Cultivating Microbial Dark Matter

A combination of traditional and molecular techniques

A thesis

submitted partial fulfilment

of the requirements for the degree

of

Master of Environmental Science (Research)

Microbial Ecology

at

The University of Waikato

by

Charlotte King

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2024

Summary

The services bacteria offer the world are immense and impact people's lives daily. They were first described by Antoni van Leeuwenhoek in 1676, but bacteria have been utilised in the medical, agricultural, and food and beverage industries even before they were formally identified. Microbes' roles in the environment for cycling of nutrients, climate regulation and pollution are becoming more well understood, and they are also utilised to answer a wide variety of research questions in disparate fields. Traditional cultivation techniques were initially developed to study bacteria and understand their functions as isolates. Such cultivation techniques have become rarer as new molecular and bioinformatic tools have been developed. These cultivation-independent techniques have allowed the study of microbial life genetic data which led to the finding that 99% of all microbial species were yet to be cultured and studied in a laboratory. The applications of these unstudied organisms could be unprecedented, though their cultivation remains challenging due to the requirement of specific nutrients that are hard to identify, very slow growth rates, environmental conditions that are difficult to replicate in a laboratory setting, live in symbiosis with other organisms, or live in extreme climates. Therefore, much of the information may be overlooked and hidden in this microbial dark matter. Methods such as metagenomics, metagenome-assembled genomes (MAGs), and single amplified genomes (SAGs) have provided us with genomic data for previously uncharacterised microbes from diverse environments. However, such data is still limited. Omics tools can provide insights into what is living there and what it can potentially do. However, this data is regularly incomplete due to the processes involved in sequencing, and conclusions are often drawn from associations relating to cultured organisms. This can lead to potentially spurious conclusions as a gene may behave differently in different species, and the absence of genes may depend on how the data was analysed. Therefore, cultured isolates of the organism are vital. To combat this issue of an uncultivated majority, it is important to have a targeted approach to culture, what is wanted, not just what is easy.

This thesis aimed to provide an easy-to-follow, straightforward workflow to cultivate target environmental organisms using standard molecular tools and methods readily accessible in any standard microbiological laboratory. The idea is to isolate and culture what is specifically wanted or needed, not just what is easy. Specifically, it combined classical cultivation and molecular techniques to target a sulphur-reducing, obligate anaerobe from an environmental sample—namely a *Desulfurella sp.* from a geothermal pool in Rotorua, New Zealand.

In order to provide the desired targeted approach, specific primers needed to be designed and assessed to first find the right sampling site for the initial inoculum and second to ensure the target is not lost throughout the culturing process. I designed and evaluated primers for targeting specific taxa at different stringency levels. Initially, I tested published primers (EPS_F/EPS_R, Gittel et al., 2012) aimed at Epsilonproteobacteria, which were reclassified into the Campylobacterota phylum (including *Desulfurella sp.*). These primers did not target *Desulfurella sp.*, so I modified them (EPS_FM/EPS_RL) to produce a longer amplicon for broader Campylobacterota detection. However, these also amplified non-target species (e.g., Desulfobacterota). Finally, I designed a third set of primers (DS_F/DS_R) specific to *Desulfurella sp.*, confirmed by in silico testing. These primers allow assessment of *Desulfurella sp.* and broader Campylobacterota before sequencing.

The second experimental chapter focuses on enriching the target using the previously described primers, classical cultivation techniques, and molecular tools. Site selection was based on the 1000 Springs Project (Power et al., 2018). Culturing was performed originally with DSMZ Desulfurella medium and later modified, investigating pH (4 and 6) and temperature (50°C and 30°C) variations and adapting it for solid media by replacing elemental sulphur with sodium polysulfide. I used a vacuum food sealer to cultivate anaerobes, an accessible and portable alternative to expensive anaerobic chambers. Automated Ribosomal Intergenic Spacer Analysis (ARISA) was used to assess axenicity before sending samples for 16S rRNA gene and whole genome sequencing. *Desulfurella* was detected in each culture using species-specific primers (DS_F/DS_R), enabling its phylogenetic placement. However, despite *Desulfurella* presence in

the 16S rRNA gene amplicon sequencing, a metagenome-assembled genome (MAG) could not be produced. This thesis discusses lessons learned and suggests improvements for future studies.

Acknowledgments

I would like to express my deepest gratitude to Craig Cary. He was a true source of inspiration, always brimming with insight and enthusiasm. I will deeply miss his stories, his invaluable guidance, and the time we spent together. The knowledge he shared with me is something I will forever remember. His unwavering passion for science and for life itself left a lasting impression on everyone around him, and his memory will continue to inspire me always.

To Huw Richards, who I genuinely would not have gotten through this without, I cannot thank you enough. You have guided me in so many ways throughout this process and encouraged me every step of the way. Giving me not only practical solutions but the confidence to get it done.

To Charles Lee for coming in as my supervisor at the last minute. You have provided a fresh look over this work and made many helpful suggestions. I will miss your straightforward feedback and advice, as well as your constant encouragement to be better and achieve more. As well as everyone else in TRU for the constant support and push to succeed.

To all of my family and friends for putting up with my complaints and encouraging me to keep going.

Table of Contents

List of Figures

List of Tables

Chapter 1: Literature Review

1.1 The Earth Microbiome

The Earth microbiome (EMB) is the vast array of microorganisms that inhabit the planet, including bacteria, viruses, and fungi. These microbes are critical for nutrient cycling, decomposition, climate regulation and other processes that maintain the health and function of Earth's ecosystems. Advancements in DNA sequencing have allowed for the in-depth study of the EMB, unlocking the network of microbial interactions and functions far more complex than previously thought. This led to the concept of the "uncultured microbial majority" gaining significance in 2014 (Hedlund et al., 2014) , suggesting that most microbes are yet to be grown in a laboratory. Indeed, estimates suggest that up to 99% of microbial species have not been cultivated (Liu et al., 2022).

Microorganisms are the most abundant form of life on Earth and are found everywhere, from deep-sea hydrothermal vents to the Antarctic dry valleys and even in the atmosphere. Most of the uncultured majority is in extreme environments and low-energy ecosystems, which usually comprise the Candidate Phyla Radiation (CPR) (Naud et al., 2022). Studying the EMB provides information on the functioning of ecosystems and their applications in fields such as biotechnology, medicine, and agriculture (Lewis et al., 2021). However, the lack of cultured representatives from many environments presents issues and routinely leads to having to make inferences on associations from incomplete data.

In 2010, some of the world's leading microbial ecologists and bioinformaticians came together to discuss the challenge of how best to use next-generation sequencing technologies to characterise the microbial world. They aimed to examine the fundamental questions in microbial ecology that advances in sequencing could address, so The Earth Microbiome Project (EMP) began (Gilbert et al., 2010). The EMP is a collaborative effort to sequence and characterise microbial life on Earth. The team on the project set out to produce a "global Gene Atlas" that

would describe each biome's environmental metabolic models by analysing 200,000 samples using metagenomics, metabolomics and amplicon sequencing. This gave rise to the website, an open source for data analysis and visualisation of processed information. This project and subsequent database are an example of how sequencing has significantly changed and expanded our knowledge of the microbes that inhabit Earth. There is still a lot to achieve in this space, with the website stating that they currently have too many samples to process, so they cannot take on anything new. This shows how much information about microorganisms was inaccessible before the omics era (Gilbert et al., 2014; Thompson et al., 2017).

1.1.1 The Tree of Life

In 1859, Charles Darwin first published the tree of life idea in "On the Origin of Species" (Darwin & Kebler, 1859), which has continued to expand due to genomic sequencing from previously unexamined environments (Hug et al., 2016). With capabilities now allowing the greater generation of genome sequences, more and more organisms are being identified along with their metabolic capabilities, allowing them to be placed onto the tree. In 2016, Hug et al. used genomic data from over 1000 uncultivated organisms to generate a far more diverse tree of life, including Eukarya, Bacteria and Archaea. This provided an overview of the significant lineage's diversity and highlighted the importance of the representatives, which lack isolates as they make up a substantial radiation (Figure 1.1). Prokaryotes have increased from hundreds to thousands of phyla due to genomic data. However, this can differ significantly depending on the estimating technique. 16S rRNA gene sequencing as of 2021 has the total number of species at approximately 400,000 (Lewis et al., 2021) , which could be considerably higher, with only approximately 14,000 being cultivated isolates. Currently, GTDB has genomes for a total of 113,104 species, for both bacteria and archaea, with over 70% represented solely through MAGs and SAGs. This points to a significant bias of cultivated representatives regarding the Tree of Life, which is constantly evolving to fit new species.

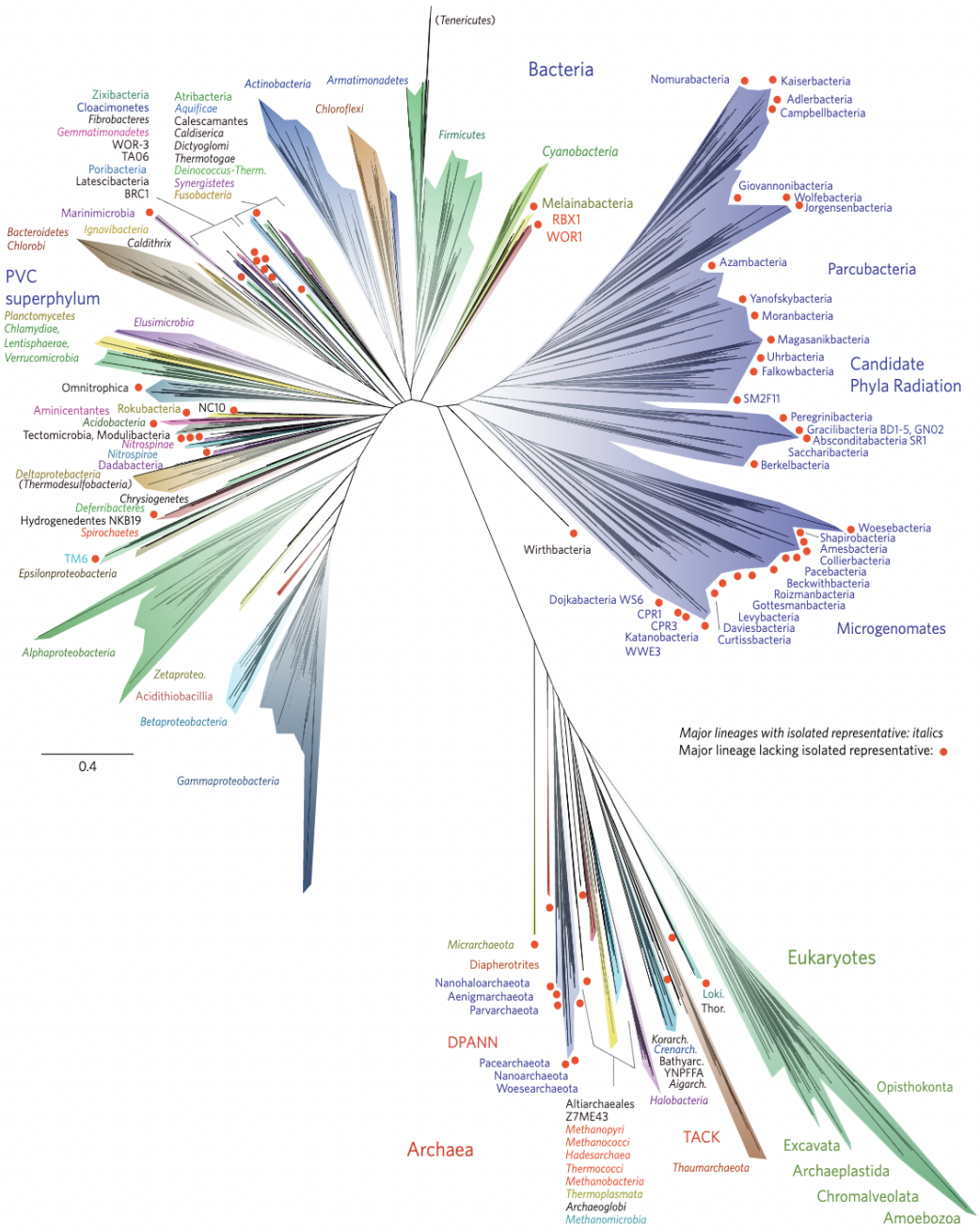


Figure 1. 1. The Tree of Life by Hug et al. (2016) depicts the total diversity shown by sequenced genomes. It includes all five Eukaryotic supergroups, 92 bacterial phyla, and 26 archaeal. Non-italicised names and red dots highlight the lineages that did not have isolated representatives (when this was published).

1.1.2 Branch Breakers

Branch beakers are newly found or classified taxa that cause a branching point (or node) on a phylogenetic tree, where a single lineage evolves into a distinct new one. Solden et al. (2016)

discuss branch breakers and how these challenge the three domains of life. The three domains of life are the model where Earth's biodiversity is resolved into monophyletic groups of Archaea, Bacteria and Eukarya. This still holds, however, there is continuing conflict to this such as Williams and Embley's proposed 'eocyte' hypothesis which is a two-domain model that has eukaryotes coming from within the archaeal radiation (Williams & Embley, 2014). Unexpected metabolisms have been found in new phylogenetic places, such as archaea, which can do things previously thought only bacteria or eukarya could do. Small genomes and metabolic interdependencies are also prevalent. Given the unusual environments that many of these microbes inhabit, these uncultivated organisms may contain proteins with unknown functions that allow them to interact with the environment in new ways (Solden et al., 2016). Branch breakers continuously change the tree and discoveries could challenge the three domains of life and are an important study area (Pavlopoulos et al., 2023).

1.2 Microbial Dark Matter

Microbial dark matter (MDM) refers to many microorganisms that cannot be easily cultured or studied using traditional laboratory methods, as previously discussed (Solden et al., 2016). These microorganisms are often found in diverse environments and are thought to play roles in various ecological processes, such as nutrient cycling, and may also have applications in biotechnology. The term "dark matter" is used because, like the hypothetical form of matter in physics, these microorganisms are difficult to detect and study, but their presence is confirmed by cultivation-independent methods such as metagenomics or single amplified genomes. They account for much of Earth's biomass, but basic ecological and metabolic principles are unknown (Yang et al., 2023). When 16S rRNA gene sequencing for environmental samples emerged in the mid-1970s, this revealed significant numbers of taxa that had yet to be cultivated, which could be made up of life forms that could provide an array of valuable properties (Handelsman, 2004). This included, in 1977, Carl Woses and his postdoc George Fox discovering the "third domain

of life", 'archaebacteria' (now called Archaea) (Woese & Fox, 1977) and shedding light on the discrepancy between those microorganisms that could be cultured in a laboratory and those that were present in the environment (Lewis et al., 2021). MDM was once said to be 'impossible' to cultivate, which has been proven wrong by labour-intensive culturing, which has seen the successful cultivation of some fastidious microbes (van Teeseling & Jogler, 2021a).

The knowledge of the microbial world is heavily skewed by what can be cultured and is therefore primarily based on well-studied model organisms such as *E. coli* (Hufnagel et al., 2015). This problem has been addressed through advances in genomic sequencing technology with metagenomics and single-cell genomics, which contain their strengths and challenges (Pavlopoulos et al., 2023). Studying the genomic data of these organisms has changed the understanding of cell biology, physiology, and evolution. However, gaining isolates for them is vital for confirming these functions. Considering the number of services microbes offer today and the large amount of understudied microbial dark matter, it is a topic that requires much further study (van Teeseling & Jogler, 2021).

1.3 Nomenclature for Uncultivated Majority

As it is now apparent that microbial diversity has previously been heavily underrepresented due to cultivation issues of microbial dark matter, there lies a problem with the nomenclature of these organisms. The International Code of Nomenclature of Prokaryotes (ICNP) only recognises cultures as 'type material', thereby preventing the naming of uncultivated microorganisms (Hedlund et al., 2022; Murray et al., 2020). This creates issues and confusion in the literature, as there is no standard way to name uncultivated prokaryotes. Murray et al. provide a consensus statement for a proposed "Roadmap for naming uncultivated Archaea and Bacteria", where they emphasise the urgent need to reconsider the rules of nomenclature to describe archaeal and bacterial diversity in their entirety accurately. By recognising this, the intent is not to rule out cultivation efforts for fastidious microbes but to create a more standardised naming system to

avoid further confusion. However, it could start a domino effect, leading to less emphasis on cultivation. The statement proposes two different ways to develop a system for naming microbial dark matter that would allow them to be classified using MAGs or SAGs with high fidelity. This would then enable such microbes to have predicted characteristics and, therefore, be able to be linked contextually to ecological and environmental scenarios (Murray et al., 2020). While the proposal of the "SeqCode" (Hedlund et al., 2022) provides recommendations on the minimal standard of DNA sequences, is reproducible and facilitates communication of all prokaryotes regardless of cultivability across disciplines, it does not fix the problems that arise with MAGs/SAGs and has the potential to make the problem bigger if misused. Sharing nomenclature could also mean no distinction between MAGs and isolates, treating them as the same.

1.4 Cultivation Independent Methods

Traditionally, bacteria were identified based on phenotypic characteristics, which is less accurate than genotypic identification methods (Clarridge, 2004). Sequencing technologies for environmental microbiology can be divided into three main methods depending on the problems/questions that need to be solved, each with its strengths and weaknesses. This thesis will focus on metagenomics and 16S rRNA gene sequencing, but single amplified genomes will also be briefly discussed.

1.4.1 16S rRNA Sequencing

Evolutionary biology, microbial ecology, human microbiome, taxonomy and environmental microbiology would not be possible without the phylogenies derived from the discovery of small subunit (16S/18S) rRNA gene sequencing (Zhulin Igor B., 2016). For which a standard workflow can be seen in Figure 1.2. Sequencing of the 16S rRNA gene was first used for phylogenetic studies in the 1970s by Carl Woese and George Fox, where later it helped to solidify the three domains of life over the previous two domains of just eukaryotes and prokaryotes (Woese &

Fox, 1977) as previously discussed. 16S rRNA gene sequence analysis has identified novel bacteria without needing cultivation. It can solve the critical question of "who is there" by utilising online databases to assign the reads to taxonomic lineages (Xu & Zhao, 2018). The choice of the 16S rRNA gene to sequence was based on work by Woese et al., where they found that very few other genes are as highly conserved as the 16S rRNA gene in bacteria and archaea and that the absolute rate of change in the gene marks evolutionary relatedness and distance of organisms. Due to these features, relationships between organisms can be measured (Clarridge, 2004).

Amplicon sequencing offers fine taxonomic resolution, but assigning metabolic functionality to this can be highly problematic and is not recommended (Xu & Zhao, 2018). While the taxonomic assignment of 16S rRNA gene sequences can give information on what species are present, even though they are taxonomically similar, microbial species can have different functional capabilities. Bacterial species may have to adapt to changing environmental conditions and utilise different nutrients, resulting in various metabolic pathways. Therefore, there is a reliance on machine learning algorithms and statistical inference when predicting functional profiles from amplicon data and should only be used as a proxy for functional potential (Monica Steffi Matchado et al., 2023). This makes 16S rRNA sequencing valuable at a diagnostic level for presence/absence detection but limited in exploring what those microbes are contributing to the environment that they are in.

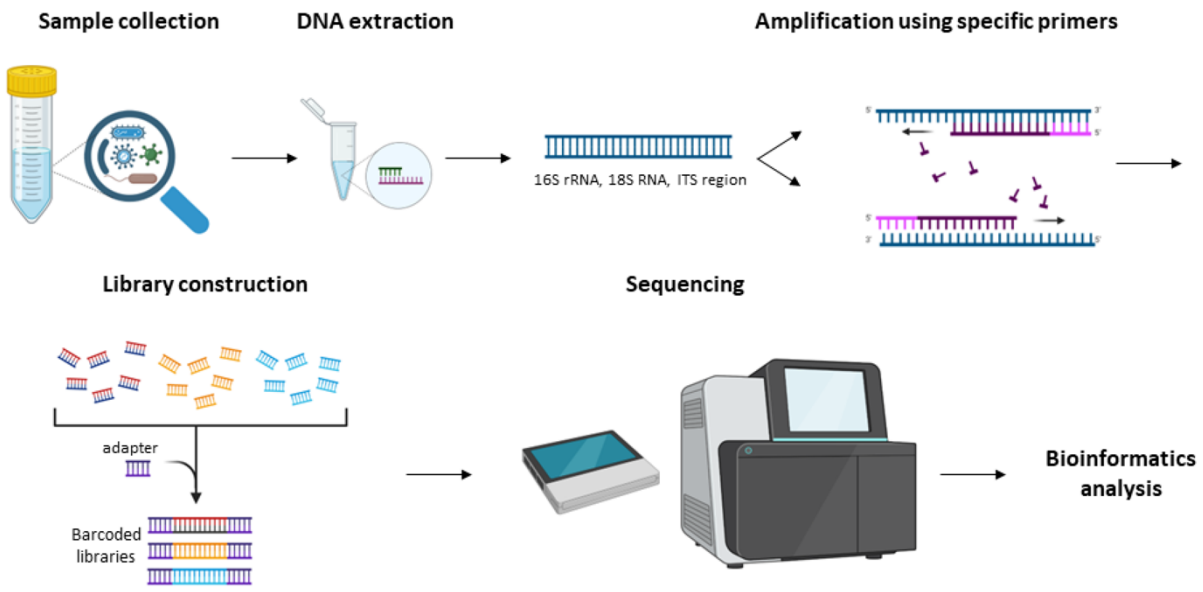


Figure 1. 2. A typical 16S amplicon sequencing workflow (taken from Athanasopoulou et al., 2023). Following sample collection, all microbes present in the sample are lysed, and their DNA is recovered. This DNA is then amplified using PCR, using primers specific to the research question. For instance, if a particular taxon is required, then primers targeting that taxa would be used, resulting in only the 16s rRNA gene from that taxon being amplified. Conversely, if a general survey of all taxa present is required, then general primers that amplify all organisms should be used. Following amplification. Sequencing adapters are then added to each amplicon before sequencing and analysis to recover the underlying sequences.

1.4.2 Single Amplified Genomes (SAGs)

SAGs involve physically separating one cell from the environment and sequencing its genome (Solden et al., 2016). The first step is to use fluorescent activated cell sorting (FACS) or microfluidics to partition the cells and then cell lysis and whole genome amplification (Figure 1.3). This is done by either PCR, isothermal or hybrid methods, such as looping-based amplification cycles and multiple annealing. It will then go on for shotgun sequence and genome assembly (Alneberg et al., 2018).

The main advantage of SAGs over 16S amplicons is that metabolic functions can be easily linked to specific species. They can be used to investigate gene insertion, duplication, rearrangement, loss and intra-species variation for uncultured microbes (Xu & Zhao, 2018). SAGs provide

information on the diversity, evolution, and adaptation of uncultivated microorganisms found in extreme environments or in low abundance (Solden et al., 2016). Some disadvantages are that cell sorting can be time-consuming and complicated, there can be highly uneven read coverage, and there is a high risk of chimeric reads (Xu & Zhao, 2018). SAGs can also be considered "good" if they are 50% complete, resulting in fragmented, less complete genomes (Solden et al., 2016). Due to low amounts of DNA in a single cell, this method is also susceptible to contamination (Sobol & Kaster, 2023). With more study in this area, SAGs will improve over time; cell sorting could also be used for cultivation, so pairing SAG methods alongside traditional cultivation could be a great way forward.

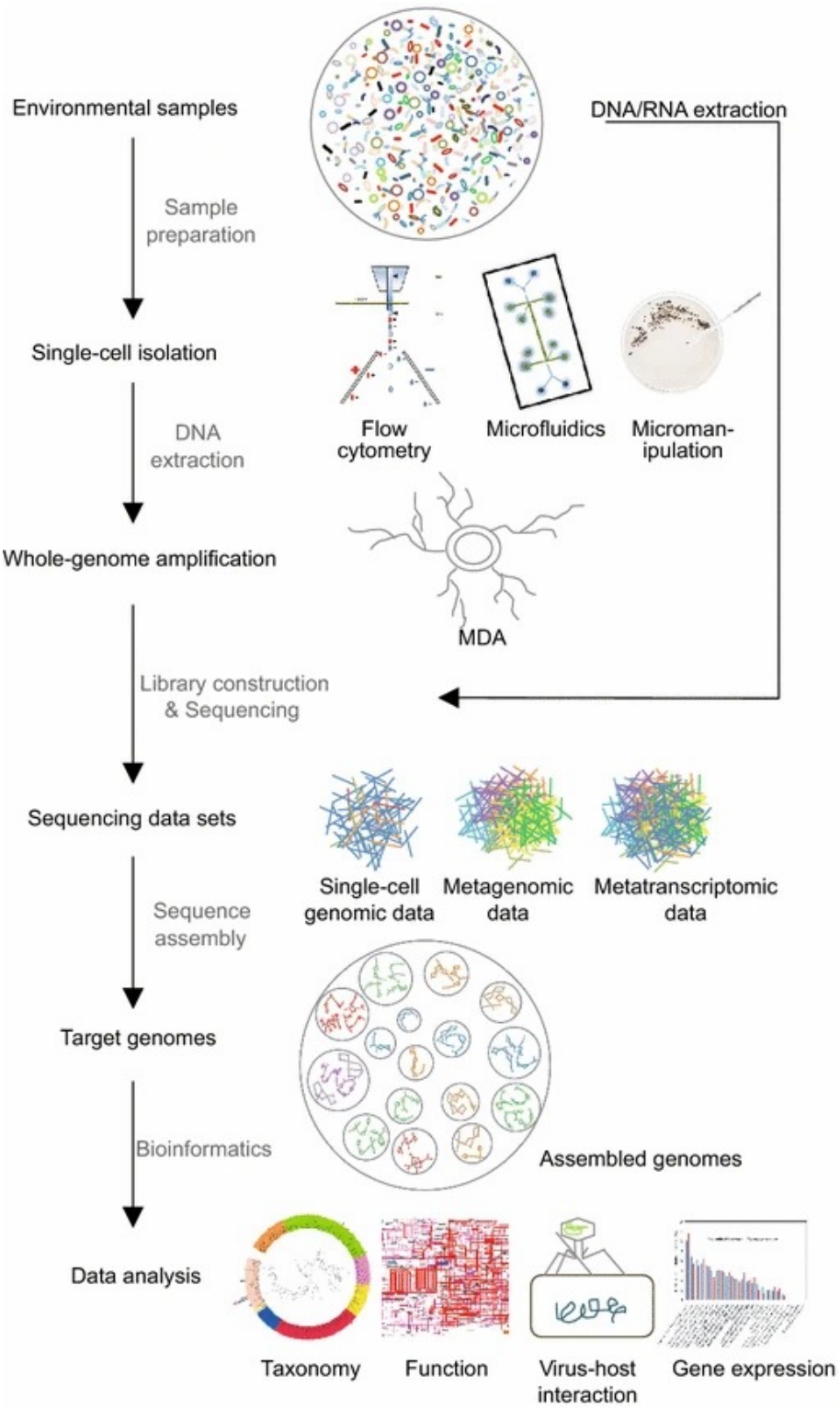


Figure 1. 3. An example of a workflow for single amplified genomes from (Xu & Zhao, 2018). Environmental samples are immediately processed or frozen until use. Cells are then stained either with a non-specific dye such as DAPI or they can be labelled specifically with fluorescence in situ hybridisation. Single cells are then physically isolated into multi-well plates most commonly with fluorescence-activated cell sorting (FACS). Single cells are then lysed to release DNA from the cell. Whole genome amplification is then required for library preparation so that there is enough DNA.

After libraries are prepared, short-read sequencing (Illumina) or long-read sequencing (Oxford nanopore) technologies are utilised. Once sequenced, bioinformatics is used to assess read quality and assemble, classify and annotate the sequences once sequenced.

1.4.3 Metagenome-Assembled Genomes (MAGs)

Metagenomics allows the study of organisms in the context of their environment, community, and chemical conditions. A typical workflow is shown in Figure 1.4; the process starts with metadata and sample collection, which then goes into the laboratory for DNA extraction, library construction, and shotgun sequencing (Kunin Victor et al., 2008). Assembly involves taking the shotgun sequences (small fragments on DNA) and stitching them together to form longer contiguous sequences (termed contigs). MAGs are generated by binning these contigs with similar characteristics (i.e. coverage, tetranucleotide frequency (Wu et al., 2016) before quality filtering and refinement to improve the final MAGs (i.e., completion/contamination statistics). While short-read technologies (e.g., Illumina) are popular due to their low cost and high throughput, long-read sequencing (e.g., Oxford Nanopore) can result in improved quality of MAGs. This is due to long reads spanning common repeat elements such as 16S rRNA genes, transposons and prophage sequences.

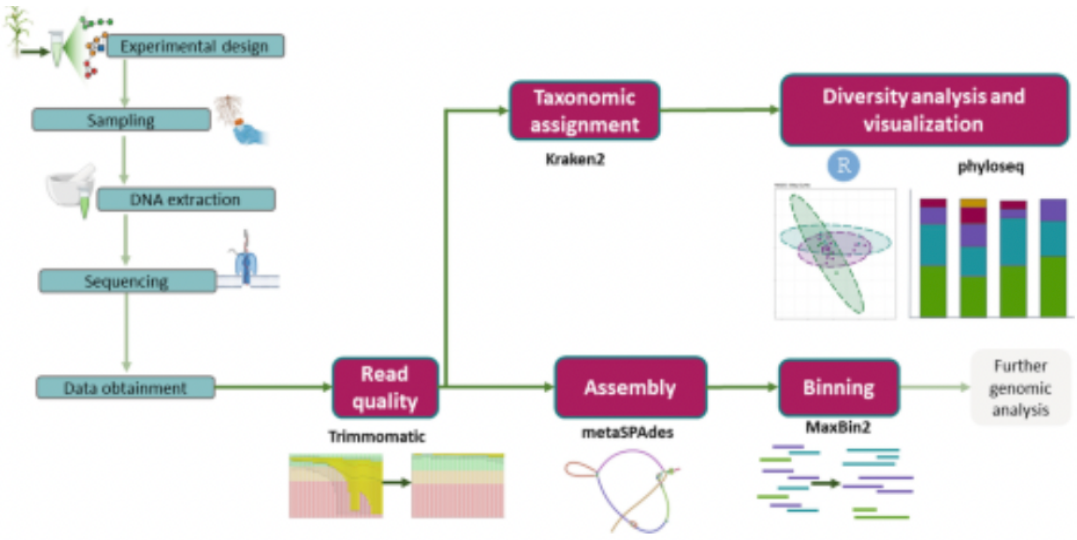


Figure 1. 4. Example of a metagenomic workflow (Nam et al., 2023). Study design and experimental protocol are important in metagenomics; it is important to decide on sequencing depth/coverage. After sampling, total DNA is extracted, quantified, and sequenced. After the data is obtained, computational quality control steps remove sequencing adaptors, duplicates, and quality trimming. Depending on experimental objectives, sequences are assembled and binned or are taxonomically assigned for whole community diversity analysis.

1.4.3.1 Benefits

Microbial dark matter (MDM) has been transformed from a data-poor topic to a data-rich one due to the current ease of DNA sequencing, which provides ample information on genomes. Genomic databases are set to increase exponentially as more microbes are discovered and studied (Jiao et al., 2021). Single-cell genomics and metagenomics allow the exploration of the genomes of these yet-to-be-cultivated thermophiles and many other difficult-to-culture microbes, enabling the study of their cell biology, evolution, and metabolic functions. The taxonomic richness of thermophiles may be underreported, with the vast majority yet to be studied with deep branching thermophiles being heavily underrepresented in cultures. Cultivation-independent studies allow the assessment of diversity in geothermal systems and other extreme environments (Hedlund et al., 2015).

MAGs expand our knowledge of yet-to-be-cultured bacteria and their potential functions and applications through gene prediction (Bodor et al., 2020). For example, thermophiles (45-80°C) and hyperthermophiles (>80°C) inhabiting geothermal systems were some of the first sites that were explored with cultivation-independent approaches using 16S amplicon sequencing. This revealed an incredible amount of diversity and unlocked a whole area of life that was not known to exist previously. In a paper by Hedlund and colleagues, they discuss how a groundbreaking finding from cultivation-independent methods was found in Yellowstone National Park, where they found deeply branching archaea (Korarchaeota) and two novel bacterial lineages (OP1-OP12). From this, a few members have been cultured, but many remain unexplored due to difficulty in figuring out growing conditions (Hedlund et al., 2015).

The main advantage of metagenomics is the comprehensive, high-resolution understanding of the community structure and possible metabolic pathways associated (at the community level) (Liu et al., 2022; Xu & Zhao, 2018). MAGs (and SAGs) can be functionally annotated for metabolic pathway reconstruction and to understand the organism's functional potential, which can help with their cultivation (Wang et al., 2017). Functional annotation involves two steps: Gene prediction and gene annotation. Databases are used to infer coding sequence function and can reveal metabolic traits, protein domains, and functional categories in yet-to-be-cultured organisms. Public databases have reference pathways to which homologous protein sequences can be mapped, which can help reconstruct metabolic pathways (Kunin Victor et al., 2008; Solden et al., 2016; Xu & Zhao, 2018). This allows for an in-depth characterisation of the molecular mechanisms of a particular organism or community, and it can be used for metagenome-informed cultivation (Kotera & Goto, 2016). This relies on the availability of well-characterised enzymatic reactions; if one enzyme is present that is already known to catalyse a reaction, its amino acid sequence is used as a reference to identify pathways in the genome interest (Kotera & Goto, 2016). Some examples are the cultivation of a critical nitrogen fixer, *Leptospirillum ferrodiazotrophum* sp., from an acid mine drainage biofilm (Tyson et al., 2005). Also, WG-1 from the *Succinovibrionaceae* group was successfully enriched from wallaby digesta samples, further understanding its contributions to lower methane emissions in wallabies (Pope et al., 2011).

1.4.3.2 Problems

Our view of microbial genetic diversity must be revised, as most sequenced prokaryotic genomes sit in only four phyla. This is because isolates are necessary for traditional whole-genome sequencing, and, as previously mentioned, culturing can be difficult for fastidious microbes. Both metagenomics and single-cell genomics bypass this step and can be applied directly to environmental samples (Rinke et al., 2014). These cultivation-independent methods provide vital information about the microbial world, unlocking a large number of diverse organisms that were never known to exist; however, MAGs being used as analogues for cultivated isolates can cause

some issues and the information that can be gained from having an isolate should not be ignored or overlooked. Even though the genomic data is available with metagenomics, it is missing the unambiguous assignment of physiological traits that cultivated isolates allow (Jiao et al., 2021). MAGs can only provide averages for the community and do not offer a fine taxonomic resolution. The direct study of cultivated microbes is used in various applications such as uses in industrial processes, new probiotics, biocontrol agents (Liu et al., 2022) and much more that may not be known yet.

If there is no isolate, there can be no laboratory experiments, and therefore, there cannot be complete confidence in the information gained from MAGs. Without cultured representatives that allow verification of taxon assignment, MDM genomes from MAGs may potentially draw misleading conclusions due to contaminated or misclassified contigs (Vollmers et al., 2022). With metagenomes, when a gene is found, it is said to have that function. However, there is still a possibility that this could be a chimeric bin or result of incorrect assembly. This shows that there are a lot of steps in these workflows where errors could arise without acknowledgement. Meziti et al. discuss how the strengths and limitations of the bioinformatics algorithms that make up metagenome-assembled-genomes (MAGs) are based on tests with mock data sets with known composition. This does not capture natural populations where diversity and complexity are common. They wanted to see if MAGs could recover population variable genes (>10% <90% of the population has the gene) as well as recover core genes (>90% of the population). To explore these issues, they compared MAGs against isolated genomes from eight diarrheal samples (Figure 1.5). They found that MAG quality may often be worse than predicted, with MAGs that had completeness near 95% only capturing 77% of core genes and 50% of variable genes. They also found possible errors at the genome binning step, with approximately 5% of the genes of the MAGs being identified as missing from the isolate and having a different taxonomic origin (Meziti Alexandra et al., 2021).

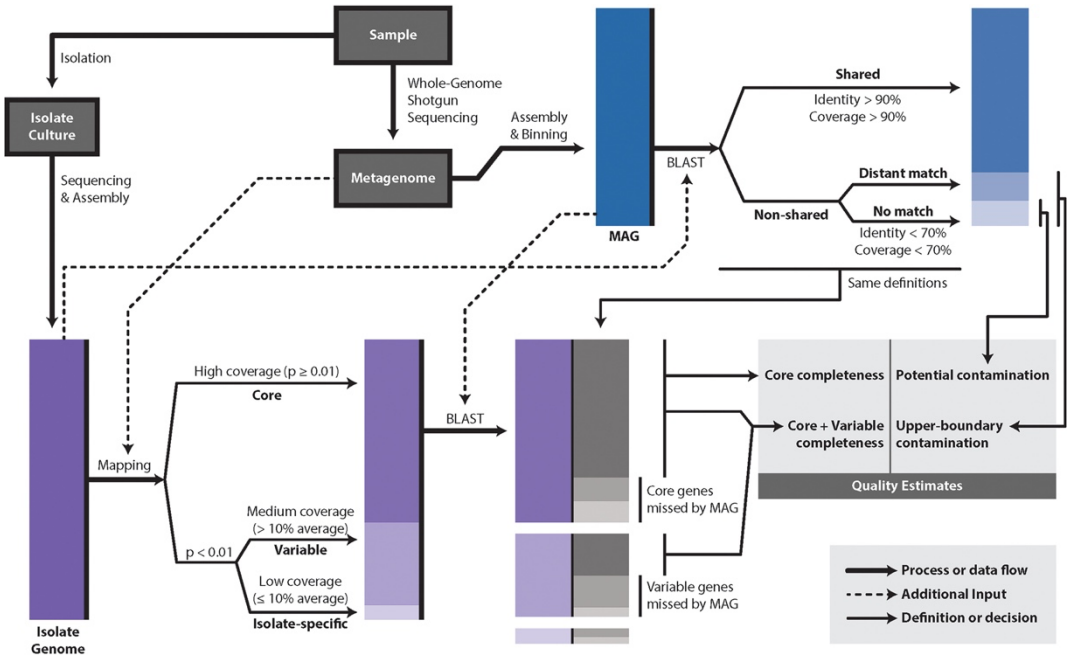


Figure 1. 5. The pipeline used by Meziti Alexandra et al. (2021) to identify the population and isolate genes captured by MAGs. From a faecal sample, isolates of E. coli (if tested positive for virulence factors by PCR) were sequenced, resulting in draft genomes. Metagenomes were assembled and binned from the same sample, and MAGs were recovered. They mapped metagenomic reads on isolate genes and contigs to identify variable and core genes and strain-specific genes. They then searched these genes against the MAG to assess the MAG for how well it represents the gene content.

As mentioned, genes can be predicted from metagenomes, and metabolic pathways can be reconstructed. There are still challenges, however, as many metabolic pathways are uncharacterised, and these methods cannot predict new enzymes or reactions that are not currently in the reference pathways. This means it cannot be recognised if a new protein, enzyme, or pathway is found. For proteins, function follows form, meaning that having access to the protein structure is essential to understanding the function of that protein (Varadi & Velankar, 2023). Emphasis was initially put on prompt public availability of protein structures, so most of them lack reliable information on their biochemical function, or no functional annotation is given. Proteins are then assigned function due to simple bioinformatics analysis based on sequence and three-dimensional structure comparisons but are often incorrect and have propagated throughout databases (Mills et al., 2015). Based on experimentally derived data, protein structures are archived and made available through the Protein Data Bank. However, the

gap between experimentally determined structures and protein sequences continues to grow. Varadi and Velankar stated that as of 2022, UniProt (protein sequence database) had over 230 million protein sequences, while the Protein Data Bank only had around 200,000 structures of 60,000 proteins (Varadi & Velankar, 2023), highlighting the many proteins where function is not known. Modelling through artificial intelligence through applications such as AlphaFold has begun to bridge this gap. Interpretation of data such as gene-gene interactions, clustering, and classification based on phylogenetic profiling and function to identify nutritional requirements to aid cultivation is a significant task and requires in-depth knowledge of physiology and metabolic pathways (Kunin Victor et al., 2008; Solden et al., 2016).1.4.5.1 Databases

The dependence on databases also creates a limitation. While it would only be possible to have the knowledge or ease of access to data available today with them, databases such as NCBI and SILVA do have limitations that researchers should be conscious of. These include but are not limited to:

- Outdated or incomplete information: As databases rely on data submitted by researchers, they may only sometimes be up to date with the latest findings. This can result in missing information or potential inaccuracies.
- Contamination: A notable portion of genomes is deposited into databases that contain foreign genome fragments. This can originate both in silico and in vitro, usually due to contamination of culture media or reagents during physical processing in the laboratory (Orakov et al., 2021). Future comparative genome analyses may be affected by gradual database contamination by past submissions, causing propagating errors. Due to this, contamination filtering must be strictly applied. However, current reporting standards tend to emphasise completeness over purity, with checkM being the most used genome assessment tool, which discriminates against fragmented genomes and, consequently, uncultured taxa (Vollmers et al., 2022).

- Taxonomic uncertainty: Knowledge of microorganisms is always rapidly evolving. This means that databases might not always reflect the most current taxonomy, leading to misinterpretation of data.
- Sequence quality and annotation: Submitted sequences can vary in quality, and the annotation of these sequences might not always be consistent.
- Sampling bias: Databases can often be biased to well-studied environments and organisms, creating a skewed representation of microbial diversity.

1.5 The Importance of Cultivation

Microbial cultivation is the most traditional method used in microbiology laboratories. It is the process of providing the proper environmental conditions and nutrients to propagate organisms. Classical cultivation strategies can be dated back to the 19th century and include methods such as selective nutrient media, microscopy and colony picking from solid media (Lewis et al., 2021). There has been a surge in genomic data since the rise of metagenomics and single-amplified genomes, but cultivation still needs to be improved. With this surge in genomic data, the bottleneck now lies with translating these sequences into functional context. The interpretation of genes is biased by the physiology of a few well-known and studied model organisms. Cultures can build bridges to help with this gap in knowledge by isolating and studying unusual microorganisms (Carini Paul, 2019; Lewis et al., 2021). Isolates are indispensable and are the foundation of microbiology research for understanding evolution, cell biology, and ecological roles and impacts. As many environmental microbes are uncultured, there is a push to find ways of bringing these poorly characterised groups into culture, as isolates can have a range of applications, such as industrial processes, medicine, and various applications in biotechnology and agriculture (Liu et al., 2022). As well as furthering our understanding of their roles in the environment, such as their impacts on biogeochemical cycles. They are invaluable tools for understanding, harnessing and utilising the microbial world for the benefit of science and society.

1.5.1 Importance for Research

Isolated microorganisms in cultures are invaluable resources for experimentation and validation of ecological value and processes, evolutionary processes, physiology, phenotypes, and pathogenicity. They can also be used in research to understand how genes and the proteins they encode for function (Gupta et al., 2021; Liu et al., 2022; Zengler, 2009). This can be done by gene knockout mutations, which is the process of changing or deleting a gene in the intact organism and seeing how differently it functions. An example of this is the Keio collection for *Escherichia coli*. Baba et al. (2006) systematically made a set of single-gene deletions of all nonessential genes in *Escherichia coli* K-12, which are precisely defined. They obtained mutants for 3985 genes out of a possible 4288; the 303 that were unable to be disrupted are candidates for essential genes. These mutants provide a resource for analysis of gene functions and regulatory networks that would be impossible without the cultivated isolate (Baba et al., 2006). The Keio Collection of gene knockouts is a critical resource in molecular biology and genetics. The collection consists of an extensive library of bacterial strains, each with a specific gene systematically knocked out. These curated mutants serve as an invaluable tool for researchers studying gene function, allowing them to investigate the roles and effects of individual genes with precision. The Keio Collection provides a standardised platform, making it easier to identify the functions of genes in various biological processes, from metabolism to pathogenicity. This resource has advanced our understanding of gene function and has widespread applications (Baba et al., 2006). *Escherichia coli* is, therefore, one of the most well-studied model organisms, and a lot is inferred from it with the assessment of how loss of a gene can impact phenotypes being extensively used in other species (Aedo et al., 2019).

1.5.2 Environmental

Human activities can adversely affect natural ecosystems from farming, industrialisation, mining, and other processes meant to enhance human life. However, this can cause changes in air, soil, and water quality. It has been shown that certain microbes can aid in controlling heavy metal

accumulation and nutrient cycling. An example is bacteria mobilising heavy metals in soil, transforming and detoxifying it (M. Chen et al., 2015; Tiwari & Lata, 2018). The mechanisms and processes involved in this are poorly understood and require further study, as this could be enhanced by adjusting the factors that control them, such as genetic and structural adaptation (Tshikantwa et al., 2018).

Bioremediation offers a promising solution for the rehabilitation of polluted environments. Industrialisation presents challenges to the environment and requires a better understanding of bioremediation strategies. Bioremediation relies heavily on the survival and effects of microbial communities and how they interact with introduced microorganisms. As most microbiomes remain uncultivated, more work needs to be done focusing on their functions in ecosystems so that they can be better utilised for environmental biotechnology (Bodor et al., 2020). An environmentally friendly and economically feasible way of treating wastewater is using microorganisms as a biological treatment. Aided with CO_2 and molecular O_2, microbes oxidise organic compounds (Oljira et al., 2018). A great example of a valuable bacteria is methanotrophs, which are interesting as they can survive on methane as their sole energy and carbon source. They are found in various environments, such as sediments, seawater, fresh water, soils, hot springs, and even the Antarctic. Methanotrophs are interesting due to their versatile applications in bioremediation, pollution degradation, and the production of chemicals, as well as their critical role in the global methane cycle (Murrell et al., 1998).

Bacteria can also be used in agriculture to modulate the microbiome by applying probiotics to enhance soil health and plant growth. Probiotic bacteria, such as those from the genera *Bacillus* and *Pseudomonas*, can be introduced into the soil to promote beneficial plant interactions. These bacteria can improve nutrient availability, protect plants from pathogens, and stimulate plant growth through various mechanisms. By establishing a healthy and diverse microbiome in the rhizosphere, where the plant roots interact with the soil, probiotics can help modulate the plant's microbiome to enhance its overall health and resilience (de Souza Vandenberghe et al., 2017; Kim & Anderson, 2018). Most soils are currently depleted in nutrients; therefore, chemical

fertilisers are used to combat this and produce healthy plants and crops. These fertilisers, in turn, harm the environment and other organisms (Prasad et al., 2021). To promote sustainable agriculture, investigating enhancing soil microbes for efficient nutrient recycling could reduce reliance on synthetic fertilisers.

1.5.3 Medicine

Mammals' gastrointestinal tract hosts a diverse number of microorganisms (Sánchez et al., 2017). This gut microbiome works with the epithelium and surrounding organs from birth to regulate the immune system and digest nutrients. Therefore, disruptions to the gut microbiome have been associated with gastrointestinal infections, inflammatory bowel disease, obesity and cancer (Fishbein et al., 2023). Other aspects, such as antibiotics and poor diet, can also harm the human gut microbiome. However, as the microbiome is a complex and diverse ecological network of microbes, the effects of these can be variable (Fishbein et al., 2023). One way to increase the health of the gut microbiome is to use bacterial isolates as probiotics. They can be introduced to the gut in various ways to promote a healthier microbiome and increase overall human health and well-being. Probiotics are perhaps the most common form of bacterial isolates used to improve gut health. These are specific strains of bacteria, often from the *Lactobacillus* and *Bifidobacterium* genera, known to confer health benefits when consumed. Probiotics are available as supplements or are found in certain foods like yogurt. They can help maintain a balanced and diverse gut microbiome by increasing the population of beneficial bacteria (Agrawal, 2005; Fijan, 2014).

1.6 Isolation Issues - Common Barriers

Difficulty isolating specific organisms from the environment can be due to, but not limited to, specific nutrients being hard to identify, growth conditions being difficult to replicate, symbiosis and competition. These are described below:

1.6.1 Nutrients and Growth Conditions

All living things, including bacteria, require certain nutrients and conditions to grow, some more specific than others. Media components can be enhanced to select a certain organism to grow by adjusting factors such as substrates and electron donors/acceptors. Growth factors (vitamins, amino acids, inorganic compounds, nucleotides, humic acids) and inorganic compounds (metals, nitrogen, sulphur) used in critical biogeochemical cycles can be difficult to identify. They may be so low in the environment that they are not recognised (Lewis et al., 2021). Extremophiles (those living in extreme environments previously thought to be uninhabitable) make up much of MDM as their environments are complex to replicate in a laboratory setting. Environmental factors include optimum growth temperature, pH, redox potential, and salt concentration. The atmosphere is another factor, as microbes can be anaerobic or aerobic (obligate or facultative), in which case, the growing environment can be artificially enhanced with different gas mixes, including varying levels of hydrogen, nitrogen, and carbon dioxide (Tshikantwa et al., 2018). These factors must be optimised to allow fastidious microbes to grow in culture.

One specific challenge is the need to grow bacteria on solid media for various reasons, including mutation and gene knockout experiments. Those obligately anaerobic and requiring insoluble energy sources, such as elemental sulphur, can be difficult to cultivate on solid media. Using anaerobic chambers can be costly, and not every laboratory can access them, so finding other ways to provide an anaerobic environment would be beneficial. Also, ways to make nutrients soluble and more bioavailable must be investigated.

1.6.2 Symbiosis

Microorganism's interactions in a community are essential for the global biogeochemical cycling of nutrients and play a role in many other processes, such as human health and disease. These heterospecific interactions can be harmful, neutral, or beneficial to each species (Zengler & Zaramela, 2018). Symbiosis is defined as the interspecies dependency of one or more microbes (Lewis et al., 2021). The convoluted networks of microbial communities not only rely on each

other for a variety of processes. These include the exchange of electron donors and signal molecules (small peptides, siderophores), which may also depend on each other for reducing ROS (reactive oxygen species) or inhibitory molecule detoxification (Zengler & Zaramela, 2018). By degrading a substrate, organisms can also rely on each other to overcome thermodynamic limitations. As this interdependency can sometimes be obligate, growing isolated cultures by separating them can be difficult (Lewis et al., 2021).

An example is methane formation and oxidation, which are essential in the global carbon cycle and are mediated by syntrophic archaeal and bacterial communities. Electron transfer between species is a critical process in these cycles, as well as the member's ability to take advantage of their syntrophic partner's metabolic abilities. This involves breaking down compounds they cannot digest and overcoming energy barriers (Stams & Plugge, 2009). For such symbionts, methods that allow co-isolation, such as cell sorting, may be beneficial for establishing a co-culture (Lewis et al., 2021).

1.6.3 Competition and Low Abundance

Some microorganisms may exist in the environment in low abundances, making them rare but still having ecological relevance and exerting substantial influences on the community. The low-abundance organism may have a faster rate of substrate metabolism, but other organisms have a higher growth rate (i.e. processes are not linked) (Lewis et al., 2021; Lynch & Neufeld, 2015). This means that when culturing, the high abundance of organisms may dominate and outcompete the target of interest. Cultivation of such rare microorganisms would benefit by choosing an environment to sample from that they have the highest natural relative abundance; information on this can be found by analysing publicly available 16S rRNA gene datasets. This may still not work as even if it is initially in high abundance, the faster-growing microbes may still outcompete once cultivation commences. This competition happens when rich substrates are added to cultivation media (such as peptone or yeast). For oligotrophs, media with low nutrients

can be successful. Slow growth can also be impractical and have economic implications for research as it involves long timescales and can be fewer appealing targets (Lewis et al., 2021).

1.7 Current Cultivation Approaches

1.7.1 iChip

The isolation chip (iChip) is an innovative tool in microbiology, revolutionising how scientists study and understand microbial diversity. It was developed by D. Nicols in 2010 to cultivate microbial dark matter in its natural environment. It comprises hundreds of miniature diffusion chambers, which are inoculated with a single cell from the environment (Nichols et al., 2010). This technology employs a semi-permeable membrane that mimics the natural environment of these microbes, allowing them to thrive under conditions similar to their habitat while protecting them from competition with faster-growing species. It is successful in the microbial recovery of species that are of notable phylogenetic novelty, such as those from the “uncultivated microbial minority”, from a variety of environments (Liu et al., 2021; Nichols et al., 2010; Zhao et al., 2023).

While the iChip has enabled the cultivation of some fastidious microbes, it does have some limitations. While the iChip has revolutionised microbial cultivation, it does have some limitations. One notable downside is the relatively low throughput compared to other culturing methods, as validation can be time-consuming. Cultivating microorganisms using the iChip is time-consuming and may not be suitable for high-volume studies or rapid screening of large sample sets. Another challenge is maintaining the conditions required to grow diverse microbial communities within the device, as environmental parameters can vary significantly between microbial species. These limitations highlight the need for continued innovation and the integration of complementary techniques to fully capture microbial diversity's richness.

1.7.2 Cell Sorting

Cell sorting methods are techniques to separate different cell types from a heterogeneous mixture. Some standard cell sorting methods include:

- Flow Cytometry: This method uses fluorescently labelled antibodies that bind to specific cell surface markers. The cells are passed through a laser beam, and detectors measure the fluorescence emitted by the labelled antibodies, allowing the sorting of cells based on their unique characteristics (McKinnon, 2018).
- Fluorescence-Activated Cell Sorting (FACS): FACS is an extension of flow cytometry that identifies and physically separates cells based on their fluorescence. It uses an electrostatic charge to deflect droplets containing specific cells into different containers, allowing the isolation of the desired cell population (Basu et al., 2010).
- Microfluidic Sorting: Microfluidic devices use small channels and chambers to manipulate cells based on their size, shape, or other physical properties. Cells can be sorted by passing through specific channels that exploit these differences (Farahinia et al., 2023).

However, these methods can be challenging for several reasons. One is that cell sorting equipment, such as flow cytometers or FACS machines, can be expensive to acquire and maintain. Additionally, the expertise required to operate these instruments effectively adds to the challenge, and they can also be prone to a lot of contamination in the negatives.

1.7.3 Metagenome-Informed Cultivation Success

The growing amount of sequence information from metagenomics and metatranscriptomics from various environments allows for the guided isolation of microbes of interest. As a result, there have been some examples of metagenome-informed cultivation success. One example of successful cultivation is the cultivation of *Artibacter laminatus* by Katayama et al. (Katayama et al., 2020). Culture-independent methods confirmed OP9 of the bacterial candidate phylum in Yellowstone National Park sediments. This phylum was found to be important for hydrocarbon

degradation and abundant in anaerobic habitats globally. It took Katayama et al. three years to enrich this axenic culture. From omics techniques, they hypothesised that they have a metabolism based on sugar degradation coupled with hydrogen production (Katayama et al., 2020). However, they found that even though it produces hydrogen, it cannot tolerate high levels of hydrogen. Therefore, it grows better in symbiosis with a hydrogen-consuming methanogen in a co-culture. If further studies confirm this, this relationship could affect the methane cycle. Understanding this syntrophic process could, therefore, be necessary for model development and future climate change predictions. This study shows how important it is to put efforts towards bringing microbial dark matter species into the culture. However, metagenomics is still primarily used for sequence information.

Another example is by Imachi et al. (2020), who isolated an archaeon at the prokaryote-eukaryote interface. This is very important in understanding how eukaryotes emerged, which remains unclear and helps provide insights into life's origin. They isolated the Asgard archaeon '*Candidatus Prometheoarchaeum syntrophicum'* strain MK-D1 after a 12-year study due to its extremely slow-growing nature. They used the metagenome to explore the metabolic potential of this archaeon and found that syntrophic growth with *Methanogenium* and *Halodesulfovibrio* can catabolise ten peptides and amino acids through hydrogen transfer. Therefore, the addition of hydrogen scavenger-inhibiting compounds and subsequent transfers allowed them to eliminate *Halodesulfovibrio* from the culture, leaving *Methanognium* and the target MK-D1 in a co-culture (Imachi et al., 2020)

1.8 Campylobacterota

1.8.1 Why are they interesting?

The novel phylum Campylobacterota was created from reclassifying the order *Desulfurellales* (previously Deltaproteobacteria) and the Proteobacterial class Epsilonproteobacteria. This was

due to recent analyses of marker genes and 16SrRNA to create bacterial trees that have failed to recover Epsilonproteobacteria as monophyletic with the other classes of Proteobacteria. Phylogenies based on the 16S rRNA gene and joined alignment of 120 conserved protein marker sequences were used to organise the phylum (D. Waite et al., 2019; D. W. Waite et al., 2017). Campylobacterota are captivating due to their significant impact on human health and the environment. Some are pathogens residing in the stomach wall, for example, *Campylobacter jejuni*, and some are also important in the sulphur cycle, living in deep-sea hydrothermal vents and geothermal sites, *Sulfurospirillum* and *Sulfurimonas spp*. The ancestor of this new phylum, as suggested by genomic analysis, was a thermophilic, autotrophic, motile, chemolithotroph (D. Waite et al., 2019; D. W. Waite et al., 2017). They all share common energy conservation mechanisms, and although they have a small genome, they have remarkably large amounts of electron transport chain enzymes (van der Stel & Wösten, 2019). Campylobacterota also undertakes a lot of horizontal gene transfer, increasing antimicrobial resistance (L. Ma et al., 2021; Yoshida-Takashima et al., 2022).

Cultivating Campylobacterota in a laboratory setting poses considerable challenges. One key reason for this difficulty lies in their fastidious growth requirements. These bacteria demand specific nutrients, growth temperatures, and microaerobic or fully anaerobic conditions that are difficult to replicate accurately in vitro, as previously described. This leads to the species in the Camplyobacterota phylum being underrepresented in databases, with genomes being recovered without isolates. Consequently, the mysterious nature of Campylobacterota makes them fascinating and requires more research, as studying this could improve our understanding of bacterial physiology and evolution. Further, there is an unequal nature of Camplyobacterota in culture collections as there are a lot of *Camplyobacteria* and *Helicobacters* represented as these are human pathogens. Still, there is a much smaller sample size of environmental Camplyobacterota, including orders of Camplyobacterales, Desulfurellales and Nautiliales.

Desulfurella has been chosen as a test case as they have garnered attention due to their unique metabolic capabilities, specifically their ability to participate in sulphur cycling and iron

reduction processes. *Desulfurella* are obligately anaerobic and can utilise a variety of organic compounds as energy sources while coupling their metabolism to the reduction of sulphur compounds like sulphate or thiosulfate (A. P. Florentino et al., 2017). As mentioned above, *Desulfurella* and other environmental Camploybacterota are underrepresented in databases and culture collections, which makes them an interesting target. Thanks to the 1000 Springs Project, they are known to reside in hot springs in Rotorua, New Zealand, making them an easily accessible test case. They also offer a lot of challenges for cultivation on solid media, such as utilising an insoluble energy source (elemental sulphur) and being obligately anaerobic.

1.9 1000 Springs Project

The 1000 Springs Project is a project where the team sampled 1000 geothermal sites in New Zealand and subsequently generated a catalogue of information on the unique microbial diversity in the ecosystems and other important geochemical information. The most important part of the data for this project was the 16S rRNA sequencing, which allows for the assessment of microbial diversity in each pool. This thesis leveraged this data for site selection to target pools with Campylobacterota, specifically *Desulfurella.*

1.10 Conclusion

As mentioned, there are considerable downfalls in treating MAGs as analogues for isolates. While genetic data can provide a lot of information on microorganisms, their isolation is still essential to study and use in industries. As microbial dark matter is difficult to grow in axenic culture, it could be beneficial to use reconstructed metabolic pathways from metagenomes to inform the culturing conditions. This thesis will attempt to combine classical techniques with omics tools to provide a workflow for isolating a targeted organism.

1.11 Aims of This Thesis

This thesis aims to provide a targeted approach to cultivating bacteria, enabling the acquisition of isolates of species of interest rather than those that are just easy to culture. This aims to target the issue of the cultivated majority and the skewed view of the microbial world. It also aims to provide cultivation approaches for fastidious microbes that are more accessible to standard laboratories without needing expensive, specialised equipment.

Chapter 2: Primer design and optimisation

2.1 Introduction

For any targeted cultivation project, like that presented in this thesis, appropriate polymerase chain reaction (PCR) primers must be used to amplify the target species whilst minimising erroneous (i.e. non-target) amplification. These primer pairs can be found in the literature, modified from published primers, or designed de novo, depending on the specificity needed for the project. Initially, the target for this study was the class Epsilonproteobacteria; therefore, I first used published primers, EPS_F/ESP_R (Gittel et al., 2012). The EPS_F/ESP_R primers were developed for the Proteobacterial class Epsilonproteobacteria before this was reclassified into the novel phylum Campylobacterota, combining the Epsilonproteobacteria class with the order *Desulfurellales* (previously in Deltaproteobacteria class). This reclassification was due to the analysis of marker gene and 16S rRNA phylogenies failing to recover Epsilonproteobacteria as a monophyletic clade from other classes of Proteobacteria. Phylogenies based on the 16S rRNA gene and a joined alignment of 120 conserved marker sequences were used to organise this novel phylum (D. Waite et al., 2019; D. W. Waite et al., 2017). Due to this reclassification and narrowing down to a more specific target of the species *Desulfurella*, new primers were needed for this work, as EPS_F/ ESP_R did not include any *Desulfurella sp.* sequences in their design process. In addition, I also sought to create a longer amplicon for sequencing since EPS_F/EPS_R only amplify approximately 150 base pairs. This chapter will outline the primers designed for this study with these factors in mind and their optimisation for PCR.

2.2 Methods

2.2.1 Original Published Primers

The EPS_F/EPS_R primers were designed by Gittel et al. (2012) with Primrose based on 11,141 public domain sequences, with their specificity checked in silico against the Ribosomal Database Project and ARB.

2.2.2 Primer Design

Geneious Prime was used to make an alignment of all available Desulfurella 16s rRNA sequences from NCBI and SILVA, using a global alignment with free end gaps at 93% similarity with a gap open penalty of 12 and a gap extension penalty of 3. Primers were edited or designed based on this alignment and following basic primer design rules (https://www.thermofisher.com/blog/behindthebench/pcr-primer-design-tips/).

All primers were tested in silico on the *Desulfurella* alignment created previously using Geneious Prime using the Test with Saved Primer's function, allowing for binding anywhere on the sequence and for one mismatch in the binding region. They were also tested using the TestPrime function on the SILVA database ssu-138.1, sequence collection RefNR, which allowed for one mismatch.

2.2.3 Optimisation

All primers were optimised for the optimal annealing temperature by running temperature gradients of 56°C, 58°C, 60°C, 62°C, 63°C, and 64°C. The Platinum Direct PCR Universal Master Mix by Invitrogen was used for this step. *Alvinella pompejana* microbial community DNA was used to optimise the EPS_F/EPS_R and EPS_FM/EPS_RL primers as well as a positive control for future PCRs. This was used as the microbial community inhabiting the back of *Alvinella pompejana* has a highly diverse Epsilonproteobacteria community, making it useful for optimising these primers (Campbell Barbara J. & Cary S. Craig, 2001). DSMZ526

Desulfurella acetivorans DNA was used for the same purpose for the DS_F/DS_R primers. These were then run out on a 1% agarose gel for visualisation.

2.2.4 Universal Primers

The choice of which universal primers to use for 16S rRNA sequencing was also based on the *Desulfurella* 16s rRNA gene alignment on Geneious Prime (Figure 2.1). Using the same Test with Saved Primers function, EubB (AGAGTTTGATCMTGGCTCAG) and EubA (AAGGAGGTGATCCANCCRCA) and the Earth Microbiome primers, 515F (GTGYCAGCMGCCGCGGTAA) and 926R (CCGYCAATTYMTTTRAGTTT) were tested.

2.3 Results

2.3.1 Primer Design

To begin the primer design process, Geneious Prime was first used to align published *Desulfurella species* 16s rRNA sequences (Figure 2.1) to create a consensus sequence. 16S rRNA sequences were sourced from NCBI and SILVA online databases. Primers were tested and designed based on the alignment created. The consensus sequence for this alignment shows a high level of agreement amongst all those sourced, presenting a good resource for primer design.

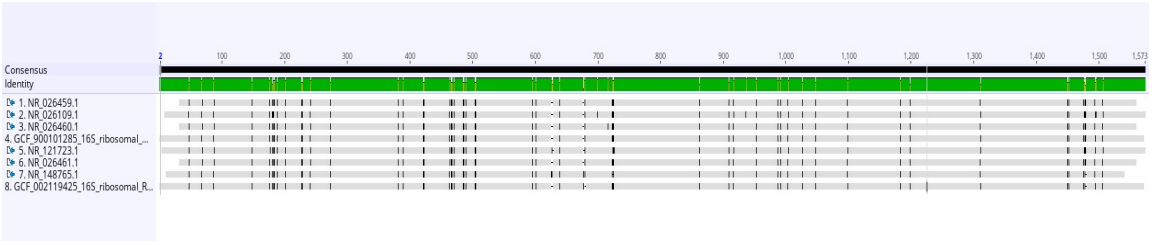


Figure 2. 1. Geneious Prime alignment of Desulfurella 16S rRNA sequences from NCBI and SILVA databases.

Table 2.1. Original primers and primers developed for this project. EPS-F mod edits from the literature are shown in red. Version will be used to reference these primers throughout the rest of this study.

Primer	Sequence (5' – 3')	Estimated annealing temperature °C	Method	Reference
EPS_F	CGTTACTCGGAATCACTG	54	Published	(Lin et al., 2006)
EPS_R	GATTTTACCCCTACACCA	54	Published	(Gittel et al., 2012)
EPS_FM	CGTTRYTCGGAATCACTG	55.1 - 60.5	Modified from published EPS_F primer	This study
EPS_RL	CATTGTAGCACGTGTGTMG	53.9 - 57.1	De novo	This study
DS_F	GGTGCGCTAATACCGGATAC	58.3	De novo	This study
DS_R	GCTCCTACTACTAGCATGTC	54.5	De novo	This study

EPS_F/EPS_R primers hit 92.1% of Camplyobacterota, but as expected, they do not amplify published *Desulfurella sp.* sequences based on the SILVA TestPrime function (Table 2.2). In the Geneious alignment created, the forward (EPS_F) binds with one mismatch, but the reverse (EPS_R) does not bind at all. A new primer pair (EPS_FM/EPS_RL) was created to include *Desulfurella sp* into the species targeted.

EPS_FM was modified from the EPS_F primer to target *Desulfurella,* and EPS_RL was designed de novo for a longer amplicon, better suited for sequencing, of 880BP. This primer pair successfully targets the Geneious alignment, and the result from the SILVA search was that they target 96.9% of *Desulfurella sp.* (Table 2.2). However, they also target 95.8% of all Campylobacterota sp. and 16.1% of all bacterial species, making them unsuitable for this work. DS_F/DS_R primers are much more specific and produce a longer amplicon than the original published EPS_F/EPS_R primer pair. They target no other bacterial or Campylobacterota species other than *Desulfurella sp.*, successfully hitting 88.2% of published *Desulfurella sp.* (Table 2.2). Whilst this is lower than the EPS_FM/EPS_RL primer pair, specificity was the focus of the design for these primers, with the drop in sensitivity (96% to 88.9%) being acceptable for this project.

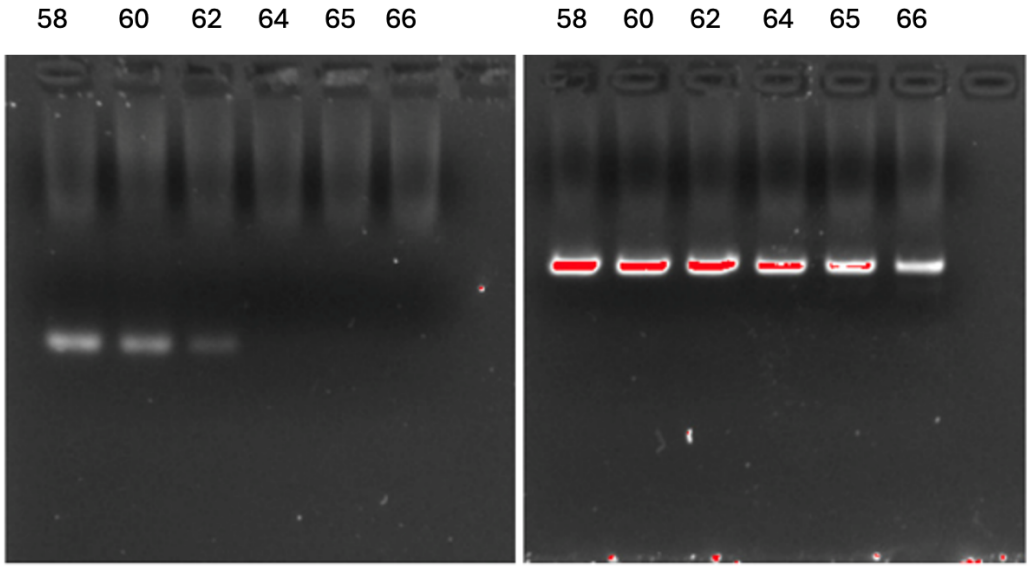
Table 2.2. Primers checked for specificity in silico on the SILVA ref NR, SSU r138.1 database, allowing for one mismatch.

Primer pair	**Bacteria targeted (total)**	**Number of different phyla targeted**	**Campylobacterota targeted (total)**	***Desulfurella* targeted (total)**
EPS_F/EPS_R	1%	1	92.1%	0%
EPS_FM/EPS_RL	16.1%	89	95.8%	96%
DS_F/DS_R	0%	1	0.4%	88.9%

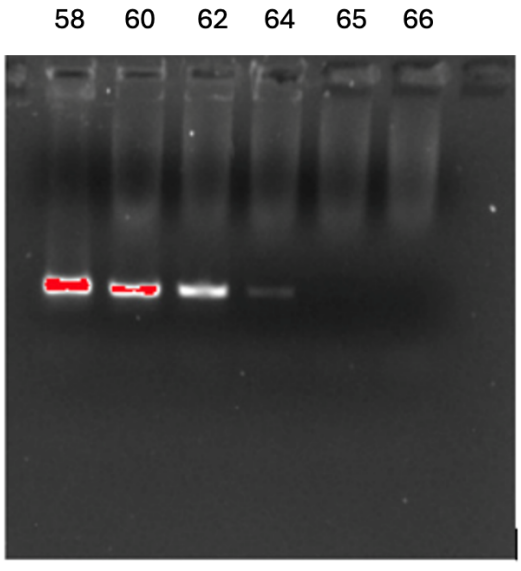
2.3.2 Optimisation

Whilst the design process can give predicted characteristics of these primers, such as annealing temperature, experimental optimisation is a crucial aspect of primer design. All primer pairs presented in Table 2.1 were optimised for annealing temperatures using a temperature gradient and running the resulting amplicons on a 1% agarose gel (Figure 2.2). When imaging the

resulting gels, the optimum annealing temperature is determined by a drop-off in amplification indicated by band brightness. Figure 2.2 shows that EPS_F/EPS/R primers displayed a significant drop off in amplification when the annealing temperature surpassed 60 degrees, with no amplification above this. Therefore, an annealing temperature of 60 degrees was chosen for this primer pair as the highest temperature is best for specificity. The same is true for the DS_F/DS_R primers; there is slight amplification at 62 degrees, but 60 degrees was chosen for this primer pair as amplification drops significantly here. EPS_FM/EPS_RL primers amplify up to 66 degrees with a slight drop off in amplification here, so this temperature was chosen.



2a 2b



2c

Figure 2. 2. Optimising annealing temperature of all primer pairs. The temperature is °C. 2a. EPS_F/EPS_R primers. 2b. EPS_FM/EPS_RL primers. 2c. DS_F/DS_R primers.

2.3.3 Universal Primers

Although this chapter has focused on developing and optimising specific primers, using a more general universal primer can still prove valuable. First, through the sequencing of universal primers, it can become clear whether the targeted isolation has successfully recovered a clean isolate. This is because while a specific primer pair determines the presence of the targeted microorganism, its specificity removes its ability to assess if any other organisms are present. Therefore, I sought to use a combinatorial approach to this project with specific and universal primer pairs. The universal primer pairs can be used at two different taxonomic levels. The first is at the phylum level, with the other being at the domain level.

The EPS_F/EPS_R primer pair can be considered the phylum level universal primers as, while they do not target any *Desulfurella sp.,* they target 92.1% of all other Campylobacterota species. Therefore, the combination of EPS_F/EPS_R and DS_F/DS_R provides good coverage of the entire phylum. Indeed, the lack of any *Desulfurella sp.* amplification can be seen as an advantage as it would inform of any Campyobacterota other than *Desulfurella* present in the isolation. These phylum-level universal primers were optimised in the previous section.

The domain-level primer pair is helpful in assessing whether any other bacterial species outside the phylum are present. Numerous published bacterial domain-level primers have been created and published. I therefore decided to use the published EubA/EubB and Earth Microbiome Project (EMP) primer pairs as domain-level universal primers. These universal primers were checked in silico on Geneious Prime with the *Desulfurella* alignment to assess the best suited for 16S rRNA sequencing. From the alignment, I noticed that the EubA/EubB primer pair is problematic as it does not hit the start or end of several of the published sequences. The EMP primers, however, successfully hit the alignment with the 515F and the 926R primers, both of which had no mismatches. The PCR protocol used was as previously described.

Figure 2. 3. Universal primers EubB/EubA tested in silico on the Desulfurella 16S rRNA gene alignment on Geneious Prime Showing that these primers do not target the published sequences as they appear to be truncated on their end of the 16S gene.

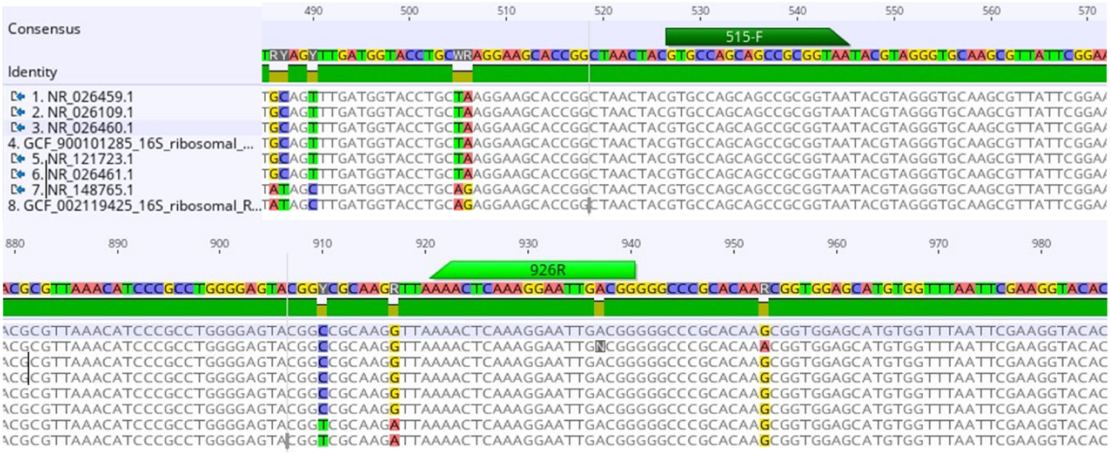


Figure 2. 4. The Earth Microbiome primers tested on the Desulfurella 16S rRNA gene alignment on Geneious Prime. This shows much better predicted binding.

2.4 Discussion

This chapter aimed to determine a suite of primer pairs to be used throughout the isolation protocol to inform on both the presence of the target organism and any contaminating organisms. The initial target for this thesis was the class Epsilonproteobacteria; thus, the EPS_F/EPS_R primers were used.

This was reclassified into the novel phylum Campylobacterota with the inclusion of Desulfurellales, and subsequently, the target was narrowed down to *Desulfurella sp*. This necessitated designing novel primers suitable for this work. EPS_FM/EPS_RL were designed to target *Desulfurella* while creating a longer amplicon for sequencing. The forward primer was edited from EPS_F/EPS_R, but a new primer was designed to make a longer amplicon for future sequencing efforts instead of just editing the reverse. The EPS_F/EPS_R primers only amplify a 150BP region, which is too short for Sanger sequencing. A conserved region among the alignments was used to make EPSL-R2 amplify a region of 880BP.

These EPS_FM/EPS_RL primers were utilised initially, but it became clear from *in silico* analysis that a far more specific primer pair would be more beneficial to this project. While these primers can be utilised initially to check for Campylobacterota more broadly, more specific primers are needed to confirm the target is present at the site and for each passage to ensure it is not being lost. As shown in Table 2.2, the EPS_FM/EPS_RL primers may amplify a fair amount of non-target bacteria as well, which would be detrimental to the targeted approach of this thesis. Out of the 89 other phyla that are targeted by EPS_FM/EPS_RL, some have a low percentage of species hit (<30%). Other phyla targeted by these primers have a higher number of species predicted to amplify (>50%). For example, they target 52% of species in the Desulfobacteria phylum. This is unsurprising as the Desulfobacteria phylum occurred through the reclassification of the Deltaproteobacterial class into four new phyla. Therefore, the high recovery within this phylum is recovering previously closely related taxa (D. W. Waite et al., 2020). This result demonstrates the need to query the designed primers over a broad taxonomic range to account for reclassifications in the database.

The DS_F/DS_R primers were designed only to amplify *Desulfurella sp.* These primers do not amplify any other Campylobacterota and, as a result, are highly specific to *Desulfurella.* Table 2.2 summarises this, with 88% of *Desulfurella* targeted with no other species. These primers can be used with EPS primers for initial site inspection for different stringency levels. EPS_FM/EPS_RL can be used as a "universal" for Campylobacterota in general, DS_F/DS_R

tells us if *Desulfurella sp.* is present, and EPS_F/EPS_R is used to identify if it is another Campylobacterota (negative for *Desulfurella).*

The EubB/EubA primers are often used to amplify bacterial SSU genes as they are "universal" and target the whole 16S rRNA gene. However, against the *Desulfurella sp*. alignment in Geneious Prime, these primers look as though they do not hit all of the target sequences. Some of the *Desulfurella* 16S rRNA sequences seem truncated at the start and end, meaning these primers may be missing it (Figure 2.3). Therefore, The Earth Microbiome primers (515F/926R) will be used for whole 16S rRNA sequencing which look to amplify the target perfectly (Figure 2.4). "Universal" primers/published primers, such as those mentioned above, all have limitations. No universal primer can target all bacteria, so it is important to be aware of what might be missed.

In conclusion, the *Desulfurella* primers (DS_F/DS_R) designed de novo are highly specific for this study. They will be used in the coming chapter to check for the presence of the target at the initial site and at each passage step to ensure the target is not lost. EPS_F/EPS_R will be used for an initial site inspection to assess if any other Campylobacterota are present and before sequencing as a negative control for *Desulfurella sp.* EPS_FM/EPS_RL were also used initially as a broader, more "universal" Campylobacterota primer pair.

Chapter 3: Targeted cultivation of a sulphur-reducing, obligate anaerobe

3.1 Abstract

Microorganisms are ubiquitous and play essential roles in ecosystems, yet approximately 99% remain uncultivated, limiting their study and application. Traditional cultivation methods primarily succeed with easy-to-cultivate species, leaving the majority—often termed "Microbial Dark Matter" (MDM)—unexplored. The ability to target and grow what is specifically wanted from an environment and not just what is easy would be an essential step forward to studying this MDM. This chapter aims to outline a workflow combining classical cultivation and molecular techniques to provide a general approach to isolate a sulphur-reducing obligate anaerobe, *Desulfurella sp*., from a geothermal pool in Rotorua, New Zealand. Initially, the target was Epsilonproteobacteria, which was later narrowed down to, more specifically, *Desulfurella sp.* The sites were first selected based on varying abundances of Epsilonproteobacteria from data collected by the 1000 Springs Project, which, upon re-testing, showed marked differences in community structure since the original study. Site 36 still possessed *Desulfurella,* although at a significantly reduced relative abundance, this site was chosen as the inoculum for culturing. The medium was adjusted based on the sampling environment and published genomes, changing pH and temperature, respectively, to attempt to isolate the target organism. As one of the aims of this work was to domesticate the target onto solid media, I also made an alteration by replacing elemental sulphur with sodium polysulphide to increase the bioavailability of sulphur in the solid media. To avoid the need for full-sized anaerobic chambers, making cultivation of obligate anaerobes more accessible to various laboratories, a vacuum food bag sealer was used to culture these solid media plates in an anaerobic headspace. Automated Ribosomal Intergenic Spaces Analysis (ARISA) was used throughout cultivation to assess the complexity of the cultures before sending them for shotgun and 16S rRNA gene sequencing with universal and species-

specific primers designed in Chapter 2. 16S rRNA sequencing recovered a single *Desulfurella* ASV in each of the culture conditions, which, when phylogenetically placed in the species 16S rRNA gene tree, supports the genome recovered from the metagenome assembly. Due to time constraints, shotgun sequencing was conducted before confirming the successful isolation of cultures, treating them as mini-metagenomes. Unexpectedly, the shotgun sequencing did not detect *Desulfurella* in the cultures, which was perplexing. Species-specific primers (DS_F/DS_R) are highly specific and consistently showed amplification throughout the isolation process.

Furthermore, 16S rRNA gene sequencing of these amplicons from the isolation cultures revealed a match to *Desulfurella* when compared with the NCBI 16S rRNA database. The most likely explanation is that *Desulfurella* was initially present but was eventually outcompeted by other taxa, with PCR detecting DNA from dead cells. The results from this chapter reveal the potential pitfalls and challenges of trying to target cultivation for a novel environmental isolate. It also provides a starting point to begin the culturing process for those looking to culture what they want, not just what is easy, with the necessary molecular tools and media alterations.

3.2 Introduction

Microorganisms are in every environment globally and play crucial roles in maintaining ecosystem functions in these habitats. Yet, estimates suggest that up to 99% of microbial species have not been cultivated in a laboratory (Liu et al., 2022; Lloyd Karen G. et al., 2018; Steen et al., 2019). This dates back to the 19th century and includes techniques such as selective nutrient media, colony picking from solid media and microscopy (Lewis et al., 2021). However, these techniques pose several problems. First, selective nutrient media without prior knowledge of a microorganism's genome may not target the organism of interest, meaning that only easy-to-cultivate microorganisms have been studied as isolates (Hedlund et al., 2014). Slow growth rates, competition, low abundance, and symbiosis of the target microorganisms all contribute to the difficulty in recovering a single fastidious microbe from the environment. These problems create

a skewed knowledge base for the tree of life of those with cultivated representatives compared with the underrepresented, uncultivated majority (Lewis et al., 2021).

The advent of cultivation-independent techniques (Voit, 2020) allowed the study of the genetic material of community members without cultivation. Initially, 16S rRNA sequencing for environmental samples revealed large numbers of taxa that had yet to be cultivated (Solden et al., 2016). This shed light on the discrepancies between cultivated isolates and what is potentially living in an environment (Lewis et al., 2021) and has revealed that the vast majority of microorganism species remained hidden due to the previously mentioned drawbacks of cultivation (Solden et al., 2016). This uncultured majority has been termed the Microbial Dark Matter (MDM) (Schultz et al., 2023). Following these initial 16s rRNA metagenomic sequencing studies, their presence is now regularly confirmed by cultivation-independent methods such as whole shotgun metagenomics.

Metagenomes can decipher microbial communities' functional capacities and reconstruct their roles in biogeochemical processes without the need for any cultivation. This entails bulk sequencing of all DNA from an environment before analysis at either the whole community level (Bhattacharjee et al., 2017; B. Ma et al., 2024) or genomic level through binning into metagenome-assembled genomes (MAGs) (Cárdenas-Hernández et al., 2023; Li et al., 2023) or both (Maruyama et al., 2024; Walters et al., 2023). However, whilst metagenomes and MAGs have undoubted benefits, they each have weaknesses that need to be considered, as issues can arise if MAGS are treated as analogues for cultivated isolates. The first potential pitfall when using metagenomics is the creation of a fragmented assembly that can break up genes. Fragmented assemblies commonly occur due to low-coverage genomes, intragenomic repeats, and elevated levels of strain heterogeneity in the environment. These scenarios all present difficulties for the assembler of choice to recover the correct sequence, thus creating a multitude of short contigs. This problem carries over to the creation of MAGs as the binning algorithms rely on the underlying assembly and often have a minimum contig size of at least 1000 bases

(e.g. MetaWRAP (Uritskiy et al., 2018)). This combination leads to the recovery, if at all, of low completeness and contaminated MAGs.

Furthermore, repetitive sequences, such as ribosomal RNA, are fragmented in the original assemble and so missing in the final MAGs, whilst certain exogenous sequences, such as horizontally acquired genes, are present in the assembly but possibly binned incorrectly. Additionally, Mise and Iwasaki (2022) found that ribosomal genes are often missing from MAGs, even though they are neither exogenous nor repetitive, and results from special evolutionary patterns in ribosomal proteins and metagenomic binning algorithm characteristics (Mise & Iwasaki, 2022). This is particularly problematic as tools for taxonomic (Parks et al., 2018) and completeness (Parks et al., 2015) analysis of MAGs routinely use these genes in their marker gene sets. The interpretation of genes is biased by the physiology of a few well-known and studied model organisms, and with MAGs, there is a tendency to see a gene is present and assume it has that function. Tools for the annotation of genomes and reconstruction of metabolic functions are available to some extent; however, currently, there are no standardised approaches for the comprehensive analysis of microbial interactions and their contributions to biogeochemical cycling, metabolic prediction, and metabolite exchanges (Zhao et al., 2023).

There has, therefore, been a surge in genomic data since the rise of metagenomes and MAGs, but the cultivation of these organisms still needs to be improved. With this surge in genomic data, the bottleneck now lies with translating these sequences into functional context. Some efforts to cultivate fastidious microbes have been successful but have been labour-intensive (van Teeseling & Jogler, 2021b). Cultures can build bridges to help gaps in knowledge by isolating and studying unusual organisms (Carini Paul, 2019; Lewis et al., 2021) , so combining traditional cultivation methods and omics tools would be very beneficial in solving the issue of the uncultivated majority. Isolates of bacteria are indispensable and are the foundation of microbiology research for understanding evolution, cell biology, and ecological roles and impacts. There is a push to find ways of bringing these poorly characterised groups into culture, as isolates can have a range of applications, such as industrial processes, medicine, and various applications in biotechnology

and agriculture (Liu et al., 2022). As well as furthering our understanding of their roles in the environment such as impacts on biogeochemical cycles. They are invaluable tools for understanding, harnessing, and utilising the microbial world for the benefit of science and society. Research must have cultivated isolates to provide unambiguous assignment of genes and their function. Isolated microorganisms provide unambiguous assignment of genes and are used in research to understand how the proteins they encode function (Gupta et al., 2021; Liu et al., 2022; Zengler, 2009). This can be achieved by gene knockout mutations, which is the process of changing or deleting a gene in the intact organism and seeing how differently it functions. An example of this is the well-known Keio collection for *Escherichia coli* (Baba et al., 2006).

Although bringing microbial dark matter into culture is challenging, it expands our knowledge base and provides many benefits. This chapter aims to provide an easy-to-follow, straightforward workflow to cultivate targeted environmental organisms using common molecular tools and methods that are readily accessible in any standard microbiological laboratory. The overarching goal was to culture specifically what is wanted, not what is common. Therefore, I will combine classical cultivation and molecular techniques to culture a sulphur-reducing, obligate anaerobe from an environmental sample. Namely a *Desulfurella sp.* from a geothermal pool in Rotorua, New Zealand. This was chosen as a target as it is a basal lineage that is underrepresented in databases online and there is a lack of cultured representatives in culture collections. This thesis is presented in two parts: 1) using molecular tools to ensure the inoculum contains the target and 2) enriching the target using knowledge of genomes, published literature, and the environment in which it was sampled. Geothermal sites in Kiurau Park, New Zealand, were targeted based on varying abundances of *Desulfurella sp.* and different microbial complexities from the information presented in the 1000 Springs Project (https://1000springs.org.nz/). I confirmed the presence of our target taxa by 16S rRNA libraries, metagenomes and 16S rRNA gene amplicon PCRs with primers designed at different stringency levels (chapter 2). ARISA (automated ribosomal intergenic spacer analysis) and microscopy were utilised to assess how close the culture was to being axenic. Liquid cultures were then domesticated onto solid media for colony

picking to improve the axenic nature of these cultures, and DNA was extracted for genome sequencing.

3.3 Methods

3.3.1 Site Selection

I used the data from the 1000 Springs Project (https://1000springs.org.nz/) (Power et al., 2018) to identify geothermal pools in Kuirau Park, Rotorua, with Campylobacterota present, as this class was the original broader target, which was later narrowed down to a more specific target of the species *Desulfurella sp*. Table 3.1 shows all these sites with the Campylobacterota genera present, Campylobacterota's percentage of the community, alongside metadata of the pool and published media recipes for the specific genera. Figure 3.1 shows an example of the 1000 Springs website for Kuirau Park. The website shows a map with easy-to-follow markers for each feature (Figure 3.1a). For each site, the website provides a general overview (Figure 3.1b), with subsequent pages providing detailed geochemistry (Figure 3.1c) and taxonomic (Figure 3.1d) data. We selected four of these sites for preliminary trials: sites 50, 39, 22 (Figure 3.2) and 36 (Figure 3.1b-d), as these represent different levels of Campylobacterota community complexity: complex (site 22) and simple (36). Specifically, I decided to focus on *Desulfurella sp.* as a group, which is the root group of the Campylobacterota and has limited species isolated in culture collections.

Table 3.2. Sites of interest at Kuirau Park, Rotorua, New Zealand. With associated data from the 1000 Springs Project. Sites marked in bold are those selected for initial trials.

Site	Genus	Percentage	pH	Temperature (°C)	Complex?	Recommended media (DSMZ)

Kuirau Park Feature 2	Nitratiruptor	59.6	7.1	56.8	No	1024
Kuirau Park Feature 16	Nitratiruptor	46.5	5.43	39.7	No	1024
	Sulfuricurvum	1.65				1020
Kuirau Park Feature 22	**Sulfuricurvum**	**50.2**	**5.16**	**31.1**	**Yes**	**1020**
	Sulfurimonas	**0.163**				**1011**
	Desulfurella	**2.89**				**480**
Kuirau Park Feature 27	Sulfuricurvum	28.7	6.13	33.6	Yes	1020
Kuirau Park Feature 28	Sulfuricurvum	41.9	6.07	62.6	No	1020
Kuirau Park Feature 30	Sulfuricurvum	57.2	6.47	26.4	Yes	1020
	Sulfurimonas	2.85				1011
	Sulfurovum	0.382				1570
	Sulfurospirillum	3.95				541
Kuirau Park Feature 36	**Desulfurella**	**32.3**	**2.9**	**30.8**	**No**	**480**
Kuirau Park	**Sulfurimonas**	**29.2**	**4.27**	**35.4**	**Yes**	**1011**
	Sulfuricurvum	**6.99**				**1020**
	Desulfurella	**4.46**				**480**

Feature 39						
Kuirau Park Feature 43	Sulfurimonas	29.3	4.5	29.3	No	1011
	Desulfurella	5.44				480
Kuirau Park Feature 50	**Sulfurovum**	**47.6**	**5.97**	**29**	**Yes**	**1570**
	Sulfuricurvum	**19.1**				**1020**
	Sulfurimonas	**0.2**				**1011**
Kuirau Park Feature 58	Nitratiruptor	80	5.75	54.5	No	1024

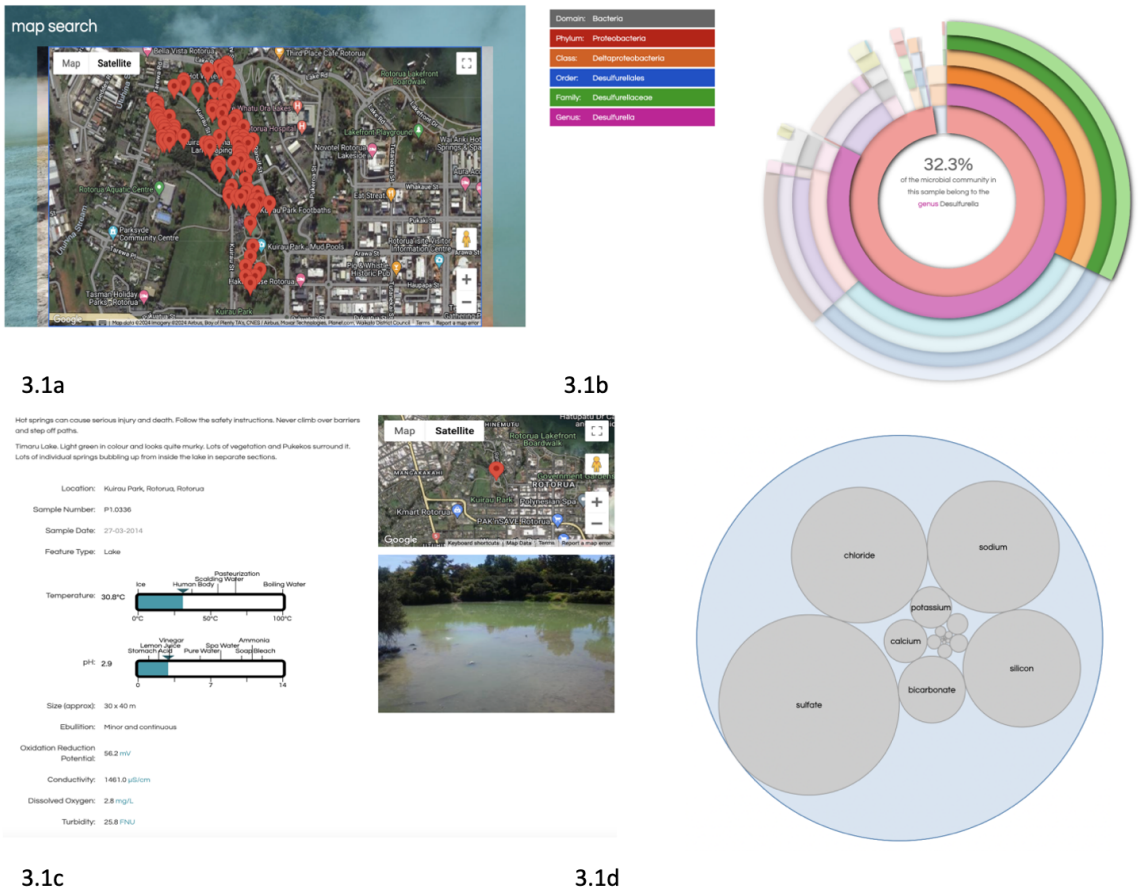


Figure 3. 1. 1000 springs data for feature 36 at Kuirau Park. 3.1a. Map of all the sites sampled. 3.1b. Microbial diversity in feature 36. 3.1c. Feature 36 data with a map of the site and photo. 3.1d. Feature 36 chemical properties.

Figure 3. 2. 1000 springs data for feature 22 at Kuirau Park. 3.2a. Microbial diversity in feature 22. 3.2b. Feature 22 data with a map of the site and photo. 3.2c. Feature 22 chemical properties.

3.3.2 Sample Collection

Five litres of water were collected in sterile Nalgene bottles for metagenome sequencing, and three 50mL falcon tubes were collected for culturing, PCR and pH analysis. These were stored at room temperature, and the head space in the tubes was minimal to avoid excess oxygen presence until use. Culturing was performed as soon as possible once back in the laboratory. Serum vials containing the correct media (Table 3.1) and atmospheric conditions (each site and media type had one vial prepared under H2:C02 and one under N2:CO2 at 80:20) being prepared before sampling. At sampling, water temperature was also measured using a digital thermometer.

3.3.3 PCR with Platinum Direct PCR Universal Master Mix

The Invitrogen Platinum Direct PCR Universal Master Mix was utilised in this project as it is designed to amplify DNA directly from various samples without the need to purify DNA. There

is a direct protocol that allows samples to be aliquoted directly into the master mix without extracting the DNA first. The kit also provides a lysis step that can be performed before amplification to easily release DNA from the cell if preferred. Initially, the lysis step was utilised before primer conditions were optimised. However, after optimising the primers, I found that the direct protocol was able to be used successfully. All results presented later in this thesis used the direct protocol unless otherwise specified. Primers optimised for annealing temperature in Chapter 2.

Table 3.3. Volumes used in PCR reaction.

Component	**20uL reaction**	**Final concentration**
Platinum Direct PCR Universal Master mix 2X	10uL	1X
Forward primer (10ng/uL)	0.4uL	0.2uM
Reverse primer (10ng/uL)	0.4uL	0.2uM
Nuclease-free water	8.2uL	-
Sample	1uL	-

Table 3.4. Recommended universal PCR thermocycling conditions.

Step	**Cycles**	**Temperature (°C)**	**Time**
Activation	1	94	2 min
Denaturation	35-40	94	15 sec
Annealing		60	15 sec
Extension		68	20 sec/kb
Hold	1	4	Hold

3.3.4 Water Sample Processing and DNA Extraction

As the water collected was murky and contained sediment, it was spun down in 50mL falcon tubes to pellet the biomass. The supernatant was filtered through a Sterivex filter to recover any biomass missed during centrifugation. This resulted in two separate tubes for DNA extraction for each site, 22.1 and 36.1 for the Sterivex extraction and 22.2 and 36.2 for the pellet extraction. Once pelleted, total DNA was extracted from each using the Cary lab CTAB soil extraction protocol for low biomass samples. Briefly, 0.5mL of the sample was put in a tube containing 0.1 mm and 2.5 mm silica-zirconia beads. 300 µL phosphate buffer (100mM NaH_2PO_4) and 300 µL SDS lysis buffer (100mM NaCl, 500mM Tris pH 8.0, 10% SDS) was added to the tube. Samples were then bead-beaten for 10 minutes on a vortex genie. Samples were then centrifuged at 10,000rpm for 3 minutes. The supernatant was mixed with 200uL CTAB buffer + 0.5% BME, incubated at 300rpm at 60°C for 30 minutes, and centrifuged at 10,000rpm for 30 seconds. I added 350 µL chloroform: isoamyl alcohol (24:1), vortexed the sample for 15 sec and then centrifuged at 10,000rpm for 5 minutes. After this, I removed the upper aqueous layer into a new 1.5mL sterile microfuge tube, added 500 µL chloroform: isoamyl alcohol (24:1), vortexed for 10 seconds, and then left it on a rocking bed for 20 minutes at 20°C. This was then centrifuged at 13,200rpm for 5 min, and I removed the upper aqueous layers into a new 1.5mL sterile microfuge tube. 10M ammonium acetate was added to the sample to a final conc. of 2.5 M, vortexed and then centrifuged at 13,200rpm for 5 min. The upper layer was then transferred to a new sterile microfuge tube, and I added 0.54x volume of isopropyl alcohol, mixed by repeated inversion (20 times), and then incubated at -20°C for 48 hours. Following this, I centrifuged at 13,200rpm for 20 min and discarded the supernatant to leave the pellet, which I washed with 1mL 70% ethanol. This ethanol wash was centrifuged at 13,200rpm for 1 min, and then the ethanol was pipetted off. The resulting pellet was dried in a speed vacuum for approximately 4-15 min (medium drying setting). Finally, I re-suspended the DNA pellet in 20 µL sterile ultra-pure water, then stored at -80°C until use.

3.3.5 DNA Sequencing Processing

The 16S rRNA library was prepared using the universal EMP16S NexteraXT-V2 fusion primer set. The resulting amplicons went through SequelPrep and were sequenced on the NovaSeq 6000. In this study, sequence analysis was performed using BioConductor 3.17 (Huber et al., 2015), RStudio 4.3.0 (R Core Team, 2023), and DaDa2 v1.28 (Callahan et al., 2016). The data were processed through the DADA2 pipeline using default options presented in their tutorial (https://benjjneb.github.io/dada2/tutorial.html). The quality profiles of forward and reverse reads were visualised to determine the desired truncation length. Following quality assessment, filtering and trimming processes were implemented to enhance data integrity. Error rates were estimated, and sample inference was performed based on the generated error model. Subsequently, paired-end reads were merged, and an amplicon sequence variant (ASV) table was constructed. Chimeric sequences were removed, and tracking metrics were generated to monitor the analysis pipeline's performance. After this site 22.1 had no reads and 36.1 had only 36, compared to 22.2 and 36.2 having 18344 and 3018 reads respectively. This is likely because 22.2 and 36.2 were extracted from pelleted samples, while 22.1 and 36.1 were filters of the supernatant from this, so all of the cells appear to have been spun down into this pellet. Due to this, it was decided only the pelleted DNA-extracted sequences would be used in this study. Taxonomic assignment was conducted using the SILVA138 training set, refined to the species level. Finally, readable information arrays containing metrics, taxonomic assignments, and ASV distributions were exported for further analysis.

3.3.6 Metagenome Analyses

Prior to assembly, metagenomic reads (Illumina 150bp paired) were trimmed with bbduk from BBMap (v39.01) (Bushnell, 2014) to remove Illumina sequencing adapters, poly-Gs, and to a quality above 20 over a sliding window. These quality trimmed paired reads were then further cleaned before being assembled with SPAdes (v3.15.4) (Nurk et al., 2017, 2017). Specifically, BayesHammer (Nikolenko et al., 2013) was used to error correct the input read set using default

settings, before assembly using specific kmers (21, 33, 55, 77, 99 and 127) with the --meta setting. Assembled contigs were binned into MAGs using the metaWRAP pipeline (v1.2.1) (Uritskiy et al., 2018). MetaWRAP uses three different binning tools in its pipeline: MetaBat2 (v2.12.1) (Kang et al., 2015), MaxBin2 (2.2.6) (Wu et al., 2016), and CONCOCT (v1.0.0) (Alneberg et al., 2014). This creates three separate bin sets that are then dereplicated to the final bin set. The final dereplicated bin set was assessed for completeness with CheckM2 (Chklovski et al., 2023) and taxonomically placed using GTDB-tk (v2.3.2) (Chaumeil et al., 2022) that used the GTDB r214 database (Parks et al., 2022). The multiple sequence alignments outputted from the GTDB-tk pipeline was subsetted to contain all representative species of the class Desulfurellia, as well as one species from the phylum preceding the Campylobacterota in the GTDB tree (DQVG01 sp015661865 from the phylum Aquificota) to act as a root for the phylogeny. This multiple sequence alignment was input to FastTree (Price et al., 2009, 2010) (v.2.1.10) using the –lg –wag –cat options. The phylogeny was viewed and rerooted in FigTree (v1.4.4) before being plotted using ggtree (Yu et al., 2017) (v3.8.0) in R v4.3.2 (R Core Team, 2023). The community profile of the metagenome was estimated using SingleM (Woodcroft et al., 2024) (v0.18.1) using the raw reads, which uses the GTDB r220 database. First, singlem pipe was run to estimate the read coverage of single copy marker genes for all taxonomic levels. Secondly, singlem summarise was run on the profile output from singlem pipe to give an estimate of relative abundance at desired taxonomic levels (phylum and family). Had this also been done before commencing culturing, metabolic pathway reconstruction could have helped with altering culturing conditions best suited for the target species.

3.3.7 Campylobacterota Presence in Water Samples

Camplyobacterota presence was confirmed in the water sample by 16S rRNA amplification using EPS_F/EPS_R (Chapter 2). These primers were used directly on 1uL of the water samples using the Platinum Direct PCR Universal Master Mix by Invitrogen. As discussed in Chapter 2 (primer design), more primers were designed (EPS_FM/EPS_RL and DS_F/DS_R) in this study and were later applied to the same sample for sites 22 and 36.

3.3.8 Culturing Conditions

Initially, the culturing was targeted for *Desulfurella* and *Sulfuricurvum* using modified DSMZ480 *Desulfurella* medium (A. P. Florentino et al., 2015) and modified DSMZ1024 *Nitratiruptor* and *Nitratifactor* medium (appendix) to target *Sulfuricurvum* (Nakagawa et al., 2007). Gas mixes of 80:20 H_2:CO_2 and N_2:CO_2 were tested as well as two different temperatures of 30°C and 50°C based on the pool temperature at sampling and optimal temperature of *Desulfurella* in the literature (A. P. Florentino et al., 2015). These conditions were also originally used to target *Sulfuricurvum* as previously stated, the target had not been narrowed down to just *Desulfurella* so culturing started out more broadly for Epsilonproteobacteria in general.

3.3.9 Transfers

Passages to fresh media were done every two weeks, originally transferring 3mL in 20mL of broth but changing down to approximately 100uL into 20mL to speed up isolation by reducing the amount of dead cell debris that non-target bacteria could use as an energy source. Every transfer was checked through PCR with DS_F/DS_R primers to ensure the target was not lost and to visually inspect community diversity microscopically. This process spanned over two years.

3.3.10 Culture Media Adjustments

Sodium polysulphide was used to make sulphur bioavailable in solid media. However, this caused the media to become too alkaline. Phosphate and MOPs buffers were compared in liquid media to see which buffer would work best to combat this and keep the pH at 6. Alongside the transfers every two weeks, samples of different media conditions for buffers were also trialled for three months. The control was keeping the media unchanged; the other conditions were MOPs + regular media, MOPs + sodium polysulphide + exclusion of elemental sulphur, phosphate buffer + regular media, phosphate buffer + sodium polysulphide + exclusion of elemental sulphur. MOPs were then used throughout.

To help speed up the isolation process, other conditions for liquid media were trialled out. First, the pH of the media was adjusted to pH 4 as it is closer to the environmental conditions at the sampling site, and also a pH more favourable to *Desulfurella*. In the literature, *Desulfurella* has also been found to have the optimum growth temperature of 50°C (A. P. Florentino et al., 2015), which was also tested. An increase in temperature also correlates with a decrease in pH, which is likely to have an effect as well. All three versions of primers from Chapter 2 were tested on these adjustments.

3.3.11 Solid Media

Initially, stab cultures were utilised to assess which viscosity of agar was best to grow the cultures on. Tubes with 40%, 60%, 80%, and 100% agar (where 100% is the recommended viscosity of 15g/L) were made into the modified 480 media by keeping the nutrients at the same concentration and decreasing the amount of agar used (3mg/mL, 6mg/mL, 9mg/mL, 12mg/mL and 15mg/mL respectively). 2mL of the highest concentration of agar was put in the bottom of a tube sitting in an ice bath so that the agar set quickly. Once set, 2mL of the next highest concentration was put on top to create the gradient. A wire was flamed and put into the culture, then stabbed down through the centre of the different viscous solid media. The tube was capped, sealed and incubated at 30°C for four weeks.

A food-grade vacuum bag sealer was used to seal the plates in the preferred anaerobic conditions. The plates were put into a bag with an anaerobic gas-generating sachet and anaerobic indicator to ensure all residual oxygen was removed. All the air was drawn out using the vacuum sealer and was sealed off twice to ensure no holes might let air in. Using a disposable 25g x 1" Microlance 3 needle, the gas mix (N_2:CO_2 at 80:20) was then added, which was also sealed off twice. In a few hours, the indicator was checked to see if it had gone from bright pink to white, confirming that the atmosphere was anoxic. Plates were incubated at 30°C for four weeks minimum. This was used to further isolate by colony picking, taking one small, frosted colony

(FC) and one bigger colony (BC) and growing them back in liquid media. I transferred these onto solid media again and left them for six weeks to grow enough biomass to extract DNA.

3.3.12 Automated Ribosomal Intergenic Spacer Analysis (ARISA)

The automated method of ribosomal intergenic spacer analysis (ARISA) can be used to estimate microbial diversity and community composition (Fisher & Triplett, 1999). In this case, it was used to check the axenicity of culturing efforts. 1µL of cultures were directly put into the Platinum Direct Universal Master Mix as described above, and PCR amplification was performed with 1406f/23Sr (Purahong et al., 2015) these ARISA primers have a broad target range and are used to estimate microbial community diversity by amplifying the Intergenic spacer between the 16s and 23s rRNA genes. As this spacer region is under less stringent selection than rRNA genes, it is more prone to changes that can allow different taxa to be distinguished on the amplicon size without sequencing. This was then sent for Sanger sequencing (Applied Biosystems 3730 Genetic Analyzer) at the University of Auckland with ARISA-PCR fragments varying in size from 400 to 1,200 bp. A LIZ1200 ladder was used with matrix D selection DS30. This gave a good indication of how mixed communities were or if the culture appeared axenic.

This was substituted for the Agilent 4200 TapeStation System for automated electrophoresis when one became available in the laboratory. This was compared with samples that had already been run through sequencing, and it gave similar results, which appears sufficient to estimate community diversity.

3.3.13 16S rRNA Gene Amplicon Analysis of Enriched Culture

DNA was extracted from the enriched cultures using the same extraction method as above (CTAB modified for low biomass). The cultures with enough DNA yield to be sent for sequencing were BC6 and BC4 (big colony at pH6 and pH4), FC6 and FC4 (frosted colony at

pH6 and pH4), and 50°C plates were diluted where necessary to 5ng/uL and amplified using both 515F/926R (i.e., universal bacterial 16S rRNA gene) and DS_F/DS_R (Camplyobacterota-specific) primer pairs. Samples were sent to AUT for 16S rRNA gene sequencing, and the results were analysed using Geneious Prime. Chromatograms of base calls were visually inspected for quality and manually edited with the sequences' start and end being trimmed. Forward and reverse reads were aligned using global alignment with free end gaps with a cost matrix of 93% similarity, gap open penalty 12, and gap extension penalty 3. Each sample's final 16s amplicon sequence was queried against NCBI's Bacterial and Archaeal 16s ribosomal RNA sequences database using megablast (Morgulis et al., 2008) of the BLAST suite (Altschul et al., 1990). From this output, all hits with a percentage identity over 85% had their sequence downloaded. All these downloaded sequences were input to AliView (Larsson, 2014), before using MUSCLE (Edgar, 2004) within AliView to align these raw sequences. The alignment was manually checked before used as an input to FastTree (v.2.1.10) (Price et al., 2009, 2010) using the -gtr option to create a 16s rRNA phylogeny. The phylogeny was viewed and rerooted in FigTree (v1.4.4) before being plotted using ggtree (v3.8.0) (Yu et al., 2017) in R v4.3.2 (R Core Team, 2023).

3.3.14 Culture DNA Extraction and Shotgun Sequencing

DNA was extracted from the enriched cultures using the same extraction method as above (CTAB modified for low biomass). The relative abundance of all microbes within the sequencing reads was estimated with singlem pipe (Woodcroft et al., 2024) and summarised, as described in the Metagenome analyses section. As the 16s rRNA sequencing, as well as other analyses (i.e. singlem relative abundance), showed these cultures were mixed, they were therefore treated as mini-metagenomes using the same assembly and binning protocol as the metagenomes specified above. For assembly, sequencing reads (Illumina 150bp paired) were trimmed with bbduk from BBMap (v39.01) (Bushnell, 2014) to remove Illumina sequencing adapters, poly-Gs, and to a quality above 30 over a sliding window. Despite this trimming of poly-Gs, many poly-Gs were

still present so a second round of cleaning using a different program, fastp (v.0.23.4) (S. Chen et al., 2018), was conducted to improve the quality of these reads.

3.4 Results

3.4.1 Site Analysis

To compare how these sites might have changed since those recorded in the 1000 Springs Project (Power et al., 2018), I retested the temperature and pH of the sites as these were the biggest drivers of community composition in that study. It is also important to have this information to inform culturing conditions. The pH and temperature of the sites are recorded in Table 3.4, which shows that there have been subtle changes in these sites compared to Table 3.1. However, this can still have a profound influence on community composition. Therefore, I conducted PCRs with the primers described in Chapter 2 to determine that the target organism was still present. The direct PCR performed straight from the water sample with EPS_F/EPS_R primers is shown in Figure 3.3. This was done before narrowing the target down to *Desulfurella* and before creating primers for this, so only EPS_F/EPS_R primers were used. Since sites 50 and 39 did not amplify with these primers, this indicates that the target is not present, so the focus was narrowed down to sites 36 and 22. After the design of more primers, these were tested on the original site samples for 36 and 22 (Figure 3.4). This shows that site 36 amplifies for all three primer pairs, but site 22 lacks amplification for DS_F/DS_R, meaning this site is lacking *Desulfurella sp.*

Table 3.5. pH and temperature of initial sites sampled at Kuirau Park compared to the original 1000 springs study.

Site	1000 springs Temperature (°C)	1000 springs pH	Current Temperature (°C)	Current pH	Temperature Difference (°C)	pH Difference
22	31.1	5.16	31.0	6	-0.1	+0.84

36	30.8	2.9	30.2	3.5	-0.6	+0.6
39	35.4	4.27	36.1	4	+0.7	-0.27
50	29	5.97	28	6	-1.0	-0.03

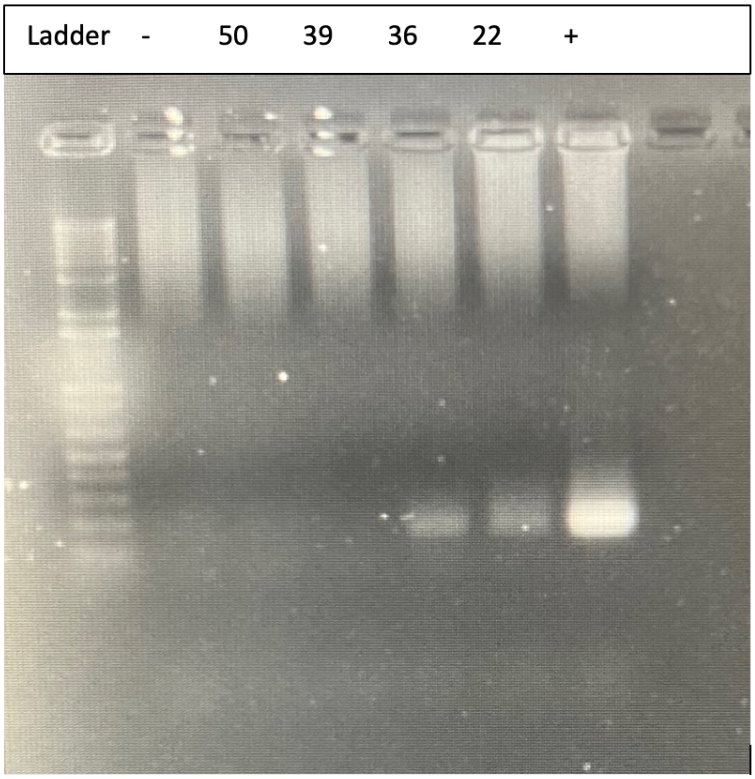


Figure 3. 3. The gel of the initial water sample of PCR with EPS_F/EPS_R primers. 1Kb+ ladder was used

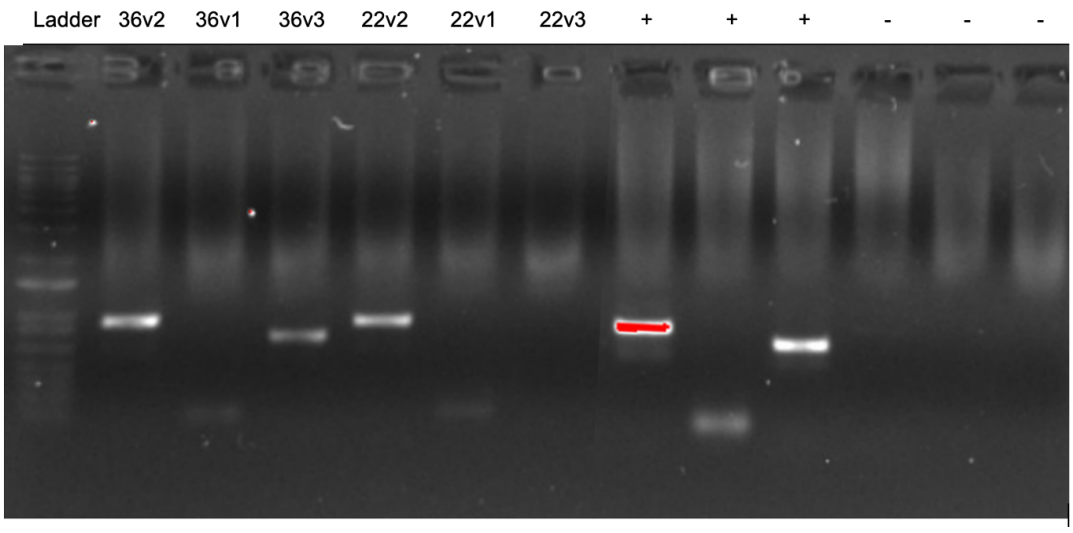
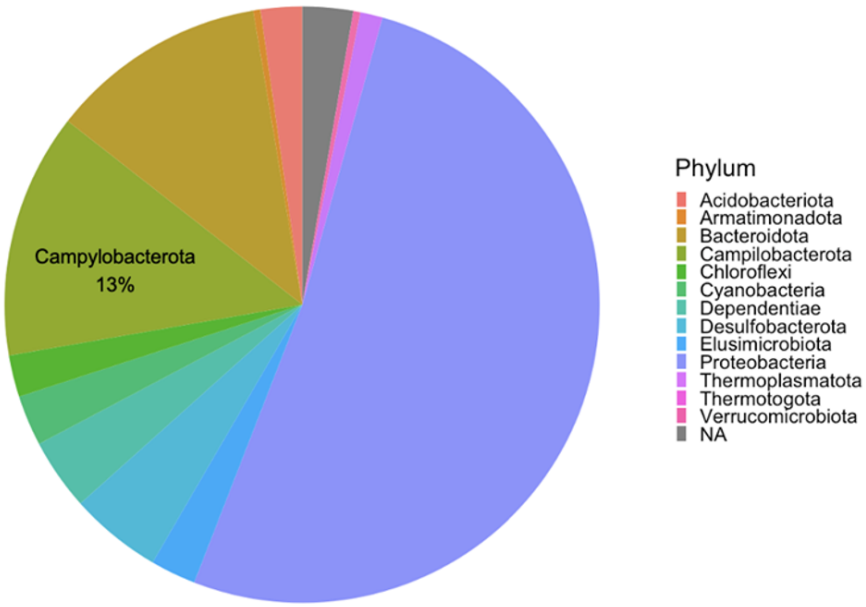


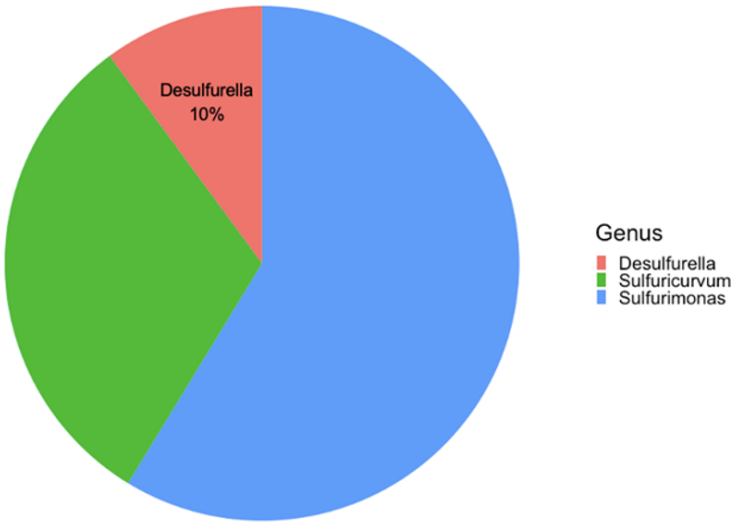
Figure 3. 4. The gel of sites 36 and 22 run with all 3 versions of primer pairs from Chapter 1. 1Kb+ ladder was used. Primer versions: V1 = EPS_F/EPS_R, V2 = EPS_FM/EPS_RL, and V3 = DS_F/DS_R.

3.4.2 16S rRNA Libraries

Similarly, to comparing the overall physiochemical properties of these sites, I also sought to compare the potential change in overall prokaryotic communities. The 16S rRNA libraries on sites 22 and 36 show that *Desulfurella sp.* is absent or exceedingly rare, respectively, as expected from the lack of amplification of DS_F/DS_R primers. Campylobacterota are present at 13% at site 36, with 10% of that being *Desulfurella,* with only one ASV present, representing 1.09% of the total community. These sites show pronounced differences in their prokaryotic community composition from the 1000 springs data. For instance, site 36 only had Desulfurella sp of Campylobacterota at 32.3% abundance, whereas the pool now has a more complex Campylobacterota community. This is expected from the change in pH since the 1000 springs (+0.6) as changing pH from an acidic to more alkaline resulted in more diverse communities (Power et al., 2018).



3.5a



3.5b

Figure 3. 5. 16S rRNA library for site 36 at Kuirau Park. 3.5a. at the phylum level, Campylobacterota makes up 13% of the total community with Desulfurella making up 1.09%. 3.5b. Pie chart of the genus level for the Campylobacterota, with Desulfurella accounting for 10%.

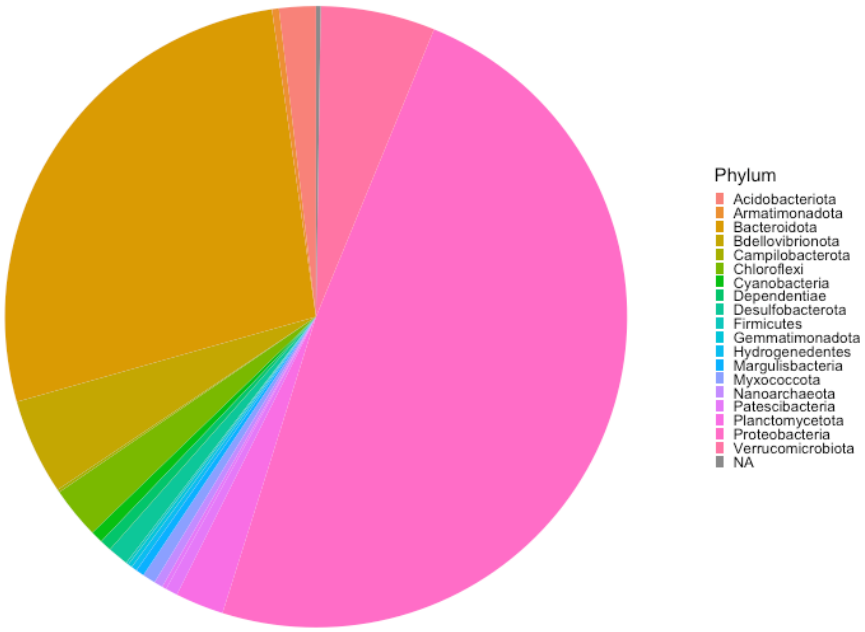


Figure 3. 6. 16S rRNA library for site 22 at Kuirua Park. No Desulfurella present.

3.4.3 Culturing

Change of Buffer

As the DSMZ 480 media was modified to use sodium polysulphide instead of elemental sulphur to increase bioavailability in solid media, this necessitated secondary changes. Specifically, a change of buffer was needed for a higher buffering capacity as the addition of sodium polysulphide increased the pH to 9 when it needed to stay at pH 6. Two buffers were trialled to keep pH in the desired ranges: phosphate (Phos) and MOPS buffers. The growth in DSMZ 480 containing elemental sulphur with either buffer was used as a control, with growth confirmed by amplification of the DS_F/DS_R primers. When grown in the controls, there is good growth, as expected, meaning these buffers have no adverse effect on the growth of the default media. However, when these buffers were tested with the DSMZ 480 media using sodium polysulphide, only successful growth was observed in the MOPS buffer. From then on, 10M MOPs buffer was used in the culturing media.

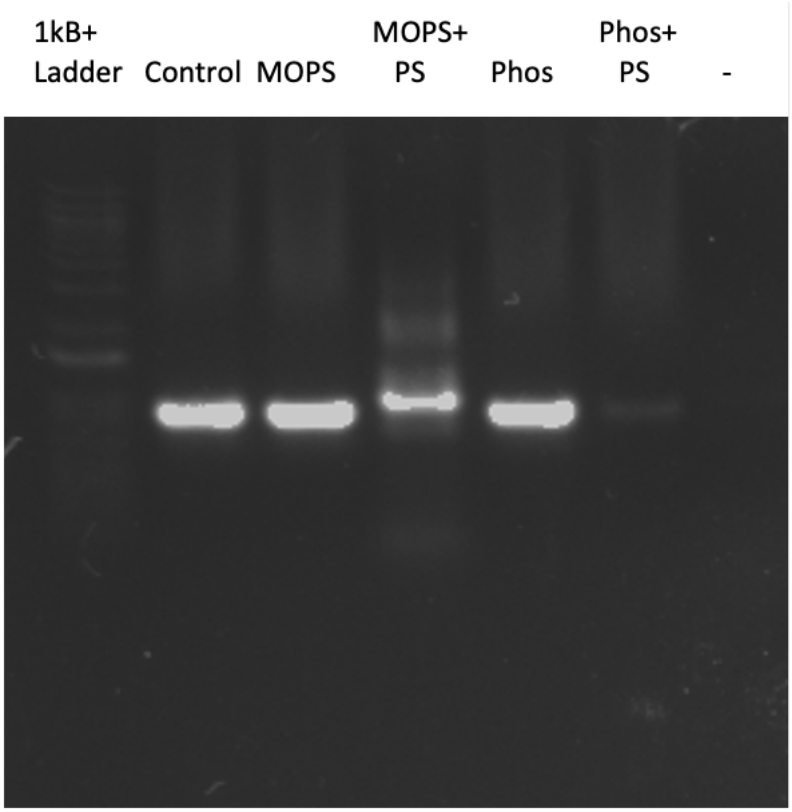


Figure 3. 7. Gel showing amplification with different buffer adjustments in the media using a site 36 isolate and DS_F/DS_R primers. PS = Polysulphide, Phos = phosphate buffer.

Solid Media

Before moving cultures onto solid media, I first wanted to determine at what agar viscosity growth was visible. For the different viscosities, I used different percentages of agar per litre of media, where 100% viscosity is 15g agar per litre of media. I created a stab culture with five different agar viscosities of 20%, 40%, 60%, 80%, and 100%. The culture was stabbed through the five viscosities, incubated for a month, and checked for visible growth. It is difficult to see in Figure 3.8, but growth was observed by visual inspection down to 80% in this stab culture method. This was then used to inform the plates, using 80% agar with the modified DSMZ480 media (sodium polysulphide + MOPS buffer).

The second aspect of this solid media protocol was to develop an easy-to-use system that can be more accessible than current anaerobic chambers. This was intended to provide a low-cost option

to easily grow obligate anaerobes or, for ease of accessibility, if culturing needed to be carried out in the field. I, therefore, tested the use of a vacuum bag sealer normally used for storing food to grow these cultures under anaerobic conditions. Figure 3.9 shows this idea of using a vacuum food bag sealer, whilst Figure 3.10 indicates the two colonies successfully grown and picked from solid media to be grown back up in liquid media. Various colony morphologies were observed, including small frosted white colonies, larger colonies, and mucoid-like larger colonies. Florentino et al. (2015) similarly reported the appearance of small (<0.5 mm) white colonies after four weeks when culturing *Desulfurella*. Based on this, I hypothesised that the small frosted white colonies were most likely the target species. To further investigate, I selected one small, frosted colony (FC) and one larger colony (BC) to grow for the rest of this project. These cultures were checked with species-specific PCR primers before each transfer. Figure 3.11 shows the last PCR run with all primer pairs after any media adjustments and just before sequencing. No cultures amplified for the broad Campylobacterota primers (EPS_FM/PES_RL), whereas all amplified for *Desulfurella* (DS_F/DS_R), and both BC treatments and the 50°C treatments amplified for Epsilonproteobacteria (EPS_F/EPS_R) but the FC treatments did not. From this, I expected the BC and 50°C treatments to contain a co-culture of *Desulfurella* and other Campylobacterota. This indicates that the FC treatment only contained *Desulfurella.*

Figure 3. 8. A stab culture of DSMZ480 with varying percentages of agar, starting at the bottom with 100%, 80%, 60% and 40%.

Figure 3. 9. Solid media is set up to create an anaerobic atmosphere of the N2:CO2 gas mix in a more accessible/portable way.

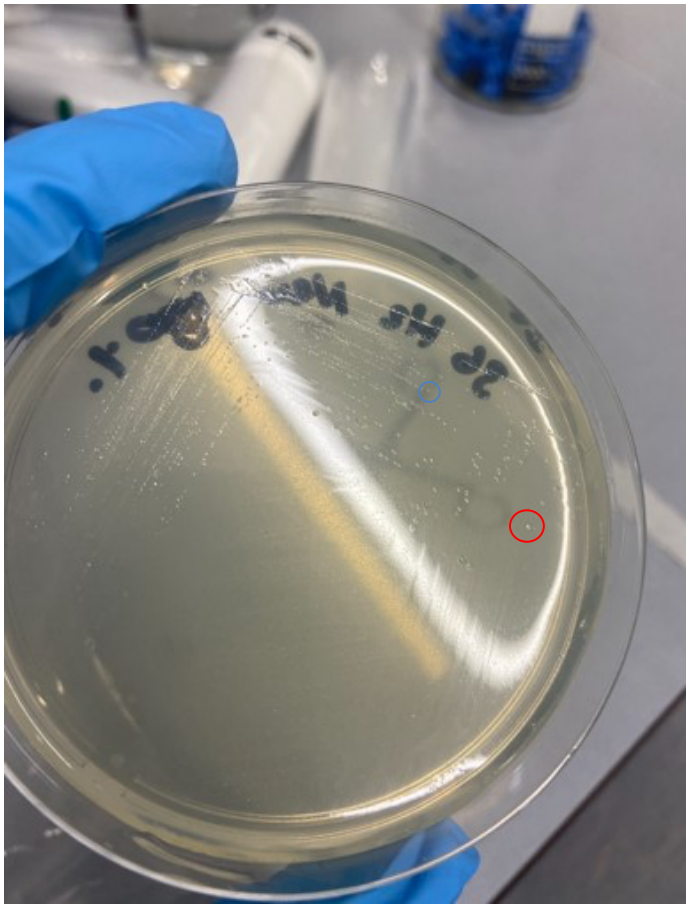


Figure 3. 10. Solid media with different morphology of colonies. Blue = small, frosted colony (FC). Red = Big, mucoid colony (BC)

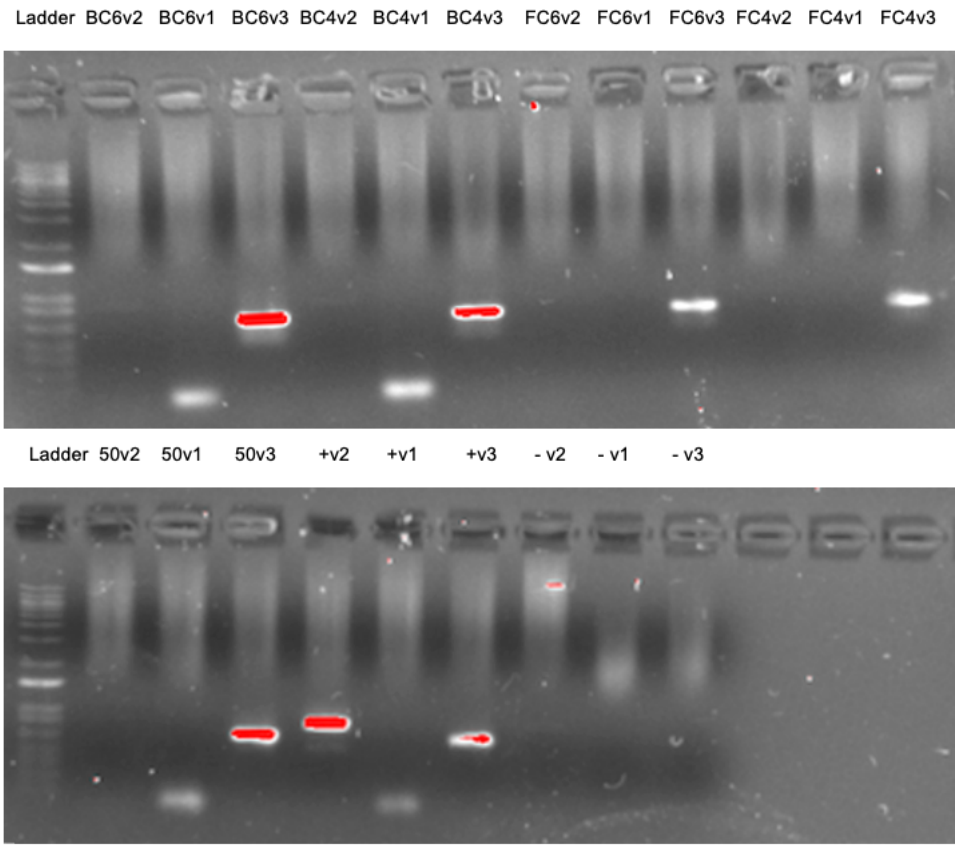
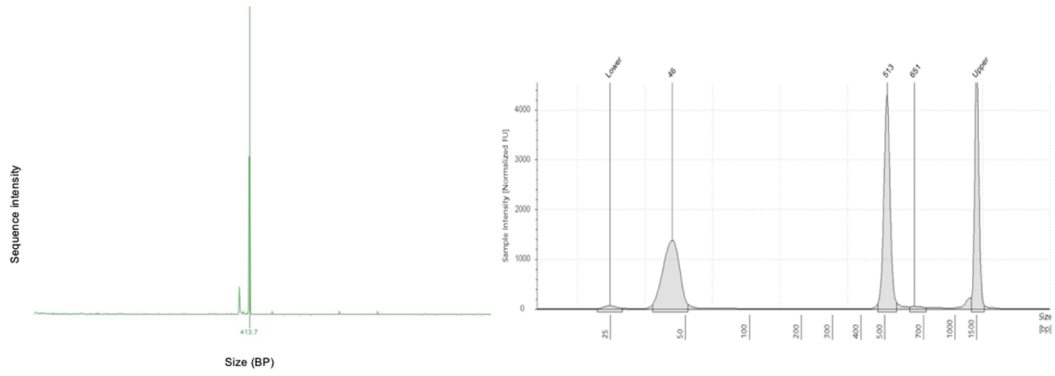


Figure 3. 11. The gel of all primer versions on all cultures. Primer versions: V1 = EPS_F/EPS_R, V2 = EPS_FM/EPS_RL, and V3 = DS_F/DS_R. +v1& v2 = Alvinella DNA positive control. +v3 = DSMZ Desulfurella positive control. -v1, -v2, & -v3 = blank negative controls.

3.4.4 Automated Approach for Ribosomal Intergenic Spacer Analysis (ARISA)

During the isolation workflow, it is important to confirm both the presence of the target organism through the use of specific PCR primers used above and determine the axenicity of the culture. ARISA was used in this study to assess how axenic the cultures were. ARISA was used multiple times during the isolation workflow to check for potential axenic cultures. I first used the DSMZ *Desulfurella acetivorans* isolate as a positive control to determine the appropriate fragment size of this species. Figure 3.12 shows the ARISA fragment analysis for DSMZ *Desulfurella* run on both the Sanger sequencer and the tape station, showing, as expected, a singular peak showing an axenic isolate.

Interestingly, the predicted fragment size differs between the two machines, so rather than focusing on the size of the fragment, I focused on the profile to determine the cleanliness of the culture. Similarly, the fragments' overall peak size is less important than the relative heights of the different peaks. ARISA, throughout the isolation workflow, shows a distinct improvement in the cleanliness of the culture (Figure 3.13). The first six months of the isolation used 3 ml between transfers, and as such, there was contamination in the culture (Figure 3.13a). This transfer volume was subsequently reduced to 100µl to speed up isolation but did not have the desired effect (Figure 3.13b). It was only when I successfully grew the culture on solid media before picking a colony into liquid media that the cleanliness of the culture improved (Figure 3.13c, d). This shows the importance of growth on solid media, as colony picking appears to speed up isolation.



3.12a 3.12b

Figure 3. 12. ARISA fragment analysis of DSMZ Desulfurella acetivorans. 3.12a. on the 3130 sequencer. 3.12b. on the tape station.

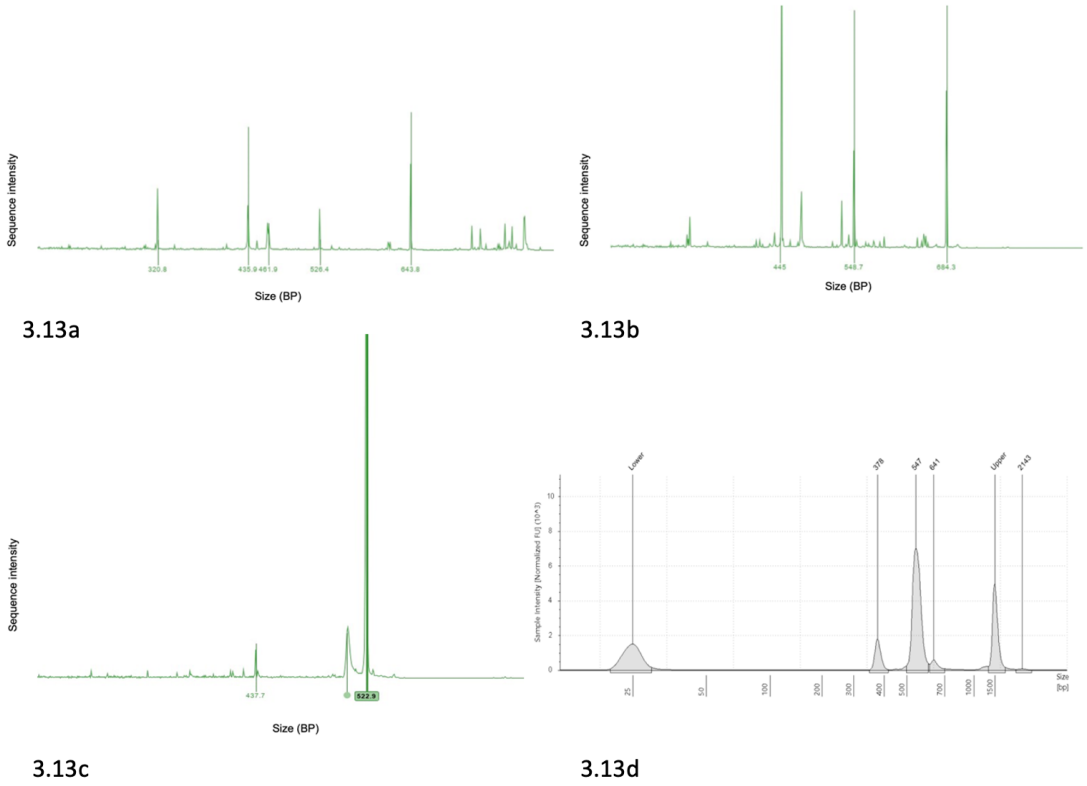


Figure 3. 13. ARISA fragment analysis for site 36. 3.13a. From liquid culture after 6 months of passages using 3mL inoculum. 3.13b. From liquid culture after a further 6 months of transfers and starting to use 100uL inoculum. 3.13c. Liquid culture after transferring from a single colony from solid media. 3.13d. The same as in 3.13c but run on the tape station instead of the 3130 sequencer.

Following the success of colony picking, I subsequently altered the growth conditions to investigate how changes in temperature and pH affect the isolation workflow. Specifically, I altered the temperature and pH to mimic both the conditions in the pool and the published optima for the genera. This entailed changing the media to approximately pH 4 to simulate the environment (pool pH is 3.5), as the standard media recipe had a pH close to the organism's optimum (pH 6). As pH decreases with increasing temperature, I increased the growth temperature of the standard DSMZ 480 media (pH 6) to 50°C to more closely match both the taxa's temperature and pH optima. Interestingly, this change did not produce the expected results, instead resulting in more peaks during ARISA (Figure 3.14). This could be explained by the

change in culture duration, which increased from two to four weeks to increase biomass for DNA extraction and genome sequencing. This increase could have resulted in several non-target taxa increasing in growth from the last ARISA (Figure 3.13).

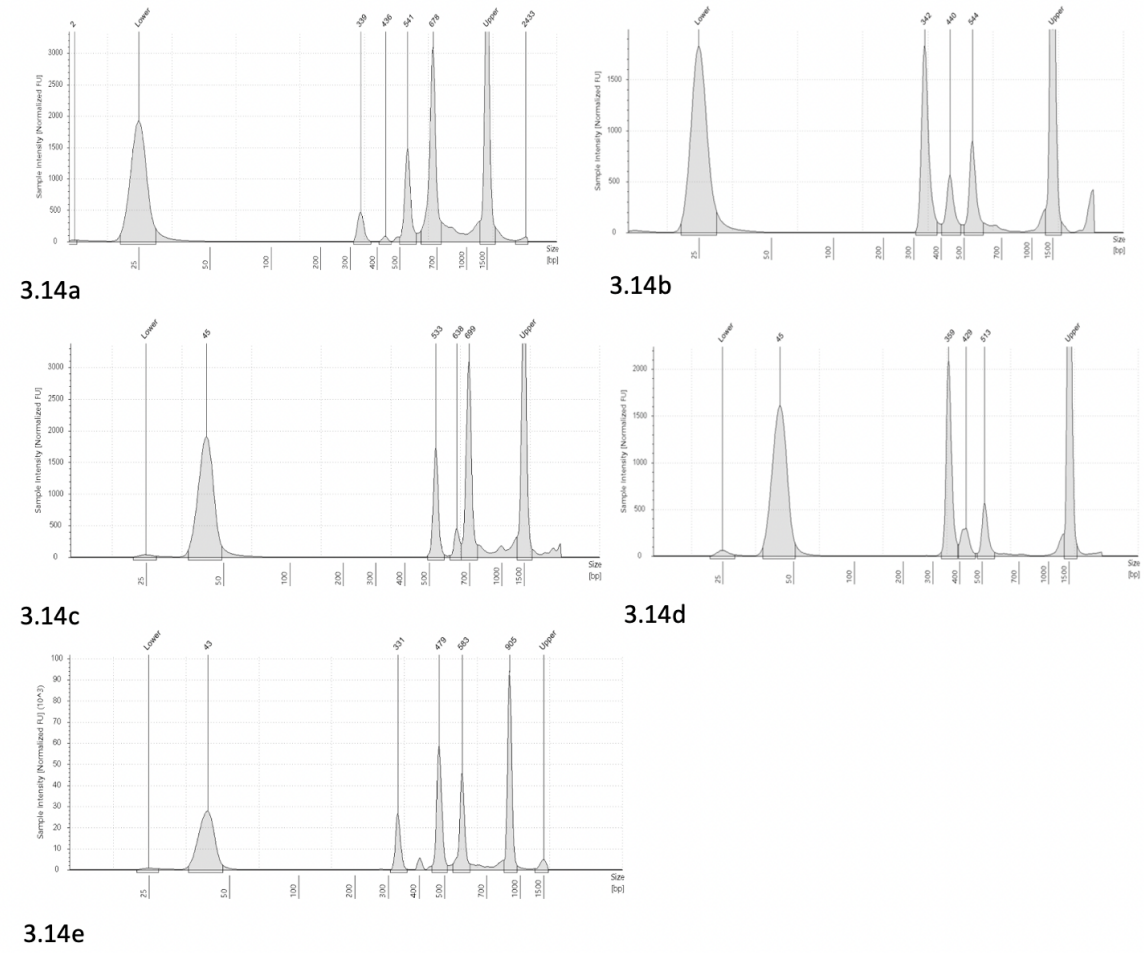


Figure 3. 14. ARISA fragment analysis of the different media alterations. 3.14a. BC pH6. 3.14b. BC pH4. 3.14c. FC pH6. 3.14d. FC pH4. 3.14e. 50°C

3.4.5 16S rRNA Gene Amplicon Analysis

To determine whether *Desulfurella* was still present in the cultures shown in Figure 3.14, the extracted DNA samples were amplified with DS_R/DS_F, which was amplified successfully, showing Desulfurella presence in these cultures. 16S rRNA sequencing of these cultures was carried out with universal primers (515F/926R) and *Desulfurella* specific primers (DS_F/DS_R).

As expected, the sequencing for the universal primers came back messy due to the mixed-species nature of the culture. The sequencing of the DS_F/DS_R amplicon came back reasonably clean (Figure 3.15). Following the recovery of a single ASV from the 16S rRNA community analysis, I expected a singular *Desulfurella* species that would give a clean sequencing run.

Additionally, despite *Desulfurella* species having multiple copies of 16s rRNA genes in their genome, these copies are nearly identical, and so would also result in a clean sequencing run. Despite this, there are regions in the sequencing run that showed low confidence in their base calling, which could be due to errors in the sequencing or the amplification of non-targets, which would indicate that multiple species could be present. The consensus sequence of each culture presented in Figure 3.14 was submitted to NCBI's BLAST webpage query using blastn. All hits with greater than 85% sequence identity were extracted to produce a phylogenetic tree of these cultures (Figure 3.16). The BLAST result and the subsequent phylogeny show that all these cultures contain *Desulfurella* species isolates, with the BLAST result bearing identical percentage identity *to D. acetivorans, D. multipotens*, and *D. kamchatkensis* (range 98.25%-99.83%).

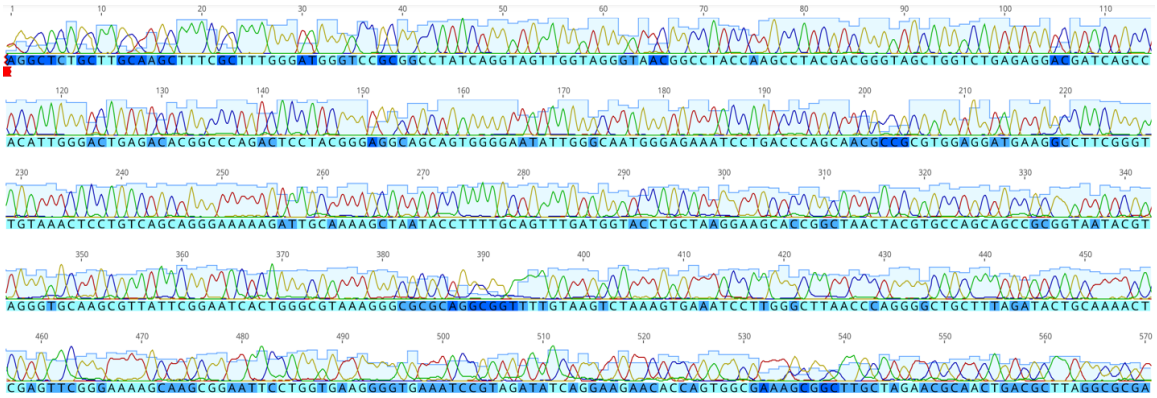


Figure 3. 15. Chromatogram 16S rRNA sequences from the FC pH4 treatment amplified with DS_F/DS_R. The colour of blue indicates confidence in the base call, light blue is very confident – and dark blue is not.

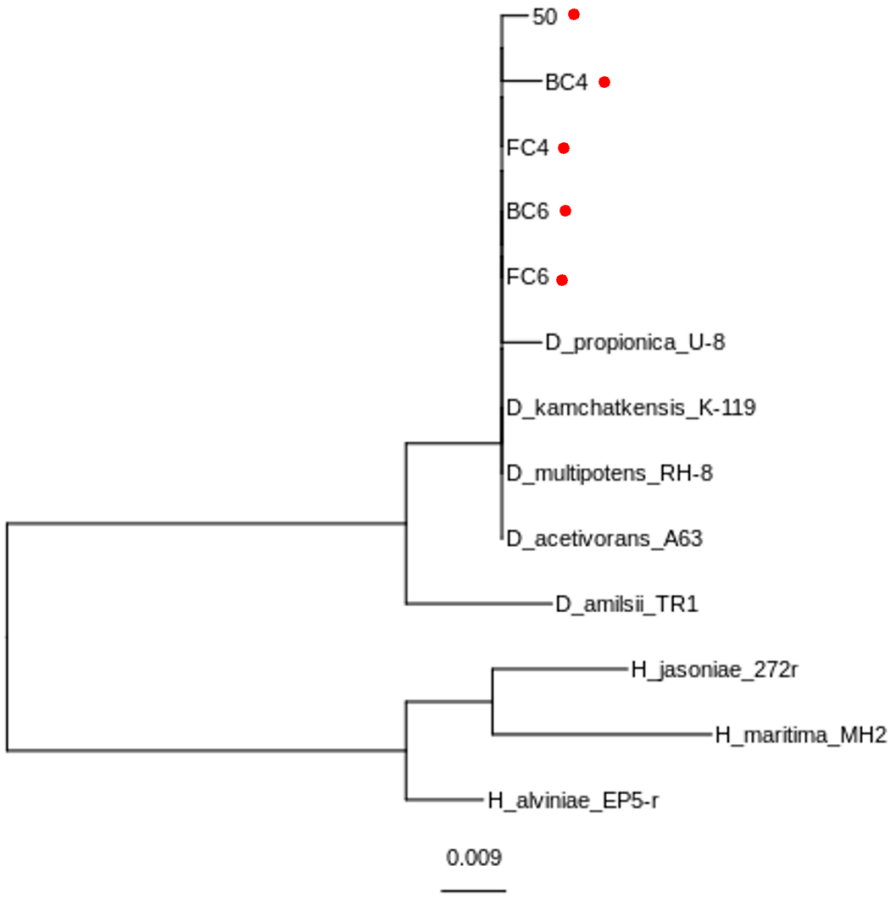


Figure 3. 16. 16S rRNA gene tree from Desulfurella alignments with DS_F/DS_R primers. Red dots indicate isolates from this study.

3.4.6 Metagenome Analysis

Community Profile

I used the shotgun metagenome to further investigate the community composition of site 36 as a comparison to the 16s rRNA amplicon results. Using singleM on the raw metagenomic sequencing reads reaffirmed the recent results of multiple Campylobacterota species present in this pool, with Desulfurella not being the dominant species from this phylum (Figure 3.17). Furthermore, this analysis suggested a singular, novel Desulfurella species in this site, supporting the previous ASV work.

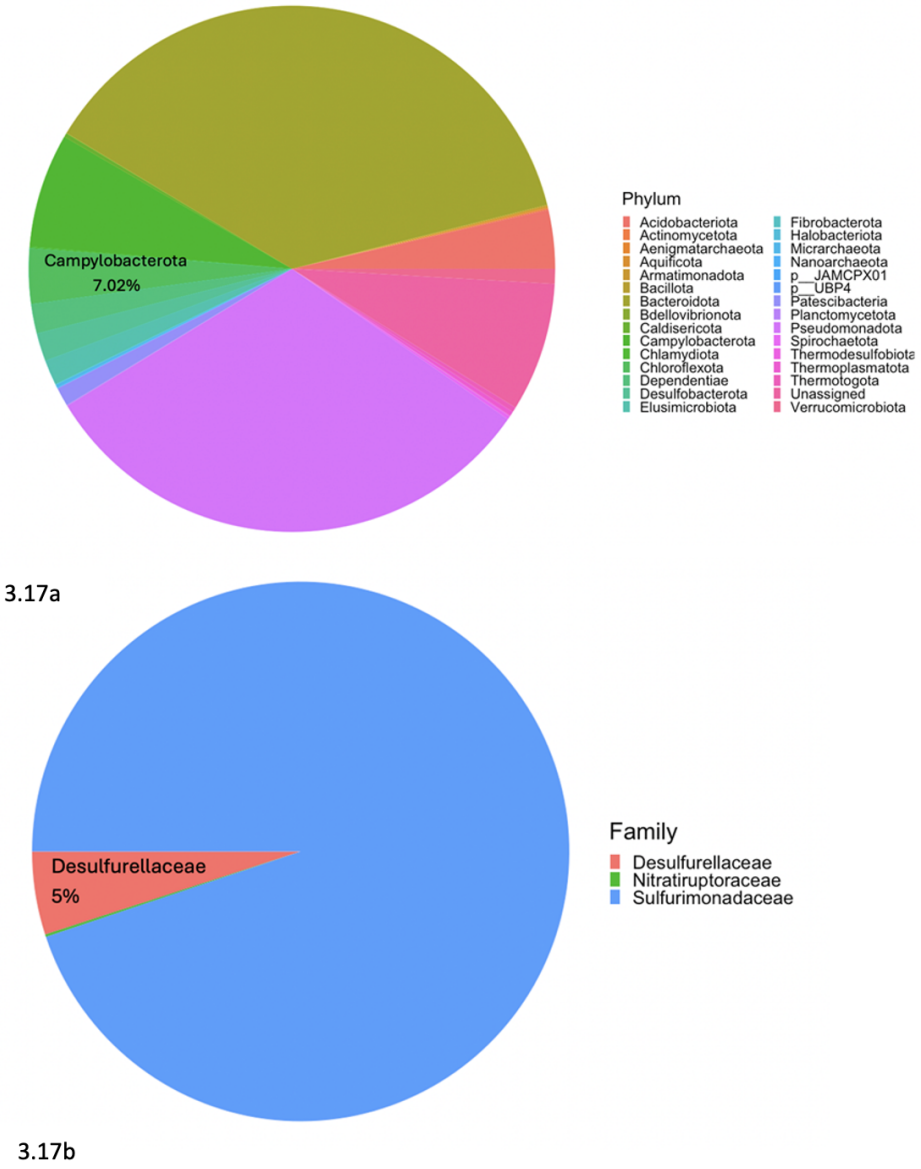


Figure 3. 17. Metagenome community profile for site 36 at Kuirau Park. 3.17a. at the phylum level, Campylobacterota makes up 7.02%. 3.17b. Pie chart of the family level for the Campylobacterota, with Desulfurellaceae making up 5% of the overall family.

Assembly and Binning

Following this community profile analysis, I assembled and binned the metagenome to recover this Desulfurella MAG and all other Campylobacterota species. I recovered three medium quality (>50% completion, <10% contamination) (Bowers et al., 2017) Campylobacterota MAGs from this metagenome. Two of these MAGs are putative novel species within the

CAITKP01 genus within the Sulfurimonadaceae family, with the third MAG being the targeted Desulfurella MAG. The binning statistics of these three MAGs are presented in Table 3.5. To determine the placement of this Desulfurella MAG into the wider Desulfurellia class, the multiple sequence alignment of the Desulfurella MAG and all GTDB Desulfurellia representative species were concatenated and input to FastTree. This phylogeny indicates that the MAG is a putative novel, ancestral species within the *Desulfurella* genus (Figure 3.18). However, this positioning could be a consequence of the incomplete nature of the MAG. Nevertheless, this confirms the previous results from the 16s rRNA analyses of a singular, novel *Desulfurella* species present at this site.

Table 3. 6. MAG statistics for Campylobacterota species recovered from enriched cultures from site 36. Bin 77 being the targeted Desulfurella.

Bin ID	**Completeness (%)**	**Contamination**	**Genome Length**	**GC (%)**
Bin. 19	81.91	3.5	1557872	40
Bin. 50	61.5	2.37	1366875	37
Bin. 77	71.91	2.41	1100160	33

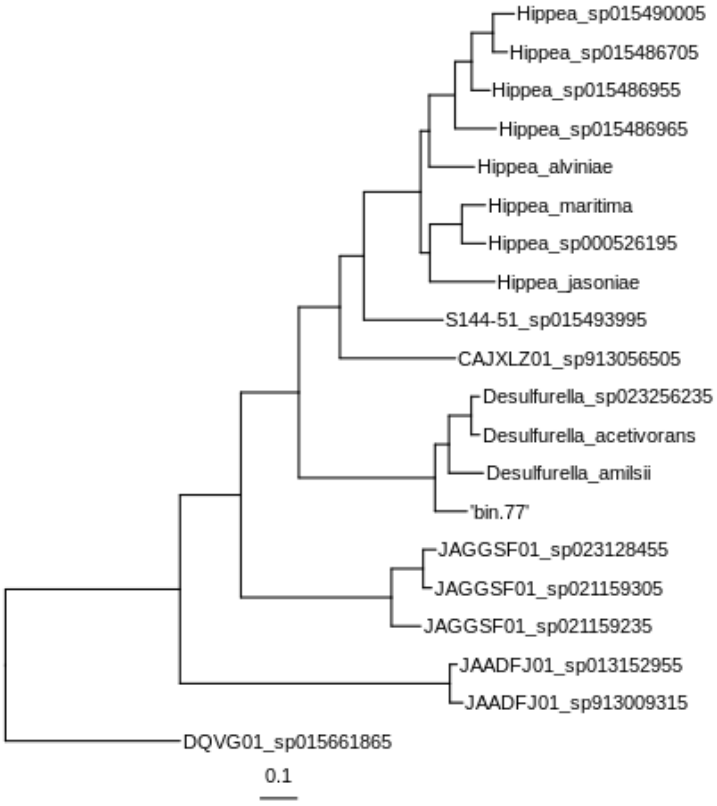


Figure 3. 18. Multiple sequence alignments outputted from the GTDB-tk pipeline subsetted to contain all representative species of the class Desulfurellia, as well as one species from the phylum preceding the Campylobacterota to act as a root. Desulfurella MAG from this study (site 36) is "bin. 77".

3.4.7 Isolation Culture Sequencing

Despite the evidence of a mixed culture in the five different isolation cultures (Figure 3.14), these cultures were sent for sequencing due to time constraints and treated as mini-metagenomes. Unfortunately, culture FC4 did not produce enough DNA for sequencing due to low growth, so only four of the five cultures were sequenced. These sequencing runs were assembled and binned as per the metagenome analysis pipeline, which successfully recovered between three and five MAGs for each of these four cultures, with these MAGs classified in the Bacillota, Bacteroidota, Campylobacterota, Desulfobacterota, and Pseudomonadota. One MAG from the Bacteroidota was present in all cultures, with a MAG from the Desulfobacterota present in three of the four cultures. Surprisingly, none of these MAGs were classified as *Desulfurella* species. This points to the fact that potentially due to the increased growth time, other species were able to

outcompete *Desulfurella* in these cultures successfully. The lack of *Desulfurella* in these cultures was confirmed by analysing the reads with singleM, as a mini-community analysis to ensure *Desulfurella* was not missed due to errors in assembly or binning. Moreover, the fact that the metagenome community analysis conducted on the site microbial community (Figure 3.17) discovered *Desulfurella* at a very low abundance (<1% total community) supports the previous assertion that *Desulfurella* was lost during one of the passages, rather than an error in the bioinformatics analyses.

3.5 Discussion

This chapter sets out to provide an easy-to-follow, straightforward workflow to cultivate targeted environmental organisms by combining standard molecular tools and classical cultivation techniques while making it accessible in any standard microbiological laboratory. The aim was to provide targeted isolation of a sulphur-reducing, obligate anaerobe from an environmental sample. Namely a *Desulfurella sp.* from a geothermal pool in Rotorua, New Zealand. This chapter planned to use published literature and the environmental conditions that it was sampled in to enrich for the target.

This chapter focuses on enriching the target species using molecular tools and traditional cultivation methods. The 1000 Springs Project (Power et al., 2018) informed the sites selected for sampling, with additional sampling and testing conducted to update 16S rRNA libraries and metagenomic community profiles. Interestingly, the results differed from those initially reported in the 1000 Springs Project. Subsequent culturing experiments were conducted by modifying existing media, tailored to the known genomes and the specific environmental conditions, focusing on variations in pH (4 and 6) and temperature (50°C and 30°C). Adjustments were also made to adapt the media for solid growth. An innovative approach involved using a vacuum food-grade bag sealer to cultivate anaerobes, aiming to make the process more accessible and portable, especially in settings without standard laboratory equipment. ARISA was employed

throughout the experiments to evaluate the community complexity in the cultures before they were subjected to whole genome sequencing. Due to time limitations, the cultures were sequenced before isolating a pure strain, resulting in mini-metagenomes that, unfortunately, did not yield a genome for *Desulfurella*. Nonetheless, 16S rRNA sequencing successfully amplified a Desulfurella sequence in each sample, enabling the phylogenetic placement of the species within the gene tree.

3.5.1 Site Selection

The sites for this project were initially selected based on the varying abundances of the class Epsilonproteobacteria, which was the original target. The 1000 Springs Project database (Power et al., 2018) was thoroughly analysed to identify pools with differing complexities and abundances of the target organisms (Table 3.1). I aimed to select sites with a spectrum of microbial community complexities—from those with low levels of Epsilonproteobacteria that are more complex to simpler communities where this group was more abundant. As a result, sites 50, 39, 35, and 22 were chosen based on their microbial compositions. However, sites 39 and 50 were later excluded due to not amplifying with the initial primers (EPS_F/EPS_R). The focus then shifted to specifically targeting the genus *Desulfurella*, a basal lineage that is underrepresented in online databases and for which there are few cultured representatives in culture collections. Site 22 starts in the 1000 Springs Project database with very low levels (2.89%) of *Desulfurella*, but it also contained some other Campylobacterota with *Sulfuircurvum* in high abundance (50.2%). This *Desulfurella* appears to have been lost by the time of sampling for this project, as indicated by the 16S rRNA library (Figure 3.5) and DS_F/DS_R primers (Figure 3.4). As a result, site 36 became the primary target, amplifying with all primer versions from Chapter 1 (Figure 3.4) and 1 ASV in the 16s rRNA library, accounting for 1.09% of the whole community (Figure 3.5). This is much lower than when the 1000 springs project was sampled at this site (Figure 3.1, Table 3.1), where they found *Desulfurella* in a much higher abundance of 32.3%, making it one of the dominant taxa in 2012. With the current abundance of *Desulfurella* being so low, this creates a few hurdles, such as pulling out a MAG from the

metagenome becomes increasingly difficult, and culturing to exclude other, more dominant taxa becomes more challenging.

When looking at the 16S rRNA library comparisons between the 1000 Springs Project and this study, there are noticeable differences in microbial community compositions, with several explanations for this. It is important to note that two different sequencing technologies were used, which could explain some of the differences in community compositions. 1000 Springs used Ion Torrent PGM, and this study used Illumina MiSeq, which was the laboratory standard at the time. Several studies have noted differences in community profiles between these two sequencing platforms when applied to the same community (D'Amore et al., 2016; Fouhy et al., 2016; Salipante et al., 2014). The choice of primers used in the sequencing is also a very significant factor when estimating community composition (Fredriksson et al., 2013). While both this study and the 1000 Springs Project used the same forward primer (515F) we differed in our reverse primer (806R and 914R). While this may not have a significant impact as the first primer is the same, it is a factor to consider, and the same primers should be used when comparing studies.

Alternatively, the community profile differences could also be explained by a legitimate change in composition due to changes in the pools over time. The 1000 Springs Project was conducted in 2014, almost ten years before this study. A change in pH in site 22 was noted at the time of sampling from 5.16 in 2014 to approx. 6 in 2022, site 36 went from 2.9 in 2014 to 3.5 in 2022. While this does not seem like a significant change, the pH scale is logarithmic, so with an increase of one integer value, concentration changes by 10-fold. pH has been proven to be the biggest determinant of microbial community composition, where decreases in pH result in reduced microbial diversity (Power et al., 2018, 2023). As most microbes are mesophilic, this is unsurprising and correlates with the increase in diversity seen here as the pH increases from 2.9 in 2014 to 3.5 in 2022. There were observed changes in the Campylobacterota abundances as well, with only one being reported in 2014 (*Desulfurella)* (Figure 3.1b), which increased to 3 in 2022 (*Sulfuricurvum, Sulfurimonas* and *Desulfurella)* (Figure 3.5b). The increase in pH supports this as while *Desulfurella* can survive at a low pH, *Sulfuricurvum* and *Sulfurimonas* are

mesophilic (Hu et al., 2021; Kodama & Watanabe, 2004), requiring a more neutral pH. The initial sites were chosen based on the 1000 Springs Project (https://1000springs.org.nz/) for different complexities and abundances of Desulfurella. Site 36 had a high amount of Desulfurella present, with it being the most dominant species. Therefore, this community composition has changed based on our 16S rRNA library and metagenome data. With this knowledge, site 36 was possibly not the best target site for Desulfurella as it is in very low abundance, with only one ASV present and accounting for 1.09% of the community. However, the use of published community data was part of this study design, revealing a possible common pitfall that could impact many other studies that use previous community analyses.

The metagenome community abundance profile agrees with the 16S library (Figure 3.17) in that the community composition now contains far less *Desulfurella* than recorded previously. Having this knowledge before selecting the site instead of selecting based on the 1000 Springs data may have given this project a greater chance of success. Choosing a site with a high abundance of the target with less community complexity is ideal. Also, had this been done before commencing culturing, metabolic pathway reconstruction could have helped with altering culturing conditions best suited for the target species, speeding up isolation.

This analysis was done by sequencing the pool's DNA twice, getting 5Gb of data each time, the Desulfurella MAG that was initially obtained from the initial sequencing run had relatively low completeness (~30%) and high contamination (~9%), which is indicative of a poor MAG (Bowers et al., 2017). However, it also had a high genome size (~1.5Mbp after accounting for contamination levels), which could mean that it is more complete than estimated, just more fragmented (N50 < 1500). Currently, MAG quality is assessed on the recovery and copy number of universal and housekeeping genes (Meziti Alexandra et al., 2021) , and if these are missing or broken into fragments, it will be reported as less complete. This could have been the case with the Desulfurella MAG in this study where initially, 5GB of data from the metagenome was acquired and pulling out the *Desulfurella* MAG from this was difficult and fragmented, as mentioned above.

Species that undergo horizontal gene transfer can pose a great challenge regarding sequence assembly and metagenomic binning. Desulfurella species are one of the lineages most impacted by this and could explain the fragmented nature of the binned MAG (Nagies et al., 2020). Additionally, the low abundance of this Desulfurella species in the community (1.09% from 16s rRNA analysis) could explain the difficulties in gaining a good quality MAG from the initial 5Gb of data (Albertsen et al., 2013). It would be unusual to delve into this in detail if a regular metagenome study was conducted, meaning that the Desulfurella MAG would only have been recovered with additional sequencing and analysis. Reconstruction of MAGs from metagenomes is a computationally difficult task; this is an example of what may be missed when interpreting the data. After more data was requested, a good-quality MAG was recovered with 71.91% completeness and 2.41% contamination. The estimated genome size is slightly lower (1.1Mbp), but with only a slightly higher N50 (>1800), it still points to Desulfurella species being difficult to recover in metagenomes, especially when at low abundance.

3.5.2 Primers

In Chapter 2, I explored the potential of using three different primer sets (published, modified, and designed *de novo*) for tracking a target organism through isolation. The published primers initially used for this study were designed in 2012 (Gittel et al., 2012) before the reclassification of *Desulfurella* and Epsilonproteobacteria into the Campylobacterota phylum (D. W. Waite et al., 2017). These primers, therefore, only amplify members of the former Epsilonproteobacteria class and not any *Desulfurella* species (Table 2.2). As this inconsistency was only picked upon after the initial sites were tested, two sites (39 and 50) were discounted from isolation without testing with primers that target *Desulfurella* species. Site 50 did not initially possess *Desulfurella* species from the 1000 springs data (Table 3.1), so the likelihood of them gaining this species is low. Site 39 originally contained a *Desulfurella sp.* (Table 3.1) at 4.46% abundance, so it is possible this pool was erroneously removed. However, the fact that the remaining pools (22 and 36) saw drastic reductions in *Desulfurella* abundance (extinct in site 22, 30-fold reduction in site 36) suggests it is possible that the same occurred in this pool. This is supported by the fact that

these sites exhibited less pronounced pH changes than sites 22 or 36, though in the opposite direction (sites 22 and 36 were less acidic, site 39 more acidic).

Furthermore, as there was no amplification from these sites using EPS_F/EPS_R, all Campylobacteria members were also missing from these sites. As a result of these primers specifically targeting the rest of the families in the Campylobacteria class but excluding members of the Desulfurellia class, I decided this would be an interesting negative control for our sites. In other words, when used in conjunction with a set of highly specific primers, I could assess the cleanliness of the culture. If both the published and specific primers amplified then that would indicate a mixed culture with several Campylobacterota members, including *Desulfurella*. Alternatively, if the published primers did not amplify, but the specific primers did, then that would mean the culture was clean from other Campylobacterota members.

3.5.3 Culturing

Passages for liquid media were performed every two weeks, and solid media was performed every four weeks. Growth was very slow; it took four weeks to see colonies on solid agar plates that were less than 0.5mm. Both visually and microscopically, pH 4 and 50-degree treatments had fewer cells than pH 6, which was to be expected as it was performed to decrease non-target cells. This also translated into a very low DNA extraction yield, resulting in one of the pH 4 treatments not having enough DNA for genome sequencing. A paper by Florentino et al. found that growth peaked at 35 days for *Desulfurella* in both the pH 6.5 and pH 3.5 treatments, with a good amount of growth at two weeks (A. P. Florentino et al., 2017). However, larger volumes may be required when it comes to the final extraction for sequencing. For the last round of solid media, the plates were left in the incubator for six weeks to grow more biomass for extracting DNA. This may have had an effect of also allowing even slower-growing microbes (i.e. non-target taxa) to grow; as seen in the ARISA, the cultures look to get messier, not cleaner, on the final extraction.

Initially, 3mL of culture was transferred to fresh liquid media every two weeks which was changed to 100uL. This was changed as Microbial EPS debris released from the cell after rupturing can play a significant role as an organic carbon source (Davis et al., 2005). This positively impacted isolation (Figures 3.13a and 3.13b), and decreasing this further could also benefit as the less organic carbon available, the fewer non-target microbes have to survive on.

3.5.4 Solid Media

One of the goals of this project was to domesticate *Desulfurella* on solid media as colony picking can aid in gaining an axenic isolate. It is also useful in research, for example for performing genetic mutation experiments. Therefore, a substitute for elemental sulphur was required as it is not soluble and would not be bioavailable in solid media. Sodium polysulphide was used to replace elemental sulphur. However, this increased the pH of the media to an undesirably alkaline pH. The buffer in the original media was sodium bicarbonate, but it did not have a buffering capacity high enough to keep the media at the desired pH.

For this reason, MOPs and phosphate buffers were assessed as substitutes. Figure 3.6 shows that phosphate buffer with polysulphide (without elemental sulphur) seems to have little to no amplification, whilst MOPs under the same conditions continue to have growth. Cultures were also checked microscopically under phase contrast with MOPS buffer, showing a similar number of cells as the control and phosphate having significantly less. MOPs was therefore chosen to add to the media to keep the pH in the correct range. Like sodium bicarbonate, phosphate buffer has a weaker buffering capacity than MOPS, so it may not have been strong enough to counteract the strong alkaline nature of the polysulphide. The trial went on alongside regular passages every two weeks for six months before deciding it was successful, and MOPs were included in the media.

When modifying the media, measuring the impact on cell growth is crucial. As mentioned earlier, PCR and microscopy were employed, each with distinct advantages and drawbacks. PCR, while helpful, cannot differentiate between live and dead cells, leading to potential false positives or

negatives (Nocker & Camper, 2006). Microscopy, although requiring considerable expertise to distinguish between species, can determine cell viability through motility observation. Optical density measurement could be another method for assessing culture growth; however, the turbidity of this media likely precludes its effectiveness. Resazurin staining is commonly used to evaluate cell viability, shifting from blue to pink when cells are alive. However, in this study, its function as an anaerobic indicator means this is also not viable for anaerobic microbes. Therefore, 16S rRNA gene PCR remained the standard method for determining the growth of the target taxa despite the known drawbacks.

When attempting to culture an organism on solid media for the first time, transitioning directly from liquid to 100% agar could be too drastic. A more gradual, gradient-based approach is essential to determine the optimal agar viscosity for the target species. As detailed in the methods, a stab culture technique was used, with growth monitored down to 80% agar (Figure 3.7). Consequently, all subsequent plates were prepared with 80% granulated agar. Growth on solid media was notably slow, taking about four weeks to produce colonies smaller than 0.5 mm. Various colony morphologies were observed, including small frosted white colonies, larger colonies, and mucoid-like larger colonies. Figure 3.13c shows the ARISA for a colony picked, indicating a possible axenic culture. However, these colonies were then transferred back to liquid media for further cultivation and either outcompeted or contaminated.

3.5.5 Media adjustments

To speed up the isolation process, some different culturing conditions were trialled out. The pH of the media for one condition was lowered to pH 4 as this is closer to the environmental conditions when sampling the site (pH 3.5 - 4). The 1000 Springs Project also found pH to be the biggest driver in determining microbial community composition (Power et al., 2018). Additionally, while DSMZ culturing conditions advise a pH of 6, the optimum pH range for Desulfurella is said to be 3-7 (A. Florentino et al., 2015). This could mean that lowering the pH may be able to eliminate any non-target bacteria that cannot withstand the lower pH. The second condition was to grow the cultures at 50°C instead of 30°C, which was the temperature of the

pools at sampling. 50°C is near the optimum growth temperature for *D. acetivorans* and *D. multipotens* (Bonch-Osmolovskaya et al., 1990; A. P. Florentino et al., 2017; Rainey & Hollen, 2015). These were passed alongside the original cultures for six months. In hindsight, this work should have been done right at the beginning, or a fresh inoculum should have been sampled.

These cultures were then confirmed to grow *Desulfurella* with DS_F/DS_R primers, so they were grown on solid media for six weeks to grow enough biomass for DNA extraction. Figure 3.11 shows all the primer versions from Chapter 2 on each media adjustment. BC6, BC4, and the 50-degree treatment returned positive for EPS_F/EPS_R and DS_F/DS_R primers. This could mean that all these cultures are co-cultures of *Desulfurella* with one or more Campylobacterota, with the possibility of them living symbiotically and not being ruled out. It could also mean that the original primers (EPS_F/EPS_R) are still amplifying *Desulfurella,* as previously discussed, and the mismatches in the primer sequence to the target are not in detrimental binding locations. FC6 and FC4 are interesting as they only amplify for DS_F/DS_R primers, indicating no other Campylobacterota are present. It is also interesting to note that the EPS_FM/EPS_RL primers did not amplify at all. These are supposed to be the broadest primers targeting Campylobacterota and other bacterial classes. These cultures were all sent for genome sequencing except for FC4, which had a very low DNA extraction yield. This was done prematurely due to time constraints in order to look at the genomes and assess if cultivation strategies were working.

3.5.6 ARISA

After running an in-silico PCR with the ARISA primers, it is shown that *Desulfurella* has the smallest amplicon out of the Campylobacterota order (607-610BP); anything below this may be contamination from other species. Anything bigger may be a co-culture, possibly with other Campylobacterota. Interestingly, when running the DSMZ *Desulfurella acetivorans* with the ARISA primers on the Applied Biosystems 3730 Genetic Analyzer, the peak came out at a much lower size than expected (607-610BP) of around 413 BP (Figure 3.12a) This was also run on the

Agilent 4200 TapeStation System and was 513 BP (Figure 3.12b). Due to this discrepancy, the ARISA and tape station results were only used as a proxy to indicate whether a culture was isolated. Figure 3.13 shows the progression of the community composition throughout different passages. Transferring 100uL of inoculum instead of 3mL has a positive effect (Figure 3.13b), with colony picking even more successful (Figures 3.13c and 3.13d). Figure 3.14 then shows the fragment analysis of the different media changes. These look to have more peaks than the original, which could indicate contamination or be due to other factors, such as the longer incubation time used to create more biomass for extraction. This could have allowed more slow-growing bacteria that were already present time to grow. Both pH 4 treatments follow a similar trend, losing some larger peaks and increasing in the smaller ones. Papers such as Fisher & Triplett (1999) described anything below 390BP as artifacts (Fisher & Triplett, 1999) that also potentially applies here. While ARISA could be a valuable and fast way to see how many species are present, it could also provide false confidence.

3.5.7 16S rRNA Sequencing

After visually inspecting the sequence chromatograms for the 16S rRNA gene sequencing using universal primers 515F/926R, all the sequences look to be mixed communities with two or more species present. This is not surprising based on the results from the ARISA. Ideally, the best practice is to wait until only one peak is seen on the ARISA before sequencing. However, under time constraints for this project, I had to do this prematurely. Due to this, 16S rRNA sequencing was also conducted with the *Desulfurella* specific primers (DS_F/DS_R) to assess if the target was present and where it would be placed in the phylogenetic tree. These sequences came back relatively clean (Figure 3.15), and the 16S rRNA gene tree was created in Figure 3.16. All the samples in this study are likely the same species of *Desulfurella,* which corroborates with only one ASV from the initial 16S rRNA library.

3.5.8 Isolation Culture Shotgun Sequencing

The shotgun genome sequencing results for all cultures did not detect *Desulfurella,* which is confounding. The species-specific primers (DS_F/DS_R) consistently amplified throughout the isolation protocol, and when these amplicons were sequenced, they matched a *Desulfurella* when blasted against the NCBI 16S rRNA database. There could be several explanations for this. First, *Desulfurella* possibly did not lyse as effectively as other taxa but is still present. However, the same DNA extraction method was used for the environmental sample 16S rRNA library and metagenome where *Desulfurella* was successfully detected (Figures 3.16, 3.18), so this is unlikely. Second, since PCR amplifies DNA from both viable and non-viable cells, the amplification could target DNA from dead *Desulfurella* cells as the target may have been outcompeted and died off. Finally, *Desulfurella* could be present but in extremely low abundance. However, singleM (Woodcroft et al., 2024) analysis is highly sensitive, so this is likely improbable as it would have detected even minimal *Desulfurella* reads. Also, a *Desulfurella* MAG (Table 3.5 and Figure 3.18) was recovered from the environmental sample at very low abundance, and the same method was used here, suggesting that the method was valid. Therefore, the most plausible explanation is that *Desulfurella* was present initially but was outcompeted over time by other taxa with the PCR-detected DNA from dead cells.

3.6 Conclusion

This chapter discusses the challenges involved with isolating fastidious environmental microorganisms to specifically target sulphur-reducing obligate anaerobe Desulfurella sp. from a geothermal pool in Rotorua, New Zealand. Traditional cultivation methods have often failed to capture the diversity of this microbial dark matter, with molecular tools shedding light on this discrepancy. The work presented here combines classical cultivation techniques with molecular tools to enhance the isolation process. Although the end sequencing results did not confirm the isolation of the target *Desulfurella,* the use of species-specific primers and media modifications

provided insights and knowledge into the necessary steps for future cultivation efforts. This study shows the importance of adapting cultivation strategies and utilising molecular tools to explore microbial dark matter further.

Chapter 4: Discussion

As discussed in previous chapters, cultivating microbial dark matter is challenging, but isolates provide many benefits and can expand our knowledge of the microbial world. The current literature provides some insights into this problem, but there are yet to be any straightforward workflows to combat this issue. This thesis aimed to develop a simple and accessible workflow for cultivating specific environmental organisms by integrating widely used molecular tools with traditional cultivation methods. This approach was designed to be easily implemented in any standard microbiological laboratory. A test case was to target a sulphur-reducing, obligate anaerobe from an environmental sample. Namely a *Desulfurella sp.* from a geothermal pool in Rotorua, New Zealand. This was planned in two parts. The first was to develop and use molecular tools to ensure the target is present in the inoculum and cultures. Second, to enrich the target, culture media was adjusted based on knowledge from published literature and the environmental conditions it was sampled in.

I designed and assessed primers based on the target taxa at different stringency levels in the first experimental chapter. I first assessed primers from the literature (EPS_F/ESP_R (Gittel et al., 2012)) that targeted Epsilonproteobacteria, which was the initial target. However, Epsilonproteobacteria was reclassified into the novel phylum Campylobacterota, combining the Epsilonproteobacteria class with the order *Desulfurellales* (previously in the Deltaproteobacteria class). *Desulfurella sp. s*ubsequently became the specific target for isolation. After this change, I found that, unsurprisingly, EPS_F/EPS_R did not target *Desulfurella* s*p.* at all, as these sequences were not included in their initial design. Therefore, I designed new primers by modifying these published primers (EPS_FM/EPS_RL). I aimed to create a primer pair to produce a longer amplicon for sequencing and target the entire Campylobacterota phylum more broadly. However, after *in silico* testing of the specificity of this primer pair, it was apparent that they also amplified a lot of non-target species (i.e. members of the Desulfobacterota phylum), making them unsuitable for this work. Finally, I designed a third set of primers *de novo* to target

Desulfurella sp. with higher specificity (DS_F/DS_R). The aim for this primer pair was similar to EPS_FM/EPS_RL (e.g. longer amplicon for sequencing) but to be highly specific, which *in silico* testing confirmed. These primers can all be used together to assess the site and also the isolates before sequencing: DS_F/DS_R tells us if *Desulfurella sp.* is present, and EPS_F/EPS_R is used to identify if it is another Campylobacterota (negative for *Desulfurella).*

The second experimental chapter attempts to enrich this target using the primers described above and classical cultivation strategies. Site selection was based on the 1000 Springs Project (Power et al., 2018), but I retested the microbial community composition with current 16S rRNA gene sequencing and metagenome community abundance profiles. This showed pronounced differences from the original study, which is to be expected with the time that had passed since that original study was conducted. I subsequently began culturing these samples by modifying published media based on published literature and the sampling environment, specifically investigating changes in pH (4 and 6) and temperature (50°C and 30°C). Alterations were also made to the media to allow growth on solid media by replacing elemental sulphur with polysulphide. A vacuum food-grade bag sealer was used as a new way of cultivating anaerobes to make it more accessible to a range of laboratories and to make it portable if the cultivation was to take place without a standard laboratory. For example, in the laboratory where this research occurred, having a working anaerobic chamber was very expensive to keep operational over two years. I used Automated Ribosomal Intergenic Spacer Analysis (ARISA) throughout to assess the axenicity of the cultures before sending them for both 16s rRNA and whole genome sequencing. 16S rRNA gene sequencing amplified one *Desulfurella* in each culture from the different conditions with the species-specific primers (DS_F/DS_R), which allowed for the phylogenetic placement of the species in the gene tree, supporting the genome recovered from the metagenome assembly. Due to time constraints, these cultures were sent for sequencing before a pure isolate was obtained and treated as mini-metagenomes. Intriguingly, this failed to produce a MAG for *Desulfurella* despite being the same culture that recovered *Desulfurella* from

the 16s rRNA sequencing. In this chapter, I will discuss what I learnt from undertaking this study and suggest improvements if this were to be carried out again or utilised for a different organism.

In the future, Figure 4.1 shows what I would recommend when isolating a species of interest again. Firstly, I would suggest that 16S rRNA library analysis is carried out before selecting a site and choosing one with the best possible chance of success (i.e. where the target is the dominant species and the community is less complex, instead of relying on a database or previous studies results). Using a database could guide the initial site selection to remove sampling and sequencing pools known to be not of interest, but pools of interest should be re-evaluated. Altering the media based on the metagenome and literature should also be done immediately, and there should be more variations and safety guards to ensure the target is not lost, for example, Quantitative PCR (qPCR) at each transfer or 16S rRNA gene libraries. Also, inoculating in the field would decrease damage to the inoculum and create a better chance for success.

An improvement I would make to my isolation protocol is to use a method to quantify the amount of the target, not just if it is present or absent, such as using qPCR. qPCR with taxa specific versus a broad bacterial primer set would determine if the target taxa increased or decreased in abundance over the isolation protocol. This has been used by Jian et al. 2020 alongside next-generation sequencing to accurately estimate taxon abundances, which could be applied here (Jian et al., 2020) A decreasing abundance would indicate that other taxa are outcompeting the desired target or perhaps even died off. Alternatively, DNA could be extracted every few passages to conduct a 16S rRNA library to determine the community composition of the isolation culture. Similarly to the qPCR protocol, this would see if the desired taxa are increasing in abundance and close to isolation, though it is a more costly and time-consuming option.

Using more primer sets during the isolation protocol would be beneficial (i.e. having 23S and ribosomal protein primers, rather than only targeting 16S) to reduce the chance of amplifying not target taxa. As mentioned earlier in the thesis, designing specific PCR primer sets relies upon the database chosen, which is problematic for taxa of which have few representatives in that

database. This could lead to an overestimation of that PCR primer set's specificity. Furthermore, overreliance on rRNA genes for primer design could limit the sample size due to the low likelihood of recovering these genes from metagenomes. Using other housekeeping genes, as they are more vertically transferred (Nagies et al., 2020), could expand the sample size for primer design, giving greater confidence in their specificity. I also learnt that colony picking to liquid media, leaving the incubation time constant, and then extracting DNA is the best option. Doing this with higher volumes of liquid rather than putting this back onto solid media to grow more biomass creates less room for contamination and for slower-growing microbes to out-compete the target. I could have also utilised methods such as dilution to extinction to gain pure cultures.

4.1 Modified Isolation Workflow

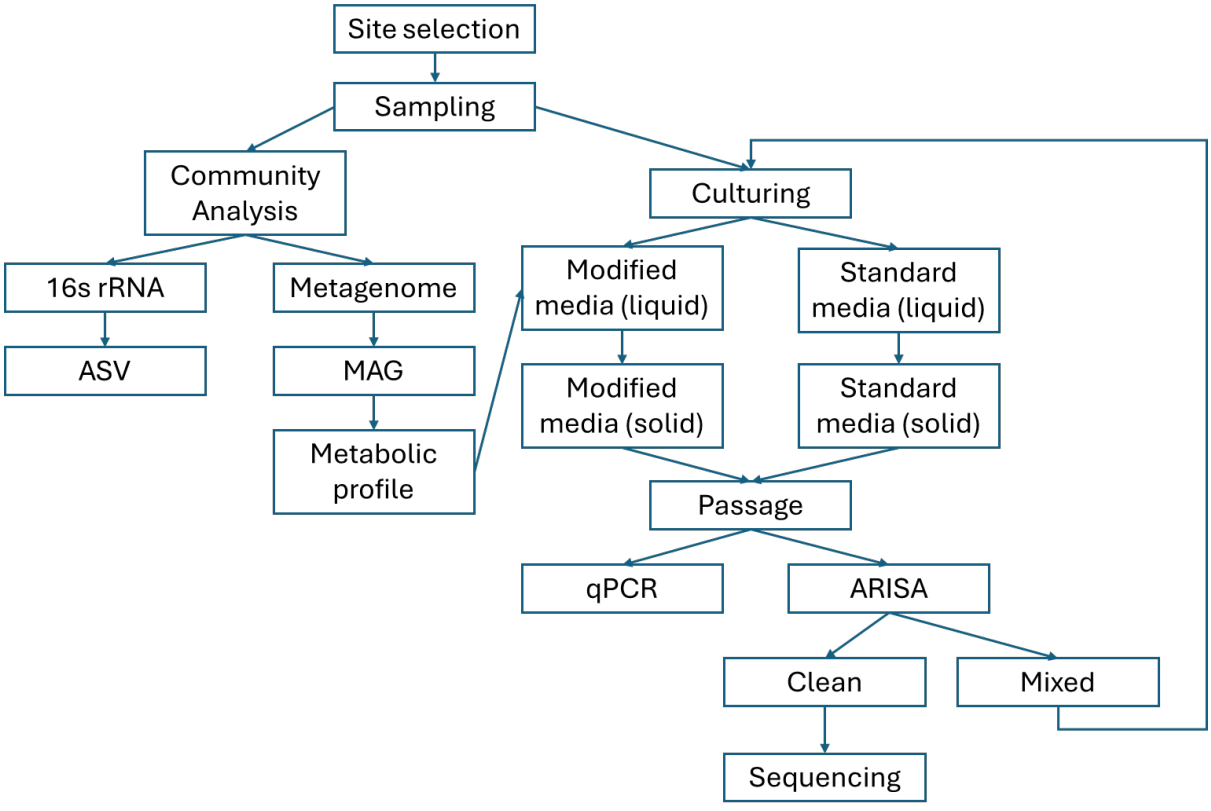


Figure 4. 1. Modified isolation workflow with the main differences being community analysis informing cultivation while keeping standard media to run in parallel. The addition of environmental conditions could also be run alongside these if they do not fall into the same categories. The addition of qPCR when passaging is a key addition to ensure the target is growing and it is not just ancient DNA being amplified. Solid media would still need to be optimised for agar viscosity and sulphur bioavailability as mentioned in the text.

4.2 Final Thoughts

This thesis faced several challenges, primarily due to the limited initial study design and the lack of early access to metagenomic data, which hindered the cultivation process and affected the project's overall direction. This initial setback, in my view, compromised the foundation of the research as the media was not able to be edited originally as planned. Consequently, adjustments to the culture media were made later in the process despite recognising that these modifications should have been implemented earlier. Unfortunately, time constraints became a significant factor.

This work demonstrates the complexities involved with targeting and isolating uncultivated microorganisms. Despite difficulties encountered, there were also some successes, such as specific primer design, domesticating onto solid media, enabling better accessibility to cultivate anaerobes and the use of ARISA to assess isolation success. This work also highlights the importance of combining classical cultivation techniques with molecular approaches to explore the world's microbial diversity and provides an optimised workflow from the learnings gained throughout. While the target organism appears to not be alive in the final cultures, this study contributes insights into the isolation process. It aims to aid future research in overcoming barriers to cultivating "Microbial Dark Matter."

This being said I have dedicated substantial effort to this project. While the final result may have been less desirable, the experience has been extremely valuable, particularly in understanding the research process and the critical importance of a well-thought-out study design

References

Aedo, S. J., Ma, H. R., & Brynildsen, M. P. (2019). Checks and Balances with Use of the Keio Collection for Phenotype Testing. In C. N. S. Santos & P. K. Ajikumar (Eds.), *Microbial Metabolic Engineering: Methods and Protocols* (pp. 125–138). Springer New York. https://doi.org/10.1007/978-1-4939-9142-6_9

Agrawal, R. (2005). Probiotics: An emerging food supplement with health benefits. *Food Biotechnology*, *19*(3), 227–246. https://doi.org/10.1080/08905430500316474

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., & Nielsen, P. H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology*, *31*(6), 533–538. https://doi.org/10.1038/nbt.2579

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F., & Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, *11*(11), 1144–1146. https://doi.org/10.1038/nmeth.3103

Alneberg, J., Karlsson, C. M. G., Divne, A.-M., Bergin, C., Homa, F., Lindh, M. V., Hugerth, L. W., Ettema, T. J. G., Bertilsson, S., Andersson, A. F., & Pinhassi, J. (2018). Genomes from uncultivated prokaryotes: A comparison of metagenome-assembled and single-amplified genomes. *Microbiome*, *6*(1), 173. https://doi.org/10.1186/s40168-018-0550-0

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. *Molecular Systems Biology*, *2*, 2006.0008. https://doi.org/10.1038/msb4100050

Basu, S., Campbell, H. M., Dittel, B. N., & Ray, A. (Directors). (2010, July 10). *Purification of specific cell population by fluorescence activated cell sorting (FACS).* (Issue 41) [Video recording]. https://doi.org/10.3791/1546

Bhattacharjee, A. S., Wu, S., Lawson, C. E., Jetten, M. S. M., Kapoor, V., Domingo, J. W. S., McMahon, K. D., Noguera, D. R., & Goel, R. (2017). Whole-Community Metagenomics in Two Different Anammox Configurations: Process Performance and Community Structure. *Environmental Science & Technology*, *51*(8), 4317–4327. https://doi.org/10.1021/acs.est.6b05855

Bodor, A., Bounedjoum, N., Vincze, G. E., Erdeiné Kis, Á., Laczi, K., Bende, G., Szilágyi, Á., Kovács, T., Perei, K., & Rákhely, G. (2020). Challenges of unculturable bacteria: Environmental perspectives. *Reviews in Environmental Science and Bio/Technology*, *19*(1), 1–22. https://doi.org/10.1007/s11157-020-09522-4

Bonch-Osmolovskaya, E. A., Sokolova, T. G., Kostrikina, N. A., & Zavarzin, G. A. (1990). Desulfurella acetivorans gen. Nov. And sp. Nov. —A new thermophilic sulfur-reducing eubacterium. *Archives of Microbiology*, *153*(2), 151–155. https://doi.org/10.1007/BF00247813

Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., Schulz, F., Jarett, J., Rivers, A. R., Eloe-Fadrosh, E. A., Tringe, S. G., Ivanova, N. N., Copeland, A., Clum, A., Becraft, E. D., Malmstrom, R. R., Birren, B., Podar, M., Bork, P., … The Genome Standards Consortium. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, *35*(8), 725–731. https://doi.org/10.1038/nbt.3893

Bushnell, B. (2014, March 17). BBMap: A Fast, Accurate, Splice-Aware Aligner. *Report Number: LBNL-7065E*. https://www.osti.gov/biblio/1241166

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Campbell Barbara J. & Cary S. Craig. (2001). Characterization of a Novel Spirochete Associated with the Hydrothermal Vent Polychaete Annelid, Alvinella pompejana. *Applied and Environmental Microbiology*, *67*(1), 110–117. https://doi.org/10.1128/AEM.67.1.110-117.2001

Cárdenas-Hernández, V., Velasco-Cuervo, S., & Toro-Perea, N. (2023). Whole metagenome sequencing reveals host plant influences microbial community associated with larvae of Anastrepha obliqua. *Entomologia Experimentalis et Applicata*, *171*(9), 668–680. https://doi.org/10.1111/eea.13333

Carini Paul. (2019). A "Cultural" Renaissance: Genomics Breathes New Life into an Old Craft. *mSystems*, *4*(3), 10.1128/msystems.00092-19. https://doi.org/10.1128/msystems.00092-19

Cary, S. C., McDonald, I. R., Barrett, J. E., & Cowan, D. A. (2010). On the rocks: The microbiology of Antarctic Dry Valley soils. *Nature Reviews Microbiology*, *8*(2), 129–138. https://doi.org/10.1038/nrmicro2281

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2022). GTDB-Tk v2: Memory friendly classification with the genome taxonomy database. *Bioinformatics (Oxford, England)*, *38*(23), 5315–5316. https://doi.org/10.1093/bioinformatics/btac672

Chen, M., Xu, P., Zeng, G., Yang, C., Huang, D., & Zhang, J. (2015). Bioremediation of soils contaminated with polycyclic aromatic hydrocarbons, petroleum, pesticides, chlorophenols and heavy metals by composting: Applications, microbes and future research needs. *Biotechnology Advances*, *33*(6 Pt 1), 745–755. https://doi.org/10.1016/j.biotechadv.2015.05.003

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chklovski, A., Parks, D. H., Woodcroft, B. J., & Tyson, G. W. (2023). CheckM2: A rapid, scalable and accurate tool for assessing microbial genome quality using machine

learning. *Nature Methods*, *20*(8), 1203–1212. https://doi.org/10.1038/s41592-023-01940-w

Clarridge, J. E. 3rd. (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clinical Microbiology Reviews*, *17*(4), 840–862, table of contents. https://doi.org/10.1128/CMR.17.4.840-862.2004

D'Amore, R., Ijaz, U. Z., Schirmer, M., Kenny, J. G., Gregory, R., Darby, A. C., Shakya, M., Podar, M., Quince, C., & Hall, N. (2016). A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genomics*, *17*(1), 55. https://doi.org/10.1186/s12864-015-2194-9

Darwin, C., & Kebler, L. (1859). *On the origin of species by means of natural selection*. J. Murray.

Davis, K. E. R., Joseph, S. J., & Janssen, P. H. (2005). Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria. *Applied and Environmental Microbiology*, *71*(2), 826–834. https://doi.org/10.1128/AEM.71.2.826-834.2005

de Souza Vandenberghe, L. P., Garcia, L. M. B., Rodrigues, C., Camara, M. C., de Melo Pereira, G. V., de Oliveira, J., & Soccol, C. R. (2017). Potential applications of plant probiotic microorganisms in agriculture and forestry. *AIMS Microbiology*, *3*(3), 629–648. https://doi.org/10.3934/microbiol.2017.3.629

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Farahinia, A., Zhang, W., & Badea, I. (2023). Recent Developments in Inertial and Centrifugal Microfluidic Systems along with the Involved Forces for Cancer Cell Separation: A Review. *Sensors*, *23*(11). https://doi.org/10.3390/s23115300

Fijan, S. (2014). Microorganisms with claimed probiotic properties: An overview of recent literature. *International Journal of Environmental Research and Public Health*, *11*(5), 4745–4767. https://doi.org/10.3390/ijerph110504745

Fishbein, S. R. S., Mahmud, B., & Dantas, G. (2023). Antibiotic perturbations to the gut microbiome. *Nature Reviews Microbiology*, *21*(12), 772–788. https://doi.org/10.1038/s41579-023-00933-y

Fisher, M. M., & Triplett, E. W. (1999). Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Applied and Environmental Microbiology*, *65*(10), 4630–4636. https://doi.org/10.1128/AEM.65.10.4630-4636.1999

Florentino, A., Brienza, C., Stams, A., & Sánchez-Andrea, I. (2015). Desulfurella amilsii sp. Nov., a novel acidotolerant sulfur-respiring bacterium isolated from acidic river sediments. *International Journal of Systematic and Evolutionary Microbiology*, *66*. https://doi.org/10.1099/ijsem.0.000866

Florentino, A. P., Stams, A. J. M., & Sánchez-Andrea, I. (2017). Genome Sequence of Desulfurella amilsii Strain TR1 and Comparative Genomics of Desulfurellaceae Family. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.00222

Florentino, A. P., Weijma, J., Stams, A. J. M., & Sánchez-Andrea, I. (2015). Sulfur Reduction in Acid Rock Drainage Environments. *Environmental Science & Technology*, *49*(19), 11746–11755. https://doi.org/10.1021/acs.est.5b03346

Fouhy, F., Clooney, A. G., Stanton, C., Claesson, M. J., & Cotter, P. D. (2016). 16S rRNA gene sequencing of mock microbial populations- impact of DNA extraction method, primer choice and sequencing platform. *BMC Microbiology*, *16*(1), 123. https://doi.org/10.1186/s12866-016-0738-z

Fredriksson, N. J., Hermansson, M., & Wilén, B.-M. (2013). The choice of PCR primers has great impact on assessments of bacterial community diversity and dynamics in a wastewater treatment plant. *PloS One*, *8*(10), e76431. https://doi.org/10.1371/journal.pone.0076431

Gilbert, J. A., Jansson, J. K., & Knight, R. (2014). The Earth Microbiome project: Successes and aspirations. *BMC Biology*, *12*(1), 69. https://doi.org/10.1186/s12915-014-0069-1

Gilbert, J. A., Meyer, F., Antonopoulos, D., Balaji, P., Brown, C. T., Brown, C. T., Desai, N., Eisen, J. A., Evers, D., Field, D., Feng, W., Huson, D., Jansson, J., Knight, R., Knight, J., Kolker, E., Konstantindis, K., Kostka, J., Kyrpides, N., … Stevens, R. (2010). Meeting Report: The Terabase Metagenomics Workshop and the Vision of an Earth Microbiome Project. *Standards in Genomic Sciences*, *3*(3), 243–248. https://doi.org/10.4056/sigs.1433550

Gittel, A., Kofoed, M. V. W., Sørensen, K. B., Ingvorsen, K., & Schramm, A. (2012). Succession of Deferribacteres and Epsilonproteobacteria through a nitrate-treated high-temperature oil production facility. *Systematic and Applied Microbiology*, *35*(3), 165–174. https://doi.org/10.1016/j.syapm.2012.01.003

Gupta, G., Ndiaye, A., & Filteau, M. (2021). Leveraging Experimental Strategies to Capture Different Dimensions of Microbial Interactions. *Frontiers in Microbiology*, *12*. https://www.frontiersin.org/articles/10.3389/fmicb.2021.700752

Handelsman, J. (2004). Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews : MMBR*, *68*(4), 669–685. https://doi.org/10.1128/MMBR.68.4.669-685.2004

Hedlund, B. P., Chuvochina, M., Hugenholtz, P., Konstantinidis, K. T., Murray, A. E., Palmer, M., Parks, D. H., Probst, A. J., Reysenbach, A.-L., Rodriguez-R, L. M., Rossello-Mora, R., Sutcliffe, I. C., Venter, S. N., & Whitman, W. B. (2022). SeqCode: A nomenclatural code for prokaryotes described from sequence data. *Nature Microbiology*, *7*(10), 1702–1708. https://doi.org/10.1038/s41564-022-01214-9

Hedlund, B. P., Dodsworth, J. A., Murugapiran, S. K., Rinke, C., & Woyke, T. (2014). Impact of single-cell genomics and metagenomics on the emerging view of extremophile "microbial dark matter." *Extremophiles*, *18*(5), 865–875. https://doi.org/10.1007/s00792-014-0664-7

Hedlund, B. P., Murugapiran, S. K., Alba, T. W., Levy, A., Dodsworth, J. A., Goertz, G. B., Ivanova, N., & Woyke, T. (2015). Uncultivated thermophiles: Current status and spotlight on 'Aigarchaeota.' *Environmental Microbiology • Extremophiles*, *25*, 136–145. https://doi.org/10.1016/j.mib.2015.06.008

Hu, Q., Wang, S., Lai, Q., Shao, Z., & Jiang, L. (2021). Sulfurimonas indica sp. Nov., a hydrogen- and sulfur-oxidizing chemolithoautotroph isolated from a hydrothermal sulfide chimney in the Northwest Indian Ocean. *International Journal of Systematic and Evolutionary Microbiology*, *71*(1). https://doi.org/10.1099/ijsem.0.004575

Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., … Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, *12*(2), 115–121. https://doi.org/10.1038/nmeth.3252

Hufnagel, D. A., Depas, W. H., & Chapman, M. R. (2015). The Biology of the Escherichia coli Extracellular Matrix. *Microbiology Spectrum*, *3*(3). https://doi.org/10.1128/microbiolspec.MB-0014-2014

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., Hernsdorf, A. W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D. A., Finstad, K. M., Amundson, R., Thomas, B. C., & Banfield, J. F. (2016). A new view of the tree of life. *Nature Microbiology*, *1*(5), 16048. https://doi.org/10.1038/nmicrobiol.2016.48

Imachi, H., Nobu, M. K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M., Matsui, Y., Miyazaki, M., Murata, K., Saito, Y., Sakai, S., Song, C., Tasumi, E., Yamanaka, Y., Yamaguchi, T., … Takai, K. (2020). Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature*, *577*(7791), 519–525. https://doi.org/10.1038/s41586-019-1916-6

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., & Korpela, K. (2020). Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. *PloS One*, *15*(1), e0227285. https://doi.org/10.1371/journal.pone.0227285

Jiao, J.-Y., Liu, L., Hua, Z.-S., Fang, B.-Z., Zhou, E.-M., Salam, N., Hedlund, B. P., & Li, W.-J. (2021). Microbial dark matter coming to light: Challenges and opportunities. *National Science Review*, *8*(3), nwaa280. https://doi.org/10.1093/nsr/nwaa280

Kang, D. D., Froula, J., Egan, R., & Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*, *3*, e1165. https://doi.org/10.7717/peerj.1165

Katayama, T., Nobu, M. K., Kusada, H., Meng, X.-Y., Hosogi, N., Uematsu, K., Yoshioka, H., Kamagata, Y., & Tamaki, H. (2020). Isolation of a member of the candidate phylum 'Atribacteria' reveals a unique cell membrane structure. *Nature Communications*, *11*(1), 6381. https://doi.org/10.1038/s41467-020-20149-5

Kim, Y. C., & Anderson, A. J. (2018). Rhizosphere pseudomonads as probiotics improving plant health. *Molecular Plant Pathology*, *19*(10), 2349–2359. https://doi.org/10.1111/mpp.12693

Kodama, Y., & Watanabe, K. (2004). Sulfuricurvum kujiense gen. Nov., sp. Nov., a facultatively anaerobic, chemolithoautotrophic, sulfur-oxidizing bacterium isolated from an underground crude-oil storage cavity. *International Journal of Systematic and Evolutionary Microbiology*, *54*(Pt 6), 2297–2300. https://doi.org/10.1099/ijs.0.63243-0

Kotera, M., & Goto, S. (2016). Metabolic pathway reconstruction strategies for central metabolism and natural product biosynthesis. *Biophysics and Physicobiology*, *13*, 195–205. https://doi.org/10.2142/biophysico.13.0_195

Kunin Victor, Copeland Alex, Lapidus Alla, Mavromatis Konstantinos, & Hugenholtz Philip. (2008). A Bioinformatician's Guide to Metagenomics. *Microbiology and Molecular Biology Reviews*, *72*(4), 557–578. https://doi.org/10.1128/MMBR.00009-08

Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*(22), 3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Lewis, W. H., Tahon, G., Geesink, P., Sousa, D. Z., & Ettema, T. J. G. (2021). Innovations to culturing the uncultured microbial majority. *Nature Reviews Microbiology*, *19*(4), 225–240. https://doi.org/10.1038/s41579-020-00458-8

Li, C., Li, X., Guo, R., Ni, W., Liu, K., Liu, Z., Dai, J., Xu, Y., Abduriyim, S., Wu, Z., Zeng, Y., Lei, B., Zhang, Y., Wang, Y., Zeng, W., Zhang, Q., Chen, C., Qiao, J., Liu, C., & Hu, S. (2023). Expanded catalogue of metagenome-assembled genomes reveals resistome characteristics and athletic performance-associated microbes in horse. *Microbiome*, *11*(1), 7. https://doi.org/10.1186/s40168-022-01448-z

Lin, X., Wakeham, S. G., Putnam, I. F., Astor, Y. M., Scranton, M. I., Chistoserdov, A. Y., & Taylor, G. T. (2006). Comparison of vertical distributions of prokaryotic assemblages in the anoxic Cariaco Basin and Black Sea by use of fluorescence in situ hybridization. *Applied and Environmental Microbiology*, *72*(4), 2679–2690.

Liu, S., Moon, C. D., Zheng, N., Huws, S., Zhao, S., & Wang, J. (2022). Opportunities and challenges of using metagenomic data to bring uncultured microbes into cultivation. *Microbiome*, *10*(1), 76. https://doi.org/10.1186/s40168-022-01272-5

Lloyd Karen G., Steen Andrew D., Ladau Joshua, Yin Junqi, & Crosby Lonnie. (2018). Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. *mSystems*, *3*(5), 10.1128/msystems.00055-18. https://doi.org/10.1128/msystems.00055-18

Lynch, M. D. J., & Neufeld, J. D. (2015). Ecology and exploration of the rare biosphere. *Nature Reviews Microbiology*, *13*(4), 217–229. https://doi.org/10.1038/nrmicro3400

Ma, B., Wang, Y., Zhao, K., Stirling, E., Lv, X., Yu, Y., Hu, L., Tang, C., Wu, C., Dong, B., Xue, R., Dahlgren, R. A., Tan, X., Dai, H., Zhu, Y.-G., Chu, H., & Xu, J. (2024). Biogeographic patterns and drivers of soil viromes. *Nature Ecology & Evolution*, *8*(4), 717–728. https://doi.org/10.1038/s41559-024-02347-2

Ma, L., Konkel, M. E., & Lu, X. (2021). Antimicrobial Resistance Gene Transfer from Campylobacter jejuni in Mono- and Dual-Species Biofilms. *Applied and Environmental Microbiology*, *87*(15), e0065921. https://doi.org/10.1128/AEM.00659-21

Maruyama, R., Yasumoto, K., Mizusawa, N., Iijima, M., Yasumoto-Hirose, M., Iguchi, A., Hermawan, O. R., Hosono, T., Takada, R., Song, K.-H., Shinjo, R., Watabe, S., & Yasumoto, J. (2024). Metagenomic analysis of the microbial communities and associated network of nitrogen metabolism genes in the Ryukyu limestone aquifer. *Scientific Reports*, *14*(1), 4356. https://doi.org/10.1038/s41598-024-54614-8

McKinnon, K. M. (2018). Flow Cytometry: An Overview. *Current Protocols in Immunology*, *120*(1), 5.1.1-5.1.11. https://doi.org/10.1002/cpim.40

Meziti Alexandra, Rodriguez-R Luis M., Hatt Janet K., Peña-Gonzalez Angela, Levy Karen, & Konstantinidis Konstantinos T. (2021). The Reliability of Metagenome-Assembled Genomes (MAGs) in Representing Natural Populations: Insights from Comparing MAGs against Isolate Genomes Derived from the Same Fecal Sample. *Applied and Environmental Microbiology*, *87*(6), e02593-20. https://doi.org/10.1128/AEM.02593-20

Mills, C. L., Beuning, P. J., & Ondrechen, M. J. (2015). Biochemical functional predictions for protein structures of unknown or uncertain function. *Computational and Structural Biotechnology Journal*, *13*, 182–191. https://doi.org/10.1016/j.csbj.2015.02.003

Mise, K., & Iwasaki, W. (2022). Unexpected absence of ribosomal protein genes from metagenome-assembled genomes. *ISME Communications*, *2*(1), 118. https://doi.org/10.1038/s43705-022-00204-6

Monica Steffi Matchado, Malte Rühlemann, Sandra Reimeiter, Tim Kacprowski, Fabian Frost, Dirk Haller, Jan Baumbach, & Markus List. (2023). On the limits of 16S rRNA gene-based metagenome prediction and functional profiling. *bioRxiv*, 2023.11.07.564315. https://doi.org/10.1101/2023.11.07.564315

Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T. L., Agarwala, R., & Schäffer, A. A. (2008). Database indexing for production MegaBLAST searches. *Bioinformatics (Oxford, England)*, *24*(16), 1757–1764. https://doi.org/10.1093/bioinformatics/btn322

Murray, A. E., Freudenstein, J., Gribaldo, S., Hatzenpichler, R., Hugenholtz, P., Kämpfer, P., Konstantinidis, K. T., Lane, C. E., Papke, R. T., Parks, D. H., Rossello-Mora, R., Stott, M. B., Sutcliffe, I. C., Thrash, J. C., Venter, S. N., Whitman, W. B., Acinas, S. G., Amann, R. I., Anantharaman, K., … Reysenbach, A.-L. (2020). Roadmap for naming uncultivated Archaea and Bacteria. *Nature Microbiology*, *5*(8), 987–994. https://doi.org/10.1038/s41564-020-0733-x

Murrell, J. C., McDonald, I. R., & Bourne, D. G. (1998). Molecular methods for the study of methanotroph ecology. *FEMS Microbiology Ecology*, *27*(2), 103–114. https://doi.org/10.1111/j.1574-6941.1998.tb00528.x

Nagies, F. S. P., Brueckner, J., Tria, F. D. K., & Martin, W. F. (2020). A spectrum of verticality across genes. *PLOS Genetics*, *16*(11), e1009200. https://doi.org/10.1371/journal.pgen.1009200

Nakagawa, S., Takaki, Y., Shimamura, S., Reysenbach, A.-L., Takai, K., & Horikoshi, K. (2007). Deep-sea vent ε-proteobacterial genomes provide insights into emergence of pathogens. *Proceedings of the National Academy of Sciences*, *104*(29), 12146–12150. https://doi.org/10.1073/pnas.0700687104

Naud, S., Ibrahim, A., Valles, C., Maatouk, M., Bittar, F., Tidjani Alou, M., & Raoult, D. (2022). Candidate Phyla Radiation, an Underappreciated Division of the Human Microbiome, and Its Impact on Health and Disease. *Clinical Microbiology Reviews*, *35*(3), e0014021. https://doi.org/10.1128/cmr.00140-21

Nikolenko, S. I., Korobeynikov, A. I., & Alekseyev, M. A. (2013). BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics*, *14*(1), S7. https://doi.org/10.1186/1471-2164-14-S1-S7

Nocker, A., & Camper, A. K. (2006). Selective removal of DNA from dead cells of mixed bacterial communities by use of ethidium monoazide. *Applied and Environmental Microbiology*, *72*(3), 1997–2004. https://doi.org/10.1128/AEM.72.3.1997-2004.2006

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: A new versatile metagenomic assembler. *Genome Research*, *27*, 824–834. https://doi.org/10.1101/gr.213959.116

Oljira, T., Muleta, D., & Jida, M. (2018). Potential Applications of Some Indigenous Bacteria Isolated from Polluted Areas in the Treatment of Brewery Effluents. *Biotechnology Research International*, *2018*, 9745198. https://doi.org/10.1155/2018/9745198

Orakov, A., Fullam, A., Coelho, L. P., Khedkar, S., Szklarczyk, D., Mende, D. R., Schmidt, T. S. B., & Bork, P. (2021). GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome Biology*, *22*(1), 178. https://doi.org/10.1186/s13059-021-02393-0

Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P.-A., & Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research*, *50*(D1), D785–D794. https://doi.org/10.1093/nar/gkab776

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*, *36*(10), 996–1004. https://doi.org/10.1038/nbt.4229

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Pavlopoulos, G. A., Baltoumas, F. A., Liu, S., Selvitopi, O., Camargo, A. P., Nayfach, S., Azad, A., Roux, S., Call, L., Ivanova, N. N., Chen, I. M., Paez-Espino, D., Karatzas, E., Acinas, S. G., Ahlgren, N., Attwood, G., Baldrian, P., Berry, T., Bhatnagar, J. M.,

… Novel Metagenome Protein Families Consortium. (2023). Unraveling the functional dark matter through global metagenomics. *Nature*, *622*(7983), 594–602. https://doi.org/10.1038/s41586-023-06583-7

Pham, V. H. T., & Kim, J. (2012). Cultivation of unculturable soil bacteria. *Trends in Biotechnology*, *30*(9), 475–484. https://doi.org/10.1016/j.tibtech.2012.05.007

Pope, P. B., Smith, W., Denman, S. E., Tringe, S. G., Barry, K., Hugenholtz, P., McSweeney, C. S., McHardy, A. C., & Morrison, M. (2011). Isolation of Succinivibrionaceae Implicated in Low Methane Emissions from Tammar Wallabies. *Science*, *333*(6042), 646–648. https://doi.org/10.1126/science.1205760

Power, J. F., Carere, C. R., Lee, C. K., Wakerley, G. L. J., Evans, D. W., Button, M., White, D., Climo, M. D., Hinze, A. M., Morgan, X. C., McDonald, I. R., Cary, S. C., & Stott, M. B. (2018). Microbial biogeography of 925 geothermal springs in New Zealand. *Nature Communications*, *9*(1), 2876. https://doi.org/10.1038/s41467-018-05020-y

Power, J. F., Lowe, C. L., Carere, C. R., McDonald, I. R., Cary, S. C., & Stott, M. B. (2023). Temporal dynamics of geothermal microbial communities in Aotearoa-New Zealand. *Frontiers in Microbiology*, *14*. https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1094311

Prasad, S., Malav, L. C., Choudhary, J., Kannojiya, S., Kundu, M., Kumar, S., & Yadav, A. N. (2021). Soil Microbiomes for Healthy Nutrient Recycling. In A. N. Yadav, J. Singh, C. Singh, & N. Yadav (Eds.), *Current Trends in Microbial Biotechnology for Sustainable Agriculture* (pp. 1–21). Springer Singapore. https://doi.org/10.1007/978-981-15-6949-4_1

Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution*, *26*(7), 1641–1650. https://doi.org/10.1093/molbev/msp077

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE*, *5*(3), e9490. https://doi.org/10.1371/journal.pone.0009490

Purahong, W., Stempfhuber, B., Lentendu, G., Francioli, D., Reitz, T., Buscot, F., Schloter, M., & Krüger, D. (2015). Influence of Commonly Used Primer Systems on Automated Ribosomal Intergenic Spacer Analysis of Bacterial Communities in Environmental Samples. *PLOS ONE*, *10*(3), e0118967. https://doi.org/10.1371/journal.pone.0118967

R Core Team. (2023). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria.* <https://www.R-project.org/>.

Rainey, F. A., & Hollen, B. (2015). Desulfurella. In *Bergey's Manual of Systematics of Archaea and Bacteria* (pp. 1–4). https://doi.org/10.1002/9781118960608.gbm01037

Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., Dmitrieff, E., Malmstrom, R., Stepanauskas, R., & Woyke, T. (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS–based single-cell genomics. *Nature Protocols*, *9*(5), 1038–1048. https://doi.org/10.1038/nprot.2014.067

Salipante, S. J., Kawashima, T., Rosenthal, C., Hoogestraat, D. R., Cummings, L. A., Sengupta, D. J., Harkins, T. T., Cookson, B. T., & Hoffman, N. G. (2014). Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Applied and Environmental Microbiology*, *80*(24), 7583–7591. https://doi.org/10.1128/AEM.02206-14

Sánchez, B., Delgado, S., Blanco-Míguez, A., Lourenço, A., Gueimonde, M., & Margolles, A. (2017). Probiotics, gut microbiota, and their influence on host health and disease. *Molecular Nutrition & Food Research*, *61*(1), 1600240. https://doi.org/10.1002/mnfr.201600240

Schultz, J., Modolon, F., Peixoto, R. S., & Rosado, A. S. (2023). Shedding light on the composition of extreme microbial dark matter: Alternative approaches for culturing

extremophiles. *Frontiers in Microbiology*, *14*, 1167718. https://doi.org/10.3389/fmicb.2023.1167718

Sobol, M. S., & Kaster, A.-K. (2023). Back to Basics: A Simplified Improvement to Multiple Displacement Amplification for Microbial Single-Cell Genomics. *International Journal of Molecular Sciences*, *24*(5). https://doi.org/10.3390/ijms24054270

Solden, L., Lloyd, K., & Wrighton, K. (2016). The bright side of microbial dark matter: Lessons learned from the uncultivated majority. *Current Opinion in Microbiology*, *31*, 217–226. https://doi.org/10.1016/j.mib.2016.04.020

Stams, A. J. M., & Plugge, C. M. (2009). Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nature Reviews Microbiology*, *7*(8), 568–577. https://doi.org/10.1038/nrmicro2166

Steen, A. D., Crits-Christoph, A., Carini, P., DeAngelis, K. M., Fierer, N., Lloyd, K. G., & Thrash, J. C. (2019). High proportions of bacteria and archaea across most biomes remain uncultured. *The ISME Journal*, *13*(12), 3126–3130. https://doi.org/10.1038/s41396-019-0484-y

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., Prill, R. J., Tripathi, A., Gibbons, S. M., Ackermann, G., Navas-Molina, J. A., Janssen, S., Kopylova, E., Vázquez-Baeza, Y., González, A., Morton, J. T., Mirarab, S., Zech Xu, Z., Jiang, L., … The Earth Microbiome Project Consortium. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, *551*(7681), 457–463. https://doi.org/10.1038/nature24621

Tiwari, S., & Lata, C. (2018). Heavy Metal Stress, Signaling, and Tolerance Due to Plant-Associated Microbes: An Overview. *Frontiers in Plant Science*, *9*, 452. https://doi.org/10.3389/fpls.2018.00452

Tshikantwa, T. S., Ullah, M. W., He, F., & Yang, G. (2018). Current Trends and Potential Applications of Microbial Interactions for Human Welfare. *Frontiers in Microbiology*, *9*, 1156. https://doi.org/10.3389/fmicb.2018.01156

Tyson, G. W., Lo, I., Baker, B. J., Allen, E. E., Hugenholtz, P., & Banfield, J. F. (2005). Genome-directed isolation of the key nitrogen fixer Leptospirillum ferrodiazotrophum sp. Nov. From an acidophilic microbial community. *Applied and Environmental Microbiology*, *71*(10), 6319–6324. https://doi.org/10.1128/AEM.71.10.6319-6324.2005

Uritskiy, G. V., DiRuggiero, J., & Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*, *6*(1), 158. https://doi.org/10.1186/s40168-018-0541-1

van der Stel, A.-X., & Wösten, M. M. S. M. (2019). Regulation of Respiratory Pathways in Campylobacterota: A Review. *Frontiers in Microbiology*, *10*, 1719. https://doi.org/10.3389/fmicb.2019.01719

van Teeseling, M. C. F., & Jogler, C. (2021a). Cultivation of elusive microbes unearthed exciting biology. *Nature Communications*, *12*(1), 75. https://doi.org/10.1038/s41467-020-20393-9

van Teeseling, M. C. F., & Jogler, C. (2021b). Cultivation of elusive microbes unearthed exciting biology. *Nature Communications*, *12*(1), 75. https://doi.org/10.1038/s41467-020-20393-9

Varadi, M., & Velankar, S. (2023). The impact of AlphaFold Protein Structure Database on the fields of life sciences. *PROTEOMICS*, *23*(17), 2200128. https://doi.org/10.1002/pmic.202200128

Voit, E. O. (2020). 27C3The—Omics revolution. In E. O. Voit (Ed.), *Systems Biology: A Very Short Introduction* (p. 0). Oxford University Press. https://doi.org/10.1093/actrade/9780198828372.003.0003

Vollmers, J., Wiegand, S., Lenk, F., & Kaster, A.-K. (2022). How clear is our current view on microbial dark matter? (Re-)assessing public MAG & SAG datasets with MDMcleaner. *Nucleic Acids Research*, *50*(13), e76. https://doi.org/10.1093/nar/gkac294

Waite, D., Chuvochina, M., & Philip, H. (2019). *Road Map of the Phylum Campylobacterota* (pp. 1–11). https://doi.org/10.1002/9781118960608.bm00040

Waite, D. W., Chuvochina, M., Pelikan, C., Parks, D. H., Yilmaz, P., Wagner, M., Loy, A., Naganuma, T., Nakai, R., Whitman, W. B., Hahn, M. W., Kuever, J., & Hugenholtz, P. (2020). Proposal to reclassify the proteobacterial classes Deltaproteobacteria and Oligoflexia, and the phylum Thermodesulfobacteria into four phyla reflecting major functional capabilities. In *International Journal of Systematic and Evolutionary Microbiology* (Vol. 70, Issue 11, pp. 5972–6016). Microbiology Society. https://doi.org/10.1099/ijsem.0.004213

Waite, D. W., Vanwonterghem, I., Rinke, C., Parks, D. H., Zhang, Y., Takai, K., Sievert, S. M., Simon, J., Campbell, B. J., Hanson, T. E., Woyke, T., Klotz, M. G., & Hugenholtz, P. (2017). Comparative Genomic Analysis of the Class Epsilonproteobacteria and Proposed Reclassification to Epsilonbacteraeota (phyl. Nov.). *Frontiers in Microbiology*, *8*, 682. https://doi.org/10.3389/fmicb.2017.00682

Walters, K. A., Mohan, G., Myers, K. S., Ingle, A. T., Donohue, T. J., & Noguera, D. R. (2023). A metagenome-level analysis of a microbial community fermenting ultra-filtered milk permeate. *Frontiers in Bioengineering and Biotechnology*, *11*. https://www.frontiersin.org/articles/10.3389/fbioe.2023.1173656

Wang, L., Dash, S., Ng, C. Y., & Maranas, C. D. (2017). A review of computational tools for design and reconstruction of metabolic pathways. *Synthetic and Systems Biotechnology*, *2*(4), 243–252. https://doi.org/10.1016/j.synbio.2017.11.002

Williams, T. A., & Embley, T. M. (2014). Archaeal "Dark Matter" and the Origin of Eukaryotes. *Genome Biology and Evolution*, *6*(3), 474–481. https://doi.org/10.1093/gbe/evu031

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(11), 5088–5090. https://doi.org/10.1073/pnas.74.11.5088

Woodcroft, B. J., Aroney, S. T. N., Zhao, R., Cunningham, M., Mitchell, J. A. M., Blackall, L., & Tyson, G. W. (2024). SingleM and Sandpiper: Robust microbial taxonomic profiles from metagenomic data. *bioRxiv*, 2024.01.30.578060. https://doi.org/10.1101/2024.01.30.578060

Wu, Y.-W., Simmons, B. A., & Singer, S. W. (2016). MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, *32*(4), 605–607. https://doi.org/10.1093/bioinformatics/btv638

Xu, Y., & Zhao, F. (2018). Single-cell metagenomics: Challenges and applications. *Protein & Cell*, *9*(5), 501–510. https://doi.org/10.1007/s13238-018-0544-5

Yang, Z.-W., Lian, Z.-H., Liu, L., Fang, B.-Z., Li, W.-J., & Jiao, J.-Y. (2023). Cultivation strategies for prokaryotes from extreme environments. *iMeta*, *2*(3), e123. https://doi.org/10.1002/imt2.123

Yoshida-Takashima, Y., Takaki, Y., Yoshida, M., Zhang, Y., Nunoura, T., & Takai, K. (2022). Genomic insights into phage-host interaction in the deep-sea chemolithoautotrophic Campylobacterota, Nitratiruptor. *ISME Communications*, *2*(1), 108. https://doi.org/10.1038/s43705-022-00194-5

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Zengler, K. (2009). Central role of the cell in microbial ecology. *Microbiology and Molecular Biology Reviews : MMBR*, *73*(4), 712–729. https://doi.org/10.1128/MMBR.00027-09

Zengler, K., & Zaramela, L. S. (2018). The social network of microorganisms—How auxotrophies shape complex communities. *Nature Reviews Microbiology*, *16*(6), 383–390. https://doi.org/10.1038/s41579-018-0004-5

Zhao, J., Shakir, Y., Deng, Y., & Zhang, Y. (2023). Use of modified ichip for the cultivation of thermo-tolerant microorganisms from the hot spring. *BMC Microbiology*, *23*(1), 56. https://doi.org/10.1186/s12866-023-02803-2

Zhulin Igor B. (2016). Classic Spotlight: 16S rRNA Redefines Microbiology. *Journal of Bacteriology*, *198*(20), 2764–2765. https://doi.org/10.1128/jb.00616-16