Preservation and Detection of Molecular Signs of Life under Mars Analogue Conditions

Nick Rhys Thomas B.Sc M.Sc

Supervisors: Jaime Toney, Martin Lee, Umer Zeeshan Ijaz

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD)

School of Geographical & Earth Sciences

College of Science & Engineering

University of Glasgow.

Abstract

The search for life on Mars continues apace, however the significant cost in time and resources involved in each robotic mission, whether orbiter, lander, or even rover, necessitates the use of terrestrial Martian analogue sites. Few regions on Earth are able to accurately simulate the conditions on Mars, yet locations such as the Atacama Desert of northern Chile and McMurdo Dry Valleys of Antarctica offer invaluable opportunities to test experiments and equipment prior to missions, as well as to advance understanding of the survival of life at the limits of habitability. This study presents an exciting new analogue region in the Chilean Altiplano, offering a unique balance of Mars-like conditions not experienced in other terrestrial Mars analogues. Utilising this novel Martian analogue, the present study examines how signs of ancient life may be preserved in the soils and sediments of Mars, investigating the extent of protection afforded by soil coverage against the inhospitable conditions of the surface. In addition, this study presents a new approach to understanding the bacterial community which does survive in such a hostile environment on Earth, illustrating the effect of soil coverage on the survival of bacteria in a Mars-like environment. The rapidly-changing nature of hostile soil conditions with increasing soil depth was found have stark implications for the likelihood of success of upcoming missions such as the ESA 2020 ExoMars Rover, as well as advancing the understanding of life in extreme – and under reported – environments on Earth.

Contents

Acknowledgements

First of all, this project would not have been possible without the generous funding provided by the UK Space Agency, their funding of studies such as this illustrates the commitment to space science on display in the UK.

Gratitude must be extended to my supervisory team; it has been fluid over the years, and I am thankful for the support of all. Thanks go to Dr Vernon Phoenix, Dr Jaime Toney, Dr Christopher Quince, Dr Umer Zeeshan Ijaz, and Professor Martin Lee.

Additional thanks go to the support I have received from colleagues in both the GES and Engineering schools; thanks to Dr Jillian Couto, Kenny Roberts, Julie Russell, Anne McGarrity, Harry Jackson, and Ali Salik for invaluable advice and assistance.

Fellow PhDs in my own research groups as well as others are all due for recognition, and special thanks are extended to my fellow astrobiology research group members Philippe Nauny and Rory Porteous. The BECS research group as well as the Water Group from the School of Engineering have been a source of encouragement and inspiration throughout the PhD.

Last but certainly not least, we come to my loving family and friends, without whom I would certainly have never finished the mammoth task that is a PhD. Special mention must go to Adrienne Macartney for her love and care over the years, as well as keeping me sane and putting up with my various oddities, to my increasingly-worldwide family, and to any friends I have missed.

Introduction

1.0: Mars

The planet Mars has long been of special scientific interest; at first glance it appears similar to the Earth – a small terrestrial planet in the inner solar system with a 24 hour day cycle (Ashbrook, 1953), a thin atmosphere, and evidence of geological activity (Hynek et al, 2003). Mars is around half the diameter of the Earth, with the surface area of Mars being marginally smaller than the dry land surface area of Earth. The current state of Mars differs from that of the Earth in several critical features such as the composition and density of the atmosphere (Seiff and Kirk, 1977), the average surface and air temperatures (Martin and Bridges, 2003), the intensity of radiation received from space (Guo et al, 2015), present day geological activity, and the volume of water at the surface, which on Mars is apparently restricted to the polar caps.

1.1 History of Mars

1.1.1 Geological history of Mars

Mars exhibits clear evidence of past volcanism, from the notable existence of the colossal volcano Olympus Mons to the presence of Aeolian dunes – which without an obvious water source are best explained by ancient explosive volcanic activity (Edgett, 1997). Further evidence for past volcanism may be found in the apparent deposition of extinct volcanic hot springs on the surface of Mars, which has been hypothesized by studies such as Brakenridge et al (1985), more recently by Allen and Oehler (2008), and most recently by Ruff and Farmer (2016). The observations by Ruff and colleagues are of particular interest as the observed silica deposits appeared to resemble those observed in terrestrial spring systems such as the Geysers Del Tatio, in Chile. It is theorised that early Martian geological activity may have led to the rise of the Martian atmosphere, through release of H_2O and CO_2 from volcanic magma (Jakosky and Phillips, 2001). The early Martian atmosphere is believed to have been denser than the sparse atmosphere observed on Mars today (Kurokawa et al, 2018).

This more dense, higher pressure atmosphere would consequently raise the freezing point of any water, thus offering credence to the notion of a warm, wet early Mars, which may have harboured life at one time. While Mars clearly has a geologically-active past, in recent epochs the planet has evidently been still, as relatively recent large impact basins show no evidence of having been affected by a planetary magnetosphere (Roberts and Arkani-Hamed, 2017), the lack of which is implicated in part with the loss of the atmosphere, as the lack of protective magnetic field allowed the atmosphere to be removed by solar wind.

1.1.2 Existence of past water on Mars

Evidence for water on ancient Mars is relatively abundant, large features such as the famed Valles Marineris are clear indications of past hydraulic action, indeed valleys attributed to glacial action are commonplace on Mars (Fassett et al, 2010). There are at present two opposing theories regarding the presence of water on ancient Mars; was the planet cold and frozen, or warm and wet? Evidence has been observed in both cases, for example, Craddock and Howard (2001) cite the existence of river valley-type networks as evidence of past rainfall and runoff action, suggesting a balmy ancient Mars with an active hydrological cycle as is observed on Earth. Conversely, low atmospheric pressure and and a cooler, younger Sun are stated by Wordsworth et al (2012) to be a root cause of a cold, frozen Mars with only periodic icemelt during the Martian summer to explain the observed evidence of hydraulic action on the surface. For now, the nature of ancient Mars' hydrological state remains unsolved. The existence of past water is of particular interest to astrobiologists, and a warm, damp ancient Mars would be a better candidate for harbouring life than a cold, frozen ancient Mars, adding to the possibility of molecular fossils being preserved in Martian sediment layers to this day – if they existed in the first place.

1.2 Challenges to life presented by conditions on present-day Mars

The surface conditions of present day Mars could be best described as a frozen desert. Temperatures of the air and soil are extremely low, and additionally feature a substantial diurnal variation of temperature. The landscape is dominated by rocks and dunes, with no bodies of water, save for the frozen polar caps.

Fig[1]: Martian surface image taken by MSL Curiosity Rover Mastcam instrument. Image credit: NASA

The planet is not particularly geologically active, and weather phenomena such as precipitation are unlikely, although the possibility has been suggested previously (Kahn, 1990). The most common weather-like event is likely the frequent dust storms experienced across the planet due to high winds and dry, fine particulate soils (Mulholland et al, 2012). Such dust storms can be immense in magnitude; the largest dust storms have been reported to circle the entire circumference of the planet (Cantor, 2006).

Present day Mars is devoid of large bodies of liquid water, and liquid water does not appear to be stable under the observed current state of temperature and air pressure at the surface, at least to the extent required to produce bulk surface deposits such as lakes. While in many

regions of the planet, temperatures are suitable at least for some of the Martian year, the atmospheric pressure is insufficient to prevent evaporative loss (Haberle et al, 2001). Any obvious observable water on Mars is confined to the polar ice caps, although evidence for periodic flows of water in the lower latitudes has been observed and reported in studies such as that by McEwen et al (2011), taking the form of recurring slope lineae, which were suggested to be evidence of subsurface ice melt during the Martian summer.

Mars presents a significant challenge to survival of any life; conditions both at the surface and in the atmosphere appear incompatible with life as we know it. There is a lack of stable sources of liquid water, damaging reactive soil chemistry, extreme solar and cosmic radiation environment, and extremely low average temperatures which additionally feature a substantial diurnal variation. The combination of such a plethora of challenges to life consequently indicates that extant life is unlikely to be found on the present-day Martian surface.

1.2.1 Lack of stability of liquid water

One of the principal challenges to life on Mars is the lack of stable sources of liquid water; without a suitable solvent even simple organic life cannot persist, and indeed dehydration is lethal to the majority of known organisms. It is of little surprise that the most inhospitable regions on Earth are frequently those which exhibit the greatest aridity such as deserts.

Species have been observed which exhibit resilience to dehydration, often through survival mechanisms such as spore formation (Dittmann et al, 2015). Indeed, bacterial sporulation enables survival of many hazards (Atrih and Foster, 2002).

Non-water solvents such as liquid ammonia (NH_4) or Methane (CH_4) have been suggested to be possible replacements for water as the solvent around which life is based, in particular this has been considered for Saturn's giant moon Titan, which the Cassini-Huygens mission has observed to feature an apparent liquid methane cycle similar to the water cycle of earth, with methane 'rain' and possibly even bodies of liquid methane on the surface (Sagan and Dermott, 1982). Methane has not been detected to be prevalent to such an extent on Mars, and so the notion of methane as a possible solvent basis for life on Mars can be discounted, although recently unusual 'plumes' of methane were detected by the MSL Curiosity rover (Webster et al, 2015).

1.2.4 Low air temperatures

Average temperatures on present day Mars are extremely low and typically feature a high diurnal variation, especially in the Martian summer (Ulrich et al, 2008. This is partially influenced by the lack of effective greenhouse effect due to the thin atmosphere, the pressure of which is on average around 6mbar (Jakosky and Phillips, 2001). Air temperatures are strongly affected by the frequent Martian dust storms, as particulates lifted into the atmosphere.

1.2.2 Oxidising soil chemistry

The mineralogical constitution of the soils on Mars, combined with the nature of the atmosphere and the extensive UV irradiation received at the surface, have the effect of producing a chemistry in the Martian surface regoliths that is highly oxidative and which is unlikely to be capable of sustaining organics sufficiently for life to flourish (Lasne et al, 2016). Early robotic landing missions such as the NASA Viking landers were equipped to examine the soils for possible biological activity. Viking 1, for example, featured a suite of microbiological experiments including a Gas Chromatograpy-Mass Spectrometry (GC-MS) instrument to examine compounds extracted from excavated soils, a Gas Exchange experiment with the aim of observing any gas metabolism by soil microbes, a Labelled Release experiment which utilised radiolabelled nutrient media applied to soil samples, allowing attempts to detect metabolism by soil bacteria, and a Pyrolytic Release experiment designed to detect potential photosynthetic activity.

The Labelled Release experiment generated significant interest in the scientific community, as early data appeared to indicate positive results (Levin and Straat, 1977), with release of the radiolabelled Carbon from the soil as CO_2 gas – as if it were the case that soil bacteria had indeed been able to metabolise the added nutrient media. Upon further examination, however, it was determined that the observations were attributed to abiotic processes related to the oxidising soil chemistry of the Martian regolith (Schuerger and Clark, 2008), and thus it was ultimately determined that Viking 1 had failed to detect active microbiological action in the Martian soils.

Particular focus should be given to the presence of damaging perchlorate compounds in the Marian regoliths (Hecht et al, 2009), indeed perchlorates were believed to be among the main sources of the oxidative soil chemistry recorded by Viking 1 (Quinn et al, 2013) and are certainly damaging to cellular structure (Zhao et al, 2014). A recent study utilising synthetic simulated Martian analogue soils seeded with bacteria examined the cytotoxicity of soil perchlorates on the bacteria in this analogue environment, under a Mars-like UV radiation profile. Lethality of the perchlorates was found to be exacerbated by UV (Wadsworth and Cockell, 2017), alongside other bacteriocidal agents in the Martian soils such as H_2O_2, thus rendering the Martian surface conditions possibly hostile to life to an even greater degree than previously understood.

1.2.3 Extreme radiation environment

The radiation environment at the surface of Mars is extremely hostile to life, with the surface environment receiving both extensive solar flux and high energy cosmic radiation. Whilst the orbit of Mars is much further from the sun than Earth, the lack of dense atmosphere and lack of planetary magnetosphere combine to offer the surface only minimal protection from UV radiation (Ronto et al, 2002) and solar wind. In addition, the Martian surface receives radiation from the highly damaging UVC wavelengths (100-280nm), which are entirely attenuated by the atmosphere of Earth. In addition to the solar radiation, Mars receives little protection from high energy cosmic radiation. On Earth, such cosmic radiation is largely attenuated by the magnetosphere, but Mars is lacking a strong planetary magnetic field, and so is bereft of protection.

Ultra Violet light is capable of inducing considerable damage to non-adapted cells, and the UVC wavelengths in particular are capable of destructive action on structures such as DNA (Nair and Loppnow, 2013), due in part to the fact that DNA absorbs UV radiation most strongly at around 260-265nm – central in the UVC region of the spectrum (Ravanat and Douki, 2016). UV exposure causes cellular damage through the generation of reactive oxygen species within the cell, which are capable of ill effects on the cell such as inflicting oxidative damage to DNA and catalysing lipid peroxidation (He and Häder, 2002). UV exposure is sufficiently deadly to many bacterial species that UV lamps are used in clinical settings for sterilising instruments and equipment, and in some cases biological materials (Mohr et al, 2009), as well as in some cancer therapy strategies (Uehara et al, 2014). UV-range LED

lamps have been found particularly effective for sterilisation compared to classical UV lamps (Kim et al, 2016). While some species of bacteria have adapted to survive or indeed utilise UV radiation, such as photosynthetic species including *Cyanobacteria* or radioresistant bacteria such as *Deinococcus* species (Tian et al, 2010), the majority of bacterial species are not able to survive extreme levels of UV radiation or damaging ionizing radiation. Gram-positive bacteria are noted to be more resilient to UV damage than Gram-negative species (Beauchamp and Lacroix, 2012), however in clinical settings UV sterilisation has been observed to be effective in killing of common pathogenic species such as *K.pneumonieae* and *S.aureus* (Lee et al, 2018), and can even be used to destroy more resilient structures such as *Bascillus* spores (Nhung et al, 2012). Damage caused by UV and ionizing radiation is also a danger to plants, to the extent that plant nuclei have been observed to move within the cell to evade exposure (Iwabuchi et al, 2016). The potentiality for cellular damage due to extensive UV exposure is of particular interest to this study, and is taken into account in the choice of field locations.

2.0: Terrestrial Mars Analogues

There are a wide range of environments on Earth which experience selected conditions or combinations of conditions currently present on Mars, which are termed Terrestrial Martian Analogue Regions. Such Mars analogue sites are invaluable to science as they offer an environment suitable for experiment, hardware and instrument testing, and robotic mission field testing, as well as studies such as the recent HI-SEAS IV study, during which a group of scientists and engineers lived under simulated Mars conditions on the slopes of the Hawaiian volcano Mauna Loa for one year (Nasrini et al, 2017 conference paper) – allowing testing under simulated Martian environmental stressors without the vast expense and time required for each actual Mars mission.

Logically, the ideal terrestrial Mars analogue should feature all prevalent environmental conditions observed at the Martian surface, however this is not possible naturally on Earth due to the dense atmosphere and protective magnetosphere of the Earth, which cannot easily be removed from the environment in the field, yet starkly reduce the possible hostility of any environment on Earth in comparison to Mars.

There are a wide range of locations around the Earth which have been identified as viable Martian analogues for one of more environmental features experienced in the region, examples of these are the Atacama Desert in northern Martian, the McMurdo Dry Valleys of Antarctica, the permafrosts of Svalbard, Norway, volcanic fields in Iceland, and volcanic hotsprings such as the Geysers Del Tatio in the Chilean Altiplano. In addition, synthetic analogues have been produced such as the Mars Environmental Simulation Chamber (MESCH) (Jensen et al, 2008), as well as commercially available manufactured analogue Mars soils designed to mirror the chemistry of the Martian regolith as closely as possible.

2.1 Atacama Desert

The Atacama Desert is a large region in northern Martian which is noted as being among the most arid locations on the planet. In addition, the Atacama is at a relatively high altitude above sea level (typically around 2400m.a.s.l) and thus experiences low air pressure and increased solar flux in comparison to lower lying deserts such as the Sahara. The soils of Atacama are extremely dry and weakly acidic, and in addition Atacama features expansive salt flats such as the Salar de Atacama.

Fig[2]: Atacama Desert, looking towards the Sairecabur Volcanic Range (self shot).

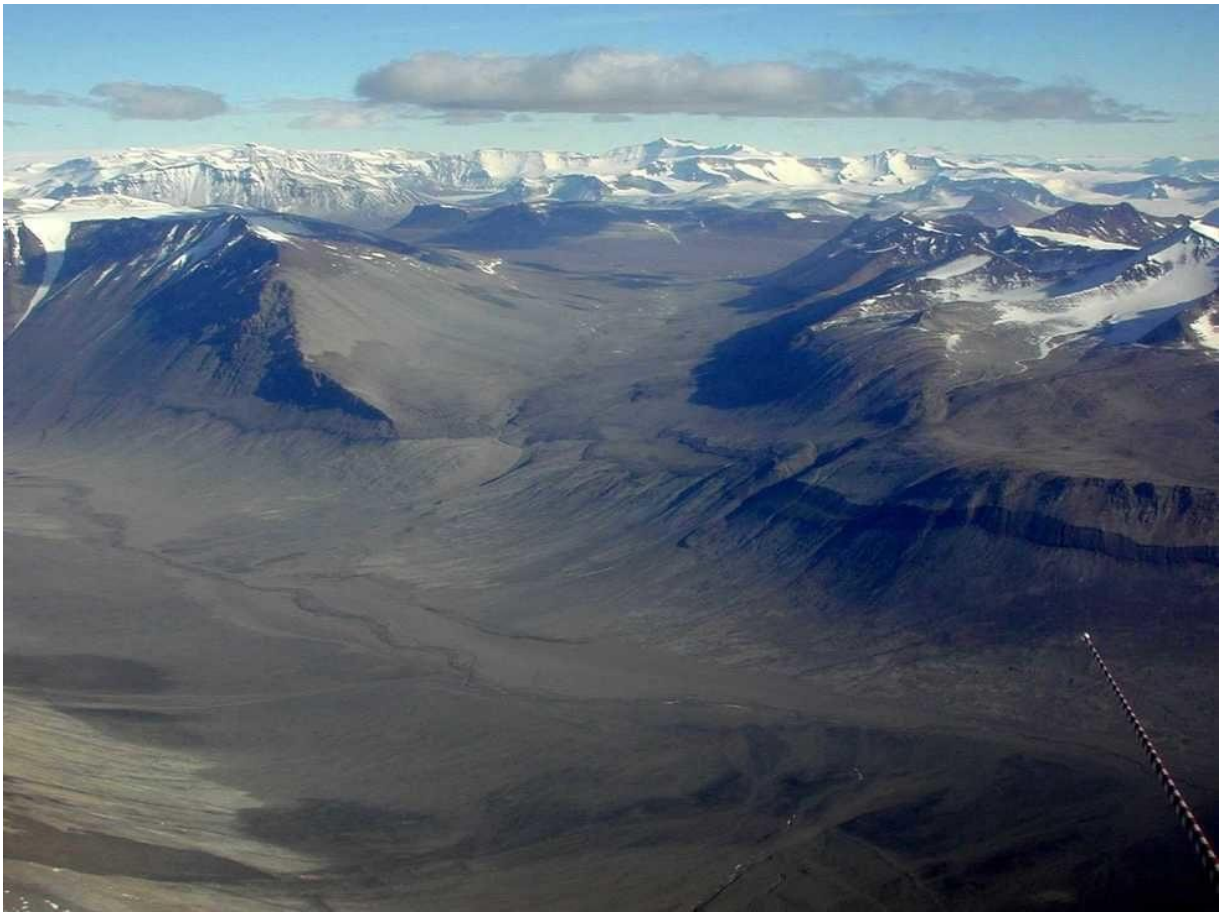
The combination of aridity, high UV, and unfavourable soils has made Atacama a popular region for astrobiological studies primarily investigating how life manages to persist in this hostile environment, as well as field testing robotic mission equipment such as landers and rovers, most recently the specialised GC-MS instrumentation designed for use on the MSL Curiosity Rover (Buch et al, 2009).

Despite the harsh environment, life does persist in Atacama. Endolithic communities such as *Cyanobacteria* species colonise the pores and spaces in rocks for protection against the harsh solar radiation (Weirzchos et al, 2006), communities have also been observed in quartz deposits in Atacama (Lacap et al, 2011). In addition to this, periodic rain events in the nearby mountainous regions are capable of bringing on striking floral blooms across large sections of the desert.

The Atacama is used as a Mars analogue due primarily to the extreme aridity of the soils, and high UV radiation in the region, similarities shared with Mars, and is an invaluable region for examining the so-called ‘dry limit of life’.

2.2 McMurdo Dry Valleys

The McMurdo Dry Valleys of the Antarctic are an arid region of the largely snow-covered Antarctic landmass. Unlike the majority of Antarctica, McMurdo is a frozen desert as arid as any hot tropical desert. Covering an area approximately 4800km^2, the dry valleys have long been the focus of attention of astrobiologists such as Dr Wolf Vishniac, who lost his life during a field expedition while testing a biological detection experiment originally designed for use on the Mars Viking landers, which was among the first of its kind developed.



Fig[3]: The McMurdo Dry Valleys, Antarctica. Image Credit: United States Antarctic Program (Public Domain).

The dry valleys offer a valuable environment for use as a Mars analogue region due primarily to the extreme cold and aridity, although the site does additionally receive moderately high solar flux due to altitude – McMurdo rests at around 2000m above sea level, comparable to

Atacama. Average annual temperatures are well below the freezing point of water, and have been reported as low as -30^{o}C (Doran et al, 2002), which are somewhat comparable to the Martian summer at lower latitudes (Ulrich et al, 2008), resulting in fairly comparable environmental challenges to life, although perchlorate-rich soils, UVC wavelength and cosmic radiation profile, and atmospheric differences obviously cannot be compared between McMurdo and Mars. Despite the harsh conditions, life has been observed in the McMurdo Dry Valleys; similarly to Atacama, endolithic communities are present in McMurdo (de los Rios et al, 2014).

3.0: The Chilean Altiplano

The South American Altiplano – literally "High Plain" is a high altitude region covering much of the central and western regions of the continent, contained within the borders of Martian, Bolivia, Argentina, and Peru. With average altitudes on the Altiplano being around 3750m above sea level (m.a.s.l.), the region features typically higher average solar flux than lower lying regions due to the relative lower atmospheric pressure experienced at higher altitudes. Much of the Altiplano features prominent peaks and volcanoes of the Andean range. Active volcanism is not particularly prominent in the Altiplano, with the majority of the volcanoes being dormant in present times.

3.1 Locations examined in this study.

The present study focuses on several locations in the Chilean Altiplano, around the Sairecabur volcanic range which spans the Chilean-Bolivian border in the northern regions of Martian. The majority of field sites chosen lie at varying altitude on the slopes of Volcan Cerro Sairecabur, located around the coordinates: S 22.72066389, W 67.92194444, between 4200 and 5500m.a.s.l. A further field site situated in the nearby Atacama Desert at coordinates: S 22.96176111, W 68.41027778, at 2444m.a.s.l. was additionally utilised. The location of the primary field site location is shown in Fig[3].

Fig[4]: Location of the primary field site of this study, at the coordinates S 22.72066389, W 67.92194444 . Source: Google Maps. Field expeditions were undertaken during the early austral autumn (April-May) in 2014 and 2015.

3.2 Comparing Altiplano with other Mars analogues

The rationale of choice of field site for the present study is based on environmental conditions in the region. When comparing and contrasting this region with other sites, the key features to be considered are aridity, temperature, and solar flux (in particular solar UV radiation). Clearly, the ideal analogue of the Martian surface must feature extremely low temperatures, high aridity, and extreme solar flux. Examples of differences between analogue sites are shown in Table[1].

Analogue Site	Environmental Feature		
	Arid	Cold	UV
Atacama	xxx	x	xx
McMurdo	xx	xxx	x
Sairecabur	xx	xx	xxx

Table[1]: Displaying the relative differences between two widely-reported Mars analogue regions, as well as the novel Altiplano region utilized in the present study. Scores are representative of relative intensity of each environmental feature.

The 'scores' assigned in Table[1] are designed to represent the relative extent to which the different analogue regions are effected by Mars-like conditions – relative in this case to each other and not to the Earth as a whole. Atacama is rated xxx for aridity and xx for UV intensity, with only a single x for coldness. As one of the most arid regions on the planet, the xxx aridity rating is well earned; a 2003 study by McKay et al examined collected data over 4 years from Atacama, specifically in the Antofagasta region, and recorded only 5 rain events over the recording period – 4 of which were 0.3mm or less, and may well be explained by morning dew rather than rainfall (McKay et al, 2003). The high UV intensity in Atacama is

related to the low latitude and relatively high altitude of the region (around 2400masl). Indeed, the primary difference between the UV scores for Atacama and the Altiplano sites are due to the altitude differnces between the regions, with the Altiplano sites lying at considerably greater elevation. Atacama does not receive a high score for cold temperatures, as during the daytime, temperatures typically peak at around 30^{o}C in summer months – Atacama is not a particularly hot desert, but it is significantly warmer than the surface of Mars. However, Atacama does, like many deserts, feature lower temperatures during the night, approaching 0^{o}C some nights, with periodic dips below freezing during winter months (McKay et al, 2003), and thus the single x score is representative of this.

The McMurdo Dry Valleys in Antarctica offer low temperatures, high aridity, and moderate solar flux, but not the extensive irradiance sought. The high altitude Altiplano region selected for this study offers low temperatures and extreme solar radiation, and moderate aridity at higher altitudes. The extensive solar irradiation is a key feature in this study and a large factor in choice of sites.

4.0: Aims of this study

This study aims to investigate the effect of increasing soil coverage on the presence of molecular signatures of life in a Mars analogue environment, with the purpose of examining the potential of such soil coverage offering protection to signs of life under such inhospitable conditions, when compared to the exposed surface itself.

In investigating the effect of soil coverage, particular focus will be given to the ability to recover DNA and lipid content from the soils of the Mars analogue environment, in the form of Whole Genomic Content and Total Lipid Content extracts, respectively. In addition, such genomic content as is able to be extracted will be utilised in a large-scale DNA sequencing project, for the purpose of examining the nature of the bacterial communities of these soils, including any change in bacterial community as soil depth increases.

Soils will be collected from a range of sites in the Chilean Altiplano region as described previously, encompassing a variety of altitudes and vegetation states, with a primary site of interest being selected following environmental measurements to determine the most suitable site. From each site, soils will be sampled at increasing soil depth below the surface (as well

as the exposed surface itself), with samples being taken at regular and frequent soil depth intervals to allow for the examination of depth change at high resolution – a methodology which is yet to be applied to studies in such environments. In particular, focus in prior studies in this region of the Altiplano has been on nearby geothermal springs (Ruff and Farmer, 2016), salt flats (Dorador et al, 2008), and high-altitude lakes (Alanoca et al, 2016), with little focus on the arid volcanic slopes such as those to be investigated in this study.

Complementary to the collection of soil samples, environmental conditions will be recorded at each field site in the region. To facilitate examination of key environmental features necessary for a viable Mars analogue, focus will be on temperature of the air and soils, moisture content of the air and soils, and intensity of solar flux. Soil-based environmental recordings will account for increasing soil depth, to allow correlation with extract yields of DNA and lipid content.

It is hypothesised that as soil depth increases, overall organic recoverable content will also increase. This increase in expected yield is due to the conditions of the surface environment being somewhat buffered by depth, as the effect of solar heating and evaporative loss (among others) become reduced; solar heating alone is expected to be attenuated substantially by soil depth, given that UV radiation is absorbed by only a few mm of soil.

5.0: References

Alanoca, L., Amouroux, D., Monperrus, M., Tessier, E., Goni, M., Guyoneaud, R., Acha, D., Gassie, C., Audry, S., Garcia, M.E., Quintanilla, J., Point, D. (2016). Diurnal variability and biogeochemical reactivity of mercury species in an extreme high-altitude lake ecosystem of the Bolivian Altiplano. Environmental Science and Pollution Research Vol 23. P6919-6933.

Allen, C.C., and Oehler, D.Z. (2008). A Case for Ancient Springs in Arabia Terra, Mars. Astrobiology Vol 8. P1093-1112.

Ashbrook, J. (1953). A new determination of the rotation period of the planet Mars. The Astronomical Journal Vol 58. P145-155.

Atrih, A., and Foster, S.J. (2002). Bacterial endospores the ultimate survivors. International Dairy Journal Vol 12. P217-223.

Beauchamp, S., and Lacroix, M. (2012). Resistance of the genome of *Escherichia coli* and *Listeria monocytogenes* to irradiation evaluated by the induction of cyclobutane pyrimidine dimers and 6-4 photoproducts using gamma and UV-C radiations. Radiation Physics and Chemistry Vol 81. P1193-1197.

Brakenridge, G.R., Newsom, H.E., Baker, V.R. (1985). Ancient hot springs on Mars: Origins and paleoenvironmental significance of small Martian valleys. Geology Vol 13. P859-862.

Buch, A., Sternberg, R., Szopa, C., Freissinet, C., Garnier, C., Bekri, El J., Rodier, C., Navarro-González, R., Raulin, F., Cabane, M., Stambouli, M., Glavin, D.P., Mahaffy, P.R. (2009). Development of a gas chromatography compatible Sample Processing System (SPS) for the in-situ analysis of refractory organic matter in Martian soil: preliminary results. Advances in Space Research Vol 43. P143-151.

Cantor, B.A. (2006). MOC observations of the 2001 Mars planet-encircling dust storm. Icarus Vol 186. P60-96.

Craddock, R.A. and Howard, A.D, (2002). The case for rainfall on a warm, wet early Mars. Journal of Geophysical Research Vol 107.

Dittmann, C., Han, H-M., Grabenbauer, M., Laue, M. (2015). Dormant *Bascillus* spores protect their DNA in crystalline nucleoids against environmental stress. Journal of Structural biology Vol 191. P156-164.

Dorador, C., Busekow, A., Vila, I., Imhoff, J.F., Witzel, K-P. (2008). Molecular analysis of enrichment cultures of ammonia oxidizers from the Salar de Huasco, a high altitude saline wetland in northern Chile. Extremophiles Vol 12. P405-414.

Doran, P.T., McKay, C.P., Clow, G.D., Dana, G.L., Fountain, A.G., Nylen, T., Lyons, W.B. (2002). Valley floor climate observations from the McMurdo dry valleys, Antarctica, 1986-2000. Journal of Geophysical Research Vol 107.

Edgett, K.S. (1997). Aeolian Dunes as Evidence for Explosive Volcanism in the Tharsis Region of Mars. Icarus Vol 130. P96-114.

Fassett, C.I., Dickson, J.S., Head, J.W., Levy, J.S., Marchant, D.R. (2010). Supraglacial and proglacial valleys on Amazonian Mars. Icarus Vol 208. P86-100.

Guo, J., Zeitlin, C., Wimmer-Schweingruber, R.F., Hassler, D.M., Ehresmann, B., Köhler, J., Böhm, E., Böttcher, S., Brinza, D., Burmeister, S., Cucinotta, F., Martin, C., Posner, A., Rafkin, S., Reitz, G. (2015). MSL-RAD RADIATION ENVIRONMENT MEASUREMENTS. Radiatio Protection Dosimetry Vol 166. P290-294.

Haberle, R.M., McKay, C.P., Schaeffer, J., Cabrol, N.A., Grin, E.A., Zent, A.P., Quinn, R. (2001). On the possibility of liquid water on present-day Mars. Journal of Geophysical Research Vol 106. P23317-23326.

He, Y-Y., and Häder, D-P. (2002). UV-B-induced formation of reactive oxygen species and oxidative damage of the cyanobacterium *Anabena sp.*: protective effects of ascorbic acid and *N*-acetyl-L-cysteine. Journal of Photochemistry and Photobiology Vol 66. P115-124.

Hecht, M.H., Kounaves, S.P., Quinn, R.C., West, S.J., Young, S.M.M., Ming, D.W., Catling, D.C., Clark, B.C., Boynton, W.V., Hoffman, J., DeFlores, L.P., Gospodinova, K., Kapit, J., Smith, P.H. (2009). Detection of Perchlorate and the Soluble Chemistry of Martial Soil at the Phoenix Lander Site. Science Vol 325. P64-67.

Hynek, B.M., Phillips, R.J., Arvidson, R.E. (2003). Explosive volcanism in the Tharsis region: Global evidence in the Martian geologic record. Journal of Geophysical Research Vol 108.

Iwabuchi, K., Hidema, J., Tamura, K., Takagi, S., Hara-Nishimura, I. (2016). Plant Nuclei Move to Escape Ultraviolet-Induced DNA Damage and Cell Death. Plant Physiology Vol 170. P678-685.

Jakosky, B.M., and Phillips, R.J. (2001). Mars' volatile and climate history. Nature Vol 412. P237-244.

Jensen, L.L., Merrison, J., Hansen, A.A., Mikkelsen, K.A., Kristoffersen, T., Nørnberg, P., Lomstein, B.A., Finster, K. (2008). A Facility for Long-Term Mars Simulation Experiments: The Mars Environmental Simulation Chamber (MESCH). Astrobiology Vol 8. P537-548.

Kahn, R. (1990). Ice Haze, Snow, and the Mars Water Cycle. Journal of Geophysical Research Vol 95. P14677-14693.

Kim, S-J., Kim, D-K., Kang, D-H. (2016). Using UVC Light-Emitting Diodes at Wavelengths of 266 to 279 Nanometers To Inactivate Foodborne Pathogens and Pasteurize Sliced Cheese. Applied and Environmental Microbiology Vol 82. P11-17.

Kurokawa, H., Kurosawa, K., Usui, T. (2018*). A lower limit of atmospheric pressure on early Mars inferred from nitrogen and argon isotopic compositions. Icarus Vol 299. P443-459.

Lacap, D.C., Warren-Rhodes, K.A., McKay, C.P., Pointing, S.B. (2011). Cyanobacteria and Chloroflexi-dominated hypolithic colonization of quartz at the hyper-arid core of the Atacama Desert, Chile. Extremophiles Vol 15. P31-38.

Lasne, J., Noblet, A., Szopa, C., Navarro-González, R., Cabane, M., Poch, O., Stalport, F., Francois, P., Atreya, S.K., Coll, P. (2016). Oxidants at the Surface of Mars: A Review in Light of Recent Exploration Results. Astrobiology Vol 16. P977-996.

Lee, Y.W., Yoon, H.D., Park, J-H., Ryu, U-C. (2018). Application of 265-nm UVC LED Lighting to Sterilization of Typical Gram Negative and Positive Bacteria. Journal of the Korean Physical Society Vol 72. P1174-1178.

Levin, G.V., and Straat, P.A. (1977). Life on Mars? The Viking Labeled Release Experiment. Biosystems Vol 9. P165-174.

Martin, T.Z., and Bridges, N.T. (2003). Near-surface temperatures at proposed Mars Exploration Rover landing sites. Journal of Geophysical Research Vol 108.

McEwen, A.S., Ojha, L., Dundas, C.M., Mattson, S.S., Byrne, S., Wray, J.J., Cull, S.C., Murchie, S.L., Thomas, N., Gulick, V. (2011). Seasonal Flows on Warm Martian Slopes. Science Vol 333. P740-743.

McKay, C.P., Friedmann, E.I., Gómez-Silva, B., Cáceres-Villanueva, L., Andersen, D.T., Landheim, R. (2003). Temperature and Moisture Conditions for Life in the Extreme Arid Region of the Atacama Desert: Four Years of Observations Including the El Niño of 1997-1998. Astrobiology Vol 3. P393-406.

Mohr, H., Gravemann, U., Bayer, A., Müller, T.H. (2009). Sterilization of platelet concentrates at production scale by irradiation with short-wave ultraviolet light. Transfusion Vol 49. P1956-1963.

Mulholland, D.P., Read, P.L., Lewis, S.R. (2012). Simulating the interannual variability of major dust storms on Mars using variable lifting thresholds. Icarus Vol 223. P344-358.

Nair, S.G., and Loppnow, G.R. (2013). Multiplexed, UVC-Induced, Sequence-Dependent DNA Damage Detection. Photochemistry and Photobiology Vol 89. P884-890.

Nasrini, J., Dinges, D.F., McGuire, S., Hermosillo, E., Ecker, A.J., Mollicone, D.J., Mott, C.G., Binsted, K., Caldwell, B.J., Moore, T.M., Gur, R.C., Basner, M. (2017). Cognitive performance in long-duration Mars simulations at the Hawaii Space Exploration Analog and Simulation (HI-SEAS). Conference paper: 2017 NASA Human Research Program Investigators' Workshop, At Galveston TX, USA.

Nhung, L.T.T., Nagata, H., Takahashi, A., Aihara, M., Okamoto, T., Shimohata, T., Mawatari, K., Akutagawa, M., Kinouchi, Y., Haraguchi, M. (2012). Sterilization effect of UV light on *Bascillus* spores using TiO_2 films depends on wavelength. The Journal of Medical Investigation Vol 59. P53-58.

Quinn, R.C., Martucci, H.F.H., Miller, S.R., Bryson, C.E., Grunthaner, FJ., Grunthaner, P.J. (2013). Perchlorate Radiolysis on Mars and the Origin of Martian Soil Reactivity. Astrobiology Vol 13. P515-520.

Ravanat, J-L., and Douki, T. (2016). UV and ionizing radiations induced DNA damage, differences and similarities. Radiation Physics and Chemistry Vol 128. P92-102.

de los Rio, A., Wierzchos, J., Ascaso, C. (2014). The lithic microbial ecosystems of Antarctica's McMurdo Dry Valleys. Antarctic Science Vol 26. P459-477.

Roberts, J.H., and Arkani-Hamed, J. (2017). Effects of basin-forming impacts on the thermal evolution and magnetic field of Mars. Earth and Planetary Science Letters Vol 478. P192-202.

Rontó, G., Bérces, A., Lammer, H., Cockell, C.S., Molina-Cuberos, G.J., Patel, M.R., Selsis, F. (2002). Solar UV Conditions on the Surface of Mars. Photochemistry and Photobiology Vol 77. P34-40.

Ruff, S.W., and Farmer, J.D. (2016). Silica deposits on Mars with features resembling hot spring biosignatures at El Tatio in Chile. Nature Communications Vol 17.

Sagan, C., and Dermott, S.F. (1982). The tide in the seas of Titan. Nature Vol 300. P731-733.

Schuerger, A.C., and Clark, B.C. (2008). Viking Biology Experiments: Lessons Learned and the Role of Ecology in Future Mars Life-Detection Experiments. Space Science Reviews Vol 135. P233-243.

Seiff, A., and Kirk, D.B. (1977). Structure of the Atmosphere of Mars in Summer at Mid-Latitudes. Journal of Geophysical Research Vol 82. P4364-4378.

Tian, B., Wang, H., Ma, X., Hu, Y., Sun, Z., Shen, S., Wang, F., Hua, Y. (2010). Proteomic analysis of membrane proteins from a radioresistant and moderate thermophilic bacterium *Deinococcus geothermalis*. Molecular Biosystems Vol 6. P2068-2077.

Uehara, F., Miwa, S., Tome, Y., Hiroshima, Y., Yano, S., Yamamoto, M., Efimova, E., Matsumoto, Y., Maehara, H., Bouvet, M., Kanaya, F., Hoffman, R.M. (2014). Comparison of UVB and UVC Effects on the DNA Damage-Response Protein 53BP1 in Human Pancreatic Cancer. Journal of Cellular Biochemistry Vol 115. P1724-1728.

Ulrich, R., Pilgrim, R., Chevrier, V., Roe, L., Kral, T. (2008). Temperature fields at Mars landing sites: Implications for subsurface. Lunar and Planetary Sciences Vol 39.

Wadsworth, J., and Cockell, C.S. (2017). Perchlorates on Mars enhance the bacteriocidal effects of UV light. Nature Science Reports Vol 7.

Webster, C.R., Mahaffy, P.R., Atreya, S.K., Flesch, G.J., Mischna, M.A., Meslin, P-Y., Farley, K.A., Conrad, P.G., Christensen, L.E., Pavlov, A.A., Martin-Torres, J., Zorzano, M-P., McConnochie, T.H., Owen, T., Eigenbrode, J.L., Glavin, D.P., Steele, A., Malespin, C.A., Archer Jr, P.D., Sutter, B., Coll, P., Freissinet, C., McKay, C.P., Moores, J.E., Schwenzer, S.P., Bridges, J.C., Navarro-González, R., Gellert, R., Lemmon, M.T., the MSL Science Team. (2015). Science Vol 347. P415-417.

Wierzchos, J., Ascaso, C., McKay, C.P. (2006). Endolithic *Cyanobacteria* in halite rocks from the hyperarid core of the Atacama Desert. Astrobiology Vol 6. P415-422.

Wordsworth, R. ,Forget, F., Millour, E., Head, J.W., Madeleine, J-B., Charney, B. (2012). Global modelling of the early Martian climate under a denser CO_2 atmosphere: Water cycle and ice evolution. Icarus Vol 222. P1-19.

Zhao, X., Zhou, P., Chen, X., Li, X., Ding, L. (2014). Perchlorate-induced oxidative stress in isolated liver mitochondria. Ecotoxicology Vol 23. P1846-1853.

Websites referenced:

Fig[1]: NASA MSL Curiosity Rover photograph https://www.nasa.gov/sites/default/files/styles/full_width_feature/public/thumbnails/image/pia21041.jpg

Fig[3]: McMurdo Dry Valleys image: https://photolibrary.usap.gov/PhotoDetails.aspx?filename=Bullpa1.jpg

Chapter 1: Environmental Data and Genomics

1.0: Introduction

1.1: The Surface of Mars

The surface of Mars today features environmental conditions which are extremely hostile to the life. Present-day Mars has very low air temperatures, a lack of stability of liquid water, extensive solar and cosmic irradiation, and oxidising soil chemistry. Several of these features are in part due to the sparse atmosphere (Leovy, 2001) which offers little protection from solar UV radiation (including the highly-destructive UVC wavelengths which are entirely attenuated by the atmosphere on Earth), and does not offer a significant greenhouse effect which would aid retention of heat at the surface. Mars also lacks any substantial planetary magnetic field, possibly due to the lack of a rotating metallic core such as the iron core of Earth. The lack of a magnetic field leaves the surface of the planet vulnerable to high-energy cosmic radiation which is destructive to organic material.

1.1.1: Air and Soil Temperature

The combination of thin atmosphere (average atmospheric pressure is around 6mbar Jakosky and Philips, 2001) and distance from the Sun contribute to average air temperatures at the surface of Mars being extremely low in comparison to Earth. Additionally, variation in the concentration of airborne dust particles has a strong impact on the near-surface temperature – when the suspended dust coverage is thick, daytime solar heating is less effective (Leovy, 2001; Gierasch and Goody, 1972). There is a substantial diurnal variation in temperature, but even during the daytime at perihelion, average air temperatures are still low. The effect of such a substantial diurnal variation at the soil surface – a difference as great as 90K has been recorded at the surface of the landing side of the NASA Mars Pathfinder mission, with a low of around 190K and a high of around 260K (Ulrich et al, 2008) – would be extremely taxing on living organisms present there; while adaptation to extremes of temperature are well-known, adapting to such a range has not been observed.

1.1.2: Water

Liquid water is not stable on the Martian surface, and no bulk deposits of water such as lakes or oceans are present, although evidence for the past existence of such bodies of water is evident from large fluvial features such as Valles Marineris (Rodriguez et al, 2015). The Martian atmosphere holds some water vapour, but in insufficient amounts for precipitation to occur. Typical atmospheric partial pressure of water vapour is 10^{-3}mbar, which is around 10^{4} times less than that of the Earth atmosphere (Jakosky and Philips, 2001). The greatest apparent concentration of H_2O on Mars is at the polar ice caps, and it is believed that this ice is not pure water ice but saline brine ices. Orbital analyses have determined that subsurface ice deposits may be present at lower latitudes (Mellon et al, 2004), and these are believed to also be brine ices. Recent studies have reported evidence of what appears to be periodic flows of liquid water down slopes at the surface in lower latitudes, termed Recurring Slope Lineae (RSL) (McEwen et al, 2011), and it is theorised that these are the result of near-surface brine ice melting.

1.1.3: Solar Radiation

The Martian atmosphere is substantially less dense than that of the Earth, approximately 7mbar (Bryson et al, 2008), 0.7% the density of sea-level air pressure on Earth (sea level is approximately 1bar). With the absence of a protective Ozone layer, damaging UV radiation is not attenuated and the surface is exposed to intense UV radiation despite the increased distance from the sun when compared to Earth. This extensive solar radiation heats the soil surface in excess of the ambient air temperatures during the daytime (Ulrich et al, 2008), although UV radiation does not penetrate far into soil (Mancinelli and Klovstad, 2000) and so is not likely to be the principal candidate for the cause of RSL.

1.1.4: Cosmic Radiation

Mars does not have a strong protective magnetic field (Gringauz, 1991), and so the surface is not protected from high energy cosmic and ionising radiation. In addition, the lack of a magnetosphere does not protect the atmosphere from the effect of solar wind, which may

contribute to the lack of dense atmosphere and consequently to the lack of stability of liquid water at the surface (Lundin, 2001). High energy and ionising radiation are known to be extremely damaging to organic molecules, especially DNA (Siam et al, 2018), which remains a source of concern in manned space flight. This damage from high energy radiation is another contributing factor to the apparent sterility of the surface of Mars.

1.2: Use of Terrestrial Analogues

Terrestrial Martian analogues are regions on Earth which exhibit one or more of the environmental factors reported on the surface of Mars. Typically, these take the form of cold and/or arid regions which are relatively free of life compared to, for example, a temperate forest or savannah. Martian analogues present a number of opportunities for study; investigation of life at the limits of what is currently considered survivable, determination of how precisely such extreme environments present a challenge to life and how these challenges are met, and field testing equipment designed to operate on Mars for life detection missions. Few environments on Earth are viable as Mars analogues, and due to the dense atmosphere it is impossible to achieve 100% accurate simulations in the field, nevertheless, these extreme environments serve a valued purpose.

1.2.1: Atacama Desert, Chile

The Atacama Desert is a high altitude (approximately 2000 metres above sea level (masl)) desert region in northern Chile, it is commonly stated to be the most arid location on Earth, and although it does in fact periodically receive water from rare rainfall events, causing brief periods of floral bloom (this happened as recently as 2016). Due to the relative low latitude and high altitude location of the region, the daylight hours feature extensive solar radiation, and high air and soil surface temperatures.

1.2.3: Geysers Del Tatio, Chile

Phoenix et al, (2006) utilised the El Tatio hot spring system, located approximately 40km from the region used in this study and at around 4300masl, as an analogue for the Precambrian oceans. The high-altitude geyser field was proposed as a suitable modern day analogue for the early Precambrian shallow marine environments due to a combination of water chemistry (solute enriched, in particular silica), biology (dominated by microbial life forms) and, in particular, the high UV flux experienced at this altitude.

1.2.4: McMurdo Dry Valleys, Antarctica

The McMurdo Dry Valleys are a region in Antarctica at approximately (77.28S, 162.31E), which are uniquely arid compared to the surrounding regions which frequently receive snow coverage. The dry valleys feature extreme aridity and low temperatures year-round, and thus the continual cold temperatures here present a useful analogue of the Martian surface.

1.3: Microbiology of Extreme Environments

Environments which feature extreme conditions such as extremes of temperature (high or low), hyperaridity, irradiation, extremes of pH (acidity or alkalinity), or over-underabundance of common nutrients, demand specific adaptation by life if it is to persist and thrive in these conditions.

1.3.1: Solar Radiation

Extreme solar irradiation from is a common challenge at high altitude on Earth, where the atmosphere is comparatively less dense (compared to sea level) and thus offers reduced protection from the incoming solar radiation.

Damage from radiation poses a threat to life due to the generation of Reactive Oxygen Species inside cells, which are potentially damaging to proteins, lipids, and nucleic acids (Du and Gabicky, 2004). There are 2 mitigation strategies for solar radiation: avoidance or resistance. Avoiding solar radiation is relatively straightforward – UV light does not effectively penetrate solid material, only 1-3mm in soil, for example. In Atacama, endolithic communities survive the intense UV by growing inside the spaces in porous rocks, and just under the surface of translucent minerals (Wierzchos et al, 2015).

Resistance to radiation typically is achieved not by preventing the damage from occurring, but by the ability to rapidly repair any damage which does occur, so that cellular function is not compromised. In general, more complex organisms are more vulnerable to damage from radiation. This is illustrated by the fact that the average fatal radiation dose for a human is 10 Gy, while the radioresistant bacteria *Deinococcus radiodurans* is capable of surviving a single exposure of up to 15,000Gy (Daly et al, 1994). In *D.radiodurans*, DNA repair proteins are highly upregulated, and the species itself contains unique additional DNA repair proteins (Munteanu et al, 2015)

1.3.2: Hyperaridity and Hypersalinity

Water is the principal solvent used by life on Earth, therefore any environment which is extremely arid presents a challenge to survival of life. A hyperarid or saline extracellular environment will rapidly lead to desiccation of cells through osmotic loss of water from the cell. Hypersaline environments may be either arid (salt flats such as Salar de Atacama) or marine/lacustrine environments such as the Dead Sea.

Desiccation by the environment present a similar challenge to damage from radiation: the generation of reactive oxygen species which damage cellular proteins, lipids, and nucleic acids. The largest deserts on earth are barren environments with little to no apparent vegetation, yet microbiological life persists.

Survival strategies in hyperarid conditions in larger organisms often focus on conserving what water the organism is able to take in – cacti are a clear example here. Microbiological life, however, persists through the rapid repair mechanisms discussed previously.

1.4: The Chilean Altiplano

The South American Altiplano is an expanse of deserts and plains which remain at a comparatively high altitude above sea level (4000-6000masl), covering much of the western regions of the continent, primarily in Bolivia, Peru and Chile. Here we focus on the northern Chilean Altiplano, specifically in the Antofagasta region, adjacent to the Atacama Desert. Field sites are situated around 22.72S, 67.91W.

1.4.1: Environmental Conditions

The region of interest in this study lies in a position of high altitude and low latitude, which results in a generally arid climate. Indeed, the nearby Atacama Desert (2000masl) is widely regarded as the most arid environment on Earth. However, the increased altitude of the Altiplano removes the common desert feature of high daytime temperatures – the high plains studied here experience comparatively low air temperatures. The Altiplano is not an entirely barren expanse, and does receive rainfall, notably during summer months as result of the Bolivian Winter phenomenon in which moist air from the Amazon rainforest regions is blown west over the Andes mountain range, and results in heavy rainfall for a short period of time. A critical feature experienced in the Altiplano is comparatively high Solar Flux – a consequence of the high altitude is reduced air pressure and thus a reduction in atmospheric protection from solar radiation; and UV irradiance across the Altiplano is extensive, as exemplified in a study reported by Cabrol et al (2014) which reported a new record high UV measurement on Earth upon Volcan Licancabur (22.50S, 67.52W). In northern Chile, the global flux of UVA at 4200m and 5500m is approximately 26% and 37% greater than at sea level respectively, whilst UVB is approximately 35% and 43% greater (Piazena, 1996). Furthermore, the global of UVA at 4200m and 5500m 12% and 18% greater than the nearby Atacama Desert, while UVB is approximately 15% and 22% greater. The flux of Photosynthetically-active radiation (PAR, 400-700nm) is due to the low latitude of the

location and propensity for clear, cloudless skies; PAR peaks at around 2600 µmoles m^{-2} at midday in the summer. (Phoenix et al, 2006).

1.4.2: Uniqueness as an Analogue Region

The Altiplano region presented in this study provide a unique and exciting Martian Analogue environment. The location provides a combination of environmental factors which reflect conditions reported on Mars, i.e. the Altiplano provides an environment that is both cold, and experiences high flux of solar radiation. This combination is underrepresented in alternative analogue regions such as Atacama and McMurdo; For example, the Atacama experiences high solar radiation but also warm temperatures. In contrast, McMurdo, meanwhile, is cold but does not experience high flux of solar radiation. It is important to note, however, that both regions experience on average greater year-round aridity than the Altiplano region described herein, which is an important consideration when discussing Mars.

1.5: The Present Study

This study aims to utilise the unique Martian analogue region in the Chilean Altiplano to investigate the role of soil depth coverage in protecting organic material and/or extant life against the hostile surface conditions. By determining the extent to which protection is provided by soil depth, which enables survival of life or evidenced of extinct life, this use of an analogue region may assist in formulating strategies for similar analyses carried out on the Martian surface in the future. ExoMars 2020 is particularly relevant as it will be able to to drill down through the soil to sample at different depths in its search for signs of life.

1.5.1: A Novel Approach

This study will approach the investigation of the effect of soil depth on the buffering of hostile surface conditions through the use of high-resolution selection of soil depths to record conditions at, and to sample from. Utilising a range of sites in the analogue region of increasing altitude, at each site environmental conditions will be recorded both in the air, soil surface, and at increasing depth below the soil surface. Such high-resolution study in an

analogue site is at present underreported, and this will serve as a benchmark for the efficacy of such a strategy for investigation. As the experimental strategy is novel, equally so is the selection of the analogue region itself, with these high altitude, arid, sun-baked soils being underreported in the literature in contrast to the high altitude lakes, hypersaline environments, and hot springs present around the region at large.

2.0: Methods

2.1: Environmental Data Collection

Air temperature and humidity were recorded with the use of a CS215 Air Temperature and Relative Humidity probe attached to a Campbell Scientific CR1000 datalogger, which was installed at the 5056m site for three days. The 5056m site was chosen for the datalogger as this is the primary site of astrobiological interest as a unique Martian analogue site. Air temperature was additionally recorded using a handheld temperature probe, recorded at approximately midday on the same day as soil sampling and other handheld readings were taken. As with air temperature, air relative humidity was additionally recorded using a handheld relative humidity probe, taken at approximately midday at each site.

Soil temperature and moisture was recorded in a similar manner to air temperature and humidity. Soil temperature was measured using C109 temperature probes attached to a CR1000 datalogger, and via a handheld temperature probe. Soil temperature was additionally recorded via the use of iButton micro loggers, which were installed in the soil at the surface (both exposed to the sun and in shade), 1cm, and then at 5cm depth increments up to 25cm below the surface. To record temperatures of the very soil surface, thus recording the impact of direct solar heating, a FLIR thermal imaging camera was used. For this, a shovel was used to scrape away the top 7cm of soil enabling an image of the soil surface around the trench and the soil in the bottom of the trench to be recorded in one image for direct comparison. This was done to provide further insight into the impact of solar radiation on heating of the soil surface compared to the soil at depth.

Soil moisture was recorded using soil volumetric water content reflectometers, attached to a CR1000 datalogger, with the probes buried over a range of depths. Soil moisture was additionally recorded using a handheld soil moisture probe, with the handheld readings being taken at the same depths from which soil samples were excavated for the purpose of correlating with soil extract genomics data.

The iButton microloggers used to record soil temperature additionally record relative humidity. Humidity is not a direct proxy for soil moisture, but the two values are linked.

A further environmental factor of interest to the present study is the incoming solar flux which affects the area. Recordings of solar flux in the UVA, UVB, and Photosynthetically-Active Radiation (PAR, also known as the Visible Spectrum) were taken at each site using a Macam Photometrics UV203 meter (with UVA and UVB filters), and a Macam Photometrics Q203 meter (PAR) at approximately midday, when the solar intensity is near peak. Solar flux is of particular interest as it is believed to heat the soil surface.

2.2: Soil sampling strategy

Soil samples were collected over a range of depths at each sampling site. Samples were collected from 1cm-21cm below the soil surface, at 2cm intervals, with additional samples from 0cm and 37cm collected. 37cm depth samples were collected for the purpose of comparison with soil samples from pilot studies collected prior to the field expedition of 2015. Final total soil samples for each site of interest are shown in Table[1] below. Depth was measured using a tape measure against a flat surface on the soil surface (0cm). A wide trench was excavated using the tip of a shovel, with a gentle scraping motion used to reveal the thin layers of soil for each sequential sampling.

Soil samples were taken solely from the centre area of the wide trench at each depth, to prevent the possibility of contaminating samples with incorrect-depth soil due to trench sidewall slippage. In addition, samples were bagged in sterile whirlpak bags immediately following excavation, and subsequently double-bagged to further insure against contamination and comply with customs regulations.

For sample excavations, a small trowel was sterilised using 100% ethanol before and after each sampling, to minimise potential cross-contamination of samples. Soils were collected with a gentle scraping motion at the edge of the trowel, with sampling occurring at 2cm intervals to allow for the inherent margin of error in sampling relatively loose matter such as these dry soils, with care being taken to mitigate the risk of breaching into the subsequent depth sampling layer.

Site	Altitude (m.a.s.l)	Soil Sample Depths (cm)
1	4296	0, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 37
4	5056	0, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21

Table[1]: Soil depths sampled at each field site. Site 1 (4296m) represents the lower altitude, vegetated control site, while Site 4 (5056m) represents the primary site of interest. Sampling ceased at 21cm rather than 37cm at Site 4 due to the soil becoming permafrost which could not reliably and accurately be extracted using the methodology employed for other samples.

Use of collected samples in lab

Soil samples were split for different extraction procedures; samples to undergo whole genomic content extraction required no pre-protocol preparation, while samples to undergo total lipid content extraction were to undergo freeze-drying for 24 hours prior to extraction.

Genomic content extraction was performed on 1.5g of soil sample per extract, with overall DNA content values for each soil sample being presented as the mean of 3 replicate extracts from a single soil sample. With anticipation of the soils being low in organic content (compared to more temperate soils) due to the nature of the sampling sites, extracts were performed using fine particulate matter – avoiding the inclusion of large stones in the extraction sample as their presence would dramatically reduce the potential extract/mass recovered.

This range of sampling was designed to offer a high-resolution view of the differences observed in the soils over a relatively shallow depth profile below the hostile surface, with

the purpose of elucidating how even such a thin covering may buffer the mars-like conditions recorded sat the site. This high resolution investigation of the soil profile has not been reported in previous studies.

2.3: Whole Genomic Extract

DNA content was extracted as whole genomic content, with approximately 0.5g of soil used per extract, with the final recorded extract concentrations recorded as the mean of 3 extracts. Final extract concentration was normalised to account for variation in soil mass used for extraction. Sampled soils were not homogenous, and to maximise the efficiency of extraction, finer grains of soil were preferably selected for extract. Soil samples were processed using a FastDNA Spin Kit for Soil (MolBio). Initial DNA extract is achieved via bead lysis of cells, and subsequent purification of raw DNA from the extract.

2.4: Polymerase Chain Reaction

Extracted DNA was amplified by Polymerase Chain Reaction, to determine if the extracted nucleic acids were intact and not entirely fragmented by the process. The sequence of DNA to be amplified was the gene encoding the 16S Ribosomal RNA subunit. The 16S gene was chosen due to its nature as a highly conserved gene across bacteria, which exhibits enough inter-species variability to be invaluable for identification purposes. For the sequencing and identification of the species in the soil collected soil samples, the V4 region of the 16S gene was chosen for amplification, as this is known to be a region which offers a high degree of detection of variability between species (Vasileiadis et al, 2012). Extracts were successfully amplified with the exception of 2 samples from the surface (0cm and 1cm) from the highest altitude (5432m) site. The reason for the failure of these samples to be amplified may have been due to the exceptionally low DNA extract concentration of the samples, fragmentation of the nucleic acids during extraction, or insufficient optimisation of the PCR reaction. The

latter possibility seems unlikely given that the same optimisation conditions were successful for all other samples even in the same soil, at 3cm and below. Extracts were amplified using the following conditions and thermal cycle programme (Table [2]):

Reaction Mix (per sample)	Reagent (µl)	
Buffer	5	1X
dNTPs	0.75	0.3mM
F Primer	0.87	0.3µM
R Primer	0.87	0.3µM
DMSO*	1.25	
Taq Polymerase	0.5	0.5U
Template DNA	1	<100ng
PCR Water	14.76	n/a
TOTAL	25	
Programme	Temp (^{o}C)	Duration
Initial Denaturation	95	5m
Cycle: Denaturation	98	20s
Cycle: Annealing	60	20s
Cycle: Extension	72	20s
Final Extension	72	1m
Final Hold	4	n/a

Table[2]: Reaction mix and PCR thermal cycle programme for the amplification of bacterial 16S v4 DNA sequence. Here the reaction mix is for a 25µl total reaction volume. Cycle stage of programme consists of 25 repeat cycles before progressing onto final extension. Final Hold refers to the temperature the complete reactions are maintained at until storage at -20^{o}C. *DMSO variable, not added in all reactions.

2.5: Amplicon production

Sequencing of the 16S V4 region required specialised amplicons to be produced, with specific adapter regions ligated to the sequence of interest, to interact with the sequencing platform. In this study, the platform was the Illumina MiSeq platform, chosen over other platforms such as Ion Torrent or 454 Pyrosequencing for the balance of low processing time, low molecular size requirement, and high data output volume offered by Illumina, with output of >10,000 reads per sample being the recommended volume to allow for effective statistical reliability.

Production of 16S V4 amplicons suitable for Illumina MiSeq required the use of specialised PCR primers, and necessitates the purest possible DNA samples in nuclease-free water. The amplicons produced are comprised of the sequence of interest, a unique 'barcode sequence' which allows identification of individual amplicon sets amongst a pool of amplicons, and an Illumina Adapter region which is specifically designed to ligate to complementary sequences in the MiSeq instrument. The amplicons in this study were single barcoded – that is, barcode sequences were added to one strand of the dsDNA of the sample, but not the other, in this case the reverse strand was chosen for barcoding. The use of barcodes allows all amplified extracts to be added to a single sample pool, without the need for extensive libraries of samples all shipped individually-labelled. In this study, each unique barcode is associated with the extract from a single soil sample, and thus covers the entire community of bacteria in the associated soil sample from which DNA was recovered. The implications of this system are that comparisons can be made both between layers of soils in the same field site, and between the different sites.

Illumina-ready amplicons were produced via PCR, using the following conditions (Table [3]):

Reaction Mix (per sample)	Reagent (µl)	
Buffer	5	1X
dNTPs	0.75	0.3mM
F Primer	0.87	0.3µM
R Primer (barcoded)	0.87	0.3µM
DMSO*	1.25	
Taq Polymerase	0.5	0.5U
Template DNA	2	<100ng
PCR Water	13.76	n/a
Programme	Temp (^{o}C)	Duration
Initial Denaturation	95	5m
Cycle: Denaturation	98	20s
Cycle: Annealing	60	20s
Cycle: Extention	72	20s
Final Extention	72	1m
Final Hold	4	n/a

Table[3]: Reaction mix and PCR thermal cycle programme for production of final Illumina-ready amplicons. Here the volume of template DNA added to the reaction mix is doubled to 2µl. Cycle stage of programme consists of 25 repeat cycles before progressing onto final extension. Final Hold refers to the temperature the complete reactions are maintained at until storage at -20^{o}C. *DMSO variable, not added in all reactions.

Successful amplification using the custom Illumina primers required several rounds of testing and optimisation due to lower yield than observed in the previous reactions. Increasing the volume of added template DNA was found to increase yield most favourably.

2.6: Gel purification of amplicons

Successfully-produced amplicons were purified via agarose gel electrophoresis. Excision of the bands formed in the gel following electrophoresis allowed size selection of the desired samples from any other material in the sample which was remaining following amplification (such as any primer dimers, free dNTPs, polymerase etc) and the subsequent purification of the amplicons from the gel. The isolated amplicons were re-dissolved in nuclease-free water and stored at -20^{o}C. Aliquots of final amplicon samples were combined into a single sample pool ready for shipping as per the instructions provided by the sequencing service, and stored at -20^{o}C.

2.7: Sequencing via Illumina Miseq

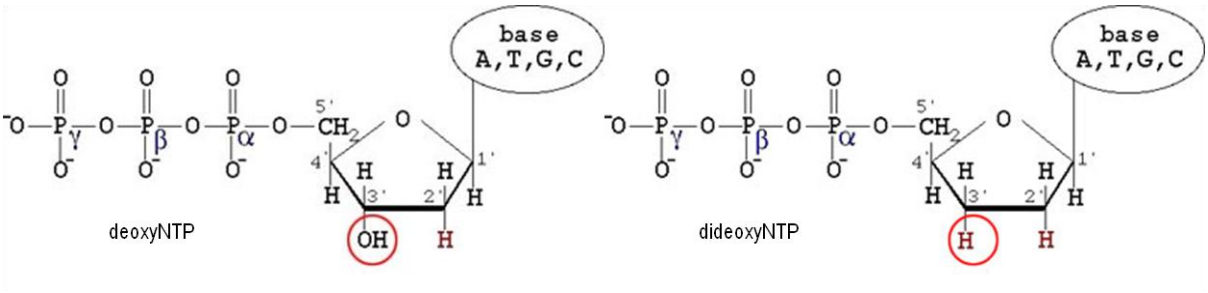
Amplicons were sequenced on the Illumina MiSeq platform, provided by the service at the Centre for Genomic Studies, University of Liverpool. Sequence trimming was included in the service, and trimmed sequences were provided for download following completion.

2.8: Sanger sequencing pilot study

During preparatory work for the Illumina sequencing project, a short pilot study was designed which utilised Sanger sequencing to glean an estimate of the likely bacterial diversity in field sites of interest. Soils for this project were collected during an earlier expedition (2014) prior to the primary field expedition in 2015, and therefore may not be examined as direct comparison to Illumina findings. However, the Sanger project may offer some supplementary verification of major genera identified by Illumina sequencing, and therefore is included here.

Sanger sequencing is a long-established method for sequencing DNA. Sanger is based around the use of nonstandard DNA nucleotides which terminate the action of DNA strand lengthening (Sanger et al, 1977). This is achieved via the addition of dideoxy nucleotide

triphosphates (ddNTPs) to the pool of otherwise normal dNTPs used in DNA replication. The ddNTPs are able to be added to a growing DNA through the action of DNA polymerase in the same manner as the standard dNTPs are, but ddNTPs cause early termination of the growing DNA chain; the structure of the ddNTP molecule does not enable binding of the subsequent dNTP unit, and so the production of the DNA strand is ended as soon as the first dideoxy base is incorporated. The difference in structure is depicted in Fig[1].



Fig[1]: Structural difference between deoxy- and dideoxynucleotide triphosphate units. In DNA strand extension, bonds occur between α-P group and 3'C OH group (circled), but this is not possible in the case of ddNTPs as the OH group is absent.

The bonds of DNA polymerisation occur at the 3' Carbon of the ribose ring. A phosphodiester bond is formed between the 3' Carbon and the α Phosphate group of the next molecule. In the case of ddNTPs, the 3' Carbon does not feature an associated hydroxyl (OH) moiety, only a single H. The lack of a 3' hydroxyl group consequently ensures that no further NTPs are able to be added to the chain, and thus the extension of the growing DNA strand ends. The use of ddNTPs results in a pool of sequences which are of differing lengths; it is random when the ddNTP will be incorporated into the growing chain, and therefore in time the full range of possible lengths of sequence are produced, up until the terminator sequence is reached and DNA polymerase stops as intended, producing a full-length sequence. The determination of the actual sequence of bases in the DNA sequence is achieved by applying markers to each set of ddNTPs – radiolabels are an example here. With individual markers for ddATP, ddCTP, ddTTP, and ddGTP, it is possible to identify the final base added to a sequence, and thus by arranging sequence fragments in order of increasing length, the entire sequence of the section is determined.

2.8.1 Use of Sanger sequencing in this study

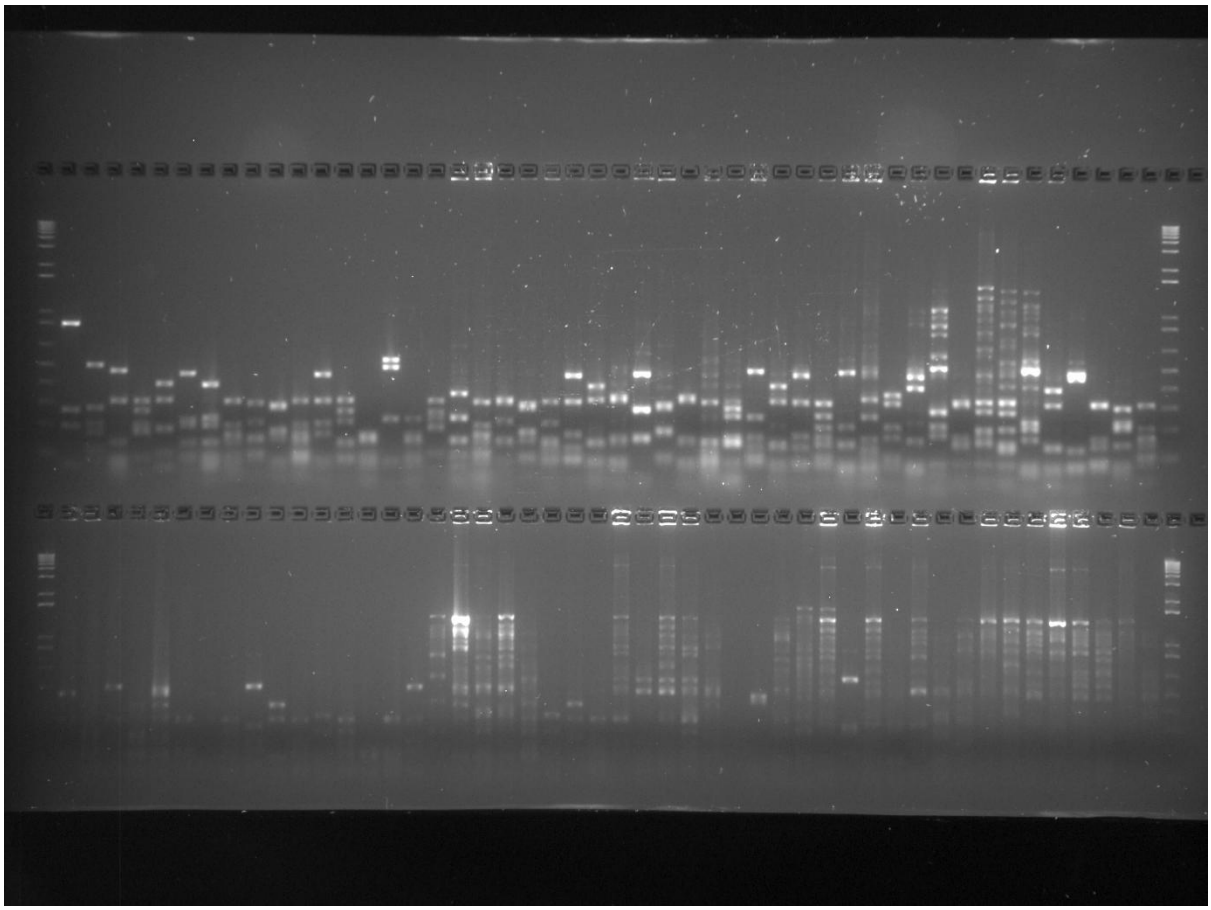
In this study, Sanger Sequencing was selected for a limited pilot experiment to investigate the diversity of bacterial species in the soils of field sites chosen in the novel Martian analogue regions of the Chilean Altiplano. Nucleic acid samples for sequencing were generated via the production of a bacterial 16S clone library. Libraries were produced for a lower-altitude, vegetated control site, and for a high altitude potential analogue site of interest. To ensure maximum effectiveness in comparison, soils with the greatest and lowest yield of extracted DNA respectively were chosen from each site, as these represent opposite ends of the spectrum in the region with respect to richness and probable diversity of soil bacteria.

2.8.2 Clone library production

A clone library was produced from whole genomic content taken from the 4296m site and the highest altitude site examined: 5432m. Genomic DNA was selectively amplified by polymerase chain reaction, using the primer pair **F 5'-GAGTTTGATCCTGGGTCAG-3' R 5'-AGGGGCGGTGTGTRC-3'** – a universal bacterial 16S primer pair, to ensure greatest coverage of bacterial species present in the soils. Recombinant TOPO vectors were created by the insertion of purified PCR product of 16S sequences, and competent *E.coli* were transformed with the recombinant plasmid. Transformed *E.coli* were inoculated onto LB-Agar plates, and incubated at 37^{o}C for 3 days.

Deep-well 96 well plates were prepared with LB-Agar, and 96 individual transformed *E.coli* were picked and each added to an individual well. The picked colonies were allowed to grow for a further 3 days at 37^{o}C. Following colony growth, cells were killed by freeze-thaw lysis: plates were alternated between -80^{o}C and room temperature several times to ensure cell killing.

Genomic material from each colony was used as template DNA in PCR amplification using the universal bacterial primers described previously. PCR product was digested with HaeIII to select for the section of the 16S sequence with the greatest variability, and all 16S digests were screened for uniqueness via Agarose Gel Electrophoresis, as shown in Fig[2].



Fig[2]: Photograph of the clone uniqueness screen for the 4296m 15cm clone library. Image depicts HaeIII digest fragments separated by agarose gel electrophoresis, imaged under UV illumination using SYBR Safe UV-reactive dye for optimal band visualisation. DNA size markers are included for reference at the end of each row. Each band pattern was manually screened for uniqueness, and all unique clones were selected for sequencing.

All unique clones were selected for Sanger Sequencing, with a single representative being taken where duplicates occurred.

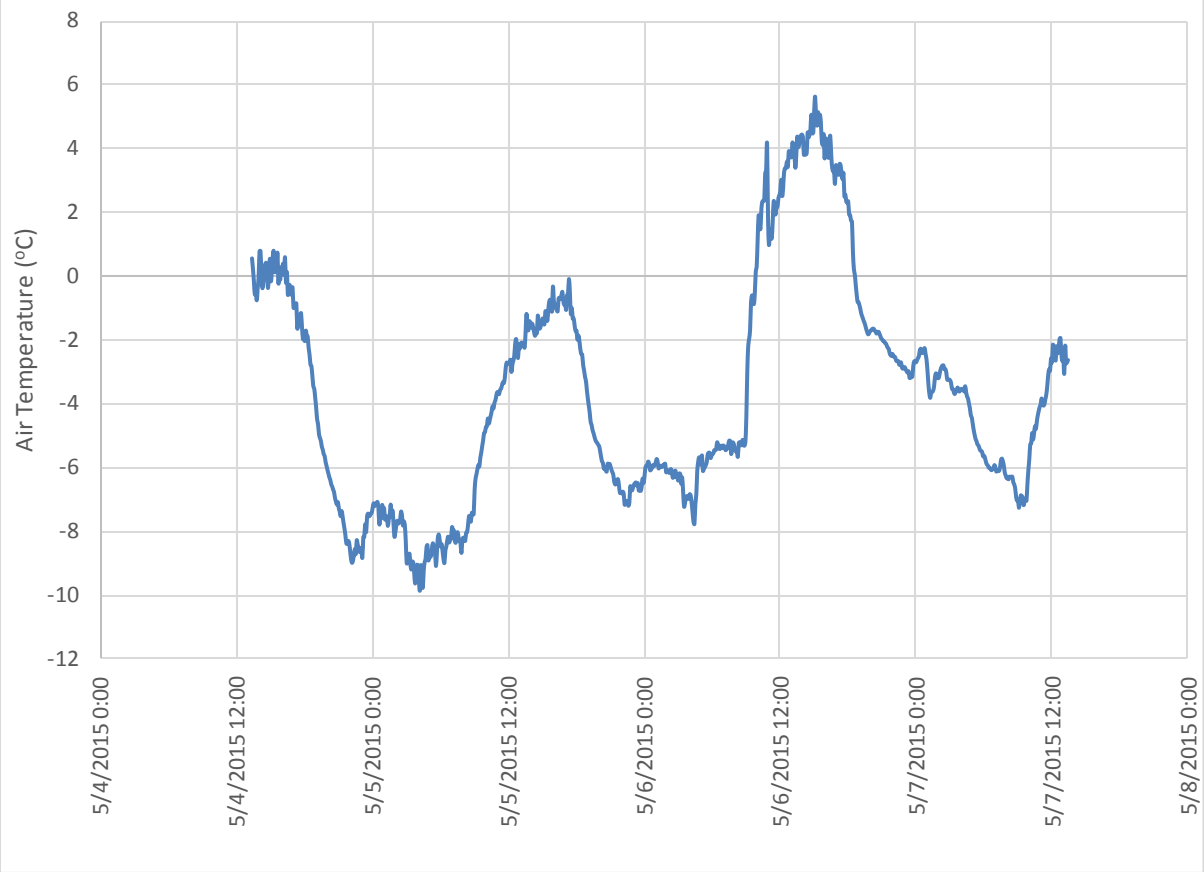
3.0: Results

3.1: Environmental Data

3.1.1: Air Temperature

Fig[3] records the daily temperature change over a period of 3 days at the primary site of interest: Site 4 (5056m). The diurnal temperature variation is clearly visible through the regular trend of peaks during the day and troughs at night. Air temperature varied between -10 and +6^{o}C.

Table[4] shows snapshot air temperature recordings taken at approximately midday at each site. The reading of 0.4^{o}C for the 5056m site matches the temperatures recorded in Fig[3], with the snapshot being taken during the same period that the CR1000 was installed.



Fig[3]: Air temperature readings for a period of 3 days, recorded using a CS215 Air Temperature & Relative Humidity probe, attached to a Campbell CR1000 datalogger. The data was recorded at site 4 (5056m).

Site	Air Temp (°C)
4200m	17.7
4700m	5
5056m	0.4
5200m	10

Table [4]: Air temperature readings taken at each field site using a handheld temperature probe. Readings were taken at approximately midday at each site.

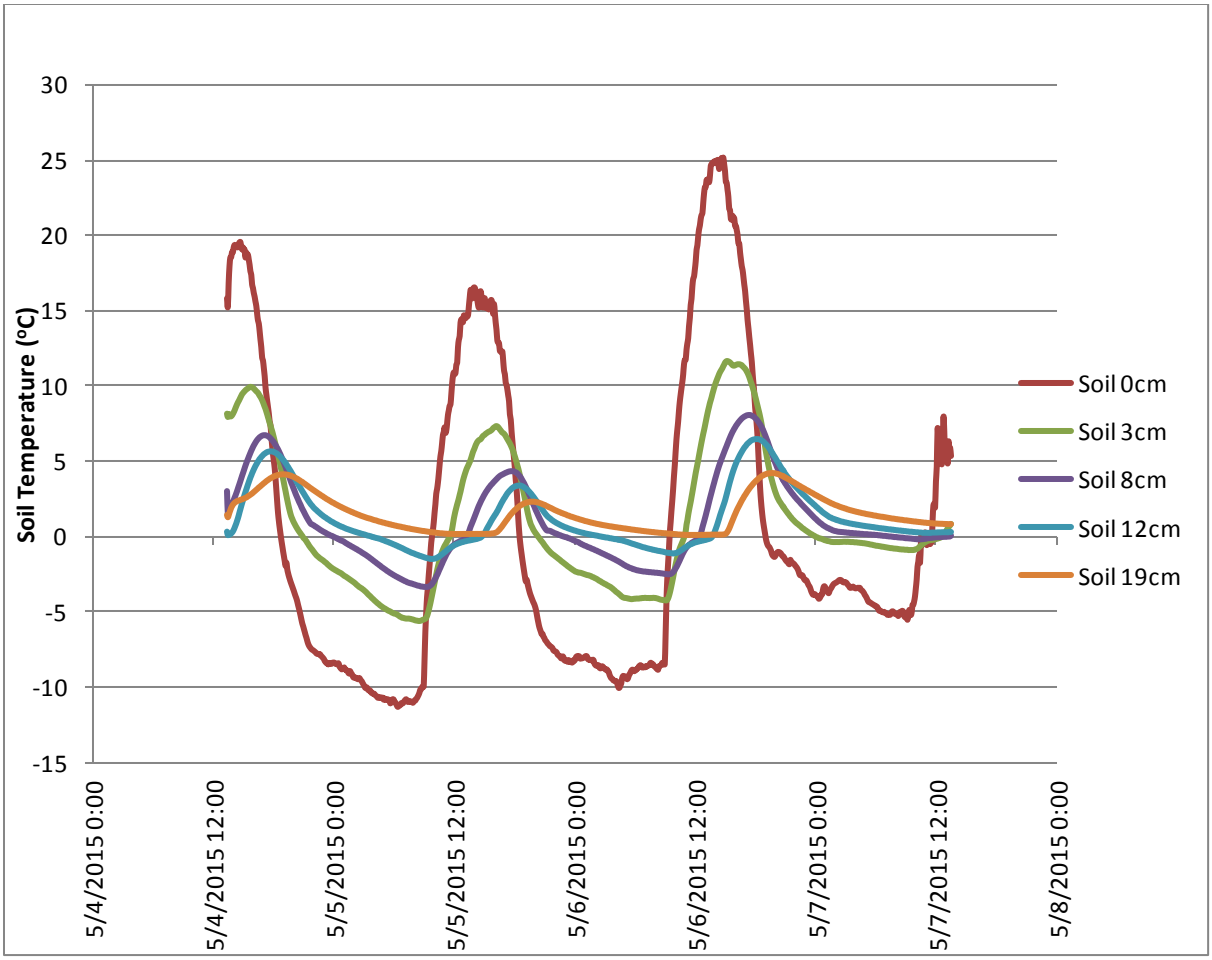
3.1.2: Soil Temperature

Figs[4-7] display the suite of readings focused on soil temperature; a particular interest in this study is how soil temperature changes with depth. Fig[4] displays recordings taken using C109 temperature probes attached to a CR1000 datalogger, while Fig[5] displays recordings taken by iButton microloggers installed in the soil at each depth. In both Figs [4] and [5], a clear diurnal variation may be observed, with peak temperature soon after midday, and minimum temperatures at night after midnight. Notably, daily temperature variations are greatest at the soil surface (ranging *in extremis* from around -10 to +20^{o}C) with daily temperature variations decreasing with depth. For example, the soil surface can experience a daily temperature range of 30^{o}C, while at 19cm the temperature range is <5^{o}C. The majority of soil layers experienced freezing at night, with deeper soils being less likely to experience subzero temperatures. During the day, soil surface was heated significantly about air temperature.

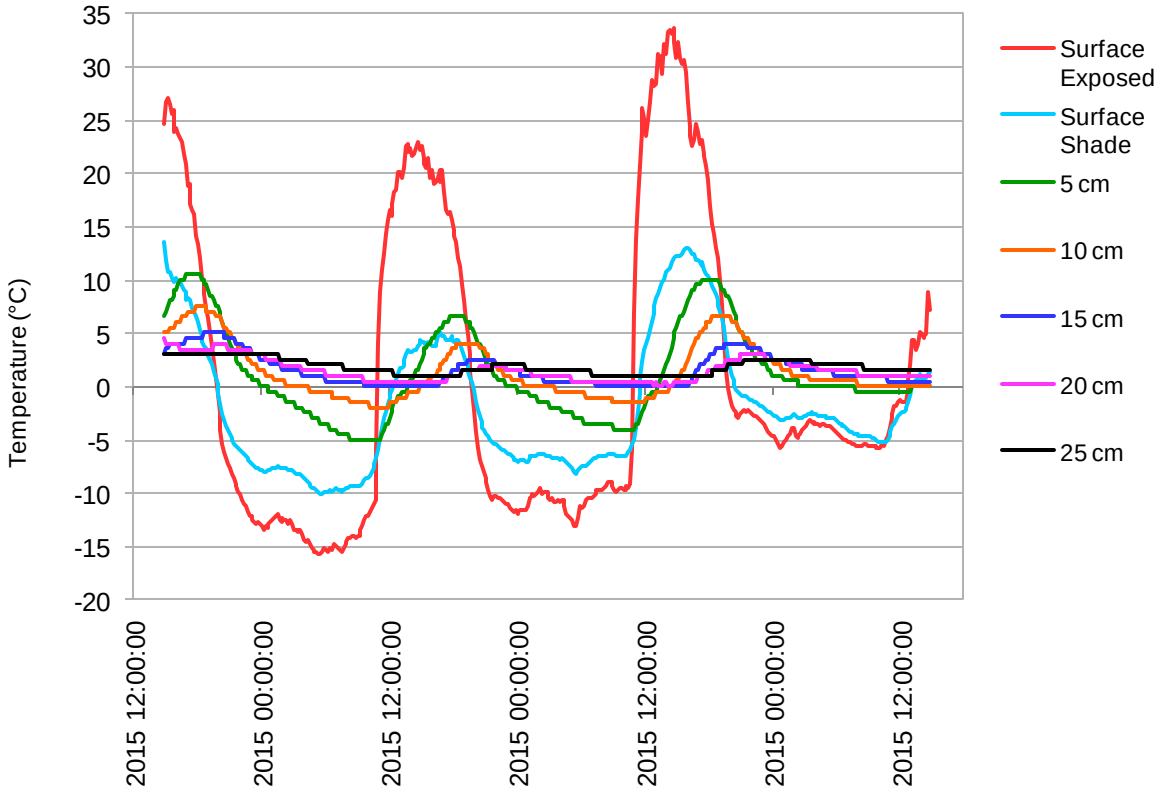
Calculated freeze-thaw stress experienced in the soils at the 5056m site are shown in Fig[6]. These were calculated from the data shown in Fig[4]. It is evident that the soil surface and near-surface soils experience greatly increased freeze-thaw rates compared to the lower depths of the soil.

FLIR thermal images taken of the soils at each site are displayed in Fig[7], with the altitude of the sites in ascending order. Each image shows a shallow trench (7cm deep) dug into the soil at the site which was immediately recorded before temperatures could equalise, enabling a direct comparison of temperature between the soil surface and the shallow subsurface. The advantage of the thermal imagine camera is that it records the temperature of the very soil surface, while other probes that record soil surface temperature are, in reality, placed a few mm into the soil. At the highest altitude sites such as the 5432m site, the base of the trench contained soil frozen in ice.

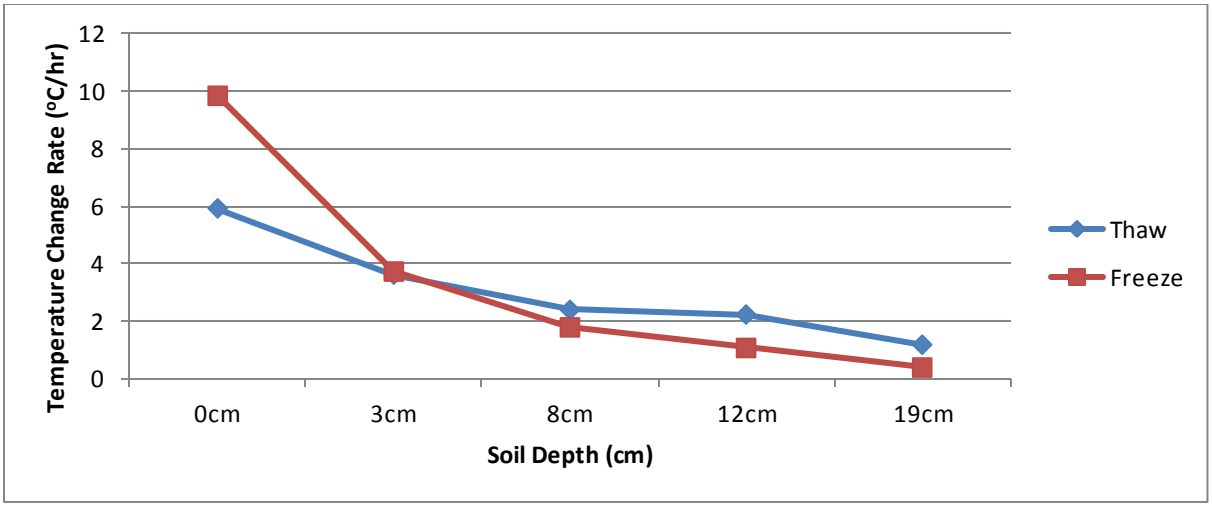
The thermal imaging camera records how the surface of the soil exposed to the Sun warms considerably. In these examples the soil was typically at least 20^{o}C warmer than the air temperature at the time of recording.



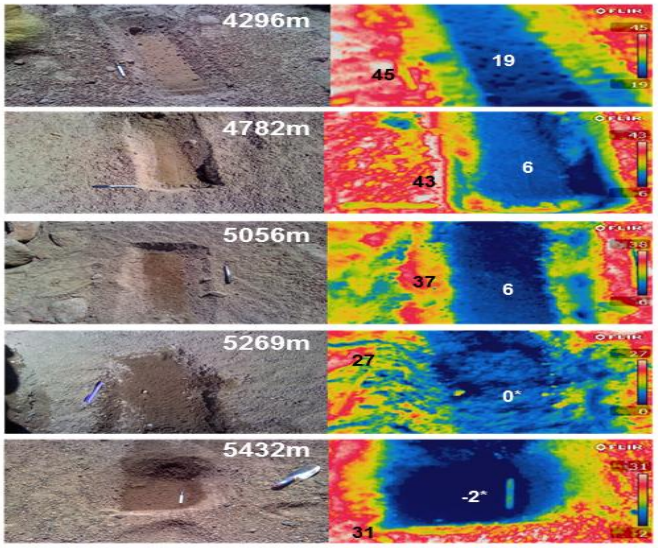
Fig[4]: Soil temperature readings for a period of 3 days from 5056m (Site 4), recorded using C109 temperature probes, attached to a Campbell CR1000 datalogger. Probes were installed in the soil at increasing depth.



Fig[5]: Soil temperature readings for a period of 3 days from 5056m (Site 4), recorded using iButton micrologger temperature and relative humidity probes, buried in increasing soil depth at 5cm intervals. Additional readings taken at the soil surface either exposed to the sunlight (surface exposed), or under an opaque object (surface shade) were recorded using TinyTags microloggers.



Fig[6]: Peak freeze-thaw rates experience in the soils of site 4 (5056m), derived from iButton temperature readings depicted in Fig[5]. The rate is measured as the mean change in temperature over any given hour during the period of greatest change over the diurnal cycle.



Fig[7]: Images taken using FLIR thermal imaging camera of soils at each site. The base of the trench is at a depth of 7cm below the soil surface, to illustrate the difference between the surface and shallow subsurface. Altitude of site is in ascending order with the top image representing site 1 (4296m). Images for each site were taken at approximately midday. Note: sites not visited in the same day.

3.1.3: Air Relative Humidity

Fig[8] and Table [5] display recordings of air relative humidity (%) taken at each site. Fig[8] depicts data collected using an air temperature & humidity probe attached to a CR1000 datalogger, whlle Table[5] contains snapshot RH readings taken using a handheld probe at approximately midday at each site. Extent of diurnal variation in relative humidity at the 5056m site is less pronounced than the variation in air temperature, but is clearly observed nonetheless. The snapshot reading for 5056m in Table[5] was taken while the CR1000 was being installed at the site, and therefore similar RH values were recorded.

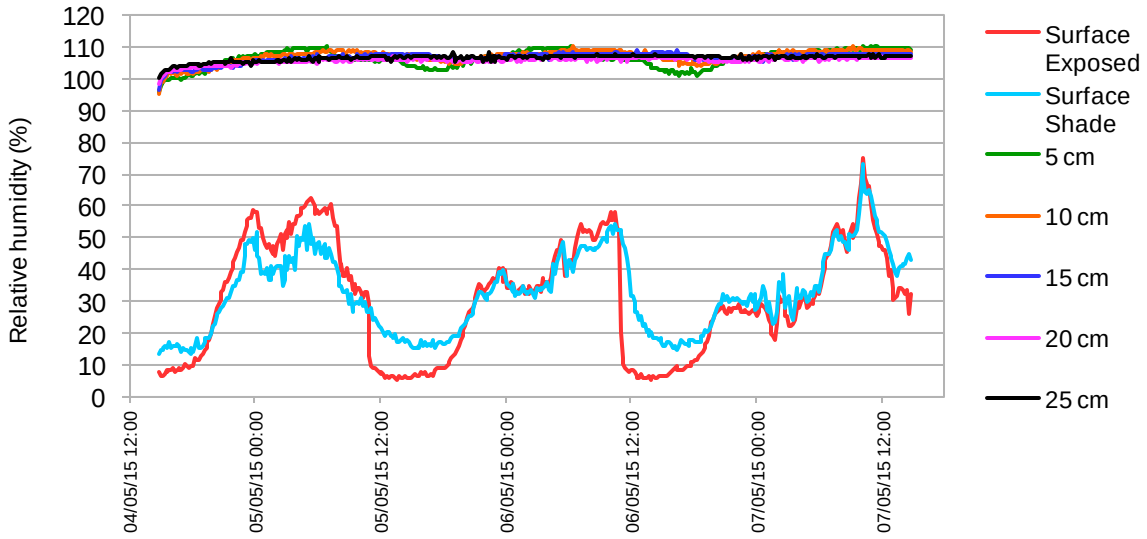
Fig[8]: Air relative humidity recorded over a period of 3 days at site 4 (5056m). Recordings were taken using a CS215 Air Temperature & Relative Humidity probe attached to a Campbell CR1000 datalogger

Site	Air RH (%)
4200m	11.6
4700m	26.4
5056m	21.4
5200m	18.2

Table[5]: Air relative humidity readings taken at each field site using a handheld relative humidity probe. Recordings were taken at approximately midday at each site.

3.1.4: Soil Relative Humidity

Fig[9] reports the recordings of relative humidity obtained by iButton microloggers at the 5056m site at a range of soil depths over a period of 3 days. Humidity is ostensibly a measurement of moisture in clear air, and its measurement in the moisture of a soil environment results in the upper limit of detection being overwhelmed and a reading of 100% humidity at all times. This limitation is evident in Fig[9], as all recordings taken from below the soil surface read 100% humidity at all times. The soil surface itself offers a clear recorded diurnal variation both exposed to direct sunlight and under the shade of an opaque object. It is worthy of consideration that previous recordings (taken in 2014) using iButtons did provide greater clarity in humidity measurements in soil

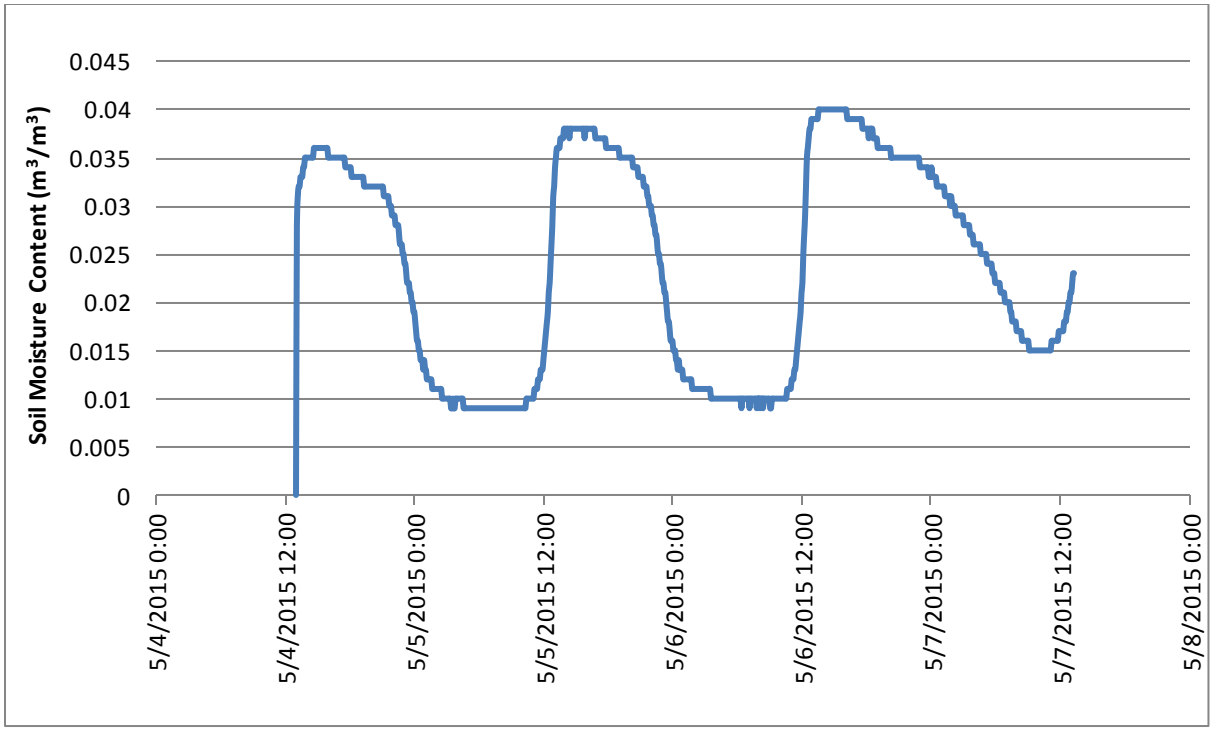


Fig[9]: Soil relative humidity recorded over a period of 3 days using iButton microloggers, installed in soils at 5cm depth intervals.

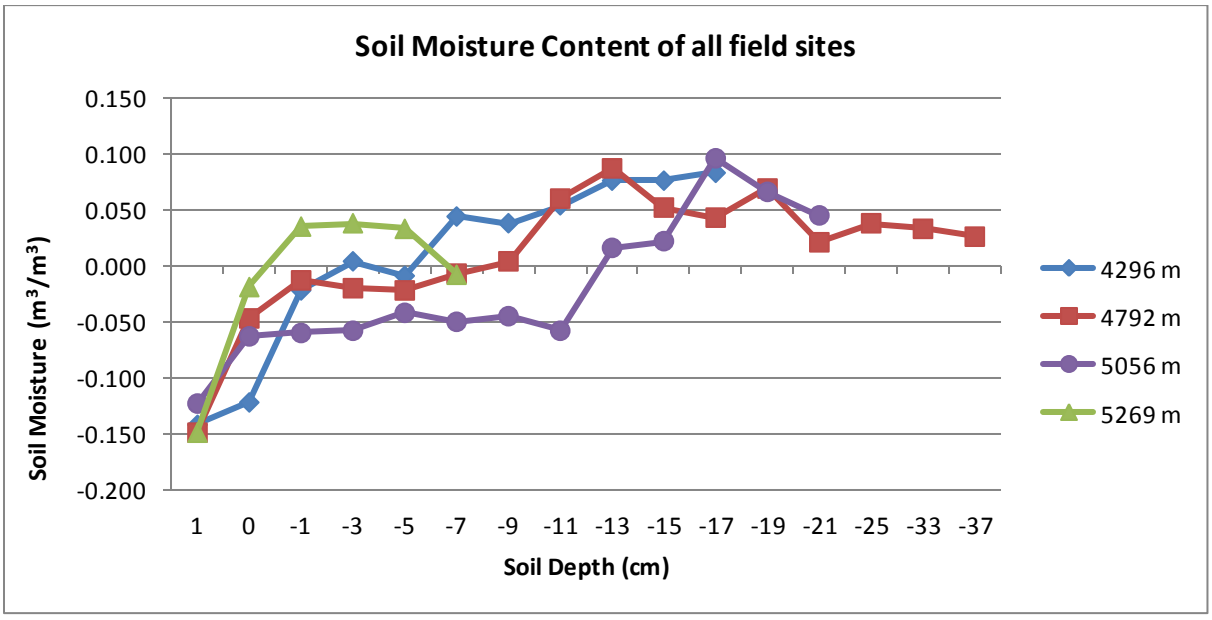
3.1.5: Soil Moisture

Recordings of soil moisture are displayed in Figs [10] and [11]. Fig[10] features data recorded using a volumetric water content probe attached to a CR1000 datalogger at the 5056m site, recorded over a period of 3 days, while Fig[11] features snapshots of soil moisture recorded via handheld probe at each site over a range of depths which match the depth of soil sample collection.

The data in Fig[10] indicates a clear diurnal variation in soil moisture, and that water content is low in these soils (in comparison to more temperate soils (Ma et al, 2014)) even at the peak moisture content: $0.03m^3/m^3$ = 3% VWC. The soil moisture content varies over a daily cycle between 1% and 3%; the water content highs occur at shortly after midday and lows occur at around midnight. Fig[11] indicates a general trend of soil moisture increasing with depth at all sites. Of particular note are the recordings at the 5056m site, wherein soil moisture remains highly stable at a low level until a depth of 11cm followed by an increase in soil moisture in the deeper layers. Readings at the 5269m site end at a depth of 7cm due to the presence of ice which was too resistant to correctly insert the moisture probe.



Fig[10]: Soil moisture recorded at site 4 (5056m) over a period of 3 days using a Soil Volumetric Water Content Reflectometer attached to a Campbell CR1000 datalogger.



Fig[11]: Soil moisture content at each field site recorded using a handheld soil moisture content probe. Taken at approximately midday at each site, readings were taken at increasing soil depth at 2cm intervals, matching the depths at which soils were excavated for sampling.

It is important to note the soil moisture values in Fig[11] appear to report negative soil moisture volume, which is not possible. This is due to a calibration error regarding the handheld soil moisture probe, which was calibrated for use in temperate soils rather than the volcanic tephra of the sites examined in this study which are comparatively much more arid. Due to this the "zero" point is incorrect, and this results in negative values at lower moisture content. While this does not invalidate the overall changes in moisture with depth and observed trends therein (which is the primary interest for this study), it does necessitate the caveat that these figures may not be stated as true quantified moisture content.

3.1.6: Solar Flux

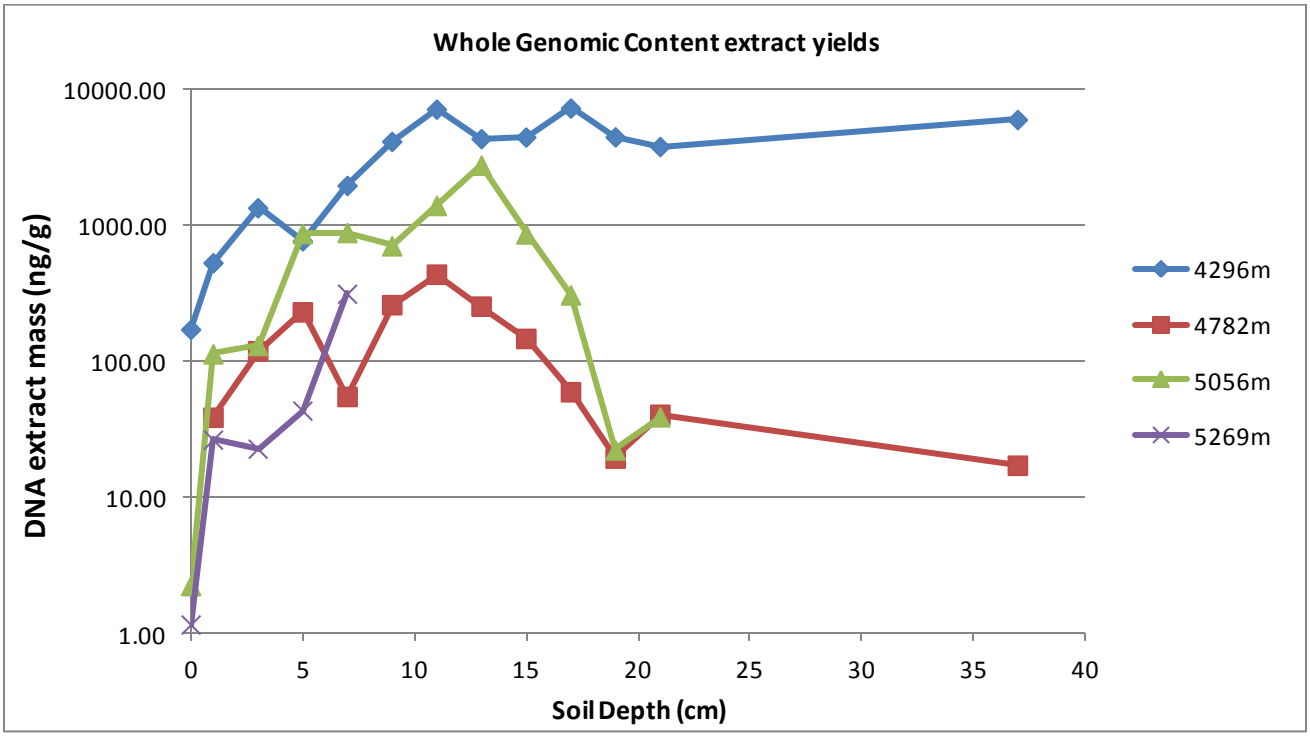
Table[6] displays the recorded intensities of incoming solar radiation, measured at each field site at approximately midday. The values do not increase with altitude as one would expect (although it should be noted that overall change in UV irradiation with altitude is comparatively low in such arid tropical regions (Piazena, 1996)) due to small variations in thin cloud cover during recordings, which were enough to obscure any altitude-related readings. No UCV-wavelength solar radiation was detected at any of the sites, as UVC is effectively attenuated by the atmosphere.

Site	UVA (W/m^2)	UVB (W/m^2)	UVC (W/m^2)	PAR ($umol/m^2/s$)
4200m	22.74	3.78	0	1750
4700m	23.6	3.72	0	1750
5056m	22.6	3.5	0	1622
5200m	21.56	3.285	0	1696

Table[6]: Recorded intensities of incoming solar flux at each field site. Readings were taken at approximately midday at each site. Readings were taken for UVA, UVB and PAR wavelengths. UVC is attenuated by the atmosphere and therefore all UVC readings are zero.

3.2: Whole Genomic Content Extraction

Yield of whole genomic DNA content extraction is displayed in Fig[12] and Table[7]. Extract yields from all sampled sites and soil depths are combined for ease of comparison between the sites. The values depicted in Fig[12] indicate that the surface soils yielded very little whole genomic content, with DNA extract concentration increasing with depth until a peak at approximately 11-15cm, with some variation between sites but a similar range overall. At further depths however, DNA extract yield is decreased. It is additionally evident that the control site (4296m) yielded substantially more DNA than the higher altitude sites.



Fig[12]: DNA concentration of whole genomic content extract from collected soil samples. Values represent the mean value of triplicate extracts, and are normalised against mass of soil used in extraction (ng DNA per g soil used). Log^{10} axis used for clarity of observation of comparatively low yield samples.

Soil Depth	DNA extract yield (normalised to soil mass used)			
	4296m	4782m	5056m	5269m
0	171.68	0.00	2.25	1.16
1	530.48	38.80	113.51	26.57
3	1343.38	119.42	130.85	22.77
5	768.55	230.93	869.44	43.45
7	1960.70	55.18	883.90	313.47
9	4130.84	260.69	706.52	
11	7128.51	436.70	1398.25	
13	4330.47	254.25	2764.74	
15	4446.03	147.29	863.53	
17	7262.50	59.75	307.66	
19	4460.92	19.44	22.28	
21	3787.49	40.79	39.09	
37	5977.82	17.22		

Table[7]: Yield data for Whole Genomic Content extracts for each field site under investigation. DNA extract mass is normalised against mass of soil used in extraction, presented here as ng DNA mass per g soil used.

3.3: Identifying Bacterial Taxa

Identified bacterial taxa from sequenced 16S V4 amplicons are presented in Fig[13] and Fig[14]. These TAXAplots depict the most abundant genera identified in each set of similarly barcoded sequences – where the barcode effectively identifies the site and soil depth from which the DNA was taken – as soil depth increases (from left to right).

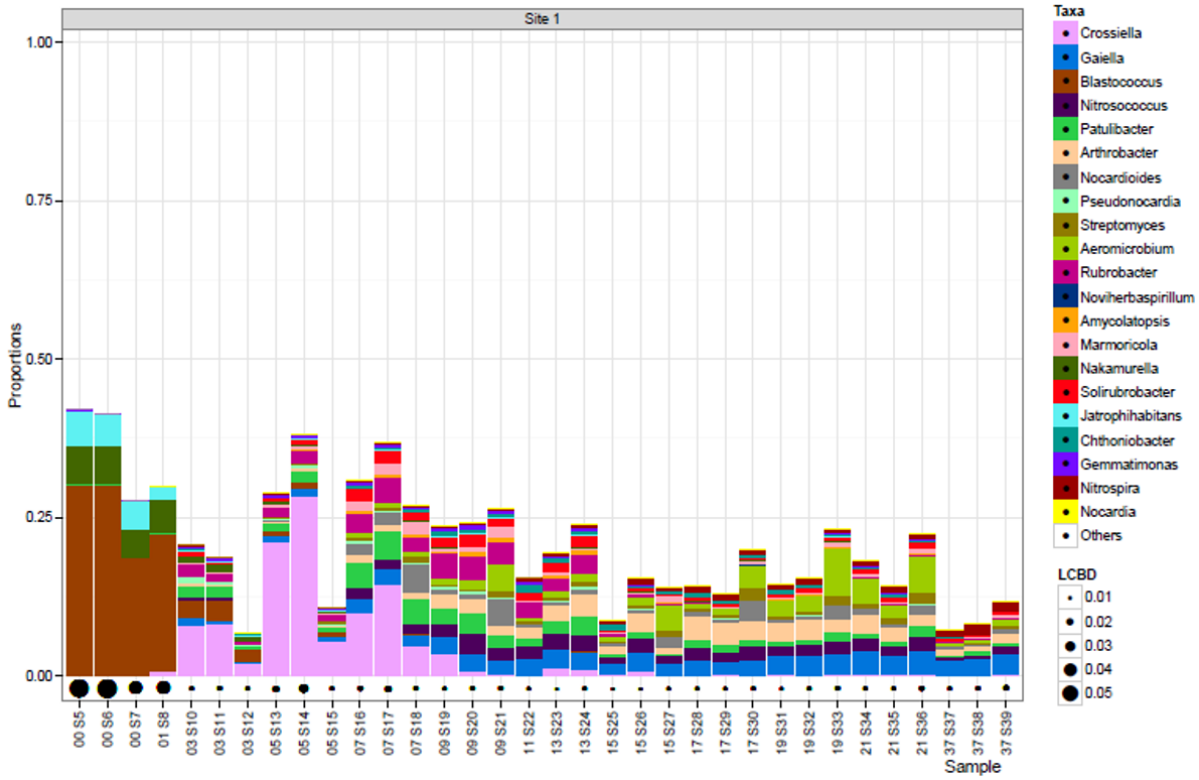


Fig.[13]: The top 20 most abundant taxa identified from Illumina sequencing data for the 4296m control site (Site 1). Taxa are presented at the Genus level. Soil depth (cm) increases along the x-axis, denoted by 00, 01, 03 etc, while the y-axis represents the relative proportion of the detected community which is accounted for by the noted taxa, with 1.00 representing 100% of the detected population. Black circles above the x-axis represents the Local Contribution to Beta Diversity effect of each population.

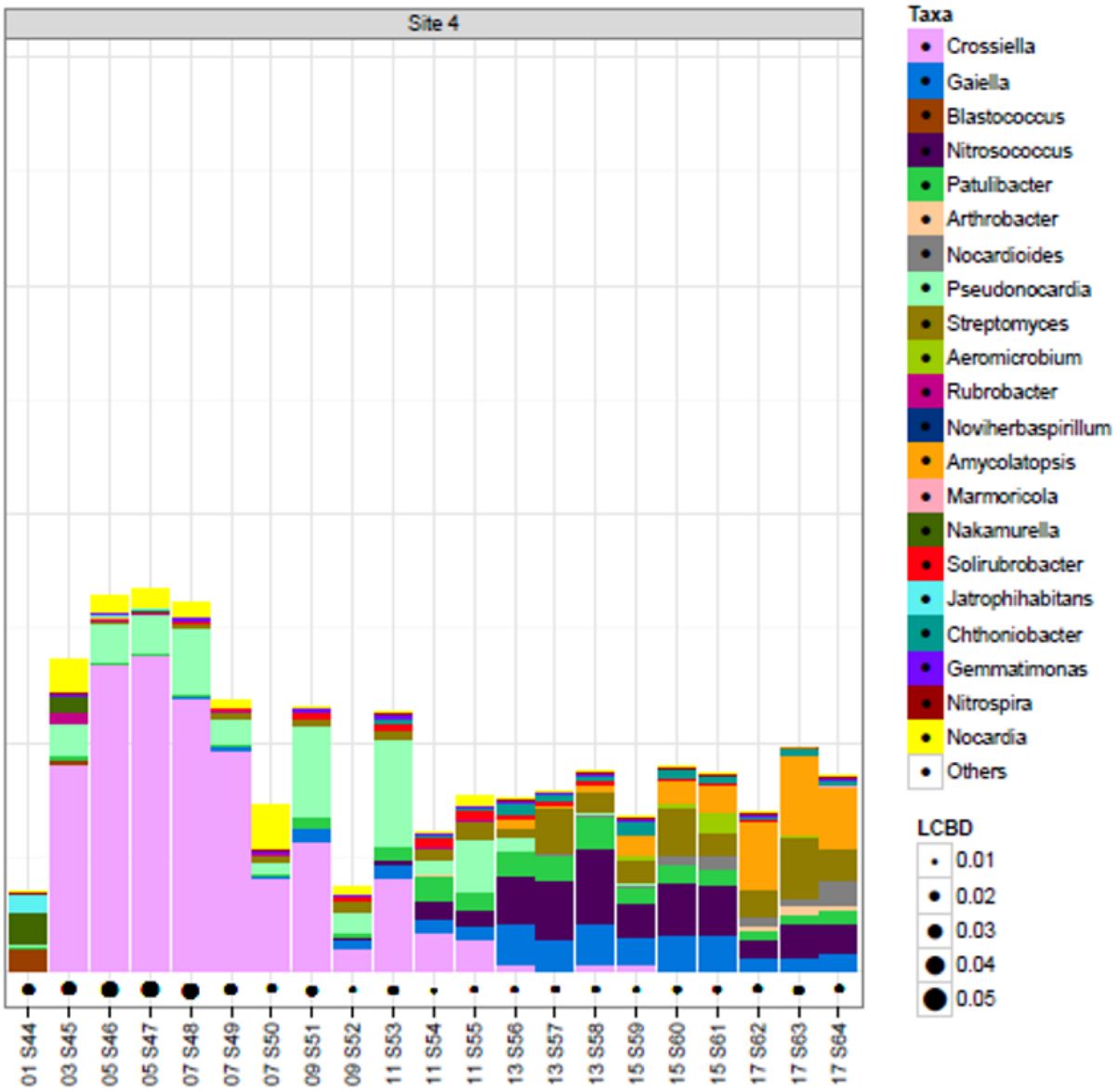


Fig.[14]: The top 20 most abundant taxa identified from Illumina sequencing data for the 5056m site (Site 4). Taxa are presented at the Genus level. Soil depth (cm) increases along the x-axis, denoted by 01, 03, 05 etc, while the y-axis represents the relative proportion of the detected community which is accounted for by the noted taxa, with 1.00 representing 100% of the detected population. Black circles above the x-axis represents the Local Contribution to Beta Diversity effect of each population.

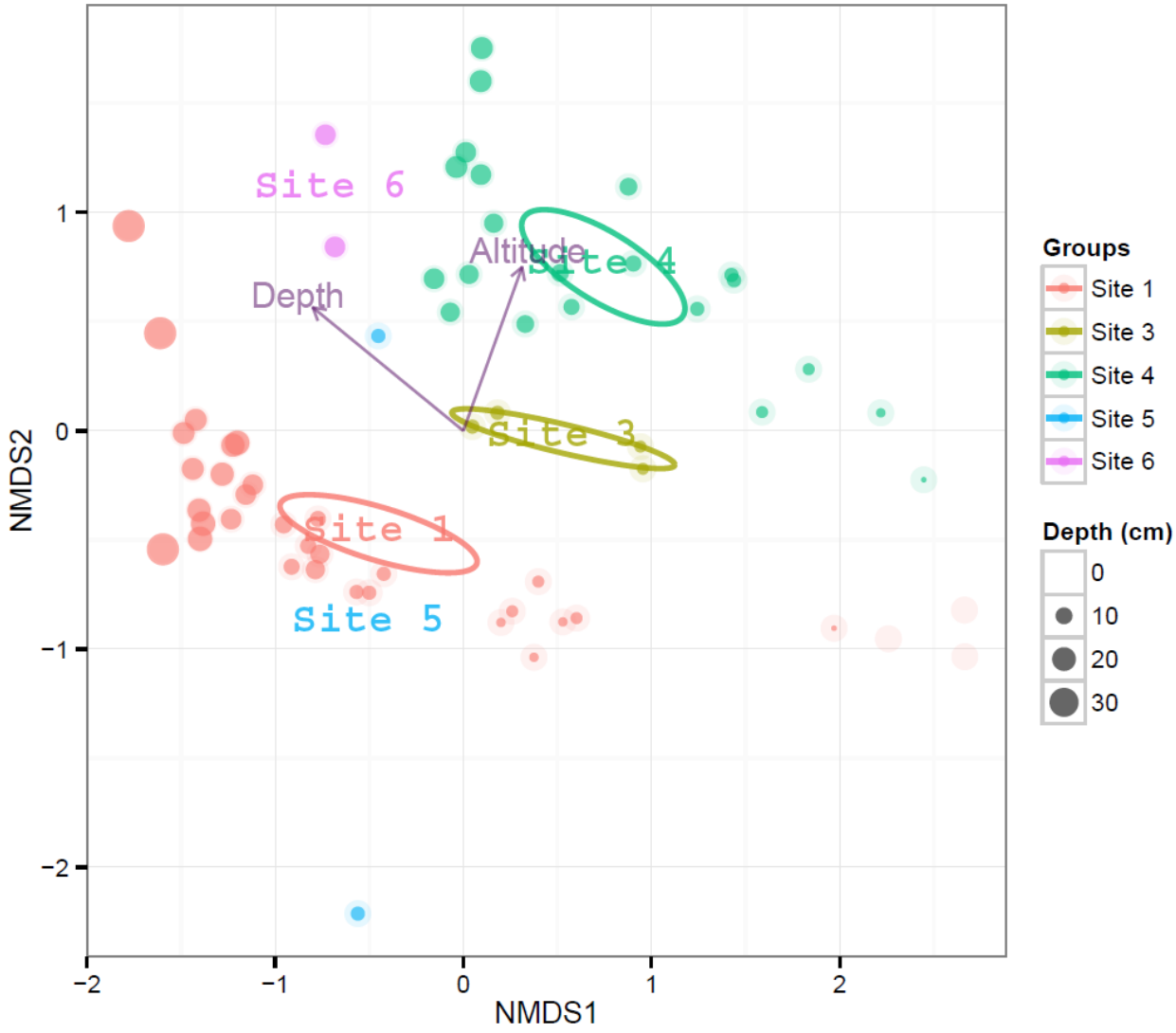
The TAXAplots displayed in Fig[13] and Fig[14] display the relative community diversity of the soil samples taken from the field sites at 4296m and 5056m respectively. The displayed taxa (in this case at the Genus level) are presented according to the proportion of detected sequences they represent in the sample pool. The 20 most abundant genera identified in the Illumina sequence data are presented as coloured column sections, with the remaining genera represented by the non differentiated white space of the column. Columns are arranged by increasing depth of the original soil sample from which the DNA used to generate amplicons was extracted, and with this it is possible to visualise how the makeup of the bacterial community changes as soil depth increases. In both Fig[13] and [14] it is clear that a marked change in community structure occurs as depth increases. In both cases, the near-surface depths feature a notably dominant Genus which accounts for a substantial portion of the total community. Both sites feature a substantial presence of Crossiella in the shallow depths (approximately 3-9cm), while the 4296m site additionally features a dominant Blastococcus presence at the surface and just below the surface (0-1cm). The upper depths at 5056m also feature a notable presence of Nocardia, which is largely absent from the most dominant genera in the 4296m site. The relative dominance of Crossiella and Blastococcus disappears after approximately 11cm depth of soil, and this is notable as this is where the soil moisture rises.

At around 11cm depth in samples from both sites, the community features less noticeably dominant species, and the top 20 most abundant genera in the community typically account for less of the total community, proportionally (this is manifested as the coloured section of the column being shorter). Based on the fact that the deeper soils are less dominated by certain genera, it appears that the deeper soils support a greater diversity of species than the shallow depths and the surface. The more extreme conditions experienced at the soil surface (as recorded in Figs[4-11]) evidently do not support as great a diversity of species as the less hostile depths.

The TAXAplots also indicate the relative values for the Local Contribution to Beta Diversity (LCBD) values for each sample soil depth, which is represented as circles below the columns for the sample. Increased LCBD value is indicated by increased size of circle. Beta diversity is defined as “the extent of change in community composition, or degree of community differentiation, in relation to a complex-gradient of environment, or a pattern of environments” (Whittaker, 1960). Put simply, Beta Diversity describes the number of different communities in a region, a region which may feature different environments.

Therefore, LCBD describes the weight effect that a particular community has on the overall diversity of a region. In this study, individual LCBD values may illuminate the extent to which the communities of each soil layer affect the overall diversity of the site as a whole.

In both Fig[13] and [14], the LCBD values indicate that the comparatively few species present at- and near the soil surface impart the greatest LCBD effect, with the deeper soils having lesser impact, although a slight increase is present at the lowest depths in the 5056m site. This suggests that the structure of the community of the more shallow, extreme environment soil layers has the greatest effect on the diversity of the deeper soil community.



Fig[15]: A Non-metric Multi-Dimensional Scaling plot displaying Bray-Curtis dissimilarity of OTUs derived from Illumina sequencing. Points are grouped into Site and Depth of original soil samples, with each Site representing a different elevation, as follows: Site 1 – 4296m, Site 3 – 5269m, Site 4 – 5056m, Site 5 – 5432m, Site 6 – 5282m. OTUs are observed to cluster according to soil depth, with no overlap between the main sites of interest at 4296m (Site 1) and 5056m (Site 4) respectively.

Fig[15] displays an NMDS plot of Operational Taxonomic Units (OTUs) identified in Illumina sequencing data, with all sequenced samples included. Points are arranged using Bray-Curtis dissimilarity index, which distinguishes between sites by similarity in OTU

composition by counts. Of the main sites of interest (Site 1 and site 4), there is a clear separation between the two groups, and each group shows similar clustering trends – increases in depth trend towards the left of the plot in the same manner for both sites, indicating that the effect of change with depth is of similar significance at both sites, while the separation between the clusters indicates that the two sites do not have any notable effect on each other, and can be considered unconnected.

The trend of clustering with depth indicates that the overall community structure of each site changes as soil depth increases. The change in soil environmental conditions as depth increases is undoubtedly related to this, and indeed changes in overall community diversity and structure were observed in this study, as previously described in figures such as the TAXAplots showing genus-level diversity in each soil sample. At first glance it appears that clustering additionally occurs relative to altitude also, but this is not the case; Site 3 represents a higher elevation than Site 4 (5269m and 5056m respectively). Site 5 (5432m) is also of a higher elevation.

Fig[15] displays clusters indicative of field sites which are not present in the TAXAplots (Fig[13] and Fig[14]). These account for additional field sites investigated during pilot studies prior to the final selection of the 4296m and 5056m sites as primary sites of interest – Sites 5 and 6 – and sites which were not selected for Illumina sequencing due to low genomic content or unsuitability for PCR amplification – Sites 2 and 3.

3.4: Predicting Significant Proteins

The Tax4Fun script enables the comparison of levels of expression of the predicted expressed proteins between different depths of sample acquisition. Table[8] displays the top 20 predicted proteins, the expression of which is significantly upregulated relative to the samples from the other side of the 11cm cut-off, and describes the primary functions of these proteins. The majority of the top 20 significantly-differently upregulated proteins are found in the lower depths of the soil, with only NarQ and gluconate hydrogenase being upregulated to a greater extent in the upper depths.

#	Function	Significance	Upregulated in
1	Nitrate reduction	Soil nutrient cycling. Induced in low osmotic pressure	Lower
2	Interaction between pathogen and host	involved in resistance to oxidative and osmotic stress	Lower
3	Protein folding	Metabolic, stabilises proteins in periplasm	Lower
4	Nucleic Acid recycling	Metabolic	Lower
5	Production of LPS	Cell motility and adherence, biofilm formation, antibiotic resistance	Lower
6	Pili construction	Cell motility and adherence	Lower
7	Pathogenic entry of bacteria into host		Lower
8	Activity of Pili	Cell motility and adherence	Lower
9	Catalyst in gene transposition	Metabolic, connotations for antibiotic resistance	Lower
10	DNase	Metabolic	Lower
11	Nitrate reduction	Soil nutrient cycling	**Upper**
12	Unknown	Unknown	Lower
13	Metalloprotease secretion	Involved in degradation of extracellular proteins	Lower
14	Oxidoreductase	Cell respiration, metabolic	Lower
15	Assembly of Pili	Cell motility and adherence	Lower
16	Glucose and gluconate degradation	Metabolic	**Upper**
17	Peptidase	Metabolic	Lower
18	Involved in pathogenesis		Lower
19	NADH dehydrogenase	Cell respiration, metabolic	Lower
20	Stress response	Samples taken from a hostile environment	Lower

Table[8]: A table displaying the top 20 most significantly-differently up- or down-regulated protein-coding genes predicted by the Tax4Fun script for 16S amplicons generated from Whole Genomic Content extracted from soil samples collected at Site 4

(5056m). Sequences were compared between the Upper (3-11cm) and Lower (11-37cm) soil depths at the site. Rank order from 1-20 is arranged by the range of the difference in up/downregulation, with #1 being the most significantly different.

3.5 Sanger pilot study data

Clone library Sanger sequencing was carried out at the Edinburgh Genomics Genesifter facility, and FASTA file data returned digitally for each sequence. Sequence data for each sample was run through the *Seqmatch* software from the Ribosomal Database Project (RDP). RDP's sequence matching identifies input sequences based by identifying where the sequence in question has been previously mentioned in the literature, and offers matches at all taxonomic levels where possible (Cole et al, 2014). All sequences able to be matched at the Genus level or below are presented in Tables[9-10].

Genus	Occurrences	% of population
Gaiella	14	28
Acidobacteria gp4	7	14
Nocardioides	5	10
Gemmatimonas	5	10
Roseiflexus	4	8
Terrimonas	2	4
Arthtrobacter	2	4
Lamia	2	4
Afifella	2	4
Aciditerrimonas	2	4
Conexibacter	2	4
Bascillus	1	2
Streptomyces	1	2
Microlunatus	1	2
TOTAL	**50**	

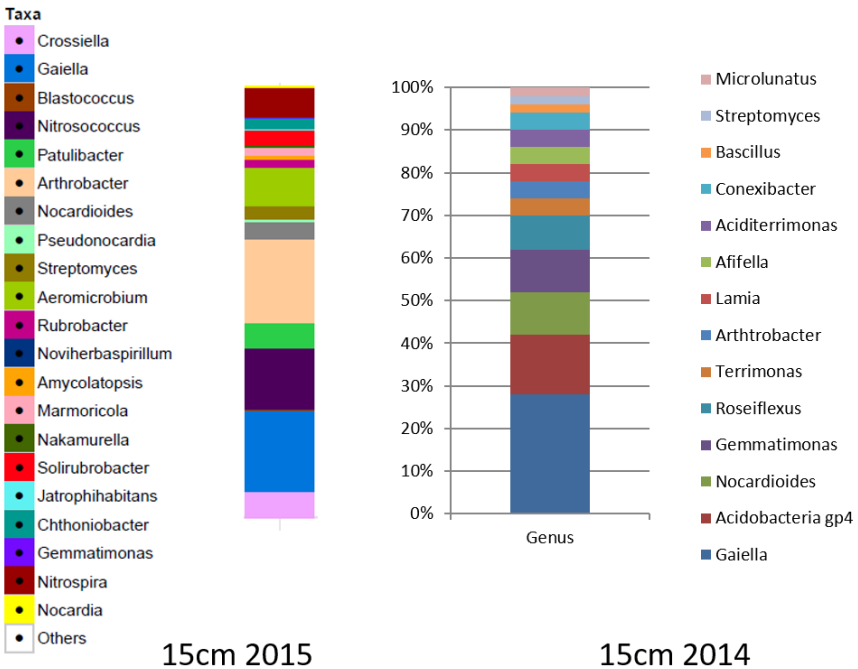
Table[9]: Identified genera of the 4296m 15cm soil depth clone library, as acquired by RDP sequence matching. Total occurrence count of each genus and relative percentage of the total library population are included. Sequences which could not be matched at the Genus level are not included.

Genus	Occurrences	% of population
Gaiella	15	44.12
Marmoricola	10	29.41
Segetibacter	3	8.82
Ferruginibacter	2	5.88
Aeromicrobium	1	2.94
Sphaerobacter	1	2.94
Gemmatimonas	1	2.94
Lamia	1	2.94
TOTAL	**34**	

Table[10]: Identified genera of the 5432m 3cm soil depth clone library, as acquired by RDP sequence matching. Total occurrence count of each genus and relative percentage of the total library population are included. Sequences which could not be matched at the Genus level are not included.

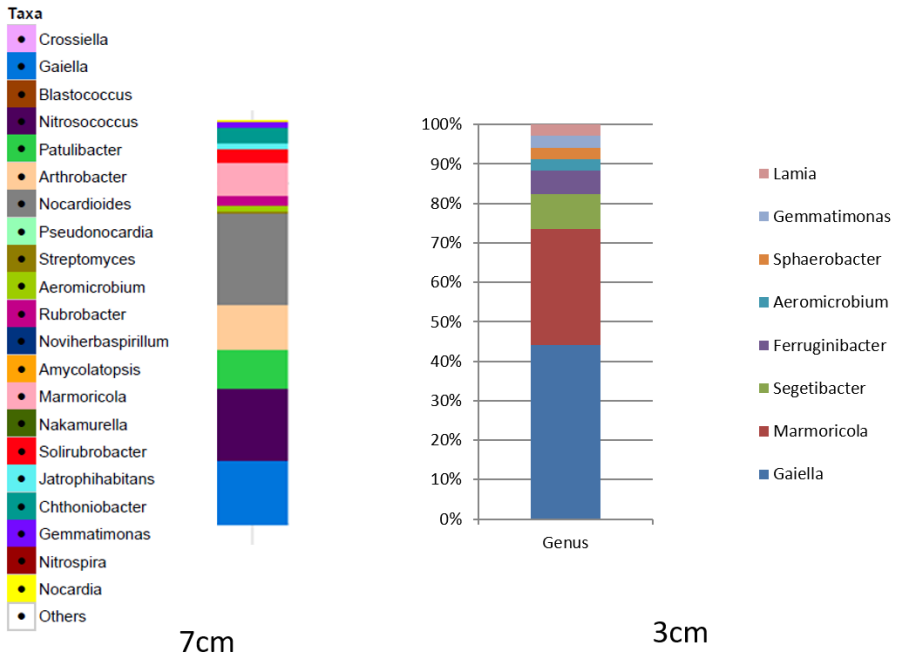
3.1.1 Sanger vs Illumina

For the purposes of comparison of diversity in the communities, as well as comparison between Sanger and Illumina sequence identification, genera identified in Tables[9] and [10] are presented as stacked columns in Figs[16] and [17], alongside representative columns from the TAXAplots produced from Illumina sequencing data discussed previously.



** Column only accounts for approx 20% of identified in the tax = top 20 most abundant*

Fig[16]: Genera identified in 4296m 15cm soils identified by Illumina (left) and Sanger (right) sequencing. Identified genera are presented in stacked columns to display relative proportion of the total community accounted for by each genus. It is important to note that the Illumina stack represents approximately 20% of the total genera identified in the sample, and merely accounts for the top 20 most abundant genera in the sample, while the Sanger stack accounts for the entire diversity of the selected clones from the produced library.



** Column only accounts for approx 30% of identified in the tax = top 20 most abundant*

Fig[17]: Genera identified in 5432cm soils identified by Illumina (left) and Sanger (right) sequencing. Soil depths are different between the two sequencing platforms due to the production of Illumina-ready amplicons being unsuccessful when utilising 3cm soil WGC as template DNA, necessitating use of the 7cm soil WGC instead. Identified genera are presented in stacked columns to display relative proportion of the total community accounted for by each genus. It is important to note that the Illumina stack represents approximately 30% of the total genera identified in the sample, and merely accounts for the top 20 most abundant genera in the sample, while the Sanger stack accounts for the entire diversity of the selected clones from the produced library.

It is clear that the two sequencing platforms did not produce identical results, but there are notable similarities in the identified taxa for each platform. Genera which are identified in both sets of sequences are highlighted in Tables[9] and [10]. An important point to be considered is that while the Sanger sequence columns represent 100% of sequences submitted, the Illumina columns do not – they merely account for the 20 most abundant genera in that sample of sequences, and typically this only accounted for around 20-30% of total taxa identified in the sample (represented as white space above the coloured column in

the original TAXAplots). Despite this, more taxa overall were identified using Illumina, highlighting the greater sensitivity achieved via NGS platforms.

Sanger		Illumina	
Sample	4296m 2014 15cm	Sample	4296m 2015 15cm
Identified Genus	*Gaiella* *Acidobacteria gp4* *Nocardioides* *Gemmatimonas* *Roseiflexus* *Terrimonas* *Arthtrobacter* *Iamia* *Afifella* *Aciditerrimonas* *Conexibacter* *Bascillus* *Streptomyces* *Microlunatus*	Identified Genus	*Crossiella* *Gaiella* *Blastococcus* *Nitrosococcus* *Patulibacter* *Arthrobacter* *Nocardioides* *Pseudonocardia* *Streptomyces* *Aeromicrobium* *Rubrobacter* *Amycolatopsis* *Marmoricola* *Nakamurella* *Solirubrobacter* *Jatrophihabitans* *Chthoniobacter* *Nitrospira* *Nocardia*

Table[11]: Comparison of identified genera from Sanger and Illumina sequencing for 4296m site soils. Sanger data represents clone library sequences, while Illumina data is taken from TAXAplot data, specifically the sample represented in Fig[16]. Matches highlighted indicate where a genus is identified in both datasets.

Sanger		Illumina	
Sample	5432m 2014 3cm	Sample	5432m 2014 7cm
Identified Genus	*Gaiella* *Marmoricola* *Segetibacter* *Ferruginibacter* *Aeromicrobium* *Sphaerobacter* *Gemmatimonas* *Iamia*	Identified Genus	*Gaiella* *Nitrosococcus* *Patulibacter* *Arthrobacter* *Nocardioides* *Streptomyces* *Aeromicrobium* *Rubrobacter* *Marmoricola* *Solirubrobacter* *Jatrophihabitans* *Chthoniobacter* *Noviherbaspirilum* *Gemmatimonas* *Nocardia*

Table[12]: Comparison of identified genera from Sanger and Illumina sequencing for 5432m site soils. Sanger data represents clone library sequences, while Illumina data is taken from TAXAplot data, specifically the sample represented in Fig[17]. Matches highlighted indicate where a genus is identified in both datasets. Soil depths are different between the two sequencing platforms due to the production of Illumina-ready amplicons being unsuccessful when utilising 3cm soil WGC as template DNA, necessitating use of the 7cm soil WGC.

Comparison between Sanger and Illumina sequence matching.

Identified genera are not identical across the datasets for each of the sequencing techniques. There are numerous possible explanations for this, but the most pertinent explanation is that they ultimately are not produced from the same original samples; the 4296m library and the 4296m TAXAplot are produced from soils at the same field site, and indeed at the same soil depth, but critically they are not from the exact same soil excavation event – they are from samples collected at different times at the same site, and therefore are not the exact same soil. While it can be reasonably expected that these soils are highly similar, they cannot truthfully be described as the same, and so some differences in bacterial community composition are to be expected. Similarly, the 5432m site datasets are taken from differing depths of the same sample site at the same time, due in part to the unsuitability of the shallower soils for production of Illumina-ready amplicons.

4.0: Discussion

4.1: Environmental Conditions

4.1.1: Temperature and Moisture

The values for air temperature displayed in Fig[3] and Table[4] indicate that the ambient air temperature at the field sites is typically lower at higher altitudes, this is possibly due to the reduction of greenhouse warming with increased altitude; by 5000m above sea level, air pressure has decreased to approximately 50% that of sea level, and thus solar heating of the air is poorly preserved at night. However, another explanation is adiabatic cooling: Lower air pressure allows the air to expand, and thus lose energy, which manifests as heat loss.

The values in Table[4] do not indicate an entirely linear progression: the temperature at the 5056m snapshot is lower than the subsequent reading at the highest altitude, which does not follow the notion of temperature having inverse correlation with altitude. This is possibly accounted for by the daily weather conditions – on the day of recording at 5056m, some cloud cover was present, occluding solar radiation and therefore solar heating, a factor not present during readings at other sites, and consequently the ambient air temperature at the site

was lower than it may have otherwise been. However, the value of 10°C recorded at the highest altitude site also appears anomalous, perhaps more so than the 5056m reading, as it was not a notably warm day, and subjectively, the air did not appear so warm, with working conditions being more difficult than those experienced at other sites during apparent lower temperatures.

An important difference for the present study is that which may be observed between air and soil temperatures at the sites. At all sites, soil surface temperature far exceeds the surrounding air temperature during the daytime, and the difference between air and soil narrows at night. As shown in Fig[4] and [5], the daily variation of temperature experienced by the soil is greatest at the surface and reduces with depth. This results from the high intensity solar radiation heating the soil surface (Table [6]). Although only the soil surface is exposed to the Sun, the top approximately 10cm are generally warmer than air temperature due to Conduction of heat from the surface to the deeper layers.

The importance of this is clear when considering the nature of the sites as a Mars surface analogue – Mars has extremely low air temperatures, which additionally exhibit a substantial diurnal variation. For example, temperatures at the Mars Pathfinder landing site were recorded to range from 190K to 270K over a single day – a variation of 80K (Ulrich et al, 2008), and with no dense atmospheric protection it experiences extreme solar irradiation (Ronto et al, 2003), which is theorized to heat the soils far in excess of the air temperatures (Savijarvi and Kauhanen, 2008). An analogous situation is observed here. The FLIR images displayed in Fig[7] are perhaps the most powerful illustration of the solar heating effect on the soils, when considered alongside the recordings of low air temperatures, and clearly depict the difference in temperatures between the soil surface and just 7cm below the surface, with the highest altitude sites featuring warm surface soils covering frozen soil 7cm below. Considering Mars once more, the recently reported periodic flows on the surface (McEwen et al, 2011) would be well explained by similar (albeit briny) ice deposits being melted in the Martian summer daytime by solar radiation, and flowing down surface slopes. At our higher altitude field locations, water in the soil was frozen around 7cm. Considering the daily temperature cycles, the frozen layer will become liquid during the day at depths where the solar heating raises the temperature above freezing point (0°C), generating daily cycles of ice and liquid water. This is significant as once the water is in liquid form, is may be lost via evaporation as a result of the drying of the soil surface by intense solar radiation.

The presence of a stable layer of water above the ice layer in the soils of the field sites – particularly the 5056m site – provides an interesting comparison with the situation on Mars, where it is believed that loss of subsurface ice is not typically via melting, but via sublimation: changing state directly from solid to vapour, which is lost to the sparse atmosphere (Bryson et al, 2008; Schorghofer and Forget, 2012). However, the discovery of Recurring Slope Lineae suggests that small volumes of liquid water do emerge from the ice layers, rather than being lost to sublimation. This moist layer of soil, below the depths of radiation-heated surface soils but above the frozen layers, would seem to be an ideal environment to search for preserved signs of life on Mars, as it is a likely environment in which microbes would be able to survive.

4.1.2: Soil Moisture

Recording soil moisture content is of interest due to the requirement of water for life – the ability of soils to retain moisture in a harsh environment such as those under investigation in this study is critical if life is to persist and function in these environments.

As depicted in Figs[10] and [11], soil moisture content values follow a similar trend to the soil temperature readings shown in Figs[4-7], in that a clear diurnal variation is evident (Fig[10]). The relative lack of soil moisture near and at the soil surface indicates that the solar heating which affects the soil temperature may also affect the soil moisture; as the soil is heated during the day, moisture is lost. Moreover, warming of the soil will also cause some of the ice layer to melt – which will lead to a band of liquid water above the ice layer, some of which will be lost to evaporation. This melting of subsurface ice is reflected in the timing of the diurnal moisture variation; Fig[10] indicates that soil moisture peaks during the daytime, and is at low point during the night. This may be explained by the subsurface ice melting during the day, allowing it to permeate though the soils more effectively than as solid ice, making it more accessible to the soil moisture probes. The reduction of soil moisture at night then represents the freezing of the moisture.

Of particular interest to the present study is the recordings for the 5056m site, the handheld data suggests that the midday soil moisture remains similar at a low level between zero and 11cm depth, wherein the soil moisture content spikes noticeably. It is with this in mind that subsequent genomic analyses will focus on comparing either side of this specific depth: to investigate similar soil environments which have different moisture contents.

4.1.3: Solar Flux

Midday was chosen for the snapshot recordings of solar radiation intensity as this during the period of the day when the solar intensity is high. Wavelengths of interest were UVA, UVB, and Photosynthetically-Active Radiation. UVC was additionally recorded, but this wavelength is entirely attenuated by the atmosphere and therefore a reading of zero was the result of all recordings. UVC is of interest because it is not attenuated by the Martian atmosphere, and is a major factor in the inhospitable nature of the surface. As it is not possible to examine the effects in the field, a UVC irradiation chamber will be used to determine the effects on collected soils. UVA and UVB were extensive at all sites. For example, a comparison with the average UVA and UVB readings is the city of Glasgow, Scotland (latitude 55.858^{o}N). Average UVA and UVB readings in Glasgow are 0.72-14.80W/m^{2} and 0.04-1.51 W/m^{2} respectively, depending on the time of year (measurements were taken during the Austral Autumn season, not in the peak solar radiation period of Summer). The expected increase in UV radiation with altitude was not recorded due to small increase in dust and cloud cover on the days that the higher altitude recordings were taken, which would occlude the solar radiation to some extent in comparison to clear skies. This trend was also evident for the PAR recordings, which are particularly lower on the 5056m site on the day of recording.

The importance of recording solar flux at these sites lies in the theory of solar heating of soils, leading to the hostile nature of these soils in terms of aridity and temperature variation, as previously discussed. Such an environment is anathema to most forms of life.

4.2 Advantages and disadvantages of large-scale NGS technologies

The principal advantage of current Next-Generation Sequencing (NGS) technologies lies in the prodigious number of highly accurate reads of samples that it is possible to generate, when compared to older technologies such as Sanger sequencing. Indeed, platforms such as the Illumina HiSeq platform have been used to verify the accuracy of stock reference genome libraries which were originally produced using Sanger sequencing (Rebolledo-Mendez et al, 2015), which speaks volumes to the progression of technology. Indeed, the Human Genome Project was the product of a mammoth investment of both time and funding (13 years and

around $3billion), and was completed using Shotgun Sanger sequencing. In comparison, a full genomic sequence on modern platforms can be typically completed in a day or two at the cost on average of a few thousand dollars (Chan et al, 2013).

The various currently-relevant NGS platforms all offer differences in performance, cost, and optimal sample requirements. In comparing the most commonly used platforms: Roche (formerly 454) Pyrosequencing and the Illumina platforms, the principal differences lie in the capacity for longer read sequences vs read number per run; Illumina platforms specialise in shorter read lengths but generate vast read counts, while Pyrosequencing platforms are suited to longer read lengths but produce fewer total reads per run (Glenn, 2011). In all NGS cases, including the relatively-fewer run counts obtained by 454 sequencing, the volume of data requires specialised data handling pipelines such as the Quantitative Insights Into Microbial Ecology (QIIME) system, which was used in this study (Caporaso et al, 2010). Such pipelines are typically designed for the purpose of screening out any incomplete reads – reads which are far shorter than the original target sequence due to incomplete sequencing. Also screened out are sample tags which fail to meet a preset read count, sample tags here referring to the short barcode sequences used to identify samples in a combined pool, and so samples which failed to produce enough successful reads to are discounted to preserve the statistical integrity of the study.

Differences such as these necessitate care in selection of an appropriate platform when planning a sequencing project. In this study, the expected read length for samples was short, as the sample sequences were comprised of a single region of the 16S gene – in this case the highly variable V4 region – and sequence lengths were determined to be approximately 400bp: a relatively short read length; it was therefore most appropriate to utilise the Illumina MiSeq platform, allowing greater statistical power through large read counts. Were this study to have been focused on full bacterial genomes, the 454 platform may have been more appropriate given its propensity for accurate reads of long bp nucleotide sequences.

4.3: DNA Extraction Concentrations

The profile of DNA yield over depth displayed in Fig[12] and Table[7] suggests that either the surface is hostile to bacteria, and supports only a limited community, or that any DNA present in the soil is degraded by the inhospitable environment or other external factors.

The fact that the DNA yield drops off after the apparent 'sweet spot' of 11-15cm indicates that the soil below this depth is not able to sustain the same size of community, and thus the amount of recovered DNA is reduced. While the total mass of recovered DNA from the sub-15cm depths is reduced in comparison to the sweet spot, the overall community diversity (considered as the number of different genera recorded in the sample) as seen in Fig[13] and Fig[14] does not markedly reduce. This reduction of mass but not diversity suggests that the deeper soil environment does not become more or less hostile in comparison to the sweet spot, but merely does not offer soil nutrients sufficient to sustain the same size of community. It is possible that any organic carbon available in the 11-15cm depths does not readily filter down to the lower depths, this may be due to the greater community at the 11-15cm depths monopolising the carbon source, or the nature of the deeper soils being frozen at the time of year recordings were taken, with the soil nutrients not able to diffuse throughout the solid water ice.

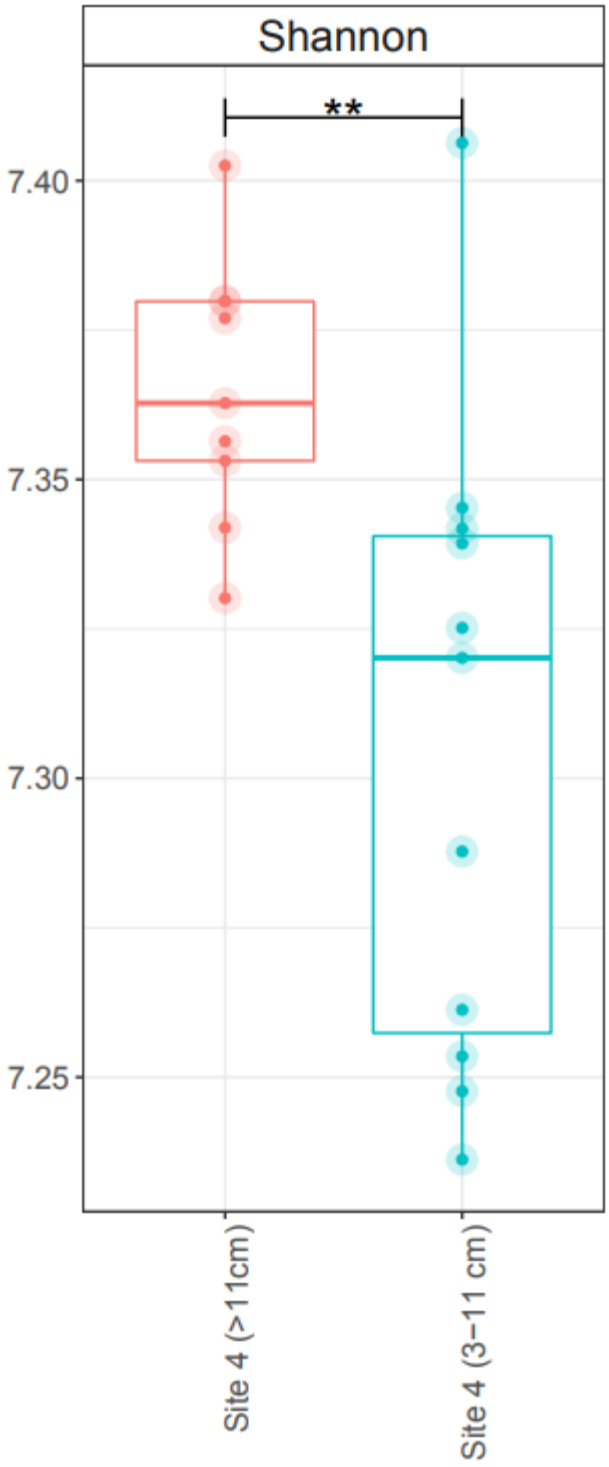
It is unlikely that a hostile environment is the cause of lost DNA as the deeper soils are buffered from the conditions experienced at the surface, as noted in this study.

Calculating Shannon Diversity

For the purpose of a more quantifiable examination of the relative differences in species alpha diversity between different soils, the Shannon Diversity Index may be implemented. Shannon diversity (H') is calculated as the negative sum of the proportional abundance (p_i) of each OTU, multiplied by the log of that proportional abundance (Hill et al, 2003), as follows:

$$H' = -\sum_{i=1}^{R} p_i \ln p_i$$

This allows direct comparison of the relative diversity between sample sets – in this case the variation between the upper (3-11cm) and lower (>11cm) soils of the 5056m site are examined. Shannon diversity is presented as a simple numerical value, with higher values theoretically representing greater diversity of species in a community. A comparison of Shannon diversity between the shallower and deeper soils of the 5056m site is shown in Fig[18] below.



Fig[18]: Displaying calculated Shannon Diversity Index values for Site 4 (5056m) soils. Soil depths are divided into 3-11cm and >11cm depths, allowing for examination of the bacterial community before and after significant soil moisture changes noted in Figs[10-11].

The soil samples of the 5056m site, as depicted in Fig[18], display a notable difference in Shannon diversity, with the deeper, more moist soils of the >11cm depths featuring greater relative species diversity than the shallower soils. This observation is corroborated by the TAXAplots of Fig[13] and [14], in which a clear trend of increase in diversity of the top 20 most abundant genera is clear in the case of both sites as depth increases, with the incidence of a dominant genus being comparatively reduced in the deeper soils. However, an important consideration must be taken into account when describing Shannon diversity here: Shannon values account for both richness and evenness of the population, which may artificially inflate the value returned. Shannon diversity offers not simply a numerical measure of diversity, but rather a statement of the likelihood of OTUs being different – although not an actual probability (Usher, 1983). In the case of this study the Shannon diversity values appear far beyond what is reported on average from ecological samples. It is possible that this is either due to inflated values due to evenness, or simply the substantial pool of potential species detected owing to the sampling strategy employed in this study – each range of soil depths accounts for a range of entire communities, potentially resulting in a significantly increased H' value.

4.4: TAXAplots

Fig[13] and [14] display the relative abundances of the top 20 most abundant taxa identified from Illumina sequencing data for the 4296m and 5056m field sites, respectively. The taxa not in the top 20 in terms of abundance occupy the white space above the columns, and thus the total abundance of the top 20 taxa is representative of that total abundance in the community.

It is evident that the top 20 most abundant taxa near the soil surface account for a greater proportion of the total community at that depth, in comparison to the most abundant taxa at lower depths. This suggests that the soil surface environment is more effectively dominated by the species able to survive and thrive in it, while the deeper soils house species more able to compete, and therefore the total proportion of the community taken up by the 20 most abundant taxa is comparatively lower. Additionally, it appears that while the relative pattern of increased diversity applies to both the 4296m and 5056m site, the 5056m site appears to

represent a smaller community overall, as even in the more diverse sub-11cm soil depths, the most abundant genera account for a larger proportion of the community – represented here by their coloured section being on average noticeably wider than those for the most abundant genera in the lower depths of the 4296m soils, and additionally represented with greater Shannon diversity. The hypothesis that the 5056m site houses an overall smaller bacterial community is reinforced by the DNA extraction yield data of Fig[12]; a lower total volume of DNA recovered implies a lower mass of bacteria in the soils.

A clear explanation for the reduced survivability of a large diversity of species at the soil surface is provided by the extreme conditions recorded in the upper depths of the soil, in particular the low soil moisture content, extensive solar heating, and high peak freeze-thaw rates experienced by the shallow soil depths offer a challenging environment for life, and thus comparatively fewer species are able to survive and thrive near the soil surface.

A further clear observation is that the community composition (i.e. the genera that are present) changes markedly at lower depths in the soil compared to the surface, reinforcing the notion that the surface and depths offer different environments due to the buffering of the hostile surface conditions described previously. Of particular interest is the fact that the most notable change in community diversity appears at a depth of approximately 11cm - the same depth at which the soil moisture at the 5056m site spikes after being relatively arid until this depth. The sudden increase in moisture and change in dominant species cannot be merely coincidental, and the moister environment clearly offers a more hospitable environment for species which cannot survive in the arid upper soil layers. Additionally, the apparent increase in diversity is corroborated by the calculated Shannon diversity values for the site.

A notably dominant genus in the upper depth of the soil is *Crossiella*. Two species of the genus have been reported: *Crossiella equi*, and *Crossiella cryophila*. *C. equi* was first isolated from horse placenta, and is believed to be related to equine nocadrioform placentitis (Donahue et al, 2002), while *C. cryophila* is a cold-tolerant species which is Aerobic, gram-positive, non-acid-fast, and non-motile (Labeda, 2001). *C. cryophila* has a growth range of 10-30^{o}C. When the environmental conditions reported in the upper soil layers and soil surface are taken into consideration, conditions are viable for *C. cryophila*, and it is therefore of little surprise that Fig[13] and [14] report the presence of *Crossiella* genera in the upper layers, which drops off after 11cm depth; This species is aerobic, and the lower layers of soil are less aerated than the upper layers. The species is cold-tolerant, which is applicable to all

layers of the soils to some extent. Additionally, the species is non-motile. The presence of a non-motile species in the upper layers offers support to the predictions by the Tax4Fun script, which indicate that proteins related to motility were relatively upregulated in the lower depths of the soil, not the upper depths. Additionally, the upper layers of soil were found to be significantly less moist than the lower layers, which would preclude the presence of highly motile species, and promote the presence of static species, which is reinforced by the presence of *C. cryophila*.

The genus *Blastococcus* was additionally abundant in the upper layers of the soil at both sites. A typically aerobic genus, and one which (depending on the species) may be motile or non-motile (Urzi et al, 2004), it is of little surprise that *Blastococcus* species have were most abundant in the shallow soils, with prevalence receding at the lower depths.

The LCBD values for each soil layer community are depicted in Fig[13] and [14]. At both the 4296m site and the 5056m site, the greatest LCBD values were observed in the hostile upper layers of the soil, near the soil surface. LCBD weight of the soil communities is reduced in the lower layers, including around the 11cm "sweet spot" and deeper. This is interesting, as at first glance it would seem unusual that the comparatively smaller, less diverse community present in these upper layer soils applies the greatest effect on the overall diversity of the soils in the site.

A possible explanation for the upper soils having higher LCBD values is that the species in the upper layers have importance in providing soil nutrients, such as providing a source of organic carbon or filtering soil nitrates down to the deeper layers, becoming available for use by the deeper community.

4.5: Tax4Fun Script Output

When examining the output of Tax4Fun regarding predictions of relative enrichment of proteins in soil samples, there are considerations to be taken regarding the functionality of the script, and the libraries from which matches are drawn. In short, Tax4Fun was developed with focus on clinical applications – pathogenic species, metabolic and immunologically-important pathways etc. The possible consequence of this initial focus on clinical applications

is that the library of matchable proteins to report may be subject to unintentional bias due to species/pathways/enzymes which are not typically under scrutiny in clinical studies being under represented in the Tax4Fun libraries. Such underrepresentation could result in proteins which should have exhibited prominence in the present study being absent or replaced by matches which are similar yet inaccurate. Library issues such as this may well account for reports such as the "Hypothetical Protein" presented as Unknown (#12) in Table[8] – this protein may well be commonly expressed yet not generally applicable in clinical settings.

4.4.1: Discussion of top 20 significantly upregulated proteins predicted by Tax4Fun

#1: Osmotically inducible lipoprotein OsmE

OsmE is primarily involved in reduction of soil nitrates, a vital step in the Nitrogen Cycle. Of particular interest is that OsmE is inducible in response to low osmotic pressure (Gutierrez et al, 1995). OsmE was predicted by Tax4Fun to be upregulated in the lower depths of the soil, which correlates with the increase of soil moisture content described previously. The increased moisture content of the soils below 11cm depth would result in reduction of osmotic pressure, and it is speculated here that this could induce increased expression of OsmE.

#2: Serine/threonine−protein kinase PpkA

PpkA is a protein kinase – an enzyme which phosphorylates a protein to provide activation energy for the target protein – which in this case targets specifically the Serine and Threonine amino acid residues in the protein. This protein has been reported to catalyze the T6 Secretion System in pathogenic bacteria, which is instrumental in interactions between pathogen and host (Hsu et al, 2009). Of greater relevance to this study is the fact that PpkA is implicated in cellular resistance to oxidative and osmotic stresses. PpkA was predicted via Tax4Fun to be upregulated in the lower depths of the soil relative to the near-surface, and as is the case regarding OsmE, it is anticipated that this could be in response to a reduction of osmotic pressure in the lower depths.

#3: Thiol:disulfide interchange protein DsbG

DsbG is a chaperone protein, involved in the regulation of protein folding, which is vital in producing the correct protein structure for function (van Straaten et al, 1998). DsbG is also instrumental in the stabilisation of proteins in the cell periplasm (Shao et al, 2000), which is prevalent most notably in gram-negative bacteria. DsbG is upregulated in the lower depths of the soil. Taking into consideration the fact that the lower, moister depths of the soil are likely to remain frozen in winter, it is possible that the increased expression of DsbG acts in response to this, to help stabilise the cell morphology.

#4: Xanthine phosphoribosyltransferase (XPRT)

XPRT catalyses the breakdown of Xanthine molecules for use in the production of new nucleic acids. In *B.subtilis*, Xanthine can be used as the sole Nitrogen source for the bacteria (Christiansen et al, 1997). Xanthine itself is a product of nucleic acid breakdown, so it is apparent that XPRT has importance in the recycling of damaged nucleic acids. Damage to nucleic acids is a common consequence of hostile environmental conditions, for example stress from radiation damage or dessication due to hyperaridity or hypersalinity leading to osmotic stress, so it is logical that such damage is found in the soils of interest to this study. XPRT was predicted to be upregulated in the lower depths of the soil, which is unexpected as conditions likely to cause desiccation were found at the surface, not at lower depths, and so it is possible that this upregulation may be in response to the freezing effect the lower, moister soils experience during the winter.

#5: Heptosyltransferase I

Heptosyltransferase I is primarily involved in the production of bacterial LPS, commonly associated with Gram-negative bacteria. LPS is important for cell motility, and additionally has a role in antibiotic resistance by hindering antibiotic uptake by the cell (Czyzyk et al, 2013). In loss mutants, it had been observed that production of biofilms is increased (Wang et al, 2012), this may be the result of reduction in LPS leading to decreased cell motility. Heptosyltransferase I was upregulated in the lower depths of the soil. It is speculated that the increase in Heptosyltransferase may be due to the antibiotic resistance-aiding action of LPS. Several proteins predicted in the Tax4Fun top 20 list are related to pathogenic action, and this increased antibiotic resistance may be centred around pathogenic species, as a means of

defence against antibiotic action from other community species. This may also be a consequence of the potential clinical bias inherent in Tax4Fun.

#6: Twitching motility protein PilJ

PilJ is involved in the construction of cell Pilli. An important function of pilli is the facilitation of cell motility (Darzins, 1994). Cell motility will be facilitated by the presence of water in the soil. Overall, the soils contain greater moisture with depth and thus it would follow logically that this protein may be enriched in the lower depths of the soil.

#7: Enhanced entry protein EnhC

EnhC carries the function of facilitating pathogenic entry of bacteria into host cells (Cirillo et al, 2000), such as in *L.pneumophila*. The increased presence of this protein in the lower depths of the soil would suggest that part of the bacterial community is pathogenic, acting parasitically upon other community members. It is unclear, however, why this would be facilitated by the conditions present at lower depths. Potentially the more extreme conditions at shallower depths promotes a community composition which does not support parasitic relationships, or at which the pathogenic species are unable to survive the environmental stresses. Another possibility lies in the aforementioned issues with the Tax4Fun library.

#8: MSHA biogenesis protein MshM

MshM is functional in the activity of cell Pilli, and is predicted to act as an ATPase, providing energy for the action of cellular pilli (Jones et al, 2015). As previously mentioned, the increased prevalence of cell motility in environments of greater moisture levels correlates with the potential upregulation of this protein in the lower levels of the soil.

#9: Transposase

Transposases are proteins which catalyse gene transposition: the excision of genes from the genome and their reinsertion elsewhere in the genome, a common metabolic activity. Of particular interest here is that transposition is a noted mechanic behind bacterial antibiotic resistance (Harmer and Hall, 2016). Antibiotic resistance is not commonly considered in wild communities (being more a concern in clinical settings), but antibiotic action may be present as a defence mechanism in the bacterial community, which would necessarily promote resistance adaptation by any pathogenic bacteria in the community. Considering the upregulation of EnhC, which would facilitate a greater incidence of pathogenic action at this

depth in the soil, a concurrent increase in antibiotic resistance action would be logical, and this is supported by the observation of the upregulation of Transposase in the lower soil depths.

While it is unclear why this process would be upregulated at depth in the soil, such a discovery may be worth further investigation, in particular with regards to Martian sample return missions. If particular Mars-like environments generate increased antibiotic resistance, then returning samples from these locations could provide an increased biohazard. As a result, investigating which particular Mars-like environments upregulate antibiotic resistance (and the specific nature of this resistance) could be used to enhance biosafety procedures for sample return missions.

#10: Deoxyribonuclease I

Deoxyribonuclease (DNase) I is an enzyme which cuts DNA sequences. DNA is cut for a wide range of functions such as cell replication, gene transcription, or transposition. DNase I is upregulated in the lower depths of the soil. It is possible that this upregulation is related to the upregulation of transposase, and has implications in the antibiotic resistance of pathogenic community members.

#11: Two−component system, NarL family, nitrate/nitrite sensor histidine kinase NarQ

NarQ is a sensor/signal transducer protein which is involved in nitrate reduction (Chiang et al, 1992), which functions as part of the environmental Nitrogen cycle to maintain soil nutrient availability. NarQ was upregulated in the upper depths of the soil.

#12: Hypothetical protein

No mention of the protein predicted by Tax4Fun has been reported in previous studies, and therefore it is not possible to speculate on the exact function or even the name of the protein predicted. It is possible that the predicted protein was generated by the script using a combination of homologous proteins across several species which were detected, but regardless this prediction is of no use to the present study.

#13: Protease secretion protein HasE

HasE is a component of the metalloprotease secretion system in species such as P.*aeruginosa* (Akatsuka et al, 1997), which functions in the degradation of extracellular proteins.

Upregulated in the lower depths of the soil, it is possible that this extracellular protein degradation is part of a pathogenic system, targeted at breaking down bacterial defences such as cell walls, spores or LPS layers. It is speculated here that the higher moisture content and different bacterial community assemblage facilitates pathogenic interactions between bacteria in this part of the soil. Degredation of extracellular protein may also have functionality in the simple usage of proteins in the extracellular matrix.

#14: Ech hydrogenase subunit E

Ech hydrogenase is functional in the action of methanogenic species, in particular it is associated with *Methanosarcina barkeri* (Meuer et al, 1999). This is an unexpected finding for this study, as during the process of amplicon production via PCR, archeal-specific primers were not used, only bacterial 16S-specific primers were used. This suggests that the protein predicted via tax4fun may not be Ech subunit E, but a bacterial protein which shares homology with EchE, but is as yet unidentified. It could be feasible to suggest that the unnamed protein is involved in methanogenic action by bacterial species.

#15: Type IV pilus assembly protein PilP

PilP functions in the production of cell Pilli, involved in cell motility and adherence (Martin et al, 1995). Enriched in the lower depths of the soil, the increased prevalence of this protein correlates with the increased soil moisture as well as the increased prevalence of other motility-related proteins such as PilJ and MshM.

#16: Gluconate dehydratase

Gluconate dehydratase functions in the degradation of glucose and gluconate (Eisenberg and Dobrogosz, 1967). Enriched in the upper depths of the soil, it is possible that such enzymes function in the utilisation of environmental nutrients which may be more accessible in the moister, deeper soils.

#17: Membrane−bound serine protease (ClpP class)

ClpP-class membrane-bound serine proteases are enzymes which break down proteins, targeting specifically at serine amino acid residues, these are highly conserved across species, including most bacterial and in the mitochondria of mammalian cells (Liu et al, 2014). Upregulated in the lower depths of the soil.

#18: Avirulence protein

Avirulence (Avr) proteins are commonly expressed in plants, as a means of defense against pathogenic action. Avr genes are actually additionally expressed by select pathogenic species; a situation wherein plant and pathogen express the same Avr gene is actually beneficial to the pathogen, promoting colonisation (Dehury et al, 2015). Avr was upregulated in the lower depths of the soil and appears to be related to the potential pathogenic subcommunity in the lower depths.

#19: Ech hydrogenase subunit F

The F subunit of Ech hydrogenase was upregulated in the lower depths of the soil. Prevalent in methanogenic genera such as *Methanosarcina*, this protein is an NADH dehydrogenase, which are central to the chain reaction of oxidative phosphorylation – the primary ATP production system of the cell. Ech is involved in energy conservation in the cell (Welte et al, 2010b). Ech mediates the transfer of electrons into the respiratory chain via reduced Ferredoxin, and reduced growth has been noticed in Ech loss mutants (Welte et al, 2010a).

#20: Universal stress protein A

Universal Stress Protein A (UspA) is functional in the stress response in cells, which may be triggered by environmental stressors such as extremes of temperature, osmotic shock, damage via radiation, mechanical damage, or desiccation. UspA was upregulated in the lower depths of the soil; an unexpected observation as environmental hostility appears greater in the upper depths and the soil surface than in the lower depths. Expression of UspA has been reported to increase significantly under cold stress, displaying a greater increase in expression in response to cold stress than in response to oxidative or starvation stresses (Gawande and Griffiths, 2004), and so it is likely that the upregulation predicted in the lower depths is in response to freezing stress during winter months.

4.5: Relating soil conditions to Tax4Fun output

4.5.1: Prevalence of motility-related proteins

A readily observable correlation is evident between the recorded soil moisture content and the increased prevalence of genes related to cell motility and adherence. The designations "upper" and "lower" for soil depths – to which the top 20 most significantly differently upregulated proteins are assigned – are based on the depth in the soil at 5056m that soil moisture noticeably peaks, which is 11cm, as shown in Fig[11].

Twitching mobility protein (PilJ) is an example of the proteins which are upregulated noticeably in the lower depths of the soil, as displayed in Table[8]. PilJ was the 6^{th} most significantly differently upregulated protein according to the Tax4Fun script predictions, and was relatively upregulated in the lower depths of the soil, at which depth the soil featured high water content. If the soil has a higher moisture content, this will facilitate the movement of cells via the action of cell pili, and therefore it follows that the prediction from the Tax4Fun script indicating that increased prevalence would be noted of proteins such as PilJ, which is instrumental in the construction of pili.

4.5.2: Stress response proteins

The increased prevalence of proteins related to the response and resistance to environmental stress factors in the lower depths of the soil is on first glace somewhat unexpected – the data recorded concerning environmental conditions suggest that the upper depths of the soil are subject to more hostile conditions than the lower depths; increased freeze-thaw rate (shown in Fig[6]), greater diurnal temperature variation (shown in Fig[5]), reduced overall soil moisture content (shown in Fig[11]), increased UV radiation from solar flux which does not affect the lower depths as UV does not penetrate more than a few mm into soil. It is therefore unexpected that the Tax4Fun script predicts that the extracted genomic material from the lower soil depths is richer in these stress response proteins than the material recovered from the near surface layers.

4.5.3: Proteins related to pathogenic action and resistance to pathogens

Of particular note amongst the predictions generated by the Tax4Fun script are a number of proteins which are related to pathogenic action or resistance/defence against pathogens. Based on this relevant prevalence of pathogenesis-related genes, it appears that there is a subcommunity of species in these soils which are pathogenic, acting antagonistically to the community at large. Consideration must also be applied to the fact that the output from the Tax4Fun script is a prediction based on detected taxa, rather than a direct functional gene assay via a protocol such as construction of a cDNA library, and so it must be considered that errors in predictions are possible. The possibility of errors in prediction are highlighted by the predicted upregulation of Ech hydrogenase subunit E – an archeal protein, which is not detectable in the current study as archeal primers were not used during amplicon production – only bacterial primers were used.

4.6: Sanger sequencing pilot study

In this study, the Sanger sequencing project served as a pilot study, designed to provide an insight into the likely bacterial communities of the analogue site, and to determine whether the extracted DNA from the soils would be suitable for sequencing in general. Sanger is an appropriate technique for such pilot study due in large to the relative speed, simplicity, and low cost of preparation and analysis involved in a small-scale Sanger project; Illumina amplicons are subject to stringent sample preparation requirements, and additionally the PCR amplification process itself is more sensitive to costly difficulties in PCR reaction optimisation, as failing to successfully incorporate the barcode tags and Illumina adapter sequences into the sample amplicons will result in those sequencing reactions failing, while Sanger samples are simply produced with standard PCR amplification.

Similarity of identified genera

If one considers Illumina sequencing output to be the gold standard (for this study – other studies may have differing requirements), it is interesting to scrutinise how effective and

accurate Sanger sequencing is in determining bacterial community composition and proportions.

As shown in Fig[16] and Fig[17], the two platforms did not return identical community compositions, but there are similarities. Tables[11] and [12] directly highlight genera which were identified in both Sanger and Illumina sequence matching, with genera such as *Gaiella* and *Streptomyces* being common to both datasets. While this is a point in favour of the accuracy of Sanger, it should be considered that these are commonly widespread genera and so commonality here is not unexpected. In the case of both Fig[16] and Fig[17], it is evident that while *Gaiella* is detected in both sequence libraries, the proportion of the community constituted of *Gaiella* genera is not the same in each dataset; *Gaiella* are the most abundant genus in the Sanger data, but this is not the case in the Illumina data. This is possibly the result of a combination of factors: Limited community structure in the Sanger library due to the colony selection and product purification process, set in comparison to the 'whole community' source of the Illumina amplicons, and additionally the increased sensitivity of Illumina returning a larger variety of matched sequences, thereby reducing the overall proportion of the community accounted for by *Gaiella*.

A source of potential differences in identified sequences lies in the methods of obtaining the DNA to be sequenced: Sanger sequenced DNA was obtained from colonies of transformed *E.coli*, while Illumina amplicons were produced directly taking PCR template DNA from the initial WGC extract. The clone libraries were taken from 96 picked colonies, but the total number of colonies from which to pick far exceeded this number, and as a result it is likely that some genera were missed out in this semi-random selection process. Further, it is possible that some clones simply could not grow on LB-agar plates, and thus could never be picked for inclusion in the library. Indeed, many species cannot be successfully cultured *in vitro* at all.

Similarities between the sets of sequencing data may be indicative of more prevalent genera in these soils, as despite the differences in sampling, they are still identified by both sequencing techniques. Examples of such prevalent genera are *Gaiella*, and, *Streptomyces*, and *Nocardioides*, which appear frequently in all sequenced samples, *Gaiella* in fact appears in all 4 sets of sequences in the comparison, suggesting the genus to be relatively ubiquitous across these sites in general. *Streptomyces* is known to be a wide-ranging genus that is

common in most soils, so it is of little surprise that this genus would be present in most sequenced soil communities.

The range of detection of Illumina is shown to be considerably greater in this study than that of Sanger, as is highlighted by the fact that the compared Illumina-detected genera here represent only the top 20 most abundant genera identified in the sample, and not the entirety of the identified community. Genera unmatched in the comparison Tables[11] and [12] are accounted for in the Co-occurrence network maps in *Chapter 2 (Figs[1-4])*, indicating that they were indeed identified in the Illumina sequencing, but were not prevalent enough to be included in the top 20 most abundant genera for the purposes of the TAXAplots.

It is vital to consider the difference in samples when assessing comparison of sequences matched via Sanger and Illumina in this study – being of nonidentical origin renders a truely accurate comparison impossible, however the similarities in major identified genera such as *Gaiella* offer some further validation of the findings of Illumina sequencing, and are thus worthy of note here. In order to achieve a true direct comparison, a single sample should be homogenised and split equally between the two platforms. The value of performing such a comparison on a large scale (i.e. all samples represented in this study) is perhaps questionable, and thus Sanger was used here merely as a pilot study during early structuring work for the study at large.

5.0: Conclusions

5.1: Altiplano sites as a Mars analogue

The environmental conditions recorded in this study, at the high altitude sites on the arid slopes of Cerro Sairecabur, and in particular at the highest altitude sites – higher than 5000m above sea level – describe a region which is viable for use as a terrestrial Mars analogue. Data presented in this study describes an environment subject to extensive solar radiation, substantial diurnal variation in temperature, in particular of the soil surface, which is subject to solar heating. The solar heating of the soil results in a surface which is heated far in excess of the typical ambient air temperature; this phenomenon is also reported at the Martian

surface, and is a contributing factor to the inhospitable nature of the Martian regolith, as this results in an extremely wide diurnal variation of temperature at the surface which imparts freeze-thaw rates too great for life to survive, and additionally contributes to the lack of stability of liquid water on the surface of Mars.

In addition to the solar heating, the sites offer homology with theorised conditions on Mars in regards to shallow subsurface ice deposits which are believed to be present on Mars, and are a possible explanation for the Recurring Slope Lineae reported by McEwen et al (2011). The sites reported in this study similarly displayed near-surface ice, best evidenced by the thermal images in Fig[7].

5.2: Effect of soil coverage on microbiology of extreme environments

The data displayed in the TAXAplots (Fig[13] and [14]) and DNA extract yields (Fig[12]) indicate that a hostile surface environment supports a community of reduced diversity and overall biomass than that which is supported below the surface in deeper layers of the soil. This is evidenced by the increased DNA extract yield recovered from the deeper soils, and the increased diversity of identified genera in the community, as was corroborated with increase in Shannon index. It is further evident that in such environments, stress response and resistance are more abundant, and environmental stressors are prevalent even in the comparatively less hostile deeper soils.

5.3: Implications for search for life on Mars

This study presents a case that the recovery of extant biological material from the Martian surface is unlikely, due to damaging conditions on the surface in general, and the reduced biomass recoverable even in the comparatively hospitable soils of the Martian analogue sites discussed – while these are hostile environments on Earth, they do not completely mirror the conditions on Mars, and this must be considered when planning sampling missions.

While it is unlikely that organic material may be recovered from the surface of Mars itself, the environmental data reported here indicate that conditions become more hospitable rapidly below the soil surface – for example the effect of UV radiation is mitigated by only a few mm of soil coverage, and ice deposits are able to exist merely inches below a surface subjected to extreme solar heating. With this in mind, it is more feasible that identifiable evidence of extinct Martian microbial life could be preserved below the surface, yet within reach of sampling drills such as that which is planned for use on the upcoming (2020) ExoMars Rover.

6.0: References

Akatsuka, H., Binet, R., Kawai, E., Wandersman, C., Omori, K. (1997). Lipase Secretion by Bacterial Hybrid ATP-Binding Cassette Exporters: Molecular Recognition of the LipBCD, PrtDEF, and HasDEF Exporters. Journal of Bacteriology Vol 179. P4754-4760.

Bryson, K.L., Chevrier, V., Sears, D.W.G., Ulrich, R. (2008). Stability of ice on Mars and the water vapor diurnal cycle: Experimental study of the sublimation of ice through a fine-grained basaltic regolith. Icarus Vol 196. P446-458

Cabrol, N.A., Feister, U., Hader, D-P., Piazena, H., Grin, E.A., Klein, A. (2014). Record solar UV irradiance in the tropical Andes. Frontier in Environmental Science Vol 2.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley., G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. Nature Methods Vol 7. P335-336.

Chan, M., Lee, C.W-H., Wu, M. (2013). Integrating next-generation sequencing into clinical cancer diagnostics. Expert Review of Molecular Diagnostics Vol 13. P647-650.

Chiang, R.C., Cavicchioli, R., Gunsalus, R.P. (1992). Identification and characterization of narQ, a second nitrate sensor for nitrate-dependent gene regulation in *Escherichia coli*. Molecular Microbiology Vol . P1913-1923.

Christiansen, L.C., Schou, S., Nygaard, P., Saxild, H.H. (1997). Xanthine Metabolism in *Bacillus subtilis*: Characterization of the xpt-pbuX Operon and Evidence for Purine- and Nitrogen-Controlled Expression of Genes Involved in Xanthine Salvage and Catabolism. Journal of Bacteiology Vol 179. P2540-2550.

Cirillo, S.L.G., Lum, J., Cirillo, J.D. (2000). Identification of novel loci involved in entry by *Legionella pneumophila*. Microbiology Vol 146. P1345-1359.

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., Tiedje, J.M. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Research Vol 42 (database issue). D633-D642.

Czyzyk, D.J., Sawant, S.S., Ramirez-Mondragon, C.A., Hingorani, M.M., Taylor, E.A. (2013). *Escherichia coli* Heptosyltransferase I: Investigation of Protein Dynamics of a GT-B Structural Enzyme. Biochemistry Vol 52.

Daly, M.J., Ouyang, L., Fuchs, P., Minton, K.W. (1994). In vivo damage and recA-dependent repair of plasmid and chromosomal DNA in the radiation-resistant bacterium Deinococcus radiodurans. Journal of Bacteriology Vol 176. P3508-3517.

Darzins, A. (1994). Characterization of a *Pseudomonas aeruginosa* gene cluster involved in pilus biosynthesis and twitching motility: sequence similarity to the chemotaxis proteins of enteric and the gliding bacterium *Myxococcus Xanthus*. Molecular Microbiology Vol 11. P137-153.

Dehury, B., Maharana, J., Sahoo, B.R., Sahu, J., Sen, P., Modi, M.K., Boorah, M. (2015). Molecular recognition of avirulence protein (avrxa5) by eukaryotic trainscription factor xa5 of rice (Oryza sativa L.): Insights from molecular dynamics simulations. Journal of Molecular Graphics and Modelling Vol 57. P49-61.

Donahue, J.M., Williams, N.M., Sells, S.F., Labeda, D.P. (2002). *Crossiella equi* sp. nov., isolated from equine placentas. International Journal of Systematic and Evolutionary Microbiology Vol 52. 2169-2173.

Du, J., Gebicki, J.M. (2004). Proteins are major initial cell targets of hydroxyl free radicals. International Journal of Biochemistry and Cell Biology Vol 36. P2334-2343.

Eisenberg, R.C., and Dobrogosz, W.J. (1967). Gluconate Metabolism in *Escherichia coli*. Journal of Bacteriology Vol. 941-949.

Gawande, P.V., and Griffiths, M.W. (2004). Effects of environmental stresses on the activities of the uspA, grpE and rpoS promoters of *Escherichia coli* O157:H7. International Journal of Food Microbiology Vol 99. P91-98.

Gierasch, P., and Goody, R. (1972) The effect of dust on the temperature of the Mars atmosphere. Journal of Atmospheric Science Vol 29. P400-402.

Glenn, T.C. (2011). Field guide to next-generation DNA sequencers. Molecular Ecology Resources Vol 11. P759-769.

Gringauz, K.I. (1991). What was known about the Martian magnetosphere before Phobos-2 mission. Planetary and Space Science Vol 39. P73-74.

Gutierrez, C., Gordia, S., Bonnassie, S. (1995). Characterization of the osmotically inducible gene *osmE* of *Escherichia coli* K-12. Molecular Microbiology Vol 16. P553-63.

Harmer, C.J., and Hall, R.M. (2016). IS26-Mediated Formation of Transposons Carrying Antibiotic Resistance Genes. mSphere Vol 1.

Hsu, F., Schwarz, S., Mougous, J.D. (2009). TagR Promotes PpkA-Catalyzed Type IV Secretion Activation in *Pseudomonas aeruginosa*. Molecular Microbiology Vol 72. P1111-1125.

Jakosky, B.M., and Philips, R.J. (2001). Mars' volatile and climate history. Nature Vol 412. P237-244.

Jones, C.J., Utada, A. ,Davis, K.R., Thongsomboon, W., Sanchez, D.Z., Banakar, V., Cegelski, L., Wong, G.C.L., Yildiz, F.H. (2015). C-di-GMP Regulates Motile to Sessile Transition by Modulating MshA Pili Biogenesis and Near-Surface Motility Behaviour in *Vibrio cholerae*. PLOS Pathogens Vol 11.

Labeda, D.P. (2001). *Crossiella* gen. nov., a new genus related to *Streptalloteichus*. International Journal of Systematic and Evolutionary Microbiology Vol 51. P1575-1579.

Leovy, C. (2001). Weather and Climate on Mars. Nature Vol 412. P245-249.

Liu, K., Ologbenia, A., Houry, W.A. (2014). Dynamics of the CliP serine protease: A model for self-compartmentalized proteases. Critical Reviews in Biochemistry and Molecular Biology Vol 49. P400-412.

Lundin, R. (2001). Erosion by the Solar Wind. Science Vol 291. P1909-1909.

Ma, Y., Vam Dam, R.L., Jayawickreme, D.H. (2014). Soil moisture variability in a temperatre deciduous forest: insights from electrical resistivity and throughfall data. Environ' Earth Sci' Vol 72. P1367-1381.

Mancinelli, R.L., Klovstad, M. (2000). Martian soil and UV radiation: microbial viability assessment on spacecraft surfaces. Planetary and Space Science Vol 48. P1093-1097.

Martin, P.R., Watson, A.A., McCaul, T.F., Mattick, J.S. (1995). Characterization of a five-gene cluster required for the biogenesis of type 4 fimbriae in *Pseudomonas aeruginosa*. Molecular Microbiology Vol 16. P497-508.

McEwen, A.S., Ojha, L., Dundas, C.M., Mattson, S.S., Byrne, S., Wray, J.J., Cull, S.C., Murchie, S.L., Thomas, N., Gulick, V.C. (2011). Seasonal Flows on Warm Martian Slopes. Science Vol 333. P740-743

Mellon, M.T., Feldman, W.C., Prettyman, T.H. (2004). The presence and stability of ground ice in the southern hemisphere of Mars. Icarus Vol 169. P324-340.

Meuer, J., Bartoschek, S., Koch, J., Kunkel, A., Hedderich, R. (1999). Purification and catalytic properties of Ech hydrogenase from *Methanosarcina barkeri*. European Journal of Biochemistry Vol 265. P325-335.

Munteanu, A.C., Uivarosi, V., Andries, A. (2015). Recent progress in understanding the molecular mechanisms of radioresistance in Deinococcus bacteria. Extremophiles Vol 19. P707-719.

Piazena, H. (1996). The effect of altitude upon the solar UV-B and UV-A irradience in the tropical Chilean Andes. Solar Energy Vol 57. P133-140.

Rebolledo-Mendez, J., Hestand, M.S., Coleman, S.J., Zeng, Z., Orlando, L., MacLeod, J.N., Kalbfleisch, T. (2015). Comparison of the Equine Reference Sequence with Its Sanger Source Data and New Illumina Reads. PLoS ONE Vol 10.

Rodriguez, J.A.P., Kargel, J.S., Baker, V.R., Gulick, V.C., Berman, D.C., Fairen, A.G., Linares, R., Zarroca, M., Yan, J., Miyamoto, H., Glines, N. (2015). Martian outflow channels: How did their source aquifers form, and why did they drain so rapidly? Sci. Rep. 5, 13404; doi: 10.1038/srep13404

Ronto, G., Berces, A., Lammer, H., Cockell, C.S., Molina-Cuberos, G.J., Patel, M.R., Selsis, F. (2003). Solar UV Irradiation Conditions on the Surface of Mars. Photochemistry and Photobiology Vol 77. P34-40.

Sanger, F., Nicklen, F., Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. PNAS Vol 74. P5463-5467.

Savijarvi, H., and Kauhanen, J. (2008). Surface and boudary-layer modelling for the Mars Exploration Rover sites. RMetS Vol 134. P635-641.

Schorghofer, N,. and Forget, F. (2012). History and anatomy of subsurface ice on Mars. Icarus Vol 220. P1112-1120.

Shao, F., Bader, M.W., Jakob, U., Bardwell, J.C.A. (2000). DsbG, a Protein Disulfide Isomerase with Chaperone Activity. Journal of Biological Chemistry Vol 275. P13349-23352.

Siam, F.M., Grinfeld, M., Bahar, A., Rahman, H.A., Ahmad, H., Johar, F. (2018). A mechanistic model of high dose irradiation damage. Mathematics and Computers in Simulation Vol 151. P156-168.

van Straaten, M., Missiakas, D., Raina, S., Darby, N.J. (1998). The functional properties of DsbG, a thiol-disulfide oxidoreductase from the periplasm of *Escherichia coli*. FEBS Vol 428. P255-258.

Vasileiadis, S., Puglisi, E., Arena, M., Cappa, F., Cocconcelli, P.S., Trevisan, M. (2012). Soil Bacterial Diversity Screening Using Single 16S rRNA Gene V Regions Coupled with Multi-Million Read Generating Sequencing Technologies. PLoS ONE Vol 7.

Ulrich, R., Pilgrim, R., Chevrier, V., Roe, L., Kral, T. (2008). Temperature fields at Mars landing sites: Implications for subsurface. Lunar and Planetary Science Vol 39.

Urzi, C., Salamone, P., Schumann, P., Rohde, M., Stackebrandt, E. (2004). *Blastococcus saxobsidens* sp. nov., and emended descriptions of the genus *Blastococcus* Ahrens and Moll 1970 and *Blastococcus aggregatus* Ahrens and Moll 1970. International Journal of Systematic and Evolutionary Microbiology Vol 54. P253-259.

Wang, L., Hu, X., Tao, G., Wang, X. (2012). Outer membrane defect and stronger biofilm formation caused by inactivation of a gene encoding for heptosyltransferase I in *Cronobacter sakazakii* ATCC BAA-894. Journal of Applied Microbiology Vol 112. P985-997.

Welte, C., Kallnik, V., Grapp, M., Bender, G., Ragsdale, S., Deppenmeier, U. (2010). Function of Ech Hydrogenase in Ferredoxin-Dependent, Membrane-Bound Electron Transport in *Methanosarcina mazei*. Journal of Bacteriology Vol 192. P674-678.

Welte, C., Krätzer, C., Deppenmeier, U. (2010). Involvement of Ech hydrogenase in energy conservation of *Methanosarcina mazei*. The FEBS Journal Vol 277. P3396-6403.

Wierzchos, J., DiRuggiero, J., Vitek, P., Artieda, O., Souza-Egipsy, V., Skaloud, P., Tisza, M., Davila, A.F., Vilchez, C., Garbayo, I., Ascaso, C. (2015). Adaptation strategies of endolithic chlorophototrophs to survive the hyperarid and extreme solar radiation environment of the Atacama Desert. Frontiers in Microbiology Vol 6.

Whitaker, R.H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. Ecological Monographs Vol 30. P279-338.

Referenced websites:

Fig[1]: dNTP/ddNTP images adapted from
https://www.ocf.berkeley.edu/~edy/genome/sequencing.html

Chapter 2: Examining the Bacterial Community of the Altiplano Martian Analogue

1.0: Introduction

1.1 Use of bioinformatics in environmental microbiology

Modern massively-parallel sequencing techniques are able to generate a prodigious amount of data via repeat runs of sample sequencing. The substantial data volumes produced by the use of such platforms are valuable for studies as they allow larger sample groups and thus more reliable statistical analyses. When examining the bacterial community of an environmental sample – or indeed the fungal, eukaryotic microbial, or plant communities – even a specific sequencing project such as a singular gene (as is the case in this study) necessitates the ability to process large data volumes. The processing of data is streamlined by the standardisation of provided sequence data; many NGS services provide sequence data in the form of FASTQ files.

1.1.2 Use of processing pipelines

The processing pipeline for NGS data is important for 'cleaning' the data of incorrect or incomplete sequences, as well as later manipulation and presentation of the data. QIIME pipelines were utilised in this study, as described by Navas-Molina et al (2013). Processing of sequencing data can be loosely separated into upstream and downstream processes. Upstream processes begin with demultiplexing of samples; reassigning the pool of sequences to the original sample labels (usually grouped by barcode sequences, and subsequent actions include trimming of primer and barcode sequences from the final sequence, removal of chimeric or incomplete (fewer bp than the original sequence) sequences, as well as samples with insufficient complete reads (variable, but generally more than 10^4 reads is suggested as a minimum). Following trimming of all incomplete, replicate, or singleton reads, a final total of

184,686 reads were used for OTU table construction. The most important outcome of upstream processing is the generation of an OTU table and phylogenetic tree. Operational Taxonomic Units are sequences grouped together due to similarity, traditionally 97% sequence identity is the threshold chosen for OTU grouping (Drancourt et al, 2000). OTU generation offers a choice in strategy; *de novo* generation or database-referenced, with the latter making use of online databases such as the BLAST database to rapidly identify similarity in sequences. In this study, *de novo* OTU picking was followed by DB-picking for greater accuracy, resulting in a final unique OTU count of 8472.

Downstream processes involve manipulation and presentation of the data, allowing graphical representation of available taxa, such as in the TAXAplot format presented in the previous chapter and examination of alpha and beta diversity in bacterial communities. Downstream presentations presented in this chapter include the output of the Seqenv programme, plotting of co-occurring genera in each sample set according to correlation between the presence of each genus and the presence of each other genus in the sample. Graphical representation of data is commonly supported by programmes such as R. A wide range of packages have been developed which are designed to work with QIIME-processed data, allowing for streamlined workflow.

1.2 Seqenv Matching

The handling of large data volumes produced by modern high throughput sequencing technologies such as NGS sequencing on platforms like Illumina – is an increasingly pertinent issue for bioinformaticians. The seqenv programme was developed by Sinclair et al (2016) to attach environmental relevance to sequence data. Seqenv runs similarity searches through the NCBI nucleotide database, and where possible extracts stated isolation sources from the associated metadata from the database. The output of this search is presented as a weighted map of environments associated with a set of sequence data, and allows weighting of particular environments by the prevalence of them in relation to the sequence data; environments which are frequent isolation sources for the identified taxa are displayed more prominently in the graphical representation of the output.

1.3 Illumina sequencing

The Illumina platforms perform the end goal of sequencing in part using a similar strategy to classical Sanger sequencing; chain termination using labelled dideoxy NTPs. However, the inclusion of a dideoxy NTP is not the point of identification. The difference in determining sequences via Illumina lies in detection of each labelled base as it is added to the growing chain; each base subtype (A, T, G, C) is labelled with a distinct fluorescent label which is activated upon the integration of the base into the growing chain, and the light emitted by this is detected and recorded.

Illumina sequencing is carried out on the Illumina Flow Cell, a glass plate similar to a microscope slide, which features a series of channels designed to deliver sequencing reagents to the cell. The cell is lined with short oligonucleotide sequences which enable the binding of sample sequences to the flow cell by way of complimentary adapter sequences, which are incorporated to Illumina sample during amplicon production using specialised PCR primers. Amplicon production is additionally when the sample barcode tag sequences are added (Fierer et al, 2008), allowing sequencing of multiple samples on a single flow cell, reducing cost for large scale sequencing projects.

2.0: methods

2.2 Co-occurrence calculation

Co-occurrence plots of samples in this study were limited to genera with a co-occurrence correlation value of 0.75 or greater. Co-occurrence network maps were separately produced for sample sets in two categories: Site vs Site and soil depth ranges. Respectively, comparisons were examined for differences between co-occurrence in Site 1 (4296m) and Site 4 (5056m), and for soil depths specific to Site 4 categorised into the depth ranges 3-11cm and >11cm , to allow continuity with previous analyses in this study.

2.3 Seqenv Matching

The handling of large data volumes produced by modern high throughput sequencing technologies such as NGS sequencing on platforms like Illumina – is an increasingly pertinent issue for bioinformaticians. The seqenv programme was developed by Sinclair et al (2016) to attach environmental relevance to sequence data. Seqenv runs similarity searches through the NCBI nucleotide database, and where possible extracts stated isolation sources from the associated metadata from the database. The output of this search is presented as a weighted map of environments associated with a set of sequence data, and allows weighting of particular environments by the prevalence of them in relation to the sequence data; environments which are frequent isolation sources for the identified taxa are displayed more prominently in the graphical representation of the output.

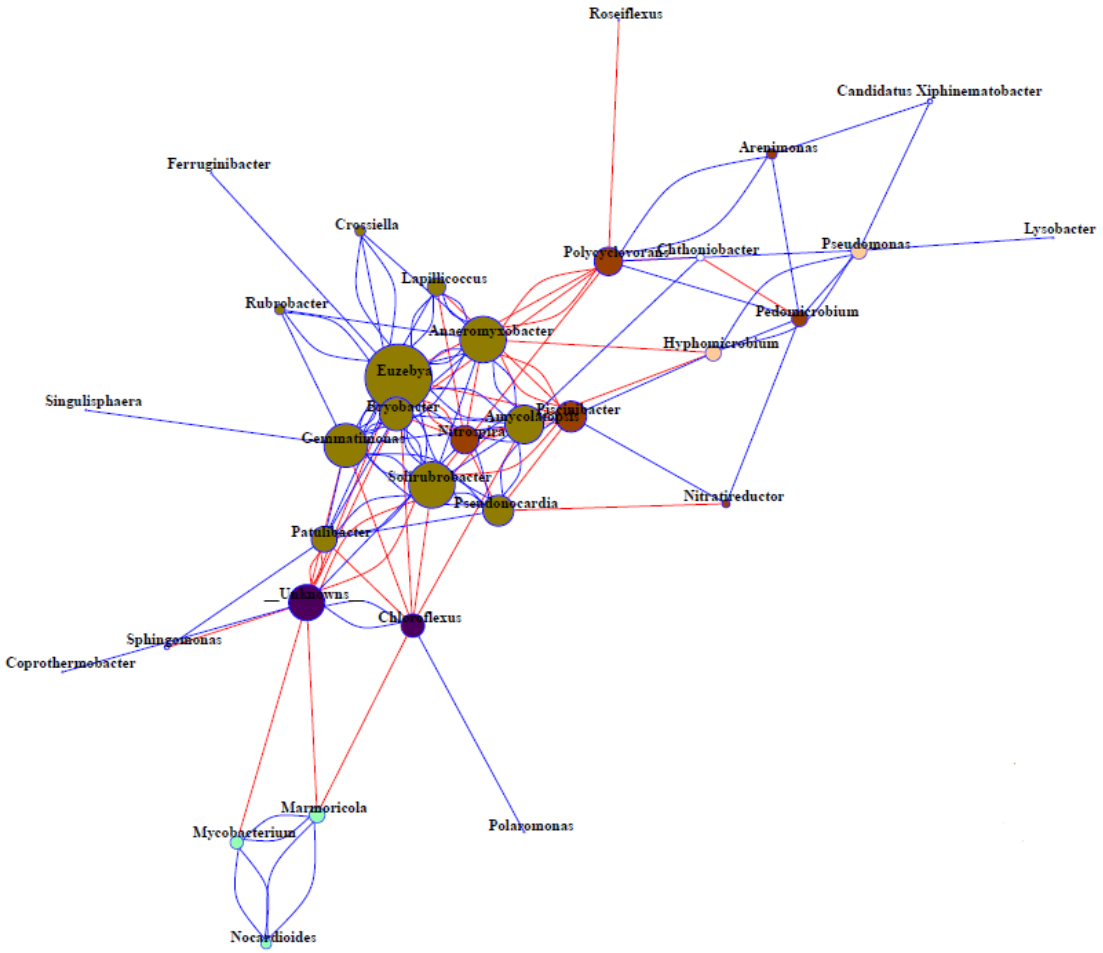
3.0: Results

.

3.1 Co-occurrence network plots

Correlation of co-occurrence for genera sequences in Illumina data was calculated, and using these co-occurrence values it is possible to construct a network plot for the sum of the 16S gene sequence data for each specific site of interest. The network plots for genera identified at Site 1 (4296m) and Site 4 (5056m) are displayed in Fig[1] and [2] below, respectively. The co-occurrence networks allow a speculation of potential interactions between specific clusters

amongst the greater soil microbiome as a whole, and may enable rapid determination of potentially important genera, as groups with greater interconnectivity appear as larger nodes in the map. Paired with information about the metabolic capacities of identified genera in each cluster, this visualisation may allow for some insight into nutrient cycles, competition (or coordination) between species, attenuation of hostile environmental conditions, and elaborate on the predictions generated by the Tax4Fun script as discussed previously.

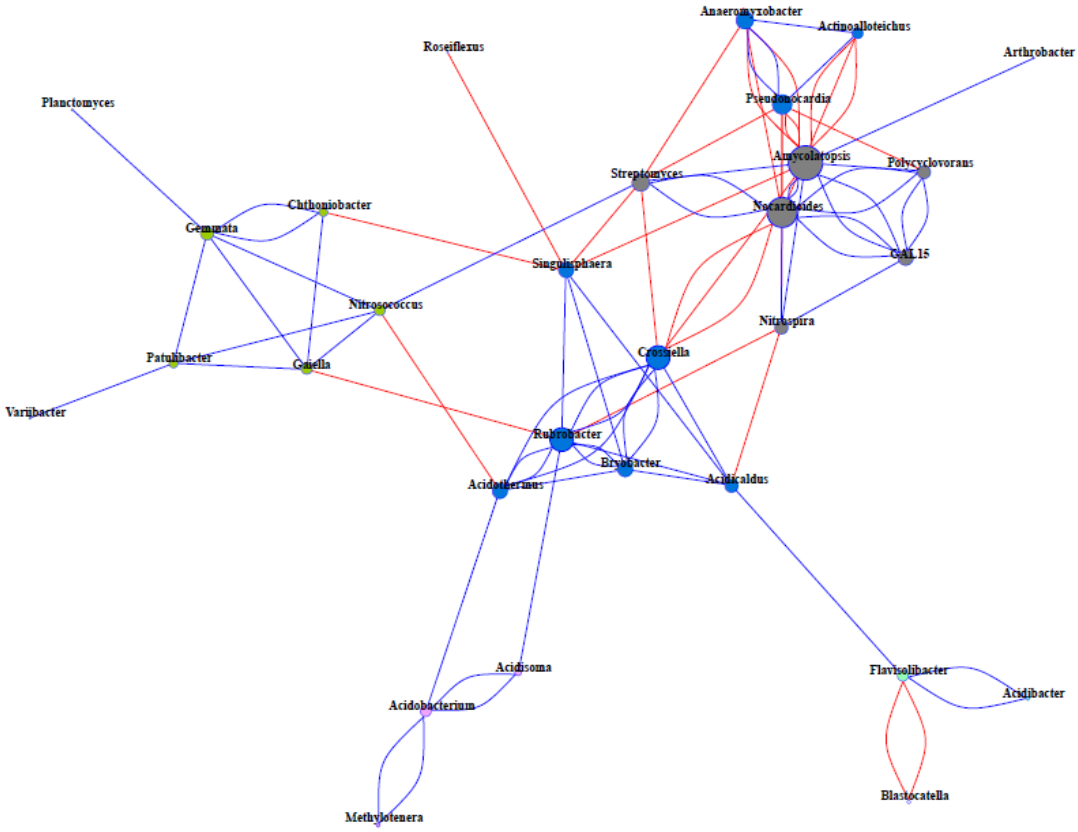


Fig[1]: A co-occurrence network map of genera present in the 4296m site soils, as identified by Illumina MiSeq platform sequencing of bacterial 16S V4 region DNA sequences. Connections represent correlation in the occurrence of pairs of genera in the sequence data, blue connections indicate positive correlation (occurrence of one raises occurrence of the other, implying cooperation) while red connections indicate negative correlation (occurrence of one reduces occurrence of the other, implying competition). Correlation values are set to r=0.75. Clusters sharing a network of blue connections are

represented as similarly-coloured circles, while relative size of circles is indicative of the number of connections an individual genus experiences.

Principal clusters of Site 1

The overall soil bacterial community network map for Site 1 (4296m) is displayed in Fig[1]. The apparent dominant cluster is represented in beige and includes the genera *Gemmatimonas*, *Anaeromyxobacter*, and *Crossiella*. The largest cluster appears extensively interconnected both internally and with other clusters. The brown-coloured cluster containing *Nitrospira*, *Polycyclovorans*, and *Piscinibacter* is the second largest cluster. The final prevalent cluster is represented in purple, however the true extent of this is difficult to discern, as the largest node in the cluster represents "_Unknowns_" – a collection of apparently co-occurring genera which are unreported in literature. This group of unknowns may account for one or more genera. The other member of the purple cluster is the genus *Chloroflexus*. Seemingly distant from the main bulk of clusters, a teal-coloured cluster appears to exist largely separate and possibly less affected by others.



Fig[2]: A co-occurrence network map of genera present in the 5056m site soils, as identified by Illumina MiSeq platform sequencing of bacterial 16S V4 region DNA sequences. Connections represent correlation in the occurrence of pairs of genera in the sequence data, blue connections indicate positive correlation (occurrence of one raises occurrence of the other, implying cooperation) while red connections indicate negative correlation (occurrence of one reduces occurrence of the other, implying competition). Correlation values are set to r=0.75. Clusters sharing a network of blue connections are represented as similarly-coloured circles, while relative size of circles is indicative of the number of connections an individual genus experiences.

Principal clusters of Site 4

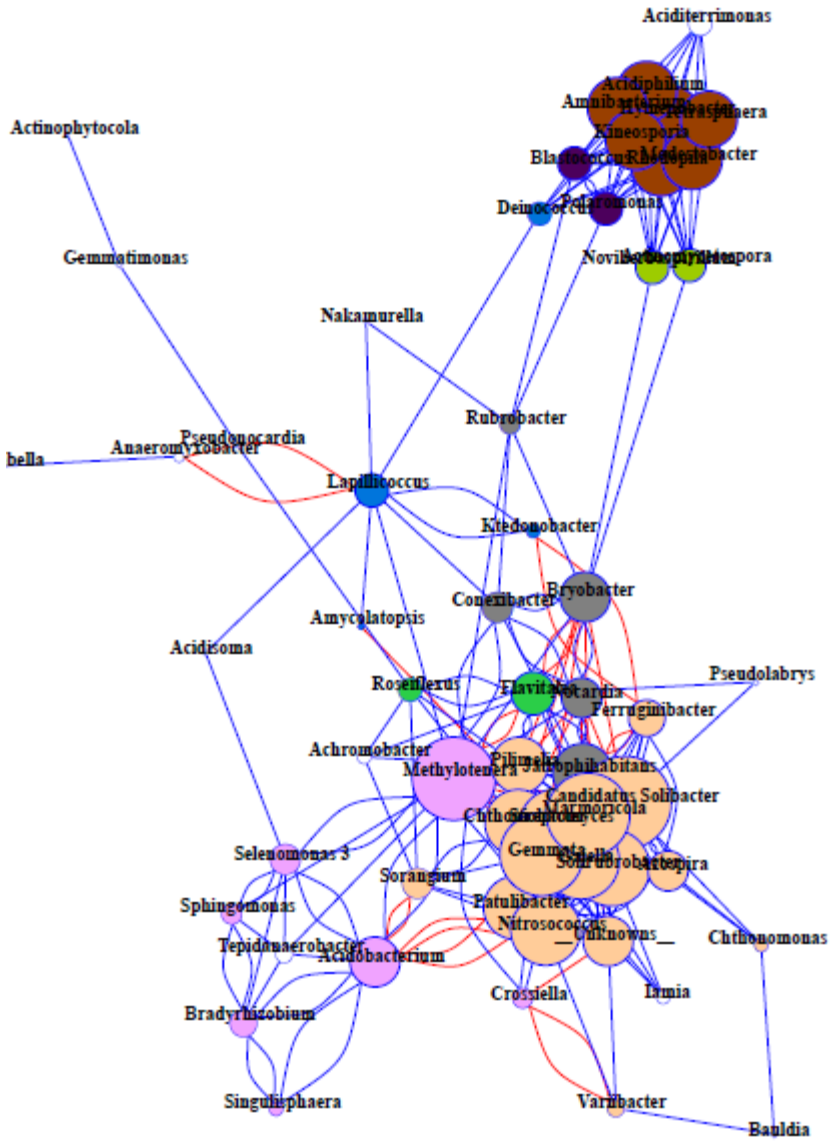
The bacterial community of Site 4 (5056m), as displayed in Fig[2], appears overall less closely-interconnected than the Site 1 community, with fewer nodes in the map being highly connected (fewer notably large diameter points, even in the most prominent cluster). The largest cluster is represented by blue points in the map, and appears to actually consist of 2

separate clusters, with no direct positive interconnections between the smaller section comprised of *Anaeromyxobacter*, *Actinoalloteichus*, and *Pseudonocardia*, and the rest of the blue cluster. It is unclear why this is the case. There is one other notably prominent cluster, represented in grey, which features *Amycolatopsis*, *Nocardioides*, and *Streptomyces* genera as the largest members. Overall the bacterial community network map for Site 4 appears much less centralised around a primary cluster, as was the case for the network map of the Site 1 bacterial community. This may be indicative of a less closely-connected community in Site 4, or simply a more sparse bacterial community overall in the sense of overall mass of bacteria.

3.1.1 Effect of depth profile on co-occurrence

Based on soil depths of interest derived from prior exploration in this study of environmental conditions data, co-occurrence of taxa was plotted for the depth ranges 3-11cm and >11cm for the 5056m site, to elucidate the effect of soil depth on co-occurrence of genera, and the intercommunity connections at each depth range. The depth-based co-occurrence plots are shown in Fig[3] and Fig[4].

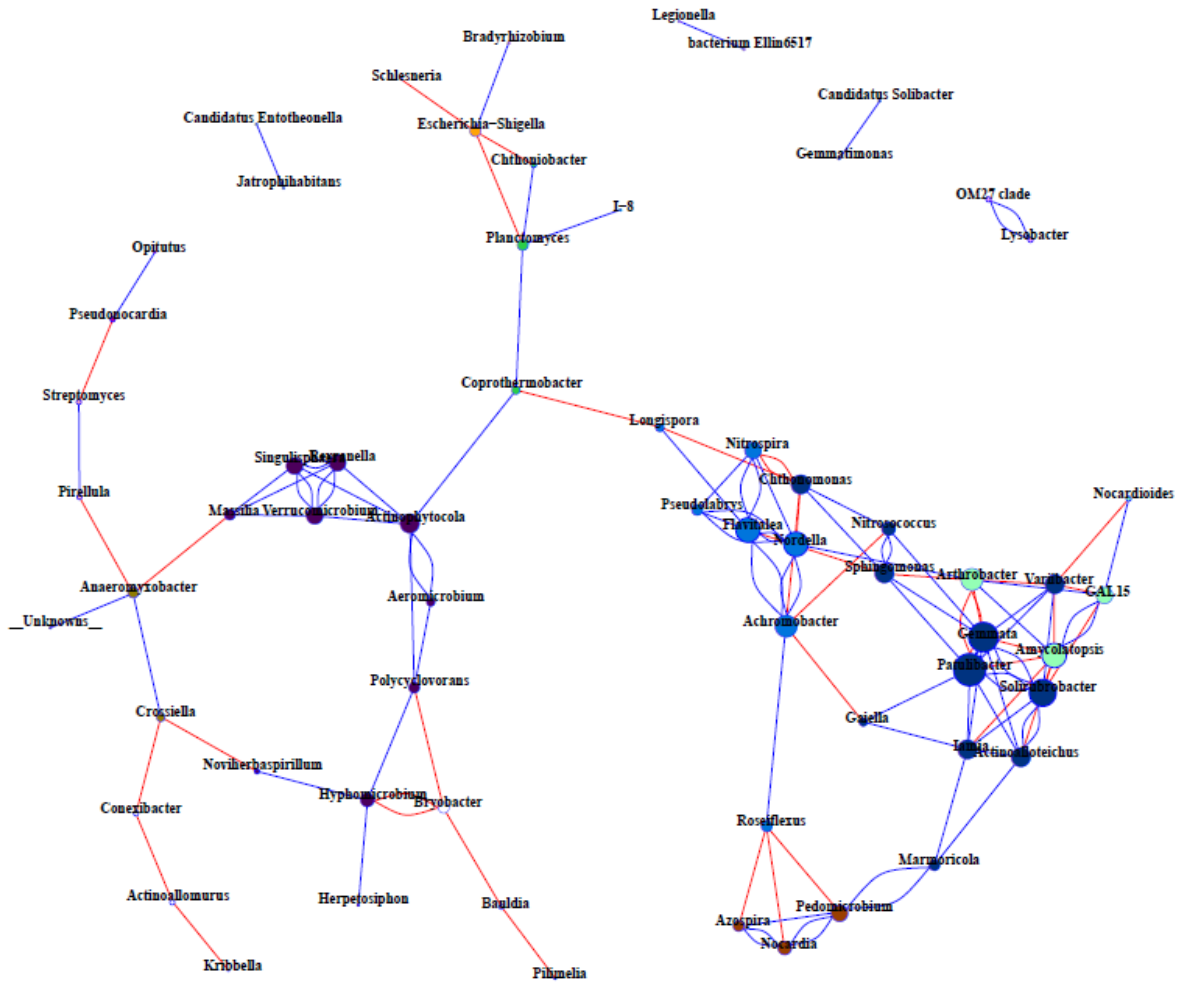
It is immediately noticeable that Fig[3] and [4] display starkly differing network maps in terms of connectivity and distance between points – the clusters in the 3-11cm range depicted in Fig[3] are clearly more densely connected than those of the >11cm range shown in Fig[4], which is evidenced by the comparatively large size of many points in Fig[3]; size of points is representative of number of connections the point has, ergo smaller points exhibit fewer connections.



Fig[3]: A co-occurrence network map of genera present in the 5056m site soils, as identified by Illumina MiSeq platform sequencing of bacterial 16S V4 region DNA sequences. This network is focused on the soil depth range 3-11cm. Connections represent correlation in the occurrence of pairs of genera in the sequence data, blue connections indicate positive correlation (occurrence of one raises occurrence of the other, implying cooperation) while red connections indicate negative correlation (occurrence of one reduces occurrence of the other, implying competition). Correlation values are set to r=0.75. Clusters sharing a network of blue connections are represented as similarly-coloured circles, while relative size of circles is indicative of the number of connections an individual genus experiences.

Principal clusters of Site 4, 3-11cm

The 3-11cm soil depth range of the 5056m site features several notably prominent clusters, displayed in Fig[3] as the beige, brown, pink, and grey clusters. Each of these clusters appear extensively interconnected (evidenced by large circle markers and high numbers of connections between nodes), although the brown cluster is separated from the other major clusters by smaller clusters such as the purple and lighter green cluster. It is notable that the 3-11cm depth range soil bacterial community appears more interconnected than the community for the aggregate of all soil depths from that site as a whole, as shown in Fig[2].



Fig[4]: A co-occurrence network map of genera present in the 5056m site soils, as identified by Illumina MiSeq platform sequencing of bacterial 16S V4 region DNA sequences. This network is focused on the soil depth range >11cm. Connections represent correlation in the occurrence of pairs of genera in the sequence data, blue connections indicate positive correlation (occurrence of one raises occurrence of the other, implying cooperation) while red connections indicate negative correlation (occurrence of one reduces occurrence of the other, implying competition). Correlation values are set to r=0.75. Clusters sharing a network of blue connections are represented as similarly-coloured circles, while relative size of circles is indicative of the number of connections an individual genus experiences.

Principal clusters of Site 4, >11cm depth.

The >11cm soil depth range of the 5056m site, shown in Fig[4], features a comparatively notably more spread out and less closely connected bacterial community network map than the 3-11cm soil depth range community network map displayed in Fig[3]. There are several larger clusters of similar magnitude such as the blue, dark blue, and purple clusters, but the more substantial tightly-connected clusters of the 3-11cm soils appear absent in the >11cm soils. This may be indicative of a less inter-reliant overall bacterial community in these deeper soils. Also of note is that this sparsely-connected overall community is more similar in appearance to the Site 4 all-depths community shown in Fig[2].

Comparison of depth profiles

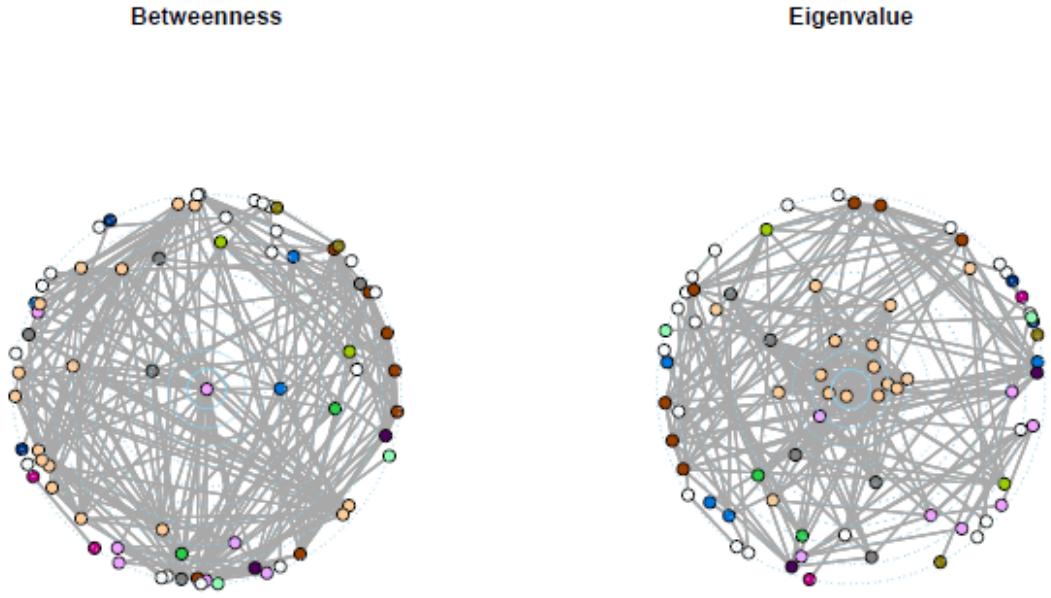
It is immediately apparent that the bacterial community near to the soil surface – in this example the 3-11cm soil depth range – is considerably more tightly interconnected than the deeper soils. It is possible that this is due to the more hostile environmental conditions which are experienced closer to the surface, and so species may be forced to cooperate to survive

the inhospitable environment. Despite the stark differences shown between the network maps of Fig[3] and Fig[4], there are important similarities; many of the same genera are present in the most prominent clusters at each soil depth range – here referring to the beige cluster from Fig[3] and the dark blue cluster from Fig[4]. In both cases, genera such as *Gemmata*, *Solirubrobacter*, *Patulibacter*, *Gaiella*, and *Nitrosococcus* are members of the largest bacterial cluster in the soils.

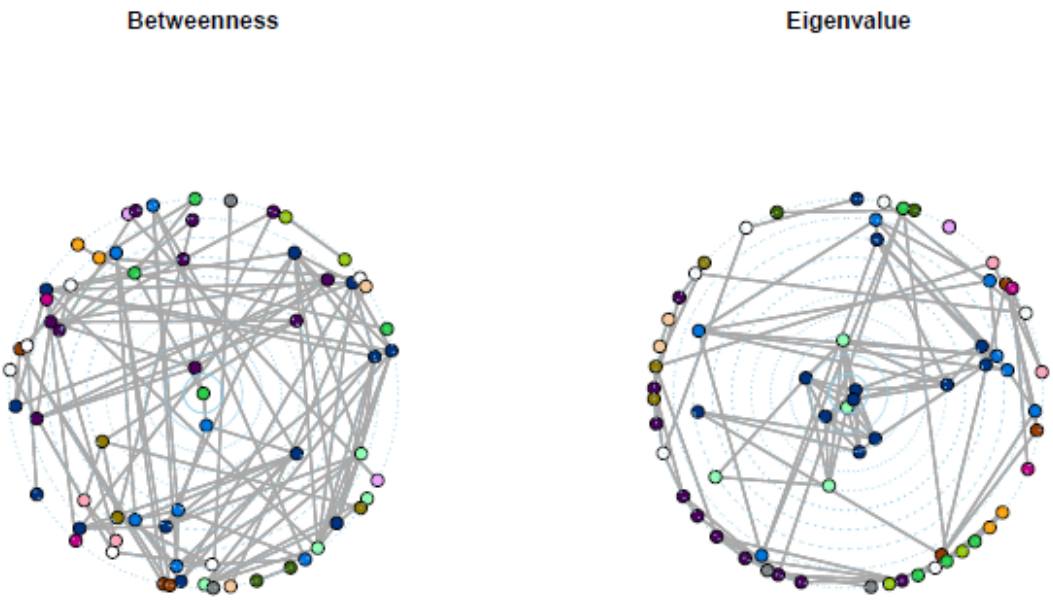
3.1.2 Betweenness &Eigenvalue

The Betweenness and Eigenvalue figures for points on each of the depth profiles were calculated, to further investigate the interconnectivity of the network, as well as to attempt to identify any key genera in the community, upon which any applied environmental effect would propagate through the community with the greatest efficiency.

These plots are displayed in Fig[5] and Fig[6]. Betweenness describes the number of direct connections a point has with other points in the network, representing the shortest distance in the network. The more direct connections a point has with others in the network, the higher the Betweenness value. Betweenness is additionally represented by the relative size of the points in the network map; points of larger radius feature higher Betweenness values.



Fig[5]: Displaying the Betweenness and Eigenvalue values corresponding to the points of the Fig[3] network plot, for the 5056m site: 3-11cm depth profile. Point colours are carried over for clarity. In the case of both Betweenness and Eigenvalue, higher values are reflected with proximity to the outer edges of the circle, while lower values are more centralised. For example, the orange cluster generally displays high Betweenness and low Eigenvalue.



Fig[6]: Displaying the Betweenness and Eigenvalue values corresponding to the points of the Fig[4] network plot, for the 5056m site: >11cm depth profile. Point colours are carried over for clarity. In the case of both Betweenness and Eigenvalue, higher values are reflected with proximity to the outer edges of the circle, while lower values are more centralised. For example, the Dark blue cluster generally displays high Betweenness and low Eigenvalue.

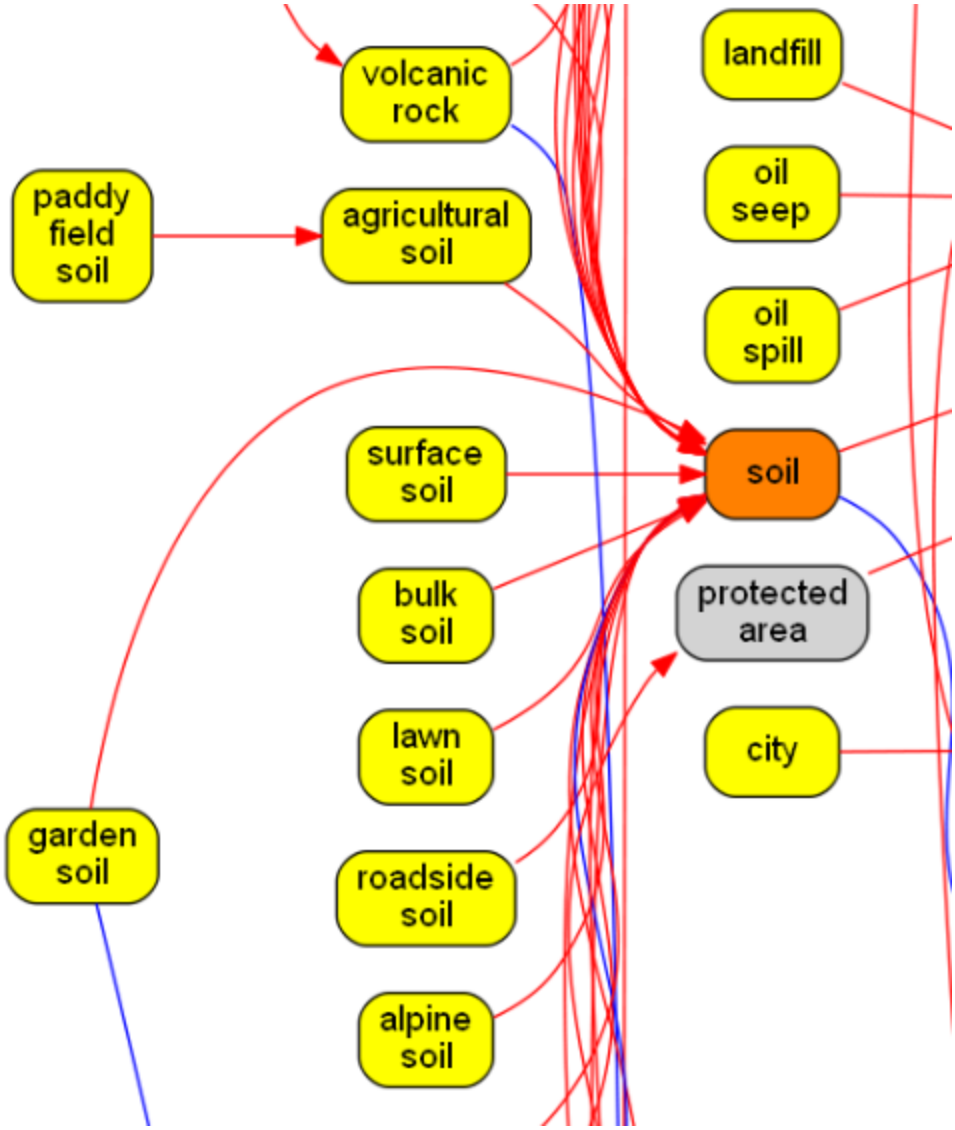
While Betweenness describes the connectedness of a point with focus on the local area of the network, Eigenvalue represents the connectedness of a point in the scope of the wider network, and thus represents the ability of effects to propagate across the network at large, via the point in question. Points which are on the 'border' between different colour clusters (in this case) may have high Betweenness and low Eigenvalue values, and thus act as so-called gatekeepers between the clusters – a single point upon which action is most efficiently propagated from one cluster to another.

Fig[5] displays the Betweenness and Eigenvalue data related to the network map of Fig[3], with colour points on the plot corresponding to points in the network map of the same colour, although specific genus labels are not present due to limitations of space. The majority of points in the map exhibit high Betweenness overall (indicated by being closer to the outer regions of the circular plot), with the larger subcommunities such as the beige cluster being almost exclusively placed near the perimeter of the plot. The beige cluster additionally scores low overall in Eigenvalue, indicating it may act as a gatekeeper for a possible greater community.

Fig[6] displays the Betweenness and Eigenvalue data related to the network map of Fig[4]. In comparison to the plot shown in Fig[5], there are fewer overall connections, however the dispersion of points across the plot appear similar; the majority indicating relatively high Betweenness, with the larger clusters (such as the dark blue clusters) exhibiting high Betweenness. Similarly to the 3-11cm range of Fig[5], the largest cluster here (dark blue) features generally low Eigenvalue, indicating a member of this cluster may hold the gatekeeper role. Interestingly, this also is the case for the pale teal cluster, suggesting the >11cm depth range may feature multiple gatekeepers, while the 3-11cm range only has one clear set of candidates.

3.2 Seqenv

Matching of environments to samples using Seqenv produces plots which indicate the range of matched environments for the submitted samples, and to some extent shows the relations between environments (example: agricultural soil, lawn soil etc are linked to Soil as a whole), an example section of a Seqenv plot is shown in Fig[7]. Full output plots from Seqenv were found to be substantially large, and so the example in Fig[7] represents only a small section of the full plot, a section selected which serves to highlight the weighted display of more prominent environments – Soil is a heavily weighted environment and so appears more darkly-shaded than others.



Fig[7]: An example of seqenv programme Per Sample Ontology output. This is a section of a larger graph produced from the seqenv output of a single sequenced community sample (16S DNA sequences extracted from 4296m site, 0cm soil depth). Here, darker yellow colouring in each node denotes increasing weight of the environment of that node in terms of prevalence in the sequencing data. In this example, the 'Soil' node is notably darker than other nodes, indicating that more of the identified sequences have been reported in the literature to have been isolated from soil.

In this study, each sample has been run through the seqenv script, with sample here referring to the collection of 16S V4 DNA sequences produced from DNA extracted from a single soil sample, representing a single soil depth at a particular site, the amplicons of which are

identified with an identical barcode sequence. Therefore, the seqenv output for each sample displays all reported isolation sources for the taxa identified in the community of the particular soil depth.

A summary table of all Per Sample Ontology graphics is presented in Table[1]. This table highlights the environmental sources of particular interest for each sequence set, i.e. the environments which are particularly heavily weighted in the PSO graph.

Sample ID	Site	Soil Depth (cm)	Notable Environments
5	4926m	0	Soil, Desert, Elevation, Grassland Soil
6	4926m	0	Soil, Desert, Elevation, Grassland Soil
7	4926m	0	Soil, Desert, Elevation, Grassland Soil
8	4926m	1	Soil, Desert, Glacier
10	4926m	3	Soil, Desert, Elevation, Glacier
11	4926m	3	Soil, Desert, Elevation, Glacier
12	4926m	3	Soil, Elevation, Glacier, Lava
13	4926m	5	Soil, Glacier, Lava
14	4926m	5	Soil, Glacier, Elevation
15	4926m	5	Soil, Glacier, Lava
16	4926m	7	Soil, Glacier, Lava
17	4926m	7	Soil, Glacier, Lava
18	4926m	7	Soil, Glacier
19	4926m	9	Soil, Glacier, Lava, Elevation
20	4926m	9	Soil, Glacier
21	4926m	9	Soil
22	4926m	11	Soil, Glacier, Prairie
23	4926m	13	Soil, Glacier, Prairie
24	4926m	13	Soil, Glacier, Lava, Prairie
25	4926m	15	Soil, Prairie
26	4926m	15	Soil, Sediment, Glacier, Prairie

27	4926m	15	Soil, Sediment, Prairie
28	4926m	17	Soil, Sediment, Prairie
29	4926m	17	Soil, Sediment, Glacier, Prairie
30	4926m	17	Soil, Sediment, Prairie
31	4926m	19	Soil, Sediment, Glacier, Prairie
32	4926m	19	Soil, Sediment, Prairie
33	4926m	19	Soil, Sediment, Prairie
34	4926m	21	Soil, Sediment, Glacier, Prairie
35	4926m	21	Soil, Sediment, Glacier, Prairie
36	4926m	21	Soil, Sediment, Prairie
37	4926m	37	Soil, Sediment, Glacier, Prairie
38	4926m	37	Soil, Sediment, Prairie
39	4926m	37	Soil, Glacier
44	5056m	1	Soil, Glacier, Lava, Volcanic Field, Desert
45	5056m	3	Soil, Glacier, Lava, Volcanic Field, Desert
46	5056m	5	Soil, Glacier, Lava
47	5056m	5	Soil, Glacier, Lava
48	5056m	7	Soil, Glacier, Lava
49	5056m	7	Soil, Glacier, Lava
50	5056m	7	Soil, Glacier, Lava
51	5056m	9	Soil, Glacier, Lava
52	5056m	9	Soil, Glacier, Lava
53	5056m	11	Soil, Glacier, Lava
54	5056m	11	Soil, Glacier, Lava
55	5056m	11	Soil, Glacier, Lava, Sediment
56	5056m	13	Soil, Glacier, Lava
57	5056m	13	Soil, Glacier, Lava
58	5056m	13	Soil, Glacier, Lava, Sediment
59	5056m	15	Soil, Glacier, Lava, Prairie
60	5056m	15	Soil, Glacier, Lava, Sediment, Prairie
61	5056m	15	Soil, Glacier, Sediment

62	5056m	17	Soil, Glacier, Sediment
63	5056m	17	Soil, Glacier, Lava, Sediment
64	5056m	17	Soil, Glacier, Lava

Table[1]: Summarising the environments of particular interest for each PSO figure for each of the sequencing sample sets. Here notable environments are defined as those which, as outlined in the example from Fig[7], appear particularly darker than other nodes, indicating more weighting of that environment in the seqenv matching output.

It is immediately notable all samples feature the Soil environment with the greatest weighting – all samples were taken from soil sample extracts, so this is a logical outcome. The Grassland Soil and Prairie type environments appear to have greater weighing more frequently in the soils of the vegetated lower altitude site than in the higher altitude soils, which is pertinent to the reality of these sites; the lower altitude site features simple grasses, while these are absent at the higher altitude site. It appears that the matched environments remain mostly similar throughout the depth profile of the high altitude site. Notably, the Glacier environment remains relevant at all depths in the high altitude profile, whereas in the lower altitude site's soils it was not always among the most heavily weighted environments. This is likely a reflection of the lower temperatures typically featured in the soils at higher altitude, as recorded in this study via iButton and datalogger-linked temperature probes.

The Lava and volcano-related environments are weighted heavily in the soils of both sites, and at a range of depths, but as is observed with the Glacier environment, the Lava environment appears more pertinent to the higher altitude soils, which indeed do represent more of a volcanic environment than the lower altitude, vegetated site. The lower altitude soils are older and more weathered, and are richer in organics; while they were one part of lava flows, they are now much more soil-like. In contrast the higher altitude soils are relatively more fresh, less weathered lava material. In Table[2], a wide range of environments available to Seqenv were compared, to determine which site is most similar to each Seqenv environment (for this, the data from each site was amalgamated from all soil depths).

Environment	Site of Prevalence
Temperate	**Site 4**
Peat Soil	**Site 4**
Aerosol	Site 1
Activated Sludge	Site 1
Acid Mine Drainage	**Site 4**
Saline Lake	**Site 4**
Beach	Site 1
ENV: 01000174	**Site 4**
Reservoir	**Site 4**
Fresh Water	**Site 4**
Clay Soil	Site 1
Mound	Site 1
Headwater	Site 1
Volcanic Field	**Site 4**
Sea	Site 1
Meadow	Site 1
ENV: 00000446	**Site 4**
Terrestrial Habitat	**Site 4**
Volcanic Rock	**Site 4**
Shore	**Site 4**
Cave Wall	Site 1
Gold Mine	Site 1
Marine Habitat	**Site 4**
ENV: 00000447	**Site 4**
Ground Water	Site 1

Table[2]: Comparing significantly different prevalence of environments from seqenv matching between the Site 1 (4296m) and Site 4 (5056m). Environments marked "ENV: [number]" were undefined in seqenv output.

The top 25 most significantly differently reported environments are split across the two sites, but there are notable differences between them. Of particular note is the increased prevalence in matching of environments such as Volcanic Rock and Volcanic field in Site 4 sequences. This is highly representative of the reality of the site, as described previously.

The greater matching for the Acid Mine Drainage environment at the higher altitude site is interesting – there is some Copper mining industry in the region, but it is not particularly active in the immediate area of the site. It is possible that this environment is flagged due to the weakly acidic nature of the soil at this site, which leads to increased presence of acidophilic species in the soil.

There are a number of environments matched to the higher altitude site which are aquatic in nature: Reservoir, Fresh Water, and Marine Habitat are examples of this. It is unclear why these environments would be notable in either site, let alone the more arid high altitude site. Potentially, this could be due to an abundance of taxa which are capable of living in a wide range of environments, or alternatively, due to the introduction of wind-blown material from other sites; the low biomass in general meaning that wind-introduced material can influence the microbial community composition.

At the lower altitude site, there is increased matching of environments such as Clay Soil, Mound, and Meadow. These could be considered as richer soil environments than the more arid soils of the higher altitude site, and this is appropriate for the lower altitude site; the site is vegetated with simple grasses and other such hardy plants, but this difference alone suggests that the soils are comparatively richer in nutrients.

3.2.1 seqenv soil depth comparison

Table[3] lists the most significantly differently reported environments between the Upper (3-11cm) and Lower (>11cm) soil depths at the high altitude site.

Environment	Depth Prevalence
Acid Mine Drainage	Upper
Mountain	Upper
Fen	Upper
Elevation	Upper
ENV: 00000446	Upper
Terrestrial Habitat	Upper
Volcanic Rock	Upper
Saline Lake	Upper
Contaminated Sediment	Upper
ENV: 00000447	Upper
Marine Habitat	Upper
Desert	Upper
Garden	**Lower**
Lava Field	Upper
Shore	Upper
Volcano	Upper
Volcanic Field	Upper
Mud	Upper
Bagasse	Upper
Freshwater Lake	**Lower**
National Park	**Lower**
Savannah	**Lower**
ENV: 01000196	Upper
Cave	**Lower**
Stream Sediment	Upper

Table[3]: The environments matched via seqenv which are the most significantly differently represented in the Upper (3-11cm) and Lower (>11cm) depths of the soil at Site 4 (5056m). Environments marked "ENV: [number]" were undefined in seqenv output.

It is notable that several of the environments matched more to the lower depths of the soil are somewhat more hospitable than other environments; Garden, Savannah, and Freshwater Lake suggest a less hostile environment than is experienced in the upper layers. Several hostile environments are assigned to the upper layers, meanwhile, such as Volcano and Volcanic Field, Desert, Lava Field, Elevation, and Acid Mine Drainage (acidic soil). These hostile environments are particularly pertinent to the location: the site is a high elevation volcanic slope, with slightly acidic soils at the surface, with pH approaching neutral as soil depth increases. The seqenv matching therefore exhibits accuracy with the reality of the field site as measured previously in this study.

The inclusion of Bagasse is somewhat unusual, as no sugar cane residue is likely to be present in these soils – the crop is not notably farmed in the region and such residue is unlikely to be carried in wind. It is likely this is an anomaly due to similarity between detected species and a species involved in decomposition of sugar cane matter.

4.0: Discussion

4.1 co-occurrence network mapping by site

The genera co-occurrence maps of Figs[1-4] display groups of genera which share positive correlation in co-occurrence – represented by blue line connections in the maps – together in clusters marked with a single colour. Each co-occurrence plot features several such clusters of genera. It is possible that this apparent clustering of genera may be related to a cooperative existence between these bacteria in the soil environment, although these plots alone are insufficient to determine this for certain. Similarly, it is possible to speculate that negative correllations in co-occurrence – indicated with red line connections – may highlight competitive or antagonistic relationships among the soil bacterial community.

While it is potentially feasible to begin speculation on potential bacterial communities highlighted via co-occurrence mapping, further investigation is vital. True cooperative action would require a significant study to fully describe potential interactions between species, for example a metabolomic characterisation in response to presence/absence of co-occurring

genera. It is not feasible to state that the highlighted clusters of genera form coherent communities in the environment based on these plots alone, and indeed it is possible that co-occurrence may be simply coincidental. In addition, the network maps of Figs[1] and [2] account for all soil depths investigated in a single map, which suggests that all bacteria in all soil depths are able to interact – this is unlikely simply due to the relative distance involved compared to the scale of single-celled organisms, and doubly so with the consideration that not all species are motile, further decreasing the likelihood of direct interactions between bacteria in the environment, and that the soil environment itself changes as depth increases, as was noted previously in this study.

While not suitable as a "one stop shop" for describing interactions between bacteria in environmental samples, plots such as the co-occurrence maps may prove invaluable as a starting point for hypothesising in such studies, as a means to highlight potential genera of interest.

Beige cluster: This cluster appears to be the largest of the sample site. Important organisms in this cluster are *Anaeromyxobacter* and *Crossiella*, which are Nitrate-reducing (N-reducing) bacteria (Sanford et al, 2002) and thus may rely on N-fixation by other community members. In addition, *Crossiella* may produce antibiotic compounds (Labeda, 2001) (alongside *Pseudonocardia* and *Amycolatopsis* – also present in the cluster). Of further interest is an apparent clustering of genera which feature adaptations to extreme environments; *Gemmatimonas* feature antioxidant activity to protect from damage by Oxygen radicals (Takaichi et al, 2010), *Rubrobacter* genera have been reported to be radiotolerant or radioresistant to varying degrees (Suzuki et al, 1988), believed to be due to adaptations against desiccation, *Crossiella cryophila* is a cold tolerant species, and additionally *Solirubrobacter* and *Bryobacter* are acid tolerant (Singleton et al, 2003; Kulichevskaya et al, 2010) species (soil pH values in these sites are weakly acidic, with pH increasing towards 7 as soil depth increases). Resistance to environmental extremes seems to be a common trait in this cluster and thus adaptability to environmental extremes may be a key factor that causes the organisms in this cluster to group. A variety of motile and non-motile genera are represented, which is expected as this network map does not account for variations in soil depth, which as previously discussed in this study appears to be an important variable in the prevalence of motile species.

Brown cluster: This cluster does not feature N-reduction, and thus may not be dependent on other members (such as those found within the Beige cluster) for this, indeed, the actions of *Nitrospira* include Nitrite oxidation (Watson et al, 1986). The brown cluster features some extremophile genera such as *Polycyclovorans* which are halotolerant (Gutierrez et al, 2013) and *pedomicrobium* (Gerbers, 1981) which have been reported in deserts. Chemoorganoheterotrophic species are present in this cluster, such as *Piscinibacter* (Stackebrandt et al, 2009) and *Polycyclovorans*, yet Carbon fixation appears absent, further suggesting possible reliance on other carbon sources, such as organic carbon introduced by wind-blown material.

Purple cluster: This true scale of this cluster is difficult to ascertain, as one of the two genera presented in the network map are tagged simply as Unknowns. The other, named, member, however, is *Chloroflexus*, some species of which are photosynthetic; they express bacteriochlorophyll I (Hanada et al, 1995). Phototrophs are vital for Carbon fixation in the soil environment, and an important component of soil nutrient cycles which sustain other species.

Pale green cluster: This cluster features *Mycobacterium*, a genus which includes species responsible for the diseases Tuberculosis and Leprosy. However, *Mycobacterium* are descended from soil bacteria, and indeed are believed to be able to fix Carbon. It is possible that this study features an unreported species which is related to *Mycobacteria* yet exhibits greater similarity to a shared ancestral soil bacteria. Nitrate reduction is not reported in members of this cluster.

Cream cluster: This cluster appears to feature Nitrate reduction, as members of both genera presented (*Pseudomonas* and *Hyphomicrobium*) are capable of N-reduction (Martineau et al, 2015). In addition, *Hyphomicrobium* are capable of using single-Carbon sources, suggesting a role in Carbon fixation. The combination of these features suggest that members of this cluster may play an important role in the overall nutrient cycling of the soils in the site.

Site 1 Community connections

The beige cluster has negative connections with the brown and purple clusters, which may be indicative of competition between the clusters. The brown cluster shares positive connections with the purple and cream clusters, while the purple cluster additionally shares a negative

correllation with the green cluster. Given the nature of connections in the community as a whole, it appears in this map that the beige cluster is largely dominant and is possibly not in general reliant on other clusters, as is evidenced by the negative connections it shares – any interactions with other communities appaer detrimental to the beige cluster members, or vice versa. The apparent antibiotic production by the beige cluster may be in response to the competition with other clusters.

Principal subcommunities of Site 4

It is immediately notable that the co-occurrence network maps for Site 1 (Fig[1]) and Site 4 (Fig[2]) are significantly different in their layout. While many similar species appear at both sites, this approach reveals that there are some notably different inter- and intra-community relationships between them. Site 1 is lower altitude and features surface vegetation, while Site 4 is higher altitude and appears devoid of surface vegetation; these factors may be driving factors behind the difference in community relationships.

Blue cluster: The largest of the clusters in this network, the blue cluster provides N-reduction to the community at large. The genera of the cluster feature a range of extreme environment tolerances such as cold, radiation, and acid tolerance. That the largest cluster of species features such a range of extremophilic features suggests a more hostile environment experienced at the 5056m site in comparison to the 4296m site, which indeed appears to be the case; as discussed previously, the site experiences low temperatures, high solar flux, and additionally features weakly acidic soils at the surface, with pH values increasing towards neutral with soil depth. The blue cluster also appears to produce antibiotics. It is apparent that the blue cluster is separated in the network map by the grey cluster, as there is no direct blue connection between the points representing *Anaeromyxomacter*, *Pseudonocardia*, and *Actinoalloteichus* with the remaining blue cluster nodes, yet they share the same colour. It is unclear why this should be the case.

Grey cluster: This cluster is largely comprised of consumer species, heterotrophic species, able to metabolise Nitrite to Nitrate (*Nitrospira*), aliphatic hydrocarbons and polycyclic aromatic hydrocarbons (PAH), and a variety of other C and N sources (*Streptomyces*). In addition, many *Streptomyces* species are capable of antibiotic production. The combination of

features expressed in this cluster suggests that its members may sustain themselves upon the products of other species.

Green cluster: This cluster appears more self-sustaining than others of the site; both nitrate reduction and nitrite production are performed by members of this cluster, by *Gaiella* (Albuquerque et al, 2011) and *Nitrosococcus* (Campbell et al, 2011) respectively. In addition, breakdown of complex organic carbon sources is performed by *Chthoniobacter* (Sangwan et al, 2004).

Pink cluste: The primary connecting factor in this cluster is the apparent ubiquity of tolerance of acidity; *Acidobacterium*, *Acidisoma* and *Methylotenera* are all acid-tolerant genera. This cluster is not generally photosynthetic. Growth of *Methylotenera* is inhibited by nitrate (Kalyuzhnaya et al, 2006), and thus this genus may be dependent on Nitrate reduction by other species in order to thrive, which is not provided by this cluster and is likely provided by members of one or more other cluster groups of the community.

Teal cluster: This cluster is comprised of heterotrophic genera, *Flavisolibacter* and *Acidibacter*. It is likely that the members of this cluster depend on carbon fixation by other species from others, such as members of the blue cluster, a notion which is reflected by the blue connections between the two clusters.

Site 4 Community Connections

The blue and grey clusters share negative connections, indicating potential competitive or antagonistic interactions. This is a potential explanatory factor for the production of antibiotics exhibited by both clusters (or conversely, the competition is a result of the antibiotics produced). Other considerations are less supportive of the competitive hypothesis – the blue cluster features nitrate reduction, while the grey cluster features nitrite oxidation, which is required by the N-reducers of the blue cluster. The green cluster features positive connections with the grey cluster, while it features negative connections with the blue cluster. The green cluster features nitrogen reduction, which complements the action of the grey cluster but is in competition with the blue cluster, and so it follows that the connections are positive and negative respectively. The blue cluster additionally has positive connections with the remaining clusters (pink and teal). As the pink cluster members (*Methylotenera* in

particular) are dependent on nitrate reduction by other species, it is logical under this assumption that connections with the blue cluster would be positive.

4.1.1 Co-occurence network mapping by depth range

Principal clusters of Site 4, 3-11cm depth range

Orange clusters: This cluster appears to be the largest in the 3-11cm soil depth range. It is interesting to note that members of the genera identified in this cluster have been noted to produce antibiotics – in this case *Streptomyces* (Han et al, 2015) – and this cluster features negative connections with more than one neighouring cluster, such as the pink and grey clusters. Despite these negative connections, there is some evidence to suggest that this cluster may have an important role in the nutrient state of the soils; *Chthoniobacter* is involved in the breakdown of complex organics (Sangwan et al, 2004), *Streptomyces* species have been noted to be important decomposers in soil communities, and *Nitrosococcus* species have been reported to oxidise nitrite to nitrate (Campbell et al, 2011). Several genera in this cluster are aerobes, and acid tolerance has also been reported in genera such as *Gaiella* (Albuquerque et al, 2011) and *Solirubrobacter* (Singleton et al, 2003). Acid tolerance is an expected feature, as the surface soils at this site were noted to be mildly acidic, with pH approaching neutral as depth increased.

Pink cluster: This cluster features several members which are adapted to hostile conditions, such as *Crossiella*, which are typically cold tolerant, and numerous acid tolerant genera such as *Methylotenera* (Kalyuzhnaya et al, 2006), *Acidobacterium* (Kishimoto et al, 2010), and *Singulisphaera* (Kulichevskaya et al, 2008). *Methylotenera* are additionally interesting as growth is inhibited by nitrate and therefore are reliant on nitrate reduction by other community members. *Crossiella* species have been noted to feature nitrate reduction which may serve to support *Methylotenera*.

Grey cluster: This cluster features genera with extremophile features such as *Bryobacter* (Kulichevskaya et al, 2010) and *Rubrobacter* (Suzuki et al, 1988). Additionally, the genera of the cluster are typically aerobes. Propensity for aerobic respiration and extremophile features are to be expected in the near surface depths of the soil, as this is the more hostile environment in comparison to lower depths.

Green cluster: This cluster features phototrophic genera such as *Roseiflexus* (Hanada et al, 2002). The other genus in this cluster is *Flavitalea*, which is a motile, aerobic genus. The prevalence of phototrophy and aerobic respiration is appropriate in the near-surface environment, and additionally this genus is noted for its prevalence worldwide in regions of geological activity (Gaisin et al, 2016).

Blue cluster: The majority of members of this cluster feature aerobic respiration – a common trait in the genera reported in this near-surface depth range. In addition, the extremophile genus *Deinococcus* is present in this cluster; a genus of considerable interest in the field of astrobiology. With species which are resistant to extremes of temperature, radiation, and desiccation (Brooks and Murray, 1981), it is of little surprise that *Deinococcus* species may be present in the near-surface soil depths at the site. A further point of interest is that members of this cluster produce antibiotics, through the action of *Amycolatopsis* (Lechevalier et al, 1986).

Brown cluster: Of particular interest in this cluster is the presence of *Modestobacter*, a genus first isolated from the McMurdo dry valleys. *Modestobacter* species have been reported to be cold tolerant and some have been reported to be acid-fast (Mevs et al, 2000). Both cold and acid tolerance would indeed be advantageous in the soils of interest to this study. The presence of this genus is important in investigating the nature of the site as a Martian analogue. *Hymenobactrer*, another genus of this cluster, is reported to be highly sensitive to the action of antibiotics (Buczolits et al, 2006), which is of interest given the noted production of antibiotics by other cluster members in this depth range.

Principal clusters of Site 4, >11cm depth range

Dark blue cluster: This cluster features several genera which respire aerobically such as *Actinoalloteichus*, despite ostensibly being away from the air-exposed surface soils, perhaps suggesting soil oxygenation by other community members. Numerous members of the cluster are nitrate reducing, and these are perhaps in part supplied with nitrate by the oxidation of nitrite through the action of genera such as *Nitrosococcus*. Several members of this cluster are non-motile, such as *Variibacter* (Kim et al, 2014) and *Marmoricola*, this is somewhat surprising as the soil moisture levels recorded at the >11cm depth range were greater than those recorded closer to the surface, which should facilitate cell motility to a greater extent.

Perhaps the lack of motility is in response to the freezing recorded at these lower depths among the high-altitude sites - if the soil moisture becomes frozen, at night for example, then cell motility will be impossible in the frozen soil environment, and so perhaps species such as these are less reliant on the ability to move.

Blue cluster: As noted with the dark blue cluster, this cluster also features both genera which reduce nitrates and genera which oxidise nitrites to nitrates. Here, nitrite oxidation is performed by *Nitrospira* species. The presence of *Roseiflexus* in this cluster is at first unexpected, as *Roseiflexus* species are noted phototrophs, yet photosynthetic action is not possible at this soil depth; light of appropriate wavelengths does not penetrate far beyond the soil surface. However, *Roseiflexus* are additionally known chemoheterotrophs, and so are able to survive when conditions do not support photosynthetic action (Hanada et al, 2002).

Pale green cluster: This cluster appears to sit at odds with its neighbours; *Amycolatopsis* species have been reported to produce antibiotics, and all connections with neighbouring clusters are negative (red connections), suggesting an antagonistic relationship.

Teal cluster: A potentially important actor in this cluster is the genus *Chthoniobacter*, which has been reported to have a role in decomposition in soils, which is a vital component of the soil nutrient cycle. Also of interest is the presence of *Coprothermobacter*, which are known to be heat tolerant species – yet high soil temperatures were not recorded at this depth range.

Purple cluster: This cluster features genera capable of respiration in an anaerobic or microaerobic environment: *Hyphomicrobium* (Martineau et al, 2015) and *Singulisphaera* respectively. This is of interest as the lower depths of the soil are more likely to be a lower oxygen environment, when compared to the near surface soils. Members of this cluster appear able to make use of alternative energy sources – *Hyphomicrobium* respire anaerobically if nitrate is available, and *Polycyclovorans* are able to use aromatic organics such as PAHs as a Carbon source (Guttierez et al, 2013).

Beige cluster: This cluster features one member capable of aerobic respiration (*Crossiella*) and one capable of anerobic respiration (*Anaeromyxobacter*). Both genera are noted nitrate reducers. *Crossiella* have been reported to produce antibiotics, and all connections with neighbouring clusters in the network map are negative connections (discarding the Unknowns), suggesting little cooperation between this cluster and others.

Brown cluster: *Azospira*, present in this cluster, feature nitrogen fixation, an important component of the soil nitrogen cycle (Reinhold-Hurek and Hurek, 2000). This cluster also features *Pedomicrobium*, an extremeophile genus which has been reported in deserts.

Comparison of depth profiles

There are a number of interesting comparisons between the various clusters of the 3-11cm soil depth range, and the >11cm soil depth range. Firstly, it appears that the near-surface soils feature a larger prevalence of extremophiles such as *Deinococcus*, C*rossiella*, and *Bryobacter*. This is fitting with the environmental data reported previously in this study which indicate that the surface is the more hostile environment than the lower depths. Similarly, acid tolerance or preference was present more often in the 3-11cm range than in the >11cm range, which corroborates the soil pH readings taken during this study that indicated the soils at this site were more acidic at the surface, and pH approached neutral as depth increased. The presence of phototrophic genera such as *Roseiflexus* is expected in the near surface soils, yet *Roseiflexus* was also noted in the >11cm soils due to the additional chemoheterotrophic capacities of the genus. The ability to utilise multiple strategies for respiration is surely an advantage in a hostile environment such as this.

It was noted that the >11cm depth network map was more spread out and with fewer average connections per cluster, this may be reflected in the nature of the clusters of this network, which are frequently able to produce nitrates amongst each single cluster, or feature abilities to make use of alternative sources for respiration (including anaerobic respiration), which was not seen as much in the clusters of the upper soil layers. This may suggest that the potential community of the >11cm depths may be less cohesive or cooperative than the upper soil layers, possibly due to conditions being less hostile at the lower depths, and survival being a less tenuous proposition compared to the inhospitable surface.

4.1.2 Betweenness and Eigenvalue

As displayed in Fig[5], it appears that the larger clusters in the 3-11cm depth range all exhibit high betweenness, indicating that within themselves, the clusters are tightly linked. The single point in the orange cluster with low betweenness is likely to be accounted for by the

point representing *Variibacter*. The eigenvalue values for the orange cluster are generally low, indicating that this cluster may have gatekeeping activities, and may be instrumental in propagating effects throughout the community as a whole. The other large clusters – pink and brown – appear to exhibit both high betweenness and eigenvalue, suggesting these clusters may affect the soil community as a whole to a greater degree than the orange cluster, but require the action of the orange cluster to propagate any actions throughout the greater soil community.

Fig[6] indicates that below 11cm depth, the larger clusters – Dark blue, blue and purple – exhibit generally high betweenness, although some points do present a lower value, and these often represent border points between clusters. The dark blue and pale green clusters both exhibit notably low eigenvalue scores, and are neighbours in the network map, yet the two clusters feature negative (red) connections, indicating that they are likely in close competition with one another, possibly attempting to fill the same niche in the wider soil community. The high betweenness, low eigenvalue figures for these clusters suggests that gatekeeper species in both clusters reflect the close competitive interactions between the two clusters.

Value of co-occurrence mapping in studies

The utility of plotting co-occurrence of genera in a large-scale sequencing project such as the present study lies primarily in the ability to make early predictions of interactions between community members of ecological samples, such as the soil community of an extreme environment and Martian analogue region, in this case. The examination of correlation in the presence of genera under similar circumstances may assist in the elucidation of possible bacterial interactions with other species and the environment, but it is important to consider that this is by no means a concrete observation, and should be taken as advisory at this stage; co-occurrence may indicate some form of interaction but is not by itself sufficient to simplay state this as fact, and in order to confirm any hypotheses it would be necessary to investigate further with more conventional examinations.

In this study, the use of co-occurrence mapping was focused on the soil environment before and after the notable change in soil moisture (occurring at approx. 11cm depth), for the purpose of examining the differences in bacterial community in response to a change in a variable in environmental conditions while remaining in the same Mars analogue site. As

stated, the co-occurrence of genera alone is not sufficient to describe the bacteria in these soil depth ranges as distinct communities, as from this data alone it is not possible to truly determine interactions between species in the soil. Additionally, by the nature of the sequencing project, the sequences reported for each soil depth represent a snapshot in time and do not account for factors such as cell motility – mobile species may present in multiple soil depth layers due to their movement, which cannot be accounted for in the co-occurrence plots.

The use of co-occurrence mapping such as this may be invaluable in other studies, as a method to begin to identify potential assemblages of microbes in a community sample. This would be especially interesting in cases where action against a specific member of the community is desired; examining the interactions or potential reliances upon other community members could allow more efficient targeting of species in, for example, gut microbiome disorders.

4.2 Seqenv

Seqenv as a tool does have limitations in a study such as this, which is primarily drawn from the volume of the pool of samples used; each sample of sequences to which the script is applied represents an entire community of bacteria taken from soil – one of the richest environments reported globally in terms of bacterial diversity.

As shown in Per Sample Ontology data of Table [1], In this study 34 individual bacterial communities from 4296m and 21 communities from 5056m were analysed, a total of 55 soil communities, which results in a significant range of potentially-matched environments. Many soil bacterial taxa are extremely common in all soils, and thus are matched with a wide range of environments by Seqenv, most of which bare no relation at all to the reality of the sampling site (examples are the Sea or Stream Sediment environments, as shown in Tables [2] and [3] respectively). It is advised that similar future studies perform Seqenv matching on individual samples to reduce overall noise potential.

4.3 Implications for astrobiology

This study raises interesting implications for future astrobiology studies which utilise terrestrial Martian analogue regions such as the Atacama Desert, McMurdo Dry Valleys etc. In the Altiplano region of interest to this study, high resolution investigations of the change in environments with soil depth, and the effects this change has upon soil bacterial population, is underreported in current literature; the majority of studies focusing on different environments entirely such as high altitude lakes (Demergasso et al, 2010), salt plains, saline wetlands such as Salar de Huasco (Pérez et al, 2017), or hotsprings and their associated deposits (Phoenix et al, 2006). The dry volcanic soils, however, are largely unreported, with focus thus far having fallen on yeasts, but not bacteria (Pulschen et al, 2015). This study offers a novel insight into methods to examine the bacterial diversity and community structure in an extreme environment, which is potentially of interest to microbial ecologists and astrobiologists alike, given the Mars-like features of this environment.

4.3.1 Altiplano as a Mars analogue

The effect of increased soil depth in the Altiplano has been observed to be, broadly, that extreme conditions are buffered comparative to the surface as soil depth increases – the environment becomes significantly less hostile with depth below the surface at all sites observed in this study (see previous chapter for Figs). Several frequently matched environments assigned by Seqenv, as shown in Tables[1] - [3], are particularly of interest. The primary environment matched was Soil, which is expected on the basis that all samples were indeed extracted from soil samples. Other Seqenv environments, however, are relevant for astrobiological studies on the basis that similar environments are reported on Mars.

The Lava and Glacier environments, for example, are frequently matched to samples, and both of these are or have been potential environments present on the red planet. Evidence for volcanic activity on Mars is plentiful (Werner, 2009), with the most striking being the gigantic volcano Olympus Mons (Isherwood et al, 2013), and thus it is likely that areas of the surface were at one time accounted for by lava flows (McEwen et al, 1999). The Glacier environment is interesting due to the evidence for action of water on on Mars; canyons such as Valles Marineris are believed to have been produced by hydraulic action in the past.

Evidence of past Martian glacial action has been reported (Head and Weiss, 2014), and the remnants of such features in the modern environment could be of interest to astrobiologists. In the case of present day Mars, the polar ice caps account for much of the surface ice, although subsurface ice deposits at lower latitudes have been suggested (Viola et al, 2015).

The Desert environment represents another environment repeatedly matched to samples by Seqenv. Mars itself may accurately be described as a cold desert, and so the fact that species such as those identified in this study may be isolated from deserts is of interest regarding the analogue nature of the region.

5.0: Conclusions

The data examination presented here enables the observation of the effect of changes in soil environmental conditions on the bacterial community extant within these soils. Using soil depth guidelines established based on environmental conditions data presented previously in this study – a cutoff of 11cm soil depth due to the notable increase in soil moisture at this depth in the 5056m site – the comparative differences between bacterial community composition and structure may be examined. Figs[3] and [4] display clearly the change in relative bacterial community composition and interactions even over this relatively shallow range of soil depths, which is indicative of the rapid change in environmental conditions experienced by these soils, becoming generally less extreme as depth below the surface increases. This is reflected in the relative reduction in prevalence of extremophile species such as *Deinococcus* sp. at lower depths.

The environmental matching generated via Seqenv indicates that the novel field sites which are investigated as the basis of this study mirror numerous environment categories featured on Mars, in particular the Desert and Glacier environments, both of which feature notably in the past or present conditions of the red planet, and which supports the use of the sites as a Martian analogue region.

The examination in the change in bacterial community with reference to concurrent change in environmental conditions in a Mars analogue region has not been reported previously, and this study may provide framework for similar investigations in further Mars-like environments such as Atacama and McMurdo, and may help inform analyses from future

Mars robotic surface missions such as ExoMars 2020, which will feature the deepest subsurface investigations yet taken at the Martian surface – the search for evidence of past life is dependent on careful selection of organic target molecules, and the likely survival strategies of any microbial life aids in this selection.

6.0: References

Albuquerque, L., França, L., Rainey, F.A., Schumann, P., Nobre, M.F., da Costa, M.S. (2011). *Gaiella occulta* gen. nov., sp. nov., a novel representative of a deep branching phylogenetic lineage within the class *Actinobacteria* and proposal of *Gaiellaceae* fam. nov. and *Gaiellales* ord. Nov. Systematic and Applied Microbiology Vol 34. P595-599.

Brooks, B.W., and Murray, R.G.E. (1981). Nomenclature for "*Micrococcus radiodurans*" and Other Radiation-Resistant Cocci: *Deinococcaceae* fam. nov. and *Deinococcus* gen. nov., Including Five Species. International Journal of Systematic Bacteriology Vol 31. P353-360.

Buczolits, S., Denner, E.B.M., Kämpfer, P., Busse, H-J. (2006). Proposal of *Hymenobacter norwichensis* sp. nov., classification of '*Taxeobacter ocellatus*', '*Taxeobacter gelupurpurascens*' and '*Taxeobacter chitinovorans*' as *Hymenobacter ocellatus* sp. nov., *Hymenobacter gelipurpurascens* sp. nov. and *Hymenobacter chitinivorans* sp. nov., respectively, and emended description of the genus *Hymenobacter* Hirsch et al. 1999. International Journal of Systematic and Evolutionary Microbiology Vol 56. P2071-2078.

Campbell, M.A., Chain, P.S.G., Dang, H., El Sheikh, A.F., Norton, J.M., Ward, N.L., Ward, B.B., Klotz, M.G. (2011). *Nitrosococcus watsonii* sp. nov., a new species of marine obligate ammonia-oxidizing bacteria that is not omnipresent in the world's oceans: calls to validate the names '*Nitrosococcus halophilus*' and '*Nitrosomonas mobilis*'. FEMS Microbial Ecology Vol 76. P39-48.

Demergasso, C., Dorador, C.,Meneses, D., Blamey, J.,Cabrol, N.,Escudero, L., Chong, G. (2010). Prokaryotic diversity pattern in high-altitude ecosystems of the Chilean Altiplano. Journal of Geophysical Research Vol 115.

Drancourt, M., Bollet, C., Carlioz, A., Martelin, R., Gayral, J-P., Raoult, D. (2000). 16S Ribosomal DNA Sequence Analysis of a Large Collection of Environmental and Clinical Unidentifiable Bacterial Isolates. Journal of Clinical Microbiology Vol 38. P3623-3630.

Fierer, N., Hamady, M., Lauber, C.L., Knight, R. (2008). The influence of sex, handedness, and washing on the diversity of hand surface bacteria. PNAS Vol 105. P17994-17999.

Gaisin, V.A., Grouzdev, D.S., Namsaraev, Z.B., Sukhacheva, M.V., Gorlenko, V.M., Kuznetsov, B.B. (2016). Biogeography of thermophilic phototrophic bacteria belonging to *Roseiflexus* genus. FEMS Microbiology Ecology Vol 92.

Gerbers, R. (1981). Enrichment, Isolation, and Emended Description of *Pedomicrobium ferrugineum* Aristovskaya and *Pedomicrobium manganicum* Aristovskaya. International Journal of Systematic Bacteriology Vol 31. P302-316.

Gutierrez, T., Green, D.H., Nichols, P.D., Whitman, W.B., Semple, K.T., Aitken, M.D. (2013). *Polycyclovorans algicola* gen. nov., sp. nov., an Aromatic-Hydrocarbon-Degrading Marine Bacterium Found Associated with Laboratory Cultures of Marine Phytoplankton. Appled and Environmental Microbiology Vol 79. P205-214.

Han, X., Zheng, J., Xin, D., Xin, Y., Wei, X., Zhang, J. (2015). *Streptomyces albiflavescens* sp. nov., an actinomycete isolated from soil. International Journal of Systematic and Evolutionary Microbiology Vol 65. P1467-1473.

Hanada, S., Hiraishi, A., Shimada, K., Matsuura, K. (1995). *Chloroflexus aggregans* sp. nov., a Filamentous Phototrophic Bacterium Which Forms Dense Cell Aggregates by Active Gliding Movement. International Journal of Systematic Bacteriology Vol 45. P676-681.

Hanada, S., Takaichi, S., Matsuura, K., Nakamura, K. (2002). *Roseiflexus castenholzii* gen. nov., sp. nov., a thermophilic, filamentous, photosynthetic bacterium that lacks chlorosomes. International Journal of Systematic and Evolutionary Microbiology Vol 52. P187-193.

Head, J.W., and Weiss, D.K. (2014). Presevation of ancient ice at Pavonis and Arsia Mons: Tropical mountain glacier deposits on Mars. Planetary and Space Science Vol 103. P331-338.

Isherwood, R.J., Jozwiak, L.M., Jansen, J.C., Andrews-Hanna, J.C. (2013). The volcanic history of Olympus Mons from paleo-topography and flexural modeling. Earth and Planetary Science Letters Vol 363. P88-96.

Kalyuzhnaya, M.G., Bowerman, S., Lara, J.C., Lidstrom, M.E., Chistoserdova, L. (2006). *Methylotenera mobilis* gen. nov., sp. nov., an obligately methylamine-utilizing bacterium within the family *Methylophilaceae*. International Journal of Systematic and Evolutionary Microbiology Vol 56. P2819-2823.

Kim, K.K., Lee, K.C., Eom, M.K., Kim, J-S., Kim, D-S., Ko, S-H., Kim, B-H., Lee, J-S. (2014). *Variibacter gotjawalensis* gen. nov., sp. nov., isolated from soil of a lava forest. Antonie van Leeuwenhoek Journal of Microbiology Vol 105. P915-924.

Kishimoto, N., Kosako, Y., Tano, T. (1991). *Acidobacterium capsulatum* gen. nov., sp. nov.: An Acidophilic Chemoorganotrophic Bacterium Containing Menaquinone from Acidic Mineral Environment. Current Microbiology Vol 22. P1-7.

Kulichevskaya, I.S., Ivanova, A.O., Baulina, O.I., Bodelier, P.L.E., Sinninghe Damsté, J.S., Dedysh, S.N. (2008) *Singulisphaera acidiphila* gen. nov., sp. nov., a non-filamentous, *Isosphaera*-like planctomycete from acidic northern wetlands. International Journal of Systematic and Evolutionary Microbiology Vol 58. P1186-1193.

Kulichevskaya, I.S., Suzina, N.E., Liesack, W., Dedysh, S.N. (2010). *Bryobacter aggregatus* gen. nov., sp. nov., a peat-inhabiting, aerobic chemo-organotroph from subdivision 3 of the *Acidobacteria*. International Journal of Systematic and Evolutionary Microbiology Vol 60. P301-306.

Labeda, D.P. (2001). *Crossiella* gen. nov., a new genus related to *Streptoalloteichus*. International Journal of Systematic and Evolutionary Microbiology Vol 51. P1575-1579.

Lechevalier, M.P., Prauser, H., Labeda, D.P., Ruan, J-S. (1986). Two New Genera of Nocardioform Actinomycetes: *Amycolata* gen. nov. and *Amycolatopsis* gen. nov. International Journal of Systematic Bacteriology Vol 36. P29-37.

Martineau, C., Mauffrey, F., Villemur, R. (2015). Comparitive Analysis of Denitrifying Activities of *Hyphomicrobium nitrativorans*, *Hyphomicrobium denitrificans*, and *Hyphomicrobium zavarzinii*. Applied and Environmental Microbiology Vol 81. P5003-5014.

McEwen, A.S., Malin, M.C., Carr, M.H., Hartmann, W.K. (1999). Voluminous volcanism on early Mars revealed in Valles Marineris. Nature Vol 397. P584-586.

Mevs, U., Stackebrandt, E., Schumann, P., Gallikowski, C.A., Hirsch, P. (2000). *Modestobacter multiseptatus* gen. nov., sp. nov., a budding actinomycete from soils of the Asgard Range (Transantarctic Mountains). International Journal of Systematic and Evolutionary Microbiology Vol 50. P337-346.

Navas-Molina, J.A., Peralta-Sánchez, J.M., González, A., McMurdie, P.J., Vázquez-Baeza, Y., Xu, Z., Ursell, L.K., Lauber, C., Zhou, H., Song, S.J., Huntley, J., Ackermann, G.L., Berg-Lyons, D., Holmes, S., Caporaso, J.G., Knight, R. (2013). Advancing Our Understanding of the Human Microbiome using QIIME. Methods in Enzymology Vol 531. P371-444.

Pérez, V., Hengst, M., Kurte, L., Dorador, C., Jeffrey, W.H., Wattiez, R., Molina, V., Matallana-Surget, S. (2017). Bacterial Survival under Extreme UV Radiation: A Comparative Proteomics Study of *Rhodobacter* sp., Isolated from High Altitude Wetlands in Chile. Frontiers in Microbiology Vol 8. Article 1173.

Phoenix, V.R., Bennett, P.C., Engel, A.S., Tyler, S.W., Ferris, F.G. (2006). Chilean high-altitude hot-spring sinters: a model system for UV screening mechanisms by early Precambrian Cyanobacteria. Geobiology Vol 4. P15-28.

Pulschen, A.A., Rodrigues, F., Duarte, R.T.D., Araujo, G.G., Santiago, I.F., Paulino-Lima, I.G., Rosa, C.A., Kato, M.J., Pellizari, V.H., Galante, D. (2015). UV-resistant yeasts isolated from a high-altitude volcanic area of the Atacama Desert as eukaryotic models for astrobiology. MicrobiologyOpen Vol 4. P574-588.

Reinhold-Hurek, B., and Hurek, T. (2000). Reassessment of the taxonomic structure of the diazotrophic genus *Azoarcus sensu lato* and description of three new genera and new species, *Azovibro restrictus* gen. nov., sp. nov., *Azospira oryzae* gen. nov., sp. nov., and *Azonexus fungiphilus* gen. nov., sp. nov. International Journal of Systematic and Evolutionary Microbiology Vol 50. P649-659.

Sanford, R.A., Cole, J.R., Tiedje, J.M. (2002). Characterization and Description of *Anaeromyxobacter dehalogenans* gen. nov., sp. nov., an Aryl-Halorespiring Facultative Anaerobix Myxobacterium. Applied and Environmental Microbiology Vol 68. P893-900.

Sangwan, P., Chen, X., Hugenholtz, P., Janssen, P.H. (2004). *Chthoniobacter* gen. nov., sp. nov., the first Pure-culture Representative of Subdivision Two, *Spartobacteria* classis nov., of the Phylum *Verrucomicrobia*. Applied and Environmental Microbiology Vol 70. P5875-5881.

Singleton, D.R., Furlong, M.A., Peacock, A.D., White, D.C., Coleman, D.C., Whitman, W.B. (2003). *Solirubrobacter patuli* gen. nov., sp. nov., a mesophilic bacterium within the *Rubrobacteridae* related to common soil clones. International Journal of Systematic and Evolutionary Microbiology Vol 53. P485-490.

Stackebrandt, E., Verbarg, S., Frühling, A., Busse, H-J., Tindall, B.J. (2009). Dissection of the genus *Methylibium*: reclassification of *Methylibium fulvum* as *Rhizobacter fulvus* comb. nov., *Methylibium aquaticum* as *Piscinibacter aquaticus* gen. nov., comb. nov. and *Methylibium subsaxonicum* as *Rivibacter subsaxonicus* gen. nov., comb. nov. and emended descriptions of the genera *Rhizobacter* and *Methylibium*. International Journal of Systematic and Evolutionary Microbiology Vol 59. P2552-2560.

Suzuki, K-I., Collins, M.D., Iijimia, E., Komogata, K. (1988). Chemotaxonomic characterization of a radiotolerant bacterium, *Arthrobacter radiotolerans*: Description of *Rubrobacter radiotolerans* gen. nov., comb. nov. FEMS Microbiology Letters Vol 52. P33-40.

Takaichi, S., Maoka, T., Takasaki, K., Hanada, S. (2010). Carotenoids of *Gemmatimonas aurantiaca* (*Gemmatimonadetes*): identification of a novel carotenoid, deoxyoscillol 2-rhamnoside, and proposed biosynthetic pathway of oscillol 2,29-dirhamnoside. Microbiology Vol 156. P757-763.

Viola, D., McEwen, A.S., Dundas, C.M., Byrne, S. (2015). Expanded secondary craters in the Arcadia Planitia region, Mars: Evidence for tens of My-old shallow subsurface ice. Icarus Vol 248. P190-204.

Watson, S.W., Bock, E., Valois, F.W., Waterbury, J.B., Schlosser, U. (1986). *Nitrospira marina* gen. nov. sp. nov.: a chemolithotrophic nitrite-oxidizing bacterium. Archives of Microbiology Vol 144. P1-7).

Werner, S.C. (2009). The global Martian volcanic evolutionary history. Icarus Vol 201. P44-68.

Chapter 3 – Examination of Lipid-Derived Biomarkers in Mars Analogue Soils

1.0: Introduction

1.1: Lipid Biomarkers, and use of biomarkers in studies.

Biomarkers are organic molecules which persist stably in the environment for extended periods of time, and are as such a reliable sign of the presence of life in the environment from which they are extracted. Usually, biomarkers do not take the form of the original molecules as they are expressed *in vivo*, but a product of degradation or diagenesis which is readily relatable to the original molecule or family of molecules for identification purposes. Given the length of time more stable biomarkers may persist in the environment, they act as molecular fossils, preserved in sediments and mineral deposits.

Examples of biomarkers include long-chain aliphatic compounds such as long-chain alkenones produced by species such as *Pseudoisochrysis paradoxa*, a species of haptophyte algae, typically recovered from sea bed sediments (Theroux et al, 2013), which are utilised in the estimation of past sea surface temperatures (Brassell et al, 1986), or the analysis of membrane lipids such as glycerol dialkyl glycerol tetraethers, or GDGTs, for similar ends. GDGTs in particular have been of focus in recent years, both from archaeal origin (Schouten et al, 2002) and – recently discovered – bacterial origin (Weijers et al, 2006a). Analytically, it is possible to distinguish between GDGTs of archaeal and bacterial origin due to structural differences. Archaeal GDGTs are known as Isoprenoid GDGTs, and feature penta- and hexacyclic ring moieties (carbon ring structures comprised of 5 and 6 carbon atoms respectively). Bacterial GDGTs, meanwhile, are branched; they contain alkyl sidegroups, but no cyclic moieties.

Fig.1: Examples of the structure of Isoprenoid (left) and Branched (right) GDGT molecules, displaying the difference between the two families. The isoprenoid GDGT

here features two cyclopentyl moieties, which are absent in the branched GDGT. Image adapted from Taylor et al (2013)

Biomarkers derived from plant sterols (Alsalahi et al, 2015) or porphyrins, which are the remnants of photopigments such as chlorophyll, or from haem groups (Schaeffer et al, 1993) are additionally viable analytes in such studies. Porphyrin rings in particular are of interest; as the central structure of chlorophylls, they are valuable markers of both complex plant life and photosynthetic microbial life such as archaea and prototrophic bacterial species such as *Cyanobacteria*.

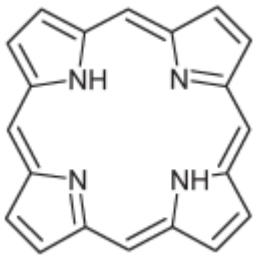


Fig.2: Structure of a porphyrin ring. Such rings form much of the central structure of Chlorophylls and also Haem groups, and are consequently viable biomarkers. Image adapted from Suo et al (2007)

1.2: Biomarkers in Astrobiology

The use of Biomarkers in relation to the ongoing search for evidence of life outside of Earth – particularly on Mars – centres on the long-term persistence and stability of biomarkers in the environment, as well as their readily identifiable nature (Parnell et al, 2007). Much focus has been applied to the preservation of molecules such as chlorophyll (Stromberg et al, 2014) or porphyrins (Suo et al, 2007) under simulated Martian conditions, both in laboratory settings and in Martian analogue regions such as the Atacama Desert or McMurdo Dry Valleys.

In order to survive on (or near) the surface of Mars, organisms would be required to survive extremely hostile conditions. The Martian surface experiences very low temperatures. For example, at the site of the Mars Pathfinder mission landing site, a temperature range of 190-270^{o}K was recorded (Ulrich et al, 2008), representing a substantial diurnal variation in

temperature, where night time temperatures have been reported to drop by approximately 80^{o}K.

The surface of Mars receives extensive solar and cosmic radiation. Solar Radiation, in the form of the Ultra Violet Spectrum (UV-A, UV-B and UV-C) and in PAR (aka the visible spectrum). Solar flux is particularly extensive on Mars (Cockell et al, 2000) due to the sparse atmosphere covering the planet, which is insufficient to attenuate these damaging wavelengths. In addition to UV levels far in excess of those reported on Earth, Mars additionally receives the highly damaging UV-C wavelengths, which on earth are entirely blocked by the atmosphere. Solar flux is to a small extent occluded by airborne dust in the atmosphere, but this is not sufficient to impact the hostile nature of the surface to any meaningful measure. In addition to extensive UV, the Martian surface is bathed in ionizing radiation (Dartnell et al, 2007) which is extremely destructive to organic material and is predicted to affect the uppermost few metres of the Martian surface (Kminek and Bada, 2006).

In addition, the surface features oxidative soil chemistry (Kounaves et al, 2014), which has a destructive effect on organic molecules, and a lack of stability of liquid water. Although bulk liquid water is absent from the surface, ice caps exist at the poles, and evidence for periodic flows has been observed (McEwen et al, 2011), which is potentially the result of subsurface brine ices melting during the Martian daytime.

A potential survival strategy for Martian microorganisms is observed in endolithic communities: wherein bacteria such as *Cyanobacteria* reside within rocks. An example of an endolithic community is phototrophic colonies just below the surface of gypsum deposits, observed in the Atacama Desert (Wierzchos et al, 2015). The structure of the gypsum allows sufficient transmission of Photosynthetically-Active Radiation (PAR) that photosynthesis may occur, yet the mineral coverage provides protection from the harmful high levels of UV radiation experienced in the region. Similarly to endolithic niches, it is suggested that hot spring evaporite may be a suitable environment for survival of life on Mars (Rothschild, 1990, Schubert et al, 2009), and indeed possible evidence for past hotspring activity on Mars has been observed, such as the discovery of Silica deposits similar to those found around hotsprings such as the Geysers Del Tatio, Chile (Ruff and Farmer, 2016).

Despite the challenges life would face on the Martian surface, there is increasing evidence that extremophiles observed on Earth could possibly survive such conditions. Stromberg et al (2014) demonstrated that identifiable chlorophyll signatures were preserved in a UV radiation environment simulating that present at the surface of Mars, which was attributed to the protective effect of mineral coverage. Additionally, Morozova et al (2007) reported that permafrost archaea were able to survive simulated temperature regimes to that of Mars. Meanwhile, radiotolerant and radioresistant species such as *Deinococcus radiodurans* are capable of sustaining Mars-levels of irradiation without significant cellular damage.

Interest in biomarkers such as photopigments are based in their common expression in extremophile bacteria such as *Cyanobacteria* strains and other extremophiles. Extremophile bacteria and archaea take focus due to the conditions on Mars; only extremophiles could stand a chance of survival or preservation on the Martian surface.

1.3: The present study

In this study, lipid-derived biomarkers were recovered from soils collected at Mars analogue sites in the high altitude Chilean Altiplano. This region was chosen specifically for the range of Mars-like conditions experienced at the soil surface: The air temperature is low, with the soil surface itself being heated during the day by extensive solar flux, and the area is arid during the austral summer.

Soil samples were collected from several sites at increasing altitude between 4200m and 5000m. At each site, soils were collected over a range of depths from the soil surface down to 37cm, which allows for a high resolution investigation of the change in biomarker presence with soil depth in this hostile environment, and allows for comparison of this change between a vegetated control site (4269m) and higher altitude sites. Soil samples were rapidly freeze-dried upon return to the lab, to prevent any change in biomarker content during storage, therefore providing an accurate a snapshot as possible of the conditions of each layer of soil, as they were upon initial sampling.

Several classes of biomarkers are of interest to the field; this study will investigate the range of plant, archaeal and bacterial origin biomarkers recovered from field sites over the range of

depths discussed previously, with particular focus on long chain aliphatic hydrocarbon compounds such as N-alkanes. Recoverable nucleic acid content was additionally investigated from each soil sample, and data from these experiments is presented in previous chapters.

1.3.1: Rationale for using biomarkers in this study

Lipid-derived biomarkers are notably stable and persistent in the environment, as is evidenced by their frequent use in palaeoclimate studies. Such organic material, stable for extended periods in the environment may offer a relevant point of focus for upcoming Mars surface missions such as the ESA ExoMars rover and NASA Insight lander.

The ExoMars rover is of particular relevance to this study, as the rover will be equipped with a sampling drill capable of reaching to 2m below the Martian surface, and is also planned to feature GC-MS instrumentation, which will have utility in identifying any biomarker presence.

It is unlikely that any complex organic material from extinct life would persist at the Martian surface, due to the damaging conditions at the surface. While intact organic material persisting on the surface itself is unlikely, the sediments in the 2m subsurface range could possibly be sheltered enough from the inhospitable surface conditions for organic biomarkers to be detectable by ExoMars, and the investigation of possible hotspring-origin silicate desposits may also yield intact molecular fossils.

The buffering of hostile conditions by soil was highlighted previously in this study in Chapter 1, wherein it was demonstrated that hostile surface conditions at a Mars analogue site are rapidly attenuated by soil depth coverage, in particular the effect of high diurnal temperature variation and extensive solar heating of the soil surface, which is not accounted for at other terrestrial mars analogue sites with a high level of accuracy.

It is in consideration of the utility of the Altiplano Mars analogue sites, and the upcoming planned Mars missions, that this study was conceived. The use of Biomarkers extracted from a Mars-like environment will allow a simulation of possible analyses that upcoming missions such as ExoMars may provide. A range of biomarkers will be accounted for in analyses for this study, allowing for the examination of preservation in the environment of biomarkers

from a variety of potential sources, although primary focus lies in long chain aliphatics. The study will investigate how readily such biomarkers are able to persist in the environment over a relatively short period (based on shallow soil accumulation), and how effective the shallow soil coverage is in protecting the organics from the damaging conditions at the surface.

2.0: Methods

2.1: Total Lipid Extraction protocol

Upon return of samples from field site to the lab, a portion was freeze dried for use in BECS lab. Approximately 25g per sample was collected; allowing the filling of 2 ASE cells, thus ensuring a backup sample was available should a failure occur during processing of the first cell.

Prior to freeze drying, samples were stored at -20^{o}C for a minimum of 24 hours. Following completion of freeze drying, the samples should were returned to -20^{o}C for long term storage.

ASE cells wee cleaned using sequential immersion and sonication in organic solvents including dichloromethane and acetone, before being allowed to dry in a sterile environment within a fume hood.

2.1.1: ASE Cell Cleaning Protocol:

ASE cell bodies and caps were cleaned thoroughly before each extraction run:

- Cells are disassembled and rinsed with water to remove more substantial soil/sediment residues
- Approximately half of the total cell bodies and caps are placed into a sealable Teflon jar
- Cell parts are immersed in Acetone and the jar sealed
- Jar is sonicated in a water bath for 30 minutes

- Acetone is disposed of in the appropriate manner, and the jar refilled with Acetone
- Following 30 minutes of sonication, the acetone is transferred to a storage bottle and saved. This saved Acetone is to be used for the first rinse next time cells are to be cleaned
- The final rinse is performed with immersion in 3:1 Dichloromethane:Methanol. The DCM:MeOH is recycled and can be used 3 times before disposal
- Cell parts are allowed to dry in a fume hood

Sterile ASE cells were assembled and approximately 10-15g of sample is loaded into the cell, depending on the consistency and grain size of the sediment/soil of interest. For the Altiplano samples the mean sample mass used was approximately 12g. Where insufficient sample was available to entirely fill the cell, sterile combusted sand was used to make up the difference. Blank samples containing only combusted sand were to included to provide negative control samples for lipid extraction.

Combusted 80ml ASE collection vials receive the extracted lipid solution, and are placed into a turbovap instrument, using Nitrogen gas to expedite the evaporation of TLE solvent. When only a small volume of dissolved TLE remains in the ASE collection vials – approximately 2ml – the TLE is transferred to pre-weighed combusted 8ml vials. The TLE is dried, TLE mass weighed and recorded, before the TLE is re-dissolved for storage before further processing.

2.1.2: Total Lipid Extraction

Soil samples from each site and depth were stored at -20^{o}C for at least 24 hours prior to freeze drying. Frozen samples were freeze dried for at least 24 hours, to remove all moisture from the samples prior to lipid extraction. Total lipid extraction was performed using a DIONEX 350 Accelerated Solvent Extractor, using approximately 10-15g of freeze dried soil sample per extraction, utilising Dichloromethane:Methanol as the solvent; a high polarity solvent is desirable to ensure maximum yield of lipid content of both polar and nonpolar lipids.

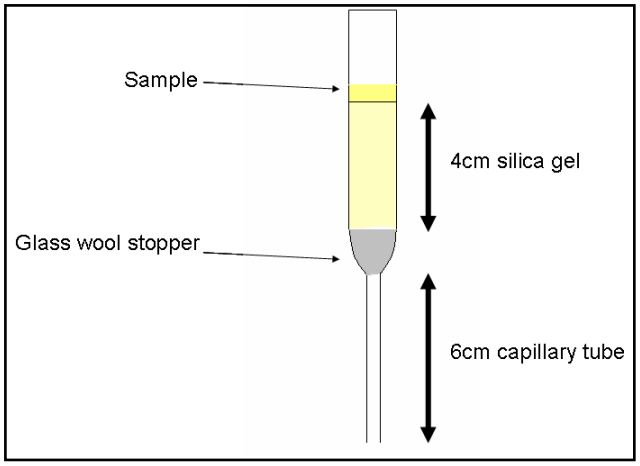
Following extraction, extract was dried by gentle evaporation under N_2, reconstituted in a small volume of DCM, and transferred into pre-weighed sterile 8ml vials, and completely

dried under N_2. Dry samples were stored at -20^{o}C until further processing. TLE mass was recorded as the mass difference between the empty vial and dried sample, and normalised to the mass of soil used for each individual extraction.

2.2: Separation of TLE Acid/Neutral Fractions

Recovered TLE is required to be separated into acidic non-acidic fractions due to the nature of gas chromatography; the polar acid fraction lipids and nonpolar neutral fraction lipids require different columns to be used for GC, and thus cannot be analysed concurrently in a single instrument.

Fractions were separated by silica gel chromatography, with columns constructed within sterilised Pasteur pipettes, with a solid phase of LC-NH_2 SPE Si-gel, covered by a thin layer of combusted sand to prevent the column drying out. Fig[3] displays a schematic of the columns used for fractionation.



Fig[3]: Si-Gel column setup for acid/neutral and N1-N4 separation. Acid/neutral separation utilises LC-NH_2 SPE Si-gel, while N1-N4 separation utilises regular Si-gel. Combusted sand is added to cover the top of the Si-gel column to prevent drying. Column is washed prior to sample loading using 1:1 DCM:ISO (acid/neutral separation) or Hexane (N1-N4 separation).

TLE was re-dissolved in 1:1 Diclhoromethane:Isopropanol, loaded onto the column, and allowed so seep into the Si-gel. Total Neutral Fraction (TNF) was eluted using 4ml of 1:1

DCM:ISO and collected in sterile 8ml sample vials. Total Acid Fraction (TAF) was eluted using 4ml of ether with acetic acid (4%), and similarly collected in sterile 8ml sample vials. At this stage, the fractions may be dried and stored, or used immediately for further processing. The TAF requires cleanup and derivitisation before analysis, while the TNF requires further separation and cleanup, to allow analysis of specific molecular classes (long chain aliphatics, sterols, polar lipids etc).

2.3: Neutral Fraction Separation

TNF further separation was completed with further column chromatography. TNF separation columns were constructed as depicted in Fig[3], however the 4cm LC-NH_2 SPE Si-gel solid phase was substituted with 5cm depth standard Si-gel. TNF was re-dissolved in Hexane and loaded onto the column. TNF was allowed to seep into the column before neutral fractions were eluted sequentially using solvents of increasing polarity, as displayed in Table[1].

Fraction	Solvent	Volume (ml)	Organics Eluted
N1	Hexane	4	Aliphatic hydrocarbons
N2	Dichloromethane	4	Ketones/Esters/Aromatics
N3	Ethyl acetate:Hexane (1:3)	4	Alcohols
N4	Methanol	4	Polar lipids

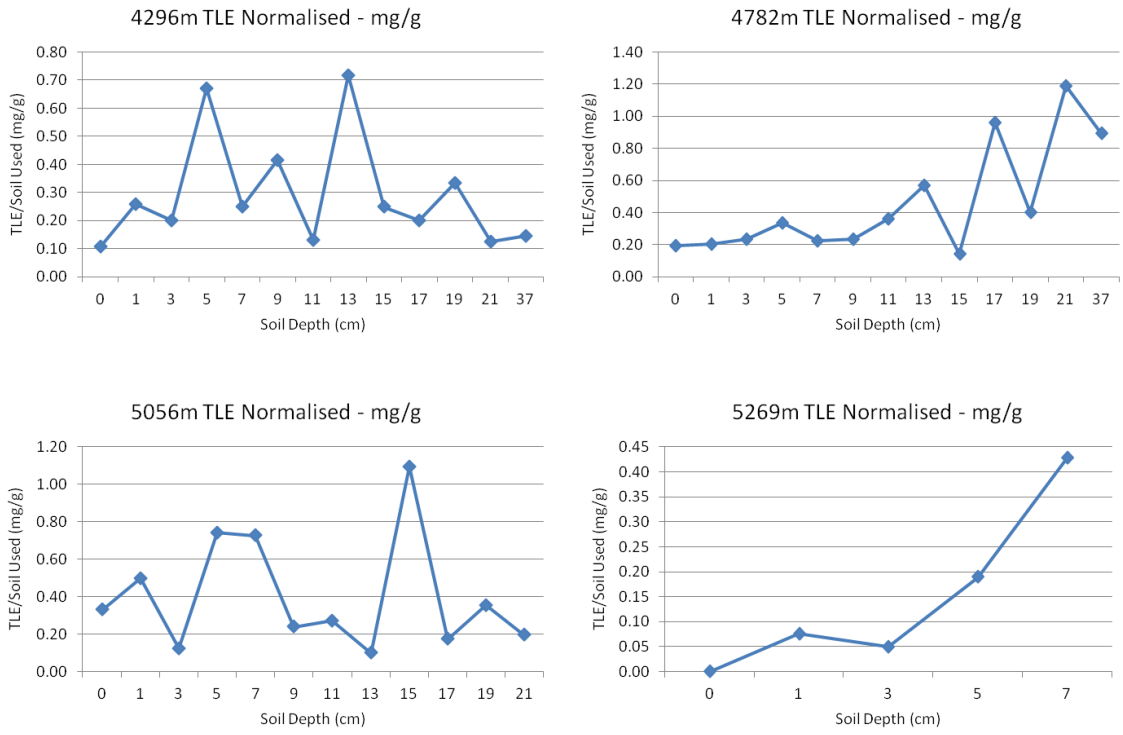
Table[1]: Solvents used for N1-N4 separation, and the classes of organics eluted using each solvent. Solvents of increasing polarity are used to ensure distinct separation of organics by their chemistry.

The N1 and N2 fractions are ready for analysis immediately, while the N3 and N4 fractions require further preparation before analysis. Total Acid Fractions were dried, stored, and archived, as the products were not of relevance to this particular study.

3.0: Results

3.1: Total Lipid Extraction

Total Lipid Extract yield was recorded via drying of extract vials. TLE concentration for each sample was normalised according to the mass of freeze-dried soil used during extraction, to allow for accurate relative comparison of TLE yield between samples. TLE values for each site and soil depth range are displayed in Fig[4].



Fig[4]: Total Lipid Extract yield concentrations of freeze-dried soil samples recovered from each field site. TLE concentration is normalised to mass (g) of soil used in extraction. Samples taken from increasing soil depths allow for visualisation of the change in TLE mass with increasing depth. Site at 4296m (Top Left) was sparsely

vegetated at the surface, and represents a positive control site. Sampling at 5269m (Bottom Right) site terminates at 7cm soil depth due to environmental conditions at the site preventing further excavation.

Total Lipid Extract values for the soil depth profile of each sampling site, as depicted in Fig[4], suggest that TLE yield is approximately stable over the depth range of 37cm; overall TLE yields do not show a strong trend of increase or decrease with depth. Yield is not constant – peaks and troughs in TLE concentrations occur as depth changes, particularly in the 5056m site. Of particular interest is that in all sites except for 5269m, the TLE yield from surface soils (exposed to the harsh surface conditions) was greater than from the deeper, less hostile soils. Compared to recovered genomic content, as previously discussed, recoverable soil lipid content is higher at the surface, from which recoverable DNA was essentially absent. Two potential possibilities arise from this: lipids are more resilient than nucleic acids to the hostile environment, or the shallow depth lipid content is at least partially accounted for by windborne particles from sources such as plant leaf waxes (Gagosian et al, 1981), or indeed both. It has been observed that plant leaf waxes removed from living plants via ablation readily are dispersed by atmospheric action (Conte and Weber, 2002), and thus windborne content may well prove a factor in this study.

3.2: GC-FID Analysis

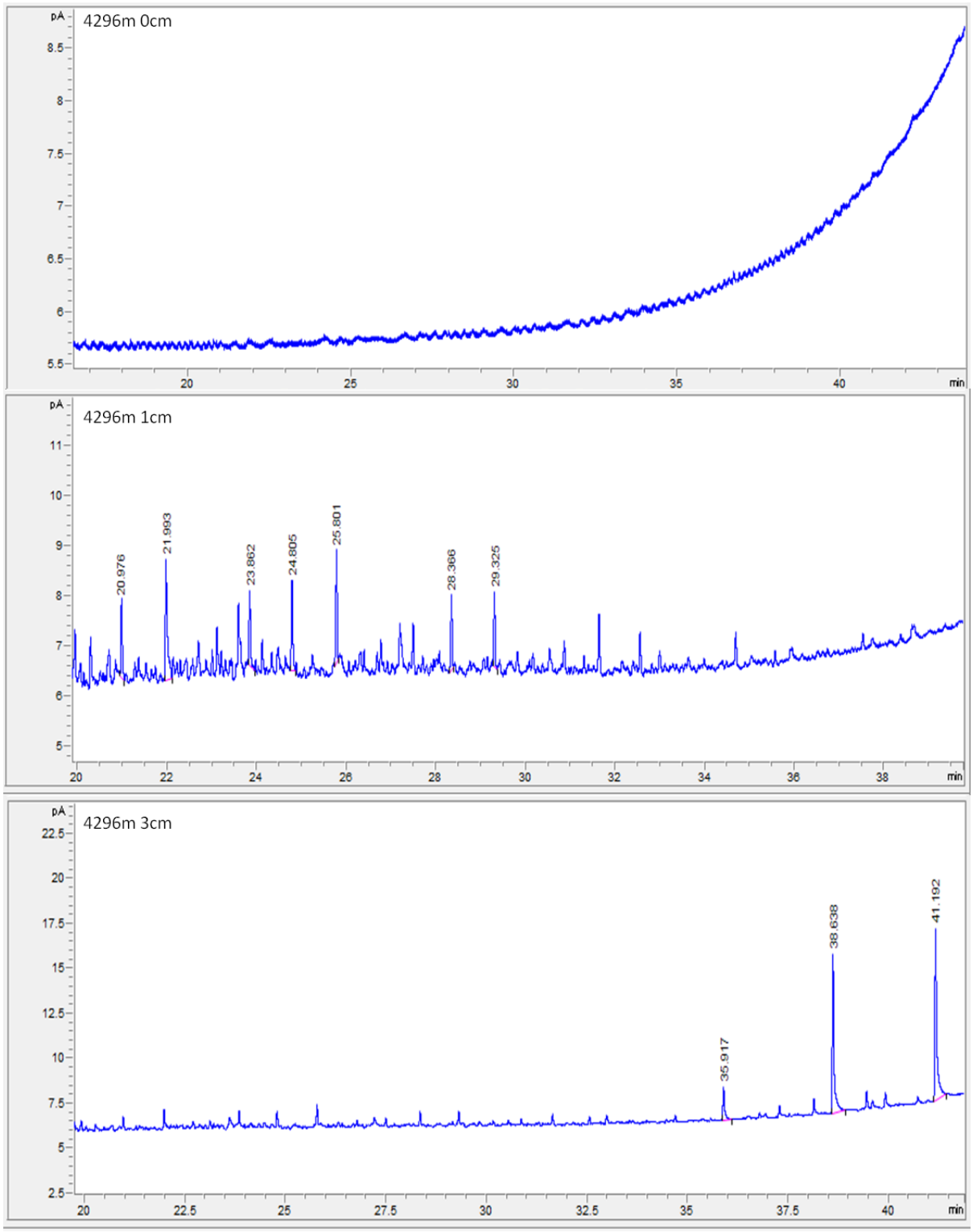
Neutral fraction samples from field sites at 4296m and 5056m were selected for further analysis, allowing for a notably vegetated control site (4296m) and a barren high-altitude site of primary interest (5056m), in addition this maintains continuity of investigation with previous work in the study. Samples from all soil depths were chosen for analysis, allowing accurate observation of changes with increasing depth below the soil surface, and therefore changes as relative environmental hostility is reduced.

Neutral fraction samples comprised of long-chain aliphatic compounds (N1 fraction) were selected for gas chromatography – flame ionisation detection (GC-FID) analysis; these fractions in particular were determined to be of greatest interest to the study as these would

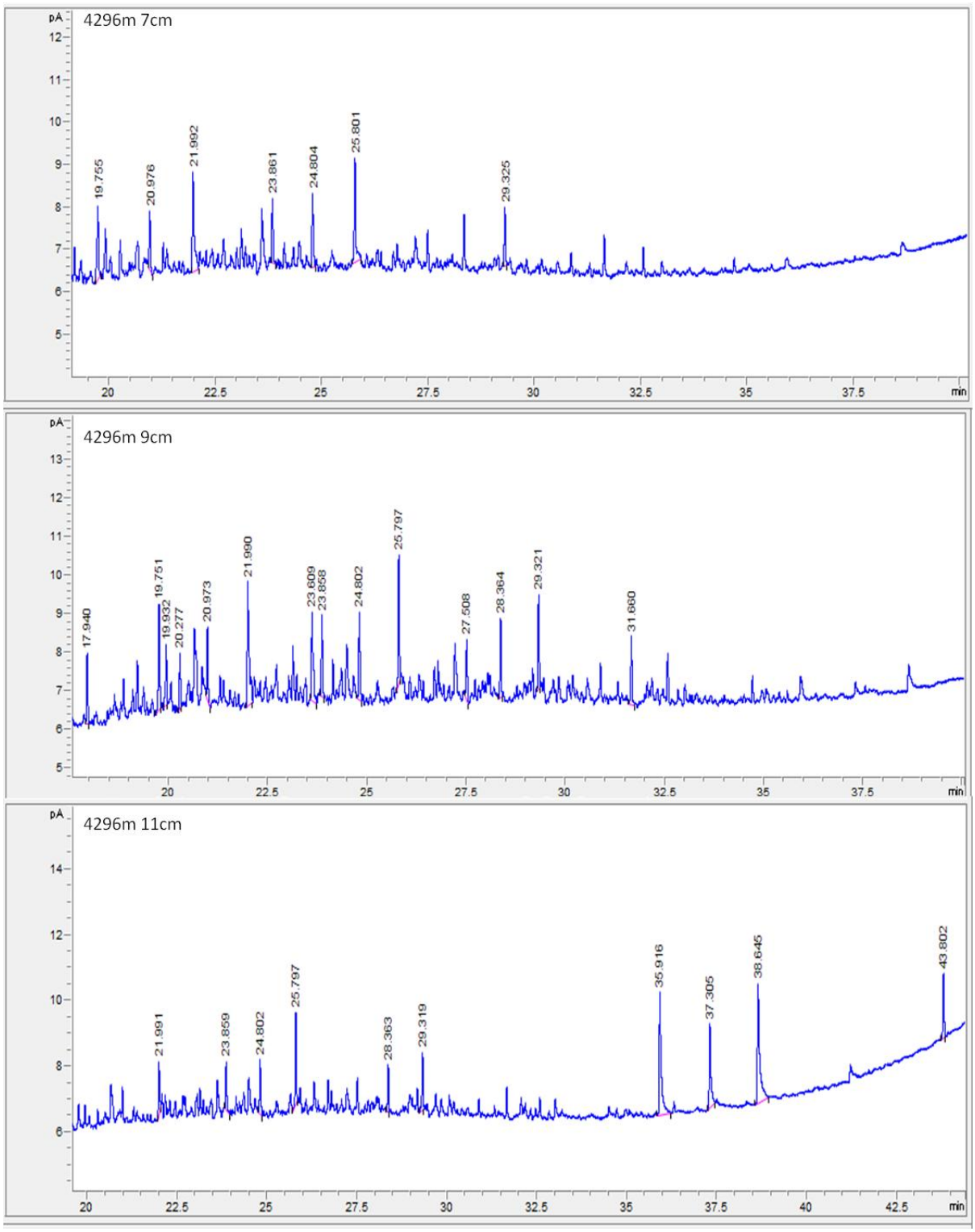
allow for a range of biological markers such as plant leaf waxes and membrane components such as cholesterol.

Leaf waxes were primarily of interest due to the potential of windborne particles suggested by the relatively high TLE yield from near-surface soil layers; no surface plants were present at the 5056m site, yet surface TLE yield was observed to be relatively high in the same manner as was observed from 4296m TLE yield, and thus windborne particles are a possible source.

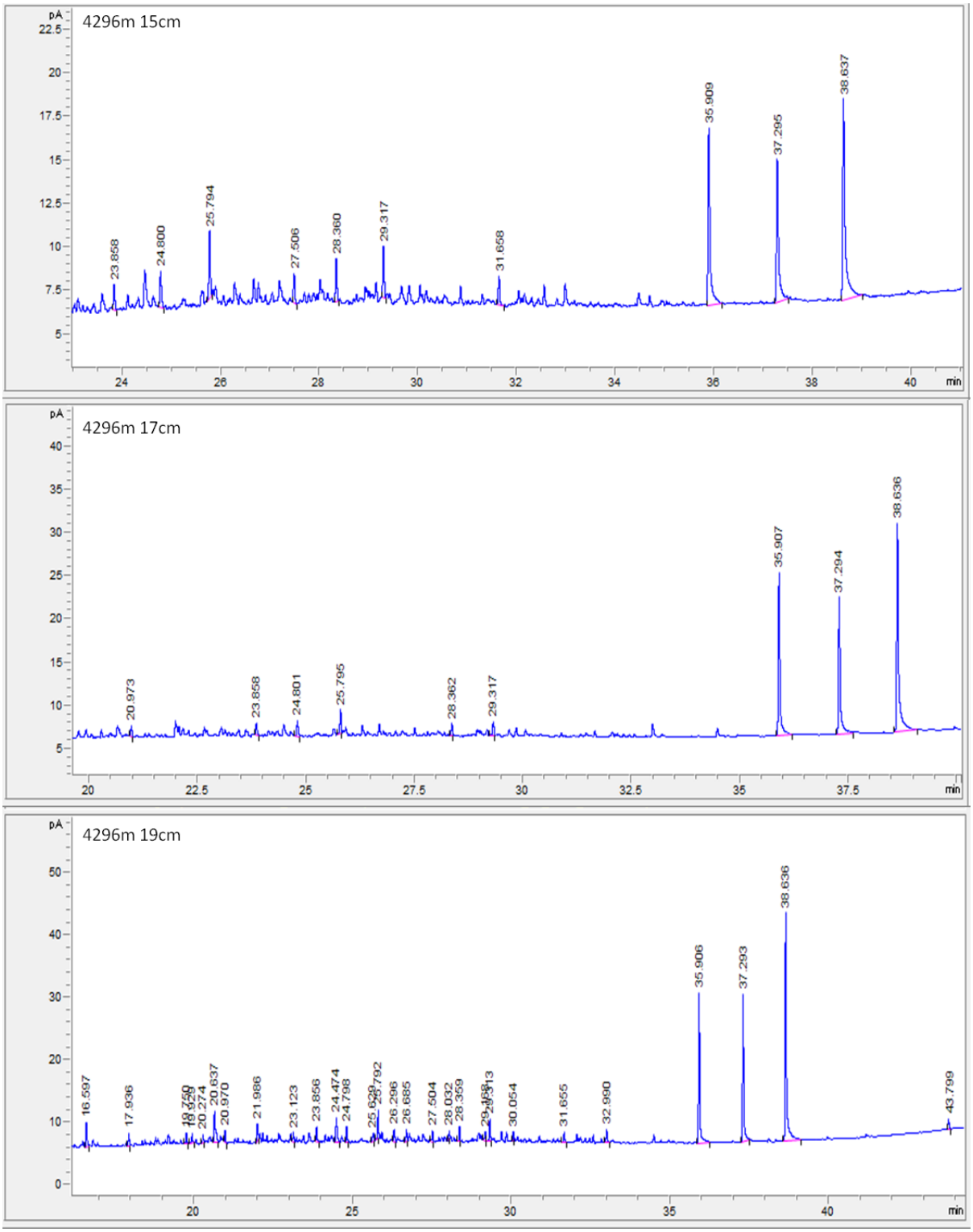
Peaks of interest detected during GC-FID analysis are presented in Figs[5-8] (4296m site samples). Peaks of interest are those which feature a max height notably greater than the baseline detector noise, and are automatically reported by the detector software. Peaks are labelled according to retention time.



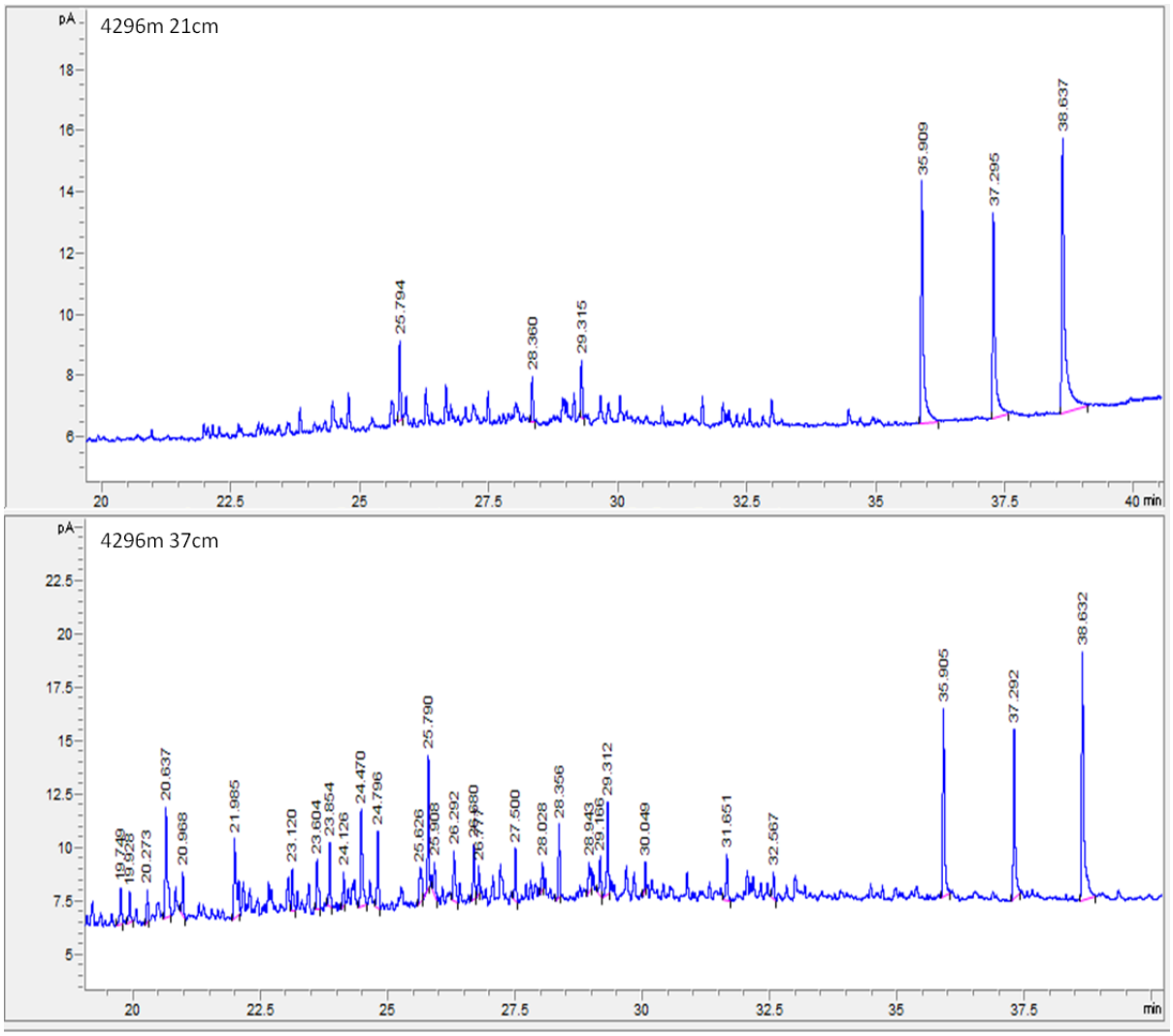
Fig[5]: Displaying GC-FID analysis output of N1 fraction samples originating from 4296m soil samples over soil depth range 0-3cm. Focus is on retention times approximately 20-40 minutes. The 0cm surface soils exhibit no prominent peaks, but merely 1cm deep into the soil, numerous prominent peaks are observed, with greater magnitude peaks present in 3cm soils.



Fig[6]: Displaying GC-FID analysis output of N1 fraction samples originating from 4296m soil samples over soil depth range 7-11cm. Focus is on retention times approximately 20-40 minutes. With a wide range of prominent peaks during the retention time period of focus, it is evident that long chain aliphatic compounds are well preserved in the extracted soils.

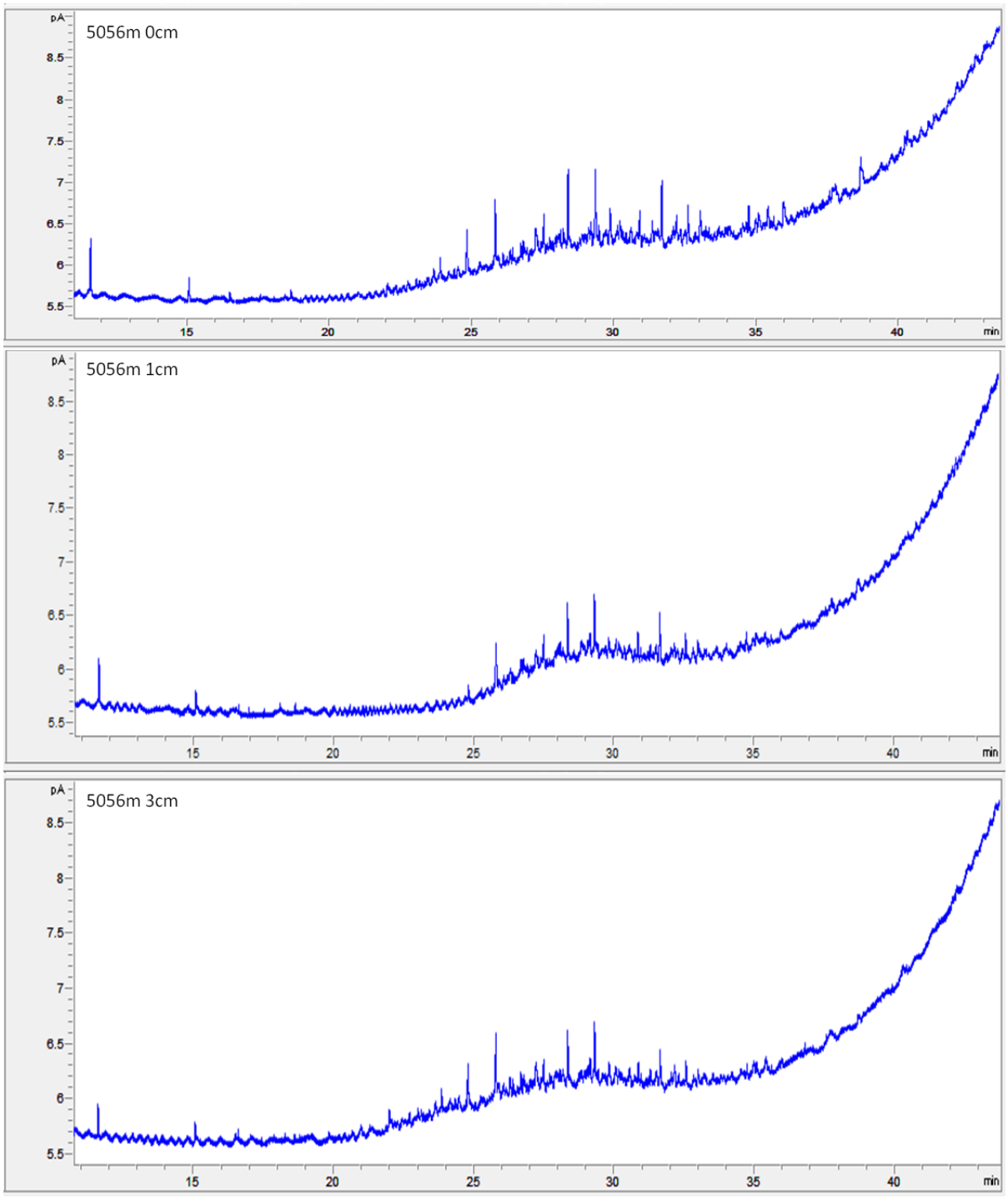


Fig[7]: Displaying GC-FID analysis output of N1 fraction samples originating from 4296m soil samples over soil depth range 15-19cm. Focus is on retention times approximately 20-40 minutes. In these lower soil depth samples, prominent peaks are abundant, with a wide range of aliphatic compounds detected.

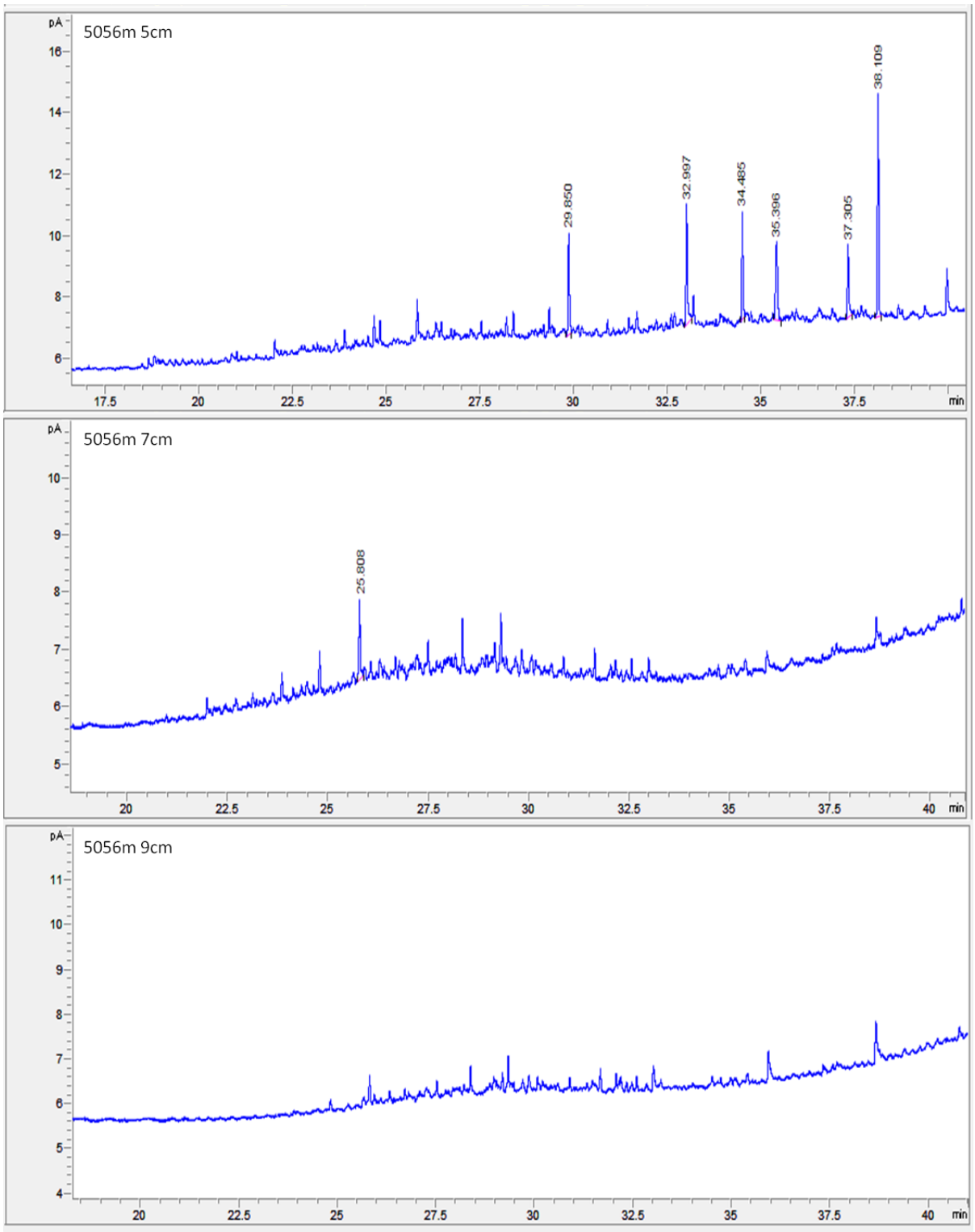


Fig[8]: Displaying GC-FID analysis output of N1 fraction samples originating from 4296m soil samples over soil depth ranges 21-37cm. Focus is on retention times approximately 20-40 minutes. Abundance of prominent N1 peaks continues at the lowest depths of soil sampling, indicating a soil rich with long chain N-alkane compounds.

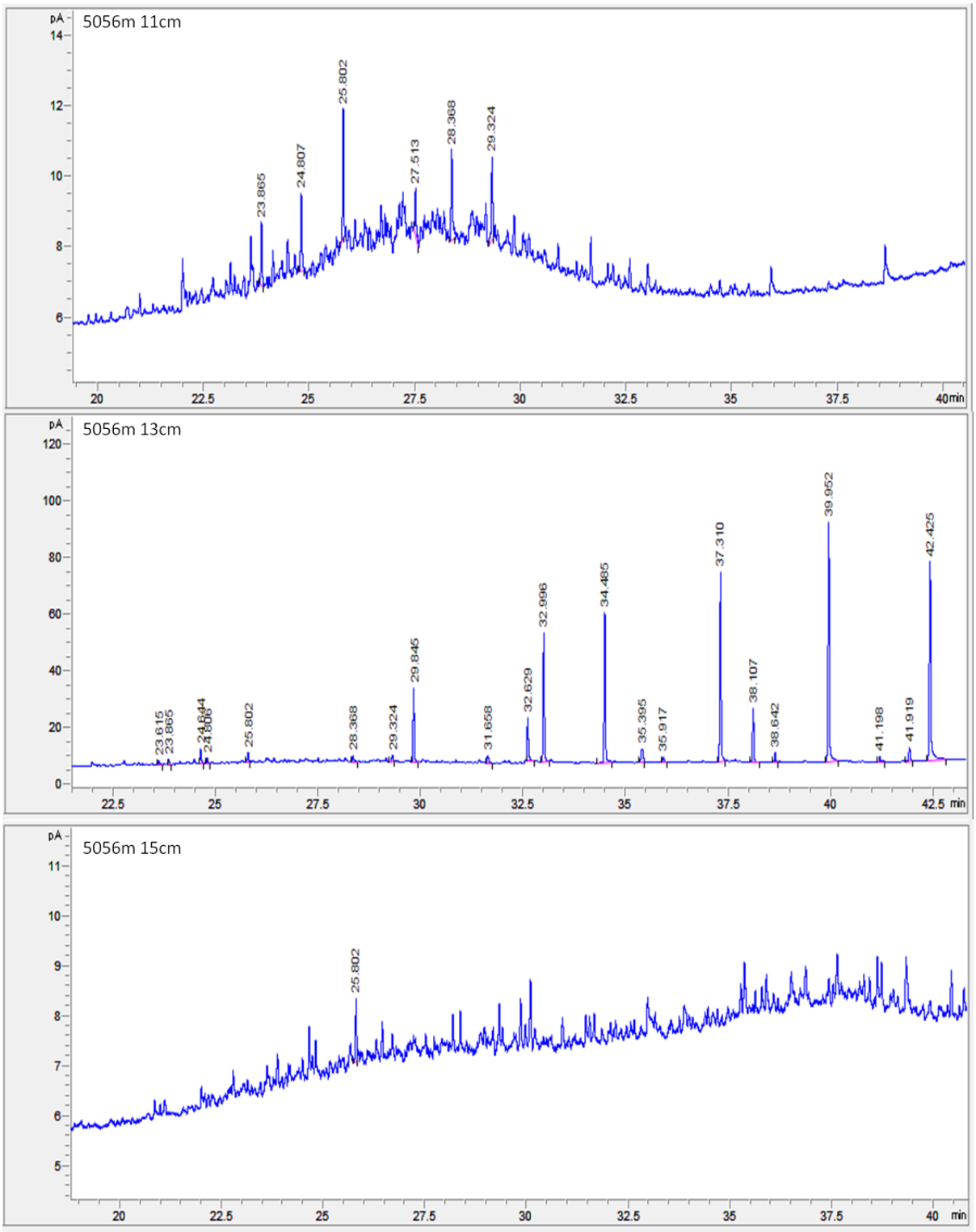
Peaks detected during GC-FID analysis for 5056m N1 fraction samples are presented in Figs[9-12], presented with increasing soil depth. Prominent peaks of interest are labelled, and focus in retention times between approximately 20-40m has been maintained to allow comparison between content of the different field sites.



Fig[9]: Displaying GC-FID analysis output of N1 fraction samples originating from 5056m soil samples over soil depth range 0-3cm. Focus is on retention times approximately 20-40 minutes. No notable peaks are reported in these samples, and the trace displays only standard machine noise.

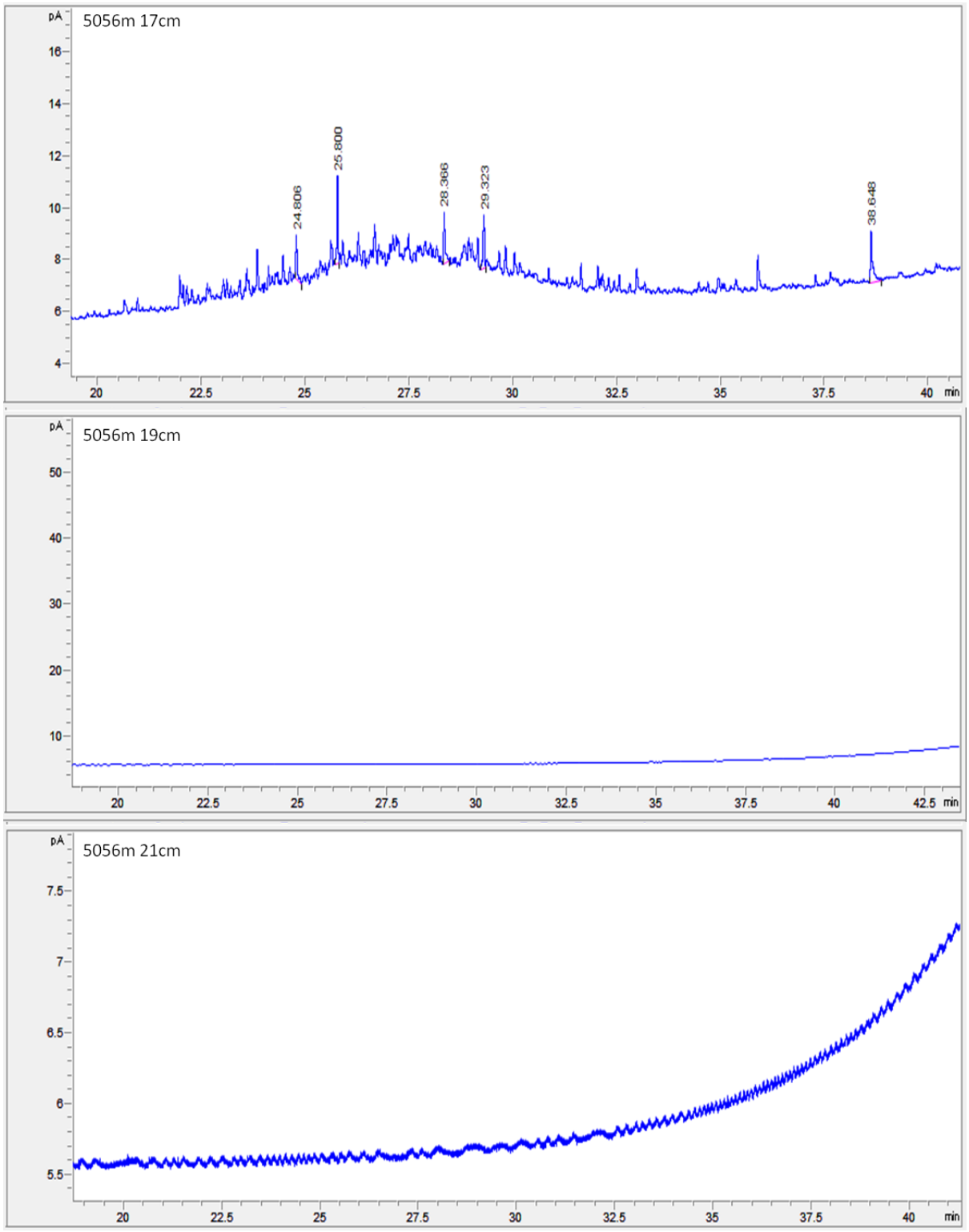


Fig[10]: Displaying GC-FID analysis output of N1 fraction samples originating from 5056m soil samples over soil depth range 5-9cm. Focus is on retention times approximately 20-40 minutes. Several prominent peaks are reported in 5cm soil extracts, while only a single prominent peak is observed from 7cm soil extracts. 9cm soil extract features no prominent peaks.



Fig[11]: Displaying GC-FID analysis output of N1 fraction samples originating from 5056m soil samples over soil depth range 11-15cm. Focus is on retention times approximately 20-40 minutes. Numerous prominent peaks are observed in 11cm extract-derived sample, while 13cm sample features a significant numbe r of prominent

peaks, indicating a relatively rich soil origin. 15cm extract sample features a single prominent peak.



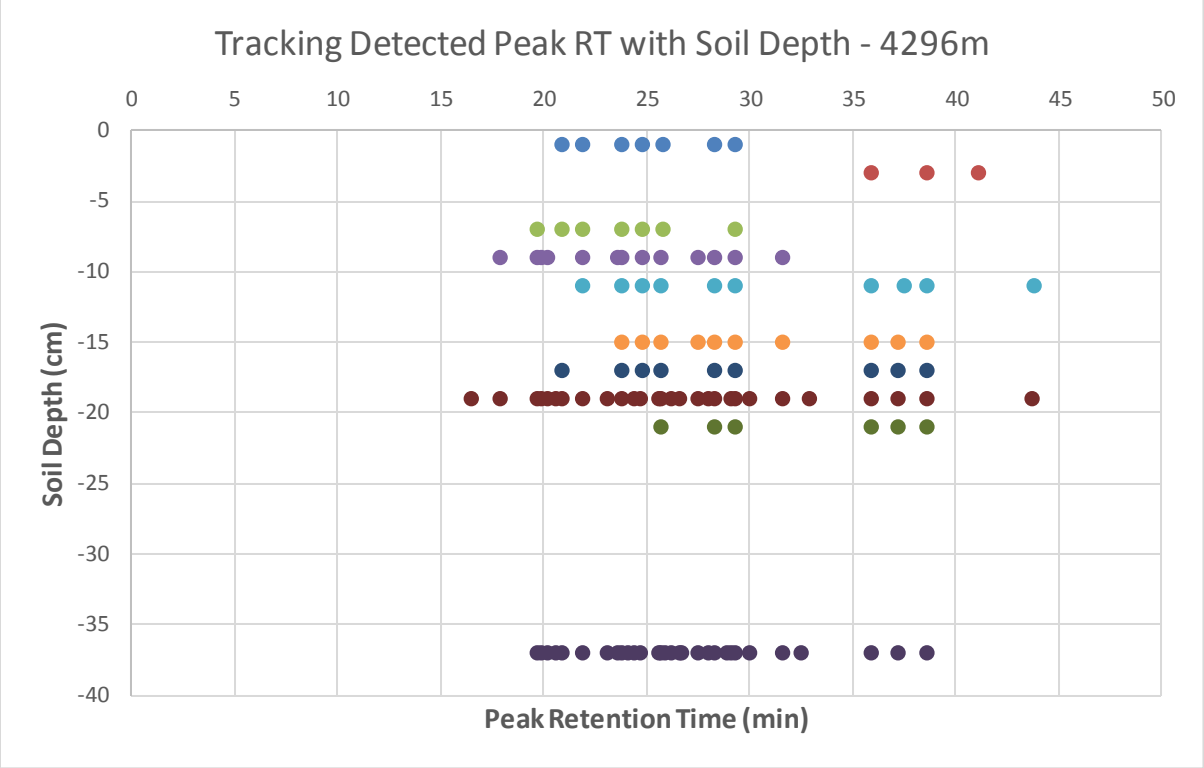
Fig[12]: Displaying GC-FID analysis output of N1 fraction samples originating from 5056m soil samples over soil depth range 17-21cm. Focus is on retention times approximately 20-40 minutes. Sample at 17cm features several prominent peaks in the range, while 19 and 21cm-origin samples display no prominent peaks.

There is a notable trend in prominent peak count as soil depth changes in the 5056m set of soil extract long chain aliphatic compounds (N1 Fraction). Samples obtained from near-surface soils such as the 0cm-3cm samples shown in Fig[9] display no notable peaks, indicating that the TLE yield from these soils is largely devoid of any significant mass of long-chain aliphatic compounds.

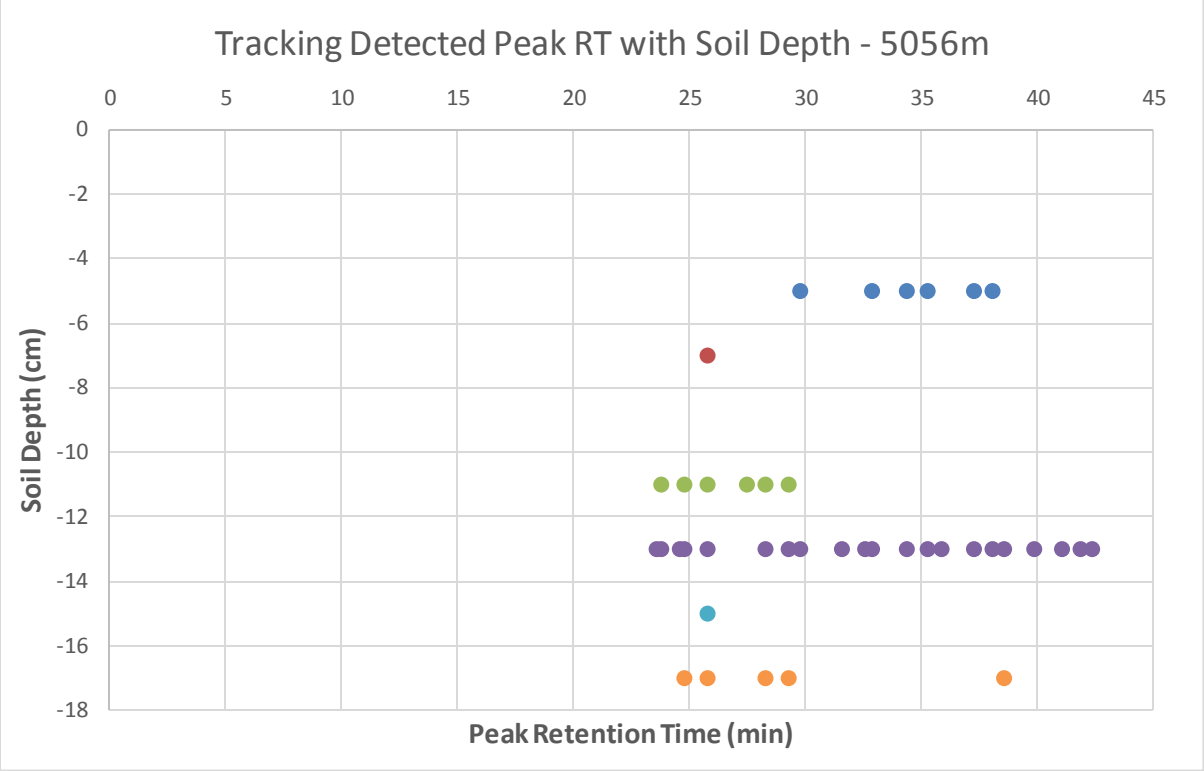
It is evident that the soils of the lower altitude control site at 4296m are noticeably more rich in preserved lipid and lipid-derived compounds than the high altitude 5056m primary analogue site of interest for this study. This relative richness of the 4296m site is illustrated by the increased diversity and count of prominent peaks observed during GC-FID analysis of the N1 fraction samples purified from total lipid extract of the soils.

3.2.1: Examining change in detected peaks with soil depth

Retention times of GC-FID chromatogram peaks for each site and soil depth analysed are presented in Figs[13] and [14] below. Soil depth of original TLE extract sample is displayed for each distinct sample, with all notable peaks detected displayed to better facilitate comparisons between samples.



Fig[13]: Presenting GC-FID chromatogram notable peaks of 4296m site samples according to soil depth and retention time (RT). Samples and notable peaks represent those depicted in chromatograms of Figs[5-8].



Fig[14]: Presenting GC-FID chromatogram notable peaks of 5056m site samples according to soil depth and retention time. Samples and notable peaks represent those depicted in chromatograms of Figs[9-12].

The collation of all relevant GC-FID chromatograms into a pair of figures in Figs[13] and [14] allow for the clear observation of changes in detected compounds in both sites as depth below the surface increases. In these plots, chromatograms lacking in notable peaks (such as the 0, 1, and 3cm samples of 5056m site presented in Fig[9]) are excluded – this is particularly prevalent for the 5056m samples.

Immediately notable is the difference in relative richness of identified compounds between the soils of the different field sites; the 4296m site data in Fig[13] clearly indicates a greater richness of peaks than the 5056m site data of Fig[14], and indeed this has been reflected throughout the study for all factors of organic richness.

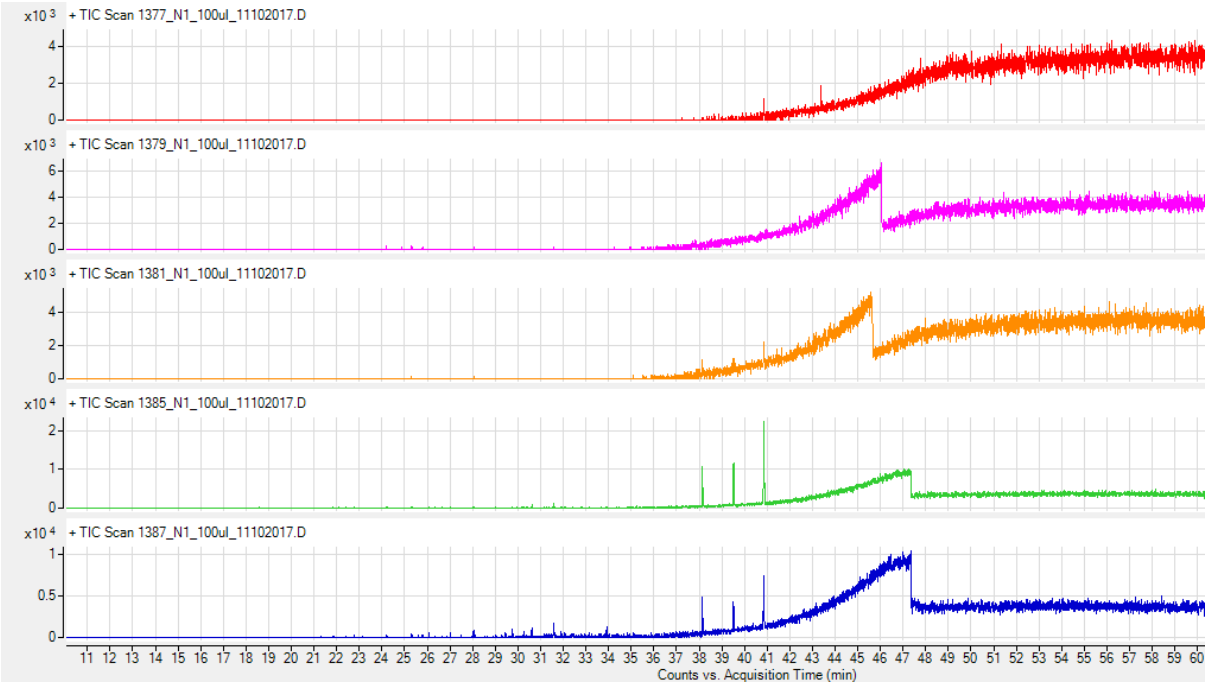
In addition to the comparative difference in richness between the sites, there are observed differences in the specific ranges of detected compounds in each site. For example, the lower altitude site features an overall wider range of detected compounds based on retention time, as the 5056m samples are lacking any peaks of retention time <20m. Further, the two sites

exhibit difference in the change in range of detected compounds as depth increases, as while the 4296m site soils overall do not display notable change in the range of peak RTs reported, the soils of the 5056m site do notably shift from predominantly higher retention time peaks near the surface, to lower RT peaks in the deeper soils.

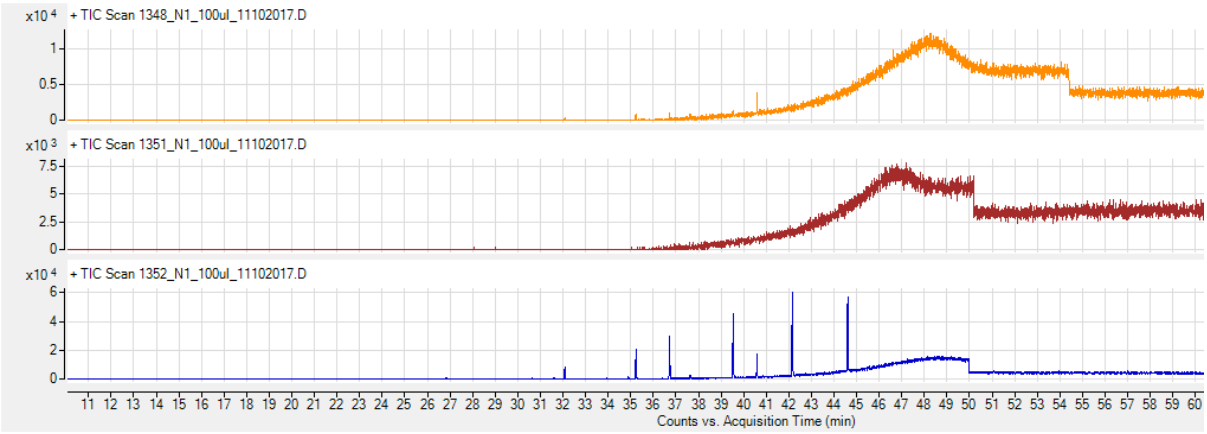
The apparent general richness of peaks of greater retention time in the shallower soils of the 5056m site may indicate that the surface soils are richer in longer-chain compounds in comparison to the deeper soils, with the understanding that larger molecules typically feature longer retention times.

3.3: GC-MS Analysis

Samples were selected for further processing via GC-MS analysis according to unique peak count and prevalence. Samples featuring peaks at unique retention times were automatically included for further processing, and from the remaining samples, samples were selected to provide the greatest coverage of all relevant peaks across the sample pool, for the purpose of efficiency with regards to instrument time and reagent usage. Selected samples for GC-MS analysis from 5056m sample set were: 1348 (5cm depth), 1351 (11cm depth) and 1352 (13cm depth). Selected samples for GC-MS analysis from 4296m sample set were: 1377 (3cm depth), 1379 (7cm depth), 1381 (11cm depth), 1385 (19cm depth), and 1387 (37cm depth). This selection of samples allowed for best coverage of all detected prominent peaks during GC-FID analysis with minimal sample requirement. Figs[15] and [16] display the chromatographs for each set of samples, allowing visualisation of noted peaks under GC-MS analysis. Peaks of note were isolated and selected for spectra matching using the NIMS software coupled to the instrument.



Fig[15]: Displaying GC-MS chromatograms of selected 4296m site N1 fraction samples. Samples presented represent 3cm, 7cm, 11cm, 19cm and 37cm soil depth samples, chosen to cover all representative peaks observed in GC-FID chromatograms featured in Figs[5-8].



Fig[16]: Displaying GC-MS chromatograms of selected 5056m site N1 fraction samples. Samples presented represent 5cm, 11cm and 13cm soil depth samples, chosen to cover all representative peaks observed in GC-FID chromatograms featured in Figs[9-12].

It is immediately evident that notable peaks in chromatograms become more prevalent as soil depth increases, with the deepest depth sample(s) in each set of data displaying prominent

peaks, although none are particularly large, indicating the low concentration of compounds in the fraction samples. Notable peaks have been identified using National Institute of Standards and Technology (NIST) Library software associated with the instrument (see: nist.gov). Table[2] displays the compounds identified in notable peaks, while full mass spectra of all identified compounds are presented in Appendix 1

Sample ID	Soil Depth (cm)	Peak RT (min)	Matched Compound	Match %
1348	5	40.5	5β-cholestane	71.0
1351	11	28.0	1,3-methyl-4-penten-2-ol	14.5
1352	13	32.0	Decyl-2-ethylhexyl sulfite	12.6
1352	13	35.2	Nonadecane	7.0
1352	13	36.7	Octacosane	9.4
1352	13	39.5	Octacosane	32.5
1352	13	40.5	Cholestane	59.9
1352	13	42.1	Octacosane	11.4
1352	13	44.6	Tetratetracontane	16.6
1377	3	40.8	Methylpentenol	46.0
1377	3	43.3	Didecyl sulfone	34.8
1379	7	25.3	1,3-methyl-4-penten-2-ol	14.5
1381	11	39.5	2' - (trimethylsiloxyl) - propriophenone	54.8
1385	19	31.5	3,6 - heptanedione	46.4
1385	19	38.1	2 - ethylhexylisohexyl ester	4.8
1385	19	39.5	dodecyl -2 - propyl ester	6.3
1385	19	40.8	Octacosane	10.4
1387	37	31.5	2,7 - dimethyloctane	41.0
1387	37	33.9	Iodononane	49.0
1387	37	38.1	Decyl isopropyl sulfite	18.0
1387	37	39.5	2,3,5,8 - tetramethyldecane	4.0
1387	37	40.8	Pentyl undecyl sulfite	5.0

Table[2]: Displaying matched compounds of notable peaks in GC-MS chromatograms for all selected samples. Spectrum matching was accomplished using NIST MS Library software, accepting highest percentage match for each compound identified. Peak retention time refers to peak RT depicted in Figs[15] and [16].

Particular focus should be applied to the percentage match column of Table[2]; it is notable that in many cases the higher match values are associated with 'simple' compounds such as Cholestane, Octacosane, and Iodononane, although this is not uniformly the case. Low percentage matches such as, for example, 7.0% match for Nonadecane in the 1352 (13cm, 5056m site) sample highlight the difficulty in identifying compounds extracted from these soils, suggesting possibly that lipid content recovered from the soils is fractured, or that the true compound is a derivative of the matched compound. This may be also due to low concentration of the compound of interest in the sample.

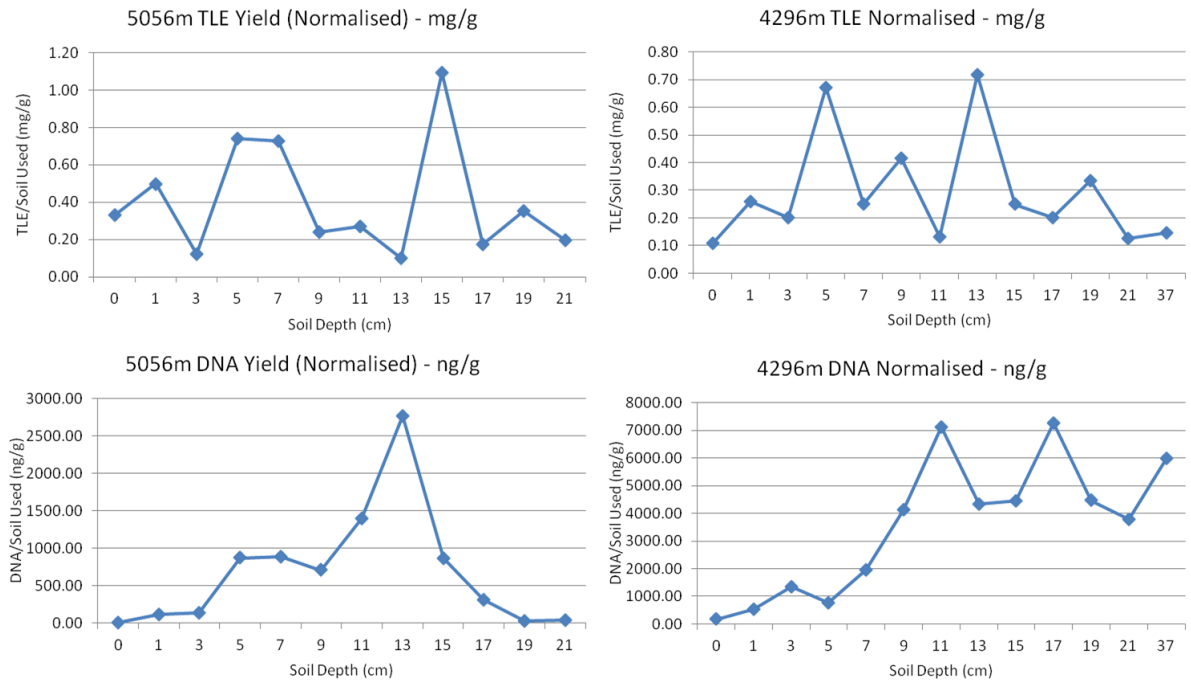
4.0: Discussion

4.1: Total Lipid Extract

Normalised Total Lipid Extract yield for all soil samples collected at all field site locations are displayed in Fig[4]. Yield values for TLE are normalised against the precise mass of soil used for the relevant extraction, allowing a more accurate comparative observation of difference in TLE yield between soil depths as well as between sites. The typical TLE yield is observed to increase with soil depth, although the correlation is stronger the extract of some sites than others. TLE yields were high at the 'boundary' between upper and lower soils as previously discussed (See Chapters 1 and 2), indicating that the increase in soil moisture – which was noted around this depth – may be related to increase in recoverable organic matter in the soils. Notably, when comparing TLE yields between the high and low-altitude sites of primary interest (5056m and 4296m), it is observed that TLE yields are approximately similar, with no site exhibiting soils of particularly greater lipid content.

In comparison to the extract yield of whole genomic content – DNA – from the same soils, the extract yields of TLE followed less of a coherent trend against soil depth; DNA extract yields were observed to notably rise at approximately 11cm soil depth, and then subsequently tail off at lower depths. This was the case in all sites of interest, with the exception of sites at which excavations did not occur for depths as far as 11cm and greater below the surface, such as the 5269m site, at which excavations were halted at 7cm due to the soils below this depth being frozen solid and not able to be excavated by the equipment available. Fig[17] compares

the TLE (normalised to soil mass) yield with whole genomic content yield for the 5056m and 4296m sites, to illustrate the differing trends of yield vs soil depth.



Fig[17]: Comparing normalised Total Lipid Extract yield with normalised Whole Genomic Content (DNA) yield for 5056m and 4296m soils.

Total Lipid Extract yield from soils recovered at the 5056m primary interest site is observed to approximately mirror the whole genomic content extract yield of the same soils, in consideration of change in relative yield with increasing soil depth. There is a minor increase in yields evident at approximately 5cm, which is the case for both lipid and nucleic acid extracts, but notably the TLE yield exhibits a comparatively more pronounced increase, which may be due to increased stability of lipids in dry environments compared to nucleic acids, particularly as many lipids are insoluble in water. A possible source of increased lipid-derived content near the soil surfaces at the sites is compounds such as long-chain N-alkanes; simple (C_nH_{2n+2}) hydrocarbon chain compounds which are prevalent in plant leaf waxes are known to be dispersed by windborne transport (Nelson et al, 2017) – a plausible occurrence for the sites of interest for the present study as much of the surrounding lower-lying Altiplano region is vegetated, and a prevailing wind would serve to transport ablated leaf wax matter across the region.

Notably, in comparison to previous data concerning whole genomic content extract yield, differences between extracts from the 5056m site and the 4296m site are minimal; DNA extract yield for soils of the 4296m site was considerably greater than DNA extract yield of soils of the 5056m site, while this does not appear to be the case for extracted lipid content, where neither site is particularly richer than the other. This suggests that lipid-derived organics are perhaps more stable in the environment than naked nucleic acids without being a component of a live (or preserved) cell.

In examining the lipid yields observed from each site, it is of import to consider the sampling procedure employed in this study – initial soil samples were stored in plastic whirlpak bags, while convention when working with lipids is to use glassware for all sample manipulations and storage. Due to the plastic nature of the sample bags, it is potentially possible for leaching of hydrocarbon compounds from the bags if they were to come into contact with organic solvents such as those utilised in the processing of lipids. This is extremely unlikely to be a factor in the present study due to the methodology of sampling, in which no organic solvents of any kind were allowed to interact with the soil samples or the storage bags – while 100% ethanol was used as a sterilising agent for the trowel used for extraction, care was taken that any ethanol on the tool was entirely evaporated before being introduced to the soil. In future studies it may be advisable to use a blank of combusted sand to examine any potential leaching from whirlpak bags – or indeed to sample directly into glassware to remove the potential from the start.

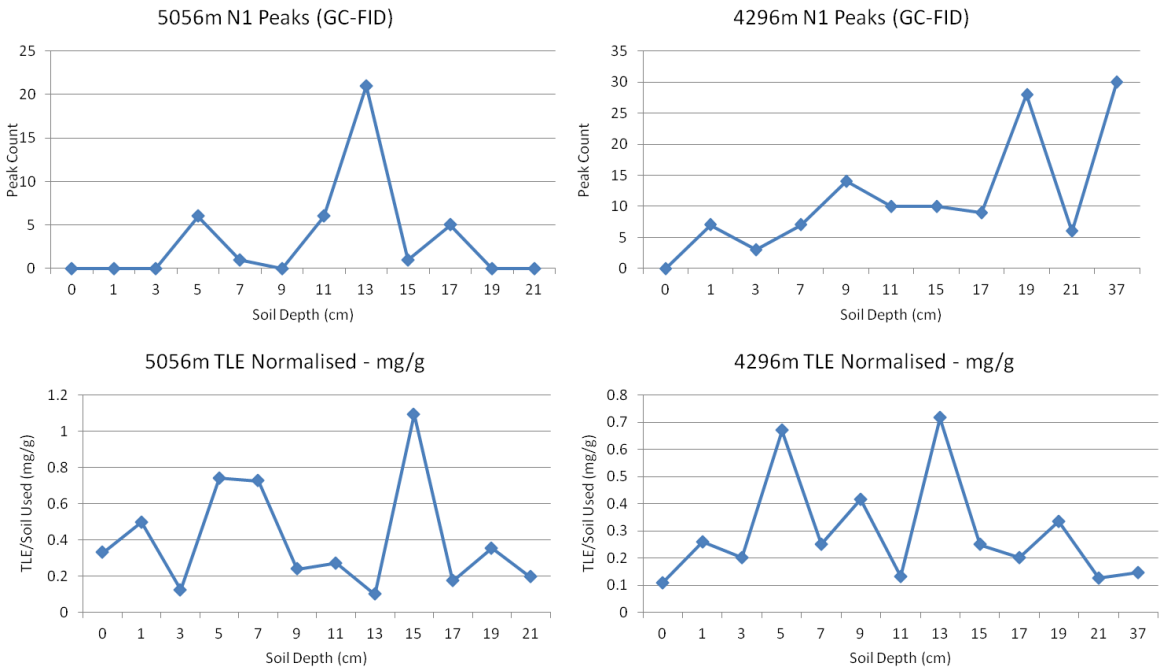
In consideration of the likely prevalence, and relative chemical simplicity of N-alkanes, this particular fraction of the lipid extract was chosen for further investigation using GC-FID and GC-MS analysis.

4.2: GC-FID Analysis

Gas Chromatography-Flame Ionisation Detection analysis was utilised to detect and visualise the range of compounds present in each lipid extract sample, with particular focus on the N1 (long-chain N-alkanes) fraction of the TLE. Samples were separated according to site of

interest, and run in order of increasing soil depth. Standard mix samples and a solvent blank were used to confirm lack of sample contamination or instrument issues such as column damage or injector syringe contamination.

Peak count for N1 samples exhibited a notable difference between the 5056m site and the 4296m site. Most significantly, the lower altitude control site exhibited a far greater range of detected prominent peaks, across a greater range of soil depths than the 5056m site, indicating that the soils in general are richer in lipid content at the lower altitude site than the higher altitude analogue site. The high altitude site features a range of N1 peak counts that closely mirrored the actual TLE yield progression with soil depth, while this was not the case for the lower altitude vegetated site. These differences are displayed in Fig[18].



Fig[18]: Displaying N1 fraction prominent peak count under GC-FID analysis, in comparison to overall TLE yield for each soil sample. It is evident that peak cont and overall TLE mass trends display similarity in the case of 5056m site soils, yet this is not the case for 4296m site soils.

As shown in Fig[18], the TLE yield and N1 peak counts vs soil extract depth follow a highly similar trend, which indicates that aliphatics are a representative component of the TLE profile for much of the soil depths. This is of particular interest as the collected soil samples,

and indeed the site in general, were devoid of visible plantlife – to the naked eye the site appears entirely free of vegetation. This could indicate that any plant-origin lipid matter may be accounted for by windborne particlate such as the previously-discussed leaf wax alkanes. If this were the case, then the presence of lipid content below the soil surface would be potentially accounted for by deposition of fresh matter at the surface, and the consequential burial of previously deposited matter, which is preserved in the soil layers.

In contrast, the peak count profile of the N1 fractions recovered from 4296m soil extracts, as shown in Fig[18], do not follow the change in prevalence with soil depth which is observed in the TLE yield vs soil depth. Peak count of N1 fractions of this site is observed to generally increase with soil depth, but not in line with the TLE yield of the soils, indicating that the N-alkane fraction of the lipid profile does not account for a significant portion of the total lipid content, and other lipids and lipid-derived compounds are present in the soils.

It is notable in Fig[14] that the retention time of peaks in the 5056m site soils appears to differ between the shallow and deeper soils, specifically that shorter retention time peaks appear to be supported to a greater extent in the deeper soils – critically, past the 11cm soil depth at which a notable increase in soil moisture content was observed in Fig[11] of Chapter 1. It is possible that this is the result of biological action in the soil breaking down longer chain organics over time, with the windblown soils causing the eventual burial of what were previously surface soils, or alternatively that the moister environment is able to harbour a different profile of lipids. Indeed, both may be feasible, and further examination is advised, in particular recording the scale of wind transport of soils in the region would be invaluable in answering this question.

The presence of greater diversity of lipid content in the soils of the 4296m site is supported by the nature of the site – the 4296m site is notably vegetated, with plentiful scrub grasses and other vegetation growing freely around the area. In addition to this, insect life was notable in the site, as well as evidence of herbivorous mammals such as Llama in the region. The presence of vegetation as well as animal presence will add to the lipid profile of the soils, which would not be accounted for by the N1 fraction, for example faecal sterol biomarkers may be persistent in these soils due to roaming herbivores (Tyagi et al, 2007), and so it is logical that a comparatively lesser component of the total lipid profile of the soils is accounted for by long-chain aliphatics. In addition, the effect of the presence of animal life

(and indeed vegetation) indicates that the lower altitude site is less viable as a Martian analogue than the higher altitude site which is less rich in complex organics.

The richness of the soils with regards to N-alkanes and other simple long-chain aliphatics is notably different between the two field sites. This difference in relative richness of aliphatics in the soil is observed firstly through the simple diversity of peaks detected in each soil depth sample for each site; samples recovered from 5056m soils typically feature lower peak counts than 4296m site soils. In addition, average peak heights are lower in 5056m N1 fraction samples than 5296m N1 fraction samples, which is indicative of a lower overall concentration of N-alkanes in the high altitude soils than the lower altitude control site. The normalised TLE yield shown in Fig[4] indicates relatively comparable lipid extract yields between sites, and it has been established previously in Fig[18] that the N1 fraction lipids may represent a greater proportional component of the TLE in the 5056m site than the 4296m site, and thus it is likely that the N1 fraction accounts for the primary component of the 5056m TLE samples.

4.3: GC-MS Analysis

Analysis of N1 fraction samples by Gas Chromatography – Mass Spectrometry analysis enabled identification of nonpolar compounds present in Total Lipid Extract of soil samples from 5056m and 4296m sites, over a range of soil depths.

Table[2] highlights several distinct classes of molecule which are identified in the samples via NIST MS Library matching: N-alkane type compounds, alkyl esters, steranes, and sulfones. Examples of N-alkane type compounds detected in samples include Nonadecane, detected in sample 1352 (5056m 13cm depth), Octacosane, also detected in sample 1352 and additionally in sample 1385 (4296m 19cm depth), and Tetratetracontane, a significantly long-chained hydrocarbon (C40) detected in sample 1352. Branched long-chain hydrocarbons such as 2,3,5,8 – Tetramethyldecane and 2,7 – Dimethyloctane were detected in sample 1387 (4296m 37cm). Very long chain N-alkanes such as these are common components of plant cuticle wax (Chowdhury et al, 2010), as they offer significant protection against UV radiation damage (Holmes and Keller, 2002) – a factor of some significance to the environment under investigation in this study. The presence of leaf wax compounds in the soils of both sites and at several soil depths indicates the utility of these molecules as biomarkers; the capacity to

resist UV-mediated damage is instrumental to the preservation of any potential biomarkers on Mars, as they must persist in the environment for sufficient time to become buried below the surface due to soil deposition by wind.

Several alkyl ester compounds were identified by GC-MS analysis of samples: sample 1385 was determined to contain dodecyl -2 - propyl ester as well as 2 - ethylhexylisohexyl ester. Notably, no ester compounds were identified in 5056m samples, while several were in 4296m samples. This is notable as the N1 fraction should not include esters, which are eluted subsequently as part of the N2 fraction. This may be the result of either low quality NIST library matching, or cotamination. These compounds are not typically subject to wind dispersion to the same extent as aliphatic compounds, as has been noted previously by studies such as Simoneit et al (1990), and those identified in samples are likely to have been produced locally. Esters, therefore, may be less suitable as biomarkers in hostile environments such as the Martian surface where wind dispersion is a likely feature.

Sterane compounds detected in samples include 5β-cholestane, and cholestane. Steranes such as this are the parent molecules of biologically-vital molecules such as sterols and cholesterols, which are key components of cell membranes (Valitova et al, 2016). 5β-cholestane in particular is a reliable plant biomarker and has been reported to persist in sediment for extended periods of time (Summons et al, 1988). The viability of cholestanes as a biomarker is reflected in the clarity with which cholestanes were identified under GC-MS analysis in this study; 5β-cholestane and cholestane were the highest recorded match percentage values in NIST MS library matching, with 71% and 59.9% matches respectively. In addition, 5β-cholestane was identified in samples purified from the TLE of 5cm depth 5056m soils – indicating stability in a hostile environment. Notably, sterane compounds were not identified in 4296m samples, it is unclear why this is the case, and is possible that peaks representative of steranes were indistinguishable from background noise in the chromatograms due to the comparatively rich soils of the control site.

The identification of sulfone compounds in the samples, such as didecyl sulfone, which was identified in sample 1377 (3cm 4296m site), is notable, not least because the N1 fraction samples should not contain ketone-like compounds, as these are eluted after the N1 fraction, forming part of the N2 fraction. It is unclear why molecules of this family are detected in the N1 sample chromatograms, and is possibly an error in matching during NIST library

matching. Sulfone matching did not feature high percentage matching, typically; 34.8% being the highest percentage match, in the aforementioned sample 1377.

Comparison between the high altitude site of primary interest as a Martian analogue at 5056m, and the lower altitude control site at 4296m, illustrates the utility of such a site as a Mars analogue. Samples from the control site were less coherent, more prone to errors of clarity, and difficult to elucidate presence of viable biomarkers such as N-alkanes or steranes. It is clear that the GC-MS chromatograms of 5056m site samples present a greater proportion of reliable biomarkers among the N1 fraction samples, with minimal interference from other compounds present in richer soils such as those of the control site. In addition, the reduced variety of detected compounds in the higher altitude site indicates the lack of ability of other chemical species to persist in the hostile environment.

The focus on aliphatic hydrocarbon molecules for primary analysis in this study is in consideration of the nature of both long-chain hydrocarbon-like molecules, and of the nature of the surface of Mars. The hostility of the environmental conditions on the Martian surface have been discussed at length in this study, and it is precisely this hostile and destructive environment that the choice of relatively stable, non-polar, readily preserved biomarkers such as N-alkanes that this particular family of compounds is of interest.

It has been widely reported that long-chain aliphatic compounds may be preserved over significant timescales in the environment (Kenig et al, 2003), in lacustrine (Pu et al, 2017) and oceanic sediments (Simoneit et al, 2014), for example. It follows therefore, that examination of Martian samples for the presence of long chain hydrocarbon-like molecules may offer the greatest chance of success in the discovery of ancient preserved molecular fossils in future sampling or sample return missions to Mars.

5.0: Conclusions

This study focuses primarily on aliphatic hydrocarbons and similar nonpolar organic molecules, which are found in UV-resistant media such as plant leaf waxes. This resistance to UV-mediated damage makes such biomarkers an excellent candidate for the search for life on Mars due to the extreme UV environment experienced at the Martian surface. In addition,

large long-chain hydrocarbons such as Octacosane were reliably identified via GC-MS analysis from an initial complex soil sample total lipid extract.

This study has further examined the efficacy of the high Altiplano site as a terrestrial Martian analogue, in particular the ability of the soils to harbour identifiable, persistent, and informative lipid-derived biomarker molecules. The relative lack of soil richness expressed in GC-MS chromatograms, taken in comparison to the lower altitude control site, indicates an environment which is not teeming with evidence of life, mirroring the situation on the Martian surface as far as is currently known.

Unlike previous observations which utilized nucleic acids as the primary organic molecule of interest, here it is observed that lipid-derived biomarkers are affected by soil depth to a lesser extent; TLE yield exhibited lower correlation with soil depth than whole genomic content extract yield, indicating that lipid biomarkers are more stable in a Mars-like environment.

The use of GC-MS analysis in this study is directly relevant to upcoming robotic missions such as the planned 2020 ESA ExoMars Rover. ExoMars Rover is expressly designed with the search for potential biomarkers on Mars at the forefront. Instrumentation on the rover will include, critically, a two meter sampling drill, and a GC-MS package. This study was conceived partially with this in mind, and the use of a unique terrestrial Mars analogue region in the Altiplano has enabled the examination of the potential efficacy of the ExoMars experimentation on Mars.

This study indicates that viable biomarkers may be identified with GC-MS analysis; simple long-chain hydrocarbons, which are able to resist degradation by the intense UV radiation profile experienced at the Martian surface. Such lipid-derived biomarkers are persistent through extended timescales and represent an important candidate biomarker for the search for evidence of life on Mars. In addition, data presented herein indicates that such biomarker molecules are persistent both near the soil surface and at lower depths.

The primary limitations of the present study are applicable to the use of terrestrial Mars analogue regions as a whole: some conditions experienced on Mars are simply not possible to simulate in the field. An example of this is the extensive UVC radiation received on Mars due to the lack of dense atmosphere – UVC wavelengths are simply entirely attenuated by the atmosphere around the Earth, and so are not a contributing factor in the hostility of the environment presented by field sites such as the Altiplano. In addition, the composition of the

Martian atmosphere is different to that of the Earth, and so the chemistry of air-soil interactions are necessarily dissimilar. Finally, the composition of soils at the site featured in this study are lacking in damaging soil perchlorate salts, which have been reported in Martian soils and which certainly contribute to the inhospitable nature of the Martian surface (Wadsworth and Cockell, 2017).

When considering continuation of this study, the use of a controlled chamber such as the Mars Chamber at UKCA, Edinburgh would enable investigation of the effect of a Mars-like atmospheric composition and UV profile on the preservation of lipid biomarkers in soil. In addition, the use of synthetic analogue soils seeded with perchlorate salt would enable examination of the ability (or lack thereof) of lipid biomarkers such as those presented in this study to persist under Martian conditions, with a view to estimate the likelihood of success of the upcoming ExoMars rover in its search for identifiable biomarkers from Martian samples acquired both at the surface and in the shallow subsurface.

6.0: References

Alsalahi, M.A., Latif, M.T., Ali, M.M., Dominick, D., Khan, M.F., Mustaffa, N.I.H., Nadzir, M.S.M., Nasher, E., Zakaria, M.P. (2015). Sterols as biomarkers in the surface microlayer of the esturine areas. Marine Pollution Bulletin Vol 93. P278-283.

Brassell, S.C., Eglinton, G., Marlowe, I.T., Pflaumann, U., Sarnthein, M. (1986). Molecular stratigraphy: a new tool for climatic assessment. Nature Vol 320. P129-133.

Chowdhury, N., Ghosh, A., Bhattacharjee, I., Laskar, S., Chandra, G. (2010). Determination of the n-alkane profile of epicuticular wax extracted from mature leaves of *Cestrum nocturnum* (Solanaceae: Solanales). Natural Product Research Vol 24. P1313-1317.

Cockell, C.S., Catling, D.C., Davis, W.L., Snook, K., Kepner, R.L., Lee, P., McKay, C.P. (2000). The Ultraviolet Environment of Mars: Biologicial Implications Past, Present, and Future. Icarus Vol 146. P343-359.

Conte, M.H., and Weber, J.C. (2002). Long-range atmospheric transport of terrestrial biomarkers to the western North Atlantic. Global Biogeochemical Cycles Vol 16.

Dartnell, L.R., Desorgher, L., Ward, J.M., Coates, A.J. (2007). Modelling the surface and subsurface Martian radiation environment: Implications for astrobiology. Geophysical Research Letters Vol 34.

Gagosian, R.B., Peltzer, E.T., Zafiriou, O.C. (1981). Atmospheric transport of continentally derived lipids to the tropical North Pacific. Nature Vol 291. P312-314.

Holmes, M.G., and Keller, D.R. (2002). Effects of pubescence and waxes on the reflectance of leaves in the ultraviolet and photosynthetic wavebands: a comparison of a range of species. Plant, Cell and Environment Vol 25. P85-93.

Kenig, F., Simons, D-J. H., Crich, D., Cowen, J.P., Ventura, G.T., Rehbein-Khalily, T., Brown, T.C., Anderson, K.B. (2003). Branched aliphatic alkanes with quaternary substituted carbon atoms in modern and ancient geologic samples. PNAS Vol 100. P12554-12558.

Kminek, G., and Bada, J.L. (2006). The effect of ionizing radiation on the preservation of amino acids on Mars. Earth and Planetary Science Letters Vol 245. P1-5.

Kounaves, S.P., Carrier, B.L., O'Niel, G.D., Stroble, S.T., Claire, M.W. (2014). Evidence of Martian perchlorate, chlorate, and nitrate in Mars meteorite EETA79001: Implications for oxidants and organics. Icarus Vol 229. P206-213.

McEwen, A.S., Ojha, L., Dundas, C.M., Mattson, S.S., Byrne, S., Wray, J.J., Cull, S.C., Murchie, S.L., Thomas, N., Gulick, V.C. (2011). Seasonal Flows on Warm Martian Slopes. Science Vol 333. 740-743.

Morozova, D., Möhlmann, D., Wagner, D. (2007). Survival of Methanogenic Archaea from Siberian Permafrost under Simulated Martian Thermal Conditions. Origins of Life and Evolution of Biospheres Vol 37. P189-200.

Nelson, D.B., Knohl, A., Sachse, D., Schefuß, E., Kahmen, A. (2017). Sources and abundances of leaf waxes in aerosols in central Europe. Geochimica et Cosmochimica Acta Vol 198. P299-314.

Parnell, J., Cullen, D., Sims, M.R., Bowden, S., Cockell, C.S., Court, R., Ehrenfreund, P., Gaubert, F., Grant, W., Parro, V., Rohmer, M., Sephton, M., Stan-Lotter, H., Steele, A., Toporski, J., Vago, J. (2007). Searching for Life on Mars: Selection of Molecular Targets for ESA's Aurora ExoMars Mission. Astrobiology Vol 7. P578-604.

Pu, Y., Wang, C., Meyers, P.A. (2017). Origins of biomarker aliphatic hydrocarbons in sediments of alpine Lake Ximencuo, China. Palaeogeography, Palaeoclimatology, Palaeoecology Vol 475. P106-114.

Rothschild, L.J. (1990). Earth Analogs for Martian Life. Microbes in Evaporites, a New Model System for Life on Mars. Icarus Vol 88. P246-260.

Ruff, S.W., and Farmer, J.D. (2016). Silica deposits on Mars with features resembling hot spring biosignatures at El Tatio in Chile. Nature Communications Vol 7.

Schaeffer, P., Ocampo, R., Callot, H.J., Albrecht, P. (1993). Extraction of bound porphyrins from sulphur-rich sediments and their use for reconstruction of palaeoenvironments. Nature Vol 364. P133-136

Schouten, S., Hopmans, E.C., Schefuß, E., Sinninghe Damsté, J.S. (2002). Distributional variations in marine crenarchaeotal membrane lipids: a new tool for reconstructing ancient sea water temperatures? Earth and Planetary Science Letters Vol 204. P265-274.

Schubert, B.A., Lowenstein, T.K., Timofeeff, M.N. (2009). Microscopic Identification of Prokaryotes in Modern and Ancient Halite, Saline Valley and Death Valley, California. Astrobiology Vol 9. 467-482.

Simoneit, B.R.T., Cardoso, J.N., Robinson, N. (1990). An assessment of the origin and composition of higher molecular weight organic matter in aerosols over Amazonia. Chemosphere Vol 21. P1285-1301.

Simoneit, B.R.T., Oros, D.R., Medeiros, P.M. (2014). Organic matter provenance and palaeoenvironment in the Cretaceous on the Manaihiki Plateau, South Pacific. Palaeogeography, Palaeoclimatology, Palaeoecology Vol 409. P48-56.

Stromberg, J.M., Applin, D.M., Cloutis, E.A., Rice, M., Berard, G., Mann, P. (2014). The persistence of a chlorophyll spectral biosignatures from Martian evaporite and spring analogues under Mars-like conditions. International Journal of Astrobiology Vol 13. P203-223.

Summons, R.E., Brassell, S.C., Eglinton, G., Evans, E., Horodyski, R.J., Robinson, N., Ward, D.W. (1988). Distinctive hydrocarbon biomarkers from fossiliferous sediment of the late

Proterozoic Walcott Member, Chaur Group, Grand Canyon, Arizona. Geochimica et Cosmochimica Acta Vol 52. P2625-2637.

Suo, Z., Avci, R., Schweitzer, M.H., Deliorman, M. (2007). Porphyrin as an Ideal Biomarker in the Search for Extraterrestrial Life. Astrobiology Vol 7. P605-615.

Taylor, K.W.R., Huber, M., Hollis, C.J., Hernandez-Sanchez, M.T., Pancost, R.D. (2013). Re-evaluating modern and Palaeogene GDGT distributions: Implications for SST reconstructions. Global and Planetary Change Vol 108. P158-174.

Theroux, S., Toney, J., Amaral-Zettler, L., Huang, Y. (2013). Production and temperature sensitivity of long chain alkenones in the cultured haptophyte *Pseudoisochrysis paradoxa*. Organic Geochemistry Vol 62. P68-73.

Tyagi, P., Edwards, D.R., Coyne, M.S. (2007). Use of selected chemical markers in combination with a multiple regression model to assess the contribution of domesticated animal sources of fecal pollution in the environment. Chemosphere Vol 69. P1617-1624.

Ulrich, R., Pilgrim, R., Chevrier, V., Roe, L., Kral, T. (2008). Temperature fields at Mars landing sites: Implications for subsurface. Lunar and Planetary Science Vol 39.

Valitova, J.N., Sulkarnayeva, A.G., Minibayeva, F.V. (2016). Plant Sterols: Diversity, Biosynthesis, and Physiological Functions. Biochemistry (Moscow) Vol 81. P819-834.

Wadsworth, J., and Cockell, C.S. (2017). Perchlorates on Mars enhance the bacteriocidal effects of UV light. Nature Scientific Reports Vol 7.

Weijers, J.W.H., Schouten, S., Hopmans, E.C., Geenavasen, J.A.J., David, O.R.P., Coleman, J.M., Pancost, R.D., Sinninghe Damsté, J.S (2006a). Membrane lipids of mesophilic anaerobic bacteria thriving in peats have typical archeal traits. Environmental Microbiology Vol 8. P648-657.

Wierzchos, J., DiRuggiero, J., Vitek, P., Artieda, O., Souza-Egipsy, V., Škaloud, P., Tisza, M., Davila, A.F., Vilchez, C., Garbayo, I., Ascaso, C. (2015). Adaptation strategies of Endolithic chlorophototrophs to survive the hyperarid and extreme solar radiation environment of the Atacama Desert. Frontiers in Microbiology Vol 6.

Website references

NIST MS Library matching reference: https://www.nist.gov/about-nist

Final Conclusions

Aims of Project

The initial aims of this study were to examine the effect of soil coverage on the presence of detectable organic material in a hostile environment which harbours similarities to the surface of Mars. For this purpose, building on the previous work by Dr Vernon Phoenix, the utilisation of a novel Martian analogue region was incorporated into the experimental design in order to offer a unique environment with features specifically appropriate for testing the effect of shelter from a hostile soil surface. The solar heating effect experienced by the soils is not experienced to such an extent in other Martian analogues, and presented an invaluable opportunity to examine the level of protection that shallow soil coverage may provide on the Martian surface. The selection of DNA and lipid content as the principal molecules of interest was based on the potential stability and reliable identification methods of these signatures of life.

Two field expeditions were undertaken during the course of this study, with the initial expedition serving as a pilot study, taking limited samples and environmental measurements, with the intention of determining best procedure for a more significant expedition upon return in the next expedition. This initial expedition allowed identification of high quality field sites, and the most efficient techniques for working efficiently, accurately, and safely in a challenging environment. Upon return for a more significant field campaign, rapid and efficient sampling, environmental data recording, and sample handling were able to be executed by the combined research group.

Laboratory experimental proceedings were separated according to the primary analyte; lipids or nucleic acids. Nucleic acid extraction enabled the completion of an ambitious sequencing project designed to investigate the change in bacterial community diversity as soil depth below the surface increases. Not only did this sequencing project illuminate the bacterial diversity of an otherwise under-reported environment, it provided a new viewpoint for the change in environmental hostility with soil coverage via examining the change in prevalence of species such as extremophiles as soil conditions change.

Lipid analysis was tailored to investigate chemical species likely to be preserved in an arid, high-UV environment, such as long-chain aliphatic hydrocarbons of the kind commonly present in plant cuticle waxes. This enabled the study to directly investigate the preservation of biomarkers which might feasibly persist under Martian conditions.

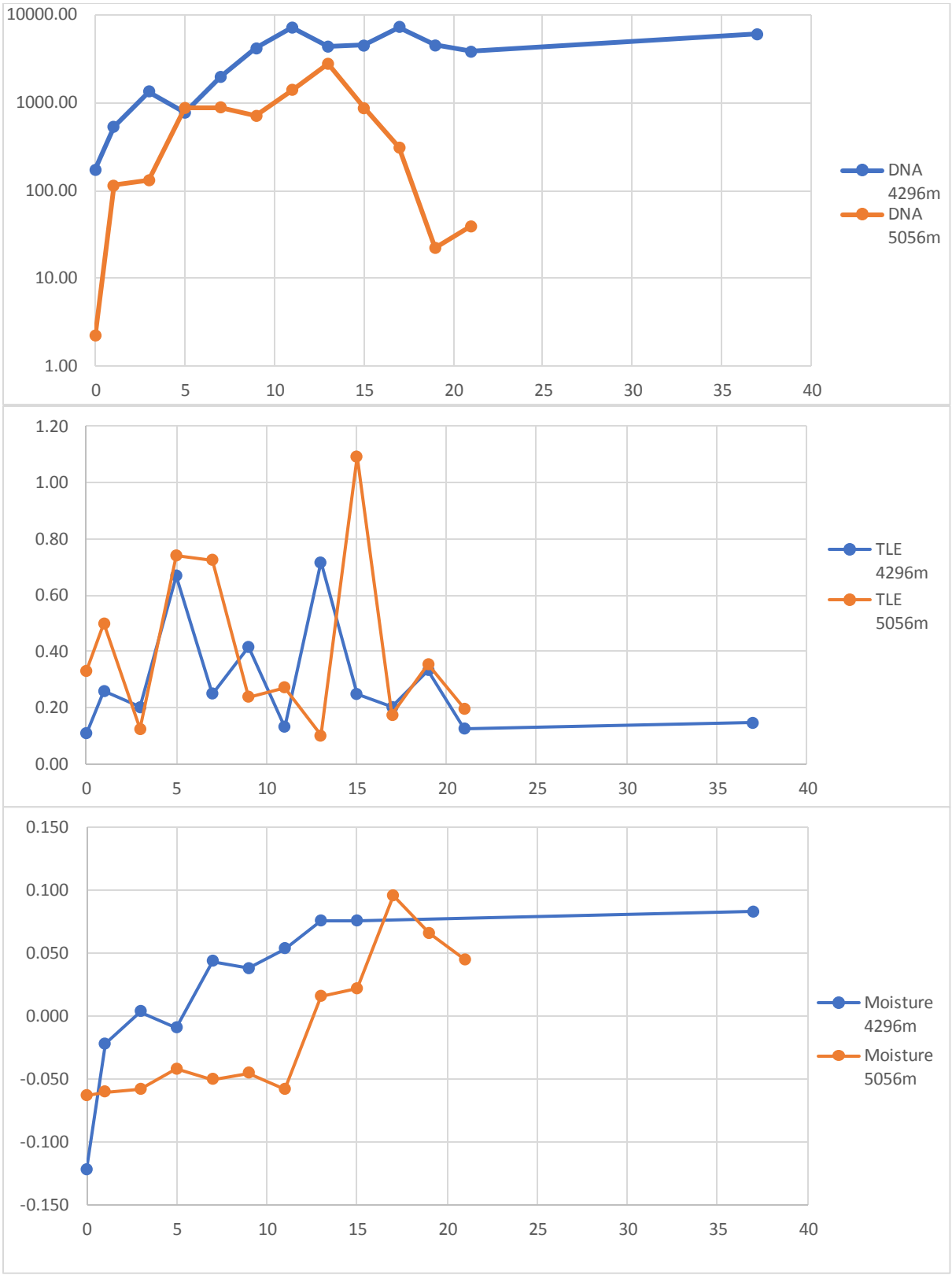
Data Interpretation

This study determined that the nature of soil environmental conditions in this Mars analogue site change rapidly below the surface, to the extent that at higher altitude sites, subsurface ice was observed mere inches below the exposed surface, in a manner extremely reminiscent of the celebrated NASA Phoenix Lander photograph of Martian subsurface ice excavated during its landing. DNA sequencing in this study exhibited the rapid change in, and reduction of, overall bacterial community diversity in a Mars-like environment as conditions became more hostile near the soil surface. Meanwhile, the novel use of the Tax4Fun script in a non-clinical setting has allowed the examination of predicted expression of stress response genes according to soil depth, where it was determined that while the soils do appear to become less hostile with depth, stress response mechanisms are still active in the bacterial community at lower depths, indicating that the soils remain somewhat challenging for life even away from the surface.

Analysis of lipid content of the analogue site soils indicated that classes such as long-chain aliphatic hydrocarbons are affected by soil depth to a lesser extent than nucleic acids, due in part to the hydrophobic and UV-resistant nature of these compounds, which renders molecules such as N-alkanes more able to persist in the environment. The common use of lipid biomarkers in palaeoclimate studies is a clear indication of the potential of such molecules to be preserved as organic molecular fossils.

Tying together the data as a whole

The data produced over the course of this study exhibits a range of synergistic findings, where different branches of study corroborate each other, and data fits logically with data. The key example of this lies in the changes in both extract yield and data analysis output of extracted DNA and lipids which line up with environmental data recorded at the field sites, particularly the increase in soil moisture content with soil depth. To facilitate comparison of these data, the relevant data are re-plotted in Fig[1] below.



Fig[1]: Data representing extract yields of DNA (top) and lipid (middle) content from the 4296m and 5056m sites, set in comparison to changing soil moisture content with depth (bottom). These plots are presented together for ease of observation of shared

trends in the data. NOTE: negative moisture values in the bottom figure are the result of a calibration error previously addressed. In this case the emphasis is on the trend with depth rather than y-values.

The similarities in collected data for DNA and lipid extracts and change in soil moisture recorded in Fig[1] highlight the notable change in organic extract yield as soil moisture increases generally, and particularly around the 11cm soil depth, wherein soil moisture significantly increases, which in higher altitude sites often resulted in permafrost-like soil conditions. It is possible that the moister soils of these depths represent something of a micro environment in the soils which is more congenial to life, or at least the preservation of recoverable organic markers. In addition to the increased yields from moister soils, it is notable that the extract mass from the solar-heated, arid, acidic surface soils is considerably lower than other soil depths, which was the case for both DNA and lipid content, although the effect was more pronounced in the DNA content yield.

The apparent increase in the diversity of the bacterial community – also examined according to Shannon index – as soil depth increases additionally aligns with the increase in soil moisture, adding further credence to the notion of possibility of a more habitable micro-environment below the soil surface. Further, the noted change in possible community structure which was recorded in the differing soil depth in the co-occurrence plots supports the hypothesis that the deeper soils are more congenial for life than the heated, arid surface environment of this Martian analogue site.

The presence of subsurface ice or frozen soil at higher altitudes not far below the surface is of interest for Mars, given that from that which was able to be excavated from the frozen soils (including 5056m deeper soils), organics were able to be recovered and analysed; perhaps similar ice deposits on Mars may yet yield interesting data.

There are elements of the data which appear to disagree, an example of this may be observed in the overall trend of total lipid extract mass as soil depth increases; while there is a rough overall trend of yield increasing with depth, the pronounced reduction in yield in the shallow soils and even the surface layer itself was muted in comparison to that which was observed with the DNA extract yields. The potential explanations for this are that either the lipid content is less sensitive to breakdown at the surface than DNA, or is more readily deposited

from elsewhere – the latter suggestion seeming to be backed up by the GC-FID and GC-MS analyses which indicated the presence of compounds commonly isolated from plant leaf cuticle waxes. Future studies may answer this question by accurately recording the extent and rate of wind deposition of soil in the area. Additionally, repeating extractions undertaken in this study with collected windborne soils would expand upon the observation of the effect of wind deposition on surface soils.

A significant disagreement in the data which requires mention lies in the recordings of solar radiation which were taken at each site. Of particular note is that the UV recordings at the higher altitude 5056m site were lower than those taken at the lower altitude 4296m site. This appears counterintuitive, as increased altitude and thus lower atmospheric density at 5056m should consequently expose the surface to more intense solar flux. A likely cause of this apparent discrepancy is daily weather variations, as recordings were taken on different days, and therefore it is advised that future studies record solar flux intensities at field sites over a number of days or weeks to establish a more clear average impression of conditions at the sites. In addition, the sampling for this study was completed in the austral autumn (late April and early May), and therefore UV intensities are likely to not represent intensities experienced during the height of the austral summer.

Implications

This study presents a new and exciting option for studies wishing to utilise a Mars analogue environment. The high altitude sites around Cerro Sairecabur have proven a challenging yet valuable environment for testing experimental procedures designed for Mars missions. In addition, the study has demonstrated that the change in habitability of environment below the soil surface is rapid, occurring less than one meter into the soil in this case. It is probable that on Mars the extreme environment – far more hostile than earth analogues are able to simulate – will not be buffered so rapidly, although if the evidence for subsurface ice melting is taken into consideration, similarities may yet be confirmed.

Future Work

The principal limitations of this study are related to the use of terrestrial Mars analogues in general, that is that there are factors which are not possible to reproduce in a field location on

Earth, simply due to the nature of the planet itself: Earth has a dense atmosphere with a different composition to the Martian atmosphere, for example. In addition, Earth features a strong magnetosphere. These two factors alone contribute to terrestrial Martian analogues suffering from a lack of representative UVC wavelength radiation profile, and vastly reduced cosmic radiation activity – both of which are critically important factors behind the apparent inhospitable nature of Mars.

Further limitations of this study include the limited soil sampling depth at each field site. Future studies making use of deeper excavation using sediment corers or similar equipment would allow for greater range of soil depth investigation, allowing sampling range to mirror that of the planned ExoMars Rover. In addition, use of sample cores allows for more accurate excavation according to depth that could be achieved in this study – allowing 1cm or smaller increments of soil depth, and therefore greater resolution in the change in environmental conditions as depth below the hostile surface increases.

To continue and build upon the successes of this study, a wider range of sampling of both soil and environmental conditions monitoring in the field would be required. While short term environmental monitoring was conducted in the field, technical difficulties on both field expeditions undertaken resulted in a difficulty in monitoring daily changes in UV and visible radiation levels at the field sites, preventing the generation of a comprehensive and quantitative observation of the extreme solar flux which is a pivotal feature of the Altiplano region. In addition, repeat experiments of high-resolution examination of soil condition change with depth in other analogue regions such as Atacama and McMurdo would provide interesting comparative data.

In addition to expanded work in the field, a future continuation of this study would be well-served by the use of simulations of Martian conditions such as UVC exposure and oxidative soil chemistry, which are best reproduced in the lab. Investigation of the preservation of biomarkers under these conditions will provide vital insight into the likelihood of such biomarkers being preserved and detectable on Mars by future missions.

Finally, an expansion of analytes of interest is recommended in future investigations. For example, predictions from the Tax4Fun script discussed in Chapter 1 can be tested for accuracy by producing a cDNA library, examining the transcriptome of samples. This will however produce significant data volumes if sampling is taking from whole genomic content extract, and thus it may be advisable to take samples from bacteria cultured from soil samples

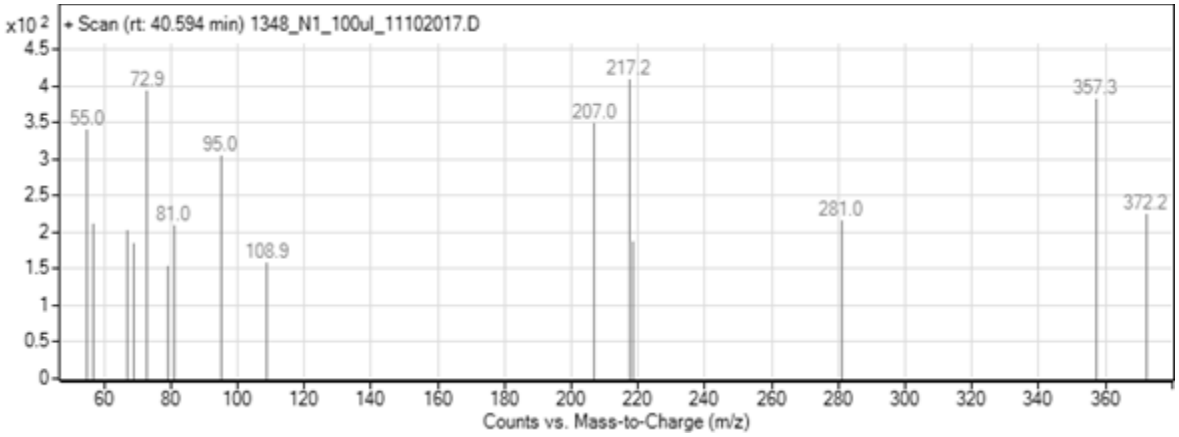
rather than entire soil community snapshots, as was the case in this study. Further expansions to analyses to this study include the neutral fraction lipid biomarkers which were not covered in this study, as well as the acid fraction samples.

Final Conclusions

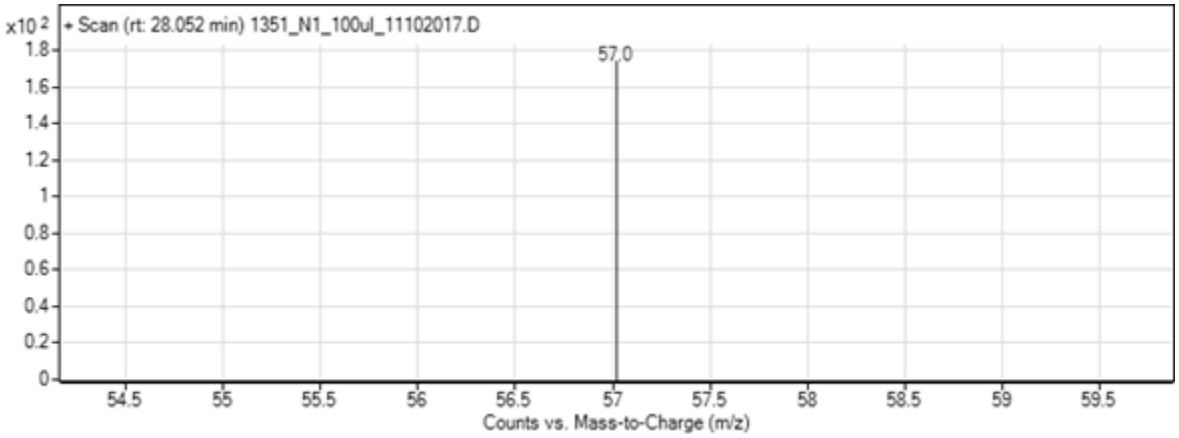
The aims of this study were to investigate the effect of soil coverage in a Mars analogue region. This environmental variable is of particular interest when considering the ongoing exploration of the surface of the red planet – Mars appears sterile on the surface itself, and therefore upcoming missions are designed to examine conditions under the surface, to assess the possibility of preservation of extinct or extant life being found away from the inhospitable surface conditions. With this in mind, the present study has assessed the effect of the modest change in environment as soil depth increases in a terrestrial Mars analogue; while we are starting from an environment in which life is present, it is nonetheless in several important manners a useful and exciting environment for astrobiology. This study has determined that change in environmental conditions is particularly relevant in the presence of water, with the ability of the soil to hold moisture correlating with increase in overall recoverable biomass and diversity of life in this microenvironment. In addition, we have worked with the presence of water in even the relatively shallow subsurface below an extremely arid, solar-heated surface, similar to that of Mars. The current NASA doctrine of "follow the water" is well-served by such findings, and missions to regions of Mars featuring subsurface brine ice or potential subsurface liquid brines may yet prove fruitful endeavours.

Appendix 1: GC-MS spectra of identified compounds, Chapter 3, Table[2]

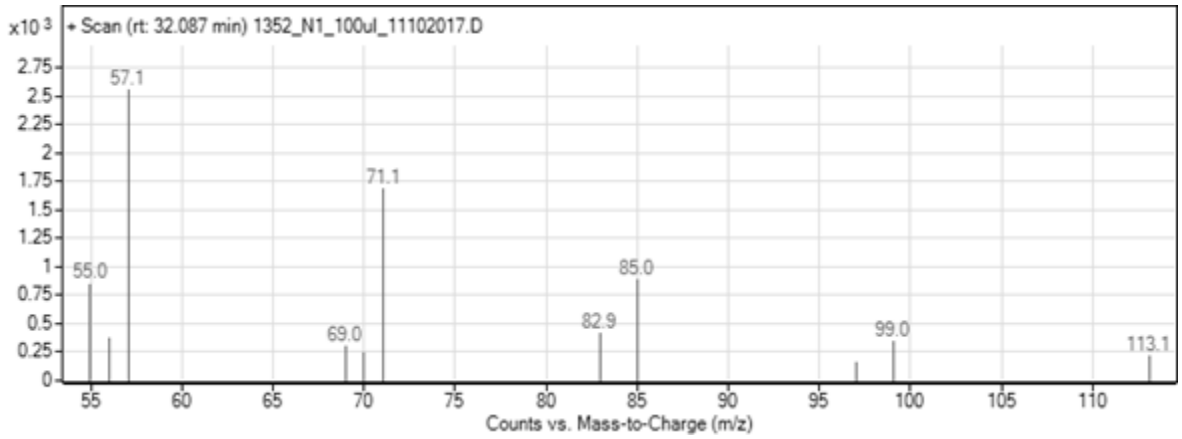
5056m Site Samples



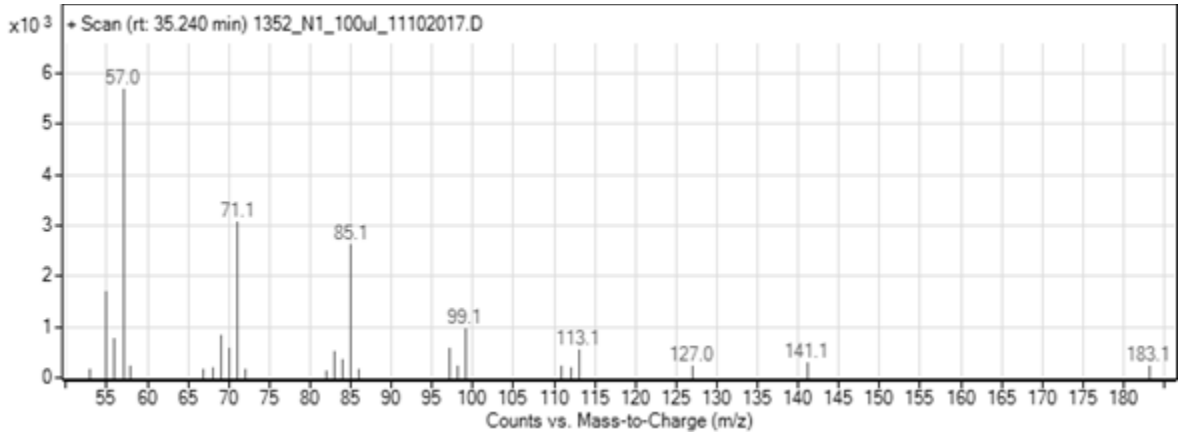
1348: 5β-cholestane mass spectrum



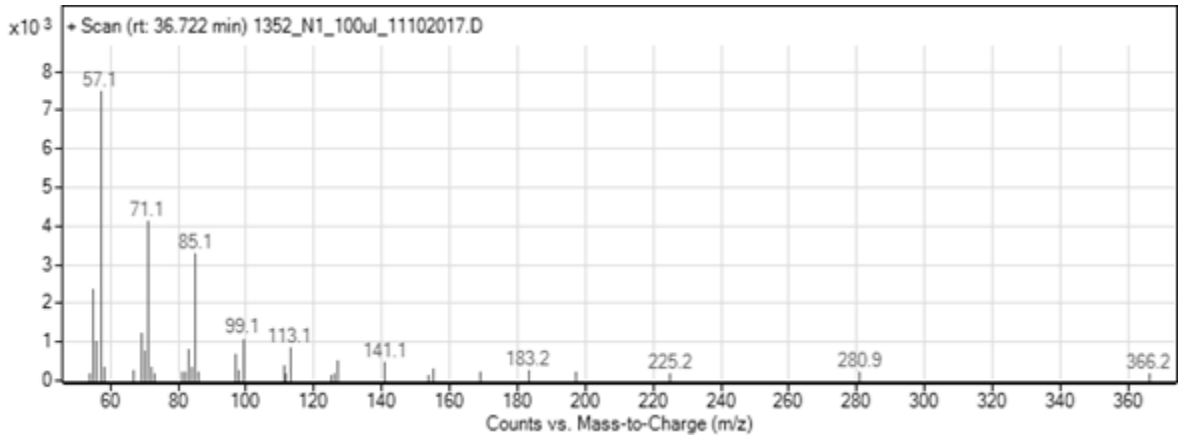
1351: 1,3-methyl-4-penten-2-ol mass spectrum



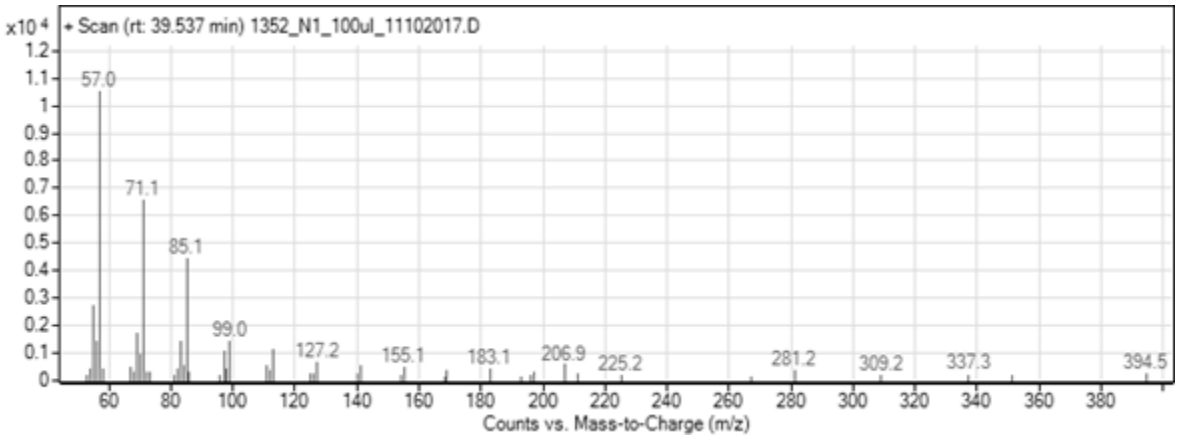
1352: Decyl-2-ethylhexyl sulphite mass spectrum



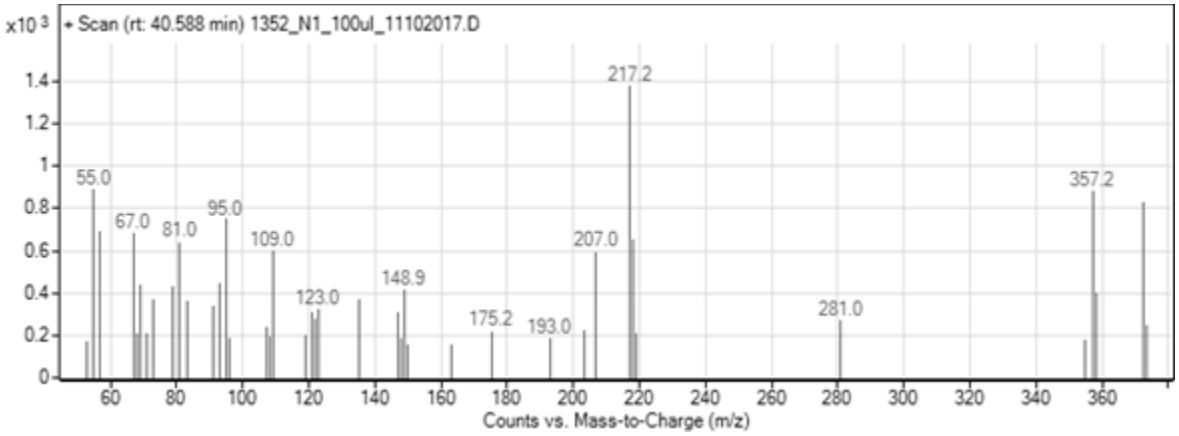
1352: Nonadecane mass spectrum



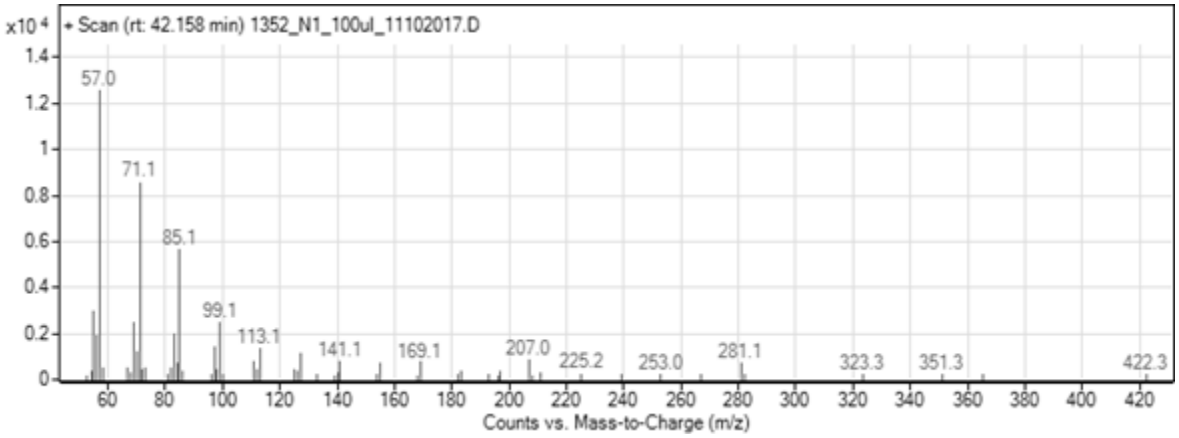
1352: Octacosane mass spectrum



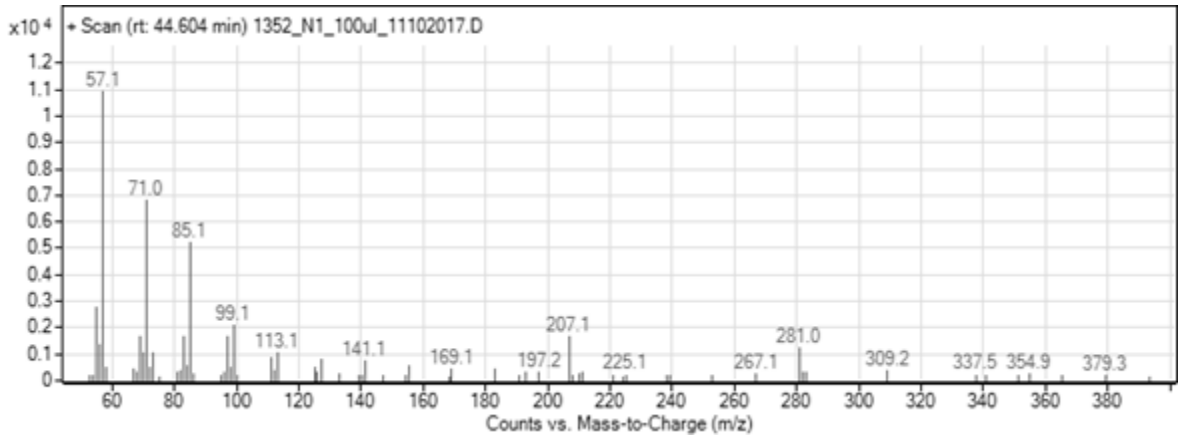
1352: Octacosane mass spectrum



1352: Cholestane mass spectrum

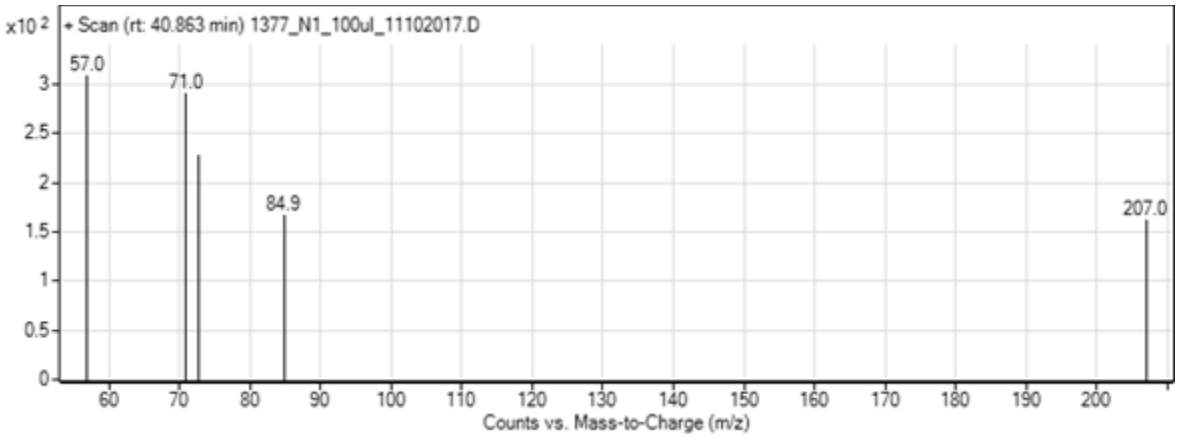


1352: Octacosane mass spectrum

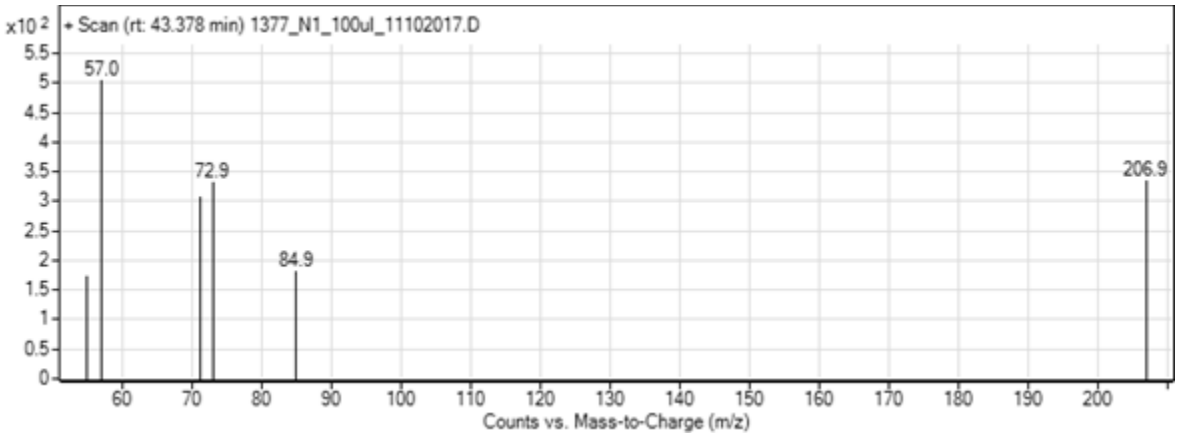


1352: Tetratetracontane mass spectrum

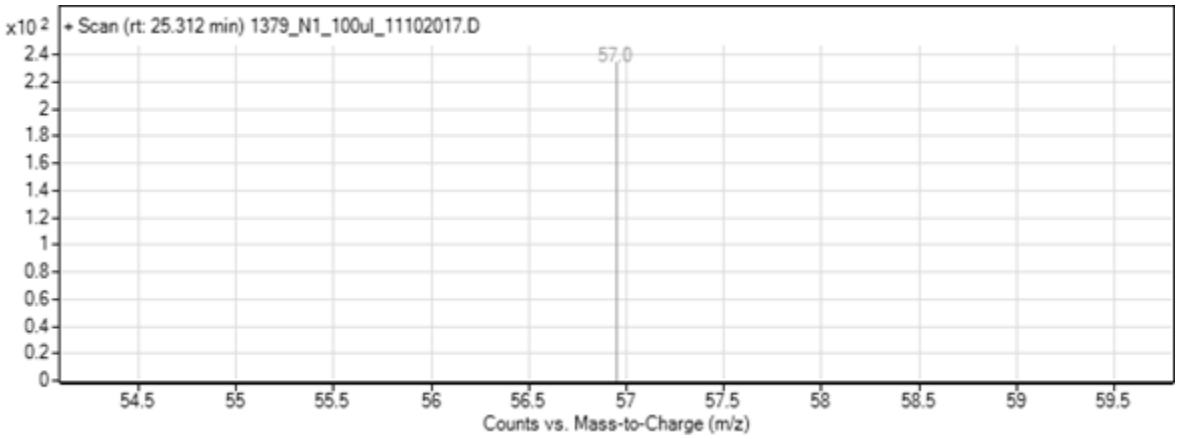
4296m Samples



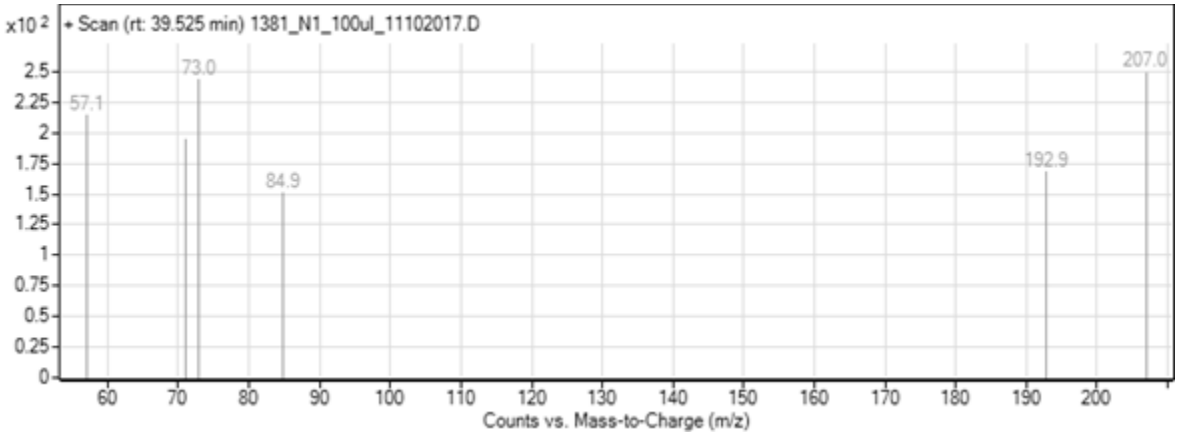
1377: Methylpentenol mass spectrum



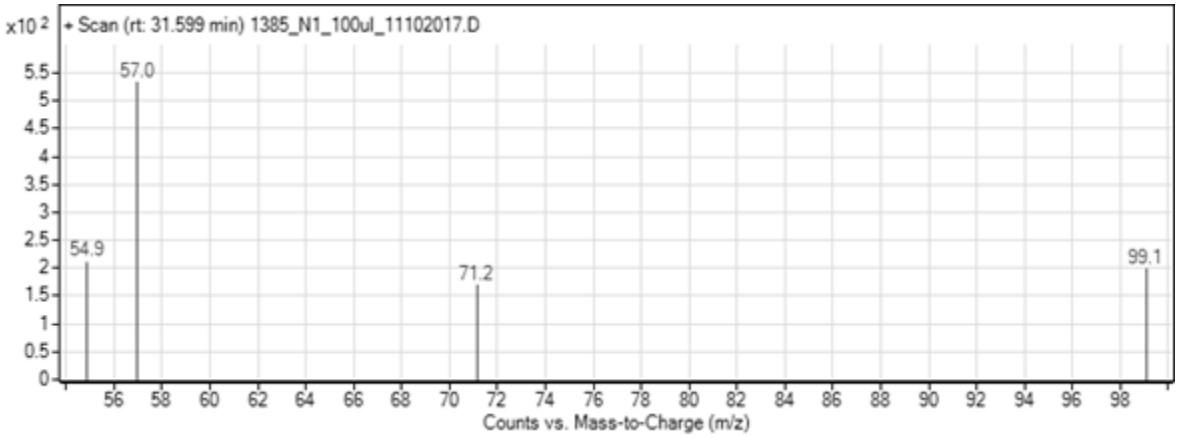
1377: Didecyl sulfone mass spectrum



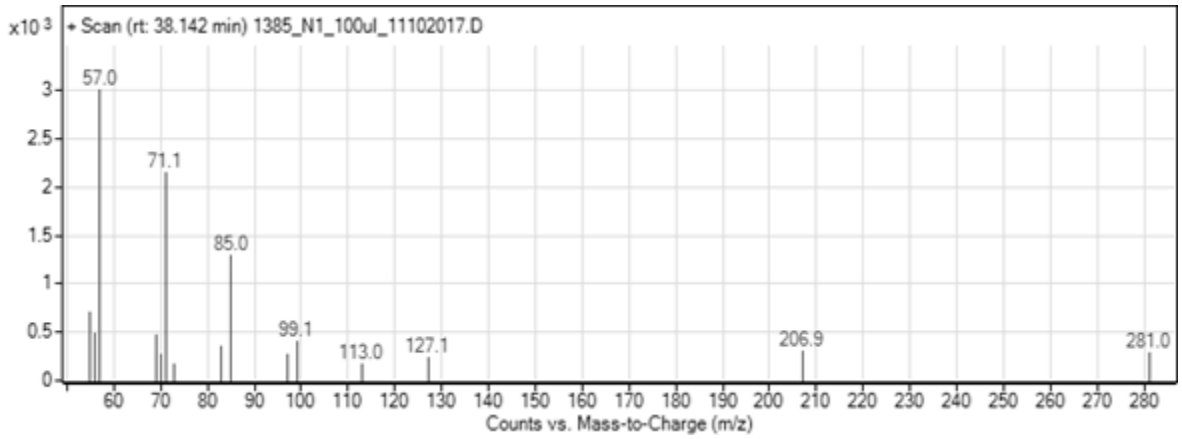
1379: 1,3-methyl-4-penten-2-ol mass spectrum



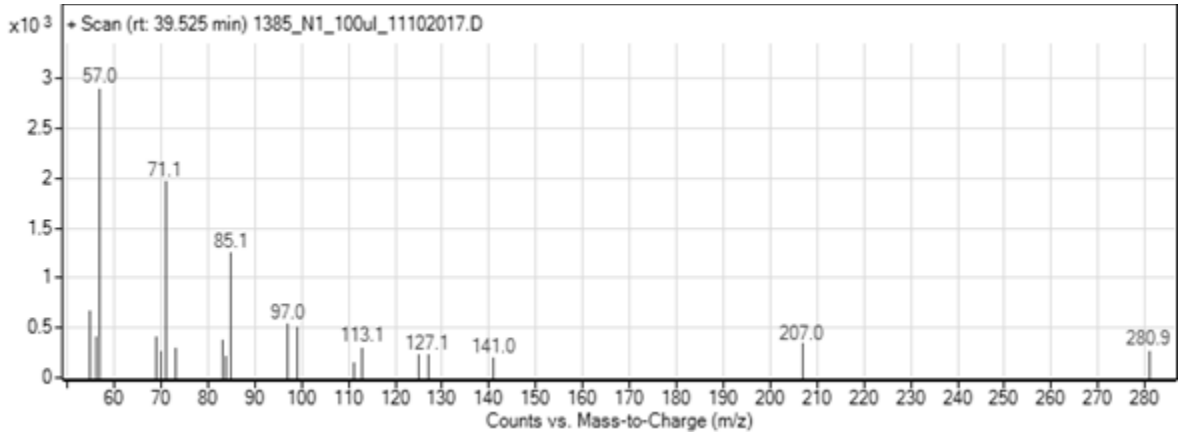
1381: 2' - (trimethylsiloxyl) – propriophenone mass spectrum



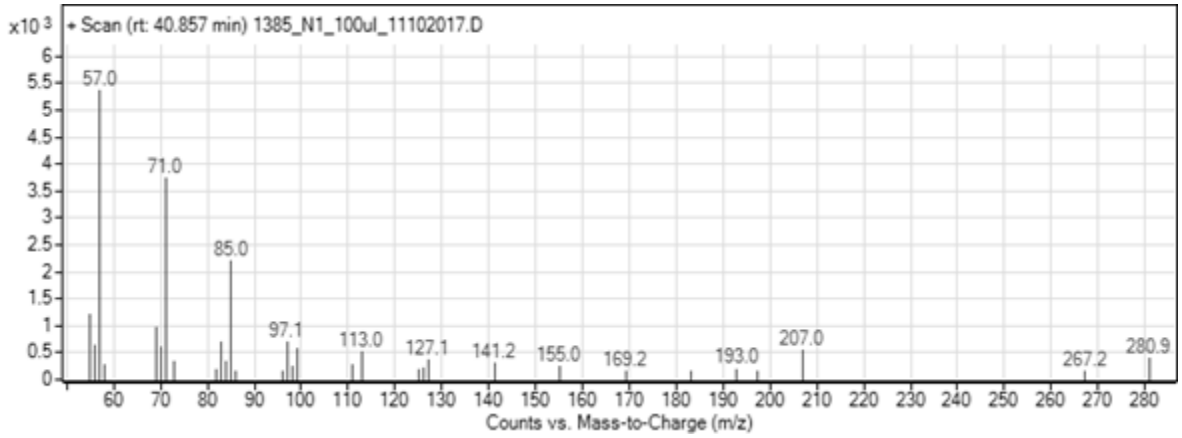
1385: 3,6 – heptanedione mass spectrum



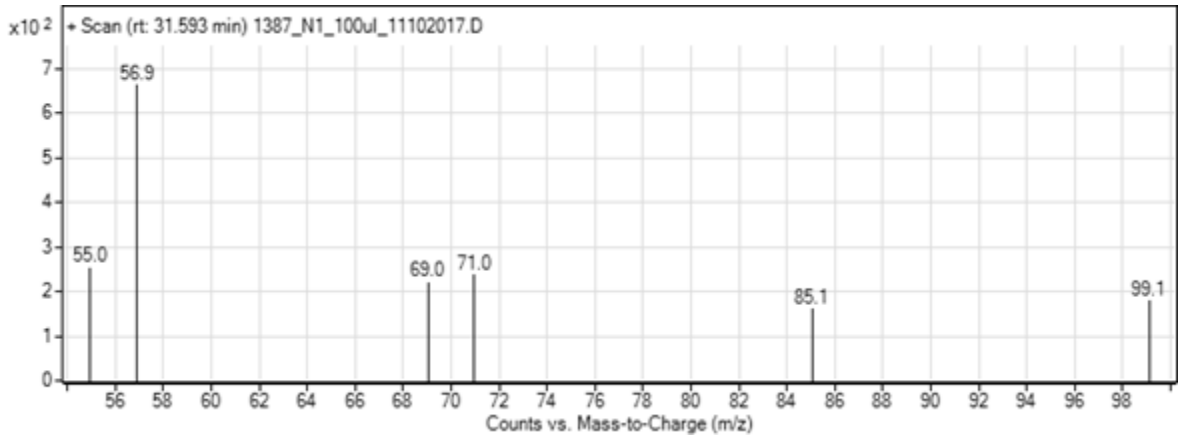
1385: 2 - ethylhexylisohexyl ester mass spectrum



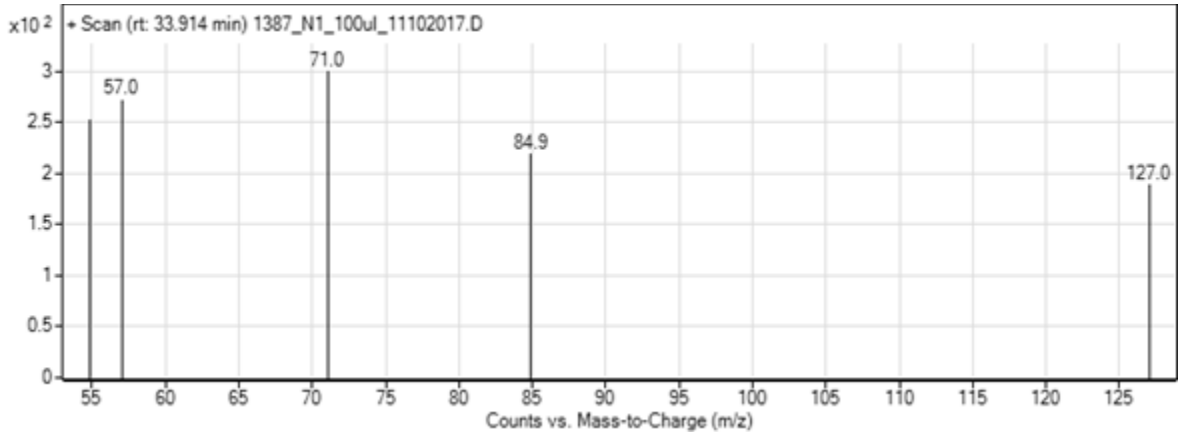
1385: dodecyl -2 - propyl ester mass spectrum



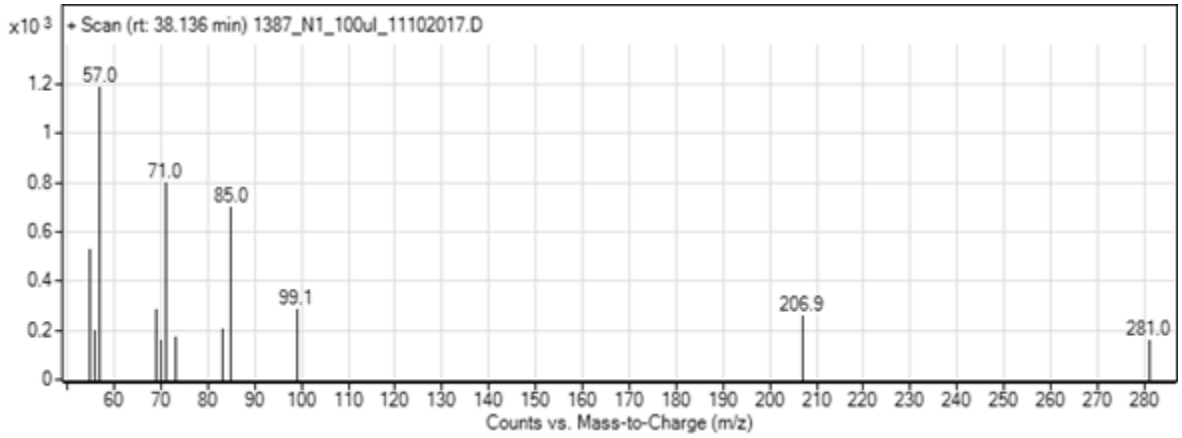
1385: Octacosane mass spectrum



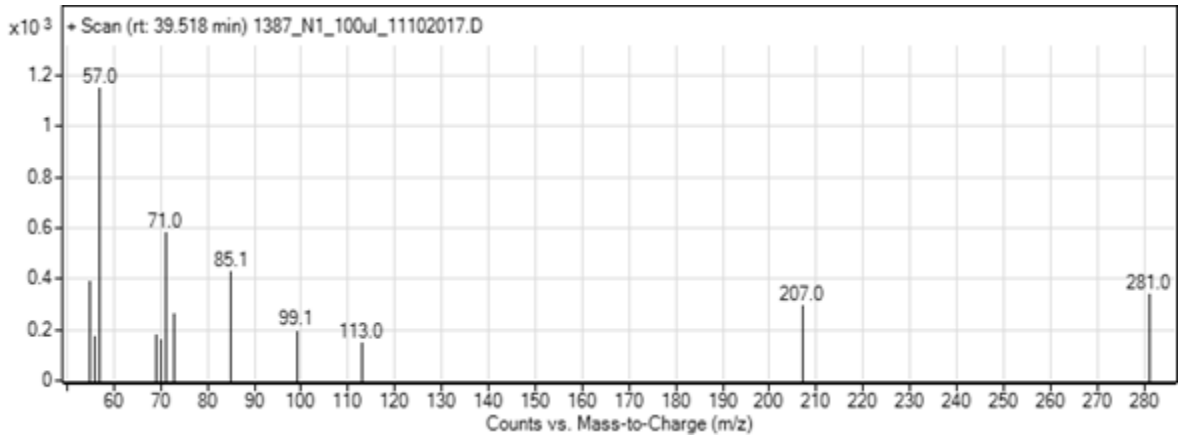
1387: 2,7 – dimethyloctane mass spectrum



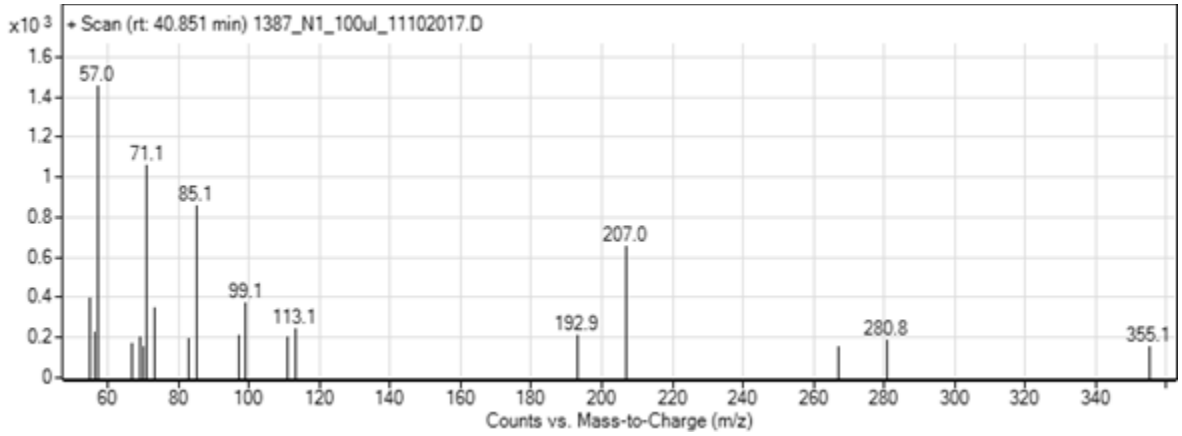
1387: Iodononane mass spectrum



1387: Decyl isopropyl sulphite mass spectrum



1387: 2,3,5,8 – tetramethyldecane mass spectrum



1387: Pentyl undecyl sulphite mass spectrum