The Physiological and Ecological Characterisation of the First Cultivated Species of the Candidate Division OP10

A thesis
submitted in partial fulfilment
of the requirements for the degree

of

Master of Science in Biology

at

The University of Waikato

by

Chien-Yu Kevin Lee

The University of Waikato

2010

Abstract

Bacteria are one of the three taxonomic domains and play a major role in the biological processes on Earth, yet their taxonomy and constituent species are poorly understood. “OP10” is a candidate division within the domain *Bacteria* that up until recently has no cultivated representative. Until the recent isolation of the cultivated OP10 strains T49 and P488 from the Taupo Volcanic Zone in New Zealand (Stott et al. 2008), little was known about this putative phylum despite the fact that its presence was detected in various environmental surveys.

This study aims to (i) characterise the physiology of one of the first isolated OP10 species, strain T49; and (ii) investigate and expand upon known OP10 diversity by using targeted oligonucleotide primers.

Optimal growth temperature and pH of T49 was determined by cultivation in liquid medium. T49 substrate utilisation was investigated by culturing T49 with various substrates, with a focus on its sugar utilisation profiles. T49 was also characterised for its cell morphology through transmission and scanning electron microscopy as well as phase contrast microscopy. The total fatty acid profile of T49 was characterised through gas chromatography and mass spectroscopy.

OP10 diversity was investigated at several geothermal sites using OP10-specific primers designed in the course of this study. The primers target conserved regions of the 16S rRNA gene sequence specific to clades within OP10. The primers were designed *in silico* by obtaining publically available OP10 16S rRNA gene sequences (through the Greengenes and SILVA databases) as references and dividing the candidate division into distinct clades through phylogenetic analysis with ARB. Primers were selected for specificity and coverage of the reference sequences within each of the clades. The performance of the primers were investigated by applying the primers against environmental DNA samples where OP10 sequences were previously identified, as well as various negative controls. The results were validated by sequencing PCR products generated to identify positive OP10 hits and false positive non-OP10 hits.

This study found that T49 had a rod-shape morphology with the dimensions of 0.5-0.7 µm by 2.5-3.0 µm, more than half (67.4%) of the total fatty acids of the bacteria consisted of stearic acid (18:0) and palmitic acid (16:0), and GC content of 54.6%. Furthermore, T49 was found to have a growth temperature range of <50 °C to 73 °C and optimum at 68 °C, growth pH range of 4.9 to 5.8 and optimum at pH 5.3. Substrate utilisation experiments showed that strain T49 grew optimally using mono- and di- saccharides such as arabinose, mannose, ribose, galactose, and maltose, as well as amorphous polysaccharides including starch, glycogen, and dextrin.

The OP10-specific primers successfully identified the presence of OP10 in environmental samples from Waikite, Waipahihi, and Mount Ngauruhoe , confirming previous unpublished 16S rRNA gene sequence surveys. Furthermore, the OP10-specific primers also identified previously undetected genetic diversity in Tikitere and Mount Ngauruhoe, demonstrating the advantage of the primers over traditional universal primers with their high specificity for OP10 DNA sequences which enabled more sensitive surveys of OP10 in the environment.

The characterisation of the first OP10 species and the development of OP10-specific primers enable further investigation into this major taxonomic group within the *Bacteria* domain. T49 and its related strains may play important role within their environments, which can now be investigated based on these known physiologies. The effectiveness of the OP10-specific primers experimented may be use to detect presence of OP10 species in the environment at a higher sensitivity and selectiveness than traditional primers. The results from this research have expanded our knowledge on this previously unknown phylum. By increasing our knowledge of the OP10 candidate division, we also increase our understanding of the global bacterial diversity and this may help bring about insights into biology and global climate processes as well generating practical solutions in these fields.

Many of these initial observations of T49 merit further investigation. These include: The production of pigments and lipids, two distinct cytosolic structures, and interesting growth behaviours such as quorum sensing and biofilm formation.

OP10-specific primers developed in this study can be improved upon through further validations with environmental DNA representing clades that were not tested during this study due to unavailability. The improved primers can be developed as a rapid diagnostic tool to detect OP10 in the environment for isolation efforts. The primers may also act as fluorescent probes to identify OP10 in microbial consortia such as biofilms while preserving the structure of the microenvironments.

Acknowledgements

I would like to thank my supervisors: Dr. Matthew Stott and Associate Professor Ian McDonald. The amount of patience and advice given through the research, and writing of this thesis can only be expressed through scientific notation. Your lively spirits made science approachable and enjoyable, it's been an honour.

I would like to thank Dr. Xochitl Morgan, who's been practically a *de facto* supervisor, looking after my well being beyond the call of duty, your reassuring words kept me grounded during this amazing/hectic process. I promise I will read Strunk and White after this.

Thanks to GNS Science for the scholarship, and the people at IRL, Jason, Kirill and Mikhail for their work on T49 lipids. Thanks to my current and past labmates Karen, Jean, Nadine, Rashidah, and Heike for the assistance and laughs we've shared, and jumping out of the plane with me just to procrastinate for half a day. Thanks to everyone at GNS Wairakei, we shall continue to rule the pub quiz with an iron fist.

Thanks to mum and dad for calling me while you're on holiday while I was knee deep in my thesis, I love you. My flatmates for having to deal with a nocturnal grad-student that haunts the flat. A big thank you to all my friends for putting up with a sleep-deprived, semi-coherent me, sorry to disappoint you, but I think it's natural.

I hereby shall be the first among many to formally acknowledge Gabe and his amazing Buffalo wings. I retract anything negative I've ever made in regards to American food. The spiciness on a stick you made made me appreciate life like never before.

Here's to sleep and everything it stands for.

Contents

ABSTRACT ..... **III**

ACKNOWLEDGEMENTS ..... **VI**

LIST OF FIGURES IX

LIST OF TABLES XII

1 LITERATURE REVIEW **1**

1.1 INTRODUCTION 1

1.2 RELEVANCE 4

1.3 CULTURE-DEPENDENT METHODS 6

1.3.1 Disadvantages of culture-dependent methods *6*

1.4 CULTURE-INDEPENDENT METHODS 8

1.4.1 Polymerase Chain Reaction (PCR) *8*

1.4.2 Clone library *9*

1.4.3 Restriction Fragment Length Polymorphism (RFLP) *9*

1.4.4 Disadvantages of PCR-based methods *10*

1.5 EMERGING APPROACHES 11

1.5.1 Culture-based approach *11*

1.5.2 Microscopy *11*

1.5.3 Culture-independent approach *12*

1.6 EXTREMOPHILES 13

1.7 CANDIDATE DIVISION OP10 15

1.8 TAUPO VOLCANIC ZONE 19

1.9 AIMS, APPROACHES 21

2 MATERIALS AND METHODS **23**

2.1 GENERAL 23

2.2 CULTURE-BASED CHARACTERISATION OF T49 23

2.2.1 Components for culture media *23*

2.2.2 Bacteriological media preparation *25*

2.3 SAMPLE COLLECTION FOR CULTURE-BASED AND CULTURE-INDEPENDENT EXPERIMENTS 28

2.4 ISOLATION AND ENRICHMENT PROTOCOLS 32
2.5 T49 CHARACTERISATION 33
2.5.1 Maintaining T49 stock on media plates 33
2.5.2 Inoculation and sampling of liquid media 34
2.5.3 Bacterial turbidity measurements 34
2.5.4 Cell counting 35
2.5.5 Substrate utilisation 35
2.5.6 Temperature optimum and range determination 35
2.5.7 pH optimum and range determination 35
2.5.8 Methane utilisation 36
2.5.9 TEM/SEM microscopy 37
2.5.10 GC content determination 37
2.5.11 Fatty acid methyl esters (FAME) profile 38
2.6 ENVIRONMENTAL SURVEYS WITH OP10-SPECIFIC PRIMERS 40
2.6.1 DNA extraction and PCR from environmental samples 40
2.6.2 OP10 primer design 42
2.6.3 Primer testing 45
2.6.4 Estimating PCR sampling depth 47
3 RESULTS 49
3.1 T49 CULTURE BASED CHARACTERISATION 49
3.1.1 General colony and culture observations 49
3.1.2 Substrate utilisation 53
3.1.3 Validation of optical density measurements 57
3.1.4 Optimal temperature and temperature range 57
3.1.5 Optimum pH and pH range 58
3.1.6 TEM/SEM analysis 59
3.1.7 Summary of characteristics of T49 and a closely related strain P488 (Stott et al. 2008) 63
3.1.8 Isolation of bacteria 65
3.1.9 GC content 65
3.1.10 T49 fatty acid methyl ester (FAME) profile 65
3.1.11 Methane utilisation 66

3.2 TARGETED ENVIRONMENTAL SURVEYS WITH OP10-SPECIFIC PRIMERS 66
3.2.1 PCR with OP10-specific primers 66
3.2.2 Sequencing of PCR products and phylogenetic analysis 67
3.2.3 PCR optimisation 72
3.2.4 Sampling coverage estimation 74
4 DISCUSSION AND CONCLUSIONS 76
PHYSIOLOGY OF T49 76
ENVIRONMENT SURVEYS WITH OP10-SPECIFIC PRIMERS 78
Future perspectives 79
CONCLUSIONS 80
BIBLIOGRAPHY 81
APPENDIX 1: LIST OF CHEMICALS AND REAGENTS 93
APPENDIX 2: PREVIOUS 16S RRNA GENE SEQUENCES FROM SAMPLE SITES 97
APPENDIX 3: REFERENCE PHYLOGENETIC TREE 101
APPENDIX 4: REFERENCE PHYLOGENETIC TREE WITH NEW SEQUENCES ... 102

List of Figures

Figure 1 - Cumulative published microbial species names from 1980 to 2009 (Euzéby, 1997). 2
Figure 2 - Phylogenetic tree of selected bacterial phyla 17
Figure 3 - Phyla closely related to Candidate Division OP10 17
Figure 4 – The environmental distribution of OP10 16S rRNA gene sequences from the Greengenes database 18
Figure 5 - The locations of the sample sites used in this study 19

Figure 6 - A Schott® bottle with a butyl septum 25

Figure 7 - Waikite Hot Spring, pink soil sample was collected from the steam affected cliff wall just above the spring. The pink soil sample was covered by a layer of green photosynthetic mat. 30

Figure 8 - Taupo lakefront at Waipahihi beach. 31

Figure 9 - Taupo wastewater treatment plant, trickling filter flow. 31

Figure 10 - Tikitere, stratified soil profile of steam affected soil. The soil sample used for inoculation was collected from the orange layer. Soil temperature was measured *in situ* at 52 °C. 32

Figure 11 - Process flowchart of 16S rRNA gene sequence analysis with OP10 specific primers 40

Figure 12 – T49 colonies, showing its round morphology and pink-orange colour. 49

Figure 13 - T49 in liquid culture settling at the bottom of the container, leaving clear medium on top 50

Figure 14 - T49 in phase contrast microscopy, showing its rod-shape morphology and a small amount of cell clumping 51

Figure 15 - T49 colonies on an AOM1 plate. A) Arrows indicate indentations around the colonies 51

Figure 16 - Liquid media A) before inoculation and B) after 1 week of growth ... 52

Figure 17 - Normalised measurements of T49 growth in a liquid culture. 57

Figure 18 - Normalised T49 growth rates at log phase at various incubation temperatures 58

Figure 19 - T49 log phase growth rate at 60 °C between pH 4.7 to pH 6.2 59

Figure 20 - T49 TEM image showing the outer A), inner membrane B), and peptidoglycan layer C) as indicated by the arrows. 60

Figure 21 – TEM image of A) Electron-light cavities and B) Electron-dense granules present in T49. 61

Figure 22 - TEM image of T49 showing full crosssection. 62

Figure 23 - SEM image of T49 with wrinkled surfaces embedded in the material carried- through from culture colonies 63

Figure 24 - Sequences obtained from PCR with 27F and OPX-1BR primer set and DNA extracted from environmental sample WPH03. All the sequences fell within clade 1B. 69

Figure 25 - Sequences obtained from PCR with OPX-2F and 1492R primer set and DNA extracted from environmental sample NGH02. The sequences fell with in multiple clades. 70

Figure 26 - Sequences obtained from PCR with 27F and OPX-4A2R primer set and DNA extracted from environmental sample TT13.4, NGH02, and WKT32. All the sequences fell within clade 4A2, with TT13.4 & WKT32 sequences clustering around T49 sequences (coloured yellow). 71

Figure 27 – Electrophoresis gel detailing primer optisation of the primer set OPX-2F and 1492R (designed to detect OP10 from Clade 2 in the NGH02 environmental sample). Lane: 1.WPH01; 2. WKT32; 3., WPH03; 4. TT13.4; 5. *M. acidophila;* 6. *Chloroflexus sp.;* 7. *T. lanuginosus;* 8. T49; 9. NGH02. ZR 1kb DNA ladder was used (Zymo Research) 73

Figure 28 – An electrophoresis gel showing the result of a temperature gradient PCR with primers 27F and OPX-4A2R with T49 DNA at 0.16 ng/mL and gradient temperature of 64 – 70 °C to demonstrate the temperature and concentration limit of the reaction. 74

List of Tables

Table 1 - Sample site locations and descriptions. 28

Table 2 - Experimental setup for methane utilisation test 36

Table 3 - PCR thermalcycling programme for bacterial 16S rRNA gene primers 9F + 1492R..... 41

Table 4 - Touchdown thermalcycling programme for M13 clone library primers 42

Table 5 - Growth reaction of T49 using various substrates and their properties . 54

Table 6 - Comparison of selected phyysiological characteristics between strain T49 and P488 (Stott et al, 2008). 64

Table 7 - Isolated and ongoing isolation cultures from two environmental samples 65

Table 8 - List of T49 fatty acid methyl esters and their relative percentages. 66

Table 9 - PCR results of OP10 specific primers against DNA samples 68

Table 10 – PCR coverage - Chao1 and ACE OTU estimation vs. number of OTUs sampled. 75

1 Literature Review

1.1 Introduction

Bacteria are the most ubiquitous life form on Earth. Where there is life, there will be bacteria surviving, if not thriving in those environments, due to their remarkable diversity and ability to adapt. Despite being the dominant life form on Earth in terms of genetic diversity (Ciccarelli et al. 2006) and biomass (Whitman, Coleman, & Wiebe, 1998), surprisingly little is known about them. This is due to the intrinsic nature of microscopic organisms and the limitations of current technologies used to study these life forms.

It has been estimated that the total number of *Bacteria* and *Archaea* are approximately 4-6 x 10^{30} with a biomass of 350–550 Pg (1 Pg = 10^{15} grams) (Whitman, Coleman, and Wiebe 1998). This biomass represents a significant pool of carbon, phosphorous, iron and other biologically active elements in the environment. Of these microorganisms, only a small fraction of species have been identified, and an even smaller fraction have been properly studied and classified. Our understanding of microbial diversity can be divided into three groups, from most- to least- characterised.

1. Microorganisms that have been identified and studied, which still require culture-dependent methods in order to understand physiological characteristics. It is worth noting that new technologies such as the use of optical tweezers and single cell sequencing are being developed to overcome this limitation. This will be discussed in the emerging section (section 1.5). For the time being, however, these are only a small fraction (roughly 0.09%) of total bacterial species estimated (Euzéby 1997; Gans, Wolinsky, and Dunbar 2005).
2. Microorganisms that have been identified through culture independent techniques, most commonly environmental 16S rRNA gene sequence analysis. With this information alone, only the phylogenetic placements of these microbes are ascertained, with little further information. High-throughput sequencing, which will be discussed in the emerging approaches section, aims to expand this category.
3. Microbial diversity estimated by C_0t analysis, also known as DNA reassociation kinetics. The method is based on the notion that, after denaturing a mixture of DNA through heating, the more common sequences are, through the chemical principals

of the collision theory and complementary pairing, faster at reassociating with their complementary strands as the temperature drops. By measuring the amount of DNA reassociating at various temperatures, an estimation of unique DNA sequences can be made(Torsvik, Goksøyr, and Daae 1990). This approach provides the least amount of information, but avoids the biases and limitations of the second approach. It is also relatively simple and affordable compared to large-scale sequencing of environmental DNA samples.

In 1980, the Bacteriological Code of Nomenclature was published. Previous species names meeting the requirements of the code, with adequate descriptions and type cultures, were published as the *Approved Lists of Bacterial Names* (which contained 1792 species). This marked a new standard for consistent publication and nomenclature (http://www.the-icsp.org/misc/Code_history.htm).

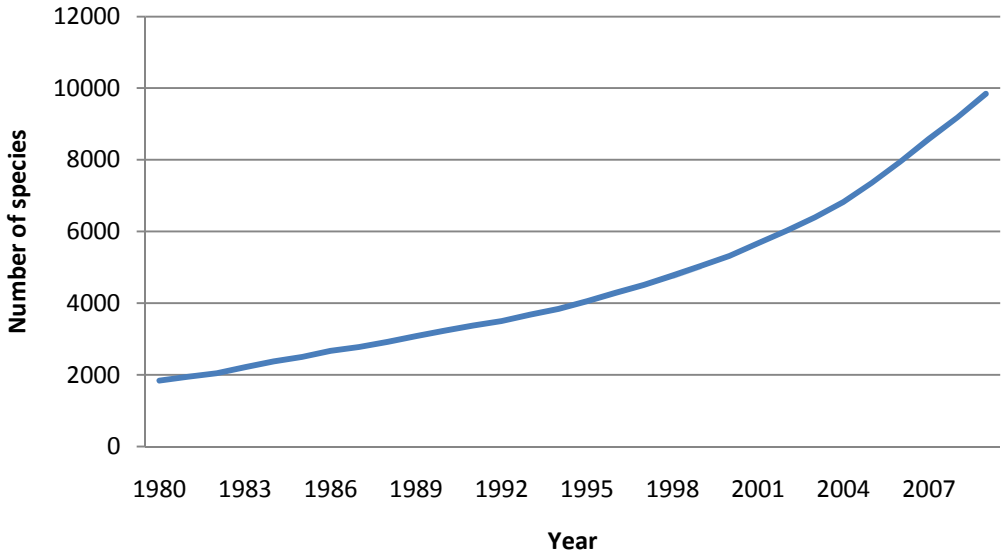


Figure 1 - Cumulative published microbial species names from 1980 to 2009 (Euzéby, 1997)**.**

The size and composition of the group 1 (microorganisms that have been studied) can be estimated from the number of formally declared species. As of 2009, there are 9841 microbial species (http://www.bacterio.cict.fr/number.html) and 33 phyla (*Bacteria* and

Archaea) published. These figures represent an approximate number of species currently known, as shown in Figure 1.

Despite the increase in known species in recent times, estimates of known species versus unknown species have mostly decreased among researchers. An optimistic early review estimated that “less than 20%” of all bacterial species in the world are known (Wayne et al. 1987) while a more recent estimate gave a figure of less than 0.09% (Euzéby 1997; Gans, Wolinsky, and Dunbar 2005).

Clone libraries constructed for environmental DNA analysis tend to behave in a predictable way in response to the amount of diversity being sampled. The less the diversity, the greater the likelihood that the clones will have similar or identical sequences. Curtis, Sloan, & Scannel (2002) used this principal to construct a mathematical model to estimate the range of bacterial taxa (here defined as any group distinguishable through 16S rRNA gene sequences) in three major environments. Bacterial taxa in the ocean were the fewest at less than 2 million, while soil and atmospheric samples each contained at least 4 million taxa. The richer biodiversity in soils compared to aquatic ecosystems is generally attributed to the higher spacial heterogeneity in soil (Stott et al. 2008). However, the study (Curtis, Sloan, and Scannell 2002) based its model on 16S rRNA gene clone library statistics and estimated the bacterial diversity of the representative samples (i.e. the bacterial diversity of a gram of soil) at a number much lower than that given by DNA reassociation analysis, the third category mentioned above.

An early C_0t analysis study (Torsvik, Goksøyr, & Daae, 1990) demonstrated the high diversity of bacterial communities in a rich soil environment, so much so that it made estimation difficult. A later study, using a less diverse environment (heavy metal-contaminated soil) as a model, extrapolated the staggering figure of 8.3×10^6 species among 10^{10} cells in the 10 g of pristine soil that Torsvik would have used (Gans, Wolinsky, and Dunbar 2005).

The only certain statement that can be made in regard to the attempts to estimate bacterial diversity is that the diversity appears to be so large, while our understanding of these organisms is so limited, that our estimations are often many orders of magnitude different from one another. Phylogenetic analysis of 16S rRNA gene sequences has shown that when

comparing cultured species (group 1) and uncultured species (group 2), the unknown diversity does not just exist at the extremities of the tree branches. Instead, major branches at phylum level are found with no cultured species and little information about the groups (Rappé and Giovannoni 2003). This shows that bacteria genetically very different from currently known species are detected, but have not been studied and characterised.

As of the time of writing, there were currently 28 bacterial and 5 archaeal phyla, with cultured and characterised species to represent the taxonomic groups (listed under http://www.bacterio.cict.fr/number.html). Phyla without any cultured species are given a putative category known as "candidate divisions" (Hugenholtz et al. 1998). The exact number of candidate divisions is difficult to determine since no single authority administers over this putative taxonomic grouping, and the number may fluctuate as new candidate divisions are found or upgraded into full-fledged phyla. A cross-referencing analysis between several different studies has shown that the number can range from 31 to 88 depending on the curators (DeSantis et al. 2006).

The figures above demonstrate that over half of the bacterial diversity at the phylum level has not been studied. The role of how the bacteria belonging to these candidate divisions impact the world around us, to how these organisms affect our understanding of life on Earth remains mostly a mystery.

1.2 Relevance

Expanding our understanding of the great microbial diversity will affect our understanding of biological processes, if what we have learnt from a small subset of bacteria currently available is any indicator. Microorganisms have exhibited adaptation due to their relative simplicity and genetic fluidity. While individual microorganisms seem to have little effect on the environment, with the correct conditions, they can reproduce at a tremendous rate; their collective effect shapes the planet we inhabit. It is thought that the oxygen generated by photosynthetic cyanobacteria changed the chemical composition of our previously anoxic planet (e.g. iron became iron oxide) and encouraged biological diversification that lead to the modern biological world (Herrero and Flores 2008).

Bacterial and archaeal influences on global biogeochemical cycles also have modern day relevance, particularly due to environmental implications such as global warming. Methane, for example, is a potent greenhouse gas that is produced by methanogens which can be found in ruminant animals (Butcher 1992). Due to emission trading regulations, countries such as New Zealand will likely suffer economically due to the greenhouse gas produced. It is therefore economically relevant to understand the complex microbial consortium in these methanogenic systems in order to address the issue.

Understanding microbial diversity is relevant for bio-prospecting for novel genes or compounds from this same diversity. The diverse metabolic capabilities of bacteria have been exploited for various applications. Bacteria have been used to treat undesirable pollutants in the environment. In applications such as the degradation of chlorinated hydrocarbons, oil spills, or even recovering uranium from contaminated ground water (Anderson et al. 2003), bioremediation appeared to be more efficient and robust than traditional methods (Swannell, Lee, and McDonagh 1996).

To effectively utilise these bacterial properties, a knowledge framework to connect this diversity as well as studies of individual species are vital. Without the framework, it would is difficult to relate one bacterial species to another. Without understanding individual species, it is much more difficult to discover valuable information on bacterial physiologies. Therefore, understanding candidate divisions helps us to connect bacterial species together as a coherent whole by filling in those large gaps in our understanding. By physiologically characterising species, features that define the taxonomic group may be found, or novel, valuable traits may be discovered.

Methods of investigating microbial diversity can be divided into three groups: culture-dependent approaches, culture-independent approaches, and emerging next-generation approaches. Techniques used in this study will be discussed in subsequent sections, while the emerging approach section will cover developments and expansions based on the previous two sections.

1.3 Culture-dependent methods

Culture-dependent methods have been the backbone of bacteriology since its inception in the 19th century. Novel bacterial species are traditionally investigated through cultivation of isolated strains (Ludwig and Klenk 2005). By enriching and eventually isolating a mixed culture down to a single species, the characteristics of the organism can easily be observed as a homogenous population. Traditional bacterial isolation techniques with a mixed culture obtained from the environment involves two strategies:

1. Physically separating bacteria through dilution. This includes, streaking, centrifuging, and serial dilution.
2. Utilising enrichment strategies in order to favour the metabolism of a subgroup within the mixed population. These include changing the condition of the entire culture or constructing a gradient across the culture (e.g. nutrient gradient across an agar plate).

Aside from visual inspection of colonies, microscopy with various staining techniques can differentiate between different species of bacteria by visualising various cellular features. Microscopy is also culture-dependent in the sense that a typical mixed culture often proves too complex to identify all the species. Bacteria in pure cultures offer not only the opportunity to directly observe the phenotypes of a species, but also an easy way to access their genetic material to evaluate their genotypes through processes such as whole genome sequencing.

Information based on culture-dependent techniques with isolated strains is still the *de facto* standard for characterising new species and related taxonomy. Both the International Committee on Systematics of Prokaryotes (ICSP), the body regulating the taxonomy and nomenclature of new species, and the International Journal of Systematic and Evolutionary Microbiology (IJSEM), where new species are often published, require pure cultures to be submitted to culture collections in order to publish the description of new bacterial species.

1.3.1 Disadvantages of culture-dependent methods

Despite the usefulness of physiological information granted by the approach, selective culturing has been noted for its limitations in capturing the wider microbial diversity in the

environment, as many species do not grow in conventional culture media (Rappé and Giovannoni 2003). The biases of culture-dependent methods are introduced by the conditions provided by culture media. Because the culture media is different from the natural habitat of the bacteria, favoured species are overrepresented in cultivation experiments. Furthermore, in a mixed culture, fast-growing bacteria tend to outcompete slower-growing species, although those minority species may have important roles in the population. Finally, formulating selective media and separating species based on growth characteristics and microscopic observations is a process of trial and error, which by itself is an inefficient way to make a dent on the great unknown diversity. Therefore, traditional methods alone could not in any practical sense begin to cover the vast microbial diversity yet to be investigated. The scope of the limitation based on numbers of known species and estimations of the "unknown diversity" is discussed in later sections.

Traditional methods also provide a limited number of physiological characteristics for taxonomy (Woese 1987). Attempts at deducing phylogenetic relationships between taxonomic groups based on these characteristics have proven generally unsuccessful; therefore, bacterial taxonomy was traditionally based on a phenetic system which attempts to numerically classify bacterial species based on their physiological similarities with little or no regards to their evolutionary history. An historical example of this is the four group bacteria taxonomy based on cell morphologies, in which bacteria are categorised as sphere, rod, filament, and spirochete (Cohn 1999).

The phenetic system based on these criteria may be practical in identifying species, especially in situations where only a few candidates are expected, but is problematic in constructing a cohesive framework as the taxonomic system expands. The phenetic system has been criticised for defining a given taxon at an arbitrary level of hierarchy based on the physiological characteristics available (Woese 1987). In the same review, to demonstrate the confusion and arbitrarity of the traditional taxonomic system, *Pseudomonas* was held as an example of an umbrella taxon that commonly includes several distinctively different groups of bacteria with very little in common. To complicate the matter further, the genetic fluidity of bacteria through processes such as convergent evolution makes grouping based on limited numbers of physiological characteristics confusing. Furthermore, the list of

defining criteria required to separate one taxon from another grew as more “exception” species were identified.

1.4 Culture-independent methods

Culture-independent techniques were developed to circumvent the limitations of cultivation techniques. Culture-independent methods involve many molecular analyses. However, 16S rRNA gene sequence analysis has become one of the most common approaches in bacteriology in investigating microbial diversity (Clarridge 2004; Janssen 2006; Weisburg et al. 1991) and has direct relevance to this study.

The basic principle behind 16S rRNA gene sequence analysis is based on genetic divergence of the sequence along with the divergence of species. The sequence therefore represents part of the evolutionary history of the organism, providing a marker for identification and phylogenetic analysis (Weisburg, Barns, Pelletier, & Lane, 1991). The 16S rRNA gene is one of the most ubiquitous and conserved genetic sequences found in bacteria (Fox, Pechman, and Woese 1977) due to its vital role in basic cellular function, thus making phylogenetic comparison within the bacterial domain possible.

The following techniques (sections 1.4.1 – 1.4.3) are all involved in 16S rRNA gene sequence analysis.

1.4.1 Polymerase Chain Reaction (PCR)

Many molecular methods rely on the DNA amplification process of the polymerase chain reaction (PCR) to generate sufficient DNA copies to visualise or use for downstream applications. PCR is a cyclic reaction, with each cycle theoretically doubling the number of DNA copies present. With such exponential growth, a small number of original DNA templates can quickly be amplified into many copies.

Stages in a PCR cycle are controlled by varying the temperature of the reaction mix to separate the double-stranded template DNA into single strands, bind the short DNA primers, and synthesise new copies complementary to those single strands with a thermostable polymerase with free nucleotides in the reaction mix. The products of each cycle then feed back to the first stage to act as templates for further amplification. The reaction kinetics for each reaction setup depend on a combination of many factors such as the template

sequences, the concentration of each reactant, the primer sequences and their specificity to the template sequence, the annealing temperature for the primers to bind to the template, contaminants in the reaction mix. All of these must be considered during the primer design process. In addition, the PCR reaction must be optimised to ensure primers perform as expected.

PCR products can be visualised through agarose gel electrophoresis, in which DNA fragments travel through the agarose gel matrix under an electrical current, and the distance travelled by the fragments reflects the size of the fragments. Bands formed by these fragments can then be visualised via DNA binding fluorescent dyes such as ethidium bromide.

Part of this study focuses on using specific primers to target the 16S rRNA gene of bacteria within the OP10 candidate division. Optimisation of primer pairs for detection threshold and boundary parameters is also described in the results section.

1.4.2 Clone library

Bacterial species in environmental samples are almost invariably mixed. Consequently, a PCR reaction performed on DNA extracted from such samples with bacterial universal 16S rRNA gene primers will result in a pool of 16S rRNA gene sequences reflecting the species present. Clone libraries are useful for separating the sequences. Separation of heterogenous sequences is essential for the downstream sequencing process. In this study, the 16S rRNA gene sequences generated provides a mean of identifying strains and elucidating phylogenetic relationships between species by comparing sequence variations.

1.4.3 Restriction Fragment Length Polymorphism (RFLP)

One method to visualise the phylotypes within a clone library or the genetic diversity between clones is to use restriction fragment length polymorphism (RFLP) (Sambrook and Russell 2001). RFLP uses restriction endonucleases to cleave a DNA fragment into various smaller fragments with lengths that are characteristic of the original DNA sequence. By using several RFLP analyses with different restriction enzymes, the collection of banding patterns formed by a DNA sequence is statistically unlikely to be recreated from a different DNA sequence (Sambrook and Russell 2001). In this study, RFLP identified clones with

similar inserts and was also used for quick comparison of an isolate's purity and identity with the RFLP profile of a known culture.

1.4.4 Disadvantages of PCR-based methods

Although PCR-based methods can help microbiology overcome the limitations of traditional culture-based approaches (as detailed in previous sections), PCR-based approaches have their own limitations. In a multiple template PCR, such as amplifying 16S rRNA gene sequences from environmental samples, PCR artefacts or biased amplifications can create challenges for environmental 16S rRNA gene surveys.

PCR bias is mainly due to the different binding affinity of primers to various sequences due to the difference in their binding energy (Kanagawa 2003). Therefore, the diversity shown by the PCR product may not reflect the actual diversity or population composition in the environment (Polz and Cavanaugh 1998).

PCR artefacts can be the result of polymerase error, chimeric sequences, and heteroduplex molecule formation (Acinas et al. 2005; Kanagawa 2003; von Wintzingerode, Göbel, and Stackebrandt 1997). Polymerases have different error rates depending on the type used. The most commonly used Taq polymerase, for example, has an error rate of 1 nucleotide for 9000 polymerised. Polymerase with higher fidelity due to proofreading ability, such as Pfu polymerase has much lower error rates. The error rate of Pfu is around 1 error in every 1.3 million nucleotides polymerised (Tindall and Kunkel 1988).

During the annealing stage of a PCR reaction, heterogeneous single stranded DNA, including primers, can inadvertently join together. Since the sequences are not completely complementary, loops may form, resulting in a molecule described as a heteroduplex. A heteroduplex may generate false sequence diversity after going through the repair process in the competent cells to generate a new linear DNA (Kanagawa, 2003).

Since 16S rRNA gene sequences are relatively short (~1.5 kb), polymerase error is a minor problem, which can also be solved through repeated sequencing. Chimeras and heteroduplex sequences require post-sequencing analyses for repeated segments within the sequence or a combination of non-chimeric sequences within a PCR process (DeSantis et al. 2006; Huber, Faulkner, and Hugenholtz 2004).

Part of this study aims to address issues related with PCR bias, especially those related to 16S rRNA universal primers, by designing specific primers that target a taxonomic group that usually makes up the minority in clone libraries (see section 1.7), thus increase the survey depth of the group of interest.

1.5 Emerging approaches

In this section, a brief description of some relevant current or emerging approaches in investigating microbial diversity is presented. These approaches are chosen for their relevance to this study, in the sense that many expand upon the techniques used in this study. Therefore, this section is further divided into sub-sections akin to those used for the methods chapter; culture-based approach, microscopy, and culture-independent approaches.

1.5.1 Culture-based approach

As mentioned in section 1.3, traditional culture-dependent methods are limited due to their inability to efficiently reproduce various growth environments (including microenvironments that bacterial communities can form). Despite this, culture-based methods still provide valuable information required to effectively characterise bacterial species. One current approach addresses this issue by simulating the complex natural environment by the use of diffusion chambers. These allow chemicals from the original environment to diffuse through membranes, while trapping the culture within a confined space and preventing contamination (Kaeberlein, Lewis, and Epstein 2002). Automated, high-throughput methods have been developed to enhance traditionally laborious cultivation techniques. For example, gel microdroplets and flow cytometry are used to parallelise and automate the inoculation and colony detection process (Zengler et al. 2002)

1.5.2 Microscopy

Microscopy-based techniques such as microinjectors and micromanipulators are shown to be effective in isolating single cells in a range of environments (Ishøy et al. 2006). Optical (laser) tweezers offer even higher degrees of precision in manipulating single cells for producing pure cultures or single cell PCR (Fröhlich and König 2000). These methods aim to reduce the bias of selective media by directly targeting observable cells.

Fluorescent *in-situ* hybridisation (FISH) was developed in the 1960s; therefore, strictly speaking, it is not a new technology. However, with the recent rise of sequencing and bioinformatic capability, the use of FISH has become a popular technique in many microbiological disciplines (Levsky and Singer 2003). Primers and sequence databases developed for ribosomal sequence surveys can be adapted for FISH use. FISH can enhance microscopy, with its dependence upon pure or enriched culture, by labelling bacteria within the target phylogenetic clade with specific fluorochrome-oligonucleotide probes. FISH can also visualise the distribution of targeted species within microbial communities such as biofilms, an important feature for ecological studies, and provide insights in bacterial interactions.

1.5.3 Culture-independent approach

As mentioned in the introduction, published species names are a good indicator of the number of known bacterial species. As shown in Figure 1, the number is increasing at an accelerating rate. The increase has been attributed to the rise in use of molecular methods such DNA-DNA hybridisation, GC content, fatty acid profilling and more recently, 16S rRNA gene sequence analysis (Hugenholtz and Pace 1996). The growth of recognised bacterial species has lead to the recent trend in high-throughput sequencing in large-scale survey projects. These new technologies combine massive parallelisation, automation, and non-Sanger sequencing-based methods such as pyrosequencing (sequencing by synthesis) and sequencing by ligation. High-throughput sequencing has the advantage of producing a very large amount of sequencing data at a lower cost than Sanger sequencing in a shorter period of time, making even full-genome sequencing possible as a commercial service. The disadvantages compared to traditional Sanger sequencing are higher setup cost and shorter reads per reaction, which makes *de novo* whole genome sequencing difficult. Improved techniques that increase read length for *de novo* sequencing incur additional costs. Although, it has been argued that short read lengths are sufficient for sequencing 16S rRNA genes, the setup cost is high for each pyrosequencing run due to the scale of the sequencing. Thus, traditional Sanger sequencing is still the preferred method for small scale 16S rRNA gene surveys, until the rapidly evolving and fiercely competitive high-throughput sequencing industry make further advancements.

With the development of new sequencing technologies, further exploration into the genetic information within a microbial ecosystem is becoming more viable. The emerging field of metagenomics studies the DNA sequences collected from the environment. Metagenomics can be divided into two different approaches, sequence-based and functional based. Sequence-based metagenomics can be viewed as an extension of the 16S rRNA gene surveys that have become a common practice; it moves forward from the identification and phylogeny of 16S rRNA gene sequence analysis to attempt reconstruction of individual genomes within an environment. This information will not only help us understand novel uncultured species at a whole new level, but also various intra- and interspecific interactions (Tyson et al. 2004).

As sequencing technologies have not yet fully reached the power and economy required to sequence all but some of the most simple bacterial communities to a sufficient depth to construct the genomes of the constituent species (Tyson et al., 2004), sequence-based metagenomic projects such as the Global Ocean Sampling Expedition have been targeting functional genes rather than reconstructing entire genomes. The pilot expedition alone discovered 148 novel bacterial phylotypes and 1.2 million novel genes (Venter et al. 2004).

In contrast to sequence-based metagenomics, functional-based metagenomics aim to explore the genetic expression with environmental DNA samples. The term “functional metagenomics” may cause confusion, as it has been used differently among researchers (Dinsdale et al. 2008; Handelsman 2004). Instead of sequencing the DNA fragments extracted from the environment, function-based metagenomics aims to discover desirable genes by expressing the gene-containing DNA fragments in transformed (Handelsman 2004). This is of particular interest for pharmaceutical and biotechnology industries as a way to survey bacterial metabolisms for molecules such as antibiotics.

1.6 Extremophiles

Despite their limited observable morphologies, bacteria are very genetically and physiologically diverse. Genetic analysis of known organisms has shown that bacteria compose the majority of branches of the evolutionary tree (Ciccarelli et al. 2006), and they are found in the most diverse and extreme environments on Earth due to a wide repertoire of physiological characteristics (Brock 1985).

Extreme environments are of particular interest to microbiologists for many reasons.

1. The environments demonstrate the ability or limit of evolution on organisms to adapt to an environment not favourable for life (Brock 1985).
2. The geographical barriers between geothermal systems provide a natural testing ground for biogeographic hypothesis on the dispersion of microbial species (Soo et al. 2009).
3. The isolation and diversity of many geothermal systems also provide environments for traits not found in other environments to evolve. Enzymes from extremophiles are more stable in various harsh environments, and since many have high thermostability, biochemical reactions at higher rates can be achieved without enzyme degradation (Morozkina et al. 2010).
4. The compositions of some geothermal systems resemble the hypothesised early conditions on Earth where abiogenesis may have occurred (Pollack, Figueroa, and Zhao 2009). Thus, it has been argued that some extremophiles may retain characteristics of these early life-forms. Indeed, many extreme environments such as underwater hydrothermal vents are noted for their richness in species that belong to deeply-rooted phylogenetic (slowest evolving) groups (Pollack, Figueroa, and Zhao 2009).

Yellowstone National Park (Wyoming, USA) was the home of much early extremophile research, which demonstrated the novel bacterial and archaeal diversity in a geothermal environment and established many common practices in molecular microbiology. This research includes the discovery of an extreme thermophile *Thermus aquaticus*, which revolutionised molecular biology (Saiki et al., 1988), the distinction between archaea and eubacteria, the identification of the archaeal domain (Woese and Fox 1977), and contributing to the development of small subunit (SSU) rRNA gene sequence phylogenetic analysis (Pace 1997; Barns et al. 1994).

1.7 Candidate Division OP10

Candidate Division OP10 was originally identified from 16S rRNA gene sequences extracted from Obsidian Pool in Yellowstone National Park. It is one of twelve previously unrecognised divisions detected in that environment (Hugenholtz et al. 1998). These putative phyla, without pure cultured type strains available to represent the taxa, are termed "candidate divisions". Even though the currently available Bacteriological Code (revision 1990) does not cover taxonomic hierarchy higher than class, it is commonly recognised as the highest mutually exclusive taxonomic unit for the bacterial and archaeal domain, based on two or more consistently monophyletic 16S rRNA gene sequences (Hugenholtz, Goebel, and Pace 1998). There are currently 28 commonly recognised bacterial and archaeal phyla in publications (http://www.bacterio.cict.fr/classifphyla.html), but many more candidate divisions exist. The phylogenetic relationships between common phyla and candidate division OP10 is shown here in Figure 2 and Figure 3.

A survey of recently published literature and the small subunit (SSU) rRNA database SILVA showed that 16S rRNA gene sequences belonging to OP10 are found in mesophilic soil (Chow et al. 2002; Lesaulnier and Papamichail 2008; Lipson and Schmidt 2004), ultraoligotrophic to hypereutrophic freshwater environments (Chow et al. 2002; Urbach et al. 2001; Wu et al. 2007), as well as extremophilic environments such as hypersaline mats (Ley et al., 2006), geothermal springs (Hugenholtz et al. 1998), geothermal soils (Portillo and Gonzalez 2008; Stott et al. 2008), contaminated environments (Dojka et al. 1998; Lipson and Schmidt 2004)and Antarctic environments (Smith et al. 2006; Soo et al. 2009). OP10 appeared to exist as a minor constituent in these environments (Hugenholtz, Pitulle, Hershberger, & Pace, 1998), although in several cases in geothermal systems (see Stott et al., 2008; Portillo & Gonzalez, 2009) OP10 strains have been detected at concentrations as high as 10% of the total clone bank populations. Many more OP10 16S rRNA gene sequences are available in databases but unpublished in journals, so understanding diversity and distribution from these data is problematic.

The lack of a dedicated publication of an OP10 species, despite many 16S rRNA gene sequences being available from various surveys, was due to the previous inability to isolate and culture OP10 species. However, descriptions of the environmental samples can often be

found along with the sequences. For example, the biggest series of submissions for soil samples came from an unpublished study, “Phylogenetic diversity and evolutionary relationships between rare and abundant members of the bacterial community in tallgrass prairie soil” by the DOE Joint Genome Institute. The sequences fall into clusters, but they are found throughout the OP10 phylogenetic tree (as seen in Appendix 3). Based on the geographical data retrieved from Greengenes 16S rRNA gene database, OP10 sequences seemed to be more abundant in wet soil high in organic plant matter (e.g. prairie, meadow, and forest wetland). The common theme of water abundance is present in other categories as well. The temperature range of OP10 varied from possibly psychrophilic (i.e. wet tundra meadow soil) to thermophilic in environments such as Mammoth Hot Spring and Obsidian Pool in Yellowstone National Park.

Combining the information from that attached to the 16S rRNA gene sequences and the data in published journals. OP10 appears to be detected in a wide temperature range, with high water availability, and often found thriving in microbial communities such as microbial mats in lakes as well as in activated sludge in various bioreactors (Crocetti et al. 2000).

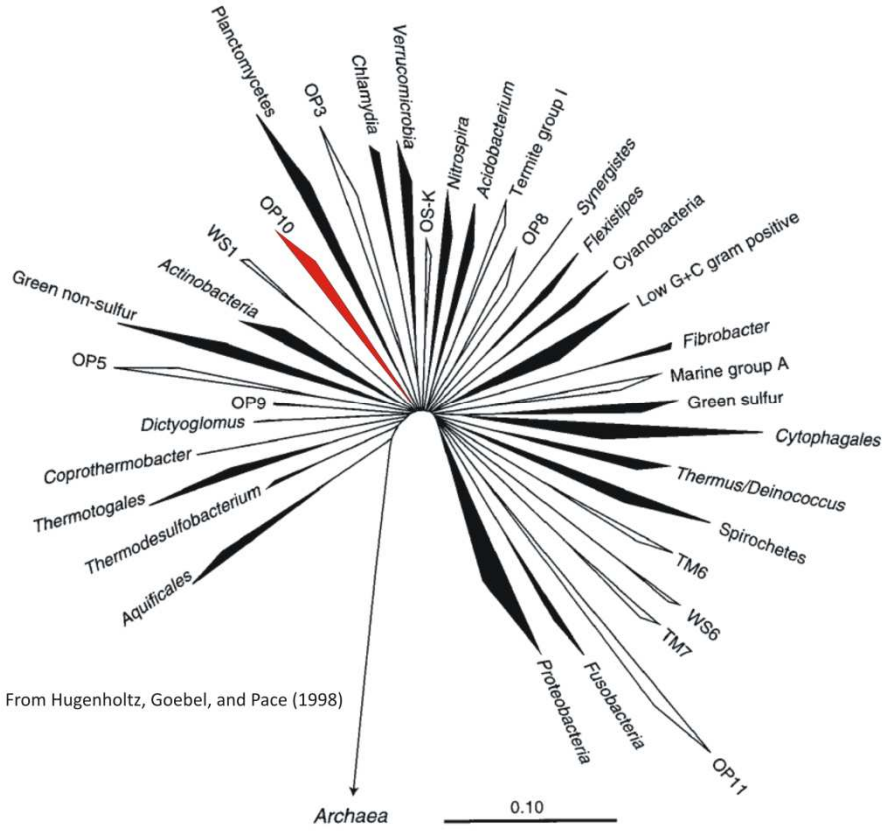
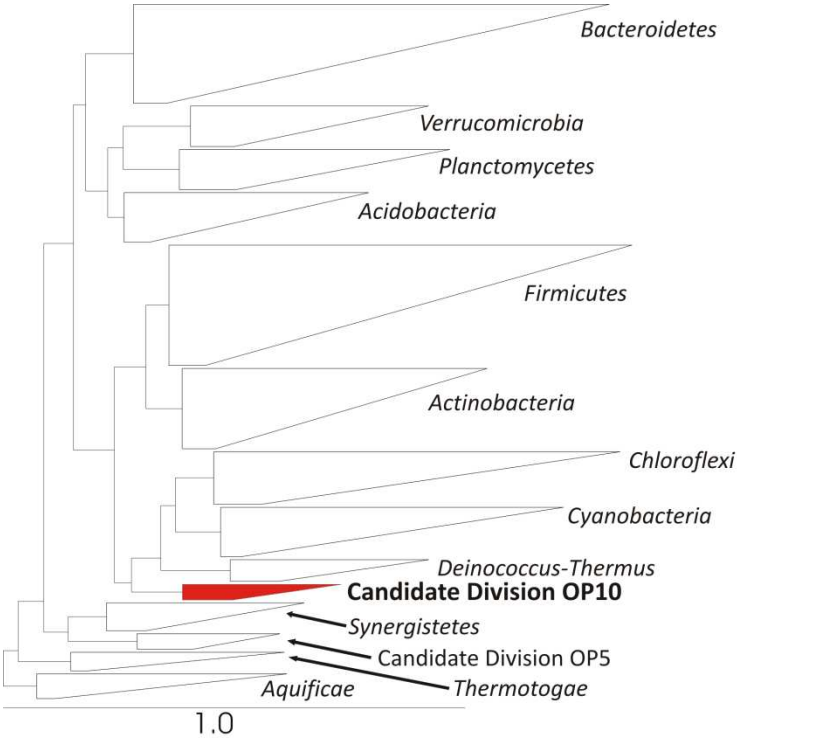


Figure 2 - Phylogenetic tree of selected bacterial phyla



Modified from SILVA release 100 SSUref guide tree

Figure 3 - Phyla closely related to Candidate Division OP10

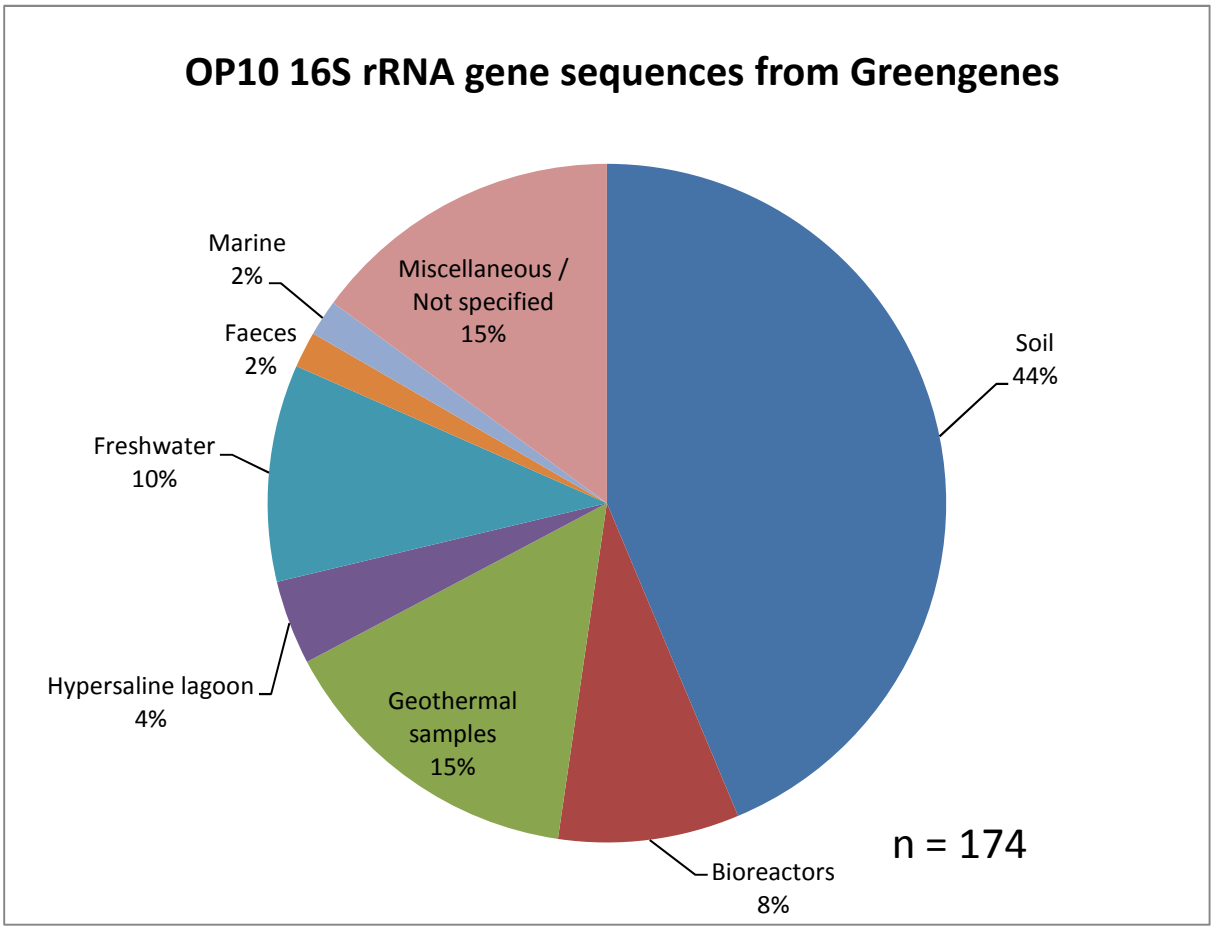


Figure 4 – The environmental distribution of OP10 16S rRNA gene sequences from the Greengenes database

The first cultivated OP10 strain, P488, was isolated from a geothermal soil sample from Waikite in the Taupo Volcanic Zone (TVZ), New Zealand (Stott et al. 2008). An additional strain, T49, has an almost-identical 16S rRNA gene sequence and was isolated from geothermal soil in Tikitere, TVZ, using the same weakly acidic minimum nutrient medium (Stott et al. 2008). The minimum nutrient media was noted for its use of gellan as a solidifying agent and growth substrate (Stott et al. 2008). Minimum nutrient media played an important part in culturing extremophiles, as it was noted that a high concentration of organic nutrients may have an inhibitory effect on organisms adapted to minimum nutrient environments (Brock and Freeze 1969).

1.8 Taupo Volcanic Zone

The Taupo Volcanic Zone is a large belt of active geothermal sites in the North Island of New Zealand, generally considered to encompass the area from Mount Ruapehu inland to White Island at the sea. It is around 300 km long and 60 km wide (Wilson et al., 1995) and includes Lake Taupo, which is a caldera formed through a large eruption 26,500 years ago. The zone is geologically active, with the Oruanui eruption noted for being the most recent eruption reaching the top (VEI 8) of the Volcanic Explosivity Index. The zone also includes submerged vents and cone volcanoes such as Ruapehu, Tongariro, and Ngauruhoe. The size, activity and diversity of activities within the TVZ make it one of the major geologically active areas in the world. All samples in this study came from within the Taupo volcanic Zone. The sites listed below are geothermal features included in this study, their locations are shown in Figure 5.

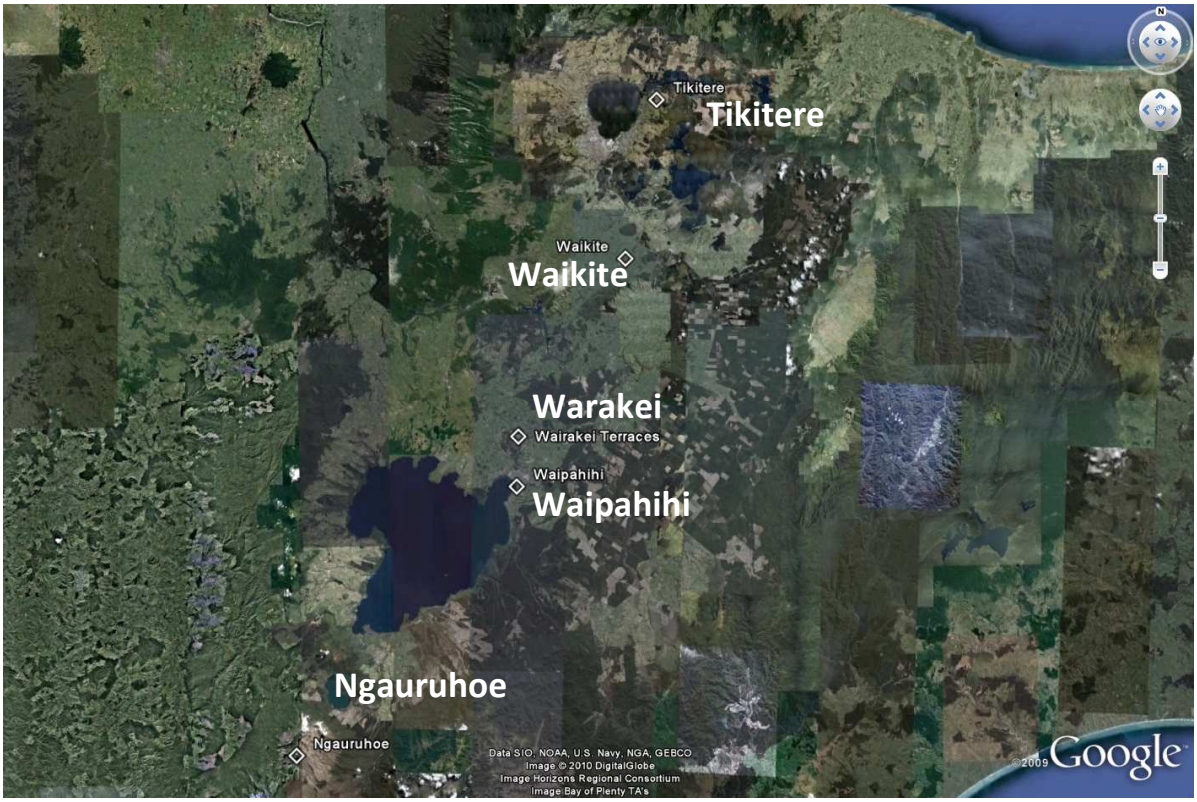


Figure 5 - The locations of the sample sites used in this study

Waikite

Waikite is located to the south of Rotorua, and is part of the Waikite-Waiotapu-Waimangu geothermal system. The spring water was measured to be close to boiling (>90 °C), weakly alkaline (pH 7.2–8.9), high in silica, but low in calcium and magnesium (Jones, Renaut, &

Rosen, 1996). Despite the high temperature, microbial slime/bioflim was clearly noticeable, suggesting an active population of thermophilic microorganisms (Jones, Renaut, and Rosen 1996).

Tikitere

Tikitere is located to the east of Lake Rotorua, approximately 16 km away from Rotorua City. The site is a geothermal tourist destination. Surface geothermal fluid has been attributed to the results of surface steam condensate caused by geothermal heating, as opposed to deep spring water (Campbell et al. 2001). Along with the steam is an abundance of hydrogen sulphide, which undergoes transformation through various oxidative states in the environment. Yellow elemental sulphur deposits were observable, and the distinctive odour of hydrogen sulphide can easily be detected. The abundance of sulphur compounds in the environment make Tikitere a potential site for sulphur biogeochemical cycle studies.

Mount Ngauruhoe

Mount Ngauruhoe, situated at the southern side of the Tongariro volcanic complex, is 2,291 m in elevation (Simkin and Siebert 1981). It is the youngest and the most active volcanic vent of the Tongariro complex, and most recently erupted in 1975 (GeoNet 2010). As of 2009, the cone is composed of two craters, with the older, larger crater surrounding the newer inner crater. The profile of the volcanic cone is round, steep and symmetrical; this was attributed to the effect of larva and pyroclastic flow on the structure. Fumaroles are observable in the craters, emitting steam and various gases.

Wairakei Geothermal Field

The Wairakei-Tauhara system is the site of a geothermal power station adjacent to the Waikato river. The power station was one of the first geothermal power stations in the world, and its continuous usage has impacted the underground geothermal system and affected the landscape (Bixley, Clotworthy, and Mannington 2009), providing a well-documented area of human impact on extremophilic environments. The artificial terraces formed at the wastewater outlet take on various colours, including pink, blue, and green, suggesting the existence of a range of microbial life-forms in this geologically young environment.

1.9 Aims, Approaches

As discussed in the literature review, bacteria play an important role in biological processes on Earth; they represent a large percentage of global biomass, as well as genetic, metabolic, and biochemical diversity. Our understanding of the bacterial domain, however, is still at an early stage compared to disciplines such as zoology and botany, as the nature of microbes hinders many research efforts. All indicators have shown that the vast majority of all bacterial species on Earth have not been studied. The development of molecular biology and the resulting techniques have finally allowed microbiologists to address these problems effectively.

Candidate division OP10 was first identified in 1998 at Yellowstone National Park and detected in a range of environments (refer to section 1.7 on OP10). It was not successfully cultured until recently, when Stott et al. (2008) isolated several pure cultures from the Taupo Volcanic Zone.

The first aim of this study was to describe the physiological characteristics of one of the first cultured OP10 isolates, strain T49. These data will be used to expand the understanding of a so far uncharacterised bacterial phylum. In doing so, it is anticipated that in this process we will generate enough data to validly publish this isolate as a type strain for this phylum. This goal will be achieved with primarily culture based methods. Specifically;

1. Determining various environmental parameters for optimal growth for the strain, such as temperature and acidity;
2. Biochemical profiles of the bacterium, such as GC ratio and fatty acid profile; and
3. Direct observation of the pure colonies and microscopic observations of the bacterium under phase contrast and electron microscopy (SEM & TEM) for morphological characteristics.

The second aim of this study was to develop a molecular approach to assist in the detection of currently uncultured OP10 species and the isolation of these novel species – in particular, to develop OP10-specific PCR primers. A culture-independent approach would complement traditional culture-dependent methods and permit surveys of environments with greater

bias and wider coverage towards OP10 phylotypes. This approach circumnavigates the need to conduct deep clone libraries using universal bacterial primers when searching for OP10 phylotypes.

The molecular approach uses oligonucleotide PCR probes/primers specific to 16S rRNA gene sequences within candidate division OP10 to amplify targeted 16S rRNA genes in environmental samples. The set of primers will be based on known 16S rRNA gene sequences publicly available in databases, and on sequences from previous 16S rRNA gene surveys of environmental samples at GNS Science. The primers will then be tested *in silico* and *in situ* against known OP10 sequences and known non-OP10 sequences. Primers will be tested by PCR and PCR parameters optimised to avoid mismatching from occurring. This protocol will serve as the primary means of detecting OP10 in the environment using the specific primers. PCR kinetics will be optimised with known environmental samples to enable reliable determinations. The goal of the primers is not to be specific to the clades within OP10, but to provide maximum coverage of OP10 as a division together, while avoiding false positives with non-OP10 groups.

2 Materials and Methods

2.1 General

Bacterial isolates

The candidate division OP10 isolate used in this study, strain T49 (to be referred to this point forward as T49) was first isolated from a geothermal soil sample from Tikitere (Stott et al., 2008). This strain was the main focus of the characterisation effort in this study. It was maintained weekly through subculturing on plates of AOM1 media; these were used as the source of all other inoculation experiments. Unless otherwise stated, inoculations were performed within a Heraeus HS12 Biological Safety Cabinet.

Reagents and Chemicals

Reagents and chemicals used in this study are listed in the Appendix 1 section along with the suppliers and, grades, and product numbers where appropriate.

2.2 Culture-based characterisation of T49

2.2.1 Components for culture media

As a general rule, in order to avoid unintended precipitations from occurring, the solutions were made by dissolving the ingredients in water then brought up to the desired volume. All media was autoclaved at 121 °C and 15 lb/in^2.

Ferric Ethylenediaminetetraacetic Acid (FeEDTA) solution

Ingredients	Amount per L
$FeSO_4 \cdot 7H_2O$	1.54 g
Sodium EDTA	2.06 g

The solution was then autoclaved and refrigerated away from light as the solution may be photosensitive (Albano & Miller, 2001).

Wolin methanogen trace elements solution

The trace element solution was modified from (Wolin, Wolin, & Wolfe, 1963) and was composed of the following (per L):

Ingredients	Amount per L
Nitrilotriacetic acid	1.500 g
$Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$	0.200 g
Na_2SeO_3	0.200 g
$CoCl_2 \cdot 6H_2O$	0.100 g
$MnSO_4 \cdot 2H_2O$	0.100 g
$Na_2MoO_4 \cdot 2H_2O$	0.100 g
$Na\ WO_4 \cdot 2H_2O$	0.100 g
$ZnSO_4 \cdot 7H_2O$	0.100 g
$AlCl_3 \cdot 6H_2O$	0.040 g
$NiCl_2 \cdot 6H_2O$	0.025 g
H_3BO_3	0.010 g
$CuSO_4 \cdot 5H_2O$	0.010 g

Nitrilotriacetic acid was first dissolved in 800 mL of DI (De-Ionised) water. The pH was adjusted to 6.5 with KOH before adding the remaining ingredients. The final pH was adjusted to 7.0 with HCl, and the final volume brought to 1 L. The solution was then autoclaved and refrigerated.

Methanotroph trace elements solution

Ingredients	Amount per L
$ZnSO_4.7H_2O$	0.44 g
$Na_2MoO_4 \cdot 2\ H_2O$	0.60 g
$MnCl \cdot 4H_2O$	0.19 g
$CuSO_4 \cdot 5\ H_2O$	0.06 g
H_3BO_3	0.10 g
$CoCl_2 \cdot 6H_2O$	0.08 g

The ingredients were dissolved in approximately 900 mL of DI water, and then made up to 1 L. The solution was then autoclaved and stored refrigerated.

Citric acid/ disodium hydrogen phosphate (Na_2HPO_4) pH buffer

Na_2HPO_4 (0.2 M) and citric acid (0.17 M) stock solutions were made and autoclaved to prevent microbial growth. Equal amounts of the two solutions were mixed together then adjusted to the target pH (refer to the section below for the target pH) with Na_2HPO_4 and citric acid. This master mix was then autoclaved again for long-term storage.

2.2.2 Bacteriological media preparation

Plate media were all stored in 4 °C cool room, while sterile liquid media was kept at room temperature and sealed within either crimp-top serum bottles or Schott® bottles with butyl septum (stopper) seals (Figure 6).

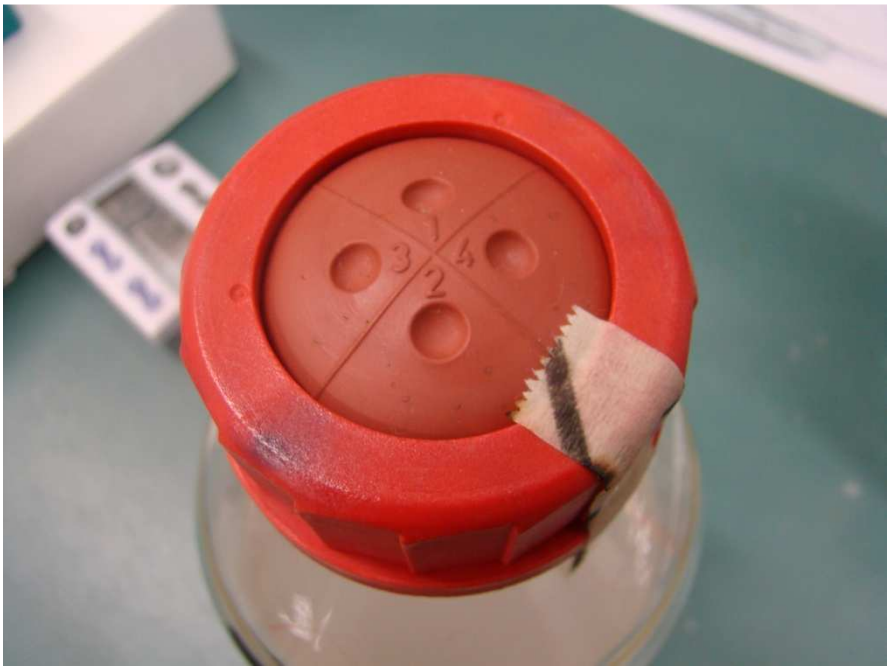


Figure 6 - A Schott® bottle with a butyl septum

AOM1 plate medium

Ingredients	**Amount per L**
FeEDTA	1.000 mL
$MgSO_4 \cdot 7H_2O$	0.656 g
K_2HPO_4	0.050 g
$(NH_4)_2SO_4$	4.000 g
$NaHCO_3$	0.050 g
$CaCl_2 \cdot 2H_2O$	0.005 g
Methanogens Trace Metals solution	1.000 mL
Phytagel™	15.000 g

The above ingredients were dissolved in 500 mL of DI water. The pH was then adjusted to 6.5 using diluted H_2SO_4**.** This made up Component A of the final medium.

Next, 15 g of Phytagel™ was rapidly added to a 1 L Schott® bottle containing 485 mL DI water while stirring with a magnetic bar. Phytagel™ acted as the gelling agent as well as the energy source for the culture. This was Component B of the final medium.

The two components were autoclaved separately, Component A was aseptically mixed with component B (using the stir bar in Component B) while the temperature was still high. When the solutions cooled down to 85 ºC, 100 mg of yeast extract and 100 mg of B Forte vitamins (Appendix 1) dissolved in 10 mL of water, were aseptically added to the warm medium through a 0.22 µm sterile filter. The medium was then poured into sterile Petri dishes to set. The plates were left overnight at room temperature with lids on to reduce the amount of condensation.

Liquid medium (4.5NZS10.2)

Ingredients	**Amount per L**
NH_4Cl	0.40 g
KH_2PO_4	0.05 g
$MgSO_4 \cdot 7H_2O$	0.02 g
$CaCl_2 \cdot 6H_2O$	0.01 g
FeEDTA solution	3 mL
Methanotroph trace elements solution	3 mL
Wolin methanogen trace elements solution	1 mL

The above ingredients were dissolved in DI water and adjusted pH to 4.3 with 1 M H_2SO_4. Phytagel™ (1 g) was added and mixed thoroughly by shaking the bottle vigorously. The mixture was then autoclaved. After autoclaving, 100 mg of yeast extract and 100 mg of B Forte Vitamins were mixed in 10 mL of DI water and injected into the autoclaved media through a 0.22 µm sterile filter.

Liquid medium for evaluating pH range of T49

This medium was based on the 4.5NZS10.2 liquid medium protocol, but KH_2PO_4 was omitted as it is known to influence pH. The buffer master mix was made at the desired pH as per buffer protocol in the components section 2.3.1. The master mix was then diluted with deionised water to make up a 1/8 strength solution. The solution was used in the place of water to make 4.5NZS10.2 media with 100 mg/L of vitamins and yeast extract (each) added. The pH of the final products (after autoclaving) is generally 0.2 to 0.3 higher than the stock; thus, it was important to re-measure the final pH after preparation.

Methanotroph medium

4.5NZS10.2 was made as described above, but Phytagel™ was omitted. Gases (methane and/or CO_2 at 10% (v/v) and 2% (v/v) over pressure respectively) were injected into the bottles through a 0.22 µm sterile filter.

2.3 Sample collection for culture-based and culture-independent experiments

Environmental samples used for culture-dependent and culture-independent analysis in this study were collected both during this MSc (Figures 7-10) and during previous sampling expeditions of the Extremophiles Research Group at GNS Science (Table 1). Environmental DNA from all samples was extracted according to the protocol outlined in section 2.6.1.

Environmental samples were aseptically collected with a sterile spatula and placed in sterile 50 mL Falcon™ centrifuge tubes. Temperatures of the sample sites were measured on-site with a Fluke 50S thermocouple, and sample pHs were measured in the laboratory at 25 °C. Soil sample pH was measured by mixing the 1 g of sample in 10 mL of DI water.

Table 1 - Sample site locations and descriptions.

Location	Code	Description	Temperature/pH	GPS – NZTM Projection
Taupo waste water treatment plant *	TWW01	Primary setting tank	Ambient, pH 7.2	N5714893, E1867933
Taupo waste water treatment plant *	TWW02	End of trickling filter flow	Ambient, pH 7.2	N5714893, E1867933
Tikitere *	TT13.3	Steaming soil with dead plant on the surface. Grey-white soil (4cm deep). Possible AH horizon.	41 °C pH 3.8	N5782155, E1894627
Tikitere *	TT13.4	Steaming soil with dead plants on the surface. Clay-like orange/ copper coloured, often fused blocks (5-10 cm deep)	52.5 °C pH 4.3	N5782155, E1894627
Mt. Ngauruhoe	NGH02	Diffused steaming soil, covered in moss, located on	68.6 °C pH 6.88	N5662800, E2827349

		the outer rim, near fumaroles.		
Wairakei (Claret Cup)	WRG1	Steaming red soil in dry mud pot	46.9 °C pH 4.8	N5721049, E1866890
Waipahihi lakefront *	WPH03	Water and fine gravel at spring outflow (in between split rock on foreshore)	63 °C pH6.5	N6273383, E2778363
Waikite cascade pool *	WKT31	3.5 metres from outlet, subaqueous gelatinous orange-red layer with outer green photosynthetic layer	~70°C pH n/a	N5752733 E1888806
Waikite Hot Spring *	WKT32	Steam-affected cliff wall just above the spring. Pink soil sample was covered by a layer of green photosynthetic mat	66.5 °C pH 4.5	N5752706, E1888857
Waikite Hot Spring *	WKT33	Water sediment from small creek with sinter edges.	77.6 °C	N5752696, E1888873

*; samples collected by author for this study. All other samples were collected previously and stored at -80 C prior to analysis.

Figure 7 - Waikite Hot Spring, pink soil sample was collected from the steam affected cliff wall just above the spring. The pink soil sample was covered by a layer of green photosynthetic mat.

Figure 8 - Taupo lakefront at Waipahihi beach.

Figure 9 - Taupo wastewater treatment plant, trickling filter flow.

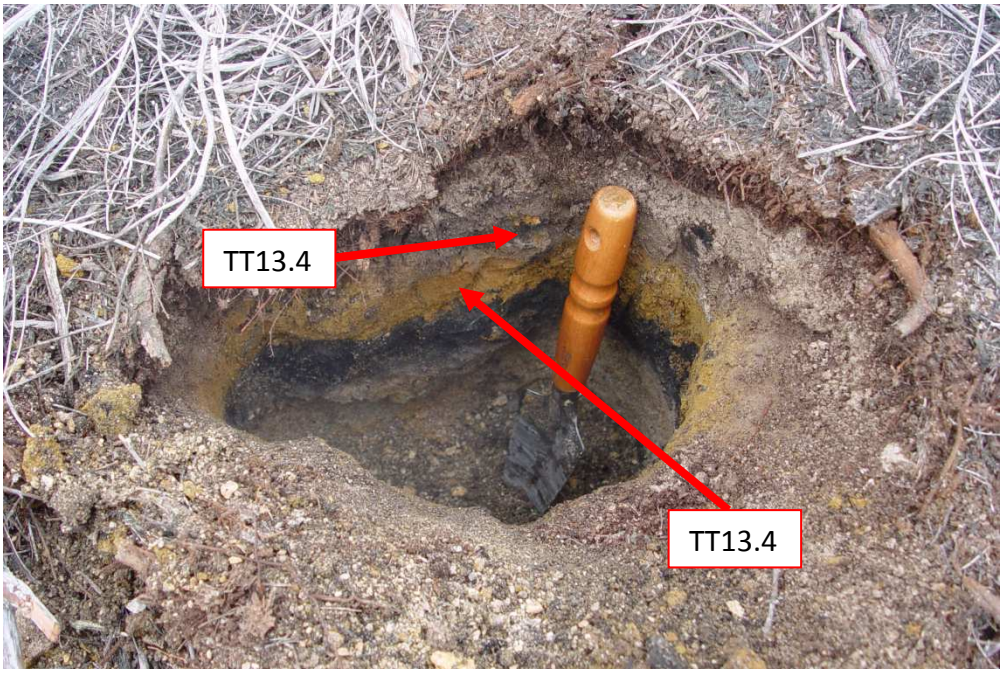


Figure 10 - Tikitere, stratified soil profile of steam affected soil. The soil sample used for inoculation was collected from the orange layer. Soil temperature was measured *in situ* at 52 °C.

2.4 Isolation and enrichment protocols

Environmental samples were cultured on AOM1 plates the same day they were collected to avoid changes in the microbial populations. The liquid phases of the samples were pipette onto AOM1 plates and streaked out using a sterile loop. The solid parts of the sample (i.e. sand, soil, and microbial mat) were placed onto plates in small amounts using a sterile spatula and also streaked out. Both standard-pH and low-pH (section 2.5.2) AOM1 plates were used in order to investigate the enrichment effect of lower pH on these microbial populations. The plates were incubated at 60 °C in an aerobic environment in anaerobic jars to prevent desiccation.

Colony formation was monitored daily under a CETI stereo microscope to identify colonies with different morphology, which were picked with a sterile toothpick and further subcultured via streaking on AOM1 plate. This isolation process was repeated until only a single type of colony morphology existed per plate. After the consistency of the colonies was further confirmed by phase contrast microscopy, colonies were picked by sterile toothpicks and inoculated into PCR mixtures for direct 16S rRNA gene sequencing. Where no

positive amplification was noted, colony DNA was extracted using Fungal/Bacterial DNA Extraction Kit (Zymo Research)(see section 2.6.1). Good sequence quality suggests colony purity, while poor quality may be a combination of several factors, including non-pure cultures.

2.5 T49 characterisation

2.5.1 Maintaining T49 stock on media plates

To maintain a culture of isolate T49 on media plates, weekly subculturing was necessary as colony viability and medium condition would degrade rapidly after that period. A minimum of two standard AOM1 plates were subcultured from one of the previous two plates. The older plate which was not used for subculturing was kept as a backup, thus maintaining a steady rotation of three plates. Additional plates were subcultured when experiments required additional T49 biomass.

The optimal period to harvest T49 cells from media plates was around 3-4 days after subculturing, before the pink/red colour of the colonies fully developed. Liquid cultures were more sensitive to this restriction, and a sufficient amount of inoculum was required (much more than using a single colony, around a whole cultured plate per 500 mL liquid culture); otherwise the culture would fail to initiate growth. This may be due to bacterial quorum sensing or other unknown mechanisms.

The primary objective of the inoculation was not to ensure that all the bottles received exactly the same amount of inoculum but rather to insure that all the bottles received sufficient inoculum, as the analysis focuses on the growth rate at log phase rather than on the starting time of particular phase. Although it was possible to inoculate from liquid culture to fresh liquid media or liquid culture to plate media, this was generally avoided during the study because liquid culture was more unstable for maintaining viable cells, and contamination was harder to identify. To avoid contamination, all work surfaces were wiped with 70% (v/v) ethanol, and T49 was always handling within a Heraeus HS12 Class 2 Laminar Flow Cabinet. Septa on liquid culture bottles were sterilised with 70% (v/v) ethanol and flamed.

Negative pressure on occasion was noted within bottles of liquid media after autoclaving due to expanding gas escaping the seals. This negative pressure can restrict growth. To detect and eliminate pressure differences, the pressure within each of the bottle of medium was neutralised with a sterile needle to atmospheric pressure.

2.5.2 Inoculation and sampling of liquid media

4.5NZS10.2 (see 2.4.2) liquid medium was used for all T49 liquid cultures in this study.

When inoculating liquid media from plates for substrate-dependent experiments, extra care was required to prevent excessive Phytagel™ carry-through. For substrate utilisation experiments, T49 colonies were grown on AOM1 plates with lower pH (4.8) to produce firmer colonies that could be lifted with loops dampened with the liquid medium creating less Phytagel™ carry-through.

Colonies were removed from plates with a sterile metal loop pre-moistened with sterile liquid media, then placed in a tube containing 1.5 mL medium. This mixture was homogenised by passing through a 21 G needle as it was taken up by a syringe. A 27 G needle was then used to inoculate the bottles; the smaller diameter of the needle further broke up the colonies. Care was taken to ensure the even distribution of colony fragments between the triplicate bottles.

Schott® bottles were used for larger volume experiments and were kept in an orbital incubator and shaken at 180 RPM. Serum bottles (50 mL) were incubated without shaking and were used for substrate utilisation experiments. Irrespective of bottle size, all liquid cultures were made up with 50% medium, 50% aerobic headspace. Volume and incubation temperature depended on the experimental aims.

2.5.3 Bacterial turbidity measurements

Cell turbidity was measured as optical density at 600 nm on a PerkinElmer Lambda 35 UV/Vis Spectrometer. Liquid cultures were briefly left to cool prior to sampling to prevent evaporation. More than 2 mL of liquid sample was extracted per bottle and placed into cuvettes, and care was taken to ensure the homogeneity of the samplings. During measurement, each cuvette from each liquid culture was measured three times, with stirring or flicking to ensure the suspension of T49.

2.5.4 Cell counting

Cell counting was performed with a haemocytometer (Hawksley BS.748 Counting Chamber) under a phase contrast microscope (Olympus AX70). Sampling followed the same procedures as the turbidity measurements.

2.5.5 Substrate utilisation

Substrate utilisation tests were performed using 4.5NZS10.2 medium (section 2.2.2) in 125 mL serum bottles. Phytagel™ was replaced with other substrates at the concentration of 0.025 g/50 mL. The serum bottles were sealed with septa and incubated at 60 °C without shaking. Each substrate was tested in triplicate with two additional uninoculated bottles, a negative control and a post-preparation pH measurement controls. Growth was visually determined by formation of biomass and pellicle at the bottom of the bottles and compared with the negative control. Substrate utilisation was ranked according to the volume of biomass generated.

2.5.6 Temperature optimum and range determination

Triplicate 4.5NZS10.2 liquid medium was inoculated and incubated at the experimental temperature while being shaken in a Ratek Orbital Mixer-Incubator® at 180 RPM. Optimum temperature and range for T49 was determined using a range of temperatures (50 °C, 60 °C, 65 °C, 68 °C, 73 °C, and 77 °C) in triplicate with an additional negative control for each temperature.

2.5.7 pH optimum and range determination

The minimum nutrient medium (4.5NZS10.2) used for growing T49 in liquid culture had a poor pH buffering capability, which would be detrimental to an accurate pH optimum and range determination experiment. Therefore, a series of qualitative pre-experiments were set up to optimise the experiment.

Candidate pH buffers near the pH range at which T49 was known to grow (~pH 5) were selected for non-toxicity to T49. These included citric acid/Na_2HPO_4 buffer, citric acid/sodium citrate buffer, sodium acetate/acetic acid buffer, and Na_2HPO_4/NaH_2PO_4 buffer. Citric acid/Na_2HPO_4 buffer was found to be suitable for this purpose at a concentration of 1/10 dilution. The buffer was further tested to determine concentration to determine

optimal buffering capability. A 1/8 dilution was capable of buffering the pH without inhibiting bacterial growth.

Buffered 4.5NZS10.2 media was produced and inoculated as per standard liquid culture. The cultures were incubated at 60 °C and at 180 RPM, and the pH tested were 4.7, 4.9, 5.3, 5.8, and 6.2.

2.5.8 Methane utilisation

Methane liquid media was made (section 2.2.2) but without the addition of Phytagel™, and Schott® (1 L) bottles were used (as opposed to the typical 500 mL) for larger headspace. Carbon dioxide (a possible additional carbon source) was added to some bottles, Yeast extract and B vitamins were not added to some bottles to compare their effect, and in all cases growth was visually determined. Standard liquid cultures with Phytagel™ were used as a positive control. The combinations of all experiments set up can be found in Table 2.

Table 2 - Experimental setup for methane utilisation test

Tests	CH_4 (v/v)	CO_2 (v/v)	Yeast extract (100 mg/L)	B vitamins (100 mg/L)	Inoculated?	Phytagel™
Triplicate	5%	2%	-	-	+	-
Duplicate	5%	2%	-	+	+	-
Duplicate	-	-	+	+	+	-
	10%	2%	-	-	+	-
	-	-	+	+	+	-
	5%	2%	+	+	+	-
	5%	2%	+	-	+	-
	-	-	-	+	+	-
	5%	2%	+	+	+	-
Negative control	5%	2%	+	+	-	-
Negative control	-	-	+	+	-	-
Positive control	-	-	+	+	+	+

2.5.9 TEM/SEM microscopy

T49 was grown in 500 mL of liquid culture for 3 days, the culture was then centrifuged at 5000 g (4 °C) for 20 minutes. The pellet was washed three times with 0.1 M Sørensen's phosphate saline buffer then homogenised in temporary fixative and centrifuged again. Sample preparation and imaging was performed with Lloyd Donaldson at Scion.

For TEM washed T49 cells were embedded in 2% (w/v) agarose gel and fixed in 4% (v/v) glutaraldehyde in 0.1 M phosphate buffer for 1 hour at room temperature followed by post fixation in 1% OsO_4 for 1 hour. Before embedding the cells in Spurr resin, the agarose blocks were washed in phosphate buffer and then dehydrated through an acetone series. The embedding blocks were sectioned with a diamond knife (90 nm thickness), then transferred on a carbon/formvar support film to grids and stained with lead citrate (6 min) and uranyl acetate (12 min) before being examined on a JEOL 6700F field emission scanning electron microscope (FESEM) equipped with a STEM detector at 25 kV.

For SEM T49 cells suspended in water were placed in a droplet on a carbon dot and allowed to air dry. The specimen was then coated with 12 nm of chromium in a sputter coater before being examined on a JEOL 6700F FESEM.

2.5.10 GC content determination

Cells were grown and harvested then washed 3 times with DI water. The pellet was then lyophilised and analysis for GC content externally by the DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH) using a HPLC (high performance liquid chromatography) method.

Briefly, the method was as follows: Cells were disrupted with a French pressure cell and the DNA was purified with a hydroxyapatite column (Cashion et al. 1977). The purified DNA was hydrolysed and dephosphorylated using P1 nuclease and alkaline phosphatase respectively (Mesbah, Premachandran, and Whitman 1989). The GC content was determined by running the sample through C18 5 µm VYDAX 201SP54 analytical column in a 0.3 M $NH_4(H_2PO_4)$/acetonitrile 40:1 (v/v) running buffer (pH 4.4). The column temperature was set at 45 °C. A 10 µL samples was injected into the running buffer at 1.3 mL/min. The GC content was calculated by the ratio of deoxyguanosine (dG) and deoxythymidine (dT) via UV

spectrophotometry (Mesbah, Premachandran, & Whitman, 1989). The standards used for GC determination used the genomes of non-methylated Lambda-DNA (Sigma), *Xanthomonas campestris* pv. *Campestris*, DSM (3586^{T}), *Bacillus subtilis* (DSM 402), and *Streptomyces violaceoruber* (DSM 40783).

2.5.11 Fatty acid methyl esters (FAME) profile

Fatty acid analysis was performed under supervision at Industrial Research Limited (IRL), Lower Hutt. Total fatty acids were extracted in the form of fatty acid methyl ester (FAME) through a two-step extraction protocol (Svetashev V et al. 1995), as follows: T49 biomass (~2 mg – grown from 12x AOM1 plate cultures) + 250 µL basic solution (prepared by dissolving 1% (w/w) metallic sodium in methanol while the acidic solution consisted of 5% HCl in methanol (v/v)were added into a small screw cap vial and incubated at 100 °C for 15 minutes. An equal amount of acidic solution was then added, and the incubation process was repeated. Sodium methoxide in the solutions acted as a catalyst for methylation of fatty acids (free or bound) to form fatty acid methyl esters (FAMEs). To extract the FAMEs from the aqueous mixture, ~0.2 mL of hexane was added and mixed. This step was repeated once. The hexane phase was carefully extracted via pipette, and the hexane was evaporated under a constant stream of argon or within a weak vacuum, leaving the FAMEs in the container. The residue in the container was redissolved in ~10 µL of chloroform, and was ready to be injected into columns for analysis. Docosanoic acid (22:0 fatty acid) was added to act as an internal standard and positive control for the methylation process. This fatty acid is not generally found in bacteria; therefore, its addition does not contaminate the results.

For Hydrogenation of the FAMEs the chloroform which contained the FAMEs was evaporated under a stream of argon gas, and then redissolved in 200 µL of methanol. Platinum (IV) oxide catalyst (~1 mg) was added into the mixture. Hydrogen gas was bubbled into the mixture for 20 minutes, and the solution was filtered to remove the catalyst for reuse. Methanol in the solution was evaporated, and the saturated FAMEs residue was dissolved in 100 µL of chloroform.

FAMEs dimethyldisulphate (DMDS) derivatisation was carried out as follows: FAMEs were placed under a stream of argon gas to remove the solvent. 40 µL of DMDS with 0.53 mg

iodine mixture was added and incubated at 35 °C for 30 min. After the incubation, 100 mL of hexane/diethyl ether (1:1 (v/v))and then ~80 µL saturated $Na_2S_2O_3$ aqueous solution were added. Excess iodine was removed by pipetting the solution, leaving behind the solid iodines. The removal of iodine could be determined by the colour change. The upper hexane layer was diluted 6x before downstream analysis.

For FAMEs N, O-Bis(trimethylsilyl) trifluoroacetamide (BSTFA) derivitisation the chloroform was evaporated from the hydrogenated sample then 30 µL of BSTFA with 1% (w/v) trimethylchlorosilane (TMCS) was added and the mixture was incubated at 60 °C for 30 min. The resulting residue was used directly for injection.

To calculate the equivalent chain lengths (ECL) of the FAMEs, gas chromatography was run in isothermal mode at 125 °C. The FAME samples were either injected via an Agilent Technologies 7683 automatic liquid sampler, or manually injected with an SGE syringe, both at a volume of 0.5 µL with a split ratio of 1:60 into an Agilent 6890N Gas Chromatography system with an Agilent 5973 Mass Selective Detector.

The GC system used a different temperature programme for downstream application with mass spectometry (MS) than for GC alone. The variable temperature gradient programme was used to provide better resolution during gas chromatography. However, results from this protocol were unsuitable for ECL analysis. The temperature programme was as follows: First, hold the temperature at 100 °C for 6 min. Second, increase the temperature to 160 °C at a rate of 5 °C/min. Finally, increase to 240 °C at a rate of 1 °C/min and hold at that temperature for 5 min. The machine was set up for split-less mode; thus, some samples were diluted to prevent the column from being saturated. Gas chromatography was performed with an Agilent Technologies 9890N Gas Chromatography System with 7683 Auto Liquid Sampler and a Solgel WAX (polyethylene glycol) 30 m x 0.25 mm column. Helium was used as the carrier gas. The sample split ratio was set to 1:60 and a flame ionisation detector (FID) was used. When manual injection was required, it was performed with a SGE syringe. Mass spectrometry was performed with an Agilent 5973 Mass Selective Detector using gas chromatography with a HP5MS 30 m x 0.25 mm nonpolar column.

2.6 Environmental surveys with OP10-specific primers

The process pipeline of the culture-independent approaches in this study can be summarised in a flowchart as shown in Figure 11.

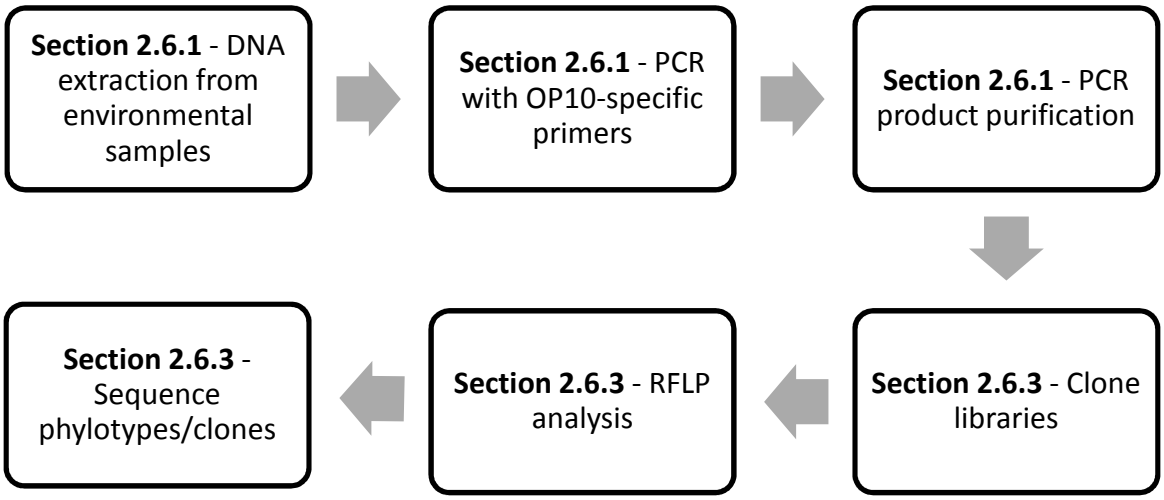


Figure 11 - Process flowchart of 16S rRNA gene sequence analysis with OP10 specific primers

2.6.1 DNA extraction and PCR from environmental samples

General PCR materials

The i-Taq™ PCR kit (iNtRON Biotechnology) was used according to the manufacturer's instructions. It included separate tubes of Taq polymerase, 10x PCR buffer, dNTP mixture (2.5mM each), and appropriate primers. To purify PCR products the PureLink™ PCR Purification Kit (Invitrogen) was used according to the manufacturer's instructions. The included High-Cutoff Binding Buffer purified mainly large DNA fragments (>600 bp), although it purified some smaller 300-600 bp dsDNA fragments as well.

All primers were manufactured by Invitrogen and purified by desalting. The primers used in this study are listed in section 2.6.2 (Bacterial universal primers and OP10-specific primers) and section 2.6.2 (M13 clone library primers).

DNA Extraction and amplification

DNA was extracted from samples collected from various environments using a ZR™ Fungal/Bacterial DNA Extraction Kit (Zymo Research) as per manufacturer's instructions. A list of sample sites where DNA were extracted is list in Table 1. PCR was performed according to the manufacturer's (iNtRON Biotechnology) instructions. Taq Master PCR enhancer was used when amplifying DNA from environmental samples. Two thermocyclers were used in this study: a Corbette Research CG1-96 and a Stratagene Mx3000p. Both have a capacity of 96 tubes, heated lids, and have essentially identical functionality in the scope of the procedures performed.

The balance between product yield and amplification stringency in PCR is influenced by parameters such as the temperature and runtime at each step, as well as the number of cycles. Generally speaking, higher stringency leads to lower product yield and lower sensitivity to the starting material. Higher annealing temperatures lead to higher stringency (Rychlik, Spencer, & Rhoads, 1990). More amplification cycles lead to more products, but less stringency with more PCR artefacts (Acinas et al. 2005). Optimisation was required (refer to PCR optimisation section below) in order to find a balance between these parameters for an efficient and reliable PCR with OP10 specific primers.

The PCR programme used for OP10-specific primers was based on the programme for universal primers for bacterial 16S rRNA sequences.

Table 3 - PCR thermalcycling programme for bacterial 16S rRNA gene primers 9F + 1492R

1 cycle	94 ^{o}C 5 min	Initial denaturation
35 cycles of:	94 ^{o}C 45s	Denaturation
	61 ^{o}C 1 min	Annealing
	72 ^{o}C 2 min	Elongation
1 cycle	72 ^{o}C 5 min	Final elongation
1 cycle	15 ^{o}C infinity	Storage

M13 primers used to amplify clones of the clone libraries used a "touchdown" programme. Touchdown PCR is a way of avoiding nonspecific amplification by slowly decreasing initial high annealing temperature as the cycles progress. The initial high annealing temperature provides the stringency and produces few but correct templates. As the number of correct

templates increases due to specific binding, decreasing annealing temperature allows more amplification to occur, thus resulting in sufficient amounts of PCR product.

Table 4 - Touchdown thermalcycling programme for M13 clone library primers

1 cycle	94 °C 10 min	Initial denaturation
20 cycles of:	94 °C 60s	Denaturation
	62-52 °C 45s 0.5s decrease each cycle	Annealing
	72 °C 90s	Elongation
15 cycles of:	94 °C 60s	Denaturation
	52 °C 45s	Annealing
	72 °C 90s	Elongation
1 cycle	72 °C 6 min	Final elongation
1 cycle	4 °C infinity	Storage

2.6.2 OP10 primer design

Selecting reference sequences

A primary aim of this study was to create OP10-specific primers, which would aid discovery of new OP10 species by identifying them in environmental samples that could then be targeted for cultivation efforts. Therefore, the primers need to collectively cover the hypothetical OP10 diversity within some practical limits (i.e. a limited number of primers to make them practical to use in diversity studies of this candidate division), yet specific enough to exclude non-OP10 sequences. That is to say, the primers were designed to be "OP10 specific" as a whole set, rather than "OP10-clade-specific". Therefore the overlapping of OP10 coverage was not selected against in the design process.

Greengenes, which had the most comprehensive OP10 taxonomic data as well as built-in chimeric sequence detection (DeSantis et al. 2006; Huber, Faulkner, and Hugenholtz 2004) was used to select suitable reference OP10 16S rRNA gene sequences. Only sequences with a minimum of 1250 bases and less than 50 ambiguity characters were chosen. The reference sequences were then manually checked and adjusted through the software package ARB

(Ludwig et al. 2004). A similar 16S rRNA gene sequence database, SILVA (Pruesse et al. 2007), was found to have a better alignment in terms of consistency and base pairing to secondary structure (Schloss 2009). Furthermore, SILVA offered direct download of aligned sequences while Greengenes utilised email delivery, which proved to impractical for large sequence files.

Since good primer design relies on an accurate reference phylogenetic tree, and a proper alignment is essential for a good phylogenetic tree, sequences were manually adjusted (using ARB) after the sequences were automatically aligned by one of the automatic aligners (SILVA and ARB's internal aligner) to eliminate some shortcomings of these algorithms.

Further attempts were made to retrieve additional OP10 sequences from the NCBI database (http://www.ncbi.nlm.nih.gov/nucleotide/) using various keyword searches. This method produced inconsistent and erroneous results. These results were in agreement with Portillo and Gonzalez (Portillo and Gonzalez 2008) who suggested that many OP10 sequences were unidentified, missing or incorrectly identified. For this reason, only sequences curated by Greengenes and SILVA were included in this study.

A total of 174 reference sequences were selected based on the "Hugenholtz taxonomy" referenced in Greengenes (DeSantis et al., 2006) and sequence alignment from SILVA release 98 (Pruesse et al., 2007). Additional manual adjustments were made within ARB to ensure alignment quality. A frequency filter was generated from these aligned sequences to ignore alignment columns with missing or ambiguous bases. A phylogenetic tree was generated from the filtered sequences using the Fitch-Margoliash distance matrix method ("fitch") of the "Phylip Interactive" module within the ARB programme to generate an additive tree. The reference tree generated is included in the Appendix 3.

Primer design from reference sequences

An ideal set of specific primers should satisfy three conditions:

1. The primers should cover their respective clades without missing any known reference sequences.

2. The primers should NOT form false positives (crosstalk) with sequences outside the OP10 candidate division, although crosstalking outside their respective clades but within OP10 is acceptable.

3. The smallest number of primers based on least subdivided, high-level clades should be used, as long as the previous two conditions were met.

The first two criteria are easily quantified, while the last criterion is a balance between the previous two and the total number of primers required to cover the diversity of the candidate division.

ARB was used to manage the sequences of the SSUref subsection of the SILVA database. For primer design, OP10 clades based on the reference tree were exported in separate FASTA files. An entire SSUref file minus OP10 sequences was also exported to act as the "outgroup". The programme PRIMROSE (Ashelford, Weightman, and Fry 2002) was used to search for suitable primers within the supplied sequences while attempting to exclude sequences supplied in the "outgroup". The main limitation of the programme was that it always attempted to cover the given sequences with only a single oligonucleotide, instead of the multi-primer approach of this study. To address this problem, OP10 sequences were divided into various clades based on the groupings of the reference tree.

The first few attempts failed because the initial clades covered too much diversity. Thus PRIMROSE gave a list of highly degenerate primers that still had poor coverage (in-group) and high mismatch (out-group). Further dividing the clades into subclades and sub-subclades eventually lead to desirable results. The primers generated by PRIMROSE all had 100% (1 dp.) hit rates of reference sequences of their respective OP10 clades, and 0.0% (1 dp.) hit outside OP10 clades against the entire SILVA98 SSUref database (368,000 total sequences).

PCR with environmental samples is difficult to optimise because different sequences have different binding energies to the primers. Therefore, it was worthwhile to design the primers to withstand certain amounts of mismatching. According to PRIMROSE, all the primers had >=0.2% hits to the outgroup at 2 mismatches. Primer degeneracy was kept to a minimum as it compounds the effect of false positives with increased mismatching. Aside

from binding specifity, primers were also selected for their compatibility with the bacterial universal primers they were paired with with in terms of similar annealing temperatures and suitable product sizes.

Universal bacterial 16S rRNA gene primers

From (Lane, Stackebrandt, and Goodfellow 1991).

Name	Orientation	5'-3' sequence
27	Forward	AGA GTT TGA TCM TGGCTCAG
1492	Reverse	TACGGYTACCTTGTTACGACTT

OP10 Primers

Forward OP10 primers pair with 1492R and reverse primers pair with 27F.

Name	Orientation	5'-3' sequence
OPX-1A	Forward	GAAAGACTTAGGACGGTACC
OPX-1B	Reverse	GTATTCACGGCGGTATGGCT
OPX-1C	Reverse	CACGGACTTCAGGTGCAGAC
OPX-2	Forward	CGAACGGTCGCGTAACACGT
OPX-3A	Reverse	GGTTCTACGGTTAGTGACG
OPX-3B	Reverse	GCTGGCAACATCYGATGAGG
OPX-3C	Reverse	CCGCTTACACGGGCAGTTTC
OPX-4A1	Forward	GGGARAAGTGGAATGGCTGG
OPX-4A2	Reverse	TTACCCCGGCAGTTTCCCAA
OPX-4A3	Forward	CGGCGAACGGTCGAGTAACA
OPX-4B	Forward	TGCCTGAGAAATGCAAGTCG
OPX-5A	Forward	GACACGGCCCTGACTCCTTT
OPX-5B	Reverse	TGACGTATTAATCCACATGC

2.6.3 Primer testing

As stated in the PCR method section, the OP10 PCR programme was based on a programme designed for universal primers targeting bacterial 16S rRNA.

Melting and annealing temperatures of the primers were estimated using Wallace's rule (Wallace et al. 1979). Annealing temperature by definition is lower than the melting temperature (Rychlik, Spencer, and Rhoads 1990), therefore it was initially chosen to be 5 °C lower than the theoretical melting temperature. Elongation time with Taq polymerase is recommended by the manufacturer (iNtRON Biotechnology) as 1000 bp/minute, and since the intended products were just short of the 1.5 kb of the 16S rRNA gene sequence, the elongation time was set at 2 minutes.

In practice, optimising PCR parameters, for specificity and good product yield required trial and error. Specific annealing temperatures used for each set of primers are listed in the results section.

Testing primers on environmental samples

Previous 16S rRNA gene sequence surveys of geothermal sites and culturing efforts (Stott et al., 2008) in the TVZ, at Waikite, Tikitere, Mount Ngauruhoe, and Waipahihi (Figure 5) identified sites where OP10 was previously detected. Environmental DNA samples from these sites therefore, can there act as positive controls to validate the newly designed primers.

Clone library

Ligation and chemical transformation of PCR products were conducted using the TOPO TA Cloning® Kit with pCR2.1-TOPO plasmid (Invitrogen) and One Shot TOP10 Chemically Competent Cells (Invitrogen) according to the manufacturer's instructions with the following exception: clonal inserts were directly amplify by picking positively transformed colonies with sterile micropipette tips and transferred to PCR master mix containing M13 primers.

M13 primer sequences:

Primers and orientation	**Sequence 5' – 3'**
M13 Forward	TGTAAAACGACGGCCAGT
M13 Reverse	CAGGAAACAGCTATGAC

Luria-Bertani (LB) broth + X-gal plate medium

Difco LB™ was made according to the manufacturer's instructions, then autoclaved. When the solution cooled down to ~60 °C but before the agar set, X-gal dissolved in dimethylformamide (concentration 80-100 mg/L) was added for a final concentration of 20 µg/mL, ampicillin for a final concentration of 100 µg/mL, and IPTG (isopropyl β-D-1-thiogalactopyranoside) for a final concentration of 0.1 mM.

Restriction digest for RFLP

Restriction enzymes - FastDigest® HinP1I (Hin6I) and MspI (Fermentas) were used according to the manufacturer's instructions.

HinP1I restriction site:

5'...G^C G C...3'
3'...C G C^G...5'

MspI restriction site:

5'...C^C G G...3'
3'...G G C^C...5'

Sequencing

All sequencing work was performed by the Waikato DNA Sequencing Facility of the University of Waikato, Hamilton.

2.6.4 Estimating PCR sampling depth

Sequence chromatograms from a clone libraries were loaded into Geneious™ (Drummond et al. 2009). After assembling each bidirectional sequence (forward and reverse), the assemblies were checked with megablast against the NCBI nr database to establish their identities, again through Geneious. Sequences identified as OP10 were manually inspected and adjusted to eliminate artefacts. The consensus sequences from the assemblies were then exported as a single FASTA file.

The FASTA file was then uploaded to the online aligner of SILVA to generate an aligned .arb file. The file was opened with ARB to calculate a distance matrix via the “ARB Neighbour Joining” option, from which ARB outputs a distance matrix in Phylip Lower Triangle format. The distance matrix file was loaded into the command-line software DOTUR (Schloss & Handelsman, 2005) with the command “dortur –l file_name” to generate various sampling statistics, including rarefraction and collectors curves i.e. Chao1 and ACE (Schloss & Handelsman, 2005).

3 Results

3.1 T49 culture based characterisation

3.1.1 General colony and culture observations

T49 followed a consistent growth pattern in plate and liquid cultures (Figure 10 and 11) as follows. Three days after inoculation, T49 on AOM1 plates exhibited a light pink colour as the colonies began to form, and the colour darkened slightly as the culture ages.

Observation under a phase contrast microscope (1000x) has shown that the morphology of the cells altered as the culture progressed. During early to mid log phase, cells were observed as individual rods with rough cell outlines. As the culture progressed, cells began to take on irregular morphologies; the rods became irregular in width and were often bent. Cell clumping was observable (Figure 14), which complicated cell counting. The morphology of T49 colonies in plate cultures was circular in form with a convex elevation and an entire margin (Figure 12). On the AOM1 plates used, T49 colonies formed indentations (Figure10-A) into the Phytagel™ medium, suggesting the actions of extracellular enzymes or the collapsing of supporting medium as T49 digested its immediate surroundings.

Figure 12 – T49 colonies, showing its round morphology and pink-orange colour.

When T49 was grown in 4.5NZS10.2 medium over the same time period transformed the clear medium (Figure 15A) to a white turbid solution when grown at 180 RPM. The colour developed to orange then pink (see Figure 16B) as the culture transitioned from log phase to stationary phase, as determined by OD 600 nm observations (refer to section 3.3.4). The amount of cell clumping in liquid culture coincided approximately with this colour formation as well. Cell clumping did not appear to be the result of further increase in biomass, as it developed during the stationary phase and made little difference to OD observed. The effect of cell clumping was apparent when cells were placed in cuvettes for OD measurements, as older cells settled, cells in younger cultures remained suspended.

When T49 was grown in 4.5NZS10.2 medium without shaking (as those used in section 3.3.6) cell aggregates tend to settle to the bottom of container regardless of growth stage or substrate used (Figure 13).

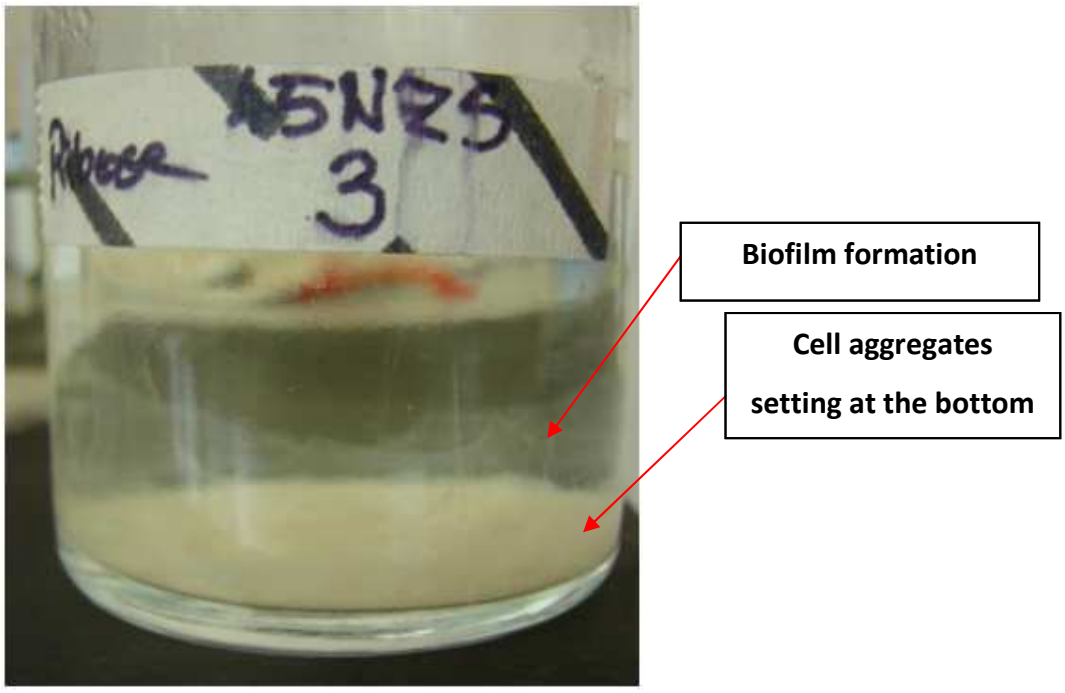


Figure 13 - T49 in liquid culture settling at the bottom of the container, leaving clear medium on top

Figure 14 - T49 in phase contrast microscopy, showing its rod-shape morphology and a small amount of cell clumping

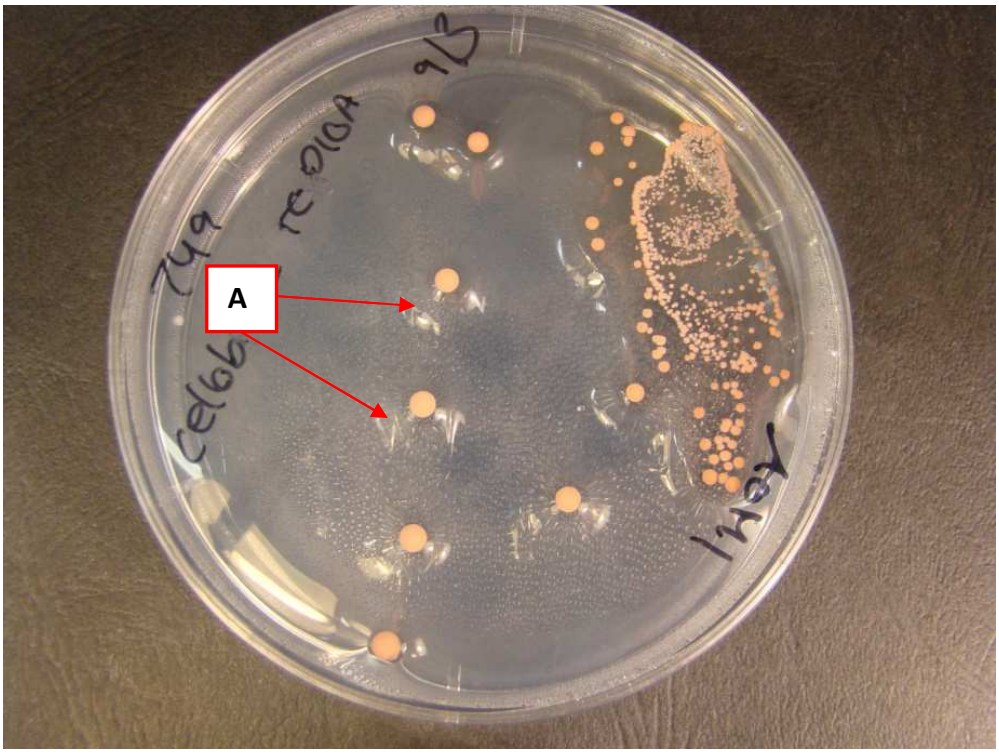


Figure 15 - T49 colonies on an AOM1 plate. A) Arrows indicate indentations around the colonies

Figure 16 - Liquid media A) before inoculation and B) after 1 week of growth

3.1.2 Substrate utilisation

T49 was tested against various substrates including monosaccharides, disaccharides, polysaccharides, and alcohols (Table 5). These substrates were targeted because Phytagel™, a polysaccharide, was used as a substrate to isolate a strain closely related to T49, and P488 growth was reported on simple sugars (Stott et al. 2008). Furthermore, T49 was isolated from an area (refer to Figure 10, TT13.4) noted for an abundance of plant matter on the surface and the accumulation of organic materials at the edge of organic rich topsoil layer. Various simple sugars were tested to investigate the selectiveness towards monosaccharides and disaccharide bond linkages, while glucose-based polysaccharides were tested to investigate the effect of polymer structures on substrate availability by T49. Other polysaccharides help to further investigate the specificity metabolic capability of T49 with various combinations of monomers, glycosidic linkages, and general structures. Alcohols were tested as they are common byproduct of carbohydrate fermentation which may be present in the environment where T49 was found.

T49 was able to utilise all monosaccarides tested, including aldopentoses, aldohexoses, and D-fructose, although the latter showed a noticeable decrease in growth. Two monosaccharide derivatives, D-galacturonic acid and D-N-acetylglucosamine were tested (Table 5). T49 did not show any growth reaction with D-galacturonic acid (a sugar derived acid), but showed weak growth witht D-N-acetylglucosamine (NAG), a common component of bacterial cellwalls.

T49 exhibited similar growth with all disaccharides tested. With the exception maltose which, which generated a particularly large amount of biomass. The growth tests that used glucose-based polysaccharides produced polarised results for T49. Starch, glycogen, dextrin and CMC had good growth, while Avicel™ and Whatman filter paper showed no growth at all. The results for other polysaccharides were equally varied; xylan, locust bean gum and xanthan showed good growth, while T49 grew poorly in pectin. and did not grow in sodium alginate, agarose, and chitin.

T49 did not grow in alcohols or the two organic salts chosen due to their role in the citric acid cycle and glycolysis. Neither did T49 grow in lignin or shown any signs of growth with the yeast extract (YE) and B vitamins added alone.

Table 5 - Growth reaction of T49 using various substrates and their properties

Substrate	Growth reaction	Structure linear/branched	Glycosidic bonds	Constituent monomers
Monosaccharides				
Aldopentoses		n/a		
DL-arabinose	+++			
D-ribose	+++			
Ketohexose				
D-fructose	+			
Aldohexoses				
D-glucose	++			
D-mannose	+++			
D-galactose	+++			
Sugar derivatives				
D-galacturonic acid	-			
D-N-acetylglucosamine (NAG)	+			
Disaccharides				
Sucrose	++	n/a	α(1→2)	glucose/fructose
Lactose	++		β(1→4)	galactose/glucose
Maltose	+++		α(1→4)	glucose/glucose
Trehalose	++		α(1→1) α	glucose/glucose
Cellobiose	++		β(1→4)	glucose/glucose
Polysaccharides				
Glucose-based polymers				
Avicel™	-	Linear	β(1→4)	glucose
Cotton	-	Linear	β(1→4)	glucose

Starch	+++	Branched	α(1→4)/ α(1→6)	glucose
Glycogen	+++	Branched	α(1→4)/ α(1→6)	glucose
Dextrin (corn, type 1)	+++	Branched	α(1→4)/ α(1→6)	glucose
Carboxymethyl cellulose (CMC)	++	Linear	β(1→4)	glucose
Whatman filter paper	-	Linear	β(1→4)	glucose
Algal and plant-derived polymers				
Xylan	+++	Mainly branched	β(1→4)	xylose
Sodium alginate	-	Linear	β(1→4)	mannuronic acid/ guluronic acid
Locust bean gum (glactomannan)	+++	Branched from a linear backbone	β(1→4)/ (1→6)	mannose/ galactose
Agarose	-	Linear	(1→3) / (1→4)	D-galactose/ L-anhydrogalactose
Pectin	+	Some branching	various	various
Animal-derived polymers				
Chitin	-	Linear	β(1→4)	NAG
Bacterial exopolysaccharides				
Phytagel™	+++	Linear	β(1→4)/ α(1→3)	glucose/ rhamnose
Xanthan	+++	Branched from a linear backbone	various/ β(1→4) backbone	glucose/ glucuronic acid / mannose
Alcohols				

Methanol (0.05%)	-	n/a		
ethanol (0.05%)	-			
ethanol (0.1%)	-			
1-propanol (0.05%)	-			
2-propanol (0.05%)	-			
Organic salts				
Sodium pyruvate	-	n/a		
Trisodium citrate	-			
Others				
Yeast extract (YE) (100 mg/l)	-	n/a		
Vitamins (100 mg/l)	-			
YE + vitamins (100 mg/l)	-			
YE + vitamins (10 mg/l)	-			
Lignin	-	Complex structure	Complex	various aromatic hydrocarbons

3.1.3 Validation of optical density measurements

Optical density (600 nm) was used to quantify growth of T49 under various conditions (section 2.5.3). To validate the measurements, T49 was cultured in 4.5NZS10.2 liquid medium at 60 °C while monitoring the cell numbers with a haemocytometer (section 2.5.4) and measuring the optical density with a spectrophotometer. The clumping nature of liquid culture in later stages made cell count measurement past later log phase difficult. Experiments were conducted in triplicate to provide accurate cell counts (Figure 17).

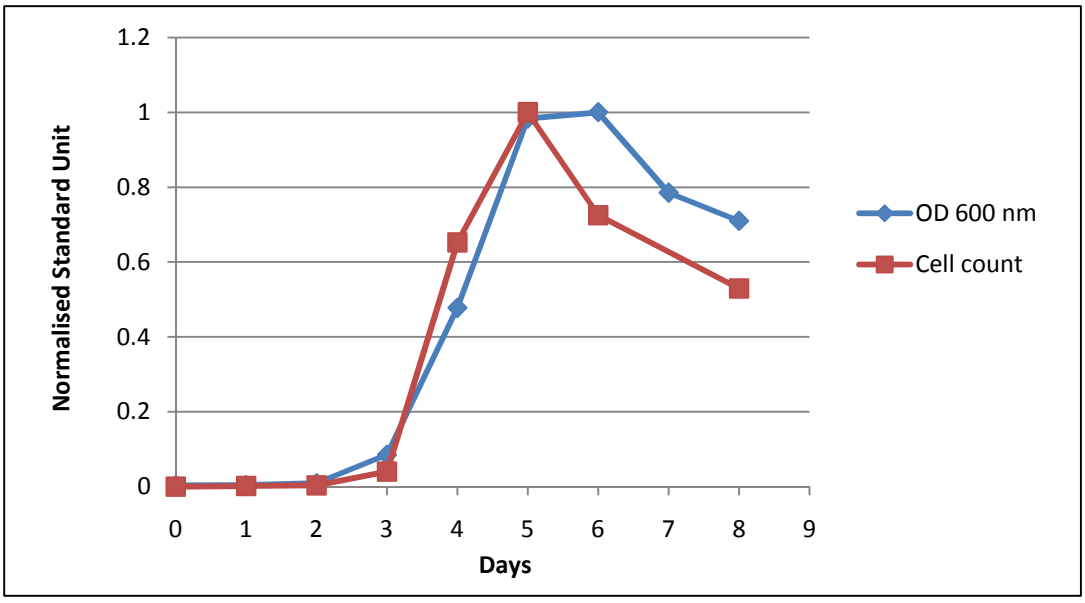


Figure 17 - Normalised measurements of T49 growth in a liquid culture.

The Pearson correlation coefficient for OD and haemocytometer measurements $r = 0.95$, indicating high correlation between the two measurements. Despite cell clumping, optical density produced consistent results when the cell clumps were temporarily resuspended via flicking of cuvettes prior to OD measurements.

3.1.4 Optimal temperature and temperature range

Due to the tendency of T49 to form pellicles in liquid culture, it was not possible to fully standardise inoculums between cultures. The determination of growth rate was instead calculated from the growth rate at log phase, which was less affected by the starting condition.

Since T49 was isolated and cultured at 60 °C, the temperatures tested were chosen around this temperature as a reasonable starting point. As detailed in Figure 18, T49 has a growth temperature range from below 50 °C to a maximum of 73 °C. The growth rate peaked at 68 °C, then declined rapidly at greater temperature. No growth was observed at 76 °C and 37 °C. The error bars were generated from the standard deviation between the triplicates used at each temperature; error range increased with growth rate. The error bars for 60 °C and 50 °C are shown but are almost negligible.

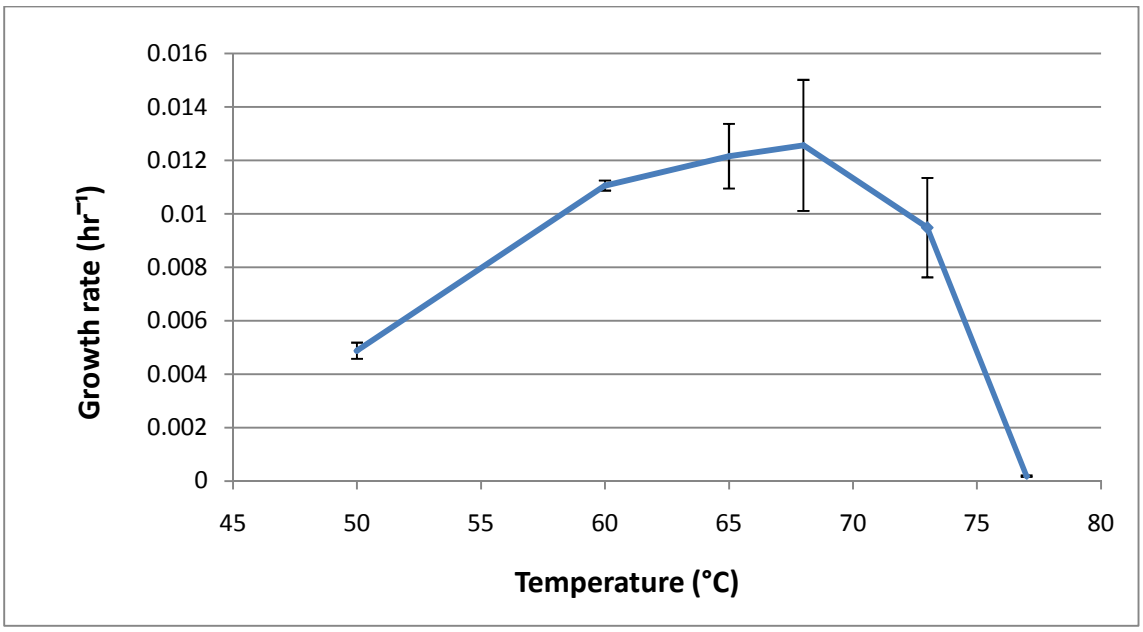


Figure 18 - Normalised T49 growth rates at log phase at various incubation temperatures

3.1.5 Optimum pH and pH range

Growth rate measurement of pH optimum and range by analysing log phase followed the same rationale as the temperature optimum and rate experiment. The medium was buffered to maintain accurate pH conditions for the cultures. The temperature used for the experiment was chosen at 60 °C because it was a known temperature for T49 growth in liquid culture as well as for the low margin of error demonstrated in section 4.1.4.

As shown in Figure 19, the pH range for the growth of T49 in liquid culture was from pH 4.9 to 5.8. The optimum pH was around pH 5.3. The growth rates dropped off to extinction at lower than 4.9, while the decrease in growth rate was less abrupt at the high end.

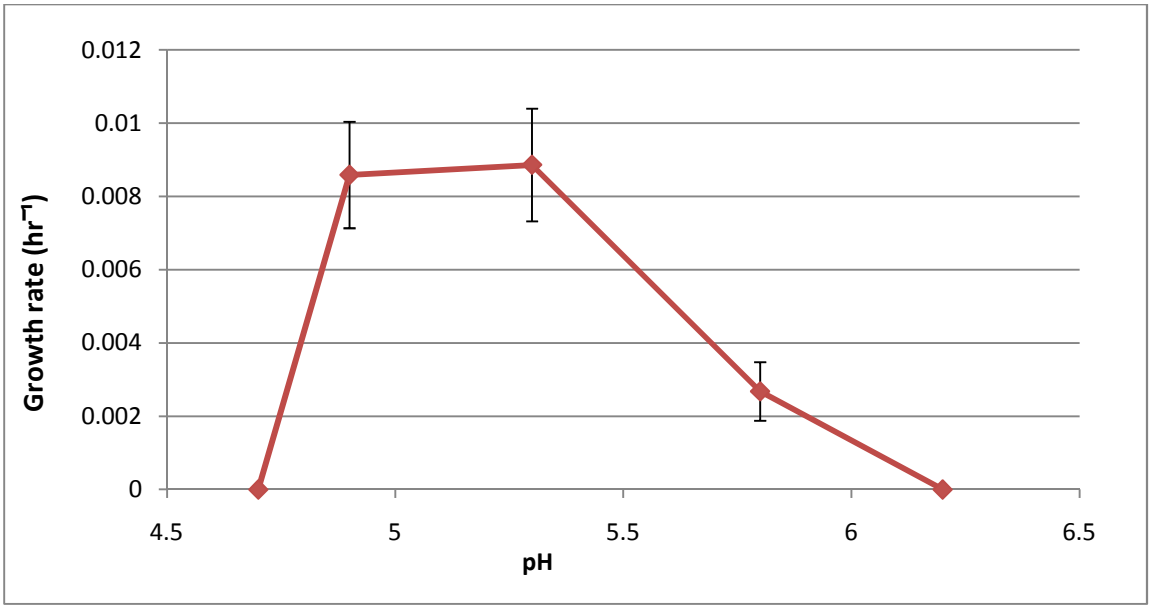


Figure 19 - T49 log phase growth rate at 60 °C between pH 4.7 to pH 6.2

3.1.6 TEM/SEM analysis

TEM images of T49 suggested the existence of two lipid bilayers interspaced by a layer of peptidoglycan (Figure 20), a characteristic of Gram-negative cell walls. The bacterium is rod-shaped, 0.5-0.7 µm in width, 2.5-3.0 µm in length, with many irregular folds in the outer membrane (Figure 22). The cells had two noticeable features in the cytosol. The first feature was an electron-light cavity which may have contained crystaline substances that were dislodged during sample preparation (Figure 21A). The cavity tended to be positioned within the centre of the bacterium, and was roughly 100 nm in diameter. The second features were electron-dense granules that tended to be positioned close to the inner surface of the bacteria, especially at the two polar ends (Figure 21B). These two features appeared to be ubiquitous in T49 cells observed (Figure 21).

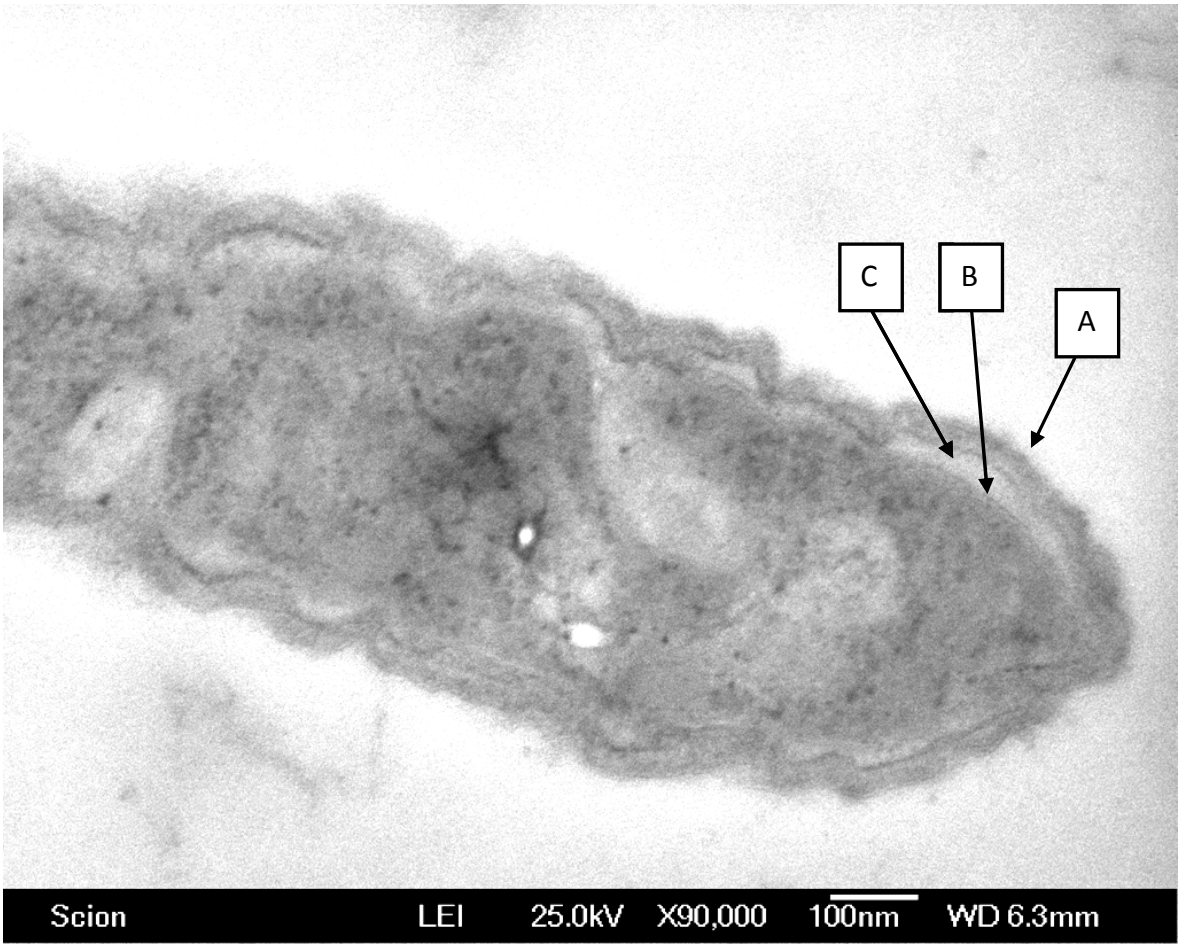


Figure 20 - T49 TEM image showing the outer A), inner membrane B), and peptidoglycan layer C) as indicated by the arrows.

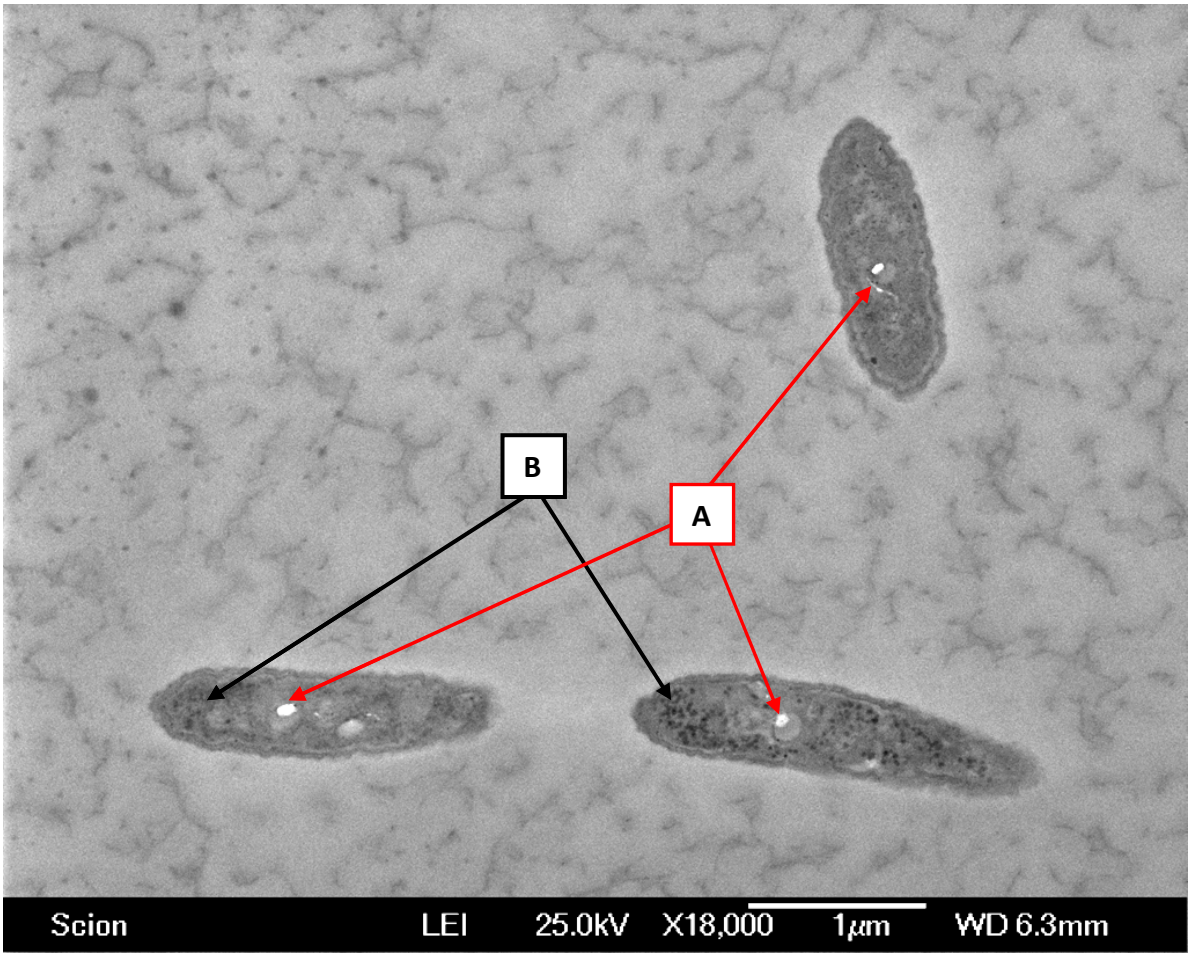


Figure 21 – TEM image of A) Electron-light cavities and B) Electron-dense granules present in T49.

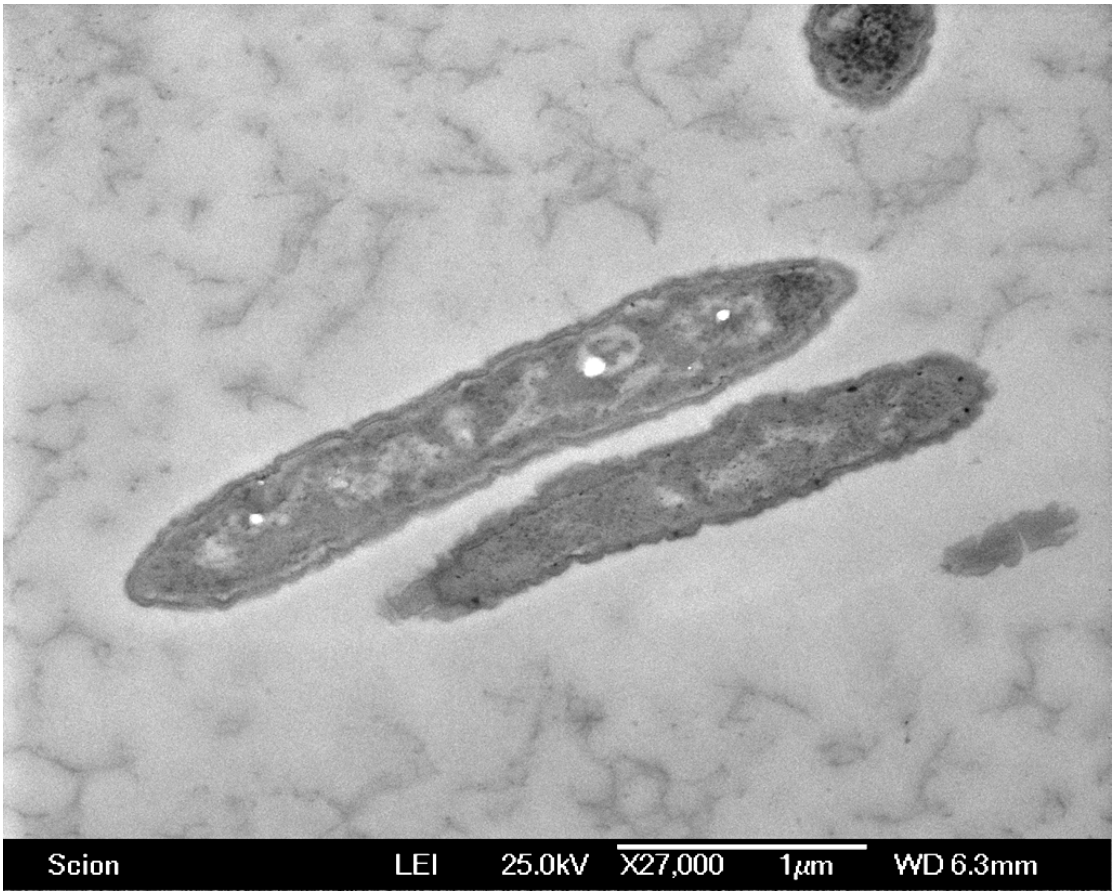


Figure 22 - TEM image of T49 showing full crosssection.

SEM images of T49 further confirmed the regular rod shape of T49, as well as the irregular surface, which appeared as ridges and grooves (Figure 23). The bacteria were found to be embedded in a material not from the preparation process therefore likely to be carry-through of extracellular substances in colonies that resisted repeated washing.

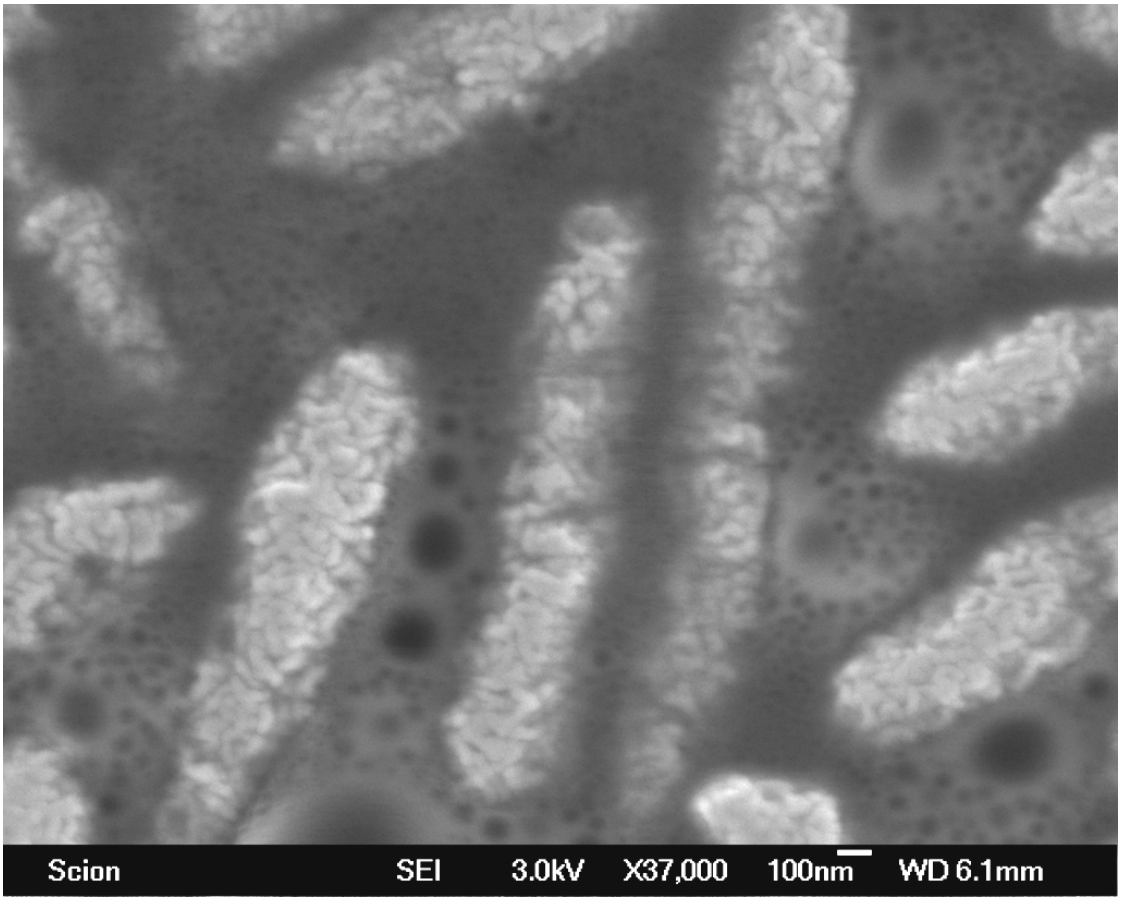


Figure 23 - SEM image of T49 with wrinkled surfaces embedded in the material carried-through from culture colonies

3.1.7 Summary of characteristics of T49 and a closely related strain P488 (Stott et al. 2008)

The phyiological data of T49 is here compared with that of OP10 strain P488, isolated from Waikite geothermal soil (Stott et al. 2008). The two strains have high 16S rRNA gene sequence similarity, with a pair wise identity of 99.9% against each other. T49 appeared to be similar to P488 in morphological features but had a higher maximum growth temperature, narrower pH range, and able to grow on pectin but not sodium alginate (the reverse was true for P488).

Table 6 - Comparison of selected physiological characteristics between strain T49 and P488 (Stott et al, 2008).

	T49	**P488 (Stott et al. 2008)**
Phylum	OP10	OP10
Isolation site	Tikitere	Waikite
Gram stain	-	-
Cell morphology	Rod	Rod
Cell dimensions (µm)	0.5-0.7, 2.5-3.0	0.5-0.6, 2.0-4.5
Colony pigmentation	Pink/orange	Pink/orange
Spore formation	-	-
Motility	-	-
Maximum growth T*	73 °C	65 °C
pH range*	4.9-5.8	3.8 –5.5
Requirement for CO2	-	-
Substrates		
Agar	-	-
Gellan	+	+
Xanthan	+	+
CMC	+	+
Xylan	+	+
Pectin	+	-
Sodium Alginate	-	+

* Cell samples were grown via two different methods for these characteristics. T49 was cultured at 180 RPM for, while P488 was grown in stationary medium.

3.1.8 Isolation of bacteria

During sample collection, DNA was extracted and cultivation attempts were made to isolate new species. The following species were isolated from two sites (Table 7).

Table 7 - Isolated and ongoing isolation cultures from two environmental samples

Culture designation	**Closest related cultured species**	**NCBI Accession #**	**Pair wise identity**
Waipahihi –WPH10			
"2"	Not yet isolated		
"DHB"	Not yet isolated		
"3.2"	*Anoxybacillus amylolyticus*	AJ618979	99
"red"	*Meiothermus ruber*	Y13597	100
Wairakei Claret Cup – WRG1			
WRG1 .1	OP10 isolate T49	AM749780	100
WRG1.2	*Geobacillus vulcani*	EU484349	100
WRG1.3	*Alicyclobacillus acidocaldarius*	AB059673	96
WRG1.4	*Bacillus tuscaiae*	Z26933	95

3.1.9 GC content

The GC content of T49 as determined by HPLC analysis (section 2.5.10) was 54.6 %. For comparison, *Deinococcus-Thermus* is one of the most closely-related phyla with cultured species; *Deinococcus radiodurans* has a GC content of 66.61% based on its genome sequence (White et al. 1999), while *Thermus thermophilus* has a GC content of 69.4% based on genome sequence (Henne et al. 2004).

3.1.10 T49 fatty acid methyl ester (FAME) profile

The FAME profile of T49 is as shown in Table 8, in ascending chain lengths. Major fatty acids of T49 included stearic acid (18:0) and palmitic acid (16:0) .

Table 8 - List of T49 fatty acid methyl esters and their relative percentages.

Fatty acids	% of total FA
14:0	1.0%
i15:0	0.4%
ai15:0	0.1%
15:0	0.2%
i16:0	0.8%
16:1	4.5%
16:0	32.2%
i17:1	2.9%
ai17:1	0.2%
i17:0	8.6%
ai17:0	4.8%
Unknown *	3.0%
17:0	0.4%
i18:0	0.4%
18:1	4.7%
18:0	35.2%

* Currently under investigation, possible cyclic fatty acid.

3.1.11 Methane utilisation

T49 was tested for methane utilisation because methanotrophs were often co-cultured with OP10 isolates and methane gas was usually in elevated concentration in environments where T49 and its closely related strain P488 were isolated (Stott et al. 2008). T49 was unable to grow in any of the medium variations used to test methane as a growth substrate (see Table 2).

3.2 Targeted environmental surveys with OP10-specific primers

3.2.1 PCR with OP10-specific primers

OP10 specific primers were designed via an interactive process using phylogenetic trees and and using PRIMROSE as described in section 2.6.2.

PCRs were performed with OP10-specific primers (as listed in section 2.6.2) on DNA extraction of environmental samples from locations listed in Table 1. In total, three OP10-specific (OPX-) primers were tested against 15 DNA samples, 11 of which were extracted from the environment, three were extracted from non-OP10 pure cultures. DNA from T49 was used as a positive control for primer OPX-4A2 as this primer was designed to cover T49 and its corresponding clade (Figure 26, yellow sequences).

As described in Table 9, primer set F27 and OPX-1BR had positive PCR results from three Waipahihi samples WPH01, WPH03, and WPH07. Primer set OPX-2F and 1492R had only one positive result in NGH02 from Mt. Ngauruhoe. Primer set 27F and OPX-4A2R had positive results from multiple locations: TT13.3 from Tikitere, WKT32 from Waikite, and NGH02 from Mt. Ngauruhoe. This primer set was also positive against the positive control (T49) as predicted.

3.2.2 Sequencing of PCR products and phylogenetic analysis

PCR products from some positive reactions (those marked red in Table 9) were sequenced to validate the performance of the primers in their ability to target OP10 sequences in environmental samples.

As reported in the previous section (3.2.1), OPX-1B had positive results from multiple samples. However, these samples were from a single site of relatively close proximity. The PCR products of WPH03 were sequenced, and the new sequences all fell in a cluster within clade 1B (Figure 24) where OP10 sequences were previously detected (Appendix 2, Figure 3). In contrast OPX-4A2R had positive results from multiple locations far apart from one another. Of the three PCR products sequenced, two (TT13.4 and WKT32) fell into tight clusters around the T49 sequence (Figure 26), suggesting strains or species closely related to T49. Sequences from NGH02, however, were unexpected, as previous 16S rRNA gene surveys only detected sequences within clade 2. These sequences fell into a wider spread within clade 4A2 (Figure 25). For the relative position of all the OP10 sequences obtained through OP10-specific primers, refer to Appendix 3 and 4.

Finally, primer OPX-2F had one positive result, which was from Mt. Ngauruhoe. The sequences were not restricted within clade 2, but fell into clades 2, 4A2, 4A1, and 4A3

(Figure 25), suggesting that the primer was not sufficiently specific, but also the sample (NGH02) contained high genetic diversity.

Table 9 - PCR results of OP10 specific primers against DNA samples

Samples	Primers		
	OPX-1BR	**OPX-2F**	**OPX-4A2R**
Environmental samples			
TWW1	-	Not tested	-
TWW2	-	Not tested	-
WPH01	+	-	-
WPH03	+	-	-
WPH07	+	-	-
TT13.3	-	-	-
TT13.4	-	-	+
WKT31	-	-	-
WKT32	-	-	+
WKT33	-	-	-
NGH02	-	+	+
Pure culture (non-OP10 species)			
Chloroflexus sp.			-
Methylocapsa acidophila B2		-	-
Thermomyces lanuginosus	-		-
	-	-	-
Pure culture (OP10 species)			
OP10 isolate T49	-	-	+

Note: the red colour indicates the PCR products were placed into clone libraries and sequenced.

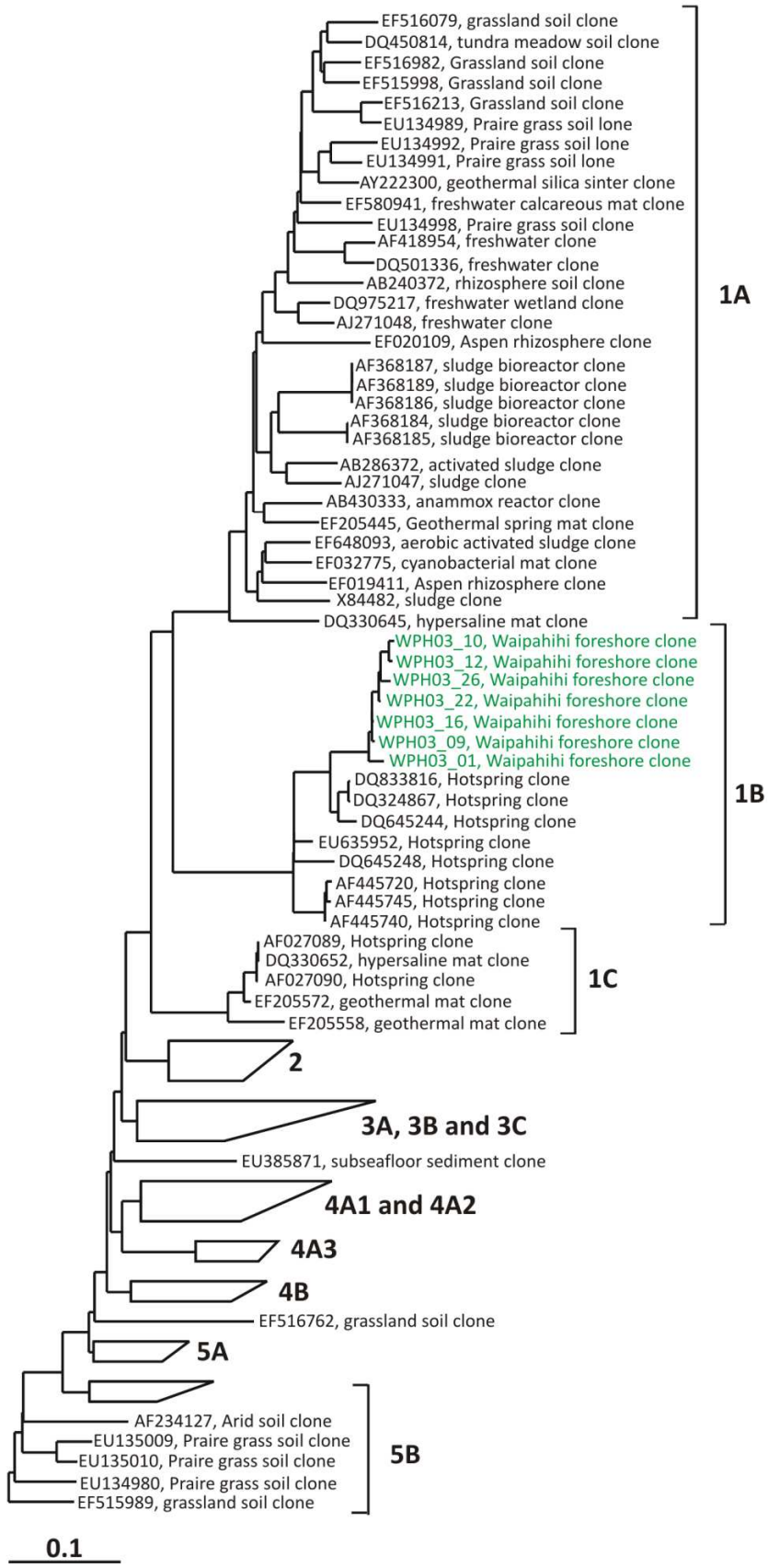


Figure 24 - Sequences obtained from PCR with 27F and OPX-1BR primer set and DNA extracted from environmental sample WPH03. All the sequences fell within clade 1B.

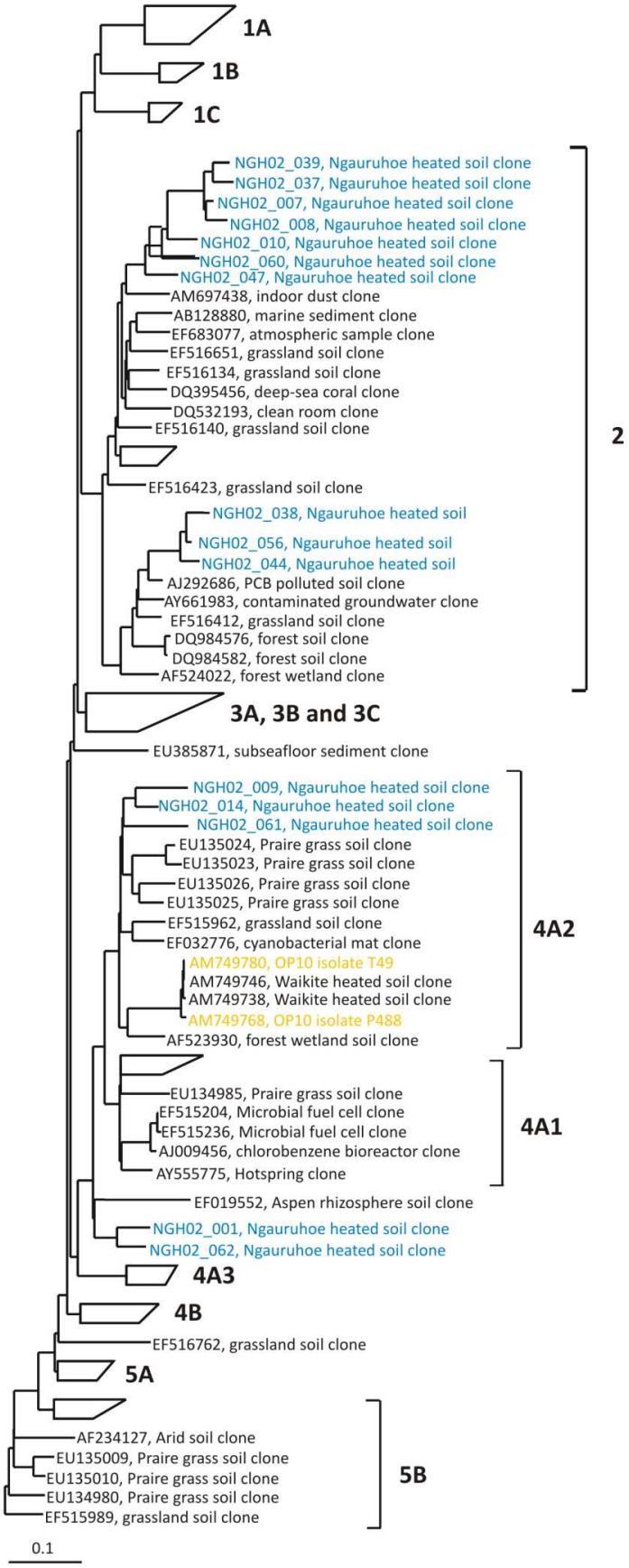


Figure 25 - Sequences obtained from PCR with OPX-2F and 1492R primer set and DNA extracted from environmental sample NGH02. The sequences fell with in multiple clades.

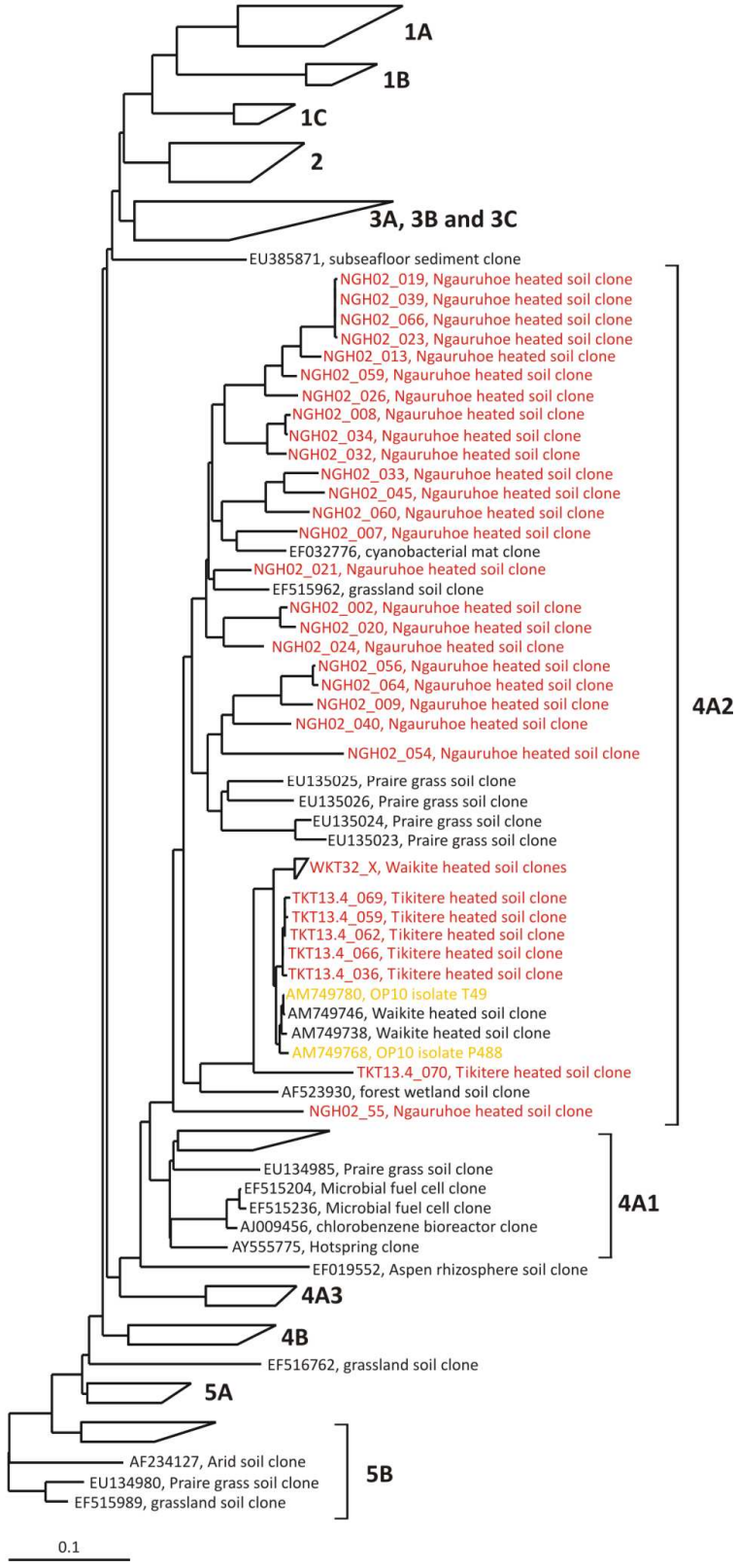


Figure 26 - Sequences obtained from PCR with 27F and OPX-4A2R primer set and DNA extracted from environmental sample TT13.4, NGH02, and WKT32. All the sequences fell within clade 4A2, with TT13.4 & WKT32 sequences clustering around T49 sequences (coloured yellow).

3.2.3 PCR optimisation

PCR optimisation was performed throughout the experimentation with OP10-specific primers, with a particular focus on OPX-4A2R as isolate T49 was within its target clade and could therefore act as a positive control of known DNA concentration. PCR sensitivity and stringency were optimised by adjusting annealing temperatures, increasing cycle number, and using PCR enhancers. Performing post-PCR cleanup to eliminate PCR residues (such as non-specific products and nucleotides) was also useful.

Negative controls included DNA extracted from an environment where previous 16S rRNA gene sequence surveys did not detect OP10, DNA from non-OP10 pure cultures of a strain frmo a closely related phylum (*Chloroflexus* strain T81), to distant phylogeny (*Betaproteobacteria Methylocapsa acidophila*), and from a different domain (Fungal specie Thermomyces lanuginosus). Figure 27 shows primer set OPX-2F and 1492R tested against a variety of DNA samples, including samples which included OP10 (T49, WPH03, and TT13.4) of a different clade than that targeted by these primers.

To determine the detection limit of 27F and OPX-4A2R primer set for pure DNA, T49 DNA was serially diluted by 100 x factors and each dilution was amplified at a PCR temperature gradient of 64-70 °C. The experiment was repeated three times as PCR became unreliable at boundary conditions (temperature and concentration), and the 1 x 10^6 dilution had a large margin of error. The detection limit of PCR with 27F and OPX-4A2R was determined to be between 0.16-1.6 ng/mL at the highest reliable temperature of 66.8 °C.

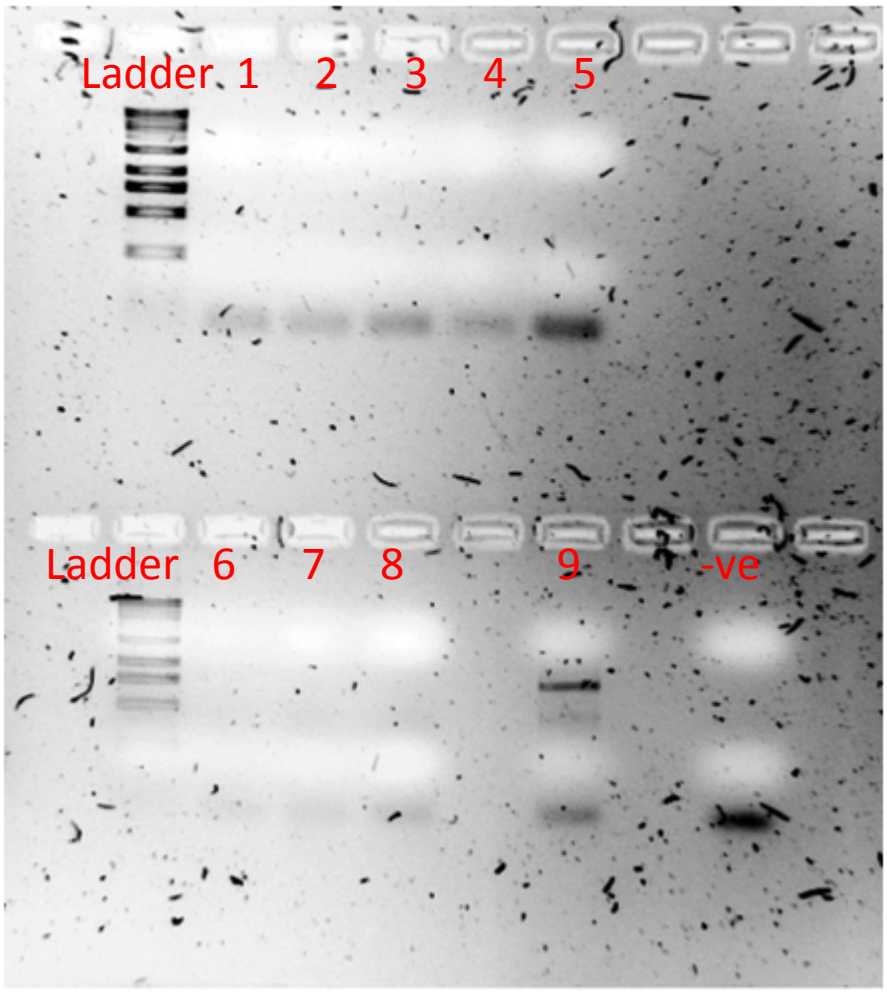


Figure 27 – Electrophoresis gel detailing primer optisation of the primer set OPX-2F and 1492R (designed to detect OP10 from Clade 2 in the NGH02 environmental sample). Lane: 1.WPH01; 2. WKT32; 3., WPH03; 4. TT13.4; 5. *M. acidophila;* 6. *Chloroflexus sp.;* 7. *T. lanuginosus;* 8. T49; 9. NGH02. ZR 1kb DNA ladder was used (Zymo Research)

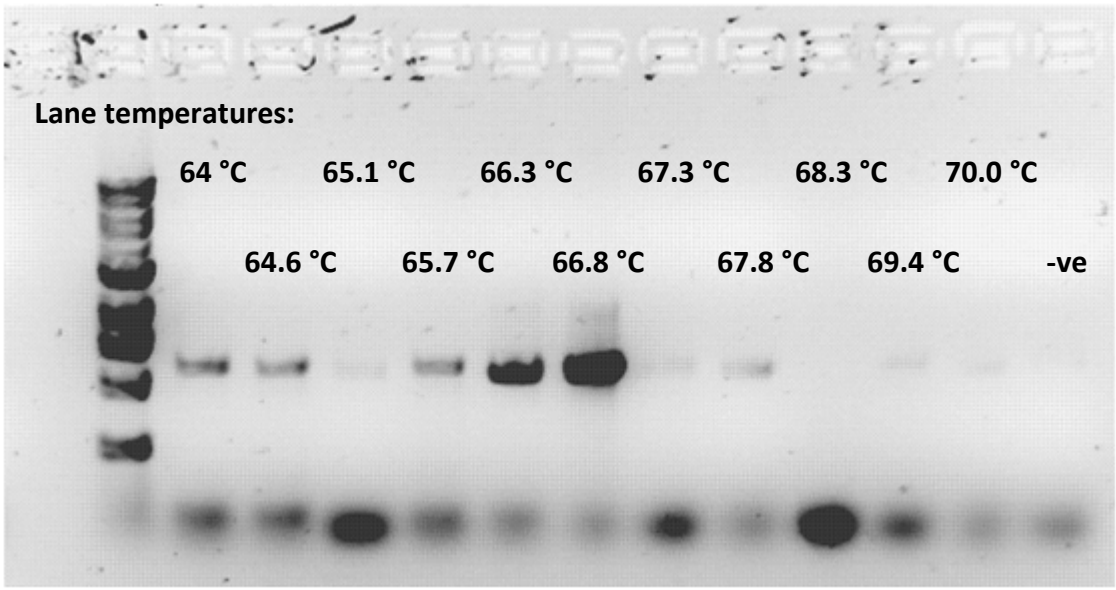


Figure 28 – An electrophoresis gel showing the result of a temperature gradient PCR with primers 27F and OPX-4A2R with T49 DNA at 0.16 ng/mL and gradient temperature of 64 – 70 °C to demonstrate the temperature and concentration limit of the reaction.

3.2.4 Sampling coverage estimation

Many of the sequences of the PCR products generated with OP10-specific primers were closely related, as shown in the clusters they form in the phylogenetic tree (Figure 24 and Figure 26). These sequences represent samples of the population diversity within the source environments. By using the Chao1 and ACE algorithms of the software package DOTUR (Schloss and Handelsman 2005), the amount of genetic diversity in an environment can be extrapolated from the sample data. Operational Taxonomic Units (OTUs) were defined as taxonomic groups organised by a chosen degree of genetic sequence similarity or dissimilarity.

Here an OTU was defined as less than 3% sequence dissimilarity for WKT32 and NGH02. Because low-diversity groups required a finer separation threshold to distinguish unique sequences, an OTU was defined as less than 1% dissimilarity for WPH03 and TT13.4. Table 10 details the Chao1 and ACE collectors curve estimations for each sample site where OP10 sequences were detected.

Sample NGH02 has shown high diversity with both primer sets, neither primer sets were able to reach a satisfying sampling depth in discovering the potential OTUs. Apart from

NGH02, the rest of the OP10 sequences from other primers have been sampled to sufficient depth, reaching the predicted number of OTUs with 95% confidence intervals.

Table 10 – PCR coverage - Chao1 and ACE OTU estimation vs. number of OTUs sampled.

Primers	Environmental DNA samples	Sequences & (OTUs)	Chao1 estimate and (±95% conf intervals)	ACE estimate and (±95% conf intervals)
OTU defined as 3% differences				
27F + OPX-4A2R	WKT32	82 (2)	2 (2-2)	2 (2-2)
27F + OPX-4A2R	NGH02	56 (29)	46 (34.5 – 81.5)	76.6 (47.4 – 151.7)
OPX-2F + 1492R	NGH02	26 (12)	32 (22.6- 66.6)	38 (29.6 – 54.2)
OTUs defined as 1% differences due to low diversity in these groups				
27F + OPX-1BR	WPH03	28 (2)	2 (2-2)	2 (2-2)
27F + OPX-4A2R	TT13.4	22 (2)	2 (2-0)	0 (0-0)

4 Discussion and Conclusions

T49 is the first characterised species in the novel phylum OP10. As one of the aims of this study was to characterise one of the first cultivated isolates of OP10, it was difficult to make conclusions based on other bacteria, as it is not closely related to other known species. Therefore, part of the research process was aimed at generating data for publication to describe the first species within a candidate phylum, in order to establish the first step towards investigating this taxonomic group from a physiological perspective. The second part of this study aimed to explore further into OP10 diversity beyond previous surveys using specific primers. From which a survey tool may also be developed to focus on further discovery of OP10 in the environment.

Physiology of T49

In order to understand a microorganism and its ecological role in an environment, how the microorganism obtains its energy from its surrounding may be the most important question of all. According to the experiments conducted in this study, T49 was capable of metabolising a range of common monosaccharides as well as disaccharides with various glycosidic bonds. This suggests that it was capable in metabolising a diverse range of simple sugars. However, this capability did not translate completely to larger polymers, including some glucose-based polysaccharides. For example, T49 could not break down cellulose, which was insoluble and highly structured. The inability to digest structured polymers was not limited to glucose-based polysaccharides, but also extended to chitin (NAG) and agarose (galactose and anhydrogalactose). The more linkages a polymer has, the more stable it is. Therefore ordered structures of cellulose that made it capable of supporting plants the size of large trees are often energetically unfavourable to breakdown (Anderson 2000) compared with digesting simple sugars when they are available . It appears that T49 does not possess cellulolytic enzymes and it was only capable of breaking down the most amorphous polymers such as Phytagel™.

In liquid cultures, T49 had a distinct tendency to aggregate and form pellicles, which made handling cells and measuring cell growth difficult. Only with aggressive was it was possible to obtain a planktonic cell suspension. This aggregation suggests T49 may be generating a

considerable amount of extracellular substances. T49 was regularly observed to aggregate and eventually settle to the bottom of the containers in the later stages of growth (Figure 13). Previous studies have linked the aggregation of cells in liquid culture with the production of extracellular polymeric substances (EPS) and the formation of biofilms (Borlee et al. 2010; Dow et al. 2003; Iwabuchi 2003). Second, observations in SEM images of T49 embedded in substances from colonies support this notion (Figure 23). Secretion of extracellular substances therefore may be part of the survival strategy of T49. Extracellular secretion may be correlated to change in cell morphology at later growth-stages with increase in cell clumping (Figure 14) and pigment (Figure 16), both of which have been previously associated with stress responses in other species (LaPaglia and Hartzell 1997; Babitha, Soccol, and Pandey 2007).

The substrate utilisation and aggregation observations are consistent with the environmental conditions where T49 was isolated. The stratification of soil at the isolation site may act as division lines of microbial communities based on the chemical composition of each layer and the effects the neighbouring layers had. For example, the degradation of complex plant-derived lignocelluloses would occur close to the soil surface, and create a downward metabolic chain as organic materials were carried down by water diffusion. T49 was isolated from just below a layer identified as possible an Ah horizon, which is known to be rich in organic materials and microbial activities (Bruneau et al. 2005). This scenario would provide substrates for T49 growth as other cellulolytic microorganisms higher in the soil profile degrade complex lignocellulose into more accessible forms. The ability of T49 to secret extracellular material may generate a competitive advantage in occupying living spaces or may benefit other organisms in a symbiotic relationship. The aerobic nature of T49 may allow the organism to form close symbiotic relationships with other cellulolytic microoranigsms. Lignocellulosic anaerobic microorganisms such as *Clostridium thermocellum*. It is feasible that T49 could “mop”-up oxygen and generate anaerobic microenvironments while savenging unutilised sugars from cellulose decomposition.

The optimum growth temperature agreed with the environment where T49 was isolated. The optimum pH, and pH range for growth of T49 measured in liquid medium disagreed with the environment measurement, as well as the measured pH of AOM1 plates. This may be due to the fact that the cells were mostly in planktonic form during the early stages of

growth in liquid culture, therefore the bacteria may be lacking in certain survival strategies such as clumping or the production of extracellular materials. The internal structures observed through TEM images (Figure 21) suggest the existence of a possible PHA inclusion body, and lipid bodies for the electron-dense areas, as osmium tetroxide used for staining binds to lipids preferentially (Bozzola and Russell 1999).

The fatty acid profile of T49 agrees with its thermophilic lifestyle, with longer-chain fatty acids than mesophilic bacteria, which is commonly believed to enhance membrane stability (Shen et al. 1970; Chan, Himes, and Akagi 1971) Furthermore, the unusual cyclic fatty acid may play an important role for the acidophilic and thermophilic lifestyle of T49 (Oshima and Ariga 1975). The GC content of T49 was found to be lower than some thermophilic species found in the neighbouring phylum *Deinococcus-Thermus*, but within the average range of GC content values for bacteria.

Environment surveys with OP10-specific primers

Despite the fact that only three pairs of primers were used in this study out of the many designed many new OP10 16S rRNA gene sequences were identified in this study. Critically, the primers were able to detect OP10 sequences from environmental samples discovered by previous surveys shown in Figure 24, Figure 25, Figure 26, and Appendix 4. The bias of the primers towards OP10 meant greater sampling depth was achieved, at many points, reaching the sampling goal of OP10 as predicted by species richness estimations (ACE & Chao1). OP10 clones are generally only a small constituent in previous 16S rRNA gene sequence surveys (Appendix 2). Therefore, 16S rRNA gene sequence surveys may fail to detect OP10 as a minor species this was demonstrated by the observation that despite strain T49 was isolated from the TT13.4 sample site, it was not detected in the inital universal 16S rRNA gene sequence survey (Appendix 2, Figure 2). Primer OPX4A2 managed to overcome this challenge and detected T49 in sample T13.4 with high specificity. The same primer was also able to detect the existence of an additional 4A2 clade in NGH02 previously undetected (Figure 25).

Despite the success, some shortcomings were noted for the primer design. Primer OPX-2A was designed to cover a wide clade (Clade 2). This has shown to be problematic as many non-OP10 sequences were picked up, among the diverse range of OP10 sequences

discovered. This may also suggest that NGH02 sample contained an unusual amount of diversity which may warrant future investigations as well as further refining the primer for higher specificity.

Future perspectives

The initial observations so far has generated many new questions. Many aspects of T49 observed warrant further investigations, including chemical analysis on the unusual cyclic fatty acid, the pink-orange pigment produced, the composition of the extracellular substances, and the possible PHA and lipid bodies within the cytosol. These features should also be investigated for their genetic basis through genome sequencing, which may yield many interesting genes that link to these physical characteristics. Transcriptomic analysis may provide insight into the genetic control and expression of these features, especially for cell clumping, quorum sensing, and pigment production as survival strategies. Ultimately, all these information will help us understand T49 from an ecological perspective within the complex soil microbial community it was found in, in order for us to answer questions such as how the bacteria interact and compete with other speices. The physiological information generated here, for example, the narrow pH range tolerated by T49 in planktonic form, may assist future efforts to isolate and characterise novel OP10 strains, such as, strain WRG1.1 isolated in the course of this study.

The investigation of OP10 through OP10-specific primers described in this study should be expanded. Specific primers not explored in this study should be validated as the design methodology has been supported by the positive results from the primers tested so far. All the OP10-specific primers can benefit from further validation and characterisation to establish their reliability as well as improve upon current designs based on new OP10 sequences discovered. From another approach, the primers can also be utilised as fluorescent probes in applications such as fluorescent *in situ* hybridisation (FISH), to identify OP10 in its native environments, which may be beneficial for structural study of bacterial communities in biofilms or aggregates found in bioreactors or aquatic environments. With recent advance in technologies, single cell multiple displacement amplification (MDA) may be combined with FISH in order to sequence OP10 species without even isolating the bacteria. Finally, the ecological information generated from detecting OP10 in various

environments may assist cultivation attempts by emulating the physiochemical factors of those environments.

Conclusions

The physiological characteristics of strain T49, one the first cultivated species within Candidate Division OP10 has been described in this study. The experiment data suggest that T49 is a thermophilic (growth detected at <50 °C to 73 °C and optimum at 68 °C), acidophilic (growth pH range of 4.9 to 5.8 and optimum at pH 5.3) , rod-shape bacterium, with dimensions of 0.5-0.7 µm by 2.5-3.0 µm, and a GC content of 54.6%. Its fatty acid profile suggest more than half (67.4%) of the total fatty acids of the bacteria consisted of stearic acid (18:0) and palmitic acid (16:0). T49 grew optimally using mono- and di- saccharides such as arabinose, mannose, ribose, galactose, and maltose, as well as amorphous polysaccharides including starch, glycogen, and dextrin. Based on these evidences, the author concluded that T49 is a scavenger of soluble sugars and amorphous polysaccharides from its organic rich natural environment. The tendency of T49 to aggregate and produce extracellular materials suggests a survival strategy of adapting to environmental fluctuations. The description provided in this study enable further studies into the first characterised species of a novel phylum.

OP10-specific primers used in this study have successfully detected new OP10 clades previously undetected by traditional 16S rRNA surveys, these include a site where OP10 was previously isolated but was not detected through a previous survey. The discovery of additional genetic diversity of OP10 in these locations suggest that despite OP10 sequences usually only made up a small proportion within environmental 16S rRNA gene seuquence surveys, may still be biologically active and more prevalent than previously thought. This study provided a means to further investigate OP10 in the environment with enhanced specificity and sensitivity.

Bibliography

Acinas, Silvia G, Ramahi Sarma-Rupavtarm, Vanja Klepac-Ceraj, and Martin F Polz. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and environmental microbiology* 71, no. 12: 8966-9. doi:10.1128/AEM.71.12.8966-8969.2005. http://www.ncbi.nlm.nih.gov/pubmed/16332901.

Anderson, Kevin L. 2000. Structural organisation of cellulose and starch. In *Glycomicrobiology*,Ronald J. Doyle, 360. Springer. http://books.google.com/books?id=641jNnIoJzgC&pgis=1.

Anderson, Robert T, Helen A Vrionis, Irene Ortiz-Bernad, Charles T Resch, Philip E Long, Richard Dayvault, Ken Karp, et al. 2003. Stimulating the in situ activity of Geobacter species to remove uranium from the groundwater of a uranium-contaminated aquifer. *Applied and environmental microbiology* 69, no. 10: 5884-91. doi:10.1128/AEM.69.10.5884-5891.2003. http://www.ncbi.nlm.nih.gov/pubmed/14532040.

Ashelford, Kevin E, Andrew J Weightman, and John C Fry. 2002. PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. *Nucleic acids research* 30, no. 15: 3481-9. http://www.ncbi.nlm.nih.gov/pubmed/12140334.

Babitha, Sumathy, Carlos Ricardo Soccol, and Ashok Pandey. 2007. Effect of stress on growth, pigment production and morphology of Monascus sp. in solid cultures. *Journal of basic microbiology* 47, no. 2: 118-26. doi:10.1002/jobm.200610261. http://www.ncbi.nlm.nih.gov/pubmed/17440913.

Barns, S M, R E Fundyga, M W Jeffries, and N R Pace. 1994. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proceedings of the National Academy of Sciences* 91, no. 5: 1609. http://www.pnas.org/cgi/content/abstract/91/5/1609.

Bixley, Paul F., Alan W. Clotworthy, and Warren I. Mannington. 2009. Evolution of the Wairakei geothermal reservoir during 50 years of production. *Geothermics* 38, no. 1: 145-154. doi:10.1016/j.geothermics.2008.12.007. http://linkinghub.elsevier.com/retrieve/pii/S0375650508000916.

Borlee, Bradley R, Aaron D Goldman, Keiji Murakami, Ram Samudrala, Daniel J Wozniak, and Matthew R Parsek. 2010. Pseudomonas aeruginosa uses a cyclic-di-GMP-regulated adhesin to reinforce the biofilm extracellular matrix. *Molecular microbiology* 75, no. January: 827-842. doi:10.1111/j.1365-2958.2009.06991.x. http://www.ncbi.nlm.nih.gov/pubmed/20088866.

Bozzola, John J., and Lonnie Dee Russell. 1999. *Electron microscopy: principles and techniques for biologists*. 2 ed. Jones & Bartlett Publishers. http://books.google.com/books?id=RqSMzR-IXk0C&pgis=1.

Brock, T D, and H Freeze. 1969. Thermus aquaticus gen. n. and sp. n., a nonsporulating extreme thermophile. *Journal of bacteriology* 98, no. 1: 289-97. http://www.ncbi.nlm.nih.gov/pubmed/5781580.

Brock, Thomas D. 1985. Life at High Temperatures. *Science (New York, N.Y.)* 230, no. 4722: 132-138. doi:10.1126/science.230.4722.132. http://www.ncbi.nlm.nih.gov/pubmed/17842674.

Bruneau, Patricia M C, Donald A Davidson, Ian C Grieve, Iain M Young, and Naoise Nunan. 2005. The effects of soil horizons and faunal excrement on bacterial distribution in an upland grassland soil. *FEMS microbiology ecology* 52, no. 1: 139-44. doi:10.1016/j.femsec.2004.10.010. http://www.ncbi.nlm.nih.gov/pubmed/16329900.

Butcher, Samuel S. 1992. *Global biogeochemical cycles (pg. 4)*. http://books.google.com/books?id=WtmAFo8xLiUC&pgis=1.

Campbell, Kathleen A, Kristen L Cook, K A Rodgers, and P R L Browne. 2001. Silica from steam condensate alteration at Tikitere, New Zealand. In *26th Stanford Workshop on Geothermal Reservoir Engineering*.

Cashion, P, M A Holder-Franklin, J McCully, and M Franklin. 1977. A rapid method for the base ratio determination of bacterial DNA. *Analytical biochemistry* 81, no. 2: 461-6. http://www.ncbi.nlm.nih.gov/pubmed/907108.

Chan, M, R H Himes, and J M Akagi. 1971. Fatty acid composition of thermophilic, mesophilic, and psychrophilic clostridia. *Journal of bacteriology* 106, no. 3: 876-81. http://www.ncbi.nlm.nih.gov/pubmed/5567555.

Chow, Monica L, Christopher C Radomski, Joseph M McDermott, Julian Davies, and Paige E Axelrood. 2002. Molecular characterization of bacterial diversity in Lodgepole pine (Pinus contorta) rhizosphere soils from British Columbia forest soils differing in disturbance and geographic source. *FEMS microbiology ecology* 42, no. 3: 347-57. doi:10.1111/j.1574-6941.2002.tb01024.x. http://www.ncbi.nlm.nih.gov/pubmed/19709294.

Ciccarelli, Francesca D, Tobias Doerks, Christian von Mering, Christopher J Creevey, Berend Snel, and Peer Bork. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science (New York, N.Y.)* 311, no. 5765: 1283-7. doi:10.1126/science.1123061. http://www.ncbi.nlm.nih.gov/pubmed/16513982.

Clarridge, Jill E. 2004. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clinical microbiology reviews* 17, no. 4: 840-62, table of contents. doi:10.1128/CMR.17.4.840-862.2004. http://www.ncbi.nlm.nih.gov/pubmed/15489351.

Cohn, Ferdinand. 1999. Studies on bacteria. In *Milestones in microbiology 1546 to 1940*,Thomas D. Brock, 266. 2, revised. ASM Press. http://books.google.com/books?id=q5JHcs8w21gC&pgis=1.

Crocetti, G R, P Hugenholtz, P L Bond, A Schuler, J Keller, D Jenkins, and L L Blackall. 2000. Identification of polyphosphate-accumulating organisms and design of 16S rRNA-directed probes for their detection and quantitation. *Applied and environmental microbiology* 66, no. 3: 1175-82. http://www.ncbi.nlm.nih.gov/pubmed/10698788.

Curtis, T.P., W.T. Sloan, and J.W. Scannell. 2002. Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences of the United States of America* 99, no. 16: 10494. doi:10.1098/rstb.2006.1921. http://www.pnas.org/content/99/16/10494.full.

DeSantis, T Z, Philip Hugenholtz, N Larsen, M Rojas, E L Brodie, K Keller, T Huber, Daniel Dalevi, P Hu, and G L Andersen. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology* 72, no. 7: 5069-72. doi:10.1128/AEM.03006-05. http://www.ncbi.nlm.nih.gov/pubmed/16820507.

Dinsdale, Elizabeth a, Robert a Edwards, Dana Hall, Florent Angly, Mya Breitbart, Jennifer M Brulc, Mike Furlan, et al. 2008. Functional metagenomic profiling of nine biomes. *Nature* 452, no. 7187: 629-32. doi:10.1038/nature06810. http://www.ncbi.nlm.nih.gov/pubmed/18337718.

Dojka, Michael A., Philip Hugenholtz, Sheridan K. Haack, and Norman R. Pace. 1998. Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Applied and environmental microbiology* 64, no. 10: 3869-77. http://www.ncbi.nlm.nih.gov/pubmed/9758812.

Dow, J Maxwell, Lisa Crossman, Kim Findlay, Yong-Qiang He, Jia-Xun Feng, and Ji-Liang Tang. 2003. Biofilm dispersal in Xanthomonas campestris is controlled by cell-cell signaling and is required for full virulence to plants. *Proceedings of the National Academy of Sciences of the United States of America* 100, no. 19: 10995-1000. doi:10.1073/pnas.1833360100. http://www.ncbi.nlm.nih.gov/pubmed/12960398.

Drummond, AJ, B Ashton, M Cheung, J Heled, M Kearse, R Moir, S Stones-Havas, T Thierer, and A Wilson. 2009. Geneious. www.geneious.com.

Euzéby, J P. 1997. List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. *International journal of systematic bacteriology* 47, no. 2: 590-2. http://www.ncbi.nlm.nih.gov/pubmed/9103655.

Fox, G. E., K. R. Pechman, and C. R. Woese. 1977. Comparative Cataloging of 16S Ribosomal Ribonucleic Acid: Molecular Approach to Procaryotic Systematics. *International Journal of Systematic and Evolutionary Microbiology* 27, no. 1: 44-57. doi:10.1099/00207713-27-1-44. http://ijs.sgmjournals.org/cgi/doi/10.1099/00207713-27-1-44.

Fröhlich, J, and H König. 2000. New techniques for isolation of single prokaryotic cells. *FEMS microbiology reviews* 24, no. 5: 567-72. http://www.ncbi.nlm.nih.gov/pubmed/11077150.

Gans, Jason, Murray Wolinsky, and John Dunbar. 2005. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science (New York, N.Y.)* 309, no. 5739: 1387-90. doi:10.1126/science.1112665. http://www.ncbi.nlm.nih.gov/pubmed/16123304.

GeoNet. 2010. GeoNet - About Ngauruhoe. http://www.geonet.org.nz/volcano/activity/ngauruhoe/about.html.

Handelsman, J. 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews* 68, no. 4: 669. doi:10.1128/MBR.68.4.669–685.2004. http://mmbr.asm.org/cgi/content/full/68/4/669?view=long&pmid=15590779.

Henne, Anke, Holger Brüggemann, Carsten Raasch, Arnim Wiezer, Thomas Hartsch, Heiko Liesegang, Andre Johann, et al. 2004. The genome sequence of the extreme thermophile Thermus thermophilus. *Nature biotechnology* 22, no. 5: 547-53. doi:10.1038/nbt956. http://www.ncbi.nlm.nih.gov/pubmed/15064768.

Herrero, Antonia, and Enrique Flores. 2008. *The cyanobacteria: molecular biology, genomics, and evolution*. http://books.google.com/books?id=xgMahO1BXrQC&pgis=1.

Huber, Thomas, Geoffrey Faulkner, and Philip Hugenholtz. 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics*

(Oxford, England) 20, no. 14: 2317-9. doi:10.1093/bioinformatics/bth226. http://www.ncbi.nlm.nih.gov/pubmed/15073015.

Hugenholtz, Philip, B M Goebel, and N R Pace. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of bacteriology* 180, no. 18: 4765-74. http://www.ncbi.nlm.nih.gov/pubmed/9733676.

Hugenholtz, Philip, C Pitulle, K L Hershberger, and N R Pace. 1998. Novel division level bacterial diversity in a Yellowstone hot spring. *Journal of bacteriology* 180, no. 2: 366-76. http://www.ncbi.nlm.nih.gov/pubmed/9440526.

Hugenholtz, Philip, and N R Pace. 1996. Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. *Trends in Biotechnology* 14, no. 6: 190–197. http://linkinghub.elsevier.com/retrieve/pii/0167779996100251.

Ishøy, Thomas, Thomas Kvist, Peter Westermann, and Birgitte K Ahring. 2006. An improved method for single cell isolation of prokaryotes from meso-, thermo- and hyperthermophilic environments using micromanipulation. *Applied microbiology and biotechnology* 69, no. 5: 510-4. doi:10.1007/s00253-005-0014-x. http://www.ncbi.nlm.nih.gov/pubmed/16034558.

Iwabuchi, N. 2003. Relationships among colony morphotypes, cell-surface properties and bacterial adhesion to substrata in Rhodococcus. *Colloids and Surfaces B: Biointerfaces* 30, no. 1-2: 51-60. doi:10.1016/S0927-7765(03)00036-5. http://linkinghub.elsevier.com/retrieve/pii/S0927776503000365.

Janssen, Peter H. 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Applied and environmental microbiology* 72, no. 3: 1719-28. doi:10.1128/AEM.72.3.1719-1728.2006. http://www.ncbi.nlm.nih.gov/pubmed/16517615.

Jones, Brain, Robin W Renaut, and Michael R Rosen. 1996. High-temperature (>90 C) calcite precipitation at Waikite Hot Springs, North Island, New Zealand. *Journal of the Geological Society* 153, no. 3: 481-496. doi:10.1144/gsjgs.153.3.0481. http://jgs.lyellcollection.org/cgi/doi/10.1144/gsjgs.153.3.0481.

Kaeberlein, T, K Lewis, and S S Epstein. 2002. Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science (New York, N.Y.)* 296, no. 5570: 1127-9. doi:10.1126/science.1070633. http://www.ncbi.nlm.nih.gov/pubmed/12004133.

Kanagawa, T. 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of Bioscience and Bioengineering* 96, no. 4: 317-323. doi:10.1016/S1389-1723(03)90130-7. http://linkinghub.elsevier.com/retrieve/pii/S1389172303901307.

LaPaglia, C., and PL Hartzell. 1997. Stress-induced production of biofilm in the hyperthermophile Archaeoglobus fulgidus. *Applied and Environmental Microbiology* 63, no. 8: 3158. http://aem.asm.org/cgi/content/abstract/63/8/3158.

Lane, D J, E Stackebrandt, and M Goodfellow. 1991. 16S/23S rRNA Sequencing. In *Nucleic Acids Techniques in Bacterial Systematics*,E. Stackebrandt and M. Goodfellow, 115-175. John Wiley and Sons.

Lesaulnier, C, and D Papamichail. 2008. Elevated atmospheric CO2 affects soil microbial diversity associated with trembling aspen. *Environmental Microbiology* 10, no. 4: 926-941. http://www3.interscience.wiley.com/journal/119419967/abstract.

Levsky, Jeffrey M, and Robert H Singer. 2003. Fluorescence in situ hybridization: past, present and future. *Journal of cell science* 116, no. Pt 14: 2833-8. doi:10.1242/jcs.00633. http://www.ncbi.nlm.nih.gov/pubmed/12808017.

Lipson, David A, and Steven K Schmidt. 2004. Seasonal changes in an alpine soil bacterial community in the colorado rocky mountains. *Applied and environmental microbiology* 70, no. 5: 2867-79. doi:10.1128/AEM.70.5.2867–2879.2004. http://www.ncbi.nlm.nih.gov/pubmed/15128545.

Ludwig, Wolfgang, Oliver Strunk, Ralf Westram, Lothar Richter, Harald Meier, Yadhukumar, Arno Buchner, et al. 2004. ARB: a software environment for sequence data. *Nucleic acids research* 32, no. 4: 1363-71. doi:10.1093/nar/gkh293. http://www.ncbi.nlm.nih.gov/pubmed/14985472.

Ludwig, Wolfgang, and Hans-Peter Klenk. 2005. Overview: A Phylogenetic Backbone and Taxonomic Framework for Procaryotic Systematics. In *Bergey's Manual® of Systematic Bacteriology*, 49-66. New York: Springer-Verlag. doi:10.1007/0-387-28021-9. http://www.springerlink.com/index/10.1007/0-387-28021-9.

Mesbah, M., U. Premachandran, and W. B. Whitman. 1989. Precise Measurement of the G+C Content of Deoxyribonucleic Acid by High-Performance Liquid Chromatography. *International Journal of Systematic and Evolutionary Microbiology* 39, no. 2: 159-167. doi:10.1099/00207713-39-2-159. http://ijs.sgmjournals.org/cgi/doi/10.1099/00207713-39-2-159.

Morozkina, E. V., E. S. Slutskaya, T. V. Fedorova, T. I. Tugay, L. I. Golubeva, and O. V. Koroleva. 2010. Extremophilic microorganisms: Biochemical adaptation and biotechnological application (review). *Applied Biochemistry and Microbiology* 46, no. 1: 1-14. doi:10.1134/S0003683810010011. http://www.springerlink.com/index/10.1134/S0003683810010011.

Oshima, M, and T Ariga. 1975. Omega-cyclohexyl fatty acids in acidophilic thermophilic bacteria. Studies on their presence, structure, and biosynthesis using precursors labeled with stable isotopes and radioisotopes. *Journal of Biological Chemistry* 250, no. 17: 6963-6968.

Pace, N R. 1997. A molecular view of microbial diversity and the biosphere. *Science* 276, no. 5313: 734. http://www.sciencemag.org/cgi/content/abstract/sci;276/5313/734.

Pollack, Gerald H, Xavier Figueroa, and Qing Zhao. 2009. Molecules, water, and radiant energy: new clues for the origin of life. *International journal of molecular sciences* 10, no. 4: 1419-29. doi:10.3390/ijms10041419. http://www.ncbi.nlm.nih.gov/pubmed/19468316.

Polz, Martin F., and Colleen M. Cavanaugh. 1998. Bias in template-to-product ratios in multitemplate PCR. *Applied and environmental microbiology* 64, no. 10: 3724-30. http://aem.asm.org/cgi/content/abstract/64/10/3724.

Portillo, M. C., and J. M. Gonzalez. 2008. Members of the Candidate Division OP10 are spread in a variety of environments. *World Journal of Microbiology and Biotechnology* 25, no. 2: 347-353. doi:10.1007/s11274-008-9895-z. http://www.springerlink.com/index/10.1007/s11274-008-9895-z.

Pruesse, Elmar, Christian Quast, Katrin Knittel, Bernhard M Fuchs, Wolfgang Ludwig, Jörg Peplies, and Frank Oliver Glöckner. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids research* 35, no. 21: 7188-96. doi:10.1093/nar/gkm864. http://www.ncbi.nlm.nih.gov/pubmed/17947321.

Rappé, Michael S, and Stephen J Giovannoni. 2003. The uncultured microbial majority. *Annual review of microbiology* 57: 369-94. doi:10.1146/annurev.micro.57.030502.090759. http://www.ncbi.nlm.nih.gov/pubmed/14527284.

Rychlik, W, W J Spencer, and R E Rhoads. 1990. Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic acids research* 18, no. 21: 6409-12. http://www.ncbi.nlm.nih.gov/pubmed/2243783.

Sambrook, Joseph, and David William Russell. 2001. *Molecular cloning: a laboratory manual*. 3 ed. Cold Spring Harbor Laboratory Press. http://books.google.com/books?hl=en&lr=&id=YTxKwWUiBeUC&pgis=1.

Schloss, Patrick D, and Jo Handelsman. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and environmental microbiology* 71, no. 3: 1501-6. doi:10.1128/AEM.71.3.1501-1506.2005. http://www.ncbi.nlm.nih.gov/pubmed/15746353.

Schloss, Patrick D. 2009. A high-throughput DNA sequence aligner for microbial ecology studies. *PloS one* 4, no. 12: e8230. doi:10.1371/journal.pone.0008230. http://www.ncbi.nlm.nih.gov/pubmed/20011594.

Shen, P Y, E Coles, J L Foote, and J Stenesh. 1970. Fatty acid distribution in mesophilic and thermophilic strains of the genus Bacillus. *Journal of bacteriology* 103, no. 2: 479-81. http://www.ncbi.nlm.nih.gov/pubmed/5432012.

Simkin, Tom, and Lee Siebert. 1981. *Volcanoes of the World: A Regional Directory, Gazetteer, and Chronology of Volcanism During the Last 10,000 Years*. Hutchinson Ross Pub. Co. http://books.google.com/books?id=HXMHfXQyS8wC&pgis=1.

Smith, Jacques J., Lemese Ah Tow, William Stafford, Craig Cary, and Donald A. Cowan. 2006. Bacterial Diversity in Three Different Antarctic Cold Desert Mineral Soils. *Microbial Ecology* 51, no. 4: 413 - 421. http://www.jstor.org/stable/25153334.

Soo, Rochelle M, Susanna A Wood, Joseph J Grzymski, Ian R McDonald, and S Craig Cary. 2009. Microbial biodiversity of thermophilic communities in hot mineral soils of Tramway Ridge, Mount Erebus, Antarctica. *Environmental microbiology* 11, no. 3: 715-28. doi:10.1111/j.1462-2920.2009.01859.x. http://www.ncbi.nlm.nih.gov/pubmed/19278453.

Stott, Matthew B, Michelle a Crowe, Bruce W Mountain, Angela V Smirnova, Shaobin Hou, Maqsudul Alam, and Peter F Dunfield. 2008. Isolation of novel bacteria, including a candidate division, from geothermal soils in New Zealand. *Environmental microbiology* 10, no. 8: 2030-41. doi:10.1111/j.1462-2920.2008.01621.x. http://www.ncbi.nlm.nih.gov/pubmed/18422642.

Svetashev V, M V Vysotskii, E P Ivanova, and V V Mikhailov. 1995. Cellular fatty acids of Alteromonas species. *Systematic and applied microbiology* 18, no. 1: 37-43. http://cat.inist.fr/?aModele=afficheN&cpsidt=3560556.

Swannell, R P, K Lee, and M McDonagh. 1996. Field evaluations of marine oil spill bioremediation. *Microbiological reviews* 60, no. 2: 342-65. http://www.ncbi.nlm.nih.gov/pubmed/8801437.

Tindall, KR, and TA Kunkel. 1988. Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. *Biochemistry* 27, no. 16: 6008. http://www.ncbi.nlm.nih.gov/pubmed/2847780.

Torsvik, V, J Goksøyr, and F L Daae. 1990. High diversity in DNA of soil bacteria. *Applied and environmental microbiology* 56, no. 3: 782-7. http://www.ncbi.nlm.nih.gov/pubmed/2317046.

Tyson, Gene W, Jarrod Chapman, Philip Hugenholtz, Eric E Allen, Rachna J Ram, Paul M Richardson, Victor V Solovyev, Edward M Rubin, Daniel S Rokhsar, and Jillian F Banfield. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428, no. 6978: 37-43. doi:10.1038/nature02340. http://www.ncbi.nlm.nih.gov/pubmed/14961025.

Urbach, Ena, Kevin L Vergin, Lei Young, Ariel Morse, Gary L Larson, and Stephen J Giovannoni. 2001. Unusual bacterioplankton community structure in ultra-oligotrophic Crater Lake. *Limnol. Oceanogr.,* 46, no. 3: 557-572.

Wallace, R B, J Shaffer, R F Murphy, J Bonner, T Hirose, and K Itakura. 1979. Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic acids research* 6, no. 11: 3543-57. http://www.ncbi.nlm.nih.gov/pubmed/158748.

Wayne, LG, DJ Brenner, RR Colwell, PAD Grimont, O Kandler, MI Krichevsky, LH Moore, et al. 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *International Journal of Systematic and Evolutionary Microbiology*. Soc General Microbiol. http://ijs.sgmjournals.org/cgi/reprint/37/4/463.pdf.

Weisburg, W G, S M Barns, D A Pelletier, and D J Lane. 1991. 16S ribosomal DNA amplification for phylogenetic study. *Journal of bacteriology* 173, no. 2: 697-703. http://www.ncbi.nlm.nih.gov/pubmed/1987160.

White, O, J A Eisen, J F Heidelberg, E K Hickey, J D Peterson, R J Dodson, D H Haft, et al. 1999. Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1. *Science (New York, N.Y.)* 286, no. 5444: 1571-7. http://www.ncbi.nlm.nih.gov/pubmed/10567266.

Whitman, William B, David C Coleman, and William J Wiebe. 1998. Perspective Prokaryotes: The unseen majority 95, no. June: 6578-6583.

Woese, C R, and GE Fox. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences* 74, no. 11: 5088. http://www.pnas.org/cgi/content/abstract/74/11/5088.

Woese, C R. 1987. Bacterial Evolution. *Microbiological reviews* 51, no. 2: 221. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=373105.

Wu, Xin, Wanyan Xi, Wenjin Ye, and Hong Yang. 2007. Bacterial community composition of a shallow hypertrophic freshwater lake in China, revealed by 16S rRNA gene sequences. *FEMS microbiology ecology* 61, no. 1: 85-96. doi:10.1111/j.1574-6941.2007.00326.x. http://www.ncbi.nlm.nih.gov/pubmed/17506827.

Zengler, Karsten, Gerardo Toledo, Michael S Rappé, James Elkins, Eric J Mathur, Jay M Short, and Martin Keller. 2002. Cultivating the uncultured. *Proceedings of the National Academy of Sciences of the United States of America* 99, no. 24: 15681-6. doi:10.1073/pnas.252630999. http://www.ncbi.nlm.nih.gov/pubmed/12438682.

von Wintzingerode, F, U B Göbel, and E Stackebrandt. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS microbiology reviews* 21, no. 3: 213-29. http://www.ncbi.nlm.nih.gov/pubmed/9451814.

Appendix 1: List of Chemicals and Reagents

Table 1. Active Ingredients per capsule of Clinicians B Complex Forte (referred to as Vitamins in text).

component	mass
Vitamin B1 (as thiamine hydrochloride)	25 mg
Vitamin B2 (as riboflavin)	10 mg
Vitamin B3 (as nicotinamide)	25 mg
Vitamin B3 (as nicotininc acid)	2.5 mg
Vitamin B5 (as calcium pantothenate)	50 mg
Vitamin B6 (as pyridoxine hydrochlride)	25 mg
Vitamin B12 (as cyanocobalamin)	12.5 µg
Biotin	12.5 µg
Choline (as choline bitartrate)	50 mg
Folic Acid	200 µg
Inositol	50 mg
PABA (para-aminobezoic acid)	25 mg
Calcium (from calcium pantothenate)	5 mg

Table 2. Reagents used in this study

Reagent	Manufacturer	Grade
Phytagel™	Sigma	analar
Sodium carboxylmethylcellulose (CMC)	Hercules	technical
Pectin type 115	CP Kelco	technical
Xanthan® CC	CP Kelco	technical
Xylan from Beechwood > 90% xylan residue	Sigma	technical
Avicel®	Fluka	analar
Starch	Sigma	technical
Alginic acid, sodium salt	Aldrich	analar
Glycogen from Bovine liver type IX	Sigma	technical
Chitin from crab shells	Sigma	technical
Dextrin from corn type I	Sigma	technical
Locust bean gum (glactomannan) from *Ceratonia siliqua* seeds	Sigma	technical
Whatman filter paper	Whatman	n/a
D-Glucose	Sigma	analar
D-(+)-maltose anhydrous monohydrate	Sigma	analar
D-(-)-fructose	Sigma	analar
D-(+)-galactose	Sigma	analar
D-(+)-mannose	Fluka	analar
DL-arabinose	Fluka	analar
D-(-)-ribose	Sigma	analar
D-(+)- cellobiose	Fluka	analar
D-(+)-trehalose dihydrate	Sigma	analar
Lactose	Difco	technical
Sucrose	Fluka	technical

D-N-acetylglucosamine	Aldrich	analar
D-(+)- galacturonic acid	Fluka	analar
Methanol	APS Chemicals	analar
Ethanol	Merck	analar
1-propanol	Merck	analar
2-propanol	Merck	analar
Sodium pyruvate	Sigma	analar
Sodium citrate	Sigma	analar
Lignin extract	Pure Power Technologies	n/a
Yeast extract	Difco	technical
Agarose	Seakem LE	analar
$FeSO_4.7H_2O$	Sigma	analar
Sodium EDTA	Sigma	analar
Nitrilotriacetic acid	Scharlan	analar
$Fe(NH_4)_2(SO_4)_2.6H_2O$	BDH	analar
Na_2SeO_3	BDH	analar
$CoCl_2.6H_2O$	APS Chemicals	analar
$MnCl_2.4H_2O$	APS Chemicals	analar
$Na_2MoO_4.2H_2O$	Hopkin & Williams Ltd	analar
$NaWO_4.2H_2O$	BDH	analar
$ZnSO_4.7H_2O$	APS Chemicals	analar
$Al_2(SO_4)_3.18H_2O$	APS Chemicals	analar
$NiCl_2.6H_2O$	APS Chemicals	analar
H_3BO_3	APS Chemicals	analar
$CuSO_4.5H_2O$	APS Chemicals	analar
KOH	APS Chemicals	analar
Hydrochloric acid	BDH	analar
Sodium citrate	Sigma	analar
Na_2HPO_4	APS Chemicals	analar
NaH_2PO_4	APS Chemicals	analar

Acetic acid	BDH	analar
$MgSO_4.7H_2O$	APS Chemicals	analar
$NaCH_3COOH$	APS Chemicals	analar
K_2HPO_4	APS Chemicals	analar
$(NH_4)_2SO_4$	APS Chemicals	analar
$NaHCO_3$	APS Chemicals	analar
$CaCl_2.2H_2O$	APS Chemicals	analar
Luria-Bertani Agar	Difco	Technical
X-Gal (bromo-chloro-indolyl-galactopyranoside)	Bioline	Analar
IPTG (isopropyl ß-D-1-thiogalactopyranoside)	Progen BioSciences	analar
Ampicillin, sodium salt	Sigma	n/a

Appendix 2: Previous 16S rRNA gene sequences from sample sites

Previous culture-independent screening of geothermal systems in the TVZ has been conducted previously by staff at the Extremophiles Research Group at GNS Science. OP10 phylotypes were detected at sample sites WKT32 (Waikite), Waipahihi (WPH03) and Mt Ngauruhoe (NGA2) in clone libraries using universal bacterial primers that target the 16S rRNA gene. In addition, OP10 strains P488 and T49 were enriched and isolated from sample sites WKT32 (Waikite) and TT13.4 (Tikitere) respectively. The general methods for generating these culture-dependent and culture-independent data are identical to those outlined in the Materials and Methods sections of this thesis except universal bacterial primers 27F (5'- AGA GTT TGA TCM TGG CTC AG -3') and 1492R ('5- TAC GGY TAC CTT GTT ACG ACT T -3') where used in place of OP10 specific primers (Lane, Stackebrandt & Goodfellow, 1991).

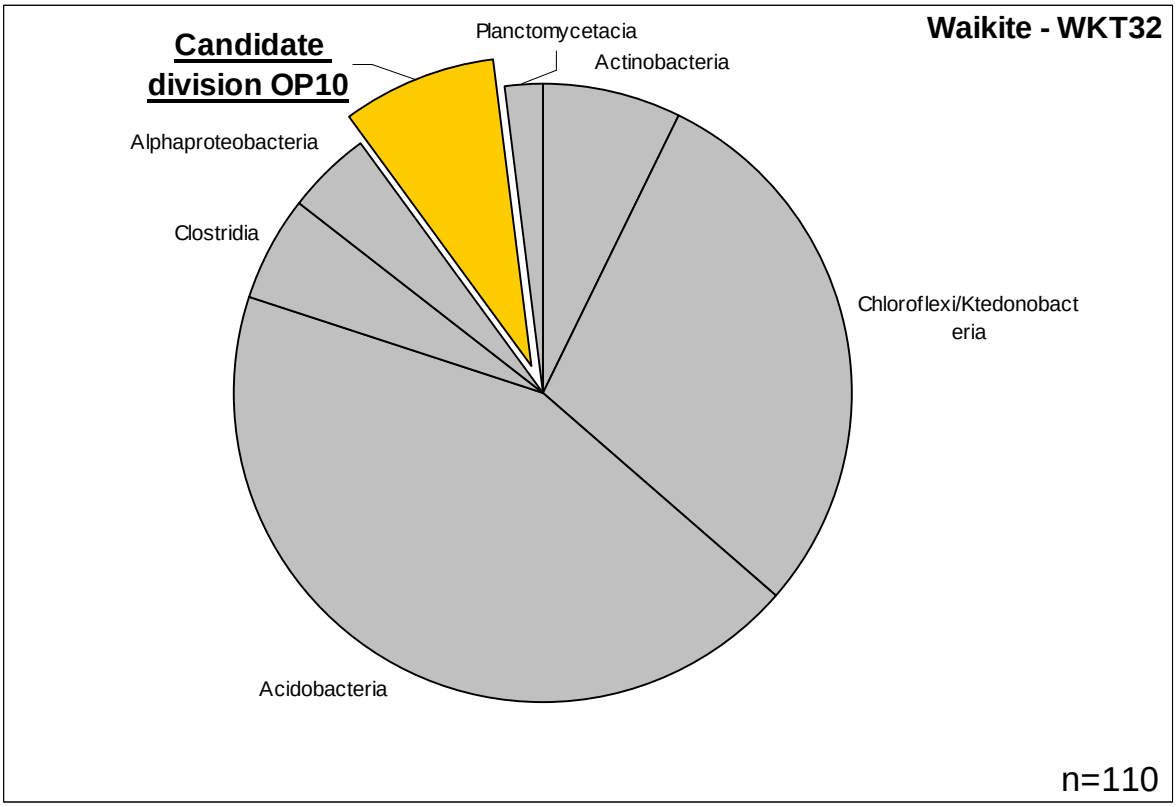


Figure 1. Pie chart showing the distribution of bacterial phyla detected at the WKT32 sample site (Waikite) (Stott *et al.*, 2008). OP10 phylotypes contributed to approximately 10% of the total phylotypes detected (n=110). OP10 strain P488 was enriched and isolated from this location.

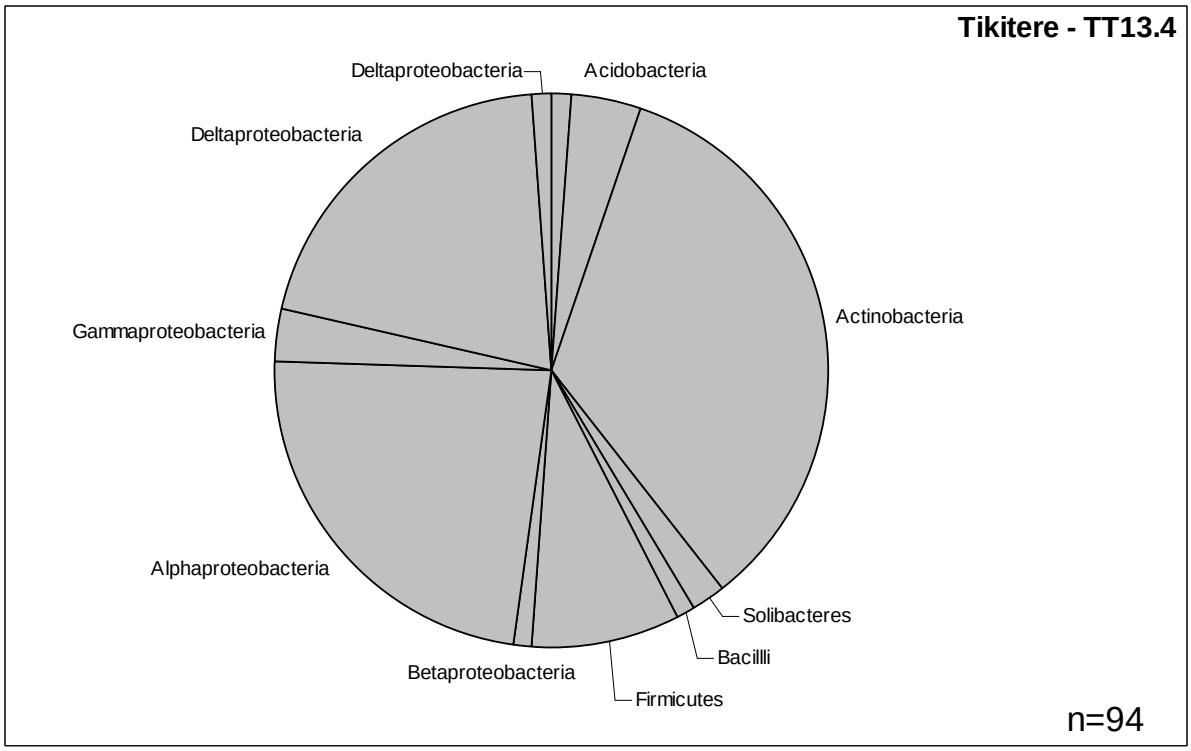


Figure 2. Pie chart showing the distribution of bacterial phyla detected at the TT13.4 sample site (Tikitere) (Stott *et al.*, 2008). No OP10 phylotypes were detected using universal bacterial 16S rRNA gene primers (n=94). However, OP10 strain T49 was enriched and isolated from this location.

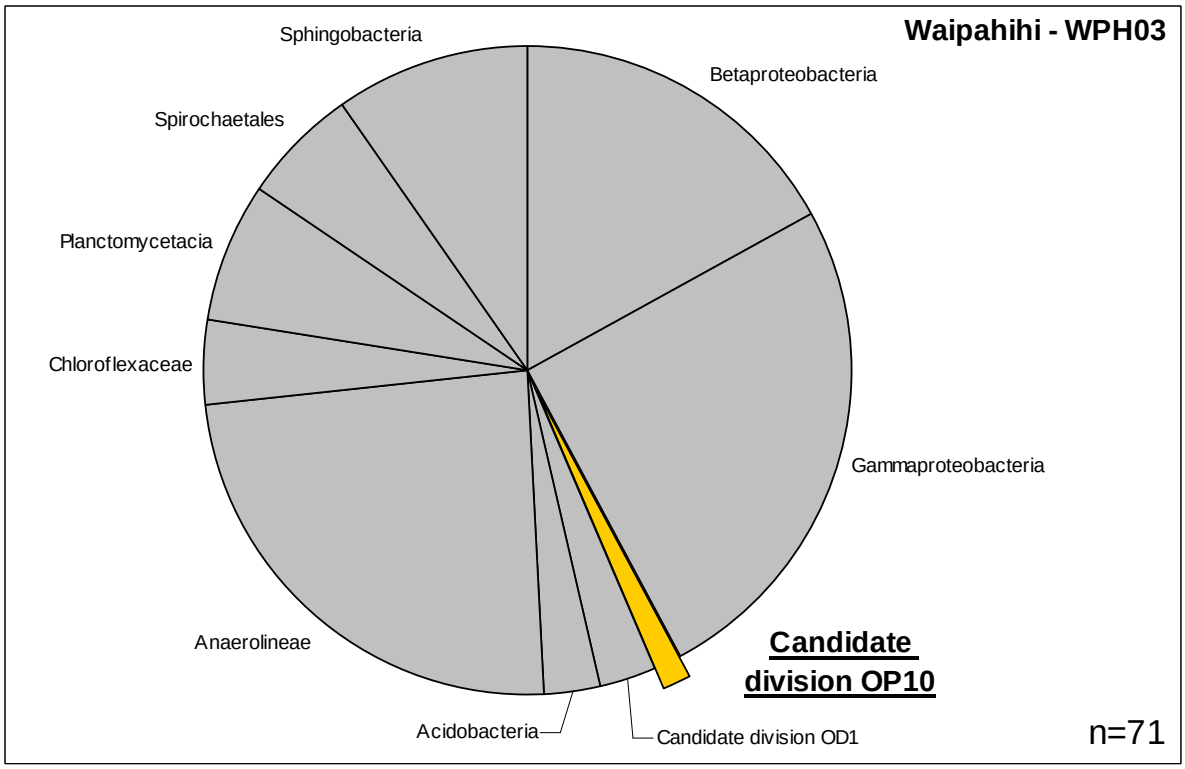


Figure 3. Pie chart showing the distribution of bacterial phyla detected at the WPH03 sample site (Waipahihi). OP10 phylotypes contributed to >1% of the total phylotypes detected (n=71).

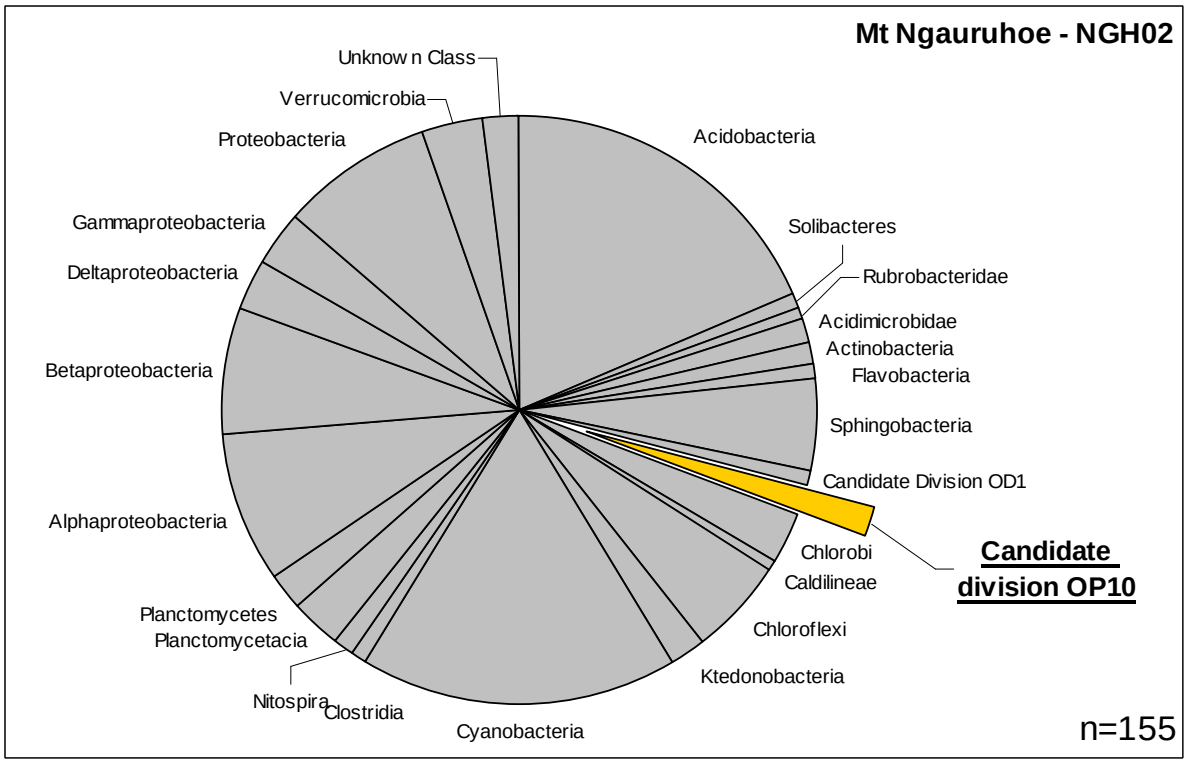


Figure 4. Pie chart showing the distribution of bacterial phyla detected at the NGH02 sample site (Ngauruhoe). OP10 phylotypes contributed to approximately 2% of the total phylotypes detected (n=155).

Appendix 3: Reference phylogenetic tree

Appendix 4: Reference phylogenetic tree with new sequences