this study, is then used to introduce the construct into the target cell where homologous recombination is achieved through the induction of the lambda phage Red operon. The Red operon encodes the nuclease inhibitor Redγ(*gam*) and the site specific recombinases Redα(*exo*) and Red β(*bet*) (Shulman *et al.*, 1970). As well as utility within *Burkholderia* sp. this system has also been used for the rapid generation of mutants of *Pseudomonas aeruginosa* (Lesic and Rahme, 2008). Any future efforts for the construction of mutants within the Bcc should ideally make use of these developments.

The second method for determining the transformation potential of Bcc members was filter transformations using the method of Ray and Nielsen, (2005) with a non-homologous plasmid that had previously shown utility with respect to Bcc

strains (Tomlin *et al.*, 2004). Here one strain, *B. multivorans* 13010, was shown to express an increased level of resistance to trimethoprim following co-incubation with pHKT2. In addition, it was possible to amplify the green fluorescent protein gene that was also located on pHKT2. Both these findings are highly suggestive of a transformation event.

However, a number of issues remain to be investigated. For example, it was not possible within the time frame of this study to determine whether the plasmid had integrated with the chromosome or existed as a replicating element within the cell. It should be noted that the frequency of uptake, integration and expression of heterologous DNA has been shown to be highly restricted in both *P. stutzeri* and *Ralstonia* sp. when compared to homology-facilitated chromosomal recombinations (Lorenz *et al.*, 1998; Mercier *et al.*, 2007). Consequently, the likelihood that this represents a chromosomal insertion in the absence of sequence homology is low. It is also possible the plasmid reformation was prevented by restriction-modification systems that inherently protect bacterial cells from foreign DNA. However, as pHKT2 was previously shown to be stable, without selection, for three days in Bcc members (Tomlin *et al.*, 2004) within the short timescales used here this would seem an unlikely outcome. Examining the expression of the green fluorescent protein would also have added additional support to the observation that the increased resistance to trimethoprim was indeed a result of pHKT2 uptake and expression. If confirmed, the finding that not all members of the Bcc are capable of transformation to the same degree mirrors the situation in *P. stutzeri*. Like the Bcc, genotypic fingerprint analysis of *P. stutzeri* has demonstrated a similar taxonomic substructure with division

into seven genomovars (Lorenz and Sikorski, 2000). These authors examined transformability across these genomovars demonstrating five (A, D, E, F1 and H) had at least one competent member. However, the transformation frequency varied by more than three orders of magnitude. Furthermore, the frequency of transformation using Rif^r mutant DNA was considerably lower between genotypic groups with some strains exhibiting reduced frequencies even with DNA from strains of the same genotypic group (strain LO179; group H) a facet of sexual isolation of strains (Lorenz and Sikorski, 2000). It was unclear, however, if strains that did not exhibit Rif^r conversion were inherently non-transformable or whether the conditions under which they would transform were not provided (Lorenz and Sikorski, 2000).

This study has produced possible evidence that at least one member of the Bcc, *B. multivorans* 13010, are transformable by free DNA. However, this result must be taken in the context of the further research required to verify this finding. Ultimately, the demonstration of transformation using homologous sequence driven allelic rescue would provide the most compelling evidence for transformation potential, as was the case for *B. pseudomallei* and *B. thailandensis*, and therefore the use of alternative knock-out systems should be investigated.

Chapter 5

Comparing 16S rRNA Variable Region Sequencing for Microbial Community Profiling of Long-term Antibiotic Amended Soil

5.1 **Introduction**

5.1.1 **Soil Microbial Diversity**

Soil represents arguably one of the most diverse environments of prokaryotes (Roesch *et al.,* 2007), with one gram of soil estimated to contain up to 10 million microorganisms and 10^3 to 10^7 different species of bacteria (Curtis *et al.*, 2002; Torsvik *et al.*, 2002; Gans *et al.*, 2005; Schloss and Handelsmann, 2006). Microorganisms active in the soil are largely responsible for the biogeochemical cycles that support life on earth (Paul and Clark, 1996). For example, they mediate the nitrogen cycle and play a central role in both carbon and sulphur cycles; thus they are crucial for both soil function and health (Tiedje *et al.*, 1999). Consequently, impacts on the survival of introduced pathogenic species in the presence of well established complex microbial communities are likely to influence the survival of introduced pathogenic species.

5.1.2 **Molecular Fingerprinting Techniques for Assessing Soil Microbial Diversity**

Given the importance of prokaryotes to soil function, there has long been a requirement for molecular methods that enable the interrogation of that community. These are aimed at determining the structure, assessing the diversity and elucidating the responses to changes in environmental conditions. To achieve this there are methods available that target either the 16S rRNA gene purely for taxonomic profiling or functional genes giving more targeted insights into community activity. Frequently used methods include denaturing gradient gel electrophoresis (DGGE), terminal restriction fragment length polymorphism

(T-RFLP) and single strand conformation polymorphism analysis (SSCP) (Muyzer *et al.*, 1993; Liu *et al.*, 1997; Schweiger and Tebbe, 1998).

More recently, metagenomic approaches that allow the total gene content of a soil to be determined have become viable alternatives and are now being implemented in initiatives such as the Terragenome project (http://www.terragenome.org) - an international consortium whose goal is to produce a reference soil metagenome. These approaches utilise significant advances in DNA sequencing technology and now enable, through the interrogation of the total gene content of the soil, simultaneous insights into community structure and function. This approach however remains highly laborious and costly, particularly where ones objectives are only for the examination of the profile of a microbial community or the determination of specific responses to particular environmental changes. Thus the use of 16S rRNA gene sequencing, in combination with the increased output enabled by current next-generation sequencing platforms, has been widely exploited for many varied environments (Charlson *et al.*, 2010; Jesus *et al.*, 2010; Bowman *et al.*, 2011; Santos *et al.*, 2011).

5.1.3 **The 16S rRNA as a Tool for Phylogenetic Analysis of Microbiomes**

The small subunit (SSU) ribosomal RNA encoding genes, of which the 16S rRNA gene is part, represent common universal genetic markers for the analysis of microbiomes. Ribosomes are large ribonucleoprotein complexes that are the site of protein synthesis (Doolittle, 1999). In prokaryotes these complexes consist of two subunits, the 50S subunit contains a 23S rRNA, 5S rRNA and over

30 proteins. The 30S subunit contains the 16S rRNA and 20 proteins (Pei *et al.*, 2009).

The diversity of microbiomes that have been analysed using high-throughput 16S rRNA sequencing approaches is testament to the utility of the method. Diverse intestinal microbiota from canines (Suchodoloski *et al.*, 2009) and humans (Stearns *et al.*, 2011), skin (Fierer *et al.*, 2008) and wounds (Dowd *et al.*, 2008), environments such as air (Bowers *et al.*, 2011), water (Bowman *et al.*, 2011) activated sludge (Zhang *et al.*, 2011) and, of course, soil (Deng *et al.*, 2012) have all been analysed using this approach. These studies have led towards greater understanding of the impacts of climate change on microbial diversity, discovery of phyla that can be assessed for the production of novel enzymes and an increased awareness of the impact of host prokaryotic communities in both health and disease.

Bacterial 16S rRNA genes contain nine 'hypervariable regions' which have been the basis for the taxonomic clarification of a number of bacterial species (Van de Peer *et al.*, 1996). The secondary structure of the 16S rRNA along with the location of these regions is shown in Figure 5.1. The design of primers for PCR amplification of the 16S rRNA gene is aided by the presence of conserved flanking sequences proximal to each hypervariable region. Once generated, 16S rRNA sequences are clustered into operational taxonomic units (OTUs) at a number of predetermined distance levels (related to percentage sequence similarity) that currently define taxonomic boundaries. These are 0.2 (80% sequence similarity) for Phylum level assignments, 0.1 (90%) for Family, 0.05

(95%) for Genus and 0.03 (97%) for Species. It is these OTUs that are then used to determine species richness, diversity, composition and structure of a microbiome.

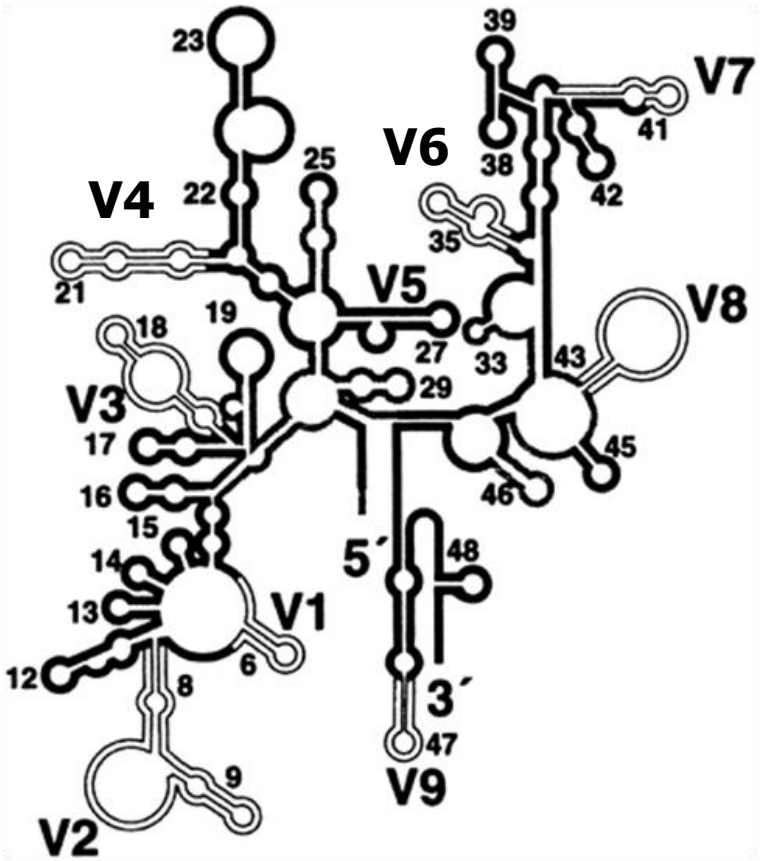

Figure 5.1: Secondary-structure model of the 16S rDNA. Double lines represent hypervariable regions with major variable regions V1 – V9 shown (taken from Tortoli, 2003).

5.1.4 **The Impact of Variable Region Selection**

An issue of note for existing taxonomic cut-offs is that they are based on analysis of full-length (~1540 bp) sequence (Schloss and Handlesmann, 2004). In light of the much shorter read lengths generated with high-throughput sequencing technologies, the accuracy of OTU assignment has therefore been re-examined.

Youseff *et al.,* (2009) compared species richness estimates and OTU assignment from 1132 16S rRNA clones from a soil sample with simulated shorter read outputs. Amplicons that cover V4, V5-V6 and V6-V7 were found to produce comparable results to full length analyses. Liu *et al.,* (2007) demonstrated fragments of 250 bp around the V1-V2, V3 and V4 region provided the best targets in this context. In a similar study Wang *et al.* (2007) demonstrated similar utility through the use of V2 and V4 regions. In one of the most thorough re-examinations of its type Kim *et al.,* (2010) compared 887 full-length sequences (>1458 bp) to a number of partial 16S rRNA sequences ranging in size from 450 to 700 bp and delineated by commonly used domain specific primers. The authors used OTU richness, accuracy of OTU clustering and microbial community structure metrics to compare the reliability of the partial fragments in comparison to full-length. The greater bacterial sequence divergence previously observed for variable regions one to four (Yu and Morrison, 2004) resulted, unsurprisingly, in the consistent overestimation of microbial diversity when compared to full-length sequences. However, using UniFrac significance and P test comparisons ($p \geq 0.25$) of depicted microbial community structure no significant differences to full-length characterisations were observed for any of

the partial sequences examined. The authors further elucidated the most accurate distance level for partial fragments by comparing diversity metric outputs at 0.02, 0.03 and 0.04 for OTU species classification. Here, the suggested distance level for upstream fragments (consisting of variable regions 1-4) that provided comparable assignments to 0.03 for full-length was 0.04 (96%). For downstream (variable regions 6-9) this was 0.02 (98%). The use of variable regions 1-4 provide an additional advantage in that phylogenetic resolution should be better than downstream regions owing to the greater sequence divergence (Kim *et al.*, 2010).

5.1.5 **The use of 16S rRNA High-throughput Sequencing for Soil Microbial Diversity Analysis**

5.1.6 **Antibiotics, Antibiotic Resistance and Soil**

The spread and evolution of antibiotic resistance genes, particularly their acquisition by bacterial pathogens, is becoming one of the most important clinical challenges (Arias and Murray, 2009). Although the exposure of humans to resistance-harbouring pathogens from the soil has yet to be fully determined it is likely to be an underestimated phenomenon (Heuer *et al*., 2011). It is known that, within the European Union, the level of total antibiotics applied per kg of meat product ranges from <20 to 188 mg of which 30 – 90% is excreted with manure to soil (Heuer *et al.*, 2011). The application of this antibiotic contaminated manure is common practice in Europe as well as the USA and other parts of the world and the transfer of sulfamethazine, chlorotetracycline and tylosin to agricultural soils in this manner has been shown (Halling-Sørensen *et al.*, 2005; Shelver *et al.*, 2010). Consequently, a range of studies have

demonstrated a link between antibiotic-amended manure and levels of resistance in soil (Heuer *et al.*, 2008; Heuer *et al.*, 2011; Knapp *et al.*, 2008). The application of manure deposits a considerable quantity of bacteria carrying antibiotic resistance determinants; even if the manure is derived from animals with no history of treatment with antibiotics. As a consequence it can be difficult to delineate the selective effects of persistent antibiotics in soil on native microflora, although it has been shown that introduced ribotypes decline within months of application.

5.2 **Aims**

The aim of this chapter was to establish the most appropriate method for the analysis of complex microbial communities in soil. The focus here was the most numerically abundant lineage, bacteria. It is likely that a released pathogenic microorganism, such as a BWA, will have a noticeable effect on the microbial community of the environment into which it is introduced. Whilst this may not extend to the long-term colonisation of that niche by the BWA, it is plausible that the persistence of nucleic acids, extracellularly or through the transfer of genetic elements to the existing community, could cause a shift in the observable community structure. Where analyses of such communities might use 16S rRNA analysis, it is important to determine the plausablity of such an approach and the impact of variable region choice. Here, this was done through the characterisation of changes in a soil microbial community under antibiotic selection. The rationale was that these samples represent a shifted microbial community and one that may have experienced HGT in response to the selective pressure of antibiotic treatment. Previous work by Gaze *et al.*, (unpublished data) suggested that antibiotic selection had a subtle but measurable effect on

prevalence of a specific mobile genetic element in the soil samples chosen. Thus the hypotheses here are that antibiotic selection has a measurable effect on community diversity as measured by 454™ sequencing and subsequent sequence analyses and that different 16S rRNA variable region produce similar estimates of diversity.

5.3 **454™16S rRNA Amplicon Assays**

As described previously, targeting partial 16S rRNA sequences (based on variable regions V1 through to V9) allow accurate reconstructions of microbial community structure when compared to the use of full length sequences. This has, however, been assessed using alpha and beta diversity metrics and may not reflect species-level accuracies of taxonomic assignment. Where this is particularly pertinent is where 16S rRNA is utilised as a tool in clinical microbiology or for pathogen detection in complex communities such as those found in soil. Here, for the identification of pathogenic species, including discrimination from closely relatives, it is important that the choice of target will enable accurate taxonomic assignment.

5.3.1 ***In silico*** **Assessment of Hypervariable Regions**

Chakravorty *et al.*, (2007) examined each variable region in isolation for their discriminatory capacity against a panel of 104 species that included common blood-borne pathogens, CDC-defined select agents as well as some environmental microflora. Using their sequence panel, a comparison was made as to the discriminatory powers of regions V1 - 3 versus V4 - 6 against the full length 16S rRNA sequence. In the context of identifying biological warfare agents and, more importantly closely related but non-pathogenic near-neighbours, a number of notable absences from this database were observed and corrected. These included *Francisella philomiragia*, *Burkholderia thailandensis*, certain *Bacillus cereus* strains such as G9241 and ATCC 14759, *Ochrobactum anthropi* and *Yersinia pseudotuberculosis*. These represent either opportunistic human pathogens, or in the case of *B. cereus* G9241, novel strains containing

pathogenicity determinants homologous to that of the near-neighbour strains (Hoffmaster *et al.*, 2004). Additionally, *Bacillus anthracis* sequences originally used did not capture the diversity of this organism and so a 16S rRNA consensus was made using *B. anthracis* AMES, Sterne and Vollum strains. To generate consensus sequences for these organisms, full length 16S rRNA sequences were first imported from GenBank into MegAlign (LaserGene 8). ClustalW (slow / accurate) was then used to align sequences and generate a consensus. The consensus sequences were then introduced into the larger data-set for use in further analyses. In total 110 sequences, including those used by Charkravorty *et al.*, (2007) and the consensi generated for the near-neighbour strains, were used to compare the efficacy of discrimination of the chosen two partial gene regions (V1 - 3 and V4 - 6) against the full length gene. This was achieved through generating alignments using ClustalW, trimming the full length sequences at recognised primer sites and comparing the sequence similarity dendrograms (Figures 5.2 – 5.4).

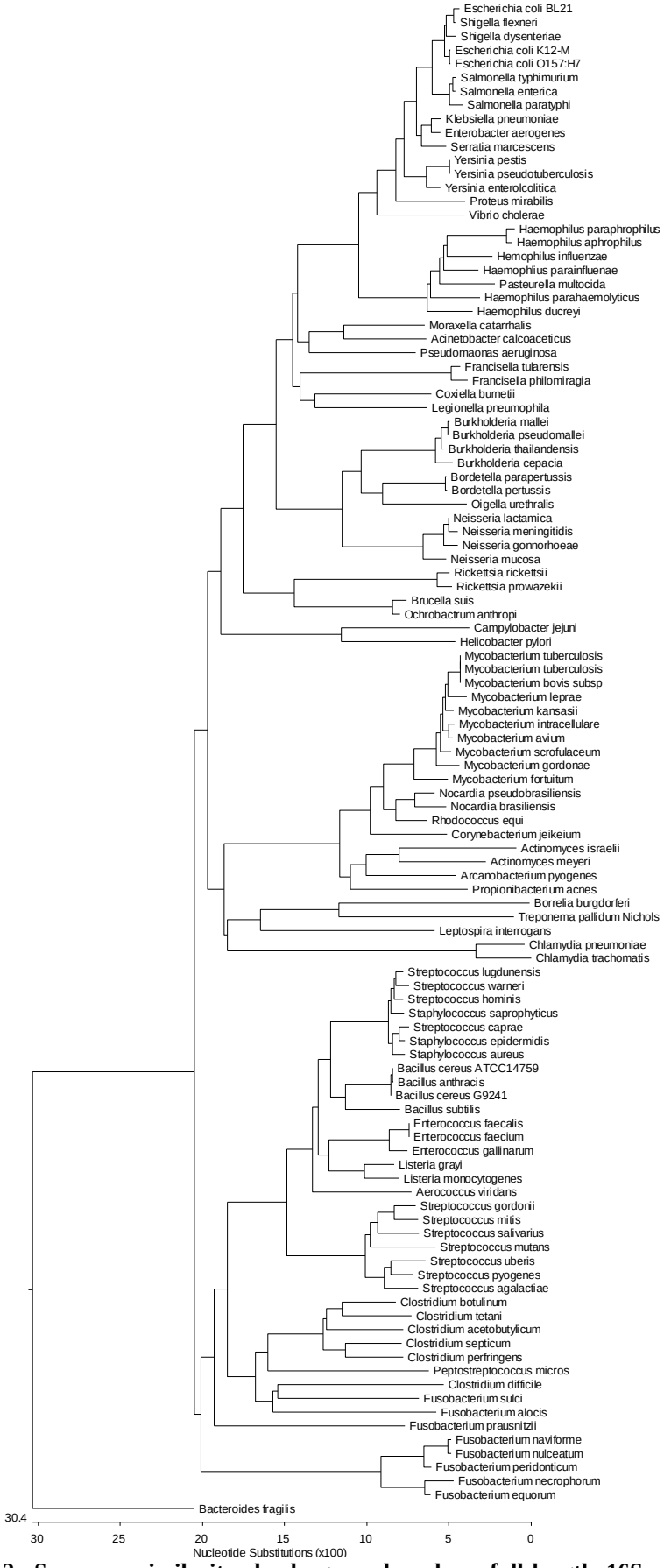

Figure 5.2: Sequence similarity dendrogram based on full-length 16S rRNA ClustalW alignments of 110 pathogenic species

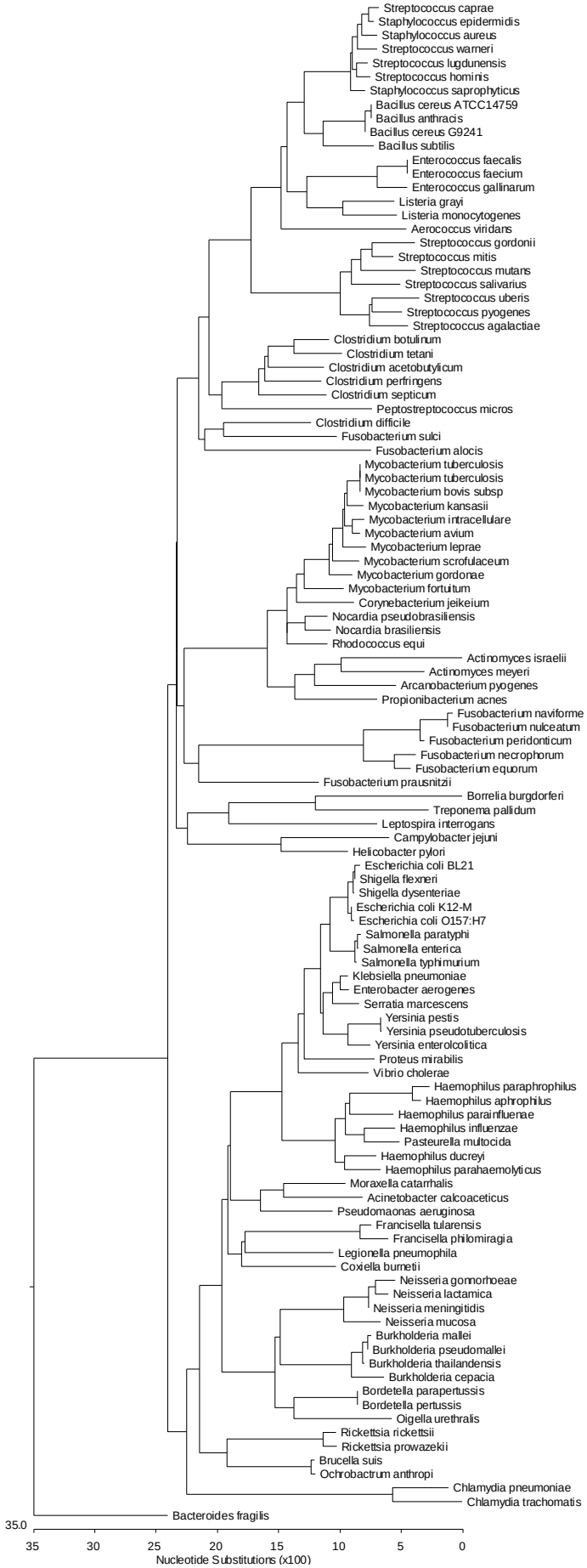

Figure 5.3: Sequence similarity dendrogram based on comparative ClustalW alignments for 16S rRNA V1 – 3 regions from 110 pathogenic species

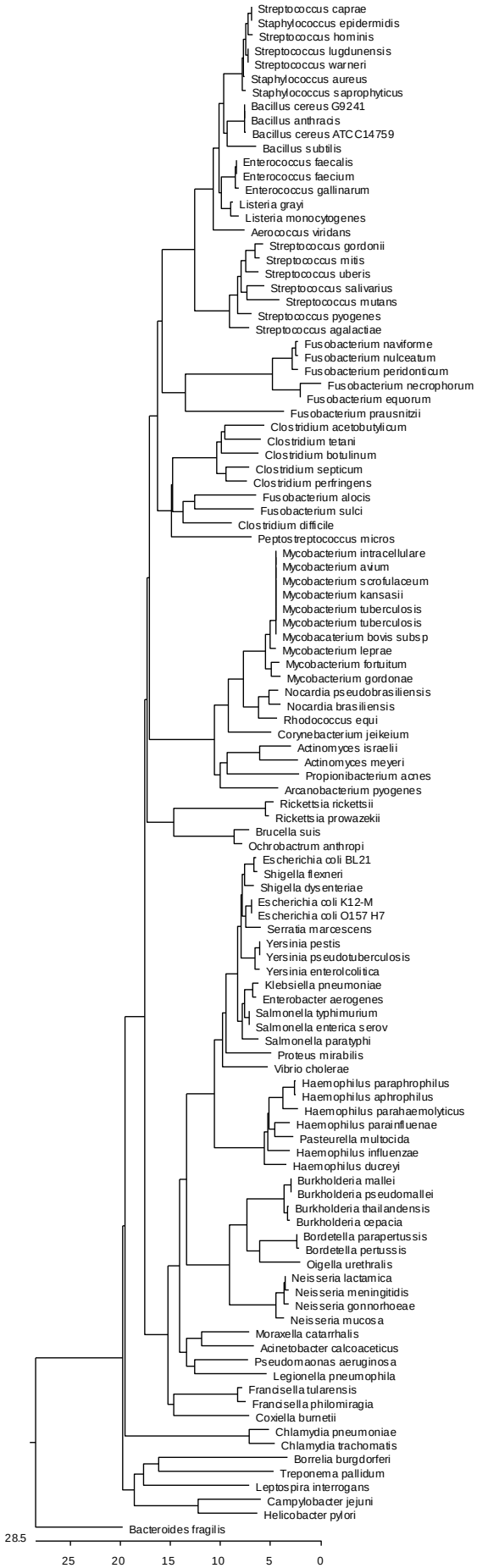

Figure 5.4: Sequence similarity dendrogram based on comparative ClustalW alignments for 16S rRNA V4 – 6 regions from 110 pathogenic species

Overall, there was good agreement between the dendrograms generated for both V1-3 and V4-6 to that of the full-length analysis. As expected, the sequence divergence of V1-3 appeared greater than V4-6 (Yu and Morison, 2004). However, there are a number of disagreements, particularly within the phylogeny generated by the V4-6 regions that are of note. Strains from the Genus *Yersinia*, within the Enterobacteriaceae, displaced the *Klebsiella* sp., *Serratia* sp., and *Enterobacter* sp. cluster with subsequent phylogenetic re-location closer to the *Escherichia* sp., *Shigella* spp., and *Salmonella* sp. grouping. Within the Gammaproteobacteria, analysis of V4-6 regions placed *Legionella pnuemophila* within the Order Pseudomonadales, grouped with *Pseudomonas aeuroginosa*. In contrast V1-3 groupings were in agreement with full-length analysis where *L. pneumophila* and *Coxiella burnetti* were clustered together. Whilst both full-length and V1-3 analysis enabled the discrimination of *Burkholderia mallei* and *Burkholderia pseudomallei*, this was not achieved using V4-6. Additionally, *Burkholderia thailandensis*, originally clustered correctly with *B. pseudomallei* and *B. mallei*, was placed with *Burkholderia cepacia*. The most notable disagreement, which occurred for both V1-3 and V4-6 analyses, was the placement of both *Campylobacter jejuni* and *Helicobacter pylori* adjacent to members of the Spirochaete family. Of the Fusobacteriaceae, *F. necrophorum*, *F. naviforme, F. equorum, F. nucleatum, F. peridonticum* and *F. prausnitzii* were placed adjacent to the Actinobacteria with V1-3 analysis; a position not mirrored in either full-length or V4-6 analysis. Finally, V4-6 was unable to split *B. cereus* G9241 from *B. anthracis* and *B. cereus* ATCC 14759.

Although there are suggested advantages for the use of V1-3 assays in the context of identifying and discriminating notable BWA pathogens, further assessment of both regions for use in complex microbiome analysis were undertaken.

5.3.2 **PCR Assessment of 16S rRNA Assays**

Selected PCR primers for 16S rRNA were based on the universal 8F / 534R and 799F / 1114R primers for V1-3 and V4-6, respectively (Table 2.6). Primer sequences were amended to include 454™ adaptor sequences (A and B) and sequencing 'key' sequence (Section 2.5). The performance of these assays was first assessed through amplification of genomic DNA from species representing a range of bacterial phyla and classes. This included *Agrobacterium tumefacians*, *Bacillus anthracis*, *Burkholderia pseudomallei*, *Deinococcus radiodurans*, *Escherichia coli*, *Salmonella enteritica* serovar *typhimurium*, *Shigella flexneri* and *Vibrio paraheamolyticus*. PCR was performed in 50 µl reaction volumes on a GeneAmp® PCR System 9700 (Applied Biosystems, UK). Full assay and cycling conditions can be found in Tables 2.6 and 2.7. PCR products (575 and 622 bp respectively, for V1-3 and V4-6) were visualised on a 2 % pre-poured E-Gel® (Invitrogen, Life Technologies™, UK). Product sizes were determined using molecular weight marker III (Roche Applied Sciences, Germany). Results are shown in Figure 5.5.

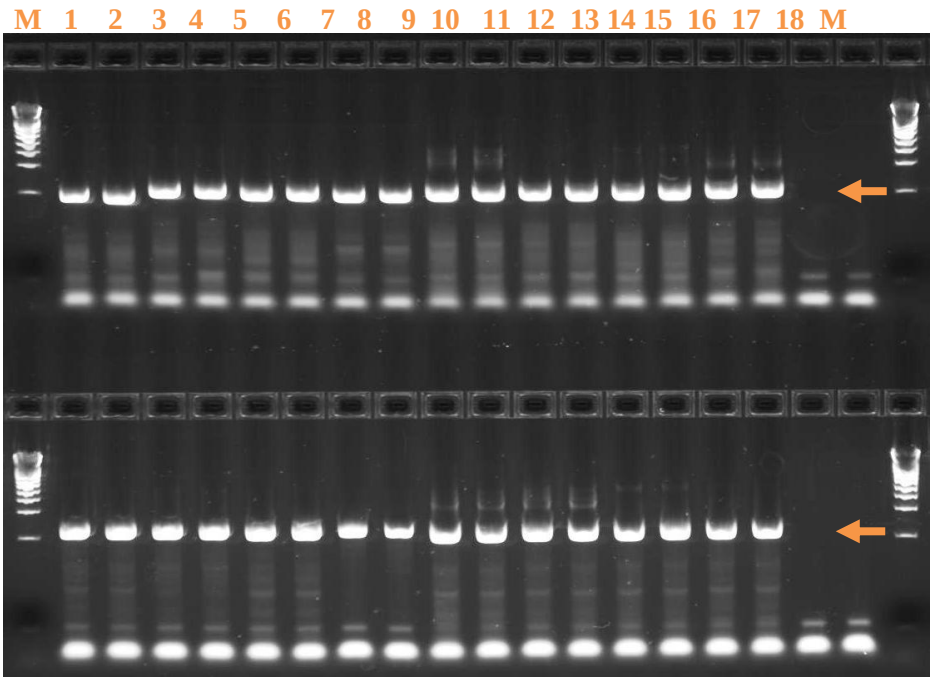

Figure 5.5: Agarose gel (2 %) demonstrating Phylum-level coverage of 16S rRNA 454™ PCR Assays targeting variable regions 1-3 (top) and 4-6 (bottom). Lane identities are as follows: 1 and 2 *Agrobacterium tumefacians*, 3 and 4 *Bacillus anthracis*, 5 and 6 *Burkholderia pseudomallei*, 7 and 8 *Deinococcus radiodurans*, 9 and 10 *Escherichia coli*, 11 and 12 *Salmonella enteritica* serovar *typhimurium*, 13 and 14 *Shigella flexneri* and 15 and 16 *Vibrio paraheamolyticus*. 17 and 18 are no template controls (ntc). M - DNA Molecular Weight Marker III. Correctly sized products (~500 bp) are denoted by the orange arrow.

5.3.3 **Assessment of Bar-coded 454™ 16S Primers**

Forward and reverse primers for both V1-3 and V4-6 were bar-coded using 454™ compatible 10mer sequence tags (Table 2.8). Prior to their use it was necessary to identify any issues, with regards to spurious products or non-amplification that arose as a result of MID tag inclusion. This was achieved through the amplification of 16S rRNA from genomic DNA of *B. pseudomallei*. PCR was performed in 50 µl reaction volumes on a GeneAmp® PCR System 9700 (Applied Biosystems, UK) as stated above. PCR products were visualised as before on a 2 % pre-poured E-Gel® (Invitrogen, Life Technologies™, UK)

with sizes determined using molecular weight marker III (Roche Applied Sciences, Germany). Results for V1-3 and V4-6 are shown in Figures 5.6 and 5.7, respectively.

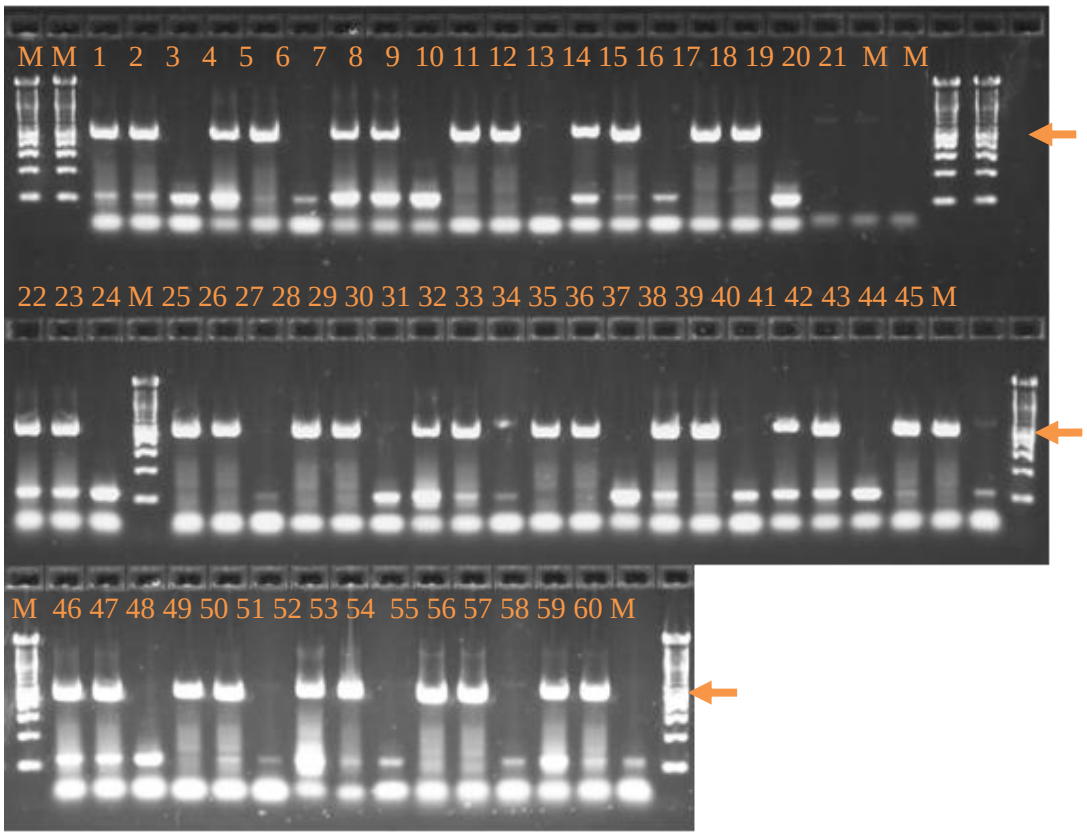

Figure 5.6: Agarose gel (2 %) Showing PCR amplification of 16S rRNA V1-3 of *B. pseudomallei* using 20 Multiplex-identifier (MID) primer pairs. PCRs for each MID primer were performed in duplicate. Lane identities are as follows: 1 and 2 - MID1, 4 and 5 - MID2, 7 and 8 - MID3, 10 and 11 - MID4, 13 and 14 - MID5, 16 and 17 - MID6, 19 and 20 - MID7, 22 and 23 - MID8, 25 and 26 - MID10, 28 and 29 - MID11, 31 and 32 - MID13, 34 and 35 - MID14, 37 and 38 - MID15, 40 and 41 - MID16, 43 and 44 - MID17, 46 and 47 - MID18, 49 and 50 - MID19, 52 and 53 - MID20, 55 and 56 - MID21, 58 and 59 - MID22. Lanes 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57 and 60 are no-template controls (ntc). M – DNA Molecular Weight Marker XIV. Correctly sized products (~500 bp) are denoted by the orange arrow.

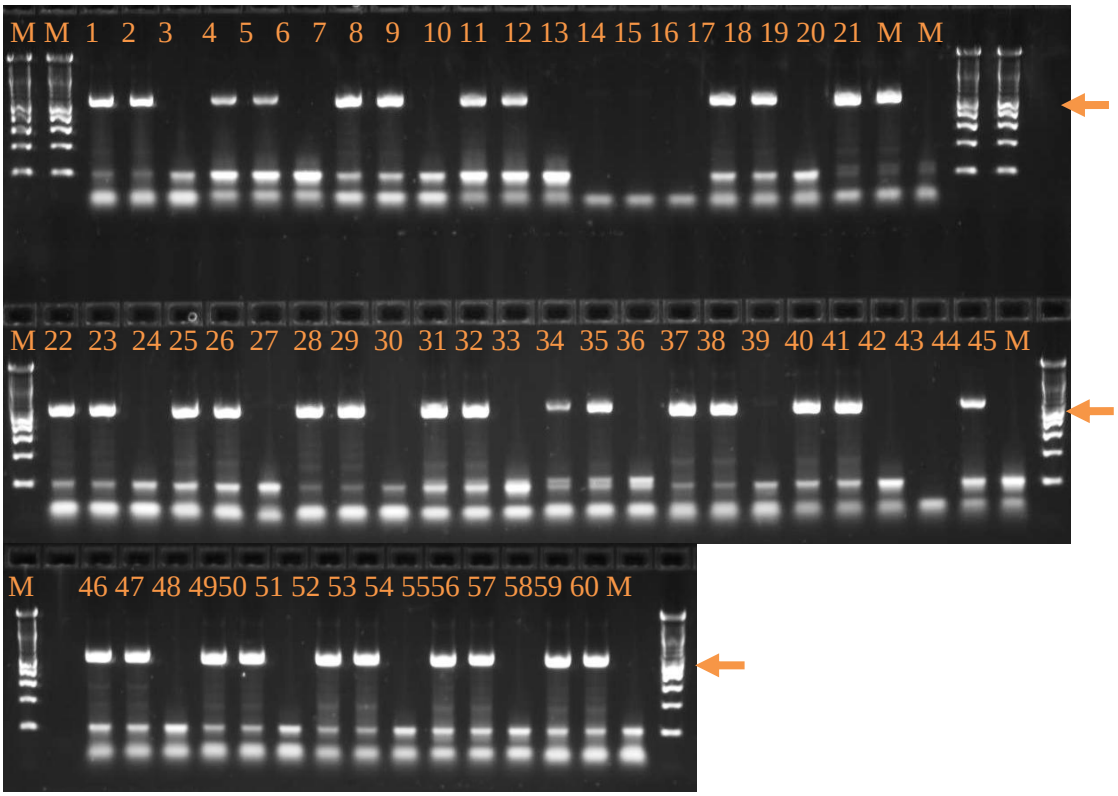

Figure 5.7: Agarose gel (2 %) Showing PCR amplification of 16S rRNA V4-6 of *B. pseudomallei* using 20 Multiplex-identifier (MID) primer pairs. PCRs for each MID primer were performed in duplicate. Lane identities are as follows: 1 and 2 - MID1, 4 and 5 - MID2, 7 and 8 - MID3, 10 and 11 - MID4, 13 and 14 - MID5, 16 and 17 - MID6, 19 and 20 - MID7, 22 and 23 - MID8, 25 and 26 - MID10, 28 and 29 - MID11, 31 and 32 - MID13, 34 and 35 - MID14, 37 and 38 - MID15, 40 and 41 - MID16, 43 and 44 - MID17, 46 and 47 - MID18, 49 and 50 - MID19, 52 and 53 - MID20, 55 and 56 - MID21, 58 and 59 - MID22. Lanes 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57 and 60 are no-template controls (ntc). M – DNA Molecular Weight Marker XIV. Correctly sized products are denoted by the orange arrow.

As shown in Figures 5.6 and 5.7, two MID adapted PCR primer pairs either yielded poor amplification (V1-3 MID7) or no amplification (V4-6 MID 5). These were not used in subsequent analyses. Although the amplification of correctly sized products appeared to mirror that of non-MID primers for both 16S targets, overall the inclusion of tag sequences did increase the formation of spurious products.

5.4 **Comparative Analysis of a Long-term Antibiotic Amended Soil**

A long-term study by Agriculture and Agri-Food Canada (AAFC) was undertaken to elucidate the impact of the application of veterinary antibiotics on soil microbial community composition and antibiotic resistance. A series of 12 plots (2 m^2) on the AAFC research farm in London Ontario received an annual application of antibiotics from 1999-2008. Each plot was isolated by means of an open fibreglass box (inserted to a depth of ~ 50 cm) with 1 m grassed strips between plots (Figure 5.8). Soil from this site was characterised as a silt-loam with a pH of 7.4. In June of each year (1999-2004), triplicate microplots received either no antibiotics, or a mix of tylosin, sulfamethazine, and chlortetracycline calculated to give a soil concentration of either 10, 100, or 1000 µg each antibiotic kg^{-1} dry weight soil. From 2005, the concentrations were increased 10-fold to 100-10000 µg kg^{-1} soil. The antibiotics were added to the plots by supplementing 1 kg portions of soil from each plot with mixtures of each antibiotic, adding the antibiotic-supplemented soil uniformly to the surface of the microplot, and manually tilling this in thoroughly to a depth of 5 cm.

Figure 5.8: Plots located at Agriculture and Agri-Food Canada (AAFC) used in the study of the impact of veterinary antibiotic application on microbial communities and antibiotic resistance in soil.

Previous work undertaken at the University of Warwick identified significantly increased integron prevalence in soil samples taken from the plots that had received antibiotic applications (Figure 5.9). Integrons, particularly those belonging to class 1, are important mobile genetic elements that confer resistance to antibiotics (Gaze *et al.*, 2011).

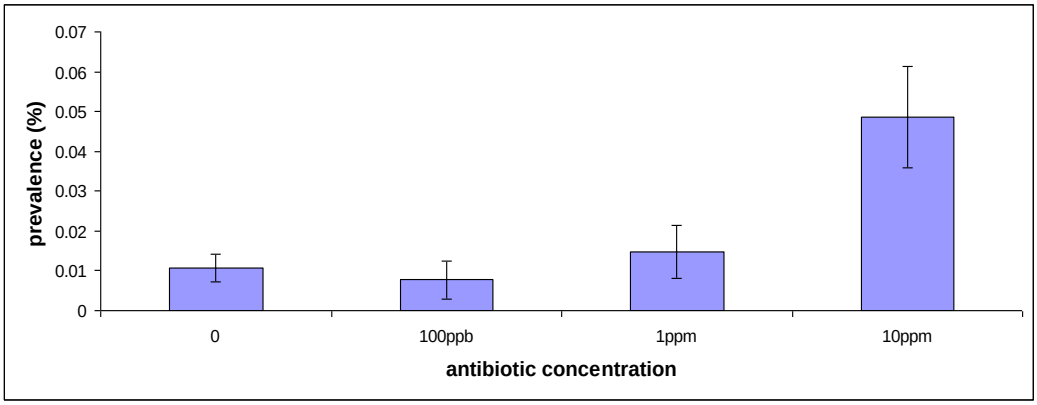

Figure 5.9: Graph showing prevalence of *intI1* in untreated and treated soils as determined by SYBR green real-time PCR. Data provided by Dr W. H. Gaze and Prof. E. M. Wellington, University of Warwick.

As a consequence, replicate soil samples (n = 4; two plots had one only two samples) from six plots (three naïve and three that had received the highest concentration of antibiotics) were chosen to determine if the application of veterinary antibiotics had an impact on the microbial community profile. Samples are listed in Table 5.1.

Table 5.1: Antibiotic-treated and Untreated Soil Samples used in this Study.

Sample Number	Antibiotic Treated Plot Reference	Sample Number	Antibiotic Untreated Plot Reference
01	Plot 4 Replicate 1	10	Plot 13 Replicate 1
02	4 R2	11	Plot 7 Replicate 1
03	4 R3	12	7 R2
04	4 R4	13	7 R3
05	Plot 8 Replicate 1	14	7 R4
06	8 R2	15	Plot 12 R1
07	8 R3	16	12 R2
08	8 R4	17	12 R3
09	Plot 11 Replicate 1	28	12 R4

5.4.1 **Extraction of DNA and Amplification of 16S rRNA**

For each of the eighteen soil samples, 3.5 g was subjected to total DNA extraction using the MoBio PowerMax® Soil DNA Isolation kit according to the manufacturer's instructions (Section 2.4.2.7). Extracted DNA (10 µl) was then analysed on an Invitrogen™ 0.8 % E-Gel® (Life Technologies™, UK) according to the manufacturer's instructions (Section 2.4.5) (Figure 5.10).

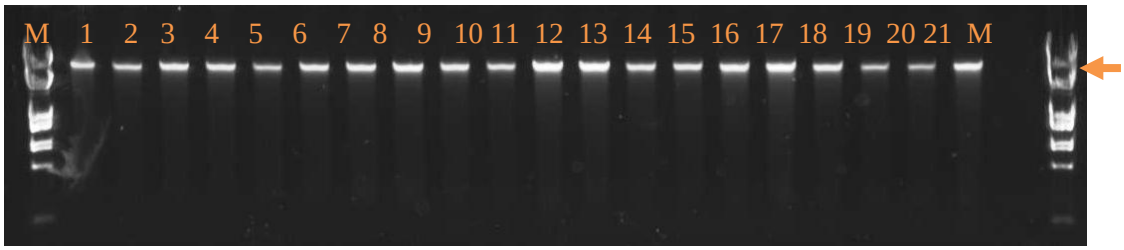

Figure 5.10: Agarose gel (0.8 %) showing DNA extraction from antibiotic amended soil samples. Lane identities are as follows 1 – 4R1; 2 – 4R2; 3 – 4R3; 4 – 4R4; 5 – 8R1; 6 – 8R2; 7 – 8R3; 8 – 8R4; 9 – 11R1; 10 – 11R2; 11 – 13R1; 12 – 13R2; 13 - 7R1; 14 – 7R2; 15 – 7R3; 16 – 7R4; 17 – 12R1; 18 – 17R2; 19 – 17R3; 20 – 17R4; 21 – no sample control; M – DNA Molecular weight marker IV. High molecular weight DNA indicated by orange arrow.

The amplification of 16S rRNA genes from each of the twenty samples was checked using the non-tagged V1-3 primer set as detailed above in section 5.3.1.2. Amplification products were analysed on a 2 % E-Gel® (Life Technologies™, UK) according to the manufacturer's instructions (Section 2.4.5). All samples produced products of the expected size (Figure 5.11).

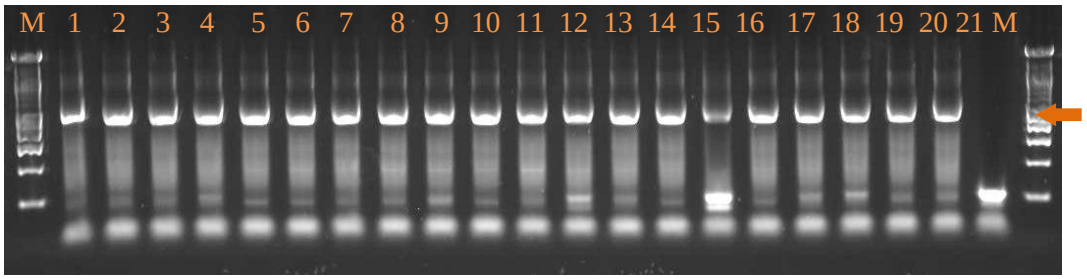

Figure 5.11: Agarose gel (2 %) showing amplification of 16S rRNA V1-3. Lane identities are as follows 1 – 4R1; 2 – 4R2; 3 – 4R3; 4 – 4R4; 5 – 8R1; 6 – 8R2; 7 – 8R3; 8 – 8R4; 9 – 11R1; 10 – 11R2; 11 – 13R1; 12 – 13R2; 13 - 7R1; 14 – 7R2; 15 – 7R3; 16 – 7R4; 17 – 12R1; 18 – 17R2; 19 – 17R3; 20 – 17R4; 21 – no sample control; M – DNA Molecular weight marker XIV. Correctly sized products are indicated by orange arrow.

5.4.2 **Comparing V1-3 and V4-6**

5.4.2.1 454™ Sequencing and Pre-processing Analysis

As a consequence of the lack of amplification from some MID-tagged primers (Figures 5.6 and 5.7) samples 10 (plot 11 replicate 2) and 12 (plot 13 replicate 2) were not taken forward for sequencing. For the remaining 18 samples, a total of 188,506 and 356,794 raw sequences were generated for V1-3 and V4-6 respectively using 454™ amplicon sequencing protocols (Section 2.5). Sequences were processed using the Quantitative Insights in Microbial Ecology (QIIME) pipeline (Caporaso *et al.*, 2010). Following clustering by barcode sequence, a number of pre-processing quality filtering steps were undertaken prior to analysis. These included selecting sequences within a length range of 200 – 600 bp, removing those with ambiguous sequences, quality scores of <25, homopolymer stretches greater than 6 bp and those with mismatches in the primer sequence. Following this, chimeric sequences were removed using the QIIME pipelines ChimeraSlayer script. The full output for each assay is shown in Table 5.2.

Table 5.2: Quality filtering outputs of 16S rRNA sequences for each sample analysed.

V1-3		V4-6	
Total Sequences 200 – 600 Ambiguous <Q25 Homopolymer Primer Mismatch	188,506 -37,597 -22,212 -1749 -3132 -74,049	Total Sequences 200 – 600 Ambiguous <Q25 Homopolymer Primer Mismatch	356,794 -189,818 -21,758 -1,591 -762 -61,178
Total	49,699	Total	76,891
Total (Post Chimera Removal)	42,162	Total (Post Chimera Removal)	63,433
Sample Number	Sequences	Sample Number	Sequences
1	1982	1	3822
2	2333	2	3588
3	375	3	2704
4	2469	4	1503
5	1954	5	2395
6	2239	6	3150
7	867	7	3556
8	1992	8	3125
9	3358	9	2651
10	1914	10	2266
11	1845	11	15718
12	2021	12	4195
13	2523	13	3556
14	1370	14	2099
15	1434	15	3278
16	1094	16	5827
17	2438	17	0
18	9954	18	0

5.4.2.2 Comparison of Dominant Bacterial Phyla

Following removal of low quality reads (score <25) and sequences <200 bp and greater than 600 bp in length, remaining sequences were clustering into OTUs using a 0.04 dissimilarity index (96% sequence similarity) as suggested by Kim *et al.*, (2010) using uclust (Edgar 2009, unpublished). Taxonomic assignments were made using the ribosomal database project (RDP) classifier. Representative sequences were then aligned using PyNAST (Caporaso *et al.*, 2010) with a pre-aligned Greengenes 16S rRNA dataset from http://greengenes.lbl.gov. Finally,

chimeric sequences were removed using ChimeraSlayer, the alignment filtered and a phylogenetic tree built by FastTree (Price *et al.*, 2009) using the remaining sequences. The resultant taxonomic summaries at phylum-level for each individual sample for both V1-3 and V4-6 are shown in Figures 5.12 and 5.13.

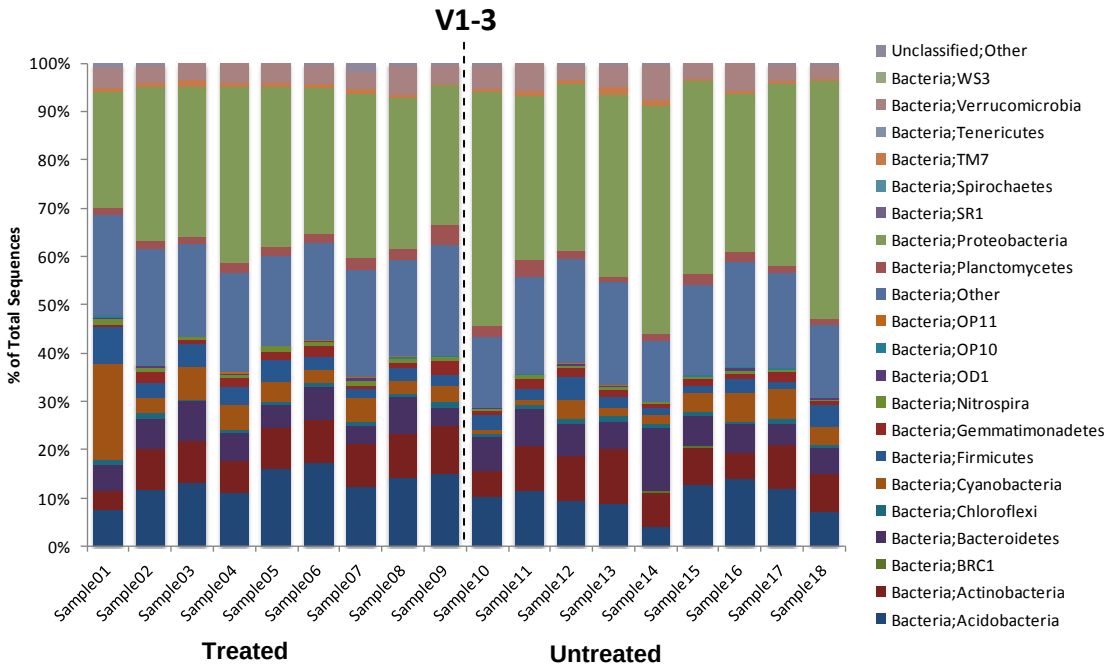

Figure 5.12: Distribution of major bacterial groups at the phylum level for 16S rRNA V1-3.

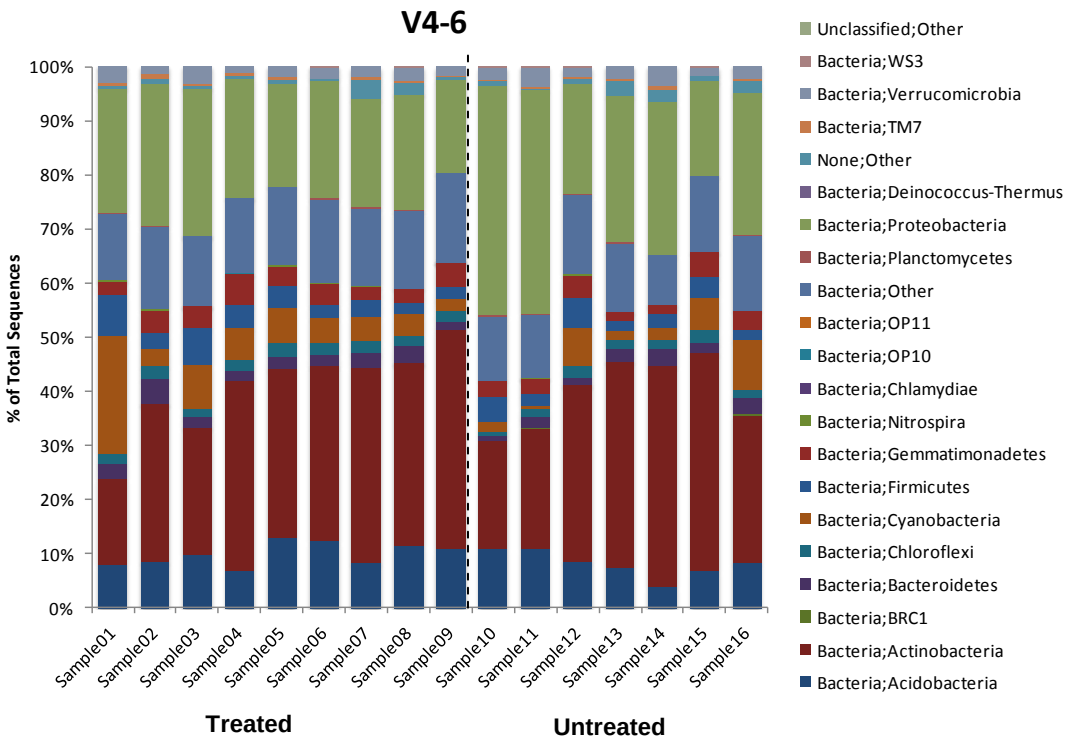

Figure 5.13: Distribution of major bacterial groups at the phylum level for 16S rRNA V1-3.

As shown above, the dominant phyla observed for V1-3 analysis were Proteobacteria (32.1 % treated; 39 % untreated), Acidobacteria (13.9 % treated; 10 % untreated), Actinobacteria (8.6 % treated; 8.4 % untreated), Bacteriodetes (5.8 % treated; 6.9 % untreated) and Verrucomicrobia (4.0 % treated; 4.4 % untreated). The dominant phyla detected by V4-6 analysis were, Actinobacteria (32.7 % treated; 33.6 % untreated), Proteobacteria (21.8 % treated; 26.8 % untreated), Acidobacteria (10.2 % treated; 7.7 % untreated), Cyanobacteria (4.9 % treated; 4.3 % untreated) and Gemmatimonadetes (3.9 % treated; 3.1 % untreated). There was a clear increase in the number of OTUs representative of Actinobacteria in the V4-6 analysis (Figure 5.13). It was unclear why samples 17 and 18 from the 'untreated' group failed to produce any sequence data.

5.4.2.3 Phylum-level Variations between Untreated and Antibiotic-treated Soils

Overall agreement between the assignment of relative abundance levels, as determined for each of V1-3 and V4-6, at phylum and class level taxonomies is shown in Figure 5.14. There was good agreement between the assays apart from those assigned to the Actinobacteria.

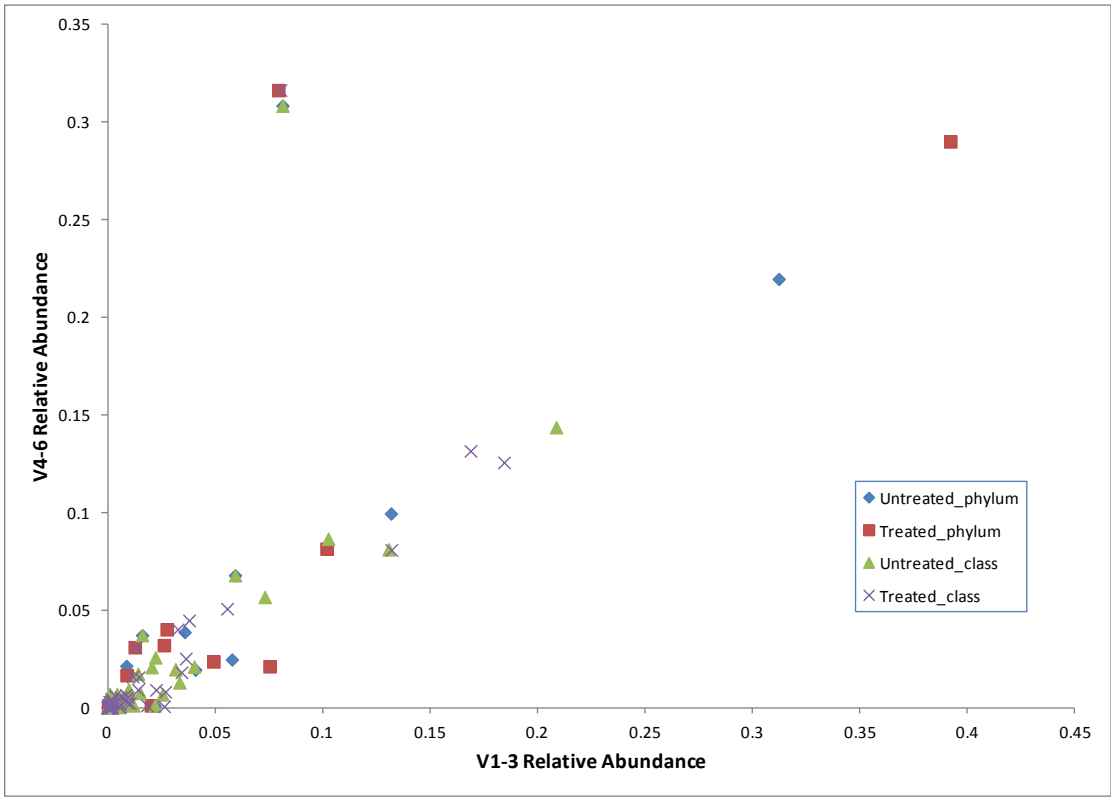

Figure 5.14: Scatter plot showing comparisons between V1-3 and V4-6 assays, of the relative abundance for each Phyla and Class observed for both untreated and treated soil samples.

Relative abundance of each phyla, compared between untreated and antibiotic amended soil samples, were analysed for statistically significant differences using two-sample students t-test assuming equal variance. Results (*p*-values) are shown in Table 5.3.

Table 5.3: Statistical comparison of changes in relative abundance of OTUs at the phylum level between untreated and treated samples undertaken using two-sample Students t-test assuming equal variance.

Phyla	*p*-value V1-3	Change Relative to Untreated	*p*-value V4-6	Change Relative to Untreated
Acidobacteria	**0.014743**	Increased	**0.056041**	Increased
Actinobacteria	0.729486	-	0.798583	-
BRC1	0.359874	-	0.750554	-
Bacteroidetes	0.364337	-	0.794469	-
Chloroflexi	0.714736	-	0.07163	-
Cyanobacteria	0.470954	-	0.722745	-
Firmicutes	0.471343	-	0.582412	-
Gemmatimonadetes	0.281546	-	0.260255	-
Nitrospira	**0.002605**	Increased	**0.034894**	Increased
OD1	0.949498	-	ND	ND
OP10	0.411028	-	0.236069	-
OP11	**0.04032**	Increased	0.524472	-
Planctomycetes	0.254797	-	0.922893	-
Proteobacteria	**0.009467**	Reduced	0.147491	-
SR1	0.177343	-	ND	ND
Spirochaetes	0.334282	-	ND	ND
TM7	0.680177	-	0.717223	-
Tenericutes	0.334282	-	ND	ND
Verrucomicrobia	0.52538	-	0.216639	-
WS3	0.422234	-	0.385308	-
Chlamydiae	ND	ND	0.272026	-
Deinococcus-Thermus	ND	ND	0.860851	-

Of the phyla examined for changes in OTU abundance between treated and untreated samples four (Acidobacteria, Nitrospira, Proteobacteria and OP11) were shown to have statistically significant variations. Of these, Acidobacteria and Nitrospira were shown to have increased abundance relative to the untreated samples using both V1-3 and V4-6 analysis. Changes in Proteobacteria and OP11 abundance were only reported through V1-3 analysis with a reduction and increased relative to untreated samples, respectively.

5.4.2.4 Comparing Alpha Diversity

The diversity and richness of OTUs within each sample (alpha diversity) was computed in QIIME. Metrics used included observed species richness (the count of unique OTUs within a sample), Chao1 (a non-parametric estimator that

calculates the minimal number of OTUs present in a sample and thus a descriptor of species richness), Phylogenetic Distance and the Shannon index (higher values indicate greater diversity and evenness). Alpha diversity results for each assay per sample are shown in Table 5.5. Figure 5.15 shows the comparison between the two data-sets (V1-3 and V4-6) for observed species.

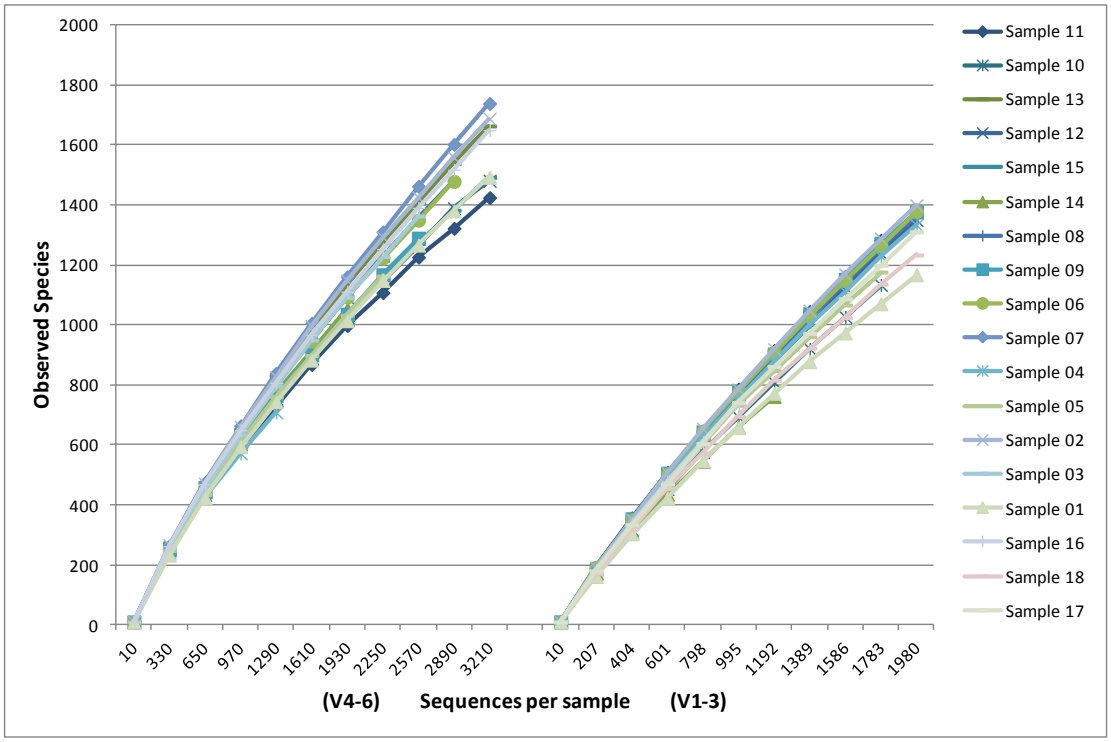

Figure 5.15: Graph showing observed species richness for soil samples amplified using V4-6 or V1-3 16S rRNA PCR assays. Richness was computed in QIIME v. 1.3.0.

Asymptotes did not occur for any sample as the number of unique OTUs did not reduce as additional sequences were introduced into the rarefaction analysis. As a consequence of the low OTU coverage for samples from both treated and untreated soils, samples were re-sequenced at a greater depth using the same sequencing libraries that were initially sequenced. Results are shown in Table 5.4. Sequencing depth (as determined by number of sequences per sample) was increased approximately 10-fold in this analysis.

Table 5.4: Quality filtering outputs of 16S rRNA re-sequencing for each soil sample analysed.

V1-3		V4-6	
Total Sequences 200 – 600 Ambiguous <Q25 Homopolymer Primer Mismatch	1,012,709 -150,724 -57,073 -4,684 -22,262 -436,984	Total Sequences 200 – 600 Ambiguous <Q25 Homopolymer Primer Mismatch	950,586 -185,033 -41,840 -2,845 -20,525 -459,190
Total	340,713	Total	227,070
Total (Post Chimera Removal)	285,406	Total (Post Chimera Removal)	186,545
Sample Number	Sequences	Sample Number	Sequences
1	13,885	1	11,261
2	14,382	2	11,002
3	3,248	3	8,734
4	16,980	4	4,324
5	11,661	5	7,274
6	13,744	6	10,546
7	6,448	7	10,674
8	13,809	8	9,793
9	22,539	9	7,838
10	13,156	10	6,956
11	12,310	11	41,255
12	13,209	12	11,891
13	16,888	13	12,050
14	10,672	14	5,645
15	12,128	15	9,441
16	8,586	16	17,861
17	15,749	17	0
18	66,732	18	0

5.4.2.5 Re-sequencing Analysis

Results of observed species richness are shown in Figure 5.16. Despite the increase in sequencing depth asymptotes were still not observed for either assay. However, what is of interest is the apparent greater diversity being observed by the V1-3 analysis. This is also the case for phylogenetic diversity, Chao1 and Shannon index metrics which are detailed in Table 5.5.

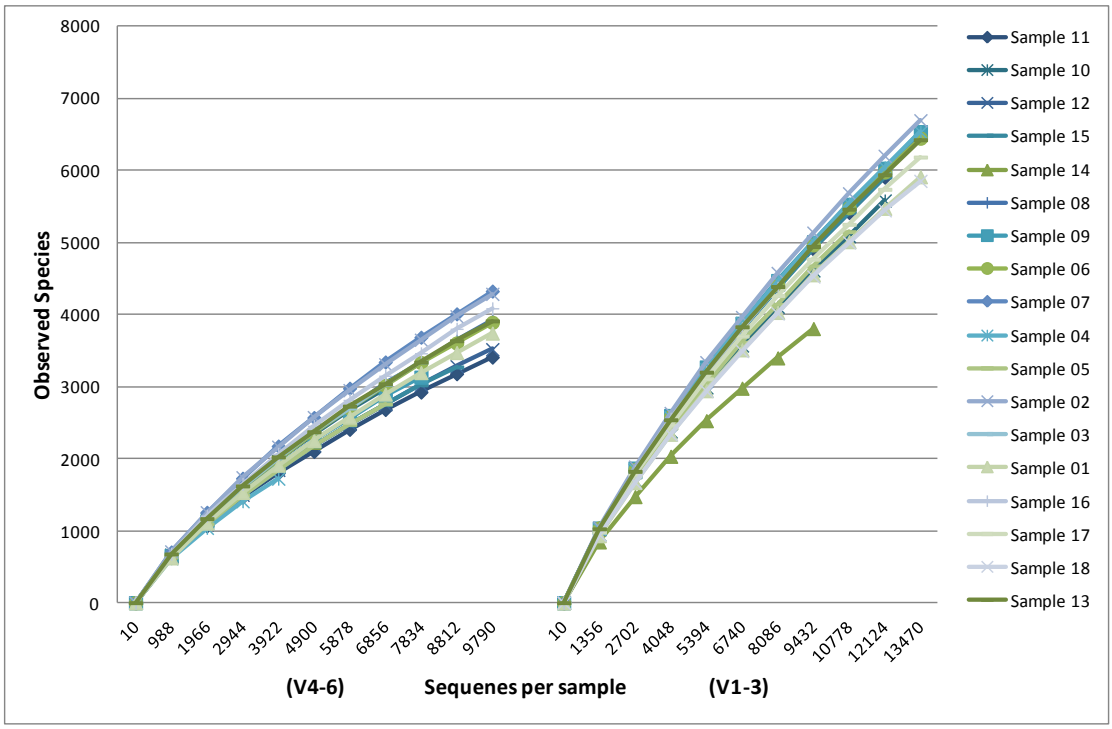

Figure 5.16: Graph showing observed species richness for re-sequenced soil samples amplified using V4-6 or V1-3 16S rRNA PCR assays. Richness was computed in QIIME v. 1.3.0.

	Observed species richness		Chao1		Phylogentic Diversity (PD)		Shannon index (*H'*)	
Sample	V1-3	V4-6	V1-3	V4-6	V1-3	V4-6	V1-3	V4-6
1	5916 (±5.69)	3747.9 (±12.91)	18211.88 (±143.96)	10295.21 (±181.23)	666.83 (±1.00)	10295.21 (±181.23)	10.73 (±0.004)	9.99 (±0.009)
2	6698.9 (±9.47)	4291.3 (±14.51)	20104.34 (±311.12)	11710.07 (±250.446)	729.24 (±1.76)	11710.07 (±250.44)	11.74 (±0.003)	10.91 (±0.009)
3	1754.2 (±10.34)	3347 (±7.07)	6712.00 (±312.84)	9278.60 (±222.44)	256.72 (±0.87)	9278.60 (±222.43)	10.10 (±0.02)	10.41 (±0.011)
4	6536.8 (±20.9)	1729.3 (±10.81)	19026.77 (±505.27)	4900.96 (±215.16)	722.49 (±3.04)	4900.96 (±215.16)	11.55 (±0.01)	9.75 (±0.010
5	5138.8 (±7.74)	2766.1 (±7.92)	15563.03 (±191.77)	7443.17 (±117.74)	568.83 (±1.37)	7443.16 (±117.74)	11.18 (±0.01)	10.15 (±0.011)
6	6450.6 (±9.00)	3897.4 (±6.20)	18339.66 (±156.33)	10628.72 (±166.22)	688.93 (±1.38)	10628.72 (±166.22)	11.61 (±0.003)	10.59 (±0.003)
7	3013.3 (±14.96)	4330.1 (±16.16)	9603.01 (±143.52)	11791.11 (±133.36)	368.04 (±2.19)	11791.11 (±133.36)	10.71 (±0.02)	10.86 (±0.008)
8	6504.5 (±5.13)	3923.1 (±0.86)	19386.66 (±136.58)	10838.42 (±8.07)	697.73 (±0.76)	10838.42 (±8.06)	11.61 (±0.003)	10.62 (±0.000)
9	6540.9 (±30.70)	3136.1 (±0.88)	19572.13 (±529.23)	8016.71 (±21.22)	703.10 (±3.43)	8016.71 (±21.21)	11.65 (±0.013)	10.38 (±0.000)
10	5586.6 (±13.53)	2978.4 (±5.56)	17537.54 (±119.33)	8031.22 (±89)	637.34 (±2.41)	8031.22 (±89.00)	11.06 (±0.01)	10.37 (±0.004)
11	5909.9 (±6.19)	3414.3 (±32.73)	17215.38 (±83.52)	8422.69 (±287.10)	649.93 (±0.78)	8422.69 (±287.09)	11.50 (±0.002)	10.35 (±0.025)
12	6034.3 (±12.89)	3524.2 (±13.11)	17392.25 (±271.97)	8782.05 (±278.90)	671.76 (±1.71)	8782.05 (±278.87)	11.52 (±0.005)	10.28 (±0.010)
13	6429.1 (±24.94)	3917.4 (±19.66)	18507.21 (±351.60)	10313 (±248.24)	704.60 (±3.51)	10313 (±248.24)	11.53 (±0.005)	10.66 (±0.013)
14	3810.7 (±16.13)	2233.5 (±10.73)	11182.68 (±192.07)	6237.64 (±122.04)	450.17 (±2.40)	6237.63 (±122.04)	10.31 (±0.01)	10.11 (±0.014)
15	5911.9 (±0.74)	3257.8 (±10.80)	18073.78 (±15.02)	8496.59 (±119.14)	657.14 (±0.29)	8496.58 (±119.13)	11.30 (±0.00)	10.24 (±0.006)

	Observed species richness		Chao1		Phylogentic Diversity (PD)		Shannon index (*H'*)	
16	4375.8 (±10.00)	4091.4 (±15.74)	13572.95 (±221.19)	11316.3 (±401.63)	506.37 (±1.74)	11316.3 (±401.63)	11.09 (±0.005)	10.64 (±0.010)
17	6188.7 (±13.11)	-	16971.62 (±307.10)	-	666.58 (±1.84)	-	11.32 (±0.009)	-
18	5858.1 (±29.89)	-	17307.43 (±439.66)	-	642.33 (±5.60)	-	11.02 (±0.02)	-
Average (Untreated)	5394.89	3463.1	16279.9	9433.7	600.2	9433.7	11.2	10.4
Average (Treated)	5567.2	3345.3	16417.9	8799.9	620.7	8799.9	11.2	10.4

Table 5.5: 16S rRNA OTU diversity and richness metrics for antibiotic treated (10-18) and untreated (1-9) soil samples

5.4.2.6 Dominant Phyla

As shown in Figures 5.16 and 5.17, the results of the dominant phyla analysis mirrored that of the initial sequencing effort. For V1-3 analysis Proteobacteria represented the dominant phyla with 32.2 % and 37.5 % of sequences for the treated and untreated samples, respectively. Acidobacteria (11.3 % treated; 9.5 % untreated), Actinobacteria (7.2 % treated; 7.5 % untreated), Bacteriodetes (8.9 % treated; 11.6 % untreated) and Verrucomicrobia (6.2 % treated; 3.8 % untreated) were the other phyla that dominated. Again, as with the initial analysis, with the exception of the Verrucomicrobia observed in the V1-3 analysis, the dominant phyla detected by V4-6 analysis were Actinobacteria (30.1 % treated; 33.5 % untreated), Proteobacteria (23.3 % treated; 27.0 % untreated), Acidobacteria (9.1 % treated; 8.2 % untreated), Cyanobacteria (7.3 % treated; 4.2 % untreated) and Gemmatimonadetes (3.0 % treated; 3.4 % untreated).

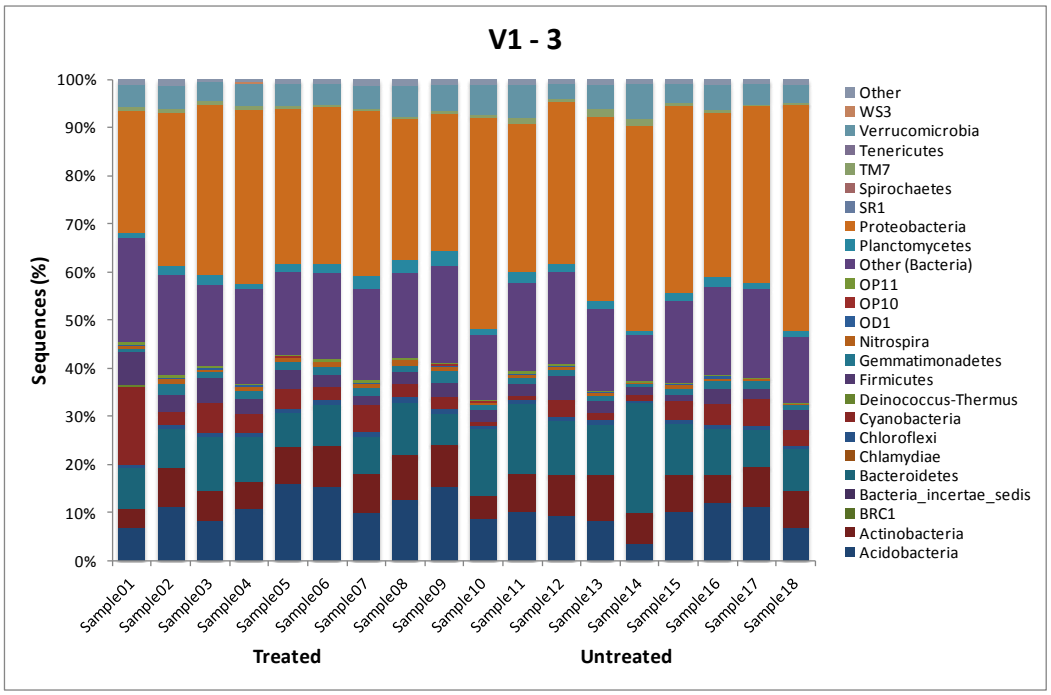

Figure 5.17: Distribution of major bacterial groups at the phylum level for 16S rRNA V1-3 of antibiotic treated and untreated soil samples.

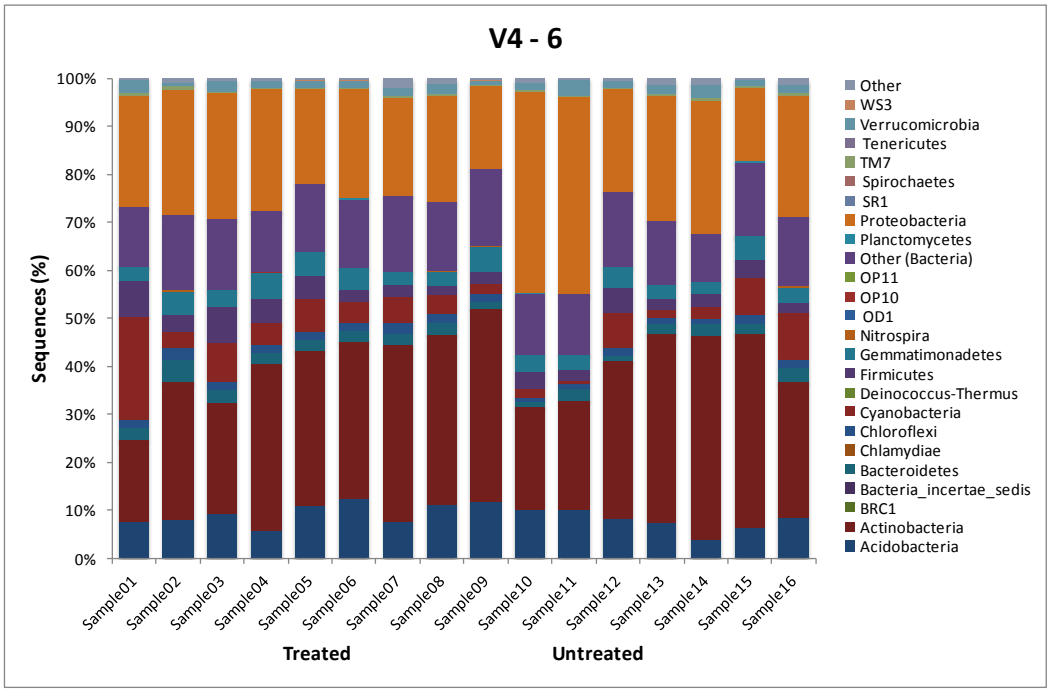

Figure 5.18: Distribution of major bacterial groups at the phylum level for 16S rRNA V4-6 of antibiotic treated and untreated soil samples.

5.4.2.7 Comparing Beta Diversity

Beta diversity was computed for the outputs of each of V1-3 and V4-6 to determine the phylogenetic relatedness of the untreated and treated samples. This was generated using the QIIME pipeline and determined using the unweighted UniFrac distance metric (Lozupone *et al.*, 2007), visualised through principal component analysis (PCA). Beta diversity was computed using an even sampling depth of 10,672 and 4,324 sequences for V1-3 and V4-6 respectively. As accurate comparisons using samples containing fewer sequences than this number is not possible, samples 3, 7 and 16 from the V1-3 group (having 3,248, 6,448 and 8,586 sequences respectively) were not used in the generation of the PCA plots. Results are shown in Figures 5.19 and 5.20.

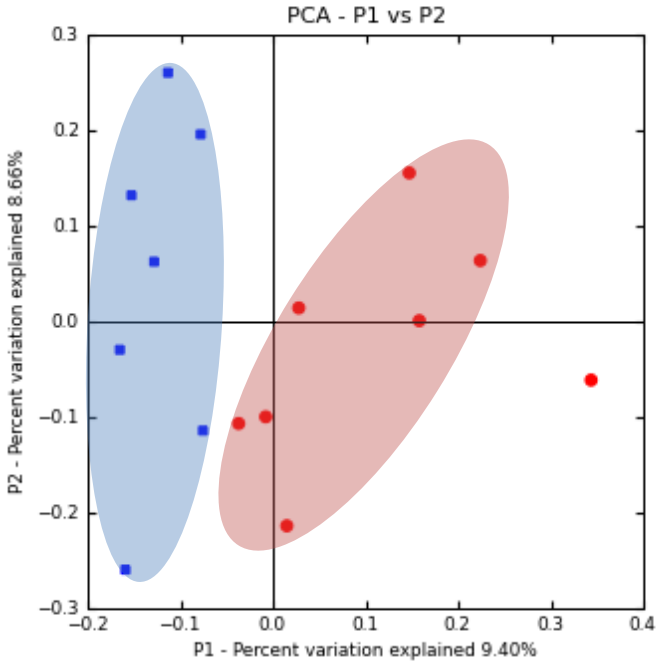

Figure 5.19: Principal component analysis plot showing OTU beta diversity of soils that have undergone antibiotic treatment (blue) compared to untreated (red). Plots are based on analysis of OTU relative abundance using V1-3 16S rRNA. Beta diversity was computed using unweighted unifrac in QIIME v 1.3.0.

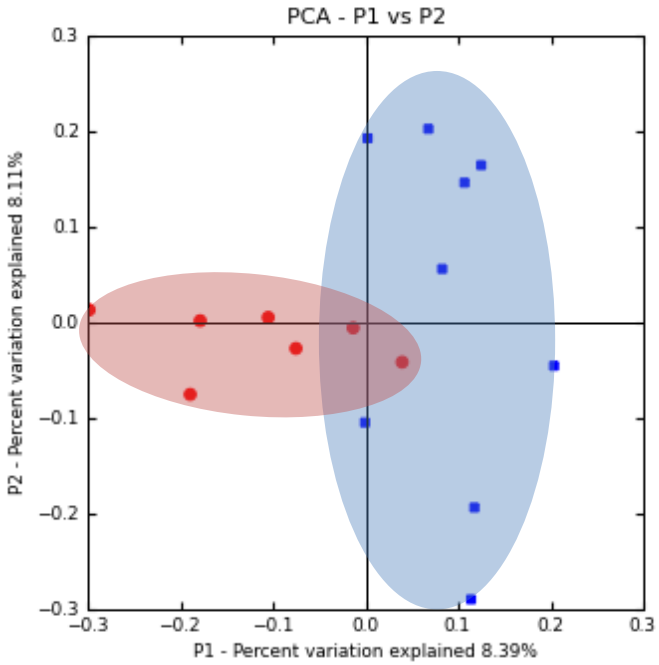

Figure 5.20: Principal component analysis plot showing OTU beta diversity of soils that have undergone antibiotic treatment (blue) compared to untreated (red). Plots are based on analysis of OTU relative abundance using V4-6 16S rRNA. Beta diversity was computed using unweighted unifrac in QIIME v 1.3.0.

Clustering of treated and untreated samples was evident for both V1-3 and V4-6 analysis however there was a lack of full segregation in the latter.

5.4.3 **Changes in OTU Relative Abundance**

ANOVA was used to determine changes in the relative abundance of OTUs between the untreated and treated soils, and subsequently between the two assays. In total, at a *p* value of <0.05, there were 277 and 98 OTUs whose relative abundance had altered following antibiotic treatment for assays V1-3 and V4-6, respectively. Of these 112 and 37 were classified to at least Family taxonomic level in QIIME using the RDP classifier. These OTUs were derived from 64 Families representing 16 Phyla of which 11 were common to both V1-3

and V4-6 assays. However, application of Bonferroni correction (where raw *p*-values are multiplied by the total number of observations that are being compared, in this case OTUs) resulted in a loss of all statistically significant OTUs.

To overcome this issue, additional filters were placed on the observed OTUs prior to ANOVA. OTUs were filtered by variance, and only those that were observed in a minimum of four samples within either treatment group were analysed. Results of these analyses are shown in Table 5.6. It should be noted however that, even after these filters were used, application of Bonferroni correction reduced statistically significant OTUs to zero.

Table 5.6: Table showing OTUs (after filtering) that demonstrated a change in relative abundance in response to antibiotic treatment.

Lineage (V1-3)	*p*-value	OTU Count (Untreated)	OTU Count (Treated)	Phylum
Comamondaceae	0.00376	0	7	Proteobacteria
Bacteria (unspecified)	0.004659	0	16	-
Bacteria (unspecified)	0.006985	0	19	-
Comamonadaceae	0.008237	1	5	Proteobacteria
Adhearibacter	0.008237	0	10	Bacteroidetes
Lineage (V4-6)	*p*-value	OTU Count (Untreated)	OTU Count (Treated)	Phylum
Dechloromonas	0.000962	23	204	Proteobacteria
Cyanobacteria	0.002827	18	0	Cyanobacteria
GpXIII (Family – Cyanobacteria)	0.004629	21	3	Cyanobacteria
Bacteria	0.009296	4	15	-

5.5 **Discussion**

The use of 16S rRNA hyper-variable region, high-throughput sequencing analysis for the characterisation of microbial communities was first introduced by Sogin *et al.*, (2006). Since this first use to explore the deep sea and rare biosphere, the technique has seen a proliferation in its application to many and varied environments including soil. Data presented here represent the use of this approach for two aims 1) the comparative performance of assays that produce amplicons spanning hyper-variable regions for the characterisation of soil communities and 2) the determination of the effects of long-term antibiotic application on microbial communities in soil.

Firstly, to enable observations of subtle changes in the relative abundance of OTUs an important consideration is that of the capacity for the chosen method of DNA extraction to present a truly representative sample of the metagenome. Although DNA extraction protocols employed here were a direct lysis only method, consideration was given to sample pooling with regards to maximising the coverage of samples taken from each antibiotic treatment regime. In this study samples were analysed without pooling prior to either PCR amplification or DNA extraction due to the recently demonstrated negative impact these have on OTU detection (Manter *et al.*, 2010). Ultimately, not pooling the samples did limit the depth of sequence coverage achieved per sample as sequencing effort was split between 18 individual samples rather than one pooled sample representing each treatment. Even so the total number of sequences for each treatment was high. In terms of DNA extraction method, Delmont *et al.*, (2011) recently quantified the benefit of applying multiple and varied methods of

sampling, fractionation and lysis stringency to a soil sample demonstrating access to a more genetically diverse proportion of the soil metagenome than with any single method. This approach has, to a certain extent, been used previously to reduce the complexity within an agricultural soil where diversity and evenness of OTUs would otherwise necessitate a much deeper sequencing strategy to fully characterise the community (Morales *et al.*, 2009). Additionally, Delmont *et al.*, (2011) noted the common practice (which was employed here) of multiple sampling from areas within a plot of interest was not the most significant factor in increasing the diversity sampled. It is clear that whilst efforts were made to ensure that diversity was not lost through the sample pooling, alternative methods of DNA extraction may have been of benefit.

The use of partial 16S rRNA fragments that span multiple, defined variable regions for the purpose of characterising microbiomes has been well validated through *in silico* comparisons to the full-length gene (Liu *et al.*, 2008; Morales *et al.*, 2009; Kim *et al.*, 2011). However, here there was apparent bias in one or both assays towards either the underestimation (V1-3) or overestimation (V4-6) of Actinobacteria. It is likely, as only the V4-6 assay was manipulated according to the ClustalW alignments (Figures 5.2-5.4), that it is an introduced amplification bias for this assay that has caused this effect. Regarding the remaining community structure of samples from both the treated and untreated groups, there are clear similarities to what has been observed before for soil. The abundance of Proteobacteria sequences observed from soils from Brazil, Florida, Canada and Illinois, representing >40 % of sequences (Roesch *et al.*, 2007), mirrors abundances shown here of between 23.3 and 37.5 % depending on the

sample and assay employed. Interestingly, the presence of Verrucomicrobia (~6.2 % of sequences in the untreated group) links these samples (representing a Canadian soil) with the Canadian and North American soils used by Roesch *et al.*, (2007) in which this Phylum was also observed (~2 %). This phylum is phylogenetically distinct within the kingdom Bacteria and includes members found in both soil and aquatic habitats (Lee *et al.*, 2009). Recent studies by Bergmann *et al.,* (2011) have suggested that the occurence of this phylum in soils has been underrepresented to date; a consequence of amplification biases using common 16S primers. The authors demonstrated detection in 180 of 181 geographically diverse soils tested, with the phylum accounting for 23 % of all sequences on average and relative abundance highest in grasslands (Bergmann *et al.*, 2011).

What is clear from the alpha diversity metrics is that the sampling effort, in terms of sequence depth per soil sample, was not sufficient to capture the full diversity of OTUs. This was complicated by computational bottlenecks that prevented the denoising of datasets. It is highly likely, therefore, that OTU richness was falsely inflated as a result of errors / artefacts generating sequences that were ultimately counted as unique. There were clear differences in observed alpha diversity metrics with V1-3 suggesting a substantial diversity that was not captured by V4-6 analysis. However, it should be remembered that the dissimilarity index (0.04) proposed by Kim *et al.*, (2010) for regions 1 through 5 does not consider region 6. Instead, for regions 6-9, a cut-off of 0.02 was suggested (Kim *et al.*, 2010). The lower diversity observed for 4-6 using 96% would suggest that use of this

assay may require a 98% similarity threshold for assigning OTUs to species instead.

Roesch *et al.*, (2007) had previously estimated the number of OTUs per gram of soil (using four geographically diverse samples) at 52,000. This exceeds that observed here using Chao1 estimator where each sample's OTU richness was determined to be 16,418 and 16,280 for treated and untreated respectively using V1-3. For V4-6 this was substantially lower at 8,800 and 9,434 for treated and untreated respectively. It has however been previously demonstrated that Chao1 underestimates richness for samples where sequencing effort is low, which is the case here (Lemos *et al.*, 2011). Regardless of this, the estimated sequencing depth required to capture 90 % of the diversity of 52,000 OTUs was determined to be 713,000 sequences (Roesch *et al.*, 2007), a number greatly exceeding that generated in this study and explains in part the reduced number of OTUs observed. In support of this requirement for greater sequencing depth, Lemos *et al.*, (2011) determined that, regardless of soil tested, at least 20,000 sequences were required to achieve coverage of ~90 % of OTUs. This could be overcome through an increase in sequencing effort to generate the required number of sequences using the assays detailed here, or through the application of shorter amplicon sequencing approaches on higher throughput instruments. The latter approach has shown utility in the analysis of complex microbial communities (Bartram *et al.*, 2011). However, careful consideration should be given as to the required length of amplicon and selection of appropriate variable region. There has, for example, been an observed inability to accurately reconstruct simulated communities *in silico* using V6 analysis (Kim *et al.*, 2010) as well as issues

related to both higher levels of taxonomic assignment using differing variable regions (Liu *et al.*, 2008) and the lack of lower levels of taxonomic classification (e.g. Genus) when compared to longer sequence reads (Claesson *et al.*, 2010).

PCA using unweighted UniFrac were used to assess the phylogenetic relatedness of the soil samples that had undergone long-term treatment with antibiotics. The antibiotics used in the study that generated these soil samples (tylosin, sulfamethazine, and chlortetracycline) are all considered to have broad-spectrum of activity. There is evidence however for increased activity of tylosin (a macrolide class of antibiotic) against the cell membrane of Gram positive bacteria, with no activity against the Enterobacteriaceae. However, as a consequence of this broad spectrum activity no clear hypothesis as to the likely impact on defined microbiome members could be made. Despite the limitation of the data presented here in terms of sampling the total community, the analysis of both V1-3 and V4-6 beta diversity using unweighted UniFrac did indicate an alteration in community structure in what can be hypothesised as a response to the application of multiple antibiotics. It should be noted that analysis of V1-3 gave a much clearer separation of samples through PCA analysis than V4-6.

The observed increase in integron prevalence within the antibiotic treated samples suggests either an expansion of species / strains harbouring these elements with antibiotic resistance determinants or the transfer of these mobilisable elements within the microbial community. Of the OTUs whose change in relative abundance was putatively determined to be statistically significant (*p*-value <0.01) *Comamonas* sp. are known to harbour IncP-1β

plasmids that had similarity to those known to harbour antibiotic resistance determinants (Tralau *et al.*, 2001). Additionally, *Dechloromonas* sp. have been shown to harbour class-1 integrons (*Int1*) (Boucher *et al.*, 2007; Nemergut *et al.*, 2008). No documented occurrence of integrons in *Adhaeribacter* sp. could be found. However, closely related members of the order Sphingobacteriales have been observed to harbour these elements (Nemergut *et al.*, 2008) as well as genes encoding resistance to tetracycline that are known to be encoded within integrons (Li *et al.*, 2010). Despite the documented occurrence of integrons within the Cyanobacteria (Boucher *et al.*, 2007; Nemergut *et al.*, 2008) it is interesting to note that two OTUs (through V4-6 analysis) were found to decline in response to antibiotic treatment. This reflects the sensitivity of Cyanobacteria which has resulted in their use in bioassays to assess the environmental toxicity of antibiotics (van der Grinten *et al.*, 2010). Here the authors showed their sensitivity was two orders greater than the classic green algae-based bioassays (van der Grinten *et al.*, 2010). The application of Bonferroni correction to these results was used to reduce type I errors (false positive *p*-values) within the OTU ANOVA outputs. However, this resulted in no OTUs remaining significantly different in response to treatment, even when the *p*-values prior to correction where <0.001. This would suggest, for these data at least, that Bonferroni correction is too conservative a method to apply. Indeed it is been highlighted previously that use of this correction increases the likelihood of type II errors (false negatives) and consequently Bayesian methods, where prior beliefs regarding the data are incorporated into the statistical analyses, are suggested for clearer interpretations (Perneger, 1998). Given the limitations of the statistical analysis of changes in the relative abundance of OTUs between untreated and

treated soils it is not possible to predict with any certainty the overall contribution of these few lineages to the observed increase in integron prevalence.

A number of issues need to be addressed to verify the initial results presented here. For a more accurate comparison as to the effectiveness of the two assays, the potential issues with the V4-6 PCR regarding the bias of amplification towards members of the Actinobacteria would require further investigation. Secondly an increased depth of sequencing would allow a better understanding of the community species richness and evenness. These estimates will be aided by accurate denoising of datasets and removal of singleton sequences. This is particularly important as it has been shown that most pyrosequencing singletons are artifactual and therefore falsely inflate the diversity of a sample under analysis (Tedersoo *et al.*, 2010). The method employed here for denoising is noted for its heavy computational requirements (Zinger *et al.*, 2011) and ultimately proved a barrier to continued analysis within this study. Lastly, although there are indications as to the impact of antibiotic treatment on the soil microbiome, the statistical tests employed here did not provide the required confidence level; although those that were identified do warrant further investigation in part due to the documented presence of class 1 integrons.

Chapter 6

Microbial Community Profiling of Wet and Dry Season Aerosol Samples from Northern Australia

6.1 **Introduction**

6.1.1 **Microbial Diversity in the Atmosphere**

Aerobiology has been of interest to scientists for centuries (Gregory, 1970). In spite of this the ecology, diversity, distribution and interactions of atmospheric microorgansism remains poorly understood. This is in part due to the difficulties in analysing low densities of microorganisms present in atmospheric samples, even with sensitive molecular biology techniques (Womack *et al.*, 2010). However, it is known that hundreds of thousands of individual microbial cells can exist in a cubic metre of air (Burrows *et al.*, 2009) representing perhaps hundreds of unique taxa (Brodie *et al.*, 2007; Fierer *et al.*, 2008; Bowers *et al.*, 2009). Although the relative importance remains unknown, there are two main terrestrial sources of bacteria in the near-surface atmosphere 1) leaf surfaces and 2) soil (Lindemann and Upper, 1985; Lighthart and Shaffer, 1994). Culture based studies have suggested, as was shown previously with terrestrial and aquatic systems, that microbial densities vary with space, time and environmental conditions. For example microbial density was shown to decrease with altitude (Fulton, 1966) and that there were both seasonal and diurnal variation (Bovallius *et al.*, 1978; Lindemann and Upper, 1985; Lighthart and Shaffer, 1995; Tong and Lighthart, 2000; Fang *et al.*, 2007). Comparative studies examining temporal variation demonstrated both pronounced differences on daily (Fierer *et al.*, 2008) and seasonal scales (Després *et al.*, 2009) to those with relatively static community structures across time (Bowers *et al.*, 2009; Pearce *et al.*, 2010). Taking into account the presence of viable but not culturable and non-viable microbes within the airborne phase, current reasoning considers that this

community can be separated into three broad groupings based upon metabolic activity. These are 1) those that are not metabolically active, for instance spores, 2) metabolically active but not reproducing – for these cells, atmospheric dispersal is seen as an accidental occurrence, and 3) metabolically active and reproducing. Evidence for the final grouping comes from the fact the large portions of the atmosphere has environmental characteristics consistent with other habitats where microbial life can thrive. Furthermore, at least some recovered microbes have been shown to be metabolically active (Womack *et al.*, 2010). However, it must be remembered that <1% of airborne microorganisms are culturable (Pace, 1997). Consequently, considerable effort has now been focussed on the application of molecular methods, in particular sequencing and microarray technologies, to aid in the better definition of community structure and changes therein. One such area has been focussed on expanding our knowledge of the occurrences of pathogens in bioaerosols. Leski *et al.*, (2011) screened airborne dust collected in nineteen locations in Iraq and Kuwait for the presence of a broad range of human pathogens; areas known to be some of the largest sources of airborne dust on earth (Washington *et al.*, 2003). Here the authors used a high-density sequencing microarray technique choosing not to pursue 16S rRNA profiling owing to the limited resolution for some species. They were able to identify *Coxiella burnetti* and *Clostridium perfringens* as well as sequences from other genera such as *Mycobacterium*, *Bacillus* and *Brucella*. The presence of *C. burnetti* was further confirmed using real-time PCR although clearly no inference as to viability and thus disease potential could be drawn. In a study of urban aerosols using custom DNA microarrays (Phylochip), Brodie *et al*., (2007) demonstrated the presence of organisms closely related to *Francisella*

(although not *F. tularensis*, the causative agent of tularaemia). Phylogenetic near-neighbours of *Bacillus anthracis, Rickettsia, Clostridium botulinum* types C and G, and both *Burkholderia mallei* and *B. pseudomallei* were also detected (Brodie *et al.*, 2007). Although no inference was made to the potential for these organisms to cause disease within a human population, there are pathogens for which this has been hypothesised where associations with weather / climatic conditions and spread within bioaerosols is well characterised. These include the epidemics of meningococcal meningitis in West Africa (Sultan *et al.*, 2005), Q fever outbreaks across Europe (Tissot-Dupont *et al.*, 2004) and melioidosis in Northern Australia (Currie and Jacups, 2003).

6.1.2 ***Burkholderia pseudomallei*** **and Community-acquired Melioidosis**

Melioidosis is an infectious disease associated with exposure to mud and pooled water and is caused by the environmental Gram negative β-proteobacterium, *Burkholderia pseudomallei* (Cheng and Currie, 2005). As discussed in Section 1.2.1, although the majority of cases result from percutaneous inoculation, it is a also a major cause of community-acquired bacteremic pneumonia and septicaemia in endemic areas such as the tropical 'Top End' of the Northern Territory of Australia (Currie *et al.*, 2000) and South-east Asia (Chaowagul *et al.*, 1989). The correlation between weather conditions and the incidence of melioidosis is well characterised with several studies linking monsoonal rains and Typhoons to cluster outbreaks (Currie and Jacups, 2003; Ko *et al.*, 2007). In Australia, for example, 85 % of cases occur during the wet season (Currie *et al.*, 2000). In a detailed epidemiological survey of patients reporting to the Royal Darwin Hospital from the Top End of the Northern Territory (Australia), Cheng *et al.*, (2008) examined the incidence of *B. pseudomallei*-associated pneumonia

and severe sepsis in 387 culture-confirmed cases. There was a clear increase in the incidence of *B. pseudomallei* bacteremic pneumonia during the wet season (60 % pneumonia; 25 % severe sepsis) compared to the dry (26 %; 13 %) (Cheng *et al.*, 2008). It is postulated that aerosolisation of *B. pseudomallei* as a result of heavy rainfall leads to a shift towards inhalation as the mode of infection and the consequences of a severe illness (Currie and Jacups, 2003). However, to date no study has specifically demonstrated the presence of *B. pseudomallei* in bioaerosols from endemic areas. With the advances in high-throughput sequencing technology that have already been discussed, there are opportunities for the application of these techniques for both the identification of aerosolised *B. pseudomallei* and the characterisation of the bioaerosol in areas where the disease is endemic.

6.2 **Aims**

The aim of this chapter was to investigate the suitability of metagenomic and prokaryotic community profiling approaches, using high-throughput sequencing, for the detection of bacterial pathogens in aerosol samples. The hypothesis here is that climatic conditions (e.g. rainfall) increase the diversity of airborne microbial communities and, additionally, the prevalence of *B. pseudomallei* in endemic regions. Planned analyses of soil taken from the same locations, using the techniques assessed in Chapter 5, were prevented as appropriate permissions to remove samples from sites of Aboriginal heritage could not be acquired. The specific aims were as follows:

- Using 16S rRNA amplicon sequencing, compare the microbial communities of aerosol samples collected during the wet and dry seasons of Northern Australia.
- Perform comparative, total community DNA profiling using shotgun sequencing approaches on aerosol samples.
- Determine the utility of shotgun metagenomic sequencing for the accurate identification of bacterial pathogens.
- Determine the prevalence of *Burkholderia* sp. in particular *B. pseudomallei*, in collected aerosol samples.

6.3 **Analysis of Aerosol Samples from Northern Australia**

6.3.1 **Samples**

DNA extractions (200 µl) from air samples, collected in Darwin, Australia were kindly provided by Dr Lois Blackman (Defence Science Technology Organisation (DSTO)). These samples were taken during trials to examine the prevalence of airborne *B. pseudomallei* during the wet and dry seasons of Northern Australia. Air samples (>1000 m^3 per 24 h) were collected at a single site using horizontal wet-walled cyclones (HWWC) into 10 mM Hepes/0.01 % tween80 buffer. Samples (500 ml) were filtered through 0.22 µM nitrocellulose filters prior to DNA extraction in 1 ml InstaGene (Bio-Rad Laboratories Ltd) in a screw cap 1.5 ml microfuge tube. Samples were vortexed and subsequently heated in an Eppendorf thermomixer with shaking (1400 rpm for 15 minutes). The tubes were placed on ice for 5 minutes before centrifugation at 16,100 g at 4 °C. The aqueous top layer was removed and used in subsequent analyses.

Samples, relating to specific days from a twenty-nine day sampling period, were selected based on the detection of *B. pseudomallei* by real-time PCR analysis targeting the type three secretion system one (TTSS1) and performed at DSTO. Here, *B. pseudomallei* had been detected following incubation of filters in Ashdown's liquid media (15 g Tryptone, 1 ml crystal violet solution, 25 mg Colimycin and 5 mg Gentamicin per litre) for 48 h at 42 ^{o}C (Blackman and Kennedy, unpublished). Chosen samples are listed in Table 6.1.

Table 6.1: DNA extracts of collected aerosol samples from Darwin, Australia. [1] DDS – Darwin Dry Season, DWS – Darwin Wet Season; numbers represent sampling days [2] Based on real-time PCR analysis of cultures following incubation of filters in Ashdown's liquid media.

Sample	Season	Sample Name[1]	Amplification of *B. pseudomallei*[2]
1	Dry	DDS10	-
2	Dry	DDS11	-
3	Dry	DDS12	-
4	Dry	DDS13	-
5	Dry	DDS14	-
6	Dry	DDS15	-
7	Dry	DDS16	-
8	Dry	DDS18	-
9	Dry	DDS24	-
10	Wet	DWS10	-
11	Wet	DWS11	+
12	Wet	DWS12	+
13	Wet	DWS13	+
14	Wet	DWS14	+
15	Wet	DWS15	+
16	Wet	DWS17	-
17	Wet	DWS18	+
18	Wet	DWS24	+

6.3.2 **Comparing full-length 16S rRNA and V1-3 for Taxonomic Assignment within the Burkholderiaceae**

The goals of this study were both to characterise the microbial community in aerosol samples and determine prevalence of *Burkholderia* sp. Consequently, the utility of partial 16S rRNA analysis for accurate species-level taxonomic assignment within the Burkholderiaceae was firstly compared to full-length equivalents. Eighty-three full-length sequences were downloaded from the Ribosomal Database Project (RDP release 10) (Cole *et al.*, 2009). These sequences represented type strains only and were >1500 bp in length. The alignment (Figure 6.1) was generated using ClustalW with Slow/Accurate pair-wise alignment in MegAlign (DNASTAR LaserGene version 8). Fragments representing variable regions V1 to V3 (*Escherichia coli* position 8 to 533) were

selected using universal primers 8F and 534R. Again alignments were generated using ClustalW and are shown in Figure 6.2.

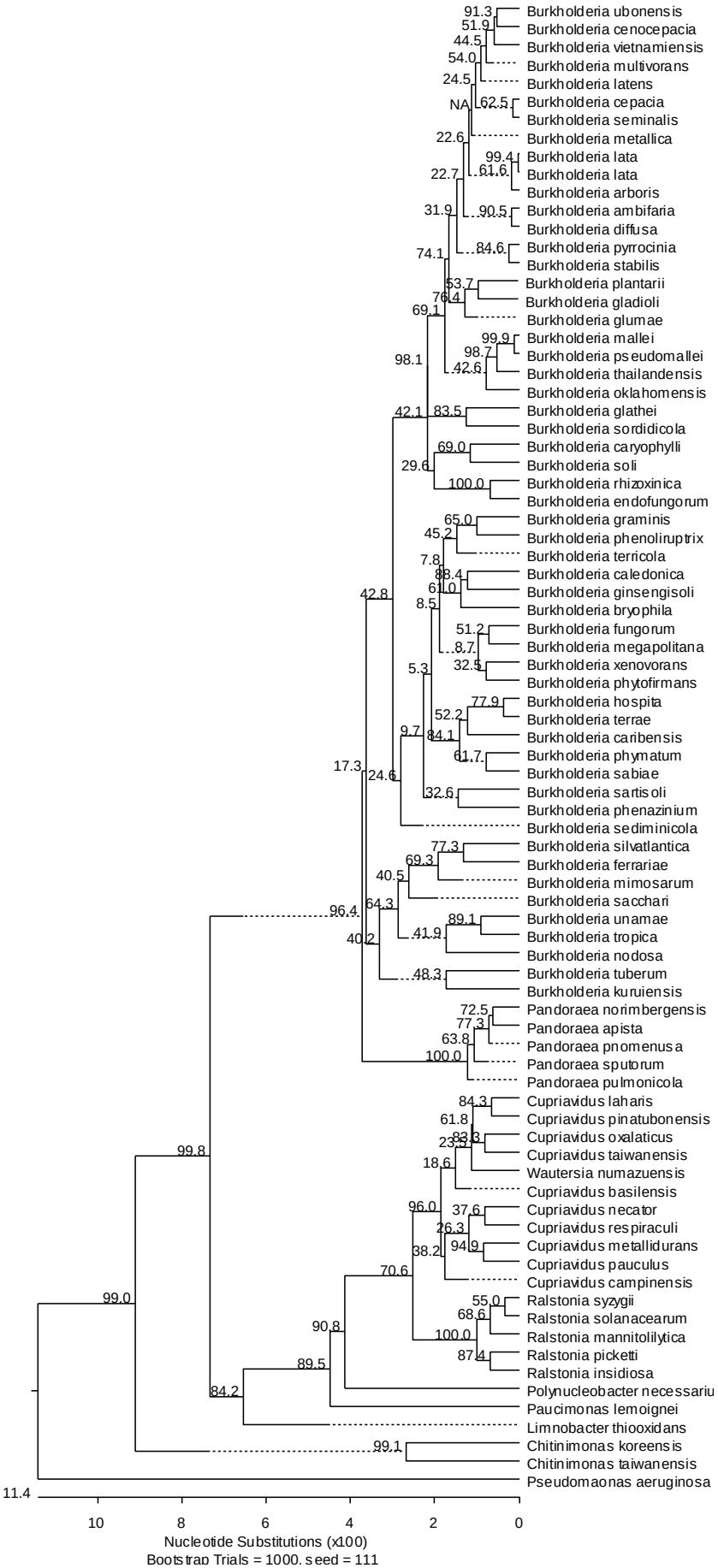

Figure 6.1: ClustalW Alignment of full-length 16S rRNA for type strains of the Burkholderiaceae.

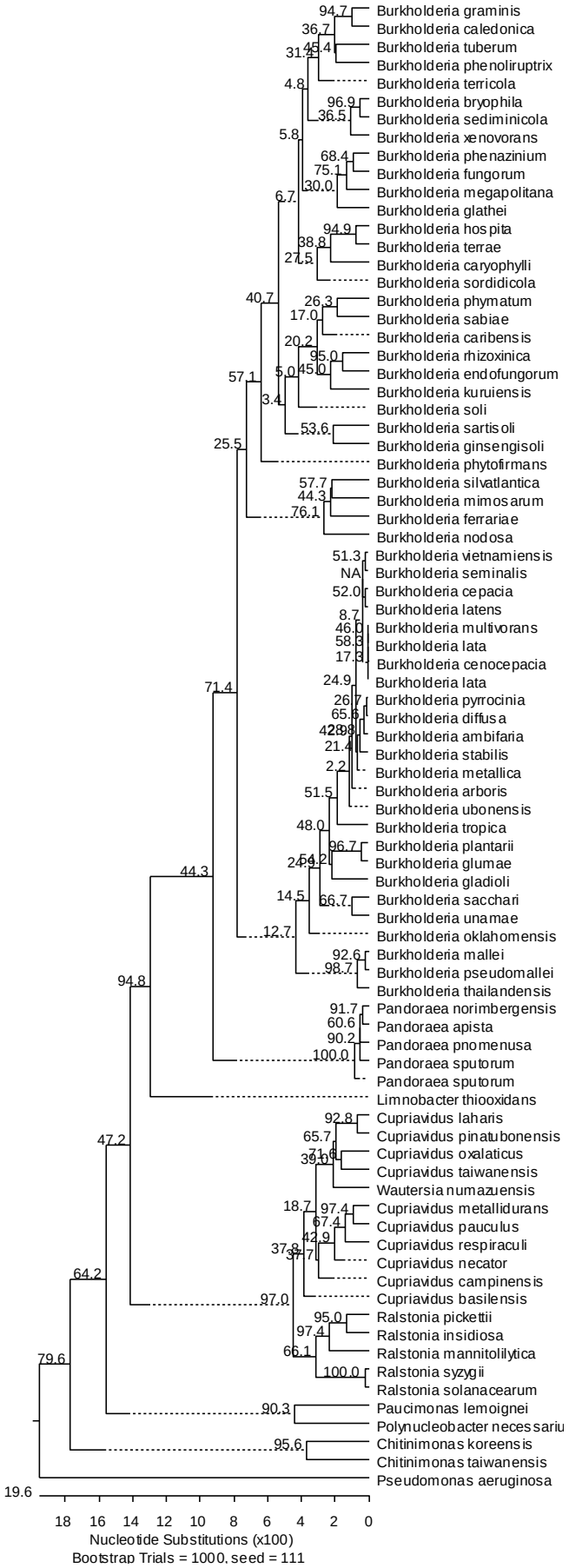

Figure 6.2: ClustalW Alignment of 16S rRNA V1-3 for type strains of the Burkholderiaceae.

There appears to be little loss in phylogenetic resolution when regions V1-3 are compared to full length 16S rRNA sequences. There is tighter grouping observed for members of the *Burkholderia cepacia* complex (*B. cepacia*, *B. cenocepacia*, *B. multivorans*, *B. vietnamiensis*, *B. stabilis*, *B. dolosa*, and *B. ambifaria*). However, importantly there remains separation for *B. pseudomallei*, *B. mallei*, *B. thailandensis* and *B. oklahomensis*.

6.3.3 **Amplification of V1-3 16SrRNA**

Selected PCR primers for amplification of V1-3 of 16S rRNA were based on the universal 8F and 534R primers (Table 2.6). PCR was performed in 50 µl reaction volumes on a GeneAmp® PCR System 9700 (Applied Biosystems, UK). Full assay and cycling conditions can be found in Tables 2.6 and 2.7. Results are shown in Figure 6.3. Samples DDS10, DDS12, DDS16 and DDS24 were not released from the ACDP/ACGM level 3 laboratories for analysis for safety reasons owing to the presence of culturable organisms within the DNA extract. These samples however were analysed using identical PCR conditions as detailed above except that the Smartcycler® (Cepheid, France) was used in place of the GeneAmp® PCR System 9700.

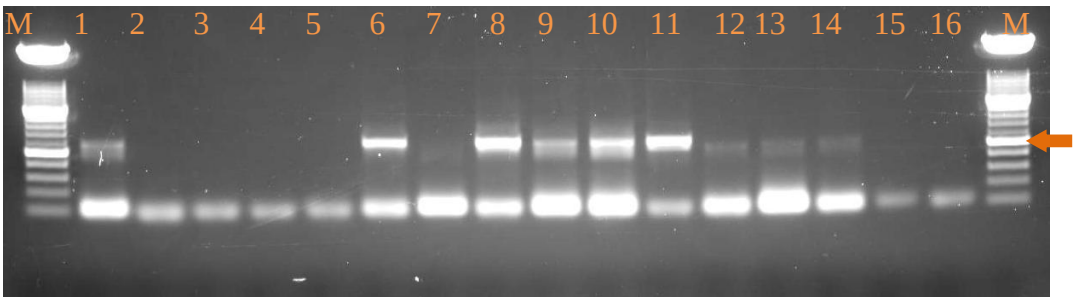

Figure 6.3 E-gel® (2 %) showing amplification of 16S rRNA V1-3 from Darwin aerosol samples. Lane identities are as follows: 1 – DDS11; 2 – DDS13; 3 – DDS14; 4 – DDS15; 5 – DDS18; 6 – DWS10; 7 – DWS11; 8 – DWS12; 9 – DWS13; 10 – DWS14; 11 – DWS15; 12 – DWS17; 13 – DWS18; 14 – DWS24; 15-16 no sample control; M – molecular weight marker XIV. Correctly sized products are indicated by orange arrow.

Amplification was only observed for one dry season samples (DDS10). In order to discount primer-related non-amplification of dry season samples, Golay-barcoded PCR primers based on 27F and 338R universal primer were also used to amplify 16S rRNA in the same samples (Fierer *et al.*, 2008). Results are shown in Figure 6.4. Again only one dry season sample produced an amplicon of the expected size.

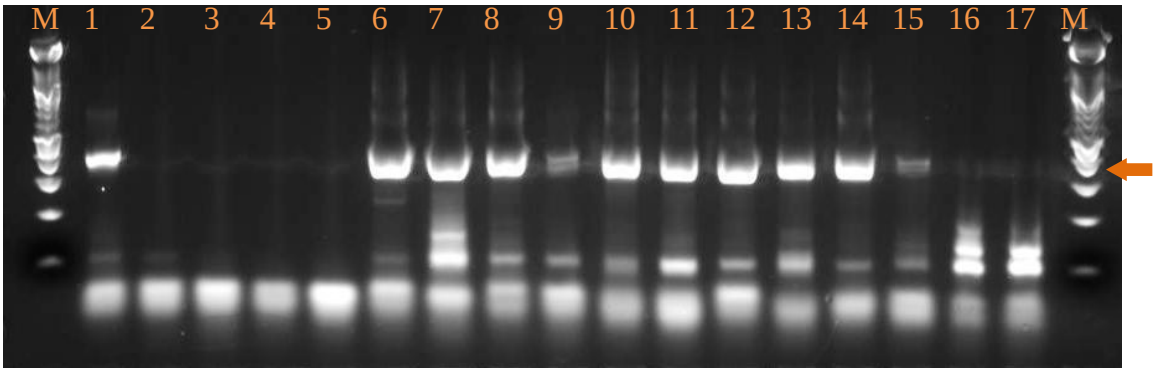

Figure 6.4: E-gel® (2%) showing amplification of 16S rRNA V1-3 using Golay bar-coded primers from Darwin aerosol samples. Lane identities are as follows: 1 – DDS11; 2 – DDS13; 3 – DDS14; 4 – DDS15; 5 – DDS18; 6 – DWS10; 7 – DWS11; 8 – DWS12; 9 – DWS13; 10 – DWS14; 11 – DWS15; 12 – DWS17; 13 – DWS18; 14 – DWS24; 15 – *Bacillus anthracis* genomic DNA; 16-17 no sample control; M – molecular weight marker XIV. Correctly sized products are indicated by orange arrow.

6.3.3.1 454™ Sequencing and Pre-processing Analysis

Amplicons from the eleven samples (generated in triplicate PCRs) from which 16S rRNA amplification was achieved were pooled and sequencing libraries constructed according to 454™ protocols. In total 497, 102 raw sequences were generated from a single sequencing run. Sequences were processed using the Quantitative Insights in Microbial Ecology (QIIME) pipeline (Caporaso *et al.*, 2010). Following clustering by barcode sequence, a number of pre-processing quality filtering steps were undertaken prior to analysis. These included selecting sequences within a length range of 200 – 600 bp, removing those with ambiguous sequences, quality scores of <25, homopolymer stretches greater than

6 bp and those with mismatches in the primer sequence. Denoising of sequences was also undertaken. Following this, chimeric sequences were removed using the QIIME pipelines ChimeraSlayer script. The full output for each sample is shown in Table 6.2.

Table 6.2: Quality filtering outputs of 16S rRNA sequences for wet and dry season aerosol samples.

Total Sequences	497, 102
200 – 600 Ambiguous <Q25 Homopolymer Primer Mismatch	- 81, 800 - 40, 429 - 10, 106 - 1, 637 -220, 265
Total	142, 444
Total (Post Chimera Removal)	140, 286
Sample Name	Sequences
DDS10	20, 262
DWS10	17, 189
DWS11	6, 279
DWS12	17, 735
DWS13	18, 424
DWS14	19, 051
DWS15	16, 165
DWS17	11, 301
DWS18	8, 838
DWS24	5, 042

6.3.3.2 Phylogenetic Diversity

Following quality checks, denoising and chimera removal, remaining sequences were clustering into OTUs using a 0.04 dissimilarity index (96 % sequence similarity) as suggested by Kim *et al.*, (2010) using uclust (Edgar 2009, unpublished). Taxonomic assignments were made using the RDP classifier, Greengenes database and standard QIIME analysis pipelines. Briefly, following OTU clustering, representative sequences were aligned using PyNAST (Caporaso *et al.*, 2010) with a pre-aligned Greengenes 16S rRNA dataset from http://greengenes.lbl.gov. Chimeric sequences were removed using ChimeraSlayer, the alignment filtered and a phylogenetic tree built by FastTree (Price *et al.*, 2009) using the remaining sequences.

For all samples Proteobacteria were the dominant phyla representing over 99 % of all OTUs identified. At class level Betaproteobacteria were dominant, followed by Alphaprotoeobacteria and Gammaproteobacteria. Betaproteobacteria represented between 85 and 96 % of OTUs for the majority of samples. Samples DWS15, 17 and 18 and 24 showed a variation in this trend with Alphaproteobacteria increasing in prevalence. For samples DWS17 and 18 Alphaproteobacteria became the dominant class with 74 and 57 % of the total OTUs respectively. Order-level OTU assignments are shown in Figure 6.3.

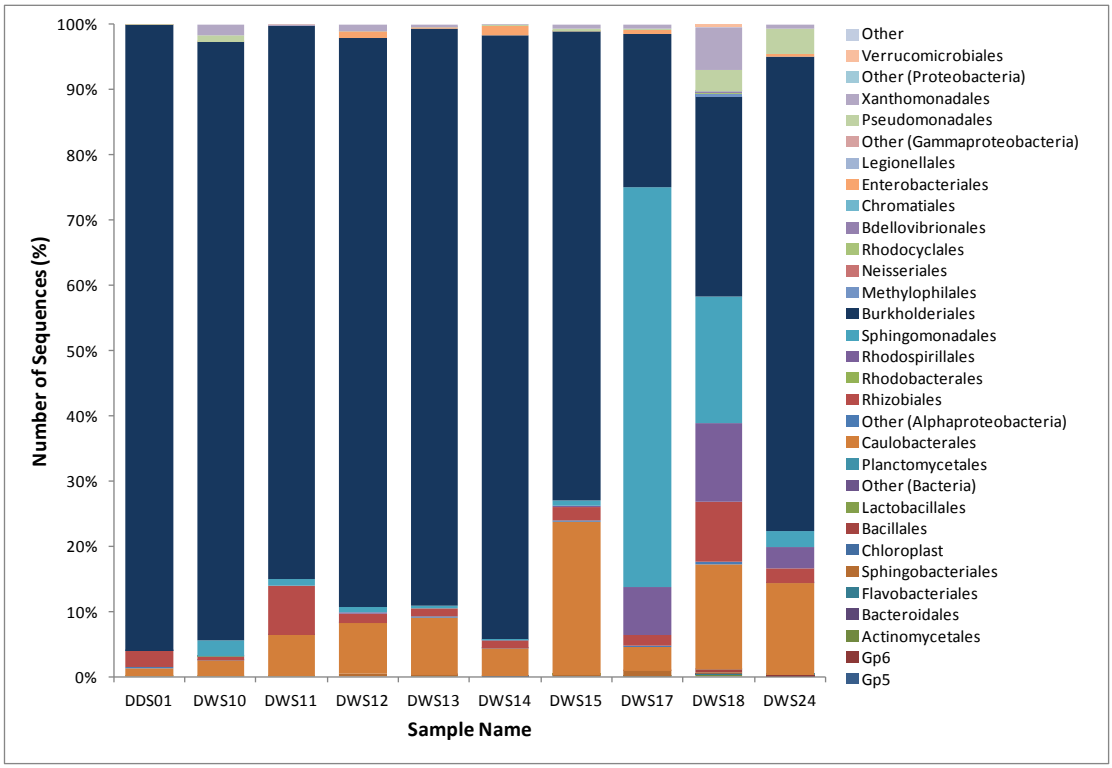

Figure 6.3: Distribution of bacterial order taxonomic assignments for collected aerosol samples as determined by 16S V1-3 analysis.

Across most samples the Burkholderiales represent the most abundant order. There are however notable increases in the relative abundance of Caulobacterales, Sphingomonadales and Rhodospirillales in samples DWS15,

17, 18 and 24. Examinations of the bacterial genus associated with these changes (Figure 6.4) reveal that Novosphingobium, and Acidello increase across these samples. Additionally, among the Burkholderiales there also appears to be a flux between the genera of *Burkholderia*, *Ralstonia* and *Herbaspirillum*.

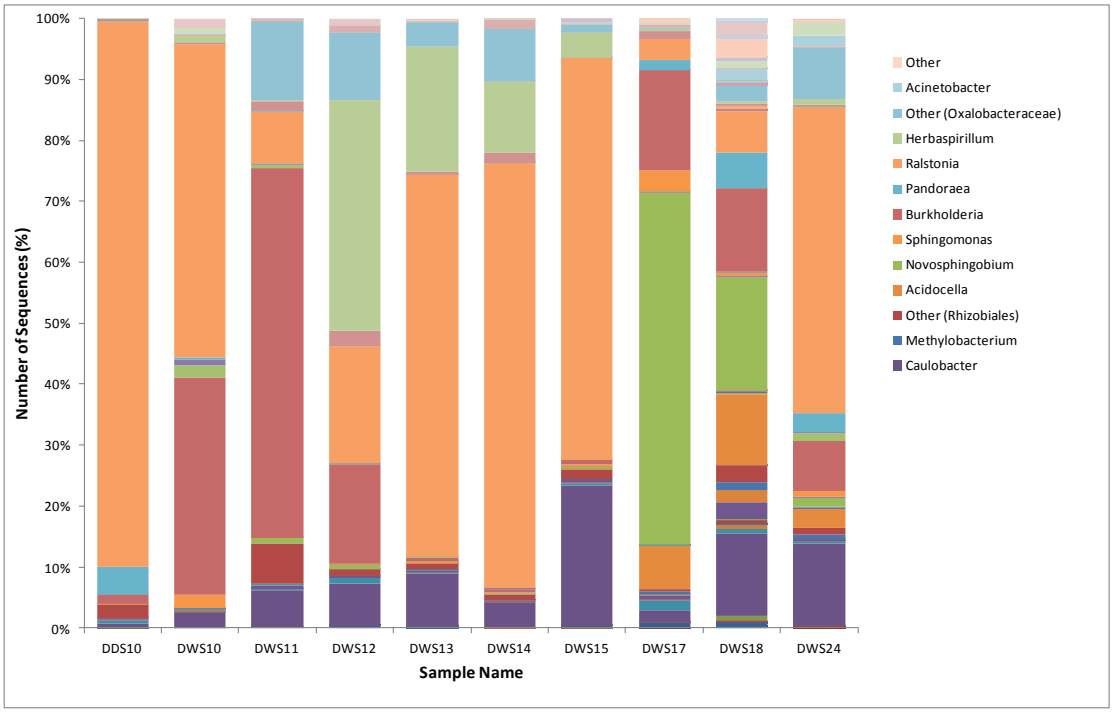

Figure 6.4: Distribution of bacterial genus taxonomic assignments for collected aerosol samples as determined by 16S V1-3 analysis. For clarity only the most abundant are named.

6.3.3.3 Alpha Diversity

The diversity and richness of OTUs within each sample (known as alpha diversity) was calculated in QIIME. Metrics used included observed species richness (the count of unique OTUs within a sample), Chao1 (species richness), Phylogenetic Distance and the Shannon index. Observed species richness for each sample is shown in Figure 6.5. Apart from samples DWS18 and DWS24 sampling depth appeared sufficient to capture the OTU richness within the remaining samples.

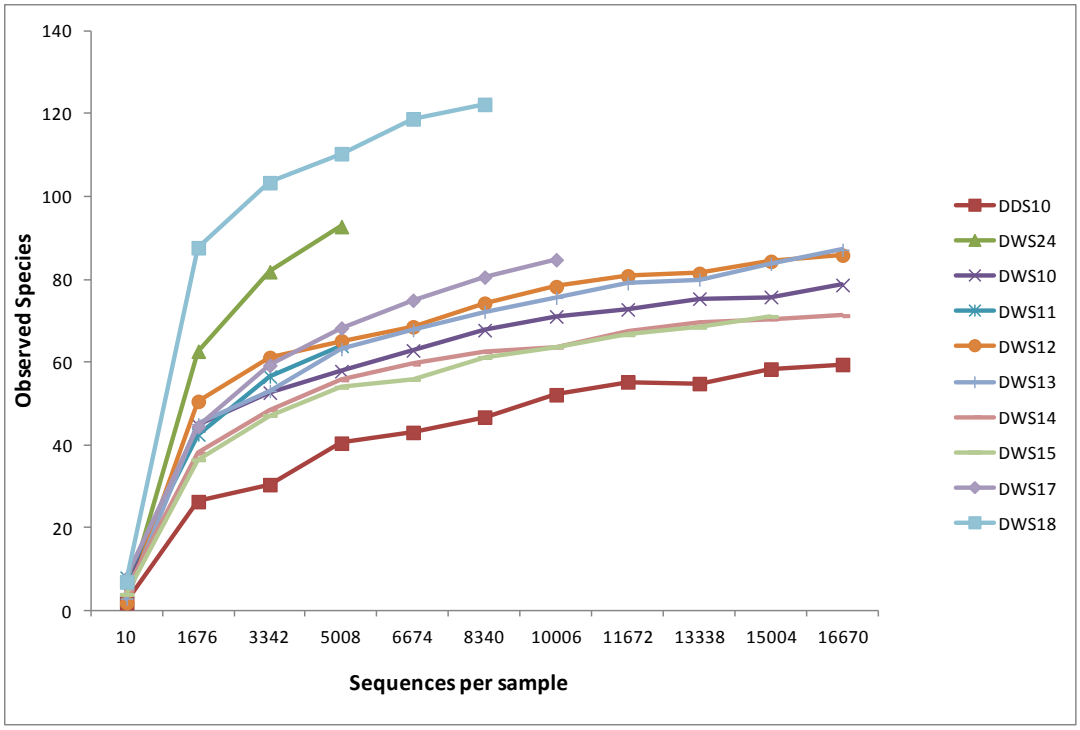

Figure 6.5: Graph showing observed species richness for wet and dry season samples as determined through 16S rRNA V1-3 analysis. Richness was computed in QIIME v. 1.3.0

Phylogenetic diversity (PD) and Chaol estimator were also used to evaluate community richness along with Shannon index for community diversity. Results from these and observed species richness are shown in Table 6.3.

Table 6.3: 16S rRNA diversity analysis of wet and dry season samples

Sample	Observed species richness	Chaol	Phylogentic Diversity (PD)	Shannon index
DDS10	59.5 (±2.07)	77.32 (±7.83)	4.35 (±0.18)	0.88 (±0.006)
DWS10	78.7 (±0.48)	89.33 (±2.13)	4.91 (±0.02)	2.51 (±0.002)
DWS11	64.0 (±2.40)	83.21 (±12.00)	4.14 (±0.10)	2.27 (±0.02)
DWS12	85.9 (±1.20)	106.94 (±4.46)	4.91 (±0.07)	3.18 (±0.02)
DWS13	87.2 (±1.14)	112.19 (±9.47)	5.90 (±0.10)	2.25 (±0.004)
DWS14	71.3 (±1.25)	85.61 (±6.76)	4.74 (±0.10)	2.15 (±0.005)
DWS15	71.1 (±0.88)	84.29 (±3.48)	5.80 (±0.09)	1.90 (±0.005)
DWS17	84.8 (±1.32)	102.74 (±4.48)	5.81 (±0.04)	2.56 (±0.006)
DWS18	122.3 (±0.67)	144.75 (±5.88)	9.35 (±0.05)	4.41 (±0.006)
DWS24	92.8 (±0.42)	121.36 (±2.73)	6.49 (±0.05)	3.30 (±0.002)

Comparing all four metrics it appears that there is an increase in diversity for samples DWS17, 18 and 24. Sample DWS18 in particular revealed the greatest levels of diversity. The lowest diversity was observed for the single dry season sample, DDS10. To determine if the diversity variations observed in Table 6.3 were a result of changes in local climate, Chao1 and observed species were plotted against rainfall (mm) for each sampling day (Figure 6.6). Rainfall data was acquired from the Australian Government Bureau of Meteorology for weather station at Darwin Airport (station number 014015).

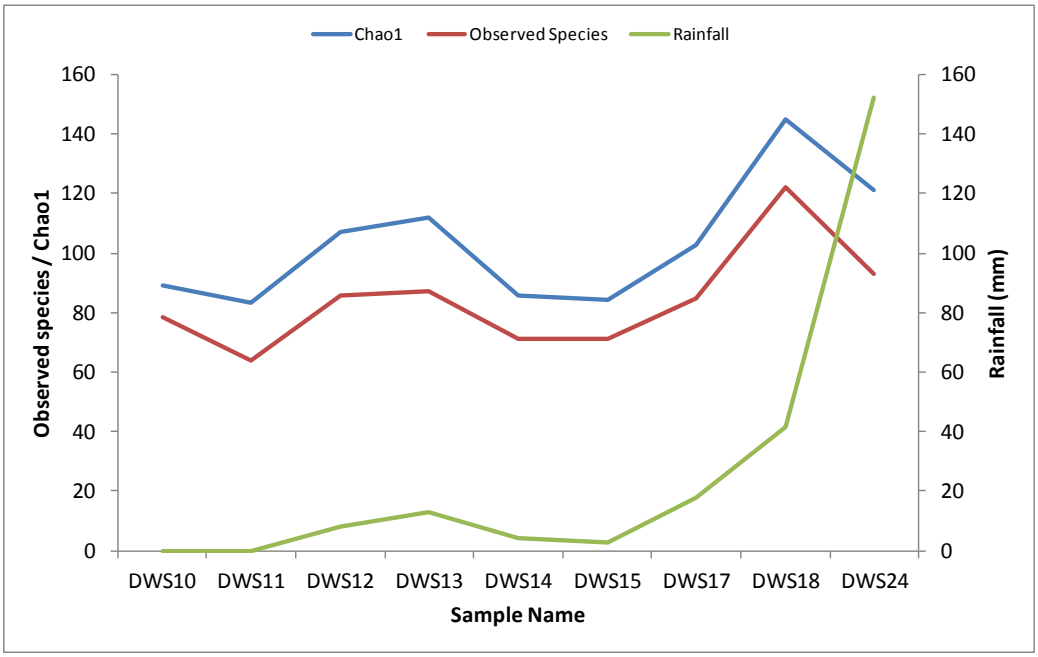

Figure 6.6: Graph showing diversity metrics (Chao1 and observed species) plotted against rainfall (mm) for wet season aerosol samples.

There is a possible correlation, based on the limited data presented here, between observed aerosol microbial diversity and rainfall. Increases in both Chao1 and observed species seen in DWS17, 18 and 24 in particular occur on days with the highest levels of rainfall (17.8, 41.8 and 152.2 mm respectively). This however does not associate with detection of *B. pseudomallei* (Table 6.1) as it was not detected in DWS17.

6.3.3.4 Beta Diversity

Comparison of samples based upon community composition was undertaken using unweighted Unifrac in QIIME. Unifrac generated a distance metric based upon the lineages contained within the samples under analysis (Lozupone and Knight, 2005). This metric was then used in hierarchical clustering analysis using unweighted-pair group method using average (UPGMA). Jackknife support was then used to assess the robustness of observed hierarchical

clustering. Ten replicates were performed using 4, 000 sequences from each sample. The resultant phenogram is shown in Figure 6.7.

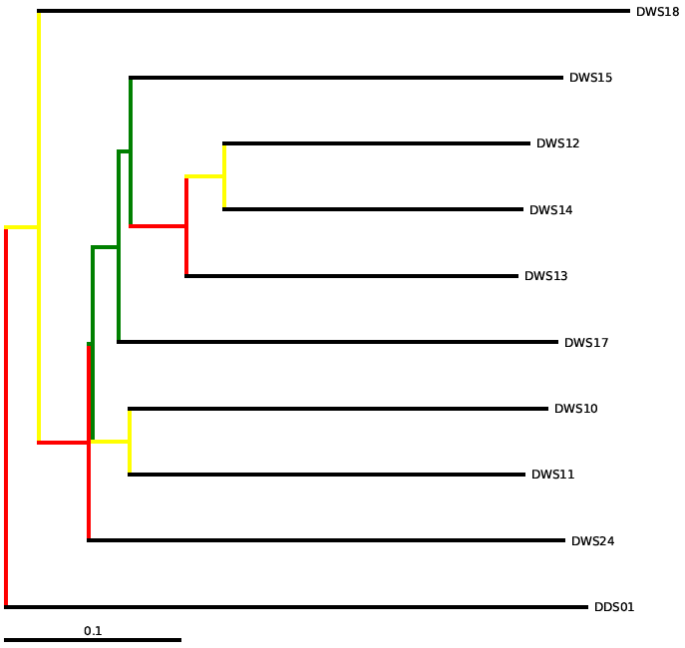

Figure 6.7: Jackknife phylogenetic tree showing hierarchical clustering of aerosol samples based on unweighted Unifrac distance metric and UPGMA. Colours represent jackknife support for nodes. Red 75 -100 %, yellow 50 – 75%, green 25 – 50 % and blue <25 %.

The clustering of samples in Figure 6.7 roughly correlates with that of the observed fluctuations in species richness and diversity (Figure 6.6). This is seen for samples DWS10 and DWS11 (forming one cluster) and DWS12, 13 and 14. The separate branching of samples DWS15, DWS17, DWS18 and DWS24 is unsurprising given the apparent changes in community structure, particularly at the genus level (Figure 6.4). However the separation of DWS15 would suggest that diversity is not fully explained by rainfall, as suggested by samples from the

preceding days. Of note is the presence of a highly supported (75 – 100 %) branch for the sole dry season sample, DDS10.

6.3.4 **Metagenomic Sequence Analysis of Simple Microbial Communities**

Shotgun metagenomic analysis for the detection of pathogenic microorganisms in environmental samples offers benefits over the use of 16S rRNA. This is foremost a result of the potential for identificaitojn of multiple gene targets that may be assigned to the pathogen of concern. However, the complexity of a metagenomic dataset is naturally greater than that of a 16S rRNA equivelant. In order to determine the extent to which accuracy of taxonomic assignment is impacted by the complexity of the sample undergoing metagenomic analysis, four mixtures of genomic DNA (gDNA) of *Burkholderia cenocepacia* J2956 and *Vibrio cholerae* N16961 were sequenced using 454 Sequencing™ technology (Roche, UK). These organisms were chosen on the basis of availability of genome sequences, phylogenetic separation at a high taxonomic level (Class) and associations with human infections; either opportunistically (*B. cenocepacia* and lung infections of cystic fibrosis patients (Holden *et al.*, 2009)) or epidemically / pandemically (*V. cholerae* (Wachsmuth *et al.*, 1994)). DNA was extracted using the Gentra Purgene Yeast/Bact. DNA Isolation kit (Qiagen, UK) according to manufacturer's instructions (Section 2.4.1) and quantified by PicoGreen® (2.4.10.2). These DNA were then mixed in final volumes of 100 µl as follows: the concentration of *V. cholerae* gDNA was maintained throughout the four samples at 2.5 µg whilst the gDNA from *B. cenocepacia* was reduced 10-fold in each sample.

6.3.4.1 454 Sequencing™

Genomic DNA mixture samples of *B. cenocepacia* and *V. cholerae* were sent to Eurofins MWG Operon (Germany) for sequencing. Each sample was loaded separately into 'medium' regions of a sequencing plate and sequenced using 454™ Sequencing Titanium chemistry. The similarities of the sequence outputs for each data-set are shown in Table 6.4.

Table 6.4: Sequence output for *V. cholerae* / *B. cenocepacia* samples as determined by analysis using MG-RAST.

Sample	gDNA concentrations: *V. cholerae* / *B. cenocepacia* (µg)	Total Bases (Mbp)	Reads	Average Fragment Length (bp)	Reads Aligned to Protein Features (%)	Reads Aligned to rRNA Features (%)
Dstl/05	2.5 / 2.5	114	208, 354	550 ± 21	85.6	1.9
Dstl/06	2.5 / 0.25	124	215, 008	578 ± 137	81.9	1.5
Dstl/07	2.5 / 0.025	110	197, 306	561 ± 123	86.4	1.6
Dstl/08	2.5 / 0.0025	136	229, 306	567 ± 124	83.4	1.4

6.3.4.2 MG-RAST Analysis

Sequenced datasets were submitted to the Metagenomics Rapid Annotation by Subsystem Technology server (MG-RAST) v 3.1.2 (Meyer *et al.*, 2008) for analysis. In order to assign *e*-values and alignment thresholds, randomised equivalents of the four datasets were generated using an in-house permutation tool, uShuffle, developed and kindly donated by Dr Carl Mayers (Dstl Porton Down). This Perl-script program randomises each sequence within the dataset whilst maintaining characteristics such as fragment length distribution, GC content and triplet codon usage. Consequently a dataset containing little or no biological relevance is created. The comparison of fragment length distribution and percent GC content for the *V. cholerae* (2.5 µg) / *B. cenocepacia* (0.25 µg) dataset is shown in Figure 6.8.

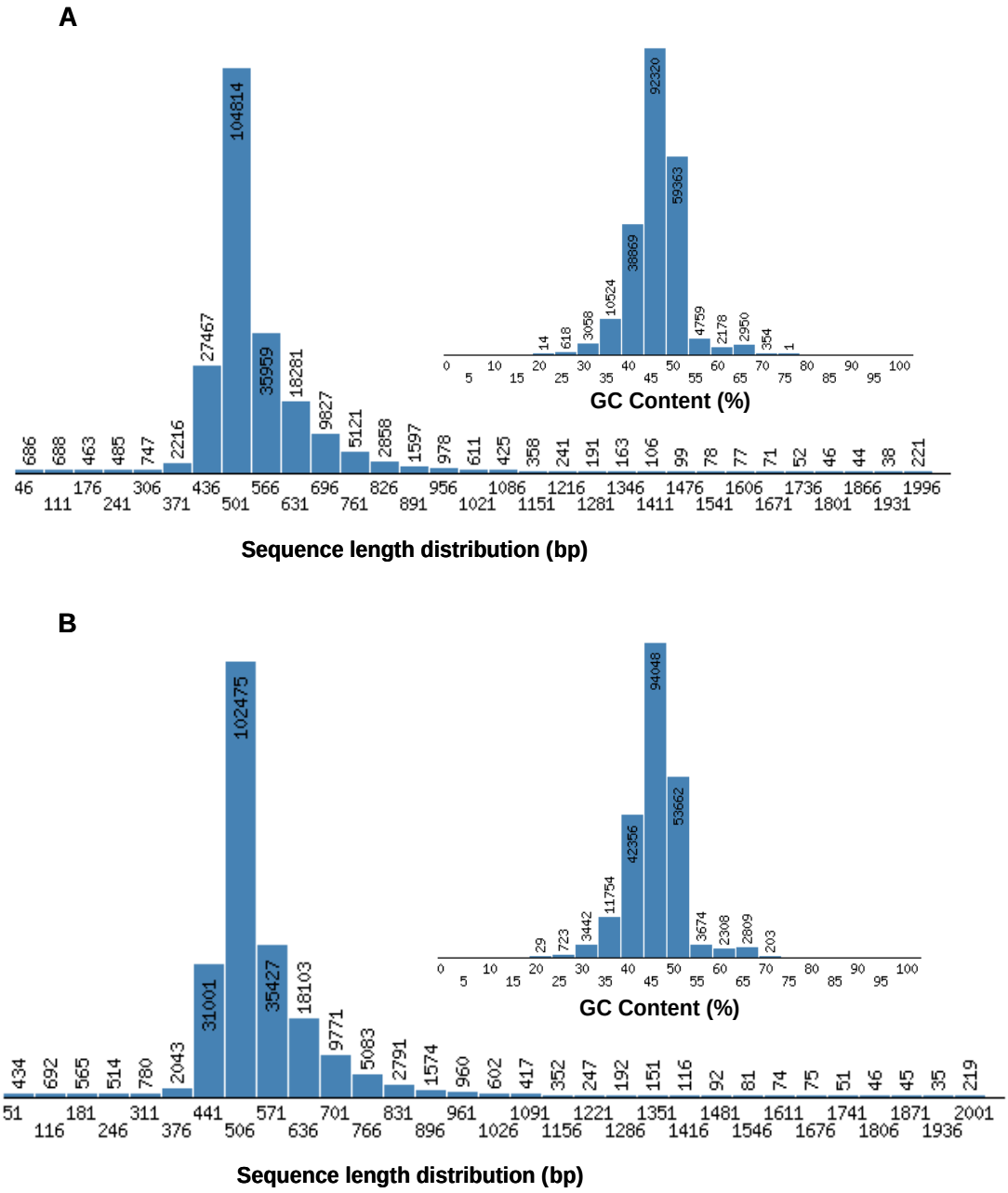

Figure 6.8: Comparisons of sequence length and GC content between the *Vibrio cholerae* / *Burkholderia cenocepacia* (2.5 µg / 0.25 µg) permuted dataset generated using uShuffle (A) and the non-permuted version (B).

These datasets were used to determine the minimum *e*-value and alignment cut-offs for sequence annotation in MG-RAST. From the analysis of permuted databases, an *e*-value of -5 and a minimum alignment length of 50 bp were determined to be the most appropriate metrics to apply.

Domain level annotation and the percentage of reads that were assigned to Vibrionaceae and Burkholderiaceae families for each sample (Dstl/05 to Dstl/06) are shown in Table 6.5. The percentage of reads that were assigned to the Burkholderiaceae followed the 10-fold dilution of *B. cenocepacia* J2956 across samples Dstl/05 to Dstl/06 with 25.3 %, 2.3 % and 0.4 % of sequences assigned respectively. Despite there being equal concentrations of genomic DNA from each organism a greater number of reads were assigned to the Vibrionaceae; a phenomenon that was maintained across all four samples.

Table 6.5: Sequence read distribution by Domain and Vibrionaceae / Burkholderiaceae Families.

Sample (gDNA concentration µg)	Read Distribution by Domain (%)			Read Distribution (%)	
	Bacteria	Eukaryota	Viruses	Vibrionaceae	Burkholderiaceae
Dstl/05 (2.5 / 2.5)	99.2	0.3	0.1	67.2	25.3
Dstl/06 (2.5 / 0.25)	99.2	0.2	0.1	92	2.3
Dstl/07 (2.5 / 0.025)	99.2	0.1	0.2	94.2	0.4
Dstl/08 (2.5 / 0.0025)	99.2	0.2	0.1	94	0.2

At genus level, *Burkholderia* sp. and *Vibrio* sp. were dominant across all four samples as expected. Results for sample Dstl/05 are shown in Figure 6.9. Vibrio genus assignments represented 66 % of the total sequences with 24.5 % of sequences assigned to Burkholderia. In contrast, when the samples were analysed using RDP only, the genus level assignments were 58 % and 37.5 % for Vibrio and Burkholderia respectively – much closer to the expected distribution.

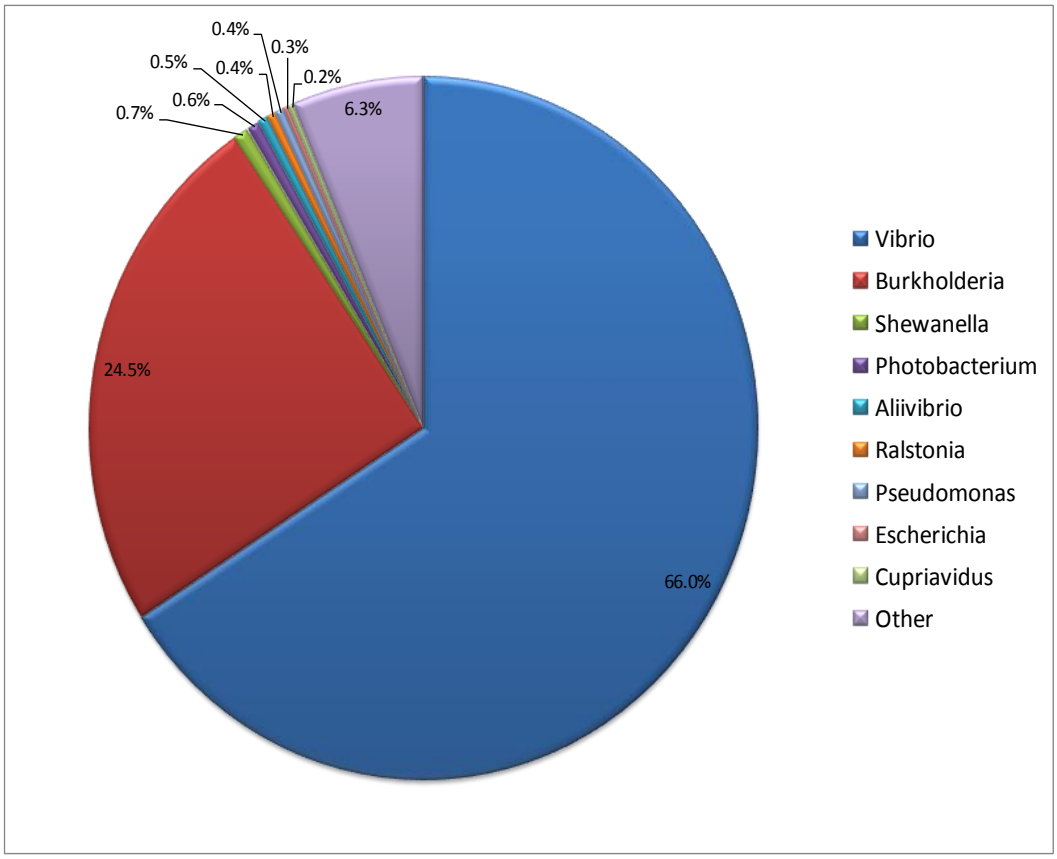

Figure 6.9: MG-RAST Genus-level Taxonomic Assignments for *Burkholderia cenocepacia* / *Vibrio cholerae* Genomic DNA (2.5 µg / 2.5 µg) Shotgun Sequence.

Finally, the distribution of sequences across species was determined in MG-RAST. Species-level taxonomies for each of the four samples are shown in Figure 6.10. For clarity only the top five species, by the total number of sequences assigned, are reported.

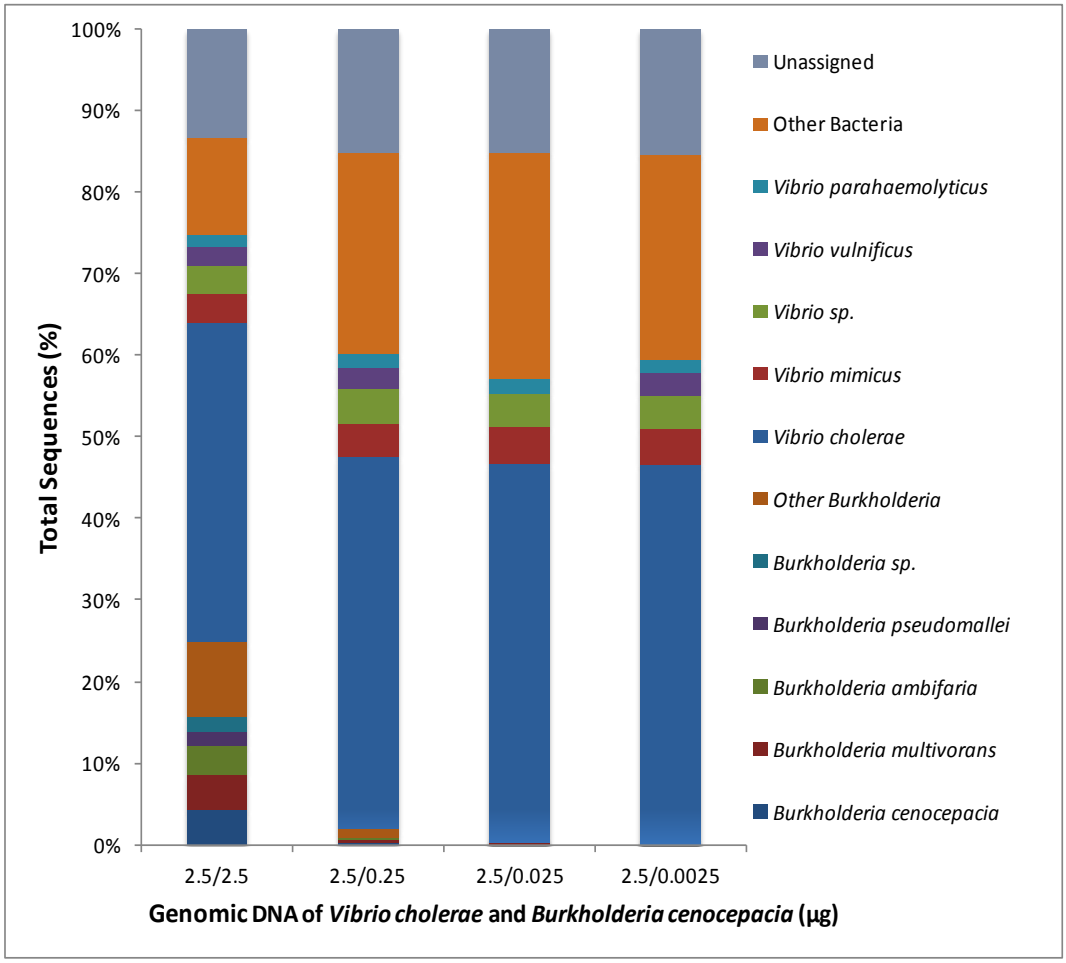

Figure 6.10: Species-level Taxonomic Assignments for Shotgun Sequences of *Vibrio cholerae* / *Burkholderia cenocepacia* Genomic DNA Mixtures.

For all four samples, *V. cholerae* was correctly identified as the dominant species of the Vibrio genus with ~272, 000 sequences. This was approximately 10-fold more than the next species, *V. mimicus*. For Dstl/05 and Dstl/08 a larger number of sequences were assigned to *B. cenocepacia* than any other *Burkholderia* sp. with 23, 194 and 45 respectively. For Dstl/06 and 07 however the species with the highest number of sequences assigned was *B. multivorans*. *B. multivorans*, like *B. cenocepacia*, is a member of the *B. cepacia* complex, a group of closely related species that share a high degree of genomic similarity (Mahenthiralingham *et al.*, 2008).

6.3.4.3 Whole Genome Amplification of Wet and Dry Season Samples

Shotgun sequencing of total community DNA from each of the wet and dry season aerosol samples was undertaken both to verify the observations of bacterial community structure and gain additional insights into structure and function. Unfortunately, concentrations within the InstaGene DNA extracts originally provided by DSTO were not sufficient to produce a pooled sample of 500 ng for the generation of a shotgun sequencing library. Consequently each sample listed in Table 6.1 underwent whole genome amplification (WGA) using the Qiagen® REPLI-g midi kit (Qiagen, UK) (Section 2.4.12). Amplification was performed on triplicate 5 µl volumes according to the manufacturer's instructions. DNA concentrations of WGA material were determined by PicoGreen® fluorometry (Section 2.4.10.2). Results are shown in Figure 6.11. One sample, DDS18, did not yield amplification products.

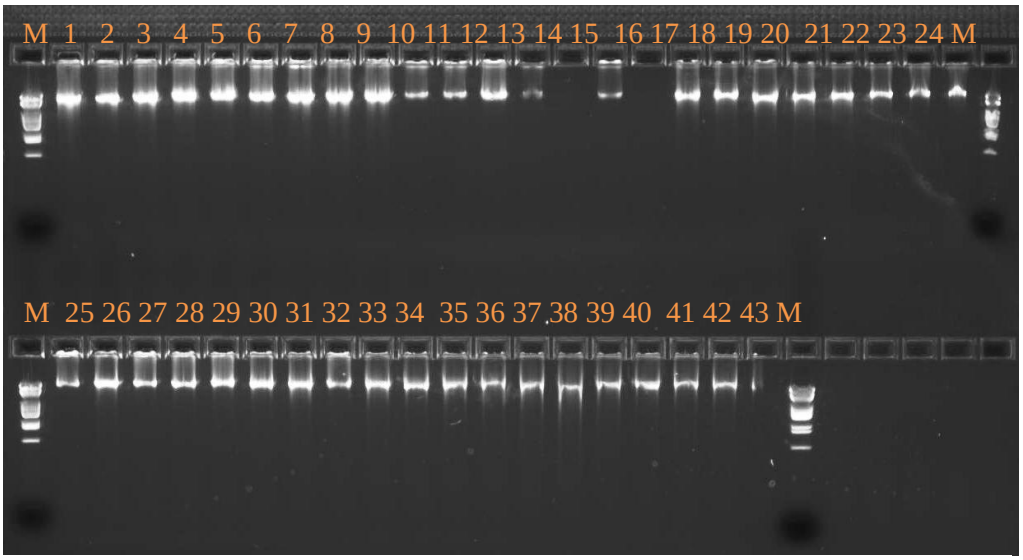

Figure 6.11: E-gel (0.8 %) showing REPLI-g whole genome amplified DNA extracts from wet and dry season aerosol samples. Samples were amplified in triplicate. Lane identities are as follows: 1-3 DDS11; 4-6 DDS13; 7-9 DDS14; 10-12 DDS15; 13-55 DDS18; 16-18 DWS10; 19-21 DWS11; 22-24 DWS12; 24-27 DWS13; 28-30 DWS14; 31-33 DWS15; 34-37 DWS17; 37-40 DWS18; 41-43 DWS24; M – molecular weight marker III.

Multiple-displacement amplification reactions produce hyperbranched structures that are not conducive to sequence analysis. These were removed using the method of Zhang *et al.*, (2006). The full method is detailed in Section 2.4.12.1. Briefly, approximately 1.5 µg of pooled WGA DNA from triplicate reactions first underwent microcon purification followed by Repli-PHI™ (Epicentre®, UK) WGA. This resulted in ~2 µg of material which then underwent digestion with S1 nuclease (NEB, UK). Lastly, the digested material underwent phenol/chloroform extraction and ethanol precipitation before visualisation on a 0.8 % E-Gel® and quantification using PicoGreen®. Results are shown in Figure 6.12. Samples were then pooled to produce a single sample for both the wet and dry season.

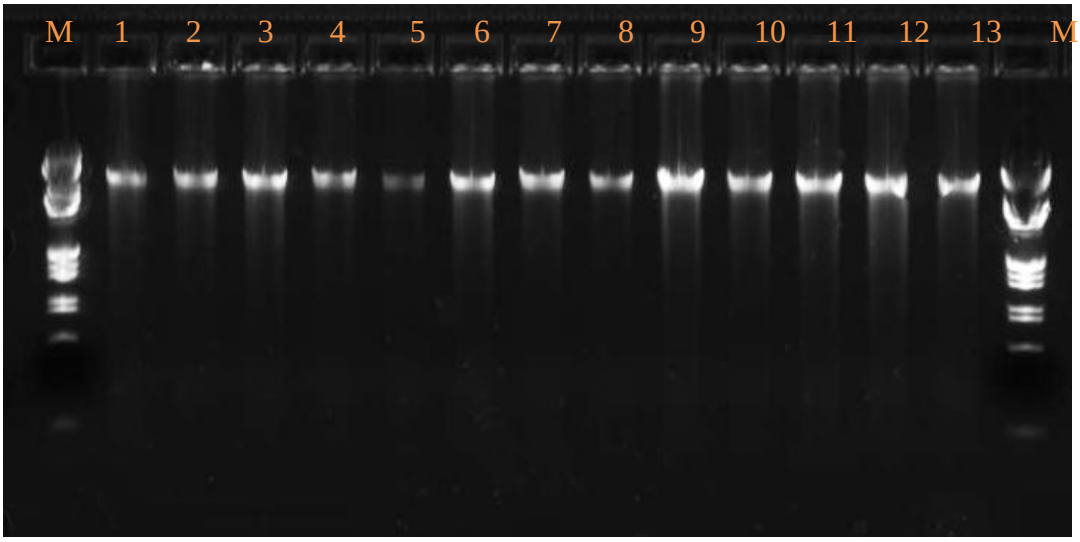

Figure 6.12: E-gel (0.8 %) showing debranched whole genome amplified DNA from wet and dry season aerosol samples. Lane identities are as follows: 1 – DDS11; 2 – DDS13; 3 – DDS14; 4 – DDS15; 5 – DWS10; 6 – DWS11; 7 – DWS12; 8 – DWS13; 9 – DWS14; 10 – DWS15; 11 – DWS17; 12 – DWS18; 13 – DWS24; M – molecular weight marker III.

6.3.4.4 454 Sequencing™ and Pre-processing Analysis

Pooled, whole genome amplified DNA (500 ng total) for both wet and dry season samples were shotgun-sequenced using 454 Sequencing™ Titanium chemistry protocols according to the manufacturer's instructions. The wet season sample yielded less sequence data than the dry and was consequently sequenced twice. Total, raw sequence output for each sample was 587 Mbp and 718 Mbp for the dry and wet season respectively. Automatic quality checking of sequences by MG-RAST v. 3.1.2 (Meyer *et al.*, 2008) left 486 Mbp (dry) and 493 Mbp (wet) of sequence per sample representing 928, 555 and 952, 173 sequences.

6.3.4.5 Phylum-level Diversity

Metagenomic analysis was undertaken using MG-RAST v. 3.1.2. The M5 non-redundant protein database (M5NR) was used with an *e*-value of -5 and a minimum alignment length of 50 bp. The diversity of bacterial phyla within both the wet and dry season samples is shown in Figure 6.13.

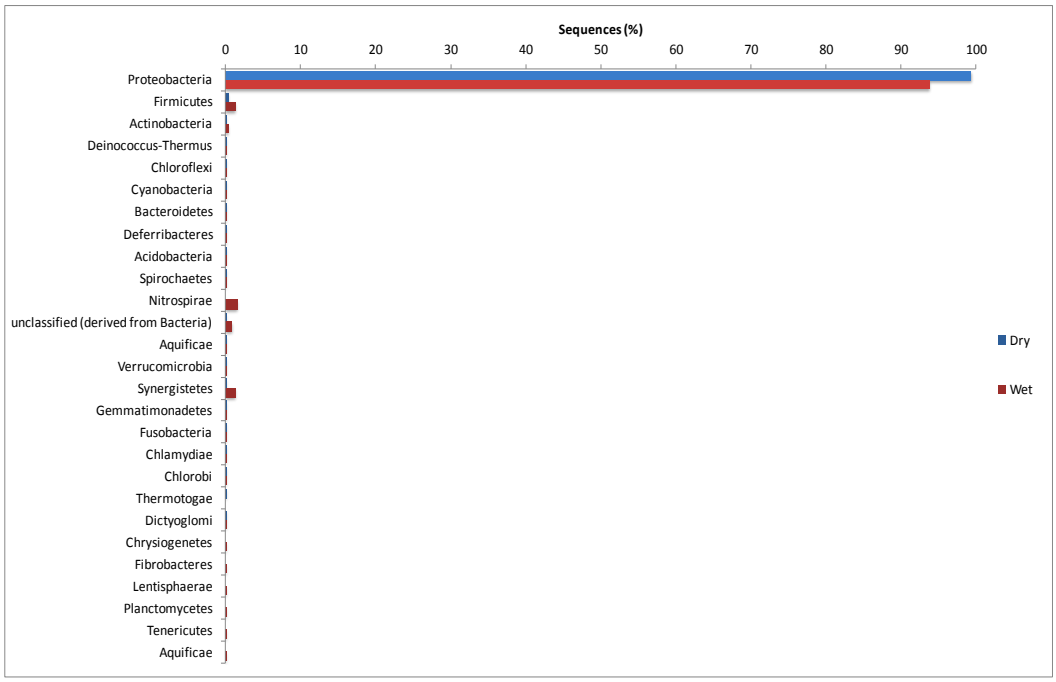

Figure 6.13: Distribution of bacterial Phyla within wet and dry season aerosol metagenomic samples. Annotation was performed using the M5NR database in MG-RAST v. 3.1.2 using an *e*-value of -5 and a minimum alignment length of 50 bp.

For both the wet and dry season aerosol metagenomic samples the Proteobacteria was the dominant phyla observed with 93.8 % (312, 739 sequences) and 99.3 % (79, 969 sequences) of reads respectively. Additional phyla were also observed but at a comparatively lower prevalence for the dry season. Sequences identified as belonging to the Nitrospirae, Synergistetes and Actinobacteria were also observed (1.5, 1.3 and 1.3 % respectively) to be more abundant in the wet season sample. The distribution of bacterial order across the Alpha, Beta and Gammaproteobacteria is shown in Figure 6.14. As expected the diversity observed for the wet season sample was greater than that of the dry. Examining the Betaproteobacteria, whilst the level of Burkholderiales is similar (42.9 % and 37 % for the dry and wet season respectively) there are notable differences in Nitrosomonadales and Neisseriales. Within the Gammaproteobacteria there were increases observed for both Methylococcales (4 %) and Pseudomonadales (35.4

%) for the dry season. In contrast the level of Enterobacteriales was 17 % for the wet season sample and just 4.6 % for the dry. Finally within the Alphaproteobacteria the wet season showed increased levels of Rhodospirillales and Rhodobacterales.

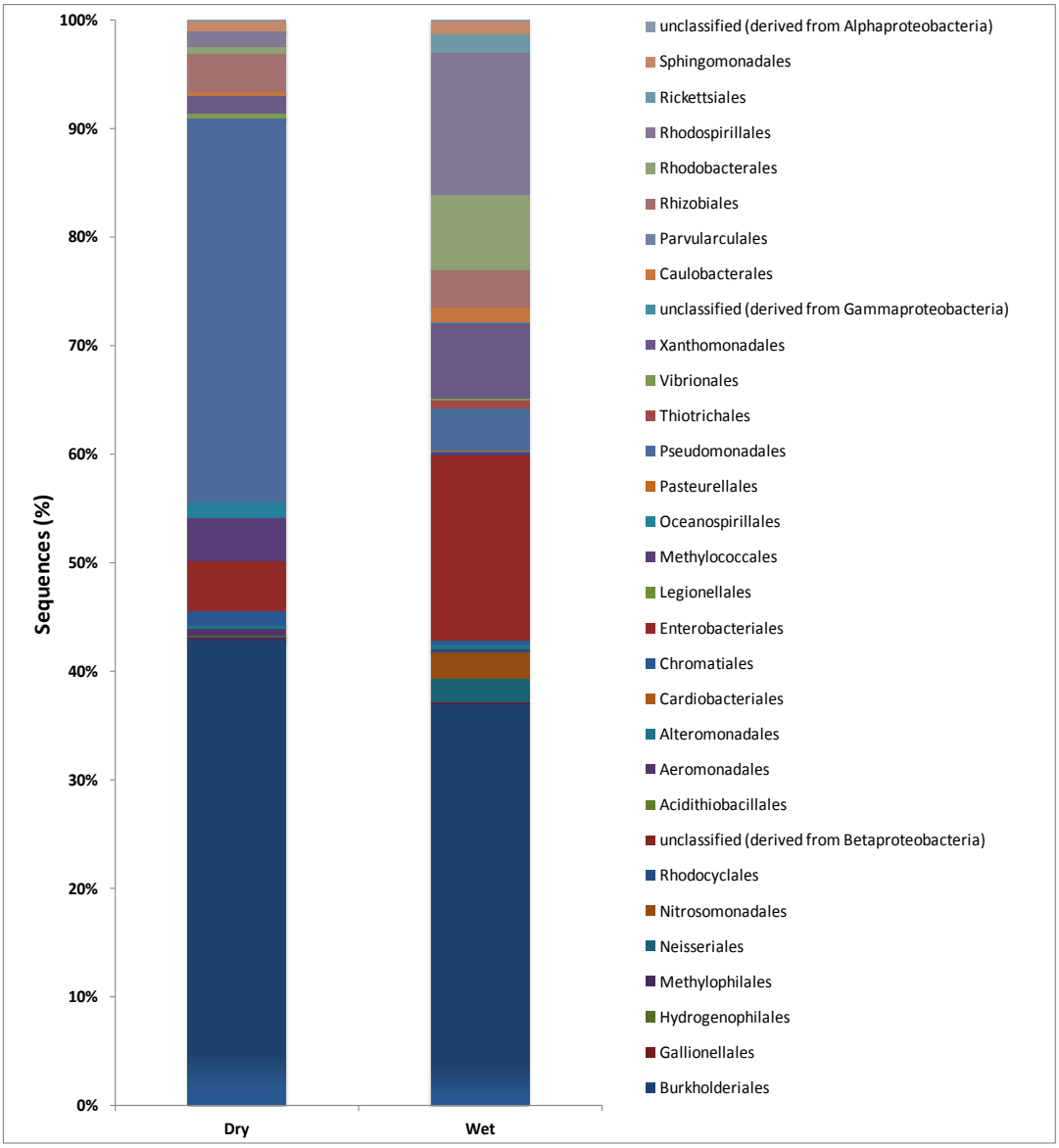

Figure 6.14: Stacked column graph showing comparative distribution of sequences by bacterial order, from the Beta, Gamma and Alphaproteobacteria from wet and dry season aerosol samples. Sequences were annotated against the M5NR database in MG-RAST (v. 3.1.2) using an *e*-value of -5 and a minimum alignment length of 50 bp.

6.3.4.6 Comparing Phyla and Functional Gene Content across Environmental Samples

To examine the relationship between environment and metagenome annotation, both wet and dry season samples were compared against a soil and indoor aerosol samples publically available on the MG-RAST server. The soil dataset (MG-RAST id: 4446153.3) was derived from a rainforest microbial community and consisted of 322 Mbp of sequence. The sampling location was described as a subtropical lower Montane forest in Puerto Rico. The indoor aerosol metagenome comprised two shotgun-sequenced, filter-collected aerosol samples (MG-RAST id: 4465946.3 and 4465825.3) sampled from a shopping centre in Singapore (Tringe *et al.*, 2008). The dataset comprised 84 Mbp and 87 Mbp for samples 1 and 2 respectively. Heatmap comparisons for all four samples were generated using and *e*-value of -5 and a minimum alignment length of 50 bp.

Comparing the samples at both the Phylum (Figure 6.15) and top level function (Figure 6.16) revealed a putative clustering of the wet season aerosol with that of the rainforest soil whilst the dry season clustered with the indoor metagenome or separately.

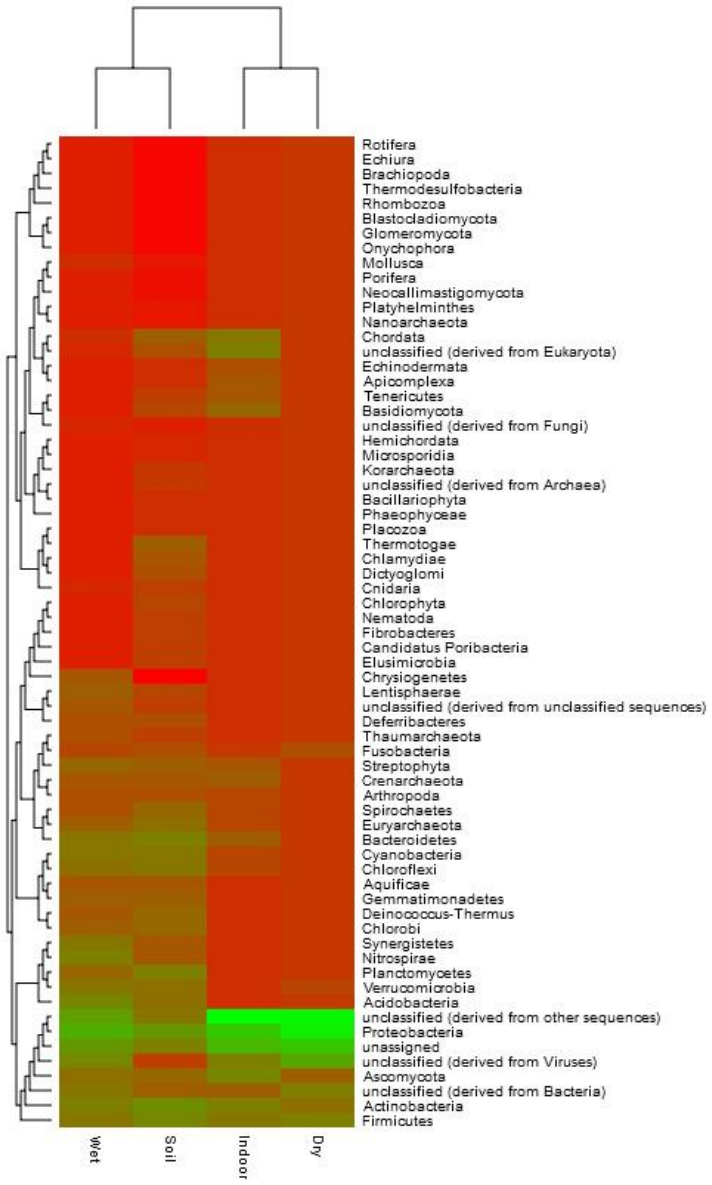

Figure 6.15: Heatmap comparison of Phylum distribution within rainforest soil, indoor, dry season and wet season shotgun sequenced metagenomes in MG-RAST. Annotations were made using the M5NR database with an *e*-value of -5 and a minimum alignment length of 50 bp.

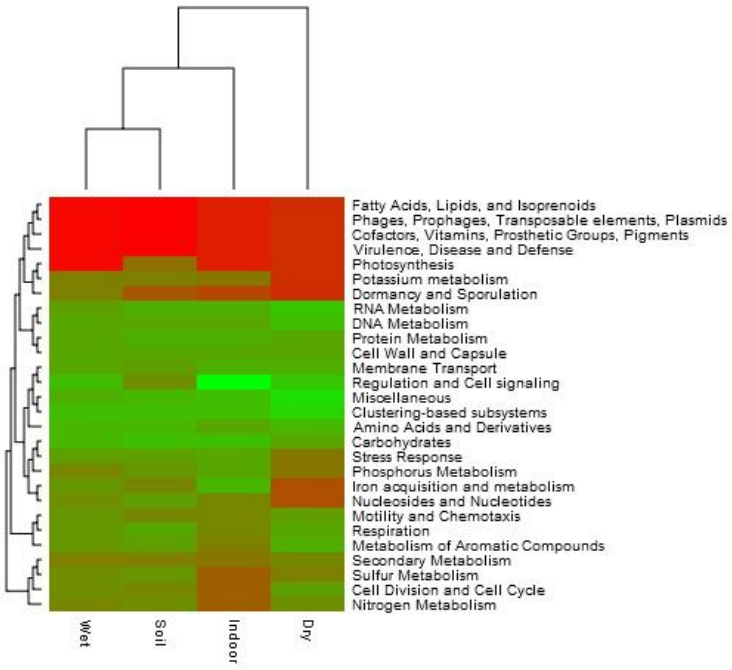

Figure 6.16: Heatmap showing comparison function distribution within rainforest soil, indoor, dry season and wet season aerosol metagenomes. Annotations were made using the M5NR database with an *e*-value of -5 and a minimum alignment length of 50 bp.

6.3.4.7 Identification of *Burkholderia* sp.

Wet and dry season metagenomic samples were compared for sequences from the genus *Burkholderia*. To ensure a greater degree of confidence associated with taxonomic classification, the data were re-analysed in MG-RAST (again using the M5NR database) using more stringent parameters (*e*-value -10, minimum alignment length 50 bp and 90 % sequence identity). Results are shown in Figure 6.17. For most *Burkholderia* sp. there was very little observed difference between each sample. In contrast, there was a clear increase in the number of sequences assigned to *B. pseudomallei* sequences for the wet season sample when compared to the dry. Here, 460 sequences were assigned to *B. pseudomallei* for the wet season from 840 that were annotated as *Burkholderia*

sp. In contrast only eight sequences were annotated as *B. pseudomallei* for the dry season sample.

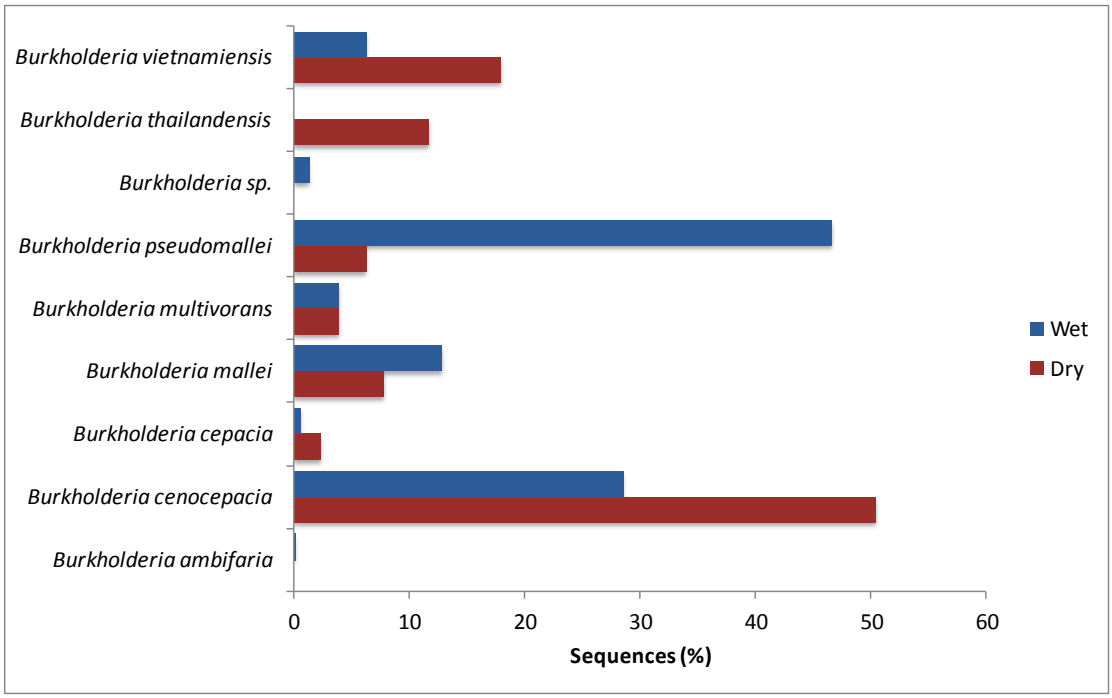

Figure 6.17: Distribution of sequences identified as belonging to the genus Burkholderia from shotgun sequenced wet and dry season aerosol samples. Sequences were analysed in MG-RAST v. 3.1.2. Annotations were generated using the M5NR database with an *e*-value of -10 for minimum alignments of 50 bp and a sequence identity of 90 %.

6.4 **Discussion**

The application of high-throughput sequencing to aerosol biogeography / bio-aerosol characterisation is a comparatively new endeavour. This is particularly true for the use of such techniques to determine the presence and/or prevalence of human pathogens. Data presented here represents the first characterisation of the aerosol metagenome from a region of known endemicity of *B. pseudomallei*. Furthermore, the combined use of both 16S rRNA amplicon sequencing and shot-gun metagenomics is shown to enable an in-depth interrogation of the microbial community with additional insights into the functional content of these samples.

Despite the use of use of two primer sets and multiple dry season samples only one (DDS10) yielded amplicons representing V1-3 of the 16S rRNA gene. Meteorological data provided by the Australian Government Bureau of Meteorology showed that this was the only day, throughout the 29 day sampling period, on which rainfall was recorded (1.2 mm). It can therefore be hypothesised that this sole observation of an airborne microbial community in the dry season is attributable to the aerosolisation of the surrounding soil microbiota. Across all diversity and richness estimates examined there is evidence to suggest a less diverse community within the dry season aerosol sample in comparison to the wet. Given the substantially greater levels of rainfall during the wet season this is not surprising and further supports a role of climatic conditions in the generation of local bioaerosols. However, it must be remembered that sampling of these communities was performed at an approximate height of 1.5 m. Consequently, this data may not reflect the total

airborne community as only the near-ground communities have been sampled. That aside, with an average of 14,000 sequences per sample, the diversity of these communities is low and likely to be a true reflection given the asymptotes present in the rarefaction curves. The number of observed OTUs was lower than expected (between 59 and 122) given the recent observations of Bowers *et al.*, (2011) who found an average of 93 OTUs using a much reduced sampling effort (500 sequences) from forest, agricultural and suburban aerosol samples. This is substantially less than Brodie *et al.*, (2007) where Chao1 indicated 1,500 16S rRNA phylotypes in an urban aerosol. Even here this was seen as an underestimate given the nonasymptotic rarefaction curves observed (Brodie *et al.*, 2007). It should be noted though that here a 99 % identity cut-off was applied and thus a direct comparison to the data presented herein is difficult. Through examinations of the structure of these communities, it is shown here that, in agreement with other studies (Bowers *et al.*, 2009; 2011), the most represented phyla is that of the Proteobacteria with the Burkholderiales as the predominant Order.

In order to determine the impact of highly abundant microbial taxa on the identification of rarer species, mixes of gDNA from two organisms (*V. cholerae* and *B. cenocepacia*) was used for metagenomic sequencing. This was important as it mirrored the approach to be taken for the detection of *B. pseudomallei* in the bio-aerosol metagenomes. The first important finding was that taxonomic assignment of these two organisms was robust for all the samples (mixtures of gDNA at different relative concentrations) using MG-RAST as the analysis pipeline. This was maintained even when *V. cholerae* gDNA was 1000-fold

greater in abundance. This may be a consequence of the use of permuted data-sets to generate thresholds (*e*-value, minimum alignment length and percentage identity) that produced more accurate assignments. Whether a similar approach is applicable across diverse environmental samples remains to be addressed.

From the poor yields of 16S rRNA amplification from dry season samples the necessity for non-specific amplification protocols to generate sufficient material for metagenomic analysis was clear. The use of whole genome amplification techniques to allow the analysis of precious or low quantity samples is well established (Lasken and Egholm, 2003). A number of studies have been conducted to determine the limitations of these techniques in terms of amplification bias and what is clear is that none are free from these artefacts. However, multiple displacement amplification using φ29 polymerase (used here) has been shown to generate amplified material with the lowest bias when used on single genomes, compared to alternative methods (Pinard *et al.*, 2006). However, care must still be exercised when interpreting the WGA data presented as it has previously been shown these issues of bias are found in analyses of more complex mixtures of bacterial genomic material (Abulencia *et al.*, 2006) and indeed aerosol samples (Ravva *et al.*, 2012). In the latter example, the authors found that diversity decreased following WGA as indicated by a reduction in Shannon-Weaver from 1.6 to 1.1 (a reduction in species richness and evenness) and Simpsons from 0.573 to 0.546 (Ravva *et al.*, 2012). The use of pooled samples representing a single dry and wet season sample was not ideal and indeed enough nucleic acid material was generated to undertake sequencing

on individual samples. Unfortunately, the scale of this endeavour was beyond the scope of this study.

Despite this there are trends in this data than can be compared to that of the 16S rRNA amplicon analysis. Firstly, the diversity of the wet season is greater than that of the dry. Secondly, the predominance of the Proteobacteria is preserved. Finally, the Burkholderiales are the most abundant Order. It should be noted though that a number of other Orders are represented in far greater numbers than previously identified. At this stage it cannot be determined if these are artefacts of bias or a true reflection of the sequences present. The heatmap analysis of the metagenomic samples revealed interesting similarities between the wet season aerosol and another, publically available, metagenome of a rainforest soil. This was observed both for functional and organism abundance analysis. Whilst only an initial analysis at this stage one can hypothesis that the similarities are linked to the presence of soil microbiota in the aerosol sample - a consequence of aerosolisation due to severe weather. The clustering of the dry and indoor metagenomes in terms of organism abundance is more likely a consequence of the lower diversity within both these samples in comparison to the wet and soil metagenomes.

The analysis of airborne metagenomes from a function perspective is still in its infancy with the only major study to date being that of an indoor environment (Tringe *et al.*, 2008). Here genes potentially involved in resistance to desiccation and oxidative stress were observed suggesting an adaptation to environmental pressures. This area is seen as one for potential expansion to give insights into

the metabolic activity and future biotechnological exploitation of these communities (Polymenakou, 2012). From a functional annotation perspective of samples analysed here, the most surprising result was the absence of sequences relating to dormancy and sporulation within the dry season aerosol. This may however simply be a consequence of the low diversity and absence of spore-forming organisms in the collected aerosols rather than an indication as to environmentally driven adaptations. To gain a greater insight, the application of metatransciptomic analyses to these samples should be considered. Here global gene expression analysis within samples would give an insight into the active organisms found and the metabolic functions therein. Perhaps the most eloquent demonstration of this technique is that of the analysis of complex marine microbial communities by Gilbert *et al.*, (2008).

A key aim of this study was to measure the prevalence of *B. pseudomallei* within the wet season aerosol samples. Whilst the use of 16S rRNA high-throughput sequencing would seem to allow the accurate identification of *B. pseudomallei* based on alignments of full-length and partial sequences, given the absence of real-time PCR detection in analyses of non-enriched samples (Blackman and Kennedy, 2008 unpublished) this was not pursued. However, what would be beneficial is further interrogation of these samples using a *Burkholderia* sp. genus-specific 16S rRNA PCR assay such as that developed by Salles *et al.*, (2002). In this instance a bespoke database could be implemented from Greengenes http://greengenes.lbl.gov/cgi-bin/nph-index.cgi containing only high-quality, fully annotated Burkholderiaceae sequences, thereby aiding taxonomic classification. From the analysis of the metagenomic samples there

does appear to be a greater number of sequences in the wet season sample, relative to the dry, that are classified as belonging to *B. pseudomallei*. Although real-time PCR analysis of the individual samples that had undergone WGA did not yield positive amplification (data not shown). At the time of writing a technical error with MG-RAST prevented the identification of where sequences from the metagenomic analysis could be mapped on a reference genome. This is an essential additional endeavour as there are a number of sequences that are annotated as *B. thailandensis*, which whilst not surprising given the high level of genomic homology between this organism and *B. pseudomallei*, does suggest certain misclassifications, even under these stringent conditions. Although it should be remembered that the geographic range of *B. thailandensis* is expanding with recent instances concerning the isolation of *B. thailandensis*-like organisms in Australia (Gee *et al.*, 2008). An interesting question that remains to be addressed is why the abundance of *B. cenocepacia* is higher than *B. pseudomallei* in the dry season. This is particularly interesting as both species are known to inhabit similar environmental niches (e.g. soil and water) and possess the capacity for intracellular survival (Landers *et al.*, 2000; Inglis and Sagripanti, 2006). It is perhaps the known requirement for moisture content of at least 10 % for the survival of *B. pseudomallei* (Tong *et al.*, 1996) that leads to its fluctuating prevalence between wet and dry season in contrast to *B. cenocepacia*.

What should be remembered, for both the metagenomic and 16S rRNA amplicon sequencing analysis is that it has been shown that a rainfall level of >125 mm is the predictor of inhalational melioidosis (Currie and Jacups, 2003). As it is likely this is related to exposure levels then, within the samples analysed here

where rainfall only exceeded this level on one day, *B. pseudomallei* would almost certainly be in low abundance and thus represent a greater challenge for detection.

Data presented here represents the first examination of the aerosol microbial community from an area where *B. pseudomallei* infection through the inhalational route is well known and associated with severe weather conditions. Although gene content was determined, additional investigations into the functional capacity of these communities using metatranscriptomic analysis would be of benefit. Future sample collection strategies should also include those with rainfall levels >125 mm. This would provide the greatest opportunity to demonstrate detection of *B. pseudomallei* within bioaerosols using high-throughput sequencing technologies. Lastly, a number of other sites were sampled in the original study by Blackman and Kennedy (2008) (unpublished). These could be used to examine the fluctuations in microbial community as a function of climatic conditions such as prevailing winds and rainfall and correlated to geographic location taking into account local land use – a feature known to influence both microbial community structure (Bowers *et al.*, 2009) and occurrence of *B. pseudomallei* in soil (Kaestli *et al.*, 2009). This would allow the characterisation of these communities across both space and time.

Chapter 7

General Discussion

7.1 **Detection of Microbial Taxa in Complex Communities**

The revolution in the application of molecular biology developments from the past decade, principally in the field of high-throughput DNA sequencing technologies, has resulted in a paradigm shift in the study of microbial ecology. Previously unprecedented insights into both the structure and function of complex environmental communities are now achievable. The application of these approaches for determining the prevalence and interactions of pathogenic species, which may have been used in the context of biological warfare, within environmental communities is no less affected and was the overarching aim of this study. The specific aims were to understand the interactions of nucleic acids derived from bacteria (in particular *Burkholderia* sp.) in soil. This included the potential for intra-species gene exchange through transformation, an hypothesised but poorly characterised phenomenon within this genus and one reliant on the availability of extracellular DNA. High-throughput sequencing techniques for the purpose of discerning shifts in community structure were also evaluated. These techniques were then used to determine the impact of local climatic conditions on the microbial diversity of previously uncharacterised bio-aerosol communities where the prevalence of the potential biological warfare agent, *B. pseudomallei*, is known to fluctuate (Currie and Jacups, 2003).

An underpinning requirement for studies into microbial ecology is the availability of nucleic acids that can be extracted from the sample matrix and interrogated to yield information about those communities. Consideration must first be given to the method through which these nucleic acids are accessed.

Comparisons of direct lysis methods undertaken here revealed clear differences in the efficiency of contrasting methods when used with contrasting soil types (Chapter 3). As a result, a simple direct lysis method using the PowerSoil™ DNA extraction kit (Cambio Ltd, UK) was applied to the analysis of all soil samples for both single species detection and total microbial community profiling (Chapter 3 and 5). Although it was beyond the scope of this study, an obvious avenue for exploration, and one that has started to be addressed for the application of sequencing techniques, is that of combining multiple methods to give a composite DNA extract (Delmont *et al.*, 2011). This enables an increase in the coverage of members of a community and consequently a more robust phylogentic reconstruction.

The use of small-scale microcosms demonstrated that bacteria-derived nucleic acids could be detected in contrasting non-sterile soil types by real-time PCR for at least 64 days. This was reduced to 32 days where extracellular DNA was used as the inocula (Chapter 3). In the contexts being addressed here nucleic acids from target organisms would be present in forms representing both extracellular DNA and associated with cellular debris or intact cells. From these results it was apparent that contrasting soil mineralogy, where purified mineral components were shown to provide protection from nuclease degradation in isolation, impacted the rate of DNA decay. Data suggested that the adsorption of DNA to these colloids *in-situ* increased the persistence of DNA for prolonged periods of time in soil, a facet of DNA dynamics in soil not fully evaluated. However, extrapolation of these results to mesocosm or field studies should be undertaken with care. Soil represents a dynamic and complex environment that can be

exposed to a myriad of climatic and anthropogenic factors that may impact on microbial communities. It was not possible to accurately replicate all these variables at the microcosm scale and as such these data represent only an indication as to the potential impacts of clay mineralogy on DNA persistence.

The persistence of extracellular DNA is of particular importance in the context of horizontal gene transfer through transformation. It is likely that the transformation potential of inoculated DNA would be much reduced compared to the duration of persistence observed here as has been noted previously (Romanowski *et al.*, 1993). Here an attempt to determine the transformation frequencies for species of the *Burkholderia cepacia* complex in soil across the time series for persistence observed for inoculated genomic DNA was undertaken. Whilst not achieved directly, there was evidence that DNA was available for transformation (at least within the initial period after inoculation) and that members of the Bcc have the mechanisms for its uptake. From the perspective of the latter phenomenon putative transformations were observed *in vitro* for *B. multivorans* 13010 where plasmid DNA was shown to be taken up and resulted in an increased resistance to trimethoprim (Chapter 4). If confirmed this would represent the first demonstration of transformation that was independent of sequence homology. It is already known that some members of the Burkholderiaceae possess mechanisms for the uptake, integration and expression of free DNA (Thongdee *et al.*, 2008). This is unsurprising given the features of the genomic architecture that indicates an evolutionary history of many gene transfer events (Mahenthiralingham and Drevenik, 2007). As naturally occurring soil microorganisms, the potential for intra-species gene

exchange within soil environments is an obvious area of future research. This is particularly pertinent where opportunistic pathogens of the Bcc may be found proximally to highly pathogenic species such as *B. pseudomallei* and the potential for gene exchange has been hypothesised (Holmes *et al.*, 1998). Moreover, transformation as a process has been shown to occur within soil (Sikorski *et al.*, 1998; Nielsen et al., 2000). An important consideration is that transformation potential is rarely distributed evenly within genera, and this may well apply to the Burkholderia genus. Certainly not all the *B. pseudomallei* strains used by Thongdee *et al.*, (2008) were transformable. Additionally, within the phylogenetic structure of *Pseudomonas stutzeri*, which arguably mirrors that of the Bcc, there exists a three-fold variation in transformation frequency between genomovars (Lorenz and Sikorski, 2000).

The application of molecular methods to the analysis of complex communities follows from understanding the availability of nucleic acids within those environments and the potential for that community to change in response to external ecological drivers. Here, the choice of technique, and the analysis that is applied, is of paramount importance. This was clearly demonstrated with the comparisons made between 16S rRNA V1-3 and V4-6 high-throughput amplicon sequencing assays (Chapter 5). Here, assays where compared using soil samples taken from one geographic location but on which existed plots that had received different long-term application regimes of veterinary antibiotic. This allowed the approaches to be compared for their power in discerning subtle shifts in highly similar complex communities. Kim *et al.*, (2012) recently demonstrated the utility of this approach to discerning shifts in complex microbial communities in

response to antibiotic use. Here analysis of human intestinal microbiota that was exposed *in vitro* to enrofloxacin was undertaken using the V1-3 of 16S rRNA. The authors showed a reduction in the proportion of Bacteroidetes and Proteobacteria and an increase in Firmicutes with increasing concentrations of enrofloxacin (Kim *et al.*, 2012). Here, the ability of each assay to discern these shifts in soil microbial communities was markedly different. This was likely to be a result of one if not all of the following factors - firstly, the 16S rRNA gene is known to exhibit different rates of sequence evolution across the nine variable regions (1 – 9) (Yu and Morrison, 2004). Where assays are used that do not target the full-length of the 16S rRNA gene, it is important to apply the most appropriate similarity cut-off for the taxonomic assignment of sequences into OTUs. The outcome of not applying the correct stringency for sequence binning is demonstrated here where diversity within soil samples was markedly reduced when V4-6 was compared to V1-3 with a 0.04 % where a 0.02 % dissimilarity index would have been more appropriate. Secondly, there is the issue regarding the accurate classification of singleton sequences. The importance of distinguishing true singletons, representative of taxa in low abundance in the rare biosphere, from sequencing errors or artefacts cannot be underestimated as the latter can inflate the apparent diversity of a community (Kunin *et al.*, 2010). Whilst there exist bioinformatic approaches to identifying and removing these artefacts (called 'denoising') these are noted for the high computational requirements (Zinger *et al.*, 2011); an issue that was encountered here and ultimately prevented the application of this process to the soil 16S rRNA sequencing data. A consequence of this was an inability to distinguish the relative merits of each assay in characterising these communities and,

consequently, difficulty in determining the true microbial diversity of the communities under investigation. Despite these limitations, the data generated in this study clearly shows a shift in the microbial community structure in response to the application of the veterinary antibiotics tylosin, sulfamethazine, and chlortetracycline (Chapter 5). These included species that increased in abundance such as *Comamoonas* sp. and *Dechloromonas* sp. which are known to harbour mobile genetic elements that have been shown to carry antibiotic resistance determinants (Tralau *et al.*, 2001; Boucher *et al.*, 2007; Nemergut *et al.*, 2008). Additionally, the observed decrease in Cyanobacteria is likely to be a consequence of antibiotic treatment owing to the well characterised sensitivities of these organisms (van der Grinten *et al.*, 2010). What remains to be discerned is whether the former changes were a result of the mobilisation of the integrons that were observed to increase in prevalence (Gaze *et al.*, unpublished data), a shift in the abundance of species / strains that already harboured the capacity to adapt to the presence of antibiotics or, what is more likely, a combined effect. The use of appropriate analysis methods to determine the statistical significance of changes in relative abundance of OTUs must also be considered, particularly where changes may be at the fine scale of community structure. Here the application of a conservative Bonferroni correction reduced the observed statistically significant trends dramatically.

Aside from the issues highlighted here, it must also be remembered that there are intrinsic limitations with regards to the use of the 16S rRNA gene. These include the presence of multiple heterologous copies within single cells as well as the issues regarding PCR amplification biases that continue to be revealed (Hong *et*

al., 2009). Provided these are understood and anticipated then methods can be implemented that limit their impact. One such strategy is the use of contrasting sequencing techniques that give a broader insight into communities. It has recently been demonstrated that use of *rpoB* pyrosequencing approach was complementary to 16S rRNA and can be used to target specific taxa of interest (Vos *et al.*, 2012). The approach used in Chapter 6 to investigate microbial communities within bio-aerosols used the 16S rRNA sequencing method evaluated in Chapter 5 and outputs of metagenomic approaches. In the latter approach the total gene content of a sample is analysed with the benefits of generating information related to both structure and functional capacity of a community. It must be remembered that only through metatranscriptomic approaches i.e. the analysis of mRNA species, can an insight into activity be discerned. This does however represent an area of bio-aerosol research with little prior effort and thus warrants a focussed level of investigation.

Using this combined sequencing method approach, a clear link between near-ground bio-aerosol diversity and the aerosolisation of microorganisms (most likely from soil) as a consequence of increased rainfall, was shown in this study (Chapter 6). The implications from a human health perspective are apparent as the shifts in diversity included an increase in the prevalence of the highly pathogenic species, *B. pseudomallei*. In addition, the comparison of aerosol communities with pre-existing metagenomic datasets for soil and indoor aerosols yielded interesting detail regarding the functional capacity of these communities. It was shown, for example, that the wet season aerosol sample was similar to that of the tropical rainforest soil. This was expected given the hypothesis

surrounding aerosolisation of soil microbes during heavy rainfall. What remains to be elucidated is what the active component of that functional capacity is in the airborne phase – a feature of bio-aerosols that could be investigated through the application of metatranscriptomics techniques. What has become clear, with regards to the interpretation of these types of data, is that the bioinformatic analysis criteria must account for the known biases and limitations that arise when any single method is used in isolation. The limitations of 16S rRNA amplicon sequencing have already been discussed. For metagenomic analysis interpretation must also be done with the same degree of scrutiny. Here defined metrics (*e*-value, minimum alignment length and similarity criteria) were generated and shown to be compatible with the datasets prior to use in the analysis of simple microbial communities. Here taxonomic assignments, using MG-RAST as the analysis pipeline, for mixtures of gDNA at different relative concentrations from two organisms (*V. cholerae* and *B. cenocepacia*) was shown to be robust even when *V. cholerae* gDNA was 1000-fold greater in abundance. Despite this, a more detailed investigation is warranted as to the validity of this approach for identifying pathogenic species within complex communities. It should be remembered that studies of bio-aerosol microbial communities such as these are still in their infancy. As such there remain a number of gaps within our knowledge that are obvious areas for future research. One such issue that remains to be addressed is the viability and potential activity of these communities within the airborne phase. These are important considerations when discussing the prevalence of pathogenic organisms and the implications of their presence to health. The difference in relative abundances within seasons that were observed for *B. cenocepacia* and *B. pseudomallei* also raises important

questions about microbial adaptation to climatic conditions. One might hypothesise that all *Burkholderia* sp., given the similarities in the environmental distribution of these organisms, would decrease to the same degree. This however does not appear to be the case and is suggestive of contrasting adaptations to these conditions. For the samples analysed here an additional important question that remains is the degree of contrast between aerosols collected at 1.5 m and those that exist at different altitudes. This is particularly pertinent where diversity in the dry season was low and it should be determined if this relates to a general reduction in microbial abundance during these climatic conditions or just a near-surface effect.

Our understanding of microbial ecology is rapidly advancing. This is in part due to the technological and computational advances that have enabled researchers to delve far deeper into the structure and function of complex communities than has previously been possible. With initiatives such as the earth microbiome project (www.earthmicrobiome.org), a multidisciplinary project designed to analyse the global microbial diversity, and the development of bioinformatic methods to attribute the proportions of communities to potential source environments (Knight *et al.*, 2011) studies such as those presented here can only benefit from our increased understanding of microbial communities, the ubiquity of taxa, temporal / spatial dynamics and the appropriateness of analysis methods.

Bibliography

Abulencia, C. B., Wyborski, D. L., Garcia, J. A. Podar, M., Chen, W., Chang, S. H., Chang, H. W., Watson, D., Brodie, E. L., Hazen, T. C., Keller, M. (2006)**.** Environmental Whole-Genome Amplification To Access Microbial Populations in Contaminated Sediments. *Appl Environ Microbiol* **72,** 3291-3301.

Allentoft, M. E., Collins, M., Harker, D., Haile, J., Oskam, C. L., Hale., M. L., Campos, P. F., Samaniego, J. E., Gilbert, M. T. P., Willerslev, E., Zhang, G., Scofield, R. P., Holdaway, R. N. & Bunce, M. (2012). The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. R. Soc. B.* **279,** 4724-4733

Alvarez, A. J., Khanna, M., Toranzos, G. A. & Stotzky, G. (1998)**.** Amplification of DNA bound on Clay Minerals. *Molecular Ecology* **7,** 775-778.

Andrews, R. E. J., Johnson, W. S., Guard, A. R. & Marvin, J. D. (2004)**.** Survival of enterococci and Tn916-like conjugative transposons in soil. *Canadian Journal of Microbiology* **50,** 957-966.

Arias, C. A. & Murray, B. E. (2009)**.** Antibiotic-Resistant Bugs in the 21st Century: A Clinical Super-Challenge. *N Engl J Med* **360,** 439-443.

Ashford D.A., Kaiser R.M., Bales M.E., Shutt K., Patrawalla A. & McShan A. (2003)**.** Planning against Biological Terrorism: Lesons from outbreak investigations. *Emerging Infectious Disease* **9,** 515-519.

Atkins, T., Prior, R., Mack, K., Russell, P., Prior, J., Ellis, J., Oyston, P. C. Dougan, G. (2002)**.** Characterisation of an acapsular mutant of *Burkholderia pseudomallei* identified by signature tagged mutagenesis. *J Med Microbiol* **51,** 539-553.

Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G. & Neufeld, J. D. (2011)**.** Generation of Multimillion-Sequence 16S rRNA Gene Libraries from Complex Microbial Communities by Assembling Paired-End Illumina Reads. *Appl Environ Microbiol* **77,** 3846-3852.

Bergmann, G. T., Bates, S. T., Eilers, K. G., Lauber, C. L., Caporaso, J. G., Walters, W. A., Knight, R. & Fierer, N. (2011)**.** The under-recognized

dominance of *Verrucomicrobia* in soil bacterial communities. *Soil Biology & Biochemistry* **43,** 1450-1455.

Bertolla, F. & Simonet, P. (1999)**.** Horizontal gene transfers in the environment: natural transformation as a putative process for gene transfers between transgenic plants and microorganisms. *Research in Microbiology* **150,** 375-384.

Bertolla, F., Van Gijsegem, F., Nesme, X. & Simonet, P. (1997)**.** Conditions for natural transformation of Ralstonia solanacearum. *Appl Environ Microbiol* **63,** 4965-4968.

Bevivino, A., Tabacchioni, S., Chiarini, L., Carusi, M. V., Del Gallo, M. & Visca, P. (1994)**.** Phenotypic comparison between rhizosphere and clinical isolates of Burkholderia cepacia. *Microbiology* **140,** 1069-1077.

Boucher, Y., Labbate, M., Koenig, J. E. & Stokes, H. W. (2007)**.** Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends in Microbiology* **15,** 301-309.

Bovallius, A., Bucht, B., Roffey, R. & Anas, P. (1978)**.** Three-year investigation of the natural airborne bacterial flora at four localities in sweden.. *Appl Environ Microbiol* **35,** 843-847.

Bowers, R. M., Lauber, C. L., Wiedinmyer, C., Hamady, M., Hallar, A. G., Fall, R., Knight, R. & Fierer, N. (2009)**.** Characterization of Airborne Microbial Communities at a High-Elevation Site and Their Potential To Act as Atmospheric Ice Nuclei. *Appl Environ Microbiol* **75,** 5121-5130.

Bowers, R. M., McLetchie, S., Knight, R. & Fierer, N. (2011)**.** Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments. *ISME J* **5,** 601-612.

Bowman, J. S., Rasmussen, S., Blom, N., Deming, J. W., Rysgaard, S. & Sicheritz-Ponten, T. (2011)**.** Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene. *ISME J.*

Brett, P. J., DeShazer, D. & Woods, D. E. (1998). *Burkholderia thailandensis* sp. nov., a *Burkholderia pseudomallei*-like species. *Int J Syst Bacteriol* **48,** 317-320.

Brim, H., Dijkmans, R. & Mergeay, M. (1994)**.** Stability of plasmid DNA of Escherichia coli C600 and Alcaligenes eutrophus CH34 inoculated in desiccating soil. *FEMS Microbiology Ecology* **15,** 169-175.

Brodie, E. L., DeSantis, T. Z., Parker, J. P. M., Zubietta, I. X., Piceno, Y. M. & Andersen, G. L. (2007)**.** Urban aerosols harbor diverse and dynamic bacterial populations. *Proceedings of the National Academy of Sciences* **104,** 299-304.

Burrows, S. M., Elbert, W., Lawrence, M. G. & Pöschl, U. (2009)**. Atmospheric bacteria in different ecosystems**. *Atmos Chem Phys Discuss* **9,** 10777-10827.

Cai, P., Huang, Q., Zhang, X. & Chen, H. (2006)**.** Adsorption of DNA on clay minerals and various colloidal particles from an Alfisol. *Soil Biology and Biochemistry* **38,** 471-476.

Cai, P., Huang, Q., Lu, Y. D., Chen, W. l., Jiang, D. H. & Liang, W. (2007)**.** Amplification of plasmid DNA bound on soil colloidal particles and clay minerals by the polymerase chain reaction. *Journal of Environmental Sciences* **19,** 1326-1329.

Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L. & Knight, R. (2010)**.** PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26,** 266-267.

Carter, G. B. (1992)**.** *Porton Down: 75 years of chemical and biological research* HMSO.

Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland, D. (2007)**.** A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods* **69,** 330-339.

Chaowagul, W. (2000)**.** Recent advances in the treatment of severe melioidosis. *Acta Tropica* **74,** 133-137.

Charlson, E. S., Chen, J., Custers-Allen, R., Bittinger, K., Li, H., Sinha, R., Hwang, J., Bushman, F. D. & Collman, R. G. (2010)**.** Disordered Microbial Communities in the Upper Respiratory Tract of Cigarette Smokers. *PLoS ONE* **5,** e15216.

Cheng, A. C. & Currie, B. J. (2005)**.** Melioidosis: Epidemiology, Pathophysiology, and Management. *Clinical Microbiology Reviews* **18,** 383-416.

Cheng, A. C., Ward, L., Godoy, D., Norton, R., Mayo, M., Gal, D., Spratt, B. G. & Currie, B. J. (2008)**.** Genetic Diversity of *Burkholderia pseudomallei* Isolates in Australia. *J Clin Microbiol* **46,** 249-254.

Chua, K. L., Chan, Y. Y. & Gan, Y. H. (2003)**.** Flagella Are Virulence Determinants of Burkholderia pseudomallei. *Infect Immun* **71,** 1622-1629.

Claesson, M. J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J. R., Ross, R. P. & O'Toole, P. W. (2010)**.** Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucl Acids Res*.

Claverys, J. P. & Martin, B. (2003)**.** Bacterial 'competence' genes: signatures of active transformation, or only remnants? *Trends in Microbiology* **11,** 161-165.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., Marsh, T., Garrity, G. M. & Tiedje, J. M. (2009)**.** The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucl Acids Res* **37,** D141-D145.

Crecchio, c. & Stotzky, G. (1998)**.** Binding of DNA on Humic Acids: Effect on Transformation of *Bacillus subtilis* and Resistance to DNase. *Soil Biology and Biochemistry* **30,** 1061-1067.

Cubillos-Ruiz, A., Junca, H., Baena, S., Venegas, I. & Zambrano, M. M. (2010)**.** Beyond Metagenomics: Integration of Complementary Approaches for the Study of Microbial Communities. In: *Metagenomics: Theory, Methods and Applications*, p. 15. Edited by D. Marco Caister Academic Press.

Currie, B. J. & Jacups, S. P. (2003). Intensity of Rainfall and Severity of Melioidosis, Australia. *Emerging Infectious Diseases* **9**.

Currie, B. J., Fisher, D. A., Howard, D. M., Burrow, J. N. C., Selvanayagam, S., Snelling, P. L., Anstey, N. M. & Mayo, M. J. (2000). The epidemiology of melioidosis in Australia and Papua New Guinea. *Acta Tropica* **74,** 121-127.

Curtis, T. P., Sloan, W. T. & Scannell, J. W. (2002). Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 10494-10499.

da, C. J., Susilawati, E., Smith, S., Wang, Q., Chai, B., Farris, R., Rodrigues, J., Thelen, K. & Tiedje, J. (2010)**.** Bacterial Communities in the Rhizosphere of Biofuel Crops Grown on Marginal Lands as Evaluated by 16S rRNA Gene Pyrosequences. *BioEnergy Research* **3,** 20-27.

Dahllof, I., Baillie, H. & Kjelleberg, S. (2000). rpoB-Based Microbial Community Analysis Avoids Limitations Inherent in 16S rRNA Gene Intraspecies Heterogeneity. *Appl Environ Microbiol* **66,** 3376-3380.

Dance, D. A. B. (2000)**.** Melioidosis as an emerging global problem. *Acta Tropica* **74,** 115-119.

Deiman, B., van Aarle, P., & Silekens, P. (2002). Characteristics and applications of nucleic acid sequence-based amplification (NASBA). *Mol Biotech* 20, 163-179

Delmont, T. O., Robe, P., Cecillon, S., Clark, I. M., Constancias, F., Simonet, P., Hirsch, P. R. & Vogel, T. M. (2011). Accessing the Soil Metagenome for Studies of Microbial Diversity. *Appl Environ Microbiol* **77,** 1315-1324.

Demaneche, S., Kay, E., Gourbiere, F. & Simonet, P. (2001)**.** Natural transformation of Pseudomonas fluorescens and Agrobacterium tumefaciens in soil. *Appl Environ Microbiol* **67,** 2617-2621.

Deng, Y., He, Z., Xu, M. & other authors (2012). Elevated Carbon Dioxide Alters the Structure of Soil Microbial Communities. *Appl Environ Microbiol* **78,** 2991-2995.

Derbise, A., Lesic, B., Dacheux, D., Ghigo, J. M. & Carniel, E. (2003). A rapid and simple method for inactivating chromosomal genes in Yersinia. *Fems Immunology and Medical Microbiology* **38,** 113-116.

DeSalle, R., Gatesy, J., Wheeler, W. & Grimaldi, D. (1992)**.** DNA sequences from a fossil termite in oligo-miocene amber and their phylogenetic implications. *Science* **257,** 1933-1936

DeShazer, D., Brett, P. J., Carlyon, R. & Woods, D. E. (1997). Mutagenesis of *Burkholderia pseudomallei* with Tn5-OT182: isolation of motility mutants and molecular characterization of the flagellin structural gene. *J Bacteriol* **179,** 2116-2125.

DeShazer, D., Brett, P. J. & Woods, D. E. (1998). The type II O-antigenic polysaccharide moiety of *Burkholderia pseudomallei* lipopolysaccharide is required for serum resistance and virulence. *Molecular Microbiology* **30,** 1081-1100.

DeShazer, D., Brett, P. J., Burtnick, M. N. & Woods, D. E. (1999). Molecular Characterization of Genetic Loci Required for Secretion of Exoproducts in *Burkholderia pseudomallei*. *J Bacteriol* **181,** 4661-4664.

DeShazer, D., Waag, D. M., Fritz, D. L. & Woods, D. E. (2001). Identification of a *Burkholderia mallei* polysaccharide gene cluster by subtractive hybridization and demonstration that the encoded capsule is an essential virulence determinant. *Microbial Pathogenesis* **30,** 253-269.

Després, V. R., Nowoisky, J. F., Klose, M., Conrad, R., Andreae, M. O. & P+Âschl, U. (2007). Characterization of primary biogenic aerosol particles in urban, rural, and high-alpine air by DNA sequence and restriction fragment analysis of ribosomal RNA genes. *Biogeosciences* **4,** 1127-1141.

Dillard, J. P. & Seifert, H. S. (2001). A variable genetic island specific for Neisseria gonorrhoeae is involved in providing DNA for natural transformation and is found more often in disseminated infection isolates. *Molecular Microbiology* **41,** 263-277.

Dineen, S. M., Aranda, R., Anders, D. L. & Robertson, J. M. (2010). An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. *Journal of Applied Microbiology* **109,** 1886-1896.

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

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science* **284,** 2124-2129.

Dowd, S. E., Sun, Y., Secor, P. R., Rhoads, D. D., Wolcott, B. M., James, G. A. & Wolcott, R. D. (2008). Survey of bacterial diversity in chronic wounds

using Pyrosequencing, DGGE, and full ribosome shotgun sequencing. *BMC Microbiology* **8**.

Droge, M., Puhler, A. & Selbitschka, W. (1998). Horizontal gene transfer as a biosafety issue: A natural phenomenon of public concern. *Journal of Biotechnology* **64,** 75-90.

Eede, G. van den., Aarts, H., Buhk, H-J., Corthier, G., Flint, H. J. Hammes, W., Jacobsen, B., Midtvedt, T., Vosse, J. van der,. & Wright, A., von. (2004). The relevance of gene transfer to the safety of food and feed derived from genetically modified (GM) plants. *Food and Chemical Toxicology* **42,** 1127-1156.

Eisen, R. J., Petersen, J. M., Higgins, C. L., Wong, D., Levy, C. E., Mead, P. S., Schriefer, M. E., Griffith, K. S., Gage, K. L., & Bear, C. B. (2008). Persistence of *Yersinia pestis* in Soil Under Natural Conditions. *Emerging Infectious Disease* **14**.

England, L. S., Vincent, M. L., Trevors, J. T. & Holmes, S. B. (2004). Extraction, detection and persistence of extracellular DNA in forest litter microcosms. *Molecular and Cellular Probes* **18,** 313-319.

Essex-Lopresti, A. E., Boddey, J. A., Thomas, R. & other authors (2005). A Type IV Pilin, PilA, Contributes to Adherence of *Burkholderia pseudomallei* and Virulence In Vivo. *Infect Immun* **73,** 1260-1264.

Fang, Z., Ouyang, Z., Zheng, H., Wang, X. & Hu, L. (2007). Culturable airborne bacteria in outdoor environments in Beijing, China. *Microbial Ecology* **54,** 487-496.

Fehlner-Gardiner, C. C. & Valvano, M. A. (2002). Cloning and characterization of the Burkholderia vietnamiensis norM gene encoding a multi-drug efflux protein. *FEMS Microbiology Letters* **215,** 279-283.

Fierer, N., Hamady, M., Lauber, C. L. & Knight, R. (2008). The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *PNAS* **105,** 17994-17999.

Fitzpatrick, K. A., Kersh, G. J. & Massung, R. F. (2010). Practical Method for Extraction of PCR-Quality DNA from Environmental Soil Samples. *Appl Environ Microbiol* **76,** 4571-4573.

Flavell, R. B., Dart, E., Fuchs, R. L. & Fraley, R. T. (1992). Selectable Marker Genes: Safe For Plants? *Nat Biotech* **10,** 141-144.

Friedlander, A. M. (1975). DNA release as a direct measure of microbial killing. I. Serum bactericidal activity. *Journal of Immunology* **115,** 1404-1408.

Fulton, J. D. (1966). Microorganisms of the Upper Atmosphere. *Applied Microbiology* **14,** 237-240.

Fushan, A., Monastyrskaya, G., Abaev, I., Kostina, M., Filyukova, O., Pecherskih, E. & Sverdlov, E. (2005). Genome-wide identification and mapping of variable sequences in the genomes of *Burkholderia mallei* and *Burkholderia pseudomallei*. *Research in Microbiology* **156,** 278-288.

Gabor, E., Liebeton, K., Niehaus, F., Eck, J. & Lorenz, P. (2007). Updating the metagenomics toolbox. *Biotechnology Journal* **2,** 201-206.

Gallori, E., Bazzicalupo, M., Dal Canto, L., Fani, R., Nannipieri, P., Vettori, C. & Stotzky, G. (1994). Transformation of *Bacillus subtilis* by DNA bound on clay in non-sterile soil. *FEMS Microbiology Ecology* **15,** 119-126.

Gans, J., Wolinsky, M. & Dunbar, J. (2005). Computational Improvements Reveal Great Bacterial Diversity and High Metal Toxicity in Soil. *Science* **309,** 1387-1390.

Garcia-Vallve, S., Guzman, E., Montero, M. A. & Romeu, A. (2003). HGT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. *Nucl Acids Res* **31,** 187-189.

Gaze, W. H., Zhang, L., Abdouslam, N. A., Hawkey, P. M., Calvo-Bado, L., Royle, J., Brown, H., Davis, S., Kay, P., Boxall, A. M., & Wellington, E. M. H. (2011). Impacts of anthropogenic activity on the ecology of class 1 integrons and integron-associated genes in the environment. *ISMEJ* **1-9**

Gebhard, F. & Smalla, K. (1999). Monitoring field releases of genetically modified sugar beets for persistence of transgenic plant DNA and horizontal gene transfer. *FEMS Microbiology Ecology* **28,** 261-272.

Gee, J. E., Glas, M. B., Novak, R. T., Gal, D., mayo, M. J., Steigerwalt, A. G., Wilkins, P. P. & Currie, B. J. (2008). Recovery of a *Burkholderia thailandensis*-like isolate from an Australian water source. *BMC Microbiology* **8**.

Gilbert, J. A., Field, D., Huang, Y., Edwards, R., Li, W., Gilna, P. & Joint, I. (2008). Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities. *PLoS ONE* **3,** e3042.

Gingues, S., Kooi, C., Visser, M. B., Subsin, B. & Sokol, P. A. (2005). Distribution and Expression of the ZmpA Metalloprotease in the Burkholderia cepacia Complex. *J Bacteriol* **187,** 8247-8255.

Graupner, S. & Wackernagel, W. (2001). Identification and characterization of novel competence genes comA and exbB involved in natural genetic transformation of *Pseudomonas stutzeri*. *Research in Microbiology* **152,** 451-460.

Graupner, S., Weger, N., Sohni, M. & Wackernagel, W. (2001). Requirement of novel competence genes pilT and pilU of *Pseudomonas stutzeri* for natural transformation and suppression of pilT deficiency by a hexahistidine tag on the type IV pilus protein PilAI. *Journal of Bacteriology* **183,** 4694-4701.

Gray, T. R. G. (1975). Survival of vegetative microbes in soil. *Symp Soc Gen Microbiol* **26,** 327-364.

Gregory, P. H. (1971). The Leeuwenhoek Lecture, 1970: Airborne Microbes: Their Significance and Distribution. *Proceedings of the Royal Society of London Series B Biological Sciences* **177,** 469-483.

Großkopf, R., Janssen, P. H. & Liesack, W. (1998). Diversity and Structure of the Methanogenic Community in Anoxic Rice Paddy Soil Microcosms as Examined by Cultivation and Direct 16S rRNA Gene Sequence Retrieval. *Appl Environ Microbiol* **64,** 960-969.

Gulledge, J. S., Luna, V. A., Luna, A. J., Zartman, R. & Cannons, A. C. (2010). Detection of low numbers of *Bacillus anthracis* spores in three soils using five commercial DNA extraction methods with and without an enrichment step. *Journal of Applied Microbiology* **109,** 1509-1520.

Hall, N. (2007). Advanced sequencing technologies and their wider impact in microbiology. *J Exp Biol* **210,** 1518-1525.

Halling-Sörensen, B., Jacobsen, A. M., Jensen, J., Sengeløv, G., Vaclavik, E. & Ingerslev, F. (2005). Dissipation and effects of chlortetracycline and tylosin

in two agricultural soils: A field-scale study in southern Denmark. *Environmental Toxicology and Chemistry* **24,** 802-810.

Hamilton, H. L., Dominguez, N. M., Schwartz, K. J., Hackett, K. T. & Dillard, J. P. (2005). *Neisseria gonorrhoeae* secretes chromosomal DNA via a novel type IV secretion system. *Molecular Microbiology* **55,** 1704-1721.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5,** R245-R249.

Herdina, Neate, S., Jabaji-Hare, S. & Ophel-Keller, K. (2004). Persistence of DNA of *Gaeumannomyces graminis* var. tritici in soil as measured by a DNA-based assay. *FEMS Microbiology Ecology* **47,** 143-152.

Heuer, H., Focks, A., Lamshöft, M., Smalla, K., Matthies, M. & Spiteller, M. (2008). Fate of sulfadiazine administered to pigs and its quantitative effect on the dynamics of bacterial resistance genes in manure and manured soil. *Soil Biology and Biochemistry* **40,** 1892-1900.

Heuer, H., Schmitt, H. & Smalla, K. (2011). Antibiotic resistance gene spread due to manure application on agricultural fields. *Current Opinion in Microbiology* **14,** 236-243.

Ho, W. C. & Ko, W. H. (1985). Soil microbiostasis: Effects of environmental and edaphic factors. *Soil Biology and Biochemistry* **17,** 167-170.

Hoffmaster A. R, Fitzgerald CC, Ribot E, Mayer LW, Popovic T. (2002) Molecular subtyping of *Bacillus anthracis* and the 2001 bioterrorism-associated anthrax outbreak, United States. *Emerg Infect Dis*. **8,** 1111-6.

Hoffmaster, A. R., Ravel, J., Rasko, D. A., et al. (2004). Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 8449-8454.

Holben, W. E., Feris, K. P., Kettunen, A. & Apajalahti, J. H. A. (2004). GC Fractionation Enhances Microbial Community Diversity Assessment and Detection of Minority Populations of Bacteria by Denaturing Gradient Gel Electrophoresis. *Appl Environ Microbiol* **70,** 2263-2270.

Holden, M. T. G., Seth-Smith, H. M. B., Crossman, L. C. & other authors (2009). The Genome of *Burkholderia cenocepacia* J2315, an Epidemic Pathogen of Cystic Fibrosis Patients. *Journal of Bacteriology* **191,** 261-277.

Holden, M. T. G., Titball, R. W., Peacock, S. J. & other authors (2004). Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 14240-14245.

Holmes, A., Govan, J. & Goldstein, R. (1998)**.** Agricultural use of *Burkholderia* (*Pseudomonas*) *cepacia*: a threat to human health? *Emerging Infectious Disease* **4,** 221-227.

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

Inglis, T. J. J. & Sagripanti, J. L. (2006). Environmental Factors That Affect the Survival and Persistence of *Burkholderia pseudomallei. Appl Environ Microbiol* **72,** 6865-6875.

Iwaki, M. & Arakawa, Y. (2006). Transformation of *Acinetobacter* sp BD413 with DNA from commercially available genetically modified potato and papaya. *Letters in Applied Microbiology* **43,** 215-221.

Johnsborg, O., Eldholm, V. & Hσvarstein, L. S. (2007). Natural genetic transformation: prevalence, mechanisms and function. *Research in Microbiology* **158,** 767-778.

Jordan, J. A., Butchko, A. R. & Durso, M. B. (2005). Use of Pyrosequencing of 16S rRNA Fragments to Differentiate between Bacteria Responsible for Neonatal Sepsis. *Journal of Molecular Diagnostics* **7,** 105-110.

Kaestli, M., Mayo, M., Harrington, G., Ward, L., Watt, F., Hill, J. V., Cheng, A. C. & Currie, B. J. (2009). Landscape Changes Influence the Occurrence of the Melioidosis Bacterium *Burkholderia pseudomallei*in Soil in Northern Australia. *PLoS Negl Trop Dis* **3,** e364.

Kang, Y., Norris, M. H., Wilcox, B. A., Tuanyok, A., Keim, P. S. & Hoang, T. T. (2011). Knockout and pullout recombineering for naturally transformable *Burkholderia thailandensis* and *Burkholderia pseudomallei*. *Nature Protocols* **6,** 1085-1104.

Khanna, M. & Stozky, G. (1992). Transformation of *Bacillus subtilis* by DNA Bound on Montmorillonite and Effect of DNase on the Transforming Ability of Bound DNA. *Appl Environ Microbiol* **58,** 1930-1939.

Kidd, T. J., Bell, S. C. & Coulter, C. (2003). Genomovar Diversity Amongst *Burkholderia cepacia* Complex Isolates From an Australian Adult Cystic Fibrosis Unit. *European Journal of Clinical Microbiology & Infectious Diseases* **22,** 434-437.

Kim, B. S., Kim, J. N., Yoon, S. H., Chun, J. & Cerniglia, C. E. (2012). Impact of enrofloxacin on the human intestinal microbiota revealed by comparative molecular analysis. *Anaerobe* **18** (3), 310-320.

Kim, M., Morrison, M. & Yu, Z. (2011). Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *Journal of Microbiological Methods* **84,** 81-87.

Kim, S. H., Schell, M. A., Ulrich, R. L., Sarria, S. H., Nierman, W. C. & DeShazer, D. (2005). Bacterial genome adaptation to niches: Divergence of the potential virulence genes in three *Burkholderia* species of different survival strategies. *BMC Genomics* **6**.

Knapp, C. W., Engemann, C. A., Hanson, M. L., Keen, P. L., Hall, K. J. & Graham, D. W. (2008). Indirect Evidence of Transposon-Mediated Selection of Antibiotic Resistance Genes in Aquatic Systems at Low-Level Oxytetracycline Exposures. *Environ Sci Technol* **42,** 5348-5353.

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., Bushman, F. D., Knight, R. & Kelley, S. T. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nature Methods* **8,** 761-763

Ko, W.-C., Cheung, B. M.-H., Tang, H.-J., Shih, H.-I., Lau, Y. J., Wang, L.-R. & Chuang, Y.-C. (2007). Melioidosis Outbreak after Typhoon, Southern Taiwan. *Emerging Infectious Diseases* **13**.

Kooi, C. & Sokol, P. A. (2009). *Burkholderia cenocepacia* zinc metalloproteases influence resistance to antimicrobial peptides. *Microbiology-Sgm* **155,** 2818-2825.

Kowalchuk, G., Speksnijder, A., Zhang, K., Goodman, R. & van Veen, J. (2007). Finding the Needles in the Metagenome Haystack. *Microbial Ecology* **53,** 475-485.

Kresk, M. & Wellington, E. M. H. (1999). Comparison of different methods for the isolation and purification of total community DNA from soil. *Journal of Microbiological Methods* **39,** 1-16.

Kunin, V., Engelbrektson, A., Ochman, H. & Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology* **12,** 118-123.

Landers, P., Kerr, K. G., Rowbotham, T. J., Tipper, J. L., Keig, P. M., Ingham, E. & Denton, M. (2000). Survival and growth of *Burkholderia cepacia* within the free-living amoeba *Acanthamoeba polyphaga*. *European Journal of Clinical Microbiology and Infectious Disease* **19,** 121-123.

Larsen, J. C. & Johnson, N. H. (2009). Pathogenesis of *Burkholderia pseudomallei* and *Burkholderia mallei*. *Military Medicine* **174**.

Lasken, R. S. & Egholm, M. (2003). Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. *Trends in Biotechnology* **21,** 531-535.

Lauber, C. L., Hamady, M., Knight, R. & Fierer, N. (2009). Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. *Appl Environ Microbiol* **75,** 5111-5120.

Lemos, L. N., Fulthorpe, R. R., Triplett, E. W. & Roesch, L. F. W. (2011). Rethinking microbial diversity analysis in the high throughput sequencing era. *Journal of Microbiological Methods* **86,** 42-51.

Lesic, B. & Rahme, L. G. (2008). Use of the lambda Red recombinase system to rapidly generate mutants in *Pseudomonas aeruginosa*. *BMC Molecular Biology* **9**.

Leski, T. A., Malanoski, A. P., Gregory, M. J., Lin, B. & Stenger, D. A. (2011). Application of broad-range resequencing array RPM-TEI for detection of pathogens in desert dust samples from Kuwait and Iraq. *Appl Environ Microbiol* **77** (13), 4285-4292.

Lessie, T. G., Hendrickson, W., Manning, B. D. & Devereux, R. (1996). Genomic complexity and plasticity of *Burkholderia cepacia*. *FEMS Microbiology Letters* **144,** 117-128.

Levy-Booth, D. J., Campbell, R. G., Gulden, R. H. *et al.*, (2007). Cycling of extracellular DNA in the soil environment. *Soil Biology and Biochemistry* **39,** 2977-2991.

Lighthart, B. & Shaffer, B. T. (1994). Bacterial flux from chaparral into the atmosphere in mid-summer at a high desert location. *Atmospheric Environment* **28,** 1267-1274.

Lighthart, B. (2000). Mini-review of the concentration variations found inthe alfresco atmospheric bacterial populations. *Aerobiologia* **16,** 7-16.

Lindemann, J. & Upper, C. D. (1985). Aerial Dispersal of Epiphytic Bacteria over Bean Plants. *Appl Environ Microbiol* **50,** 1229-1232.

Lipuma, J. J. (2001). *Burkholderia cepacia* complex: a contraindication to lung transplantation in cystic fibrosis? *Transplant Infectious Disease* **3,** 149-160.

LiPuma, J. J., Spilker, T., Coenye, T. & Gonzalez, C. F. (2002). An epidemic *Burkholderia cepacia* complex strain identified in soil. *The Lancet* **359,** 2002-2003.

Liu, L., Spilker, T., Coenye, T. & LiPuma, J. J. (2003). Identification by Subtractive Hybridization of a Novel Insertion Element Specific for Two Widespread *Burkholderia cepacia* Genomovar III Strains. *J Clin Microbiol* **41,** 2471-2476.

Liu, W. T., Marsh, T. L., Cheng, H. & Forney, L. J. (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* **63,** 4516-4522.

Liu, Z., Lozupone, C., Hamady, M., Bushman, F. D. & Knight, R. (2007). Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucl Acids Res* **35,** e120.

Lorenz, M. G. & Wackernagel, W. (1987). Adsorption of DNA to Sand and Variable Degradation Rates of Adsorbed DNA. *Appl Environ Microbiol* **53,** 2948-2952.

Lorenz, M. G. & Wackernagel, W. (1994). Bacterial Gene Transfer by Natural Genetic Transformation in the Environment. *Microbiological Reviews* **58,** 563-602.

Lorenz, M. G., Gerjets, D. & Wackernagel, W. (1991). Release of Transforming Plasmid and Chromosomal Dna from 2 Cultured Soil Bacteria. *Archives of Microbiology* **156,** 319-326.

Lorenz, M. G. & Wackernagel, W. (1990). Natural genetic transformation of *Pseudomonas stutzeri* by sand-adsorbed DNA. *Archives of Microbiology* **154,** 380-385.

Lorenz, M. G. & Sikorski, J. (2000). The potential for intraspecific horizontal gene exchange by natural genetic transformation: sexual isolation among genomovars of *Pseudomonas stutzeri*. *Microbiology* **146,** 3081-3090.

Losada, L., Ronning, C. M., DeShazer, D. *et al.*, (2010). Continuing Evolution of *Burkholderia mallei* Through Genome Reduction and Large-Scale Rearrangements. *Genome Biol Evol* **2010,** 102-116.

Lozupone, C. & Knight, R. (2005). UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Appl Environ Microbiol* **71,** 8228-8235.

Luna, R. A., Fasciano, L. R., Jones, S. C., Boyanton, B. L., Ton, T. T. & Versalovic, J. (2007). DNA Pyrosequencing-Based Bacterial Pathogen Identification in a Pediatric Hospital Setting. *J Clin Microbiol* **45,** 2985-2992.

Mack, K. & Titball, R. W. (1998). The detection of insertion sequences within the human pathogen *Burkholderia pseudomallei* which have been identified previously in *Burkholderia cepacia*. *FEMS Microbiology Letters* **162,** 69-74.

Mahenthiralingam, E., Baldwin, A. & Dowson, C. G. (2008). Burkholderia cepacia complex bacteria: opportunistic pathogens with important natural biology. *Journal of Applied Microbiology* **104,** 1539-1551.

Mahenthiralingam, E., Baldwin, A. & Vandamme, P. (2002). Burkholderia cepacia complex infection in patients with cystic fibrosis. *J Med Microbiol* **51,** 533-538.

Mahenthiralingam E., Drevenik P. (2007). Comparative genomics of *Burkholderia* species, in Burkholderia, eds Coeyne T., Vamdamme P., editors. (Norwich: Horizon Scientific Press 53–79.

Manchee, R. J., Broster, M. G., Melling, J., Henstridge, R. M. & Stagg, A. J. (1981). Bacillus anthracis on Gruinard Island. *Nature* **19,** 254-255.

Manter, D. K., Weir, T. L. & Vivanco, J. M. (2010). Negative Effects of Sample Pooling on PCR-Based Estimates of Soil Microbial Richness and Community Structure. *Appl Environ Microbiol* **76,** 2086-2090.

Mathis, L. S. & Scocca, J. J. (1982). Haemophilus influenzae and Neisseria gonorrhoeae Recognize Different Specificity Determinants in the DNA Uptake Step of Genetic Transformation. *Journal of General Microbiology* **128,** 1159-1161.

Matsui, K., Ishii, N. & Kawabata, Z. (2003). Release of extracellular transformable plasmid DNA from Escherichia coli cocultivated with algae. *Appl Environ Microbiol* **69,** 2399-2404.

McKevitt, A. I., Bajaksouzian, S., Klinger, J. D. & Woods, D. E. (1989). Purification and characterization of an extracellular protease from Pseudomonas cepacia. *Infect Immun* **57,** 771-778.

Meibom, K. L., Blokesch, M., Dolganov, N. A., Wu, C. Y. & Schoolnik, G. K. (2005). Chitin Induces Natural Competence in *Vibrio cholerae. Science* **310,** 1824-1827.

Meier, P., Berndt, C., Weger, N. & Wackernagel, W. (2002). Natural transformation of *Pseudomonas stutzeri* by single-stranded DNA requires type IV pili, competence state and comA. *FEMS Microbiology Letters* **207,** 75-80.

Mercier, A., Bertolla, F., Passelegue-Robe, E. & Simonet, P. (2007). Natural transformation-based foreign DNA acquisition in a *Ralstonia solanacearum* mutS mutant. *Research in Microbiology* **158,** 537-544.

Meyer, F., Paarmann, D., D'Souza, M. & other authors (2008). The Metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*. **9** (386).

Michel, B. (2005). After 30 Years of Study, the Bacterial SOS Response Still Surprises Us. *PLoS Biol* **3,** e255.

Milne, J. C., Furlong, D., Hanna, P. C., Wall, J. S. & Collier, R. J. (1994). Anthrax protective antigen forms oligomers during intoxication of mammalian cells. *J Biol Chem* **269,** 20607-20612.

Milne, J. C. & Collier, R. J. (1993). pH-dependent permeabilization of the plasma membrane of mammalian cells by anthrax protective antigen. *Molecular Microbiology* **10,** 647-653.

Morales, S. E., Cosart, T. F., Johnson, J. V. & Holben, W. E. (2009). Extensive Phylogenetic Analysis of a Soil Bacterial Community Illustrates Extreme Taxon Evenness and the Effects of Amplicon Length, Degree of Coverage, and DNA Fractionation on Classification and Ecological Parameters. *Appl Environ Microbiol* **75,** 668-675.

Muyzer, G., de Waal, E. C. & 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.

Nakamura, S., Maeda, N., Miron, I. M. & other authors (2008). Metagenomic Diagnosis of Bacterial Infections. *Emerging Infectious Diseases* **14,** 1784-1786.

Nemergut, D. R., Robeson, M. S., Kysela, R. F., Martin, F. P., Scmidt, S. K. & Knight, R. (2008). Insights and inferences about integron evolution from genomic data. *BMC Genomics* **9**.

Nielsen, K. M. (1998). Barriers to horizontal gene transfer by natural transformation in soil bacteria. *Apmis* **106,** 77-84.

Nielsen, K. M., Smalla, K. & van Elsas, J. D. (2000a). Natural transformation of *Acinetobacter* sp strain BD413 with cell lysates of *Acinetobacter* sp., *Pseudomonas fluorescens*, and *Burkholderia cepacia* in soil microcosms. *Appl Environ Microbiol* **66,** 206-212.

Nielsen, K. M., van Elsas, J. D. & Smalla, K. (2000b). Transformation of *Acinetobacter* sp. strain BD413(pFG4 Delta nptII) with transgenic plant DNA in soil microcosms and effects of kanamycin on selection of transformants. *Appl Environ Microbiol* **66,** 1237-1242.

Nielsen, K. M., vanWeerelt, M. D. M., Berg, T. N., Bones, A. M., Hagler, A. N. & vanElsas, J. D. (1997). Natural transformation and availability of transforming DNA to *Acinetobacter calcoaceticus* in soil microcosms. *Appl Environ Microbiol* **63,** 1945-1952.

Nielsen, K. M., Bones, A. M., Smalla, K. & van Elsas, J. D. (1998). Horizontal gene transfer from transgenic plants to terrestrial bacteria - a rare event? *FEMS Microbiology Reviews* **22,** 79-103.

Nielsen, K. M., Johnsen, P. J., Bensasson, D. & Daffonchio, D. (2007). Release and persistence of extracellular DNA in the environment. *Environmental Biosafety Research* **6,** 37-53.

Nielsen, K. M. & Townsend, J. P. (2004). Monitoring and modeling horizontal gene transfer. *Nat Biotech* **22,** 1110-1114.

Nielsen, K. M., van Elsas, J. D. & Smalla, K. (2000c). Transformation of *Acinetobacter* sp. Strain BD413(pFG4Delta nptII) with Transgenic Plant DNA in Soil Microcosms and Effects of Kanamycin on Selection of Transformants. *Appl Environ Microbiol* **66,** 1237-1242.

Nocker, A., Sossa-Fernandez, P., Burr, M. D., Camper, A. K. (2007). Use of propidium monoazide for live/dead distinction in microbial ecology. *Appl Environ Microbiol* **73.**

Nogva, H. K., Bergh, A., Holck, A. & Rudi, K. (2000). Application of the 5'-Nuclease PCR Assay in Evaluation and Development of Methods for Quantitative Detection of *Campylobacter jejuni*. *Appl Environ Microbiol* **66,** 4029-4036.

Nyrén, P. & Lundin, A. (1985). Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Analytical Biochemistry* **151,** 504-509.

Nzula, S., Vandamme, P. & Govan, J. R. W. (2002). Influence of taxonomic status on the in vitro antimicrobial susceptibility of the *Burkholderia cepacia* complex. *J Antimicrob Chemother* **50,** 265-269.

O'Sullivan, L. A., Weightman, A. J., Jones, T. H., Marchbank, A. M., Tiedje, J. M. & Mahenthiralingam, E. (2007). Identifying the genetic basis of ecologically and biotechnologically useful functions of the bacterium Burkholderia vietnamiensis. *Environmental Microbiology* **9,** 1017-1034.

Ogram, A., Sayler, G. S., Gustin, D. & Lewis, R. J. (1988). DNA adsorption to soils and sediments. *Environ Sci Technol* **22,** 982-984.

Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science* **276,** 734-740.

Paget, E., Lebrun, M., Freyssinet, G. & Simonet, P. (1998). The fate of recombinant plant DNA in soil. *European Journal of Soil Biology* **34,** 81-88.

Paget, E. & Simonet, P. (1994). On the track of natural transformation in soil. *FEMS Microbiology Ecology* **15,** 109-117.

Palacios, G., Druce, J., Du, L. & other authors (2008). A New Arenavirus in a Cluster of Fatal Transplant-Associated Diseases. *N Engl J Med* **358,** 991-998.

Palmen, P., Buijsman, P. & Hellingerf, K. J. (1994a). Physiological regulation of competence induction for natural transformation in *Acinetobacter calcoaceticus*. *Archives of Microbiology* **162,** 344-351.

Palmen, R., Buijsman, P. & Hellingwerf, K. J. (1994b). Physiological Regulation of Competence Induction for Natural Transformation in *Acinetobacter-Calcoaceticus*. *Archives of Microbiology* **162,** 344-351.

Palmen, R. & Hellingwerf, K. J. (1997). Uptake and processing of DNA by *Acinetobacter calcoaceticus* - A review. *Gene* **192,** 179-190.

Paul, E. A. & Clark, F. E. (1996)**.** *Soil Microbiology and Biochemistry*.

Payne, G. W., Vandamme, P., Morgan, S. H., LiPuma, J. J., Coenye, T., Weightman, A. J., Jones, T. H. & Mahenthiralingam, E. (2005). Development of a recA Gene-Based Identification Approach for the Entire Burkholderia Genus. *Applied and Environmental Microbiology* **71** 3917-27.

Pei, A. Y., Oberdorf, W. E., Nossa, C. W. *et al.*, (2010). Diversity of 16S rRNA Genes within Individual Prokaryotic Genomes. *Appl Environ Microbiol* **76,** 3886-3897.

Pietramellara, G., Dal Canto, L., Vettori, C., Gallori, E. & Nannipieri, P. (1997). Effects of air-drying and wetting cycles on the transforming ability of DNA bound on clay minerals. *Soil Biology and Biochemistry* **29,** 55-61.

Pinard, R., de Winter, A., Sarkis, G. J., Gerstein, M. B., Tartaro, K. R., Plant, R. N., Egholm, M., Rothberg, J. M. & Leamon, J. H. (2006). Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics* **7**.

Polymenakou, P. N. (2012). Atmosphere: A Source of Pathogenic or Beneficial Microbes? *Atmosphere* **3,** 87-102.

Pontiroli, A., Rizzi, A., Simonet, P., Daffonchio, D., Vogel, T. M. & Monier, J. M. (2009). Visual Evidence of Horizontal Gene Transfer between Plants and Bacteria in the Phytosphere of Transplastomic Tobacco. *Appl Environ Microbiol* **75,** 3314-3322.

Poretsky, R. S., Bano, N., Buchan, A., LeCleir, G., Kleikemper, J., Pickering, M., Pate, W. M., Moran, M. A. & Hollibaugh, J. T. (2005). Analysis of Microbial Gene Transcripts in Environmental Samples. *Appl Environ Microbiol* **71,** 4121-4126.

Price, L. B., Liu, C. M., Melendez, J. H. *et al.*, (2009). Community Analysis of Chronic Wound Bacteria Using 16S rRNA Gene-Based Pyrosequencing: Impact of Diabetes and Antibiotics on Chronic Wound Microbiota. *PLoS ONE* **4,** e6462.

Rainbow, L., Hart, C., & Winstanley, C. (2002). Distribution of type III secretion gene clusters in *Burkholderia pseudomallei*, *B. thailandensis* and *B. mallei*. *J Med Microbiol* **51,** 374-384.

Ramette, A., LiPuma, J. J. & Tiedje, J. M. (2005). Species Abundance and Diversity of Burkholderia cepacia Complex in the Environment. *Appl Environ Microbiol* **71,** 1193-1201.

Ravva, S. V., Hernlem, B. J., Sarreal, C. Z. & Mandrell, R. E. (2012). Bacterial communities in urban aerosols collected with wetted-wall cyclonic samplers and seasonal fluctuations of live and culturable airborne bacteria. *J Environ Monit* **14,** 473-481.

Ray, J. L. & Nielsen, K. M. (2005). Experimental Methods for Assaying Natural Transformation and Inferring Horizontal Gene Transfer. In: *Methods in Enzymology*, pp. 491-520 Academic Press.

Rebrikov, D. V., Desai, S. M., Siebert, P. D. & Lkyanov, S. A. (2004). Suppression subtractive hybridisation. *Methods Mol Biol* **258,** 107-134.

Reckseidler, S. L., DeShazer, D., Sokol, P. A. & Woods, D. E. (2001). Detection of Bacterial Virulence Genes by Subtractive Hybridization: Identification of Capsular Polysaccharide of *Burkholderia pseudomallei* as a Major Virulence Determinant. *Infect Immun* **69,** 34-44.

Reckseidler-Zenteno, S. L., DeVinney, R. & Woods, D. E. (2005). The Capsular Polysaccharide of *Burkholderia pseudomallei* Contributes to Survival in Serum by Reducing Complement Factor C3b Deposition. *Infect Immun* **73,** 1106-1115.

Reik, R., Spilker, T. & LiPuma, J. J. (2005). Distribution of *Burkholderia cepacia* Complex Species among Isolates Recovered from Persons with or without Cystic Fibrosis. *J Clin Microbiol* **43,** 2926-2928.

Ribot, W. J. & Ulrich, R. L. (2006). The Animal Pathogen-Like Type III Secretion System Is Required for the Intracellular Survival of *Burkholderia mallei* within J774.2 Macrophages. *Infect Immun* **74,** 4349-4353.

Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. (2005). REBASE-restriction enzymes and DNA methyltransferases. *Nucl Acids Res*.

Robertson, B. D. & Meyer, T. F. (1992). Genetic variation in pathogenic bacteria. *Trends in Genetics* **8,** 422-427.

Roesch, L. F. W., Fulthorpe, R. R., Riva, A. *et al.*, (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* **1,** 283-290.

Romanowski, G., Lorenz, M. G. & Wackernagel, W. (1993). Use of Polymerase Chain Reaction and Electroporation of *Escherichia coli* to Monitor the Persistence of Extra-cellular Plasmid DNA Introduced into Natural Soils. *Appl Environ Microbiol* **59,** 3438-3446.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. & Nyren, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* **242,** 84-89.

Ronaghi, M., Uhlen, M. & Nyren, P. (1998). A sequencing method based on real-time pyrophosphate. *Science* **281,** 363, 365.

Ronaghi, M. (2000). Improved Performance of Pyrosequencing Using Single-Stranded DNA-Binding Protein. *Analytical Biochemistry* **286,** 282-288.

Ronaghi, M. (2001). Pyrosequencing Sheds Light on DNA Sequencing. *Genome Res* **11,** 3-11.

Rotz, L. D., Khan, A. S., Lillibridge, S. R., Ostroff, S. M. & Hughes, J. M. (2002). Public Health Assessment of Potential Biological Terrorism Agents. *Emerging Infectious Disease* **8**.

Rudi, K. & Jakobsen, K. S. (2006). Overview of DNA purification for nucleic acid-based diagnostics from environmental and clinical samples. *Methods in Molecular Biology* **345,** 23-35.

Ruppitsch, W., Stöger, A., Indra, A., Grif, K., Scabereiter-Gurtner, C., Hirschl, A. & Allerberger, F. (2007). Suitability of partial 16S ribosomal RNA gene sequence analysis for the identification of dangerous bacterial pathogens. *Journal of Applied Microbiology* **102,** 852-859.

Saito, Y., Taguchi, H. & Akamatsu, T. (2006). Fate of transforming bacterial genome following incorporation into competent cells of *Bacillus subtilis*: a continuous length of incorporated DNA. *Journal of Bioscience and Bioengineering* **101,** 257-262.

Salles, J. F., De Souza, F. A. & van Elsas, J. D. (2002). Molecular Method To Assess the Diversity of *Burkholderia* Species in Environmental Samples. *Appl Environ Microbiol* **68,** 1595-1603.

Sanchez-Romero, J. M., Diaz-Orejas, R. & De Lorenzo, V. (1998). Resistance to Tellurite as a Selection Marker for Genetic Manipulations of Pseudomonas Strains. *Appl Environ Microbiol* **64,** 4040-4046.

Santos, A. L., Siqueira, J. F., Rôças, I. N., Jesus, E. C., Rosado, A. S. & Tiedje, J. M. (2011). Comparing the Bacterial Diversity of Acute and Chronic Dental Root Canal Infections. *PLoS ONE* **6,** e28088.

Schell, M. A., Ulrich, R. L., Ribot, W. J. *et al.* (2007). Type VI secretion is a major virulence determinant in *Burkholderia mallei*. *Molecular Microbiology* **64,** 1466-1485.

Schloss, P. D. & Handelsman, J. (2006). Toward a Census of Bacteria in Soil. *PLoS Comput Biol* **2,** e92.

Schwieger, F. & Tebbe, C. C. (1998). A New Approach To Utilize Single-Strand-Conformation Polymorphism for 16S rRNA Gene-Based Microbial Community Analysis. *Appl Environ Microbiol* **64,** 4870-4876.

Shelver, W. L., Hakk, H., Larsen, G. L., DeSutter, T. M. & Casey, F. X. M. (2010). Development of an ultra-high-pressure liquid chromatography tandem mass spectrometry multi-residue sulfonamide method and its application to water, manure slurry, and soils from swine rearing facilities. *Journal of Chromatography A* **1217,** 1273-1282.

Shimizu, T., Ohtani, K., Hirakawa, H. *et al.* (2002). Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater. *PNAS* **99,** 996-1001.

Sikorski, J., Graupner, S., Lorenz, M. G. & Wackernagel, W. (1998). Natural genetic transformation of Pseudomonas stutzeri in a non-sterile soil. *Microbiology* **144,** 569-576.

Sinclair, R., Boone, S. A., Greenberg, D., Keim, P. & Gerba, C. P. (2008). Persistence of Category A Select Agents in the Environment. *Appl Environ Microbiol* **74,** 555-563.

Smeets, L. C. & Kusters, J. G. (2002). Natural transformation in Helicobacter pylori: DNA transport in an unexpected way. *Trends in Microbiology* **10,** 159-162.

Smith, M. D., Angus, B. J., Wuthiekanun, V. & White, N. J. (1997). Arabinose assimilation defines a nonvirulent biotype of Burkholderia pseudomallei. *Infect Immun* **65,** 4319-4321.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M. & Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored rare biosphere *Proceedings of the National Academy of Sciences* **103,** 12115-12120.

Speert, D. P. (2001). Understanding *Burkholderia cepacia*: epidemiology, genomovars and virulence. *Infections in Medicine* **18,** 49-56.

Speert, D. P. (2002). Advances in *Burkholderia cepacia* complex. *Paediatric Respiratory Reviews* **3,** 230-235.

Spilker, T., Baldwin, A., Bumford, A., Dowson, C. G., Mahenthiralingam, E. & LiPuma, J. J. (2009). Expanded Multilocus Sequence Typing for Burkholderia Species. *J Clin Microbiol* **47,** 2607-2610.

Stearns, J. C., Lynch, M. D., Senadheera, D. B., Tenenbaum, H. C., Goldberg, M. B., Cvitkovitch, D. G., Croitoru, K., Moreno-Hagelsieb, G. & Neufeld, J. D. (2011). Bacterial biogeography of the human digestive tract. *Sci Rep* **1,** 170.

Steffan, R. J., Goksöyr, J., Bej, A. K. & Atlas, R. M. (1988). Recovery of DNA from soils and sediments. *Appl Environ Microbiol* **54,** 2908-2915.

Stevens, M. P., Wood, M. W., Taylor, L. A., Monaghan, P., Hawes, P., Jones, P. W., Wallis, T. S. & Galyov, E. E. (2002). An Inv/Mxi-Spa-like type III protein secretion system in *Burkholderia pseudomallei* modulates intracellular behaviour of the pathogen. *Molecular Microbiology* **46,** 649-659.

Suchodoloski, J. S., Dowd, S. E., Westermark, E., Steiner, J. M., Wolcott, R. D., Spillmann, T. & Harmoinen, J. A. (2009). The effect of the macrolide antibiotic tylosin on microbial diversity in the canine small intestine as demonstrated by massive parallel 16S rRNA gene sequencing. *BMC Microbiology* **9**.

Sultan, B., Labadi, K., Gu+®gan, J. F. & Janicot, S. (2005). Climate Drives the Meningitis Epidemics Onset in West Africa. *PLoS Med* **2,** e6.

Summer, E. J., Gonzalez, C. F., Carlisle, T., Mebane, L. M., Cass, A. M., Savva, C. G., LiPuma, J. & Young, R. (2004). *Burkholderia cenocepacia* Phage BcepMu and a Family of Mu-like Phages Encoding Potential Pathogenesis Factors. *Journal of Molecular Biology* **340,** 49-65.

Sun, Y., Exley, R., Li, Y., Goulding, D. & Tang, C. (2005). Idenitification and Characterization of Genes Required for Competence in *Neisseria meningitidis*. *Journal of Bacteriology* **187,** 3273-3276.

Tebbe, C. C. & Vahjen, W. (1993). Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Appl Environ Microbiol* **59,** 2657-2665.

Thomas, C. M. & Nielsen, K. M. (2005). Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nat Rev Micro* **3,** 711-721.

Thongdee, M., Gallagher, L. A., Schell, M., Dharakul, T., Songsivilai, S. & Manoil, C. (2008). Targeted Mutagenesis of *Burkholderia thailandensis* and *Burkholderia pseudomallei* through Natural Transformation of PCR Fragments. *Appl Environ Microbiol* **74,** 2985-2989.

Tiedje, J. M., Asuming-Brempong, S., Nusslein, K., Marsh, T. L. & Flynn, S. J. (1999). Opening the black box of soil microbial diversity. *Applied Soil Ecology* **13,** 109-122.

Tissot-Dupont, H., Amadei, A. M., Nezri, M. & Raoult, D. (2004). Wind in November, Q fever in December. *Emerging Infectious Disease* **10**.

Tomlin, K. L., Clark, S. R. D. & Ceri, H. (2004). Green and red fluorescent protein vectors for use in biofilm studies of the intrinsically resistant *Burkholderia cepacia* complex. *Journal of Microbiological Methods* **57,** 95-106.

Tong, S., Yang, S., Lu, Z. & He, W. (1996). Laboratory investigation of ecological factors influencing the environmental presence of *Burkholderia pseudomallei*. *Microbiology and Immunology* **40,** 451-453.

Tong, Y. & Lighthart, B. (2000). The Annual Bacterial Particle Concentration and Size Distribution in the Ambient Atmosphere in a Rural Area of the Willamette Valley, Oregon. *Aerosol Science and Technology* **32,** 393-403.

Torsvik, V., Goksöyr, J. & Daae, F. L. (1990). High diversity in DNA of soil bacteria. *Appl Environ Microbiol* **56,** 782-787.

Tortoli, E. (2003). Impact of Genotypic Studies on Mycobacterial Taxonomy: the New Mycobacteria of the 1990s. *Clinical Microbiology Reviews* **16,** 319-354.

Tralau, T., Cook, A. M. & Ruff, J. (2001). Map of the IncP1-β Plasmid pTSA Encoding the Widespread Genes (tsa) forp-Toluenesulfonate Degradation in Comamonas testosteroni T-2. *Appl Environ Microbiol* **67,** 1508-1516.

Trevors, J. T. (1996). Nucleic acids in the environment. *Current Opinion in Biotechnology* **7,** 331-336.

Tringe, S. G., Zhang, T., Liu, X. et al., (2008). The Airborne Metagenome in an Indoor Urban Environment. *PLoS ONE* **3,** e1862.

Tringe, S. G., von Mering, C., Kobayashi, A. *et al.*, (2005). Comparative Metagenomics of Microbial Communities. *Science* **308,** 554-557.

Trung, T. T., Hetzer, A., Gähler, A., Topfstedt, E., Wuthiekanun, V., Limmathurotsakul, D., Peacock, S. J. & Steinmetz, I. (2011). Highly Sensitive Direct Detection and Quantification of *Burkholderia pseudomallei* Bacteria in Environmental Soil Samples by Using Real-Time PCR. *Appl Environ Microbiol* **77,** 6486-6494.

Turnbull, C. B. (1990) Anthrax. In Topley and Wilson's Principles of Bacteriology, Virology and Immunity, Vol. 2, ed. Smith, G.R. & Easmon, C.R. pp. 364-377. Sevenoaks, Edward Arnold.

Tsai, Y. L. & Olson, B. H. (1992). Detection of low numbers of bacterial cells in soils and sediments by polymerase chain reaction. *Appl Environ Microbiol* **58,** 754-757.

Tsang, J. S. H. (2004). Molecular Biology of the *Burkholderia cepacia* complex. *Advances in Applied Microbiology* **54,** 71-91.

Tuanyok, A., Auerbach, R. K., Brettin, T. S. & other authors (2007). A Horizontal Gene Transfer Event Defines Two Distinct Groups within *Burkholderia pseudomallei* That Have Dissimilar Geographic Distributions. *Journal of Bacteriology* **189,** 9044-9049.

Tyler, S. D., Rozee, K. R. & Johnson, W. M. (1996). Identification of IS1356, a new insertion sequence, and its association with IS402 in epidemic strains of Burkholderia cepacia infecting cystic fibrosis patients. *J Clin Microbiol* **34,** 1610-1616.

Ulrich, R. L., DeShazer, D., Brueggemann, E. E., Hines, H. B., Oyston, P. C. & Jeddeloh, J. A. (2004). Role of quorum sensing in the pathogenicity of *Burkholderia pseudomallei*. *J Med Microbiol* **53,** 1053-1064.

Ulrich, R. L., Ulrich, M. P., Schell, M. A., Kim, H. S. & DeShazer, D. (2006). Development of a polymerase chain reaction assay for the specific identification

of *Burkholderia mallei* and differentiation from *Burkholderia pseudomallei* and other closely related Burkholderiaceae. *Diagnostic Microbiology and Infectious Disease* **55,** 37-45.

Urich, T., Lanzön, A., Qi, J., Huson, D. H., Schleper, C. & Schuster, S. C. (2008). Simultaneous Assessment of Soil Microbial Community Structure and Function through Analysis of the Meta-Transcriptome. *PLoS ONE* **3,** e2527.

Vandamme, P. A. R., Govan, J. R. W. & LiPuma, J. J. (2007). Diversity and role of *Burkholderia* spp. In *Burkholderia* Molecular Microbiology and Genomics, vol. 1, pp. 1–28. Edited by T. Coenye & P. Vandamme. Wymondham, UK: Horizon Bioscience.

Van de Peer, Y., Chapelle, S. & De Wachter, R. (1996). A Quantitative Map of Nucleotide Substitution Rates in Bacterial rRNA. *Nucl Acids Res* **24,** 3381-3391.

van der Grinten, E., Pikkemaat, M. l. G., van den Brandhof, E.-J., Stroomberg, G. J. & Kraak, M. H. S. (2010). Comparing the sensitivity of algal, cyanobacterial and bacterial bioassays to different groups of antibiotics. *Chemosphere* **80,** 1-6.

van Schaik, E. J., Giltner, C. L., Audette, G. F., Keizer, D. W., Bautista, D. L., Slupsky, C. M., Sykes, B. D. & Irvin, R. T. (2005)**.** DNA Binding: a Novel Function of *Pseudomonas aeruginosa* Type IV Pili. *Journal of Bacteriology* **187,** 1455-1464.

van Waasbergen, L. G., Balkwill, D. L., Crocker, F. H., Bjornstad, B. N. & Miller, R. V. (2000). Genetic Diversity among *Arthrobacter* Species Collected across a Heterogeneous Series of Terrestrial Deep-Subsurface Sediments as Determined on the Basis of 16S rRNA and *recA* Gene Sequences. *Appl Environ Microbiol* **66,** 3454-3463.

Vanlaere, E., Baldwin, A., Gevers, D. & other authors (2009). Taxon K, a complex within the *Burkholderia cepacia* complex, comprises at least two novel species, *Burkholderia contaminans* sp. nov. and *Burkholderia lata* sp. nov. *Int J Syst Evol Microbiol* **59,** 102-111.

Vanlaere, E., LiPuma, J. J., Baldwin, A., Henry, D., De Brandt, E., Mahenthiralingam, E., Speert, D., Dowson, C. & Vandamme, P. (2008). *Burkholderia latens* sp. nov., *Burkholderia diffusa* sp. nov., *Burkholderia arboris* sp. nov., *Burkholderia seminalis* sp. nov. and *Burkholderia metallica* sp. nov., novel species within the *Burkholderia cepacia* complex. *Int J Syst Evol Microbiol* **58,** 1580-1590.

Vermis, K., Coenye, T., Mahenthiralingam, E., Nelis, H. J. & Vandamme, P. (2002). Evaluation of species-specific *recA*-based PCR tests for genomovar level identification within the Burkholderia cepacia complex. *J Med Microbiol* **51,** 937-940.

Vettori, C., Paffetti, D., Pietramellara, G., Stotzky, G. & Gallori, E. (1996). Amplification of bacterial DNA bound on clay minerals by the random amplified polymorphic DNA (RAPD) technique. *FEMS Microbiology Ecology* **20,** 251-260.

Vos, M., Quince, C., Pijl, A. S., de Hollander, M. & Kowalchuk, G. A. (2012). A Comparison of *rpoB* and 16S rRNA as Markers in Pyrosequencing Studies of Bacterial Diversity. *PLoS ONE* **7,** e30600.

Waag, D. M. & DeShazer, D. (2004). Glanders: New Insights into and Old Disease. In: *Biological Weapons Defense: Infectious Diseases and Counterbioterrorism*, pp. 209-237. Edited by L. E. Lindler, F. J. Lebeda, G. W. Korch Humana Press Inc.

Wachsmuth, I. K., Blake, P. A. & Olsvik, Ø. (1994). *Vibrio cholerae and Cholera: Molecular to Global Perspectives*.

Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol* **73,** 5261-5267.

Washington, R., Todd, M., Middleton, N. J. & Goudie, A. S. (2003). Dust-Storm Source Areas Determined by the Total Ozone Monitoring Spectrometer and Surface Observations. *Annals of the Association of American Geographers* **93,** 297-313.

Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* **173,** 697-703.

Whitchurch, C. B., Tolker-Nielsen, T., Ragas, P. C. & Mattick, J. S. (2002). Extracellular DNA Required for Bacterial Biofilm Formation. *Science* **295,** 1487.

White, P. J., Squirrell, D. J., Arnaud, P., Lowe, C. R. & Murray, J. A. (1996). Improved thermostability of the North American firefly luciferase: saturation mutagenesis at position 354. *Biochem J* **319,** 343-350.

Whitehouse, C. A. & Hottel, H. E. (2007). Comparison of five commercial DNA extraction kits for the recovery of *Francisella tularensis* DNA from spiked soil samples. *Molecular and Cellular Probes* **21,** 92-96.

Wicker, T., Schlagenhauf, E., Graner, A., Close, T. J., Keller, B. & Stein, N. (2006). 454 sequencing put to the test using the complex genome of barley. *BMC Genomics* **7**.

Widmer, F., Seidler, R. J., Donegan, K. K. & Reed, G. L. (1997). Quantification of transgenic plant marker gene persistence in the field. *Molecular Ecology* **6,** 1-7.

Wolfgang, M., Lauer, P., Park, H. S., Brossay, L., Hebert, J. & Koomey, M. (1998). PilT mutations lead to simultaneous defects in competence for natural transformation and twitching motility in piliated *Neisseria gonorrhoeae*. *Molecular Microbiology* **29,** 321-330.

Womack, A. M., Bohannan, B. J. M. & Green, J. L. (2010). Biodiversity and biogeography of the atmosphere. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365,** 3645-3653.

Yabuuchi, E., Kosako, Y., Oyaizu, H., Yano, I., Hotta, H., Hashimoto, Y., Ezaki, T. & Arakawa, M. (1992). Proposal of *Burkholderia* gen. nov. and transfer of seven species of the genus *Pseudomonas* homology group II to the new genus, with the type species Burkholderia cepacia (Palleroni and Holmes 1981) comb. nov. *Microbiol Immunol* **36,** 1251-1275.

Yeates, C., Gillings, M. R., Davison, A. D., Altavilla, N. & Veal, D. A. (1998). Methods for microbial DNA extraction from soil for PCR amplification., pp. 40-47.

Yeh, Y. C., Lin, T. L., Chang, K. C. & Wang, J. T. (2003). Characterization of a ComE3 homologue essential for DNA transformation in *Helicobacter pylori*. *Infect Immun* **71,** 5427-5431.

Yin, H., Cao, L., Xie, M., Chen, Q., Qiu, G., Zhou, J., Wu, L., Wang, D. & Liu, X. (2008). Bacterial diversity based on 16S rRNA and gyrB genes at Yinshan mine, China. *Systematic and Applied Microbiology* **31,** 302-311.

Youssef, N., Sheik, C. S., Krumholz, L. R., Najar, F. Z., Roe, B. A. & Elshahed, M. S. (2009). A Comparative study of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-

generated fragments in 16S rRNA gene-based environmental surveys. *Appl Environ Microbiol*.

Yu, Y., Kim, H. S., Chua, H. *et al.*, (2006). Genomic patterns of pathogen evolution revealed by comparison of *Burkholderia pseudomallei*, the causative agent of melioidosis, to avirulent *Burkholderia thailandensis*. *BMC Microbiology* **6,** 46.

Yu, Z. & Morrison, M. (2004). Comparisons of Different Hypervariable Regions of rrs Genes for Use in Fingerprinting of Microbial Communities by PCR-Denaturing Gradient Gel Electrophoresis. *Appl Environ Microbiol* **70,** 4800-4806.

Zhang, K., Martiny, A. C., Reppas, N. B., Barry, K. W., Malek, J., Chisholm, S. W. & Church, G. M. (2006). Sequencing genomes from single cells by polymerase cloning. *Nat Biotech* **24,** 680-686.

Zhang, T., Shao, M. F. & Ye, L. (2011). 454 Pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. *ISME J*.

Zinger, L., Gobet, A. & Pommier, T. (2011). Two decades of describing the unseen majority of aquatic microbial diversity. *Molecular Ecology* doi: 10.1111/j.1365-294X.2011.05362.x