A Study of Iron-Oxidising Bacteria, Their Habitats and Their Associated Biogenic Iron Ochre, Sampled From the Greater Glasgow Area

Adam Drummond

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

School of Chemistry

College of Science and Engineering

University of Glasgow

February 2022

Abstract

Leptothrix ochracea is a species of iron-oxidising bacterium that is frequently found in ferrous iron rich waters throughout the world where it contributes to the production of biogenic iron ochre. When viewed microscopically, biogenic iron ochre produced by *L.ochracea* is found to comprise hollow microtubular filaments. The ability to produce iron oxide containing microtubes under ambient conditions has led to *L.ochracea* being an attractive bacterium to study however there is currently no isolated axenic culture of *L.ochracea* nor is a thorough understanding of the exopolymeric secretions which act as a scaffold for microtubular formation. This research aims to address these areas by studying biogenic iron ochre mats collected from several sample sites surrounding the greater Glasgow area.

Chapter 3 provides a provides a characterisation of the sampled biogenic iron ochre, primarily by SEM-EDX and XRD, to confirm the presence of filamentous material and investigate its composition and phase. Further to this there is also the development of a protocol to extract and characterise organic material associated with the samples along with a study of the interaction of thiol containing reducing agents with the material.

Chapter 4 provides a characterisation of the bacterial communities present at three of the sample sites via the high-throughput Illumina sequencing of the V3 – V4 region of the 16S rRNA gene. This is an important area of the study as it is essential to understand which other bacteria are found within the biogenic iron ochre mats as this will provide insight to the biogeochemistry occurring. By understanding the other bacteria present and biogeochemistry it may be possible to develop an artificial environment in which to grow and isolate *L.ochracea*.

Chapter 5 provides the development of protocols to isolate various types of bacteria from biogenic iron ochre mats. This chapter begins with the utilisation of solid agar media and solid gellan gum media supplemented with ferrous iron salts to isolate single colonies of bacteria. This chapter then investigates other isolation techniques including liquid enrichment growths and gradient tubes and culminates with the development of a protocol to isolate single filaments of bacteria via a micromanipulator. Isolated bacteria had their genomic DNA extracted which was then amplified by PCR and sequenced via high-throughput Illumina sequencing.

Chapter 6 provides a thorough characterisation of the sample sites used throughout this study. Characterisation includes multiple photographs, measurements and descriptions, the physicochemical conditions present, the inorganic species present and the

concentration of dissolved inorganic and organic carbon present. Orbitrap mass spectrometry is then used to characterise the organic material present. The recent history and underlying geology of selected sample sites is also investigated to assess whether anthropogenic activity or natural geology have a greater effect on the chemistry occurring within the sample sites. As with Chapter 4, by fully understanding the sample sites and chemistry occurring within them it may be possible to develop artificial media and environments that could be used to grow and isolate *L.ochracea*.

Table of Contents

List of Figures

List of Tables

List of Equations

Acknowledgements

Many people have helped me, both directly and indirectly, complete this work and I would like to use this space to thank them.

I would like to begin with thanking my dad, Alfie, for his support throughout everything I have ever done with my life. Without him I would not be anywhere near the person I am today, and I miss him daily. Along with him, I must also like to thank my mum for supporting me throughout my many, many years of education.

Academically, I have to begin with thanking Justin Hargreaves and Adrian Lapthorn for their guidance throughout my undergraduate and postgraduate education. They have gone above and beyond what I could ever have expected from PhD supervisors, and I am extremely grateful for the countless coffees and spur-of-the-moment excursions to muddy bogs in all weathers that we have went on. I would also like to thank all members of the Catalysis research group whom I have worked with over the years. You have made my PhD experience enjoyable daily.

Within the Joseph Black building I must firstly thank Jim Gallagher for his assistance and chats when using the SEM and I will one day eventually buy him that pint in Neeson's. Margaret Mullin has also helped greatly with SEM work and Chris Syme has contributed to my CLSM understanding. I must also thank Clare Wilson for her XRD expertise and Brian Smith for his assistance with all things NMR related.

I have made several friends throughout my time at Glasgow University; however, few have been around as long as Robert Pow, Cameron Gilroy and Marsali Paton. These friends, along with PhD addition Matt Martin, have given me some amazing memories throughout my studies and I would not nearly have enjoyed my university life as much without them. From walking the West Highland Way as a postgraduate, special shoutout to Chris Newton here, to spending umpteen Tuesday nights in Sub Club as an undergraduate, I am forever grateful for each and all these experiences.

I must also thank Claire, who was with me and supported me through the entire PhD journey.

Author's Declaration

I declare that this thesis is the result of my own work, except where reference is made to the contribution of others. This work has not been submitted for any other degree at the University of Glasgow or any other institution.

Abbreviations

ASV	Amplicon Sequence Variant
At %	Atomic Percentage
a.u	Arbitrary Units
BIOX	Biogenic Iron Ochre
BP	Base Pairs
BME	β-mercaptoethanol
CLSM	Confocal Laser Scanning Microscopy
DIC	Dissolved Inorganic Carbon
DO	Dissolved Oxygen
DOC	Dissolved Organic Carbon
DTT	Dithiothreitol
EPS	Exopolymeric Secretions
EtOH	Ethanol
EDX	Energy Dispersive X-ray Spectroscopy
FeOB	Iron Oxidising Bacteria
FeRB	Iron Reducing Bacteria
Fh	Ferrihydrite
FTIR	Fourier Transformed Infrared Spectroscopy

MeOH   Methanol

NMR   Nuclear Magnetic Resonance

SEM   Scanning Electron Microscopy

SOB   Sulfide Oxidising Bacteria

UV-Vis   Ultraviolet-visible Spectroscopy

XRD   X-ray Diffraction

Thesis Format

This thesis contains research that encompasses several different areas including inorganic chemistry, material characterisation, microscopy, isolation of bacteria, phylogenetics of environmental samples and physicochemical characterisation of environmental sample sites. This may initially seem overly broad; however, each area of research is required to build a better understanding of iron-oxidising bacteria and their associated biogenic materials.

Chapter 1 contains an introduction to the element iron and associated biogenic and abiogenic processes that it is part of. This is followed by Chapter 2 which details the experimental methodology carried out. Chapter 3 shows a study of sampled biogenic iron oxide materials using a combination of spectroscopy, x-ray diffraction and microscopy. Chapter 4 investigates the 16S rRNA phylogenetic profiles of the sample biogenic materials using high-throughput Illumina Sequencing and compares the results with literature examples that have carried out similar studies in a variety of countries. Chapter 5 then investigates and develops methods of isolating bacteria from sampled biogenic iron ochre and then includes sequencing data of selected isolates. Chapter 6 shows a study of the sample sites and contains historical information about the sample sites as well as the physicochemical characteristics and potential geological influences on the sample sites. Finally, Chapter 7 contains the conclusions and summary of the above work.

Please note that throughout this thesis, figures have been taken and occasionally modified from publications. If the figure caption contains a reference number in it, then it has been taken from a publication. If there is no reference number noted, then it is my own work.

Introduction

1.1 Iron

Iron is the fourth most abundant element in the Earth's crust where it makes up ca. 6.3 % of the elemental composition. It is also the second most abundant metal and the most abundant element by mass on Earth.[1] This ubiquity makes iron prevalent in both geological and biological systems.

1.1.1 Iron In Geological Systems

Iron exists in a variety of mineral forms in the Earth's crust. The most abundant iron minerals are generally iron oxides however iron sulfides, iron carbonates and iron silicates are also prevalent. These iron oxide minerals typically form due to the weathering of primary rocks in terrestrial and marine habitats that contain iron silicates and iron sulfides.[2] This weathering process releases Fe (II) to the environment. When the environmental conditions include pH > 2 and the presence of both oxygen and water then free Fe (II) is oxidised to Fe (III). These species are then rapidly hydrolysed to form iron oxides and oxyhydroxides.[3] Table 1 contains a selection of iron minerals and the typical environments in which they are found and Equation 1 and Equation 2 show for the formation of iron oxyhydroxides from primary rocks.

$$Fe_2SiO_4 + 0.5\ O_2 + 3\ H_2O \longrightarrow 2\ FeOOH + Si(OH)_4$$
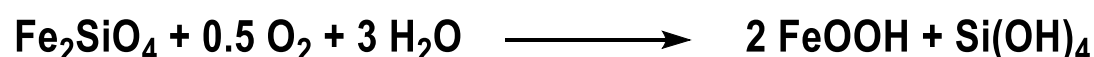
Equation 1 - Oxidation of fayalite yielding iron oxyhydroxide and orthosilicic acid.

$$4\ FeS_2 + 15\ O_2 + 10\ H_2O \longrightarrow 4\ FeOOH + 8\ H_2SO_4$$
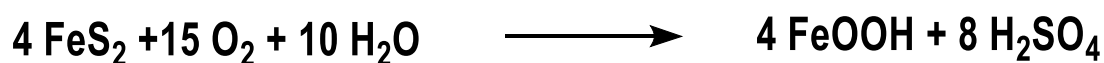
Equation 2 - Oxidation of pyrite yielding iron oxyhydroxide and sulfuric acid.

1.1.2 Iron Minerals

Table 1 contains a selection of common non-oxide iron minerals whose weathering results in the formation of environmental iron oxides. It also includes a selection of common iron oxides and the typical environments in which they are found. Following this is a brief introduction to each of the iron oxides.

Mineral	Formula	Environmental Location
Biotite	$K(Fe, Mg)_3Si_3AlO_{10}(OH)_2$	Primary Rock
Pyroxene	$(Ca, Mg, Fe, Al, Ti)_2(Si, Al)_2O_6$	Primary Rock
Fayalite	Fe_2SiO_4	Primary Rock
Pyrite	FeS_2	Primary Rock
Titanomagnetite	$Fe_{3-x}Ti_xO_4$	Primary Rock
Ilmentite	$FeTiO_3$	Primary Rock
Amphibole	$Ca_2(Mg, Al, Fe)_5(Si, Al)_8O_{22}(OH)_2$	Primary Rock
Haematite	$\alpha\text{-}Fe_2O_3$	Aerobic soils, banded iron formations and sedimentary rock
Maghemite	$\gamma\text{-}Fe_2O_3$	Aerobic soils in subtropical climates
Magnetite	Fe_3O_4	Banded iron formations and volcanic sands
Goethite	$\alpha\text{-}FeOOH$	Aerobic and anaerobic soils, bogs, wetland and cave formations
Ferrihydrite	$Fe_5HO_8.4H_2O$	Groundwater, sub-surface waters, wetland and aerobic soils

Table 1 - Iron containing minerals found in primary rocks and common iron oxides.[3]

1.1.2.1 Ferrihydrite

Ferrihydrite is an iron oxyhydroxide commonly found in areas rich in oxygen and water with a source of ferrous iron, Fe (II).[4] It is a poorly crystalline, metastable iron (III) oxyhydroxide mineral with a high density of defects, such as vacancies and stacking faults, irregular / spherical particle morphology and particle size 2 – 6 nm. These nanoparticles then aggregate to produce spheres ca. 100 nm in diameter resulting in a microporous material with surface area typically 200 – 400 m^2/g.[5] It occurs in several forms, most commonly as either 2-line ferrihydrite (2-Fh) or 6-line ferrihydrite (6-Fh) and has a composition of approximately $Fe_5HO_8.4H_2O$. The 2-line and 6-line nomenclature are indicators of crystallinity. 2-Fh has poorer crystallinity and produces two broad reflections in its X-ray diffraction (XRD) pattern while 6-Fh produces six sharper reflections. Figure 1 shows the diffraction patterns of 2-Fh and 6-Fh.
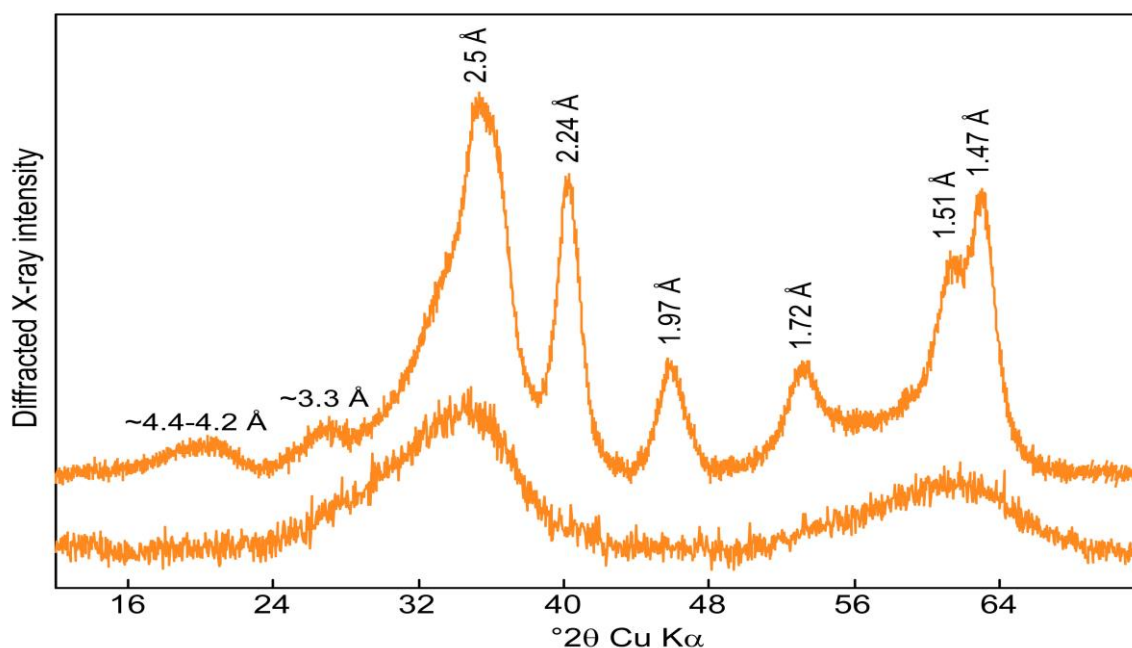
Figure 1 - XRD patterns of 2-Fh and 6-Fh from Drits *et al*. Reflections have been labelled with their corresponding d-spacing.[6]

These broad reflections arise from the fact that ferrihydrite has a nanoscale particle diameter and poor long-range ordering. This makes it difficult for techniques such as XRD to be used to determine the crystallographic structure and as such this is commonly the subject of debate in the literature.[5, 7-9]

The metastability of ferrihydrite is influenced by environmental conditions such as the concentration of organic species present. Ferrihydrite coprecipitated with environmental

organics, such as humic and fulvic acids, is found to be more stable than pure ferrihydrite.[10] Pure ferrihydrite acts as a precursor, by undergoing phase transformations, for the crystalline and more thermodynamically stable goethite and haematite phases of iron oxide.[11] Coprecipitated ferrihydrite is therefore an important environmental detoxifier due to its increased phase stability combined with its high surface area and microporous nature. Ferrihydrite will be discussed further in Chapter 3.

1.1.2.2 Haematite

Haematite (α-Fe_2O_3) is one of the oldest and most abundant iron oxides known to man. It is extremely stable under ambient conditions and is generally the final transformation product of the other iron oxide minerals.[3] It typically has a distinctive red/brown colour as a powder, although can appear black/grey, and at room temperature is anti-ferromagnetic.[12] The crystal structure is that of rhombohedral corundum (α-Al_2O_3) which is based on hexagonal close packing (ABABAB…) of O^{2-} ions with two thirds of Fe^{3+} ions occupying the octahedral holes, as illustrated in Figure 2.
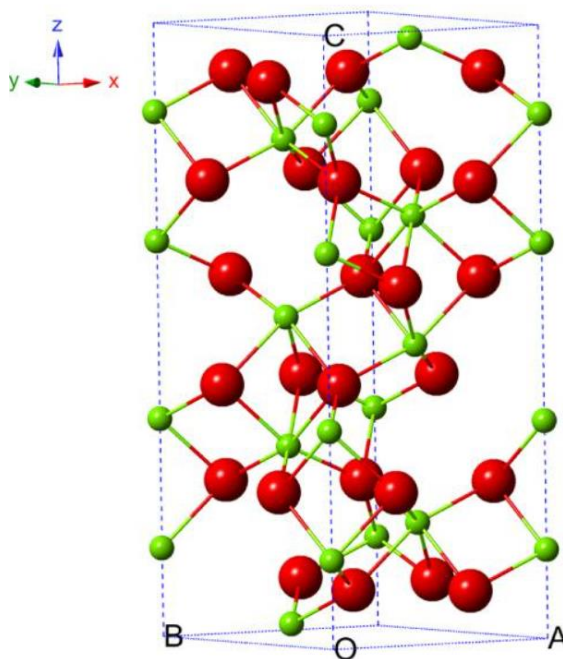
Figure 2 - Crystal structure of haematite. Green spheres = Fe^{3+} and red spheres = O^{2-}.[13]

The cation arrangement causes the formation of FeO_6 octahedra where each octahedron shares faces with one octahedron in an adjacent plane and edges with three octahedra in the same plane.

1.1.2.3 Goethite

Goethite (α-FeOOH) is highly thermodynamically stable and is regularly the most abundant iron oxide/oxyhydroxide found in wet soil and sediment systems in cooler climates.[3] Much like haematite it is anti-ferromagnetic at room temperature and in the environment forms either from the weathering of iron silicates, carbonates and sulfides or via the transformation of metastable ferrihydrite.[14]

Goethite has an orthorhombic unit cell containing hexagonally close packed anions, O^{2-} and OH^-, with Fe^{3+} ions occupying half the octahedral holes. Each Fe^{3+} ion is surrounded be three O^{2-} and three OH^- ions respectively, yielding FeO_3OH_3 octahedra.[15] Figure 3 contains a diagram of goethite's crystal structure.
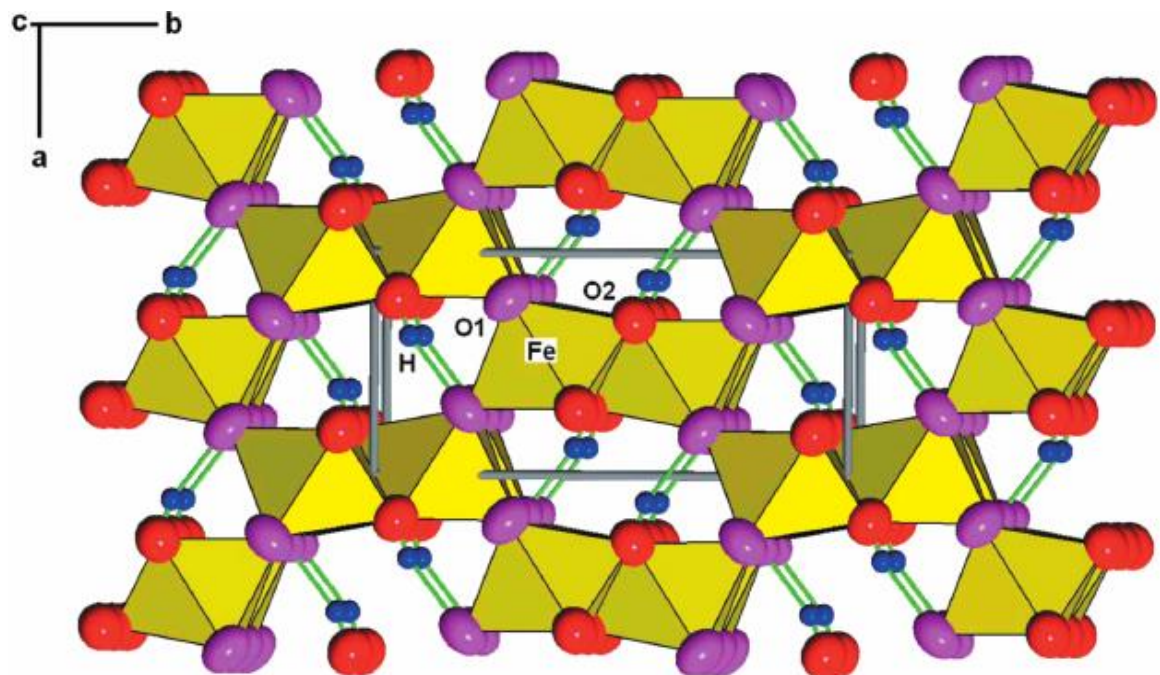
Figure 3 - Crystal structure of goethite.[16]

This shows that the crystal structure contains double chains of octahedra separated by double chains of vacant sites that appear as tunnels within the structure.

1.1.2.4 Maghemite

Maghemite (γ-Fe_2O_3) is typically formed via the weathering of magnetite (Fe_3O_4). It is a ferrimagnetic iron oxide that is polymorphous with haematite and contains a spinel like structure that is similar to magnetite (Fe_3O_4).[17] Maghemite has a cubic unit cell (ABCABC…) and can be thought of as magnetite that contains point vacancies instead of

Fe (II) sites. The ferric ions are distributed over octahedral and tetrahedral sites while the vacancies are generally octahedral. Figure 4 contains a diagram of maghemite's crystal structure.
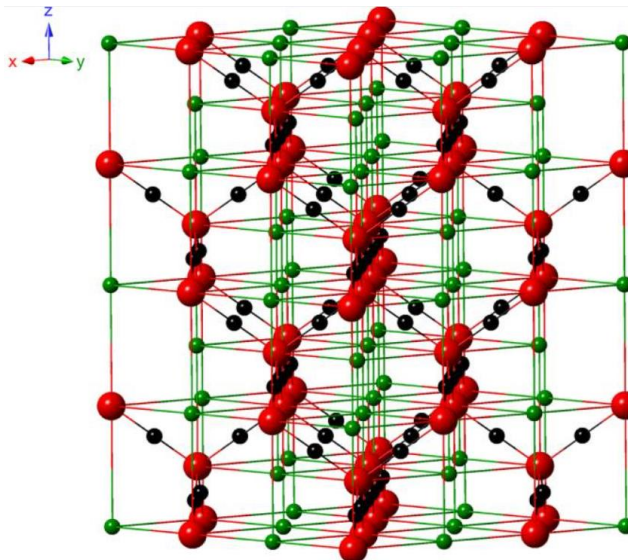
Figure 4 - Crystal structure of maghemite. Green spheres = Fe^{3+}, red spheres = O^{2-} and black spheres = vacancies.[13]

This similarity with magnetite makes it challenging to differentiate between the two using powder XRD. Instead, distinct bands in their Raman spectra are typically used for assignment.[18]

The magnetic properties of maghemite arise from two iron varying sub-lattices, one based on tetrahedral sites and the other on octahedral sites. Each sub lattice has opposing magnetic moments however the magnetic moments within each lattice are aligned anti-parallel producing an overall effective magnetic field.[17]

1.1.2.5 Magnetite

Magnetite is another ferrimagnetic iron oxide and is the most magnetic naturally occurring material on Earth.[19] It differs from other iron oxide minerals as the iron is in the form of Fe^{2+} and Fe^{3+} thus allowing magnetite the inverse spinel structure. The regular spinel structure is $A^{2+}B^{3+}O^{4-}$, where the oxide anions form a cubic close packed lattice and the cations occupy some of the interstitial octahedral and tetrahedral sites. For the inverse spinel the structural formula is written as $(Fe^{3+})_A O^{2-}[Fe^{2+}Fe^{3+}]_B O^{2-}$ where the A cations occupy tetrahedral sites and the B cations randomly occupy the octahedral sites. Figure 5

contains a diagram of magnetite's crystal structure and a diagram showing the ferrimagnetic ordering of iron electron spins.

A

B

8 Fe^{3+}$_A$

8 Fe^{3+}$_B$

Tetrahedral (A)   sites

8 Fe^{2+}$_B$

Octahedral (B)   sites
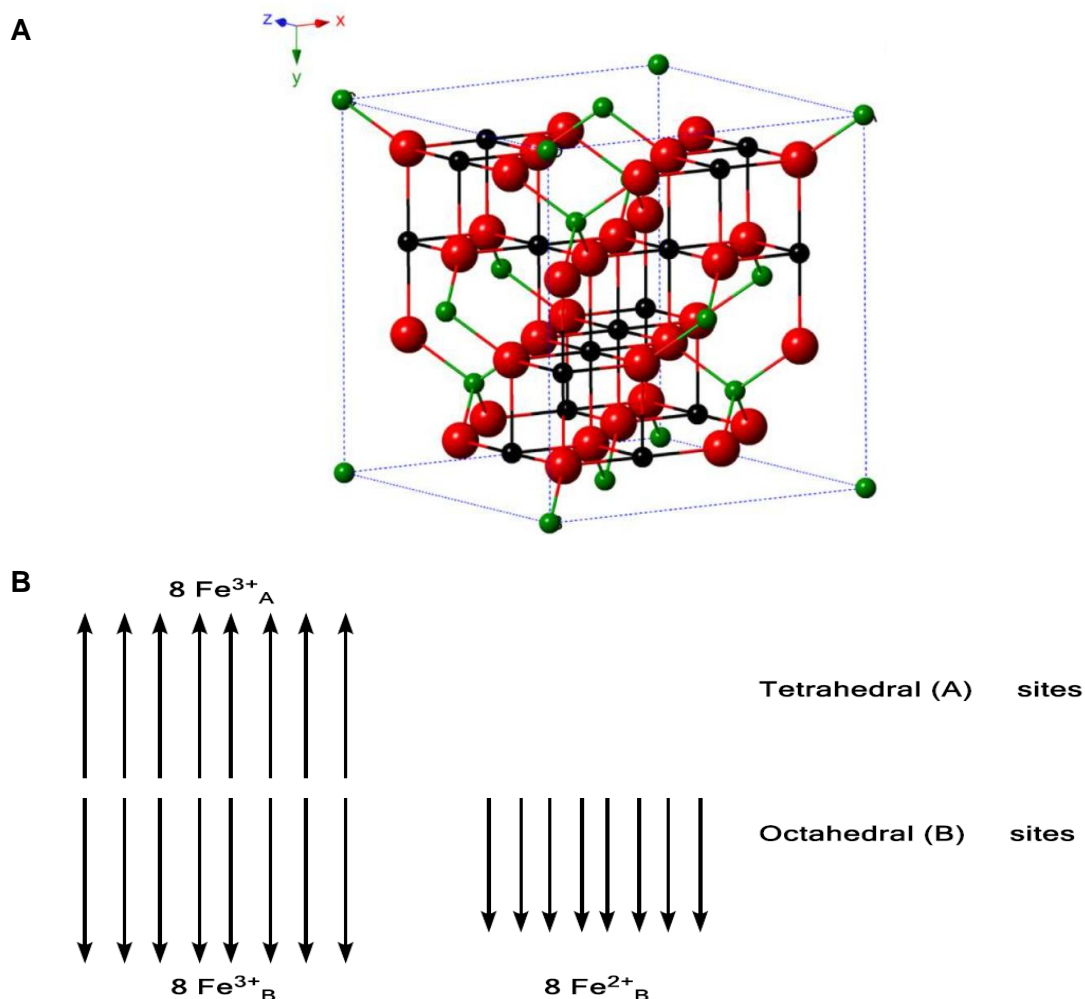
Figure 5 – A - Crystal structure of magnetite. Green spheres = Fe^{3+}, black spheres = Fe^{2+} and red spheres = O^{2-}.[13] B – Ferrimagnetic ordering of iron electron spins in magnetite.

The divalent nature of magnetite gives rise to its ferrimagnetic properties. Figure 5 image B shows that 8 Fe^{3+} spins on octahedral sites align anti-parallel with 8 Fe^{3+} spins in tetrahedral sites which cancel each other out. There are however 8 Fe^{2+} spins, also on the octahedral sites, that are not cancelled out and as such give rise to a magnetic moment.

1.1.2.6 Iron Ochre

Iron ochre is the colloquial name given to a thick orange gelatinous sludge that contains a mixture of insoluble iron oxides/oxyhydroxides and organic substances. These iron oxides/oxyhydroxides can be any of the previously discussed materials in Chapter 1 Section 1.2 however are generally 2-Fh while the associated organics can be either humic

or biological in origin. Ochre deposits are routinely found throughout the world in regions of ferrous rich groundwater and have historically polluted drainage and irrigation systems.[20-22] Ochre can either be produced by biogenic/biotic processes (BIOX) or abiogenic/abiotic processes (aBIOX). BIOX is produced by FeOB in both freshwater and marine environments across a host of pH values from acidic through circumneutral.[23] FeOB responsible for BIOX production include *Leptothrix spp.* and *Gallionella spp.* These two genera of FeOB form BIOX microstructures that include microtubular filaments and twisted stalks.[24, 25] Figure 6 contains a photograph and phase contrast micrograph of BIOX to highlight its environmental appearance and microstructure. These bacteria, and their associated microstructures, represent a method of producing metal oxide micromaterials under ambient conditions.

BIOX has been shown to be porous with high surface area and can be further functionalised as well as act catalytically. By controlling the phase of these materials it is also possible to tune their magnetic properties, for example transforming 2-Fh BIOX into magnetite or maghemite creates magnetic micromaterials that can be aligned with external fields. As such, there is currently great interest in studying these materials and the FeOB that produce them.[26-28] Interestingly, BIOX has found historical use dating back thousands of years. MacDonald *et al.* analysed the pigmentation of cave paintings near Babine Lake in the American Pacific Northwest.[29] Their study showed that the pigments used for these paintings contain microtubular filaments routinely produced by *Leptothrix ochracea*. The microstructure of these filaments is highly sintered and the phase is predominantly haematite indicating that they have been heat treated at 750 – 850 °C prior to use. This means that the hunter-gatherers responsible for these paintings have collected BIOX samples then treated them to produce different shades of red, orange and brown. Takada *et al.* and Tamura *et al.* have more recently investigated the use of BIOX and BIOX-alumina composites in creating heat stable pigments that do not lose their colour.[30, 31]
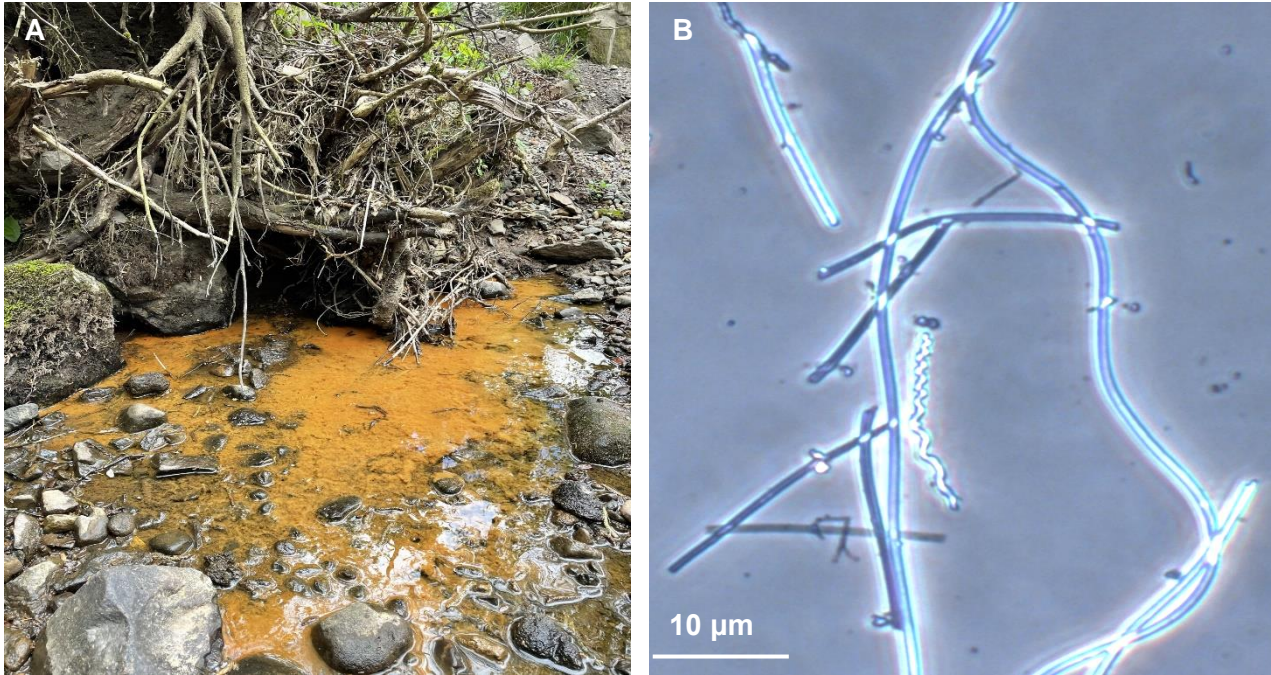
Figure 6 - A - Photograph of BIOX and B - Phase contrast micrograph of *Leptothrix spp*. filaments surrounding a *Gallionella spp*. twisted stalk.

These bacteria and their BIOX microstructures will be further discussed in Chapter 3.

1.1.3 Iron In Biological Systems

Iron is a redox active element that can exist in a variety of oxidation states from – 2 to + 6.[32] The most commonly found oxidation states of iron in biological systems are the ferrous, Fe (II), and ferric, Fe (III), oxidation states. These have a redox coupling of 0.77 V making them useful for electron transfer reactions. High valence species such as ferryl iron, Fe (IV), can also be found in biological systems as intermediates in oxidation processes.[33] Iron is an essential component of a number of proteins, such as haemoproteins, non-haem proteins and iron-sulfur proteins, where it takes part in a number of metabolic processes including electron transfer, oxygen transport, ribonucleotide reduction, dinitrogen reduction and the decomposition of peroxides.[34] Iron is so prevalent in nature that many bacteria require near-millimolar intracellular concentrations and to date the only known organisms that do not require it belong to the *Lactobacillus spp*.[35] Figure 7 shows a selection of known iron complexes found in proteins.
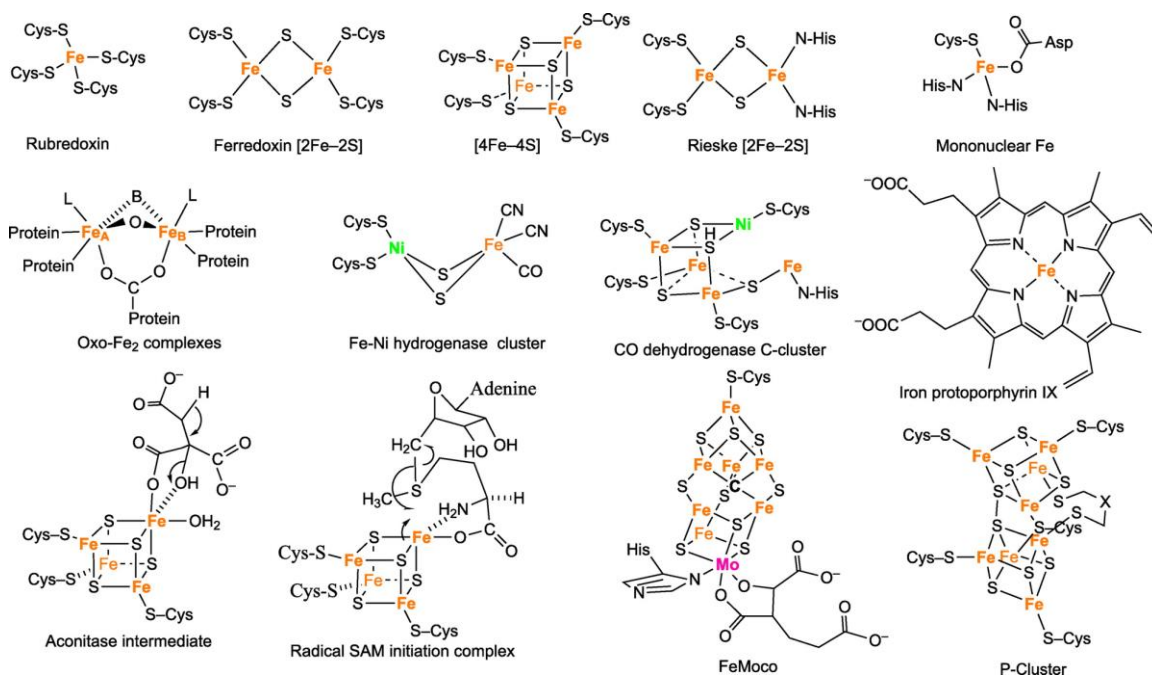
Figure 7 - Selection of chemical structures that represent known iron complexes in proteins taken from Frey and Reed.[1]

1.1.3.1 Iron Metabolism

Many types of bacteria are chemolithotrophs, meaning that they can utilise inorganic species as electron donors and subsequently acquire energy. This process can occur in both aerobic and anaerobic environments. Common inorganic species utilised include sulfur compounds, methane, hydrogen, ammonia, nitrates and iron.[36] The utilisation of non-iron species will be further discussed in Chapter 4 Section 7. The oxidation of ferrous iron yields the lowest Gibbs energy for cellular metabolism however iron-oxidising bacteria (FeOB) are a prominent group of chemolithotrophs.[23] This means that a large quantity of iron must be oxidised to make this a sufficient energy source. As such, a small number of FeOB can be responsible for the formation of large amounts of ferric material which can contribute to the formation of the previously discussed iron oxides/oxyhydroxides. Equation 3 and Equation 4 show the oxidation of ferrous iron at acidic and circumneutral pH values and their corresponding Gibbs free energy change.

$$Fe^{2+} + 0.25\ O_2 + H^+ \longrightarrow Fe^{3+} + 0.5\ H_2O \quad pH < 3,\ \Delta G^\circ = -29\ kJ\ mol^{-1}$$
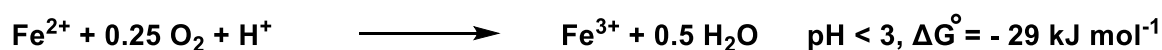
Equation 3 - Equation showing the Gibbs free energy produced during ferrous iron oxidation in acidic environments.[23]

$$\text{Fe}^{2+} + 0.25\ \text{O}_2 + 2.5\ \text{H}_2\text{O} \longrightarrow \text{Fe(OH)}_3 + 2\ \text{H}^+ \quad \text{pH ca. 7, } \Delta G^\circ = -90\ \text{kJ mol}^{-1}$$

Equation 4 - Equation showing the Gibbs free energy produced during ferrous iron oxidation in circumneutral environments.[23]

These equations show that FeOB in circumneutral environments can acquire energy more readily than FeOB in acidic environments. There are other mitigating factors however, specifically the stability of ferrous iron under these conditions. In acidic environments ferrous iron is highly stable meaning that only iron oxidation via biological processes occurs while under circumneutral conditions ferrous iron is metastable. This means that neutrophilic FeOB must outcompete abiotic oxidation processes to survive.[37] FeOB will be further discussed throughout this work and the stability of aqueous iron will be further discussed in Chapter 1 Section 2.1.

1.2 Iron Oxidation Chemistry

1.2.1 Aqueous Iron Chemistry

The oxidation of iron under aqueous conditions is a thoroughly studied and complex topic. In circumneutral waters soluble ferrous iron is oxidised to insoluble ferric iron oxyhydroxide complexes with a host of metastable intermediates also produced upon certain conditions.[38] The speciation of iron across conditions such as pH and redox potential (Eh) is commonly shown using a Pourbaix diagram (pH – Eh stability field diagram).[39, 40] Pourbaix diagrams can be thought of as diagrammatic representations of the Nernst equation, seen in Equation 5. They show the regions of thermodynamic stability of redox species for a certain element or compound.

$$E = E^\circ - \frac{RT}{nF} \ln(Q)$$

Equation 5 - Nernst equation: E = cell potential, E° = standard cell potential, R = gas constant, T = absolute temperature, n = mol of electrons transferred, F = Faraday constant and Q = reaction quotient.

The area between the upper and lower sloped lines is the stability field of water meaning all species in this area can be found in aqueous solution. Above the upper line water is oxidised while below the lower line it is reduced. Vertical lines show equilibria between species that are solely pH dependent (non-redox) whereas horizontal lines show equilibria that are solely Eh dependent (redox). Finally, sloped lines show transitions dependent on both pH and Eh. Figure 8 shows a Pourbaix diagram for iron under typical environmental

conditions and Equation 7 – 9 are example equations for the interconversion of the chemical species separated by horizontal, vertical and sloped lines.
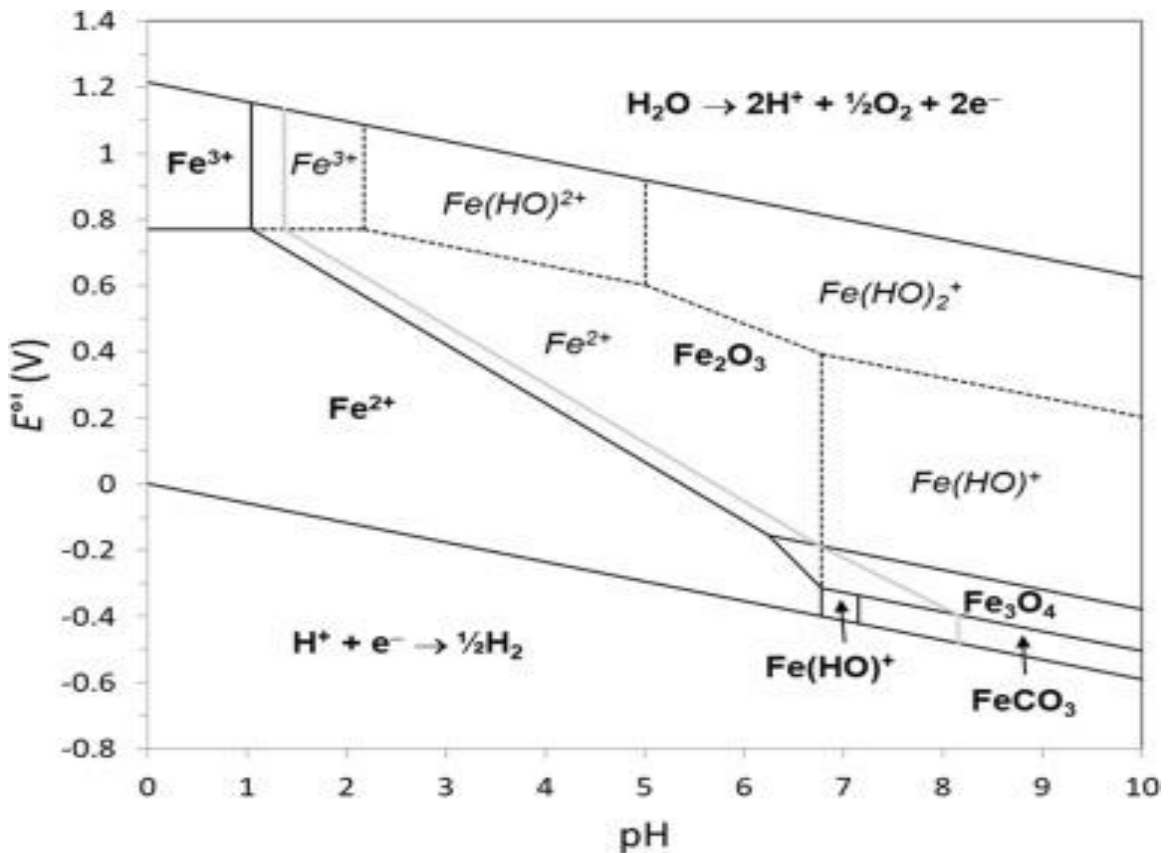
Figure 8 - Pourbaix diagram for iron under aqueous conditions. Solid lines show the stability fields of thermodynamically stable species, dashed lines those of metastable species for an iron concentration of 10 µM in black, and for 1 µM in grey.[41]

$$Fe^{3+}_{(aq)} + e^- \longrightarrow Fe^{2+}_{(aq)}$$
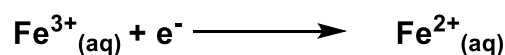
Equation 6 – Example of a redox dependent reaction that would be separated by a horizontal line.

$$2\ Fe^{3+}_{(aq)} + 3\ H_2O \longrightarrow Fe_2O_{3(s)} + 6\ H^+_{(aq)}$$
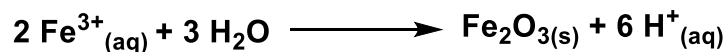
Equation 7 – Example of a pH dependent reaction that would be separated by a vertical line.

$$2\ Fe^{2+} + 3\ H_2O \longrightarrow Fe_2O_{3(s)} + 6\ H^+_{(aq)} + 2e^-$$
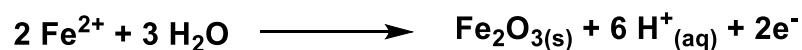
Equation 8 – Example a pH-E$_h$ dependent reaction that would be separated by a sloped line.

The above Pourbaix diagram shows that several stable and metastable species are possible. Under circumneutral and partially oxidising conditions, commonly found slightly above the oxic-anoxic boundary in many freshwater systems, the dominant species of iron is Fe_2O_3. This can be seen in the Pourbaix diagram between pH values of ca. 6 – 7 and E_h values of ca. 0.3 – 0.5. In aquatic environments this species would be hydrated so may be

written as $Fe_2O_3.3H_2O$ indicating that it is a hydrated ferric oxide.[42] As the pH decreases below 4, Fe (II) becomes the dominant species in partially reducing to oxidising solutions, whereas at a pH above 7, ferric hydroxide species and metastable species dominate. These results are seen in freshwater systems where circumneutral environments regularly contain insoluble iron ochre species while acid mine effluents contain high concentrations of soluble Fe (II).[43]

1.2.2 Abiotic Iron Oxidation

Abiotic iron oxidation is the oxidation of ferrous iron via atmospheric oxygen or other naturally occurring oxidising species such as peroxides. It occurs rapidly in circumneutral waters, with Fe (II) having a half-life of < 1 minute in fully oxygenated water.[23] Equation 9 contains a kinetic expression for this reaction in standard solution.

$$\frac{-d[Fe^{2+}]}{dt} = k[Fe^{2+}][OH^-]^2 pO_2$$

Equation 9 - Kinetic equation for Fe^{2+} oxidation with O_2 as the terminal electron acceptor. k = rate constant and p = partial pressure.[44]

This equation shows that the oxidation of iron is highly dependent on both pH and oxygen concentration. When combined with data from the Pourbaix diagram it is apparent that ferrous iron is most stable in anoxic, acidic environments that are partially reducing. These conditions are commonly associated with groundwater environments and mining effluents. This will be further discussed throughout Chapter 6. Surprisingly, it has been shown that under circumneutral conditions biotic iron oxidation processes occur at a comparable rate with abiotic processes and that each likely contributes 50 % to total iron oxidation.[45] This means that FeOB have evolved mechanisms that can compete with abiotic processes, allowing them to thrive in these environments.

1.2.3 The Fenton Reaction

The Fenton reaction, shown in Equation 10, is the ferrous iron catalysed degradation of hydrogen peroxide resulting in the formation of highly oxidising species. This reaction occurs naturally, both in water systems and *in vivo*, and is employed in wastewater management as a means of degrading pollutants.[46]

$$Fe^{2+} + H_2O_2 \longrightarrow Fe^{3+} + {}^{\cdot}OH + {}^{-}OH$$
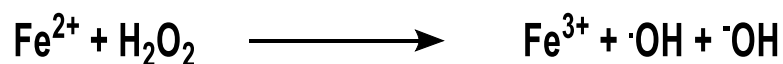
Equation 10 - Classic depiction of the Fenton reaction.

Fenton first reported the use of ferrous salts mixed with hydrogen peroxide as an oxidant of organics, specifically tartaric acid, in 1876. In 1894 Fenton then discovered that Fe was acting catalytically.[47, 48] In the 1930's Haber, Weiss and Willstätter were among the first to investigate the involvement of radical hydroxyl species in this reaction while at the same time Bray and Gorin proposed a non-radical mechanism where Fe (IV) species acted as oxidising agents instead of hydroxyl radicals. Since then, decades of conflicting mechanistic literature have been contributed to, confusing the situation further. The prevailing view currently is that both pathways are possible depending on the conditions present.[49-53] The rest of the discussion here will focus on Fenton processes in natural water systems.

Ferrous iron species can be generated at the surface of freshwater systems via the photoreduction of Fe (III) species complexed with dissolved organic matter (DOM). Strongly oxidising superoxide ($O_2^{\cdot-}$) species are also generated near the surface in a number of ways including the photoreduction of oxygen via photoexcited DOM species, the incomplete reduction of oxygen by aerobic organisms during ATP synthesis, and by some cell wall associated enzymes within blooms of microflora.[54-56] These superoxide species then rapidly disproportionate to produce hydrogen peroxide and oxygen. The resulting hydrogen peroxide may then react with the photo generated Fe (II) and begin a Fenton cycle.

$$DOM + \mathit{hn} \longrightarrow DOM^* \longrightarrow e^- + DOM^{\cdot +}$$

$$e^- + O_2 \longrightarrow O_2^{\cdot -}$$

$$DOM^* + R \longrightarrow DOM^{\cdot -} + R^{\cdot +}$$

$$DOM^{\cdot -} + O_2 \longrightarrow DOM + O_2^{\cdot -}$$

$$O_2^{\cdot -} + H^+ \longrightarrow HO_2^{\cdot}$$

$$HO_2^{\cdot} + O_2^{\cdot -} + H^+ \longrightarrow H_2O_2 + O_2$$
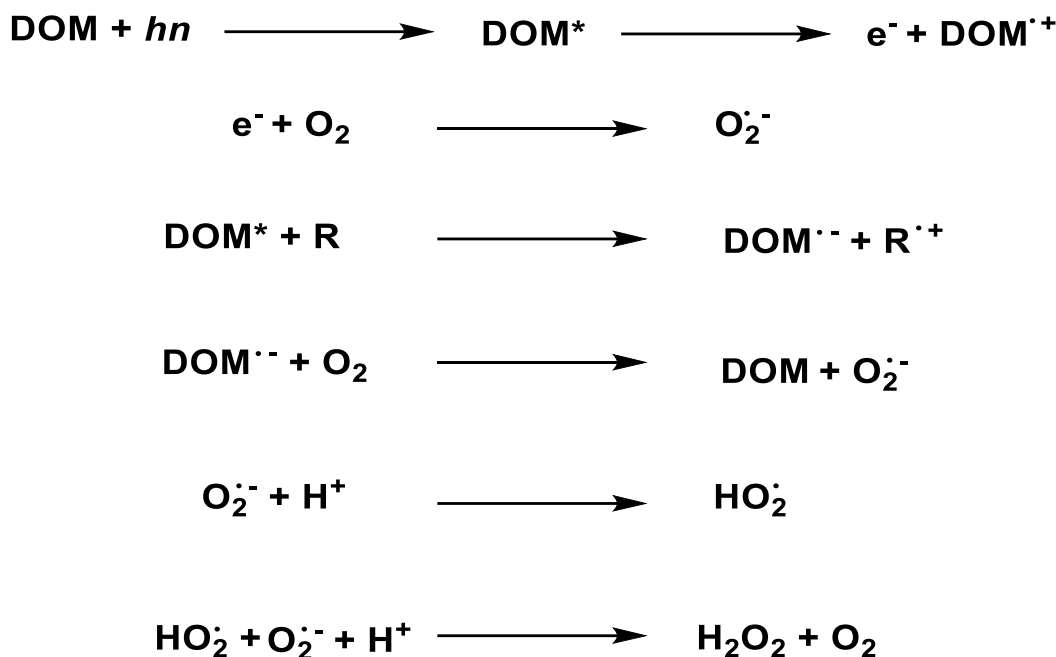
Figure 9 - Reaction scheme showing H_2O_2 production via photoexcited DOM and superoxide disproportionation. DOM – dissolved organic matter and R – lower molecular weight oxidisable organic matter.[55]

Both the photoreduction of Fe (III) and photoinitiated production of H_2O_2 have been shown to be affected by diurnal cycles. The concentration of H_2O_2 in the Kurose River in Japan was found to increase sharply from 08:00 – 12:00 and reached a daily maximum around 13:00 – 14:00 before decreasing sharply from 17:00 – 19:00 and then slowly decreasing through the night until the cycle may start over. .[57-59] These results indicate that the Fenton reaction will be much more likely to occur in the early afternoon of sunlit surface water than at other periods during the day.

1.2.3.1 Potential Fenton Mechanism

The Fenton reaction proceeds via an inner sphere mechanism where the initial step involves the formation of a ferrous hydroxperoxyl (Fe (II) – OOH$^+$) species which then further reacts to form hydroxyl radicals or ligated ferry ion species. An outer sphere mechanism would involve direct transfer of electrons from Fe (II) to H_2O_2, without the formation of any direct bonds, forming $H_2O_2^-$. This process has been shown to endergonic and is thermodynamically unfavourable when compared to an inner sphere mechanism.[60,61] Figure 10 shows a potential mechanism for the inner sphere Fenton reaction.
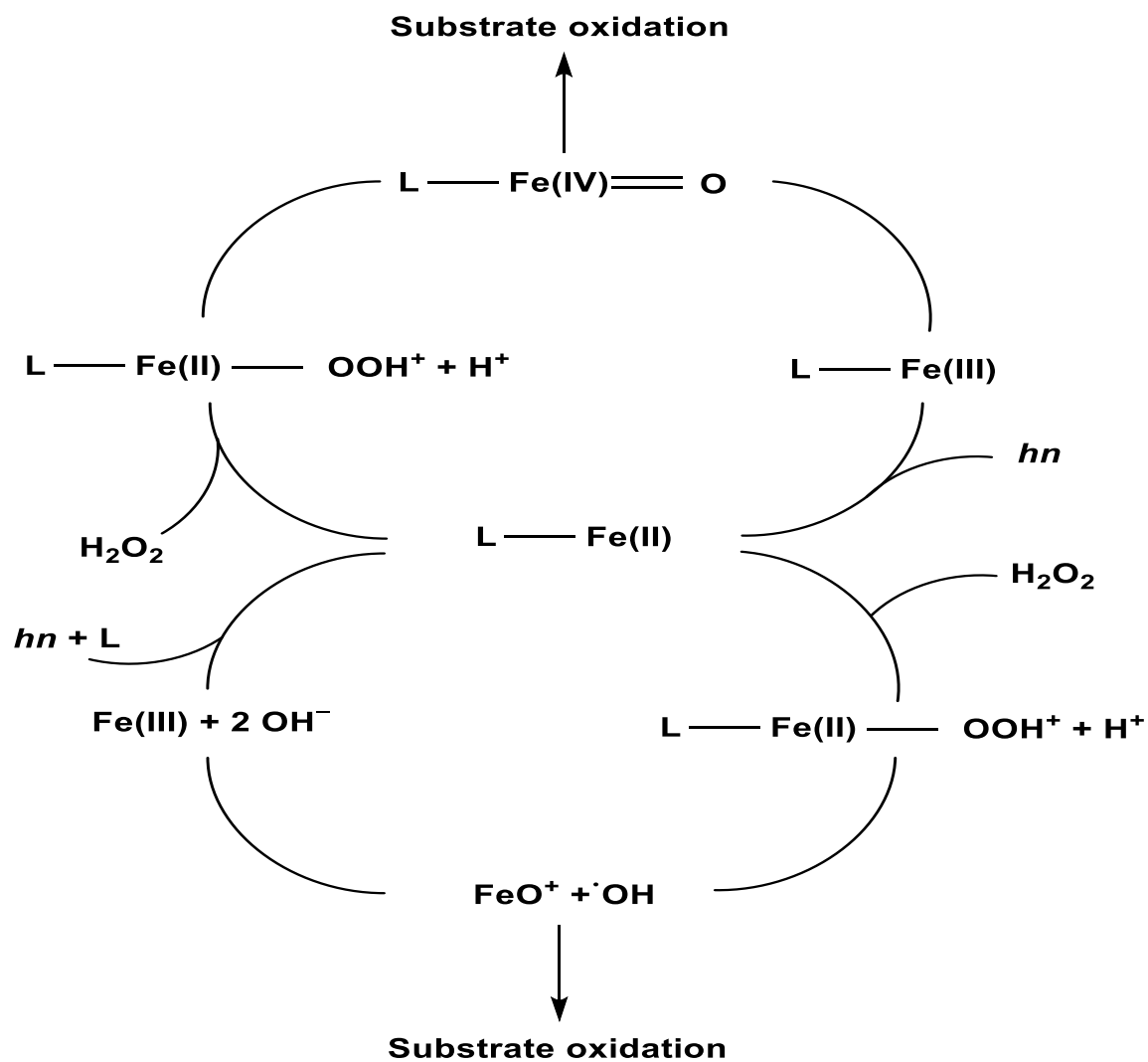
Figure 10 – Potential Fenton mechanisms to yield either highly oxidising hydroxyl radicals or ferryl ion species.

The steps following the initial complexation depend on the conditions such as pH and the ligand present. Table 2 contains information showing how the ligand present affects the oxidising species produced.

Ligand	Major Oxidising Product
No Ligand	·OH
ethylenediaminetetraacetic (EDTA)	·OH
diethylenetriaminepentaacetic acid (DTPA)	·OH
Citric Acid	Fe (IV)
Tartaric Acid	Fe (IV)
Malic Acid	Fe (IV)
Quinic Acid	Fe (IV)
EDSS (ethylenediamine-N,N'-disuccinic acid)	Fe (IV)
Nitrilotriacetate (NTA)	Fe (IV)
Tettramidomacrocyclic ligand (TAML)	Fe (IV)

Table 2 - Ligands and oxidising products formed during controlled circumneutral Fenton reactions.[46, 62]

In the presence of ligands such as EDTA and DTPA under circumneutral conditions the major oxidising product formed is the hydroxyl radical, this is also true for unchelated ferrous species. In the presence of the other tested ligands shown in Table 2 the major oxidising product is the Fe (IV) species. This suggests that under environmental conditions the Fenton reaction may produce a variety of radical species.

1.3 Project Aims

This short introduction has highlighted the importance and ubiquity of iron compounds in geological and biological systems. The metastability and aqueous chemistry of ferrous and ferric species under circumneutral conditions has been shown and the reader has been made aware of FeOB and their dependence on these ferrous and ferric compounds. It has also been highlighted that BIOX has a variety of material applications related to properties that can be tuned through phase transformations and functionalisation. This makes FeOB, and their associated BIOX, of interest to study as they potentially provide a route to metal oxide micromaterial production under ambient conditions. Currently, the production of such materials is an energy intensive process.

The following work aims to combine and expand on these topics in a study of FeOB and their habitats within the Greater Glasgow area. This begins with an investigation of the inorganic and organic composition of BIOX collected at different sample sites along with an analysis of the microscopic morphology of samples. Following this is a study of the phylogenetic profile of sample sites after 16S rRNA sequencing of extracted DNA. This data is then compared with literature examples of similar studies from around the world to explore any similarities or differences that may arise. This work is then built on with attempts to isolate filamentous bacteria and sequence their genomes. A variety of isolation and enrichment methods are employed, including the development of a micromanipulation protocol which is shown to be effective at isolating single filaments of bacteria. Finally, the sample sites themselves are studied by investigating historic anthropogenic and geological influences on them. Furthermore, the physicochemical conditions and environmental conditions at selected sample sites are considered. This is important research that provides insight to the habitat and conditions in which FeOB thrive and provides the foundation for subsequent work to isolate and functionalise BIOX micromaterials under laboratory conditions.

1.4 References

1.   P. A. Frey and G. H. Reed, *ACS Chemical Biology*, 2012, **7**, 1477-1481.
2.   U. Schwertmann, in *Iron in Soils and Clay Minerals*, eds. J. W. Stucki, B. A. Goodman and U. Schwertmann, Springer Netherlands, Dordrecht, 1988, pp. 267-308.
3.   R. M. Cornell and U. Schwertmann, *The Iron Oxides: Structure, Properties, Reactions, Occurences and Uses (First Edition),* 1996.
4.   J. L. Jambor and J. E. Dutrizac, *Chemical Reviews*, 1998, **98**, 2549-2586.
5.   F. M. Michel, L. Ehm, S. M. Antao, P. L. Lee, P. J. Chupas, G. Liu, D. R. Strongin, M. A. A. Schoonen, B. L. Phillips and J. B. Parise, *Science*, 2007, **316**, 1726.
6.   V. A. Drits, B. A. Sakharov, A. L. Salyn and A. Manceau, *Clay Minerals*, 1993, **28**, 185-207.
7.   H. Guo and A. S. Barnard, *Journal of Materials Chemistry A*, 2013, **1**, 27-42.
8.   A. Manceau, *Clay Minerals*, 2009, **44**, 19-34.
9.   F. Maillot, G. Morin, Y. Wang, D. Bonnin, P. Ildefonse, C. Chaneac and G. Calas, *Geochimica et Cosmochimica Acta*, 2011, **75**, 2708-2720.
10.   U. Schwertmann, *Nature*, 1966, **212**, 645-646.
11.   Y. Cudennec and A. Lecerf, *Journal of Solid State Chemistry*, 2006, **179**, 716-722.
12.   C. B. De Boer, T. A. T. Mullender and M. J. Dekkers, *Geophysical Journal International*, 2001, **146**, 201-216.
13.   W. Wu, Z. Wu, T. Yu, C. Jiang and K. Woo-Sik, *Science and Technology of Advanced Materials*, 2015, **16**, 023501.
14.   D. W. Strangway, R. M. Honea, B. E. McMahon and E. E. Larson, *Geophysical Journal International*, 1968, **15**, 345-359.
15.   J. F. W. Bowles, in *Encyclopedia of Geology (Second Edition)*, eds. D. Alderton and S. A. Elias, Academic Press, Oxford, 2021, pp. 442-451.
16.   H. Yang, R. Lu, R. Downs and G. Costin, *Acta Crystallographica Section E-structure Reports*, 2006, **62**, 250- 252.
17.   H. Shokrollahi, *Journal of Magnetism and Magnetic Materials*, 2017, **426**, 74-81.
18.   M. Hanesch, *Geophysical Journal International*, 2009, **177**, 941-948.
19.   R. J. Harrison, R. E. Dunin-Borkowski and A. Putnis, *Proceedings of the National Academy of Sciences of the United States of America*, 2002, **99**, 16556-16561.
20.   M. P. Denison, *Journal of the Royal Agricultural Society of England*, 1856, **17**, 625 - 629.
21.   R. E. Wheatley, *Journal of Soil Science*, 1988, **39**, 253-264.
22.   D. Vaughan and B. G. Ord, *Science of The Total Environment*, 1994, **152**, 73-80.
23.   D. Emerson, E. J. Fleming and J. M. McBeth, *Annual Review of Microbiology*, 2010, **64**, 561-583.
24.   H. Hashimoto, S. Yokoyama, H. Asaoka, Y. Kusano, Y. Ikeda, M. Seno, J. Takada, T. Fujii, M. Nakanishi and R. Murakami, *Journal of Magnetism and Magnetic Materials*, 2007, **310**, 2405-2407.
25.   T. Suzuki, H. Hashimoto, H. Ishihara, N. Matsumoto, H. Kunoh and J. Takada, *Microbes and Environments*, 2012, **27**, 338-341.
26.   J. Takada, H. Kunoh and T. Kunoh, *Journal of Microbial & Biochemical Technology*, 2015, **7**, 419-426.
27.   H. Hashimoto, T. Fujii, S. Kohara, K. Nakanishi, C. Yogi, H. Peterlik, M. Nakanishi and J. Takada, *Materials Chemistry and Physics*, 2015, **155**, 67-75.
28.   T. Ema, Y. Miyazaki, I. Kozuki, T. Sakai, H. Hashimoto and J. Takada, *Green Chemistry*, 2011, **13**, 3187-3195.
29.   B. L. MacDonald, D. Stalla, X. He, F. Rahemtulla, D. Emerson, P. A. Dube, M. R. Maschmann, C. E. Klesner and T. A. White, *Scientific Reports*, 2019, **9**, 17070.

30. J. Takada, H. Hashimoto, T. Fujii and M. Nakanishi, *ACS Applied Materias Interfaces*, 2014, **6**, 20282-20289.
31. K. Tamura, T. Kunoh, N. Nagaoka and J. Takada, *ACS Applied Bio Materials*, 2020, **3**, 27287-27294.
32. M. Ilbert and V. Bonnefoy, *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 2013, **1827**, 161-175.
33. W. Nam, *Accounts of Chemical Research*, 2007, **40**, 522-531.
34. R. R. Crichton, in *Biological Inorganic Chemistry (Second Edition)*, ed. R. R. Crichton, Elsevier, Oxford, 2012, pp. 247-277.
35. F. Archibald, *FEMS Microbiology Letters*, 1983, **19**, 29-32.
36. R. Amils, in *Encyclopedia of Astrobiology*, eds. M. Gargaud, R. Amils, J. C. Quintanilla, H. J. Cleaves, W. M. Irvine, D. L. Pinti and M. Viso, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 289-291.
37. D. Emerson, *Frontiers in Microbiology*, 2016, **6**, 1502-1502.
38. B. Morgan and O. Lahav, *Chemosphere*, 2007, **68**, 2080-2084.
39. M. J. N. Pourbaix, *Thermodynamics of Dilute Aqueous Solutions (First Edition)*, 1945.
40. B. Beverskog and I. Puigdomenech, *Corrosion Science*, 1996, **38**, 2121-2135.
41. W. H. Koppenol and R. H. Hider, *Free Radical Biology and Medicine*, 2019, **133**, 3-10.
42. J. D. Hem and W. H. Cropper, *Survey of Ferrous-Ferric Chemical Equilibria and Redox Potentials*, U.S Geological Society Report 1459A, 1959.
43. S. Panda, S. Mishra and A. Akcil, *Euro-Mediterranean Journal for Environmental Integration*, 2016, **1**, 8.
44. W. Stumm and J. J. Morgan, *Aquatic Chemistry : Chemical Equilibria and Rates in Natural Waters 3rd edition*, 1996.
45. D. Ionescu, C. Heim, L. Polerecky, V. Thiel and D. de Beer, *Geomicrobiology Journal*, 2015, **32**, 221-230.
46. G. Farinelli, M. Minella, M. Pazzi, S. Giannakis, C. Pulgarin, D. Vione and A. Tiraferri, *Journal of Hazardous Materials*, 2020, **393**, 122413.
47. H. J. H. Fenton, *Chemistry News*, 1876, **33**, 190.
48. H. J. H. Fenton, *Journal of the Chemical Society, Transactions*, 1894, **65**, 899-910.
49. F. Haber and R. Willstätter, *Berichte der deutschen chemischen Gesellschaft (A and B Series)*, 1931, **64**, 2844-2856.
50. F. Haber, J. Weiss and W. J. Pope, *Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences*, 1934, **147**, 332-351.
51. W. C. Bray and M. H. Gorin, *Journal of the American Chemical Society*, 1932, **54**, 2124-2125.
52. M. L. Kremer, *Physical Chemistry Chemical Physics*, 1999, **1**, 3595-3605.
53. P. Wardman and L. P. Candeias, *Radiation Research*, 1996, **145**, 523-531.
54. K. Barbusinski, *Ecological Chemistry and Engineering S*, 2009, **16**, 347-358.
55. J. Prousek, *Chemické Listy*, 1995, **89**, 11-21.
56. A. Vianello and F. Macri, *Journal of Bioenergetics and Biomembranes*, 1991, **23**, 409-423.
57. A. W. Vermilyea and B. M. Voelker, *Environmental Science & Technology*, 2009, **43**, 6927-6933.
58. K. M. G. Mostofa and H. Sakugawa, *Environmental Chemistry*, 2009, **6**, 524-534.
59. K. M. G. Mostofa, C.-Q. Liu, H. Sakugawa, P. Fu, M. Minella, D. Vione, D. Minakata, F.-C. Wu, L. Zhai, M. G. Mortuza, F. A. Al-Misned, J. Chen, J. Yuan and L. Si-liang, in *Surface Water Photochemistry*, The Royal Society of Chemistry, 2016, pp. 117-137.
60. C. C. Winterbourn, *Toxicology Letters*, 1995, **82-83**, 969-974.
61. S. Goldstein, D. Meyerstein and G. Czapski, *Free Radical Biology and Medicine*, 1993, **15**, 435-445.

62.   C. J. Miller, A. L. Rose and T. D. Waite, *Frontiers in Marine Science*, 2016, **3**, 134.

2 Methods

All reagents were of analytical grade and were used as received without further purification.

2.1 Sample Collection

Dr Adrian Lapthorn and Prof Justin Hargreaves kindly assisted with the collection of samples throughout this study. Water, biogenic iron ochre (BIOX) and surface film samples were collected from a variety of locations (Table 3). Water was collected by submerging glass bottles (100 ml – 1000 ml) beneath the water's surface and allowing the bottles to fill completely until there was a negligible headspace. BIOX was ladled using the lid of a bottle and was collected as close to the surface as possible to include actively growing material. Bottles of BIOX were also filled as full as possible to minimise any headspace of air. Surface film was collected in two ways. Firstly, a weigh boat that had been cut in half length-wise was used to scoop surface film whilst avoiding underlying ochre. This was then gently rinsed into glass bottles with $_dH_2O$. Secondly, surface film for SEM imaging was collected by spotting coverslips directly on to the surface film. The surface film adhered to the coverslip and was then stored in a Petri dish and submerged in a layer of 0.2 µm filtered river water to prevent drying out during transportation. Care was taken when sampling to minimise the collection of contaminants such as plant matter and sediment. Once collected, all samples were returned to the laboratory within 1 h and stored at - 80 °C for DNA extraction or 4 °C for all other uses.

Sample Site	Ordnance Survey Grid References
Allander_1	NS 5471 7573
Allander_2	NS 5471 7573
Allander_3	NS 5468 7573
Allander_4	NS 5470 7572
Dawsholm Park	NS 55751 69671
Dougalston Golf Course	NS 56641 74791
Kelvin Walkway	NS 56995 67566
Allanton	NS 86080 57798

Table 3 - Sample site names and respective Ordnance Survey grid references.

2.2 Microscopy

2.2.1 SEM-EDX

Jim Gallagher and Margaret Mullin kindly assisted with the preparation and imaging of samples via SEM-EDX.

2.2.1.1 Fixed Samples

BIOX samples had overlaying river water decanted and replaced with phosphate buffered saline (PBS) (250 ml, 10 mM). This was allowed to settle and repeated three times. Washed BIOX samples (20 µl) were then added to coverslips in 96-well plates. Coverslips with surface film adhere were also added to 96-well plates. These were gently washed twice with PBS (500 µl, 10 mM) then glutaraldehyde (1.5 %) was added and allowed to fix for 48 h. Fixed samples were then washed with sodium cacodylate buffer (0.1 M, 3 x 5 mins) and stained with OsO_4 (1 %) in sodium cacodylate buffer (0.1 M) for 1 h. Stained samples were next washed with $_dH2O$ (3 x 10 mins) and fixed in uranyl acetate (0.5 %) for

75 mins while wrapped in tinfoil. Fixed samples were then washed with $_dH2O$ to remove residual uranyl acetate and dehydrated using a gradient series of EtOH: 30 % for 10 mins, 50 % mins, 70 % for 10 mins, 90 % for 10 mins, absolute EtOH for 4 x 5 mins and dried absolute EtOH for 4 x 5 mins. Finally, samples were dried in hexamethyldisilazane (HMDS) for 2 x 5 mins, mounted on graphite tape stuck to aluminium stubs and sputtered with either Au or Au/Pd via a Polaron SC7460 sputter coater.

Scanning electron microscope (SEM) images of sputtered samples were obtained using a XL30 ESEM Phillips tungsten filament electron microscope with a secondary electron detector operating at 20 kV. Energy dispersive X-ray (EDX) analysis was performed using a connected Oxford Instruments Inca Energy 250 system with X-act 10 mm^2. Upwards of 20 images were recorded to provide a statistically representative visualisation of the samples.

SEM images of sputtered samples were also obtained using either a JEOL 6400 SEM or a JEOL IT100 SEM with a secondary electron detector operating at 10 kV. Upwards of 20 images were recorded to provide a statistically representative visualisation of the samples.

2.2.1.2 Unfixed Samples

BIOX was collected under vacuum over Whatman 1 filter paper and washed with $_dH_2O$. Large contaminants such as stones, leaves and twigs were removed with tweezers and BIOX was then suspended in EtOH. EtOH suspensions were pipetted onto graphite tape mounted on aluminium stubs, allowed to air dry and sputtered with Au via a Polaron SC7460 sputter coater. Sputtered samples were imaged using a XL30 ESEM Phillips tungsten filament electron microscope with a secondary electron detector operating at 20 kV. Energy dispersive X-ray (EDX) analysis was performed using a connected Oxford Instruments Inca Energy 250 system with X-act 10 mm^2. Upwards of 20 images were recorded to provide a statistically representative visualisation of the samples.

2.2.2 Optical Microscopy

Agar plates were viewed using an Olympus SZH10 light microscope at varying magnifications from 10 × – 70 × and images captured using a CCD camera mounted in place of the eyepiece and with the software ToupView.

Sample (15 µl – 100 µl) was added to either a microscope slide or microscope slide with a rubber gasket attached and covered with a cover slip. Microscope slides were then viewed using either a Carl Zeiss Axiovert 25, both in bright field mode and phase contrast, mode with a magnification of 50 × - 200 × or a Leica DMIRB microscope in either bright field or phase contrast mode with a magnification of 100 × - 400 ×. Images were captured using a Qimaging camera with Qcapture software.

2.2.3 Confocal Microscopy

Samples for confocal microscopy were concentrated by centrifugation (13000 RPM, 5 mins), washed twice with $_dH_2O$ (1 ml), once with NaCl (25 mM, 1 ml) and suspended in NaCl (25 mM, 0.5 ml). Samples were stained using a LIVE/DEAD ™ *Bac*Light ™ Bacterial Viability Kit from Invitrogen ™. A 1:1 mixture of Syto 9 and propidium iodide was prepared and diluted 20 ×. A 1:1 mix of diluted stain and sample was prepared and stored in the dark for 15 minutes. Sample (5 µl – 100 µl) was added to either a microscope slide or a microscope slide with a rubber gasket attached, covered with a cover slip and imaged.

A Zeiss LSM confocal microscope was used to capture confocal images and images were processed using the Zeiss ZEN software provided. An Ar laser was used to excite the sample at 488 nm. Two channels were used to collect separate live and dead images which were subsequently overlayed. The live channel measured emission at 515 nm and provided green fluorescence while the dead channel measured emission at 617 nm and provided red fluorescence. Objective lens magnification ranged from 20 × - 50 × and grayscale images were also collected.

2.3 Characterisation

2.3.1 X-ray Diffraction

2.3.1.1 Powder XRD

Powder X-ray diffraction (XRD) was performed on all pre- and post- reaction materials to identify the crystalline phases present. Ambient temperature XRD patterns were produced using an X'Pert Pro X-ray diffractometer fitted with a reflection/transmission spinning flat plate. The diffractometer had a Cu Kα source (λ = 1.5406 Å). Each pattern was produced using a step size of 0.02 ° from 5 – 85 ° 2θ for 30 minutes. ~ 0.3 g of sample was compacted into a round sample holder to prepare the material. Alternatively, a zero-

background spinner was employed where a small amount of sample was dispersed in a minimal amount of acetone, pipetted on to the spinner and dried for 2 h prior to use.

2.3.1.2 Single Crystal XRD

Dr Claire Wilson kindly assisted with the characterisation of single crystals by XRD. Single crystal XRD patterns were measured at 150 K using a Bruker D8 Venture Kappa diffractometer equipped with a Photon-II CPAD detector and dual, Cu and Mo Kα, (1.5406 and 0.7101 Å respectively) ImS 3.0 microfocus sources. Crystals were looped out of the bulk sample using a nylon loop and coated with paratone oil before mounting on the goniometer.

2.3.2 Mass Spectrometry

Dr Clement Regnault kindly assisted with the performance of Orbitrap mass spectrometry.

2.3.2.1 Orbitrap Mass Spectrometry

Hydrophilic interaction liquid chromatography (HILIC) combined with Orbitrap mass spectrometry was carried out at the University of Glasgow Polyomics facility.

HILIC was carried out on a Dionex UltiMate 3000 RSLC system (Thermo Fisher Scientific, Hemel Hempstead, UK) using a ZIC-pHILIC column (150 mm × 4.6 mm, 5 μm column, Merck Sequant). The column was maintained at 25 °C and samples were eluted with a linear gradient (20 mM ammonium carbonate in water, A and acetonitrile, B) over 26 min at a flow rate of 0.3 ml/min as follows:

Appendix IV shows a table containing all 902 identified putative compounds from this experiment along with their exact masses, formula, and number of potential isomers.

Time (minutes)	% A	% B
0	20	80
15	80	20
15	95	5
17	95	5
17	20	80
26	20	80

Table 4 - HILIC Analytical Method

The injection volume was 10 µl and samples were maintained at 5 °C prior to injection. For the MS analysis, a Thermo Orbitrap QExactive (Thermo Fisher Scientific) was operated in polarity switching mode and the MS settings were as follows:

• Resolution 70,000

• Automatic Gain Control 1e6

• m/z range 70–1050

• Sheath gas 40

• Auxiliary gas 5

• Sweep gas 1

• Probe temperature 150 °C

• Capillary temperature 320 °C

For positive mode ionisation: source voltage +3.8 kV, S-Lens RF Level 30.00, S-Lens Voltage 25.00 (V), Skimmer Voltage 15.00 (V), Inject Flatopole Offset 8.00 (V), Bent

Flatapole DC 6.00 (V). For negative mode ionisation: source voltage-3.8 kV. The calibration mass range was extended to cover small metabolites by inclusion of low-mass calibrants with the standard Thermo calmix masses (below m/z 138), butylamine ($C_4H_{11}N$) for positive ion electrospray ionisation (PIESI) mode (m/z 74.096426) and COF_3 for negative ion electrospray ionisation (NIESI) mode (m/z 84.9906726). To enhance calibration stability, lock-mass correction was also applied to each analytical run shown below. 5 Positive Mode Lock masses: Number of Lock Masses: 1 Lock Mass #1 (m/z): 144.9822 Negative Mode Lock masses: Number of Lock Masses: 1 Lock Mass #1 (m/z): 100.9856.

Water samples (1 l) were collected from respective sample sites in one litre Duran bottles. The lid of the Duran bottle was used to collect water meaning that care could be taken when sampling to minimise particulate contamination. Samples were promptly acidified to pH 2 with HCl (100 mM) and concentrated on Bond Elut PPL SPE cartridges supplied by Agilent. Cartridges were attached to a vacuum manifold and washed sequentially with one cartridge volume of $_dH_2O$, MeOH and HCl (100 mM). Samples were then loaded on to the cartridge using a peristaltic pump at a flow rate of 20 ml/min. Once loaded, samples were dried under air for 5 mins then eluted in a minimal volume of MeOH and stored at – 20 °C until needed.

2.3.2.2 Electro Spray Ionisation Mass Spectrometry

Mass spectra were collected using a Bruker microTOFq high resolution spectrometer, utilising an electrospray (ESI) source, scans 50-500 m/z and positive ion polarity. Samples were run in methanol.

2.3.3 Spectroscopy

2.3.3.1 UV-Visible (UV-Vis)

UV-Vis spectra were collected using a Jasco V550 UV-Vis spectrometer at 25 °C. H_2O (1 ml) was used as the reference solvent and samples were appropriately diluted to achieve an absorbance of less than 1.5. Each spectrum was then collected over a wavelength range of 200 – 900 nm.

2.3.3.2 Fourier Transformed Infrared Spectroscopy (FTIR)

FTIR spectra were collected using a Thermo Scientific iD5 ATR spectrometer. A small amount of dried material was loaded onto the stage and a spectrum was collected in the range of 500 – 4000 cm^{-1} for a total of 16 scans. A background was taken prior to each measurement and subtracted to give the corresponding FTIR spectrum.

2.3.3.3 Inductively Coupled Plasma Optical Emission Spectroscopy (ICP-OES)

Water samples (200 ml) were collected in plastic bottles provided by Ivario, a company based in Hamburg that assesses water quality. Collected sample were then sent to Ivario for ICP-OES analysis. Water samples were filtered through a 0.2 μm membrane to remove particulates and acidified via HNO_3 to promote solubility of constituents.

2.3.3.4 Nuclear Magnetic Resonance (NMR) Spectroscopy

Dr Brian Smith kindly assisted with the running of NMR experiments throughout this study.

2.3.3.4.1 NMR of Synthesised Ligands

^1H and ^{13}C NMR spectra were collected using a Bruker AVI 400 MHz spectrometer with chemical shift values in ppm relative to dimethylsulfoxide (DMSO) ($δH$ 2.50 and $δC$ 39.52). The sample temperature was set to 298 K. Each NMR sample was prepared in Wilmad 535-PP-7 5 mm NMR tubes to the final volume 600 μl.

2.3.3.4.2 NMR of BIOX

^1H and ^{13}C-HSQC experiments were run on a Bruker AVANCE IIIHD 600 MHz spectrometer equipped with a TCI cyroprobe using Topspin™ v3.2 software and the Fast-HSQC, phase sensitive ge-2D H – C HSQC using WATERGATE (3-9-19) pulse program. The sample temperature was set to 298 K. Each NMR sample was prepared in Wilmad 535-PP-7 5 mm NMR tubes to the final volume 600 μl in appropriate solvent.

2.3.4 Ferrozine Assay Iron Concentration

The Fe (II) and total iron (Fe$_T$) concentrations of samples were measured using an optimised ferrozine method.[1] Ferrozine (3-(2-pyridyl)-5,6-bis(4-phenylsulfonic acid)-1,2,4-triazine) (50 mg/ml) was prepared in KOAc buffer (500 mM, pH 5.5). Samples (75 μl) for

Fe (II) measurement had ferrozine (60 µl) and $_dH_2O$ (365 µl) added and were incubated at 37 °C for 1 h. Samples (75 µl) for Fe_T measurements had ascorbic acid (1 M, 15 µl), ferrozine (60 µl) and $_dH_2O$ (350 µl) added and were incubated at 37 °C for 1. After incubation samples had $_dH_2O$ (500 µl) added and their absorption profiles were measured between 650 – 500 nm. The following formula was then used to quantify the respective concentration:

$$\frac{A_{562} - A_{650}}{0.02487} \times 13.33333$$

Equation 11 - Ferrozine assay formula to measure Fe (II) and Fe_T concentrations. A_{562} = Absorbance at 562 nm, A_{650} = Absorbance at 650 nm, 0.02487 = Gradient of calibration line and 13.33333 = Dilution factor.

Calibration was carried out using $(NH_4)_2Fe(SO_4)_2.6H_2O$ with concentrations of 1 mM, 500 µM, 250 µM, 100 µM, 50 µM, 25 µM, 10 µM, 1 µM and a blank $_dH_2O$ respectively. Figure 11 shows the calibration curve for this assay.
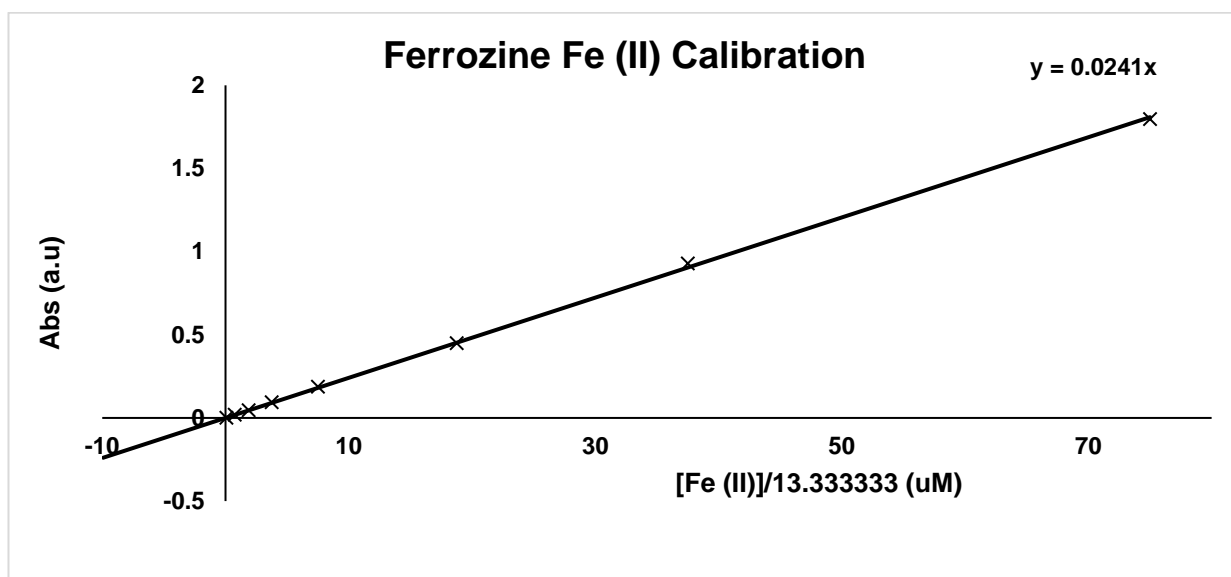
Figure 11 - Calibration curve of ferrozine assay.

2.3.5 Carbon Measurements

2.3.5.1 Dissolved Inorganic Carbon (DIC)

The DIC content of sample site water was measured at the University of Glasgow Marine Global Change Laboratory in the School of Geographical and Earth Sciences with the kind

assistance of Charlotte Slaymark. Samples (12 ml) were collected in triplicate by pipetting sample site water (24 ml) through a 0.45 µm filter into 12 ml acid washed glass vials to allow the vial to overflow. A saturated solution of $HgCl_2$ (6 µl) was then added to sterilise each sample and the vials were capped and stored at 4 °C and analysed within 24 h. Samples were analysed using an AIRICA: automated infrared inorganic carbon analyser produced by Marianda. This uses infra-red detection of the CO_2 extracted from an acidified sample. AIRICA gives a DIC reading which is within ± 1.5 to 2 µmol/kg (0.1%). The following procedure was used: A syringe pump injected the sample into a stripper which added phosphoric acid to the sample until the pH reached below 4.5. This caused CO_2 to be released. The acidified sample was then stripped of the CO_2 by a flow of nitrogen carrier gas. Water content was removed as much as possible by a Peltier-element and a Nafion dryer and the resulting gas was measured by the LI-COR 820, a non-dispersive IR (NDIR) sensor. The AIRICA analysed DIC by integration of the CO_2 ratio. Certified reference materials were used to calibrate the measurements before samples were run and also between every 10 samples. Samples were run in triplicate and the two closest repetitions taken as an average to calculate DIC concentration.

2.3.5.2 Dissolved Organic Carbon (DOC)

The DOC content of sample site water was measured at the University of Glasgow School of Geographical and Earth Sciences with the kind assistance of Kenny Wilson. Samples (250 ml) were collected in acid washed 250 ml Duran bottles, avoiding particulates and filling to the brim to avoid a headspace of air, then stored at 4 °C and analysed within 24 h. DIC was removed from the sample via acidification of the solution to pH 4 using HCl (50 mM). This acidification converted the DIC into CO_2 which was then degassed by sonication for 12 minutes. DOC concentration was then measured using a Thermalox®, from Analytical Sciences Ltd, via high temperature catalytic oxidation. The Thermalox® was fully automated and analysed multiple samples once they were added to a rack. Here, samples (100 µl) were injected into a furnace and the DOC was oxidised at 680 °C over a Pt catalyst. This results in the formation of CO_2 which is then measured via a NDIR sensor. The Thermalox® was programmed to produce standard solutions from a hydrogen potassium phthalate stock solution (1000 mg/l). These standards were interspersed with the samples being measured, as were UV-filtered $_dH_2O$ containing no DOC and an internal standard provided by Analytical Sciences Ltd. This mitigated any drift in DOC concentration that may have arisen and also prevented the accumulation of an analytical blank.

2.4 Synthesis

2.4.1 Preparation of Abiogenic Iron oxides/oxyhydroxides

Abiogenic iron oxides were prepared in accordance with procedures authored by Cornell and Schwertmann.[2]

2.4.1.1 Ferrihydrite

$Fe(NO_3)_3.9H_2O$ (8.00 g, 0.198 M) was dissolved in H_2O (100 ml). Potassium hydroxide (30 ml, 1 M) was added, and the reaction mixture was brought to pH 7-8 by further addition of potassium hydroxide (1 M) dropwise under vigorous stirring. Once neutral the mixture was centrifuged until conductivity values were consistently close to that of distilled water. The dark brown precipitate was dried under air for 18 h at 80 °C.

2.4.1.2 Goethite

The following synthesis was carried out under an inert nitrogen atmosphere. $Fe(NO_3)_3.9H_2O$ (5.00 g, 0.154 M) was dissolved in H_2O (80 ml), potassium hydroxide (20 ml, 2.5 M) was added dropwise and stirred for 30 minutes. The reaction mixture was dried under air for 18 h at 60 °C. Dilute nitric acid was added to the mixture dropwise until a pH of ~7 was reached. The reaction mixture was centrifuged washed until conductivity values were consistently close to that of distilled water. The orange solid was dried in the oven overnight (60 °C).

2.4.2 Preparation of BIOX-DTT Complex

To a stirring suspension of BIOX or ferrihydrite (0.3 g) in relevant solvent (20 ml) was added 1,4-dithiothreitol (DTT) (5ml, 0.05 M). The reaction mixture was stirred under air until a dark green suspension was produced. This suspension was collected via centrifugation and washed centrifugally with $_dH_2O$. The dark green solid was then dried under air for 18 h at 80 °C or lyophilised depending on the experiment.

2.4.3 BIOX Reduction Reactions

To a stirring suspension of BIOX in $_dH_2O$ (10 ml) was added reducing agent (25 mM). Reducing agents used were DTT, BME, TCEP, sodium dithionite and $NaBH_4$. These

stirred at RT for 10 mins or until a visible change had occurred to the BIOX. Reaction mixtures were quenched with ethanol and isopropanol and the sediment collected via centrifugation. The sediment was then washed with $_dH_2O$ and dried under air at 80 °C or lyophilised depending on the experiment.

2.4.4 Preparation of L1 Ligand

To a stirring solution of semicarbazide hydrochloride (0.911 g, 0.5 M) in MeOH (20 ml) was added salicylaldehyde (1.220 g, 0.5 M) in MeOH (20 ml). The mixture was refluxed for 4 h at 65 °C, concentrated to 10 ml *in vacuo* and stored at 4 °C for 18 h. A white precipitate was then collected, washed with MeOH and dried under vacuum. (Yield = 1.616 g, 80 %)

NMR data for (L1): δH (400 MHz, DMSO-d$_6$) 6.77–6.89 (2H, m), 7.17–7.25 (1H, m), 7.85–7.96 (2H, m), 8.09 (1H, s), 8.36 (1H, s), 9.86 (1H, s), 11.36 (1H, s); δC (101 MHz, DMSO-d6) 116.0 (CH), 119.2 (CH), 120.3 (C), 126.7 (CH), 131.0 (CH), 139.6 (C), 156.4 (CH), 177.7 (C). Mass spectrometry data: *m/z* (ESI) 218.0354 (MNa+. $C_8H_9N_3NaO$ requires 218.0359).

2.4.5 Preparation of MS1 Complex

To a stirring solution of $FeCl_3$ (0.0838 g, 0.103 M) in EtOH (5 ml) was added L1 (0.2018 g, 0.207 M). Once dissolved the stirring was turned off and the solution was stored in the dark at room temperature for 18 h to crystallize. (Yield = 0.185 g, 75 %).

2.5 BIOX EPS Extraction

Extraction protocols were based on work by Boleij *et al.* and were adjusted appropriately.[3]

2.5.1 NaOH Extraction

BIOX (5 g wet) was collected by centrifugation (5000 rpm, 10 mins, 4 °C), centrifugally washed with $_dH_2O$ twice and suspended in NaCl (25 mM, 10 ml). This suspension was shaken for 10 mins and the BIOX was collected using the same centrifugation method. NaCl washing was repeated three times after which the BIOX pellet was suspended in NaOH (10 ml, 200 mM, 4 °C) for 4 h with occasional shaking. BIOX was collected by centrifugation (20000 RPM, 20 mins, 4 °C) for further extraction. The dark red/brown

supernatant was collected, neutralised with HCl (1 M), then either precipitated with cold EtOH or had sodium dithionite (50 mM) added and was combined with a DOWEX™ MARATHON™ C (Na form) resin and shaken for 1 h to remove any soluble Fe (II). Soluble Fe (II) was also removed via dialysis against oxalic acid (1 l, 1 %) The resulting solution was then filtered under vacuum over a 0.45 µm cellulose filter and dialysed against $_d$H$_2$O using the appropriate molecular weight cut-off (MWCO) tubing, this was typically Spectrum™ Spectra/Por™ 3 RC Dialysis Membrane Tubing with a 3.5 kDa MWCO. The dialysed sample was then lyophilised and stored at 4 °C until needed.

2.5.2 Dithionite Extraction

NaOH extracted BIOX was suspended in sodium dithionite (10 ml, 200 mM, 4 °C) for 4 h with occasional shaking. BIOX suspensions were then centrifuged (20000 RPM, 20 mins, 4 °C) and the supernatant was collected and incubated on a shaker at room temperature with a DOWEX™ MARATHON™ C (Na form) resin for 1 h to remove any soluble Fe (II). The resulting pale yellow solution was filtered under vacuum over a 0.45 µm cellulose filter and dialysed against $_d$H$_2$O using the appropriate MWCO tubing, this was typically Spectrum™ Spectra/Por™ 3 RC Dialysis Membrane Tubing with a 3.5 kDa MWCO. The dialysed sample was then lyophilised and stored at 4 °C until needed.

2.6 Bacterial Isolation

All equipment used during isolation protocols was sterilised prior to use. Single use sterile loops were used when transferring bacteria between media and plates and when streaking bacteria on plates. Sterile pipette tips were also used when handling growths and media. Appendix II contains the 16S rRNA sequences of all bacteria isolated during this study.

2.6.1 Growth Media

A variety of both liquid and solid growth media were used to isolated bacteria from environmental samples. Solid growth media containing agar (3.75 g/l) is denoted – A and solid growth media containing gellan gum (6 g/l) is denoted – G. A cycloheximide (100 mg/ml, 60 % EtOH) stock solution was prepared and added to all growth media at a concentration of 1 ml/l. An ATCC MD-VS ™ vitamin supplement was also added to all media at a concentration of 1 ml/l. The composition of this supplement can be found in Table 5. These supplementary solutions were filter sterilised (0.2 µm) and added once the

media had cooled to prevent heat degradation of the supplements. Sodium pyruvate and glucose were also added in this manner to prevent the Maillard rection from occurring. Solutions of phosphate salts were autoclaved separately from agar and gellan and added afterwards to prevent the formation of hydrogen peroxide in the media. All media was prepared by dissolving components in either $_d$H$_2$O or 0.2 µm filtered sample site H$_2$O, adjusting to pH 7 with NaOH and autoclaving.

Component	Concentration (mg/l)
Folic Acid	2
Pyridoxine hydrochloride	10
Riboflavin	5
Biotin	2
Thiamine	5
Nicotinic Acid	5
Calcium Pantothenate	5
Vitamin B12	0.1
p-Aminobenzoic acid	5
Thioctic acid	5
Monopotassium phosphate	900

Table 5 - Composition of ATCC vitamin supplement.

2.6.1.1 Groundwater Phosphate (GP) Medium

GP medium was prepared in accordance with Sawayama *et al*.[4] GP media contained the following components in 0.2 μm filtered sample site water.

Component	Concentration (g/l)
$Na_2HPO_4.12H_2O$	0.076
$KH_2PO_4.12H_2O$	0.02
HEPES	2.383
$FeSO_4$ (10 mM)	1 ml

Table 6 - Composition of GP medium.

2.6.1.2 FW70 Medium

FW70 medium was prepared in accordance with Imazaki and Kobori and was supplemented with iron salts when appropriate.[5] FW70 media contained the following components in 0.2 µm filtered sample site water.

Component	Concentration (g/l)
Tryptone	1
Sodium Pyruvate	0.02
Fe Lactate (100 mM)	1 ml

Table 7 - Composition of FW70 medium.

2.6.1.3 Silicon Glucose Peptone (SGP) Medium

SGP media was prepared in accordance with Sawayama *et al*. and was supplemented with iron salts when appropriate.[4] SGP media contained the following components in $_dH_2O$.

Component	Mass (g/l)
Glucose	1.000
Soy Peptone	1.000
$NaSiO_3.9H_2O$	0.200
$CaCl_2.2H_2O$	0.044
$MgSO_4.7H_2O$	0.041
$Na_2HPO_4.12H_2O$	0.076
$KH_2PO_4.2H_2O$	0.020
HEPES	2.383

Table 8 - Composition of SGP medium.

2.6.1.4 Minerals Salts Vitamins Pyruvate (MSVP) Medium

MSVP medium was prepared in accordance with Emerson and Ghiorse and supplemented with iron salts when appropriate.[6] MSVP media contained the following components in $_dH_2O$.

Component	Mass (g/l)
Sodium Pyruvate	1
$(NH_4)_2SO_4$	0.24
$MgSO_4.7H_2O$	0.06
$CaCl_2.2H_2O$	0.06
KH_2PO_4	0.02
Na_2HPO_4	0.03
HEPES	2.383

Table 9 - Composition of MSVP medium.

2.6.2 Single Colony Isolation

Freshly collected BIOX was diluted 1:10 and 1:100 with either 0.2 µm filtered sample site water or with PBS (10 mM). Diluted BIOX (20 µl) was then streaked on the appropriate solid media and allowed to grow until distinct colonies formed. Single colonies were picked, added to appropriate media (20 ml) in 25 ml plastic universals with loose caps, and allowed to grow either statically or with shaking until appreciable growth could be seen. This was then repeated until plates contained only one colony type. Aliquots (1 ml) of single strains were frozen in liquid nitrogen, both with and without 10 % glycerol, and stored at – 80 °C.

2.6.3 Single Filament Isolation

Dr Mathis Riehle and Dr Adrian Lapthorn kindly assisted with the set-up of the micromanipulator and helped develop the protocol used.

All microscope slides were sterilised under UV light for 15 minutes prior to use and gloves and facemasks were worn throughout the experiments..

Single filaments of cells were isolated from BIOX using a Leica DMIRB microscope with a mounted micromanipulator attached (Figure 12). Objective lens magnifications were 10 ×, 20 × and 40 × and samples were viewed in brightfield, darkfield and phase contrast modes. Images and videos were captured using a Qimaging camera with Qcapture software.

BIOX samples were loaded on to 50 μm mesh netting and then spotted on a droplet of water to pull through any large particulates. The loaded netting was then flipped over so the BIOX side was face down and gently washed with 600 μl $_dH_2O$ on to a microscope slide with a rubber gasket attached. No coverslip was added, and these samples were then left for 1 h to allow filamentous bacteria to begin extending from the BIOX aggregates. Once individual filaments could be seen without satellite bacteria surrounding them a glass capillary mounted on a micromanipulator was used to extract the filament from the sample. The filament was then ejected onto a fresh microscope slide with rubber gasket attached and extracted again to remove satellite bacteria. This step was repeated to ensure that no satellite bacteria were captured. The isolated filament was then ejected into a droplet of sterile PBS (10 mM) in a PCR tube and processed to extract the DNA and amplify it for sequencing.
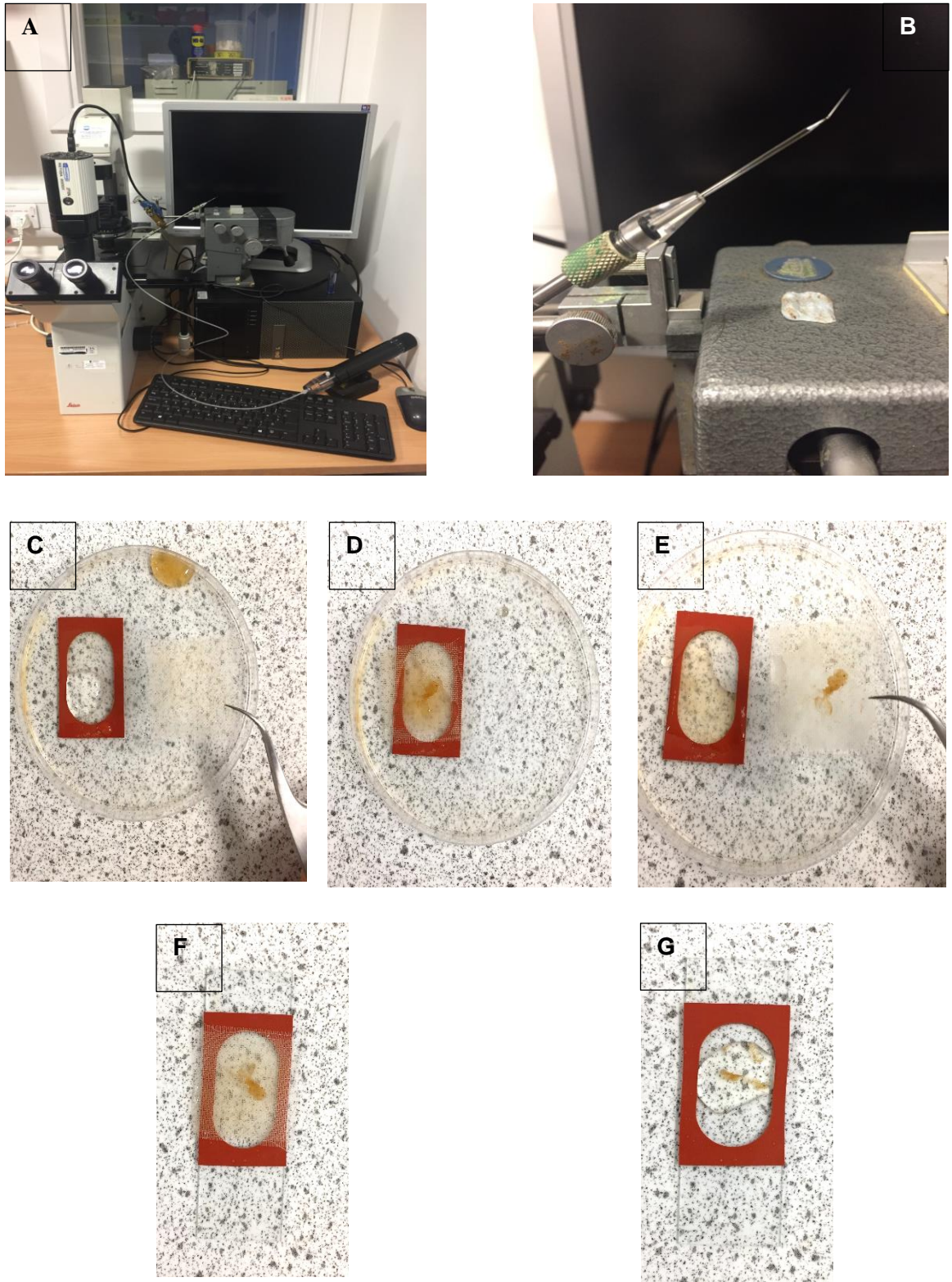
Figure 12 - A & B – Micromanipulator set up, C – Preparing to load sample onto mesh, D – Sample loaded onto mesh, E – Loaded mesh flipped, F – Sample ready to be washed onto sterilised slide and G – Filtered sample ready to have filamentous bacteria extracted.

2.7 DNA Extraction and Sequencing

DNA for sequencing was extracted from axenic cultures, single filaments and from fresh BIOX samples.

2.7.1 Axenic Cultures

Single colonies of axenic cultures were streaked on the corresponding plates to produce a lawn of bacteria. The lawns were then harvested, placed in a collection tube, and sent to MicrobesNG for high-throughput Illumina MiSeq sequencing.

2.7.2 Single Filaments

Isolated single filaments had their genomic DNA extracted and amplified using a REPLI-g Single Cell Kit supplied by Qiagen following the provided protocol.[7] Genomic DNA was amplified by PCR, washed on a 10 kDa spin column and eluted in Tris.HCl (10 mM, 500 µl, pH 8). A NanoDrop spectrophotometer was used to estimate the quantity and purity of the DNA, making sure that the concentration was greater than 10 ng/µl in 30 – 100 µl. Purified genomic DNA was then sent to MicrobesNG for high-throughput Illumina MiSeq sequencing.

2.7.3 BIOX

Prof Cindy Smith and Mr Dominic Quinn kindly assisted with the extraction of genomic DNA from BIOX samples.

Freshly collected BIOX samples were transported to the laboratory within 1 h and were stored at – 80 °C until needed. 0.5 g samples had their genomic DNA extracted using a FastDNA ™ SPIN Kit for Soil supplied by MPBio following the provided protocol.[8] Extracted genomic DNA was run on a gel to confirm that there had been little or no fragmentation. Samples were then sent to GENEWIZ for high-throughput Illumina MiSeq sequencing. Appendix III contains the 16S rRNA sequences and abundance data of the ten most abundant bacteria from each sample as well as the abundance data for the following forty most abundant bacteria from these sites respectively.

2.8 Bioinformatics

Data relating to extracted DNA was processed using a variety of software packages. Procedures specific to experimental requirements will be discussed further in the related chapters.

2.8.1 R

R is an open-source software package primarily used for statistical analysis. For this research it was used as a means of converting CSV files to FASTA files and *vice versa*. Pipelines such as DADA2 were utilised, along with the Silva nr99 v138 train set, to analyse sequencing data. This allowed the generation of amplicon sequence variants (ASVs) and assignment of taxonomy.[9, 10] The relevant R-scripts used for these experiments can be found in Appendix 1.

2.8.2 Jalview

Jalview is an open-source software package used for the visualisation and alignment of sequencing data.[11] Jalview was used in this study to truncate and align 16S rRNA sequencing data from FASTA files via the ClustalO method, calculate percentage sequence similarity between selected sequences and remove duplicate sequences from datasets. Manipulated sequences were then further analysed using MEGA-X per Section 2.8.3.

2.8.3 MEGA-X

MEGA-X was used to generate maximum likelihood phylogenetic trees of 16S rRNA sequences that had been accurately truncated and aligned as described in Section 2.8.2.[12] Relevant sequences for outgroups were chosen and included, as were sequences of closely related found via a BLAST search.

The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage

of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value. This analysis involved 62 nucleotide sequences. There were a total of 429 positions in the final dataset. Evolutionary analyses were conducted in MEGA X and bootstrap values <70 were removed from the trees.

2.8.4 Genome Analysis

Metagenomes were split into their constituent parts using the dereplication, aggregation and scoring strategy (DAS) tool.[13] The individual genome qualities were assessed using CheckM, Figure 13 and were then given individual taxonomic assignments using the genome taxonomy database GTDB-Tk.[14, 15] These bioinformatics programs are part of KBase, an online platform that integrates a variety of data and analysis tools from the DOE and other public services into an easy-to-use platform coupled with computing infrastructure.[16]
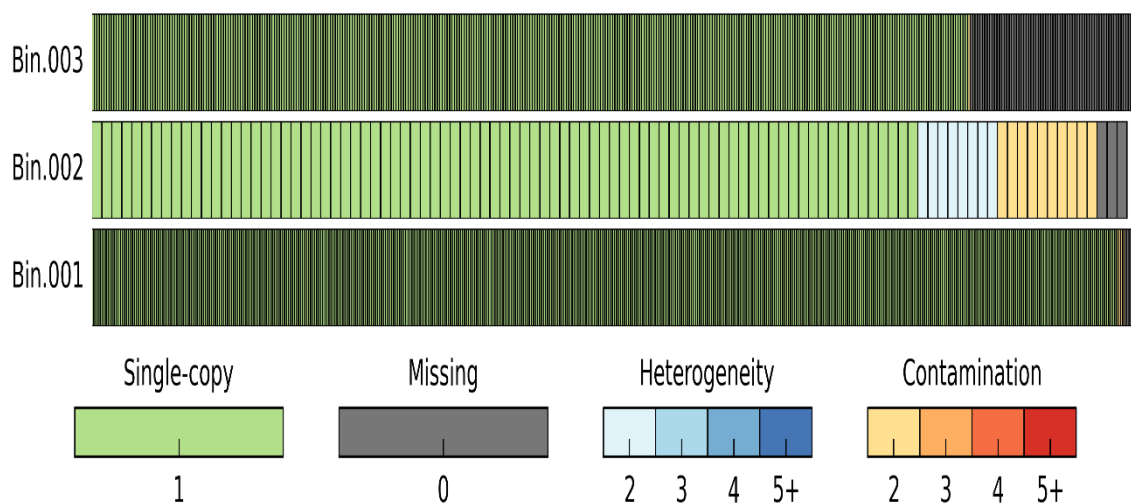
Figure 13 – Example of CheckM genome quality assessment.

2.9 References

1.   T. M. Jeitner, *Analytical Biochemistry*, 2014, **454**, 36-37.
2.   R. M. Cornell and U. Schwertmann, *The Iron Oxides: Structure, Properties, Reactions, Occurences and Uses (First Edition),* 1996.
3.   M. Boleij, T. Seviour, L. L. Wong, M. C. M. van Loosdrecht and Y. Lin, *Water Research*, 2019, **164**, 114952.
4.   M. Sawayama, T. Suzuki, H. Hashimoto, T. Kasai, M. Furutani, N. Miyata, H. Kunoh and J. Takada, *Current Microbiology*, 2011, **63**, 173-180.
5.   I. Imazaki and Y. Kobori, *Canadian Journal of Microbiology*, 2010, **56**, 333-341.
6.   D. Emerson and W. C. Ghiorse, *Applied and Environmental Microbiology*, 1992, **58**, 4001-4010.
7.   Qiagen, Repli-G Single Cell Kit For Whole Genome Amplification Handbook

<https://www.qiagen.com/us/resources/resourcedetail?id=38faca1c-64b0-4281-aab3-aa8324bbd181(=en>).
8.   M. Bio., Fast DNA SPIN Kit For Soil Manual <https://www.mpbio.com/116560000-fastdna-spin-kit-for-soil-samp-cf>).
9.   A. E. Pérez-Cobas, L. Gomez-Valero and C. Buchrieser, *Microbial Genomics*, 2020, **6**, mgen000409.
10.   B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson and S. P. Holmes, *Nature Methods*, 2016, **13**, 581-583.
11.   A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp and G. J. Barton, *Bioinformatics*, 2009, **25**, 1189-1191.
12.   S. Kumar, G. Stecher, M. Li, C. Knyaz and K. Tamura, *Molecular Biology and Evolution*, 2018, **35**, 1547-1549.

13.   Sieber, C.M.K., Probst, A.J., Sharrar, A.*et al.*Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy,.*Nat Microbiol*, 2018, **3,** 836–843.

14.   Parks, D.H, Imelfort, M. Skennerton, C.T. *et al*. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 2015, **25**, 1043-1055.

15.   Chaumeil, P.A, Mussig, A.J, Hugenholtz. P, Parks, D.H, GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database, *Bioinformatics*, 2020, **36**, 1925-1927.

16.   Arkin, A.P, Cottingham, R.W, Henry, C.S. *et al*. KBase: The United States Department of Energy Systems Biology Knowledgebase, *Nature Biotechnology*, 2018, **36**, 566-569.

3 Biogenic Iron Ochre Characterisation

3.1 Hypotheses

Prior to commencing work on this chapter, the following hypotheses were developed. Firstly, it was hypothesised that samples collected would contain microtubular filaments of 2-Fh and that these should be consistent with those seen in *Leptothrix* related literature. It was also assumed that twisted stalks consistent with *Gallionella* related literature may be present and that the elemental composition of these materials would be predominantly iron, oxygen and carbon. Secondly, it was hypothesised that by reducing ferric iron to ferrous iron it would be possible to separate the inorganic material from the organic material and subsequently characterise repeat units found within any organic backbone that is present. Finally, it was hypothesised that the interactions responsible for the formation of dark green aggregates between BIOX and DTT could be described through the development of a model system and utilisation of known dark green iron-thiol complexes.

3.2 Introduction

FeOB are found in a variety of freshwater and marine environments around the world. These environments may vary in pH, being either acidic or circumneutral, and in dissolved oxygen concentration, being either oxic or anoxic.[1, 2] The common environmental condition found in all areas where FeOB thrive is a continual source of dissolved ferrous iron typically in excess of 100 µM. At circumneutral pH in ferrous rich freshwater neutrophilic FeOB such as *Leptothrix spp.*, *Gallionella spp* are the dominant FeOB present, while acidophiles such as *Sideroxydans spp.* can also be found.[3] These bacteria are known for their production of copious amounts of BIOX, which is often a thick gelatinous sludge comprising iron oxides/oxyhydroxides and bacterial exopolymeric secretions (EPS) (Figure 14).
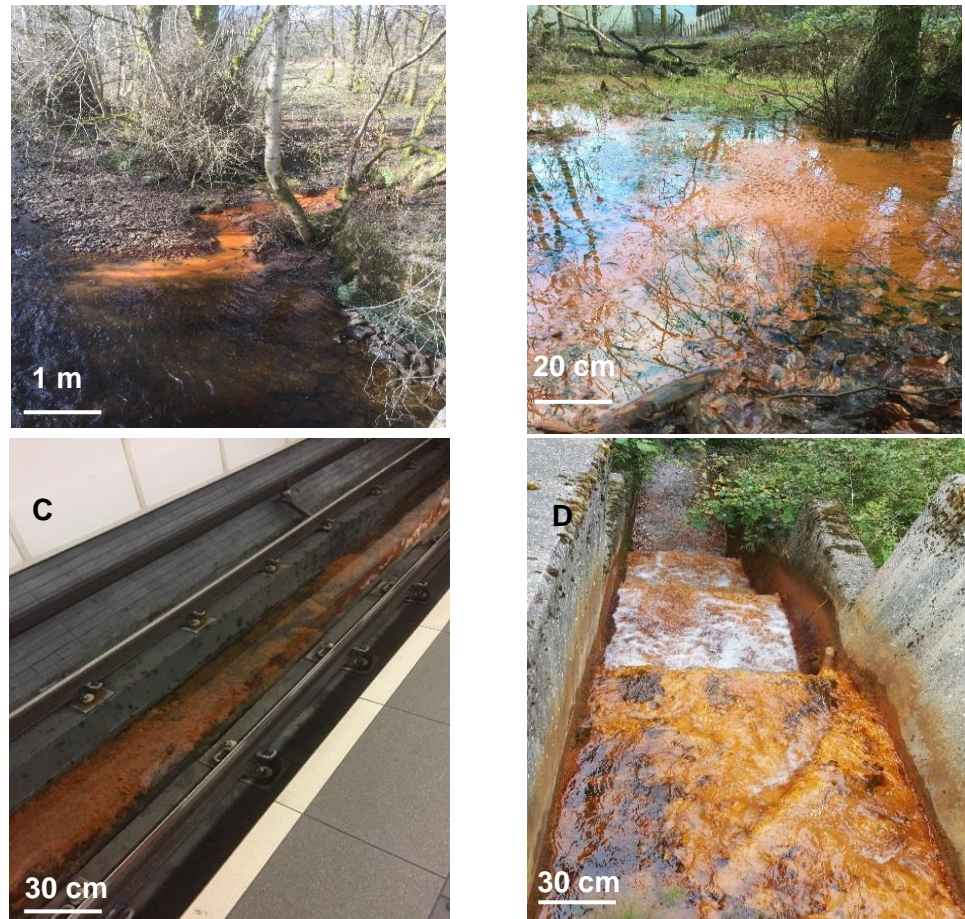
Figure 14 - Photographs of BIOX taken in Glasgow and surrounding areas. A and B - River Allander, C - Kelvinbridge subway station and D - Disused coalmine in Allanton.

This material has been known to clog irrigation systems, wells and pipes and contribute to the biocorrosion of steel of ships and nuclear plants leading to these bacteria being called nuisance bacteria and their BIOX production being known as biofouling.[4-6] These BIOX are typically seen as a waste material and an environmental contaminant however current research has shown that they possess interesting properties and may have applications for materials science.[7] The bulk material is amorphous and is composed of unique microstructures that *Leptothrix spp.* and *Gallionella spp.* produce. *Leptothrix spp.* such as *Leptothrix ochracea* are known to produce hollow microtubular filaments of BIOX while *Gallionella spp.* such as *Gallionella ferruginea* are known to produce twisted stalks of BIOX. There are also other FeOB such as *Toxothrix spp.* which are known to produce complex trichomes of BIOX (Figure 15) however these bacteria are less commonly found. These BIOX microstructures are formed under ambient conditions making them of interest to study as iron oxide microstructures are typically synthesised using energy intensive methods such as microwave assisted reactions or thermal oxidation.[8, 9] A final notable genus of FeOB that is commonly found in BIOX mats is the *Sideroxydans spp.*[1] This genus of FeOB differ with the previously described examples in that they do not produce

iron oxide microstructures, instead they contribute to the production of particulate iron oxides to which they become associated.
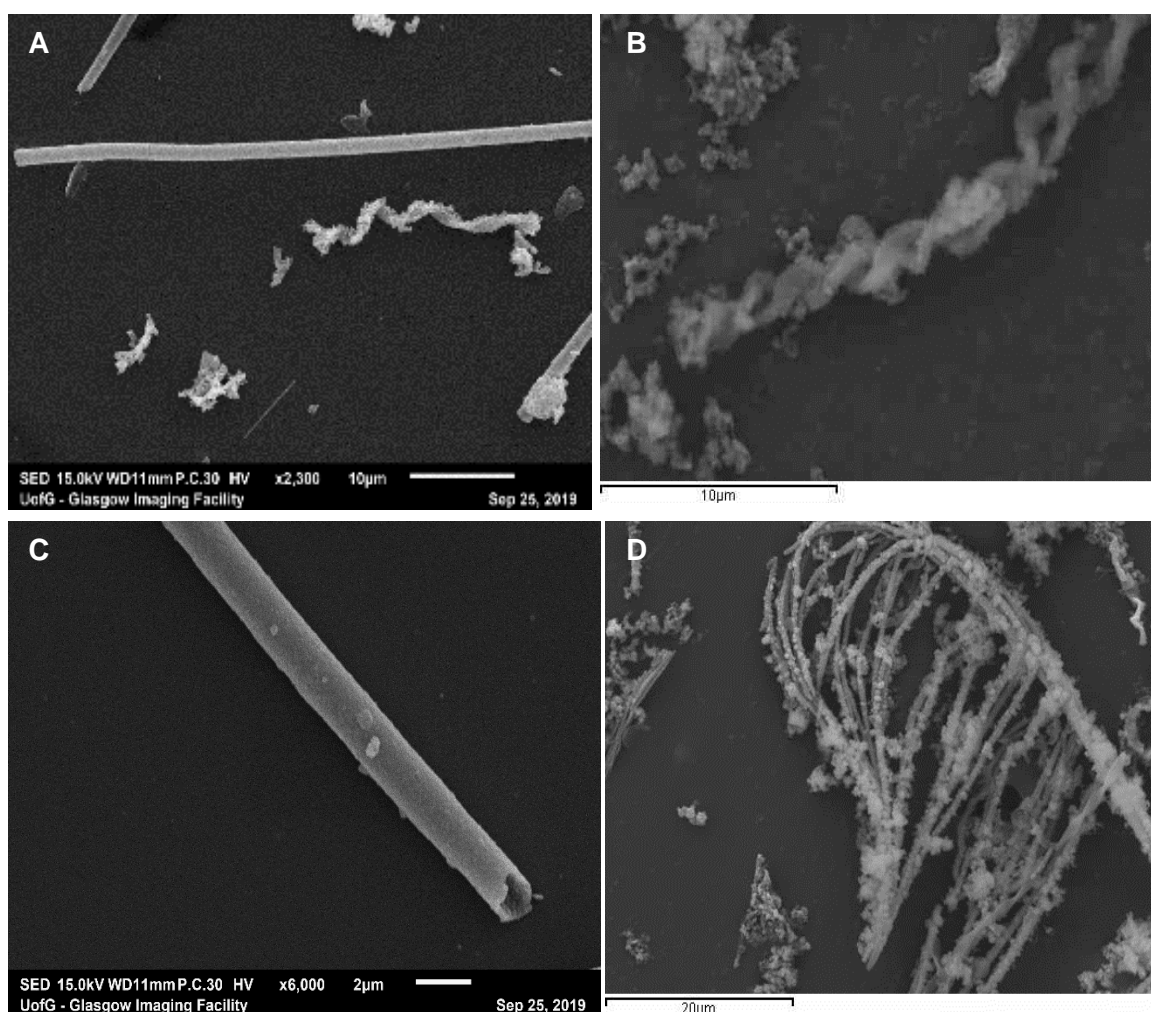
Figure 15 - SEM images of A - *L.ochracea* filament and *G.ferruginea* stalk, B - Higher magnification of a *G.ferruginea* stalk, C - Higher magnification of a *L.ochracea* filament and D - *Toxothrix* trichome.

Both *L.ochracea* filaments and *G.ferruginea* are typically ~ 1 - 2 μm in diameter and tens to hundreds of microns in length. These microstructures are not smooth but are highly porous with typical surface areas ~ 280 m^2/g and can be thought of as an inorganic/organic hybrid consisting of iron oxide/oxyhydroxide nanoparticles precipitated on an organic backbone.[10] To produce these microstructures, FeOB release EPS into the environment. When characterised, these secretions contain carboxylic acid, aldehyde, amino and occasionally sulfhydryl functionalities which chelate aqueous ferrous iron and allow a site of precipitation for insoluble iron oxides/oxyhydroxides.[11, 12] Environmental inorganics such as silicates and phosphates may also be included in the microstructure in lesser quantities (Table 10) making it a composite of the aforementioned compounds.

Element	EDX Relative Quantity
Fe	80
Si	15
P	5

Table 10 – EDX relative elemental quantities of BIOX, not including carbon or oxygen, measured by Hashimoto *et al.*[13]

The combination of structure, composition, porosity and surface area makes BIOX desirable for applications such as catalysis, where it could be used as a pre-catalyst for the cracking of methane or as a catalyst for organic reactions such as the Baeyer-Villiger oxidation.[14, 15] Hashimoto *et al.* have shown that it is possible to reduce the iron oxyhydroxide to metallic iron and remove it from the structure by acid leaching yielding a skeleton of high surface area (500 m^2/g) silica when can be used as a catalyst support.[16] As iron oxides/oxyhydroxides have tuneable magnetic properties and are non-toxic *in vivo* there is the potential for BIOX to be used for targeted drug delivery and biosensing. Finally, BIOX may also be used in water purification as iron oxides/oxyhydroxides have been shown to adsorb phosphates, lead, nickel, arsenic and palladium amongst other water borne pollutants.[17, 18]

Takeda *et al.* have characterised the EPS sheath material of *Leptothrix cholodnii*, an isolated cultured relative of *L.ochracea*, by NMR and proposed both a repeat and method of elongation of this sheath. Figure 16 shows their proposed repeat unit and method of elongation along with SEM micrographs of mats of *L.cholodnii* filaments with chains of cells inside.[19-21]
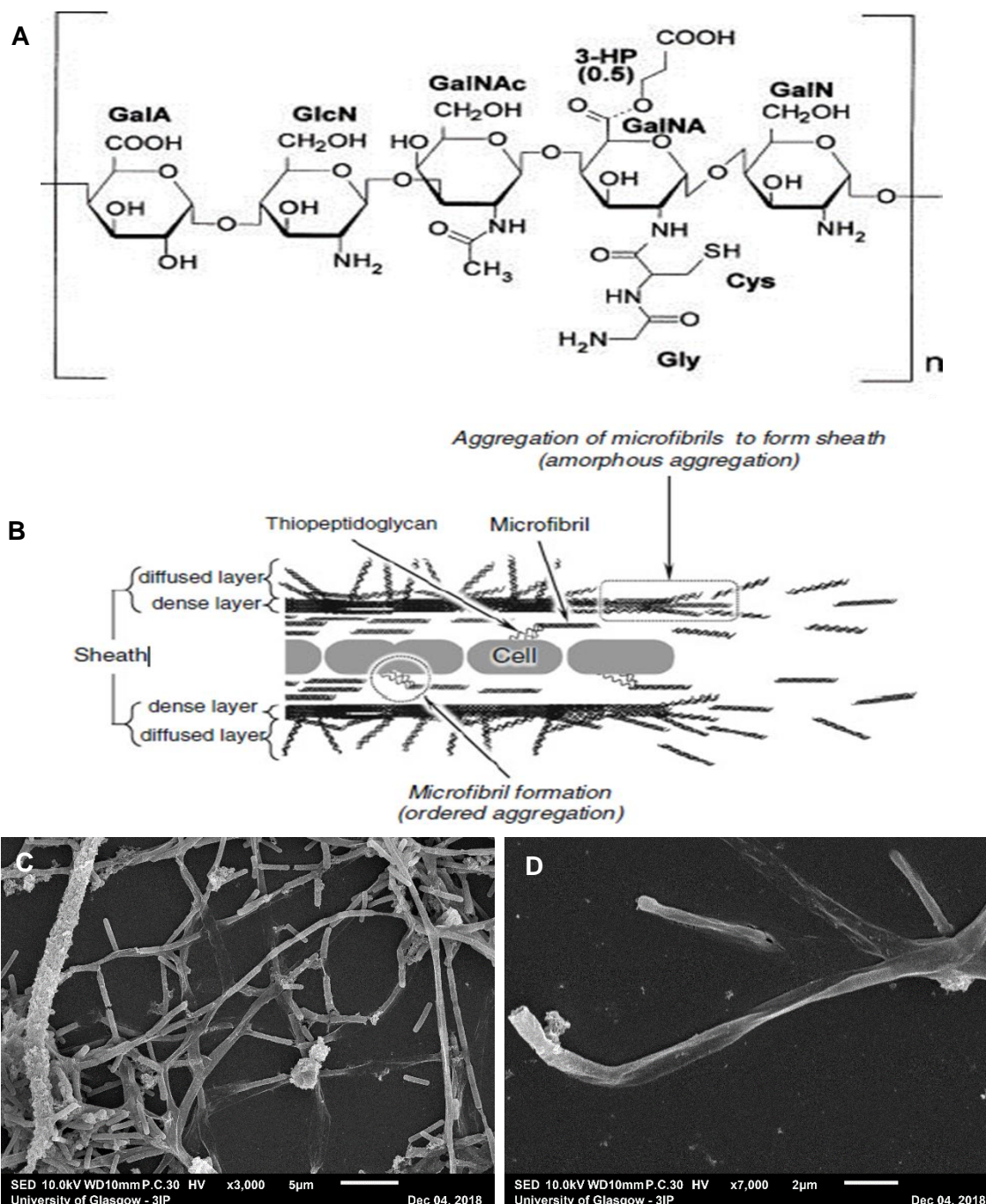
Figure 16 - A - Proposed subunit of *L.cholodnii* eps sheath material, B - Proposed mechanism of sheath elongation. GalA = galaturonic acid, GlcN = glucosamine, GalNAc = N acetylated galactosamine, GalN = galactosamine, Cys = cysteine, Gly = glycine and 3-HP = hydroxypropionic acid, C and D – SEM micrographs of *L.cholodnii* mat and filament.[19, 21]

The above micrographs of *L.choldnii* show the filaments to be similar in size to *L.ochracea* filaments however visually quite different. Image D shows a filament that appears to have collapsed, likely due to the vacuum in the microscope chamber. The repeat unit is a pentasaccharide comprising functionalised amino sugars. Functionalisation includes a hydroxypropionic acid sidechain and dipeptide sidechain of glycine and cysteine. The hydroxypropionic acid sidechain likely promotes hydrophobic interactions whereas the

cysteine allows the formation of disulfide bonds which promotes structural integrity of the material. There is also an abundance of polar moieties including carbonxylic acids, amines and alcohols. These groups may allow inter and intra molecular hydrogen bonds to form which would also promote stability of the material as well as providing suitable chemical groups for the chelation of metal cations such as iron and manganese. *L.cholodnii* secretes fibrils of this pentasaccharide that are ~ 6.5 nm in diameter. These fibrils then aggregate to form what is described as a dense inner layer and diffuse outer layer of material around the cells. As chains of cells are continually moving and secreting this material a microtubular sheath is formed.[21] This type of study is possible for *L.cholodnii* as it has been isolated as an axenic culture. Takeda *et al.* were able to enzymatically remove all proteinaceous material and completely denature the micro and ultrastructure of the sheath with the confidence that the resulting material was EPS subunits produced by *L.cholodnii*.[21] Two studies, by Chan *et al.* and Furutani *et al.* respectively, have shown that environmental BIOX samples produced by *L.ochracea* also contain polysaccharides however no structural characterisation has been performed.[22, 23] Chan *et al.* used a combination of TEM, STXM and NEXAFS to show that BIOX are precipitated on acidic polysaccharides while Furutani *et al.* used high angle annular darkfield SEM to show that the BIOX sheath is an organic-inorganic composite.

3.3 Chapter Aims

The following work looks to characterise in detail the BIOX collected from sample sites in Scotland near and around Glasgow. Characterisation will include confirmation that they contain microtubular iron oxide filaments and to compare their inorganic composition with literature examples from other sites. Methods for the isolation and purification of EPS associated with BIOX was performed to permit structural analysis by NMR. Finally, the interaction of BIOX with thiol containing small molecules is investigated as a means of characterising the accessibility and nature of iron sites of BIOX.

3.4 Results and Discussion

Before BIOX can be functionalised or applied to materials science it must firstly be fully characterised. This characterisation should include phase determination by diffraction methods, microscopic imaging, accessibility of iron sites and characterisation of associated organic material. Samples were collected from sample sites in accordance

with the protocol in Chapter 2 Section 1. Once returned to the laboratory, samples were washed and initially characterised by XRD to determine the phase of material (Figure 17).

3.5 Initial XRD Characterisation
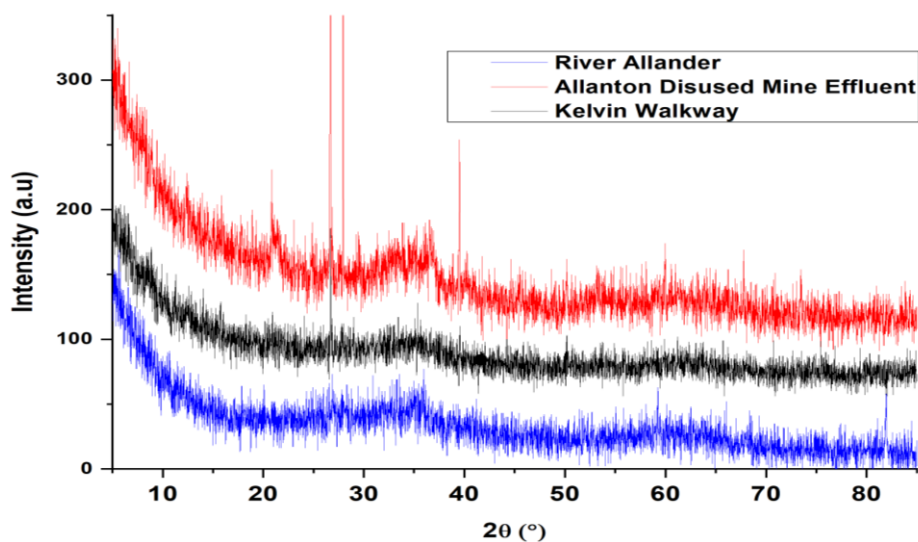
Figure 17 - XRD patterns of BIOX collected from three different sample sites.

All three samples characterised by XRD showed a high fluorescent background. Iron is known to fluoresce when irradiated with X-rays produced by a copper source meaning that this high background may be indicative of iron in the samples. A minimised fluorescence run was carried out in which the pulse height discriminator (PHD) lower threshold was lowered. This should filter out fluoresced X-rays as they will have a lower energy than the characteristic copper radiation however the samples still contained a high background. To produce ideal diffraction patterns the K_α wavelength of the X-ray source should be longer than the K absorption edge of the sample being analysed. Copper K_α X-ray radiation has a wavelength of 1.542 Å which is shorter than the iron K absorption edge of 1.743 Å. This causes fluorescent radiation to be produced by the iron. To overcome this, it would be ideal to use chromium as an X-ray source as it has a K_α wavelength of 2.291 Å.[24] Samples collected from the River Allander and Kelvin Walkway appear almost featureless with broad, shallow reflections around 30 – 40° 2θ and 55 – 70° 2θ. Broad reflections are typically indicative of amorphous materials. These reflections suggest that the samples contain predominantly 2-line ferrihydrite (2-Fh), an amorphous iron oxyhydroxide. 2-Fh shows broad reflections in these regions which can be attributed to the (110) and (330) crystal planes respectively.[25, 26] Sample collected from the runoff of a disused coal mine in Allanton showed sharper features than the other two patterns. Again, there are broad reflections indicative of 2-Fh, however the 30 – 40° feature appears

sharper. There is a reflection at ~ 22° 2θ which could not be assigned along with intense sharp reflections at ~26°, 28° and 41° 2θ respectively. It could be argued that the peak at 26° 2θ is from silica in the sample or perhaps the sample holder, however attempts at pattern matching have so for been unsuccessful. This peak may also be present in the Kelvin Walkway sample but is absent from the Allander sample. These three patterns agree with Takada *et al.* and Hashimoto *et al.* that filamentous BIOX contains predominantly 2-Fh.[7, 13]

3.6 SEM-EDX Analysis

Collected BIOX samples were next imaged and characterised by scanning electron microscopy coupled with energy dispersive X-rays (SEM-EDX) to confirm that all sample sites contained microstructures that could be attributed to FeOB (River Allander -Figure 18, Allanton mine effluent - Figure 19 and Kelvin Walkway - Figure 20).
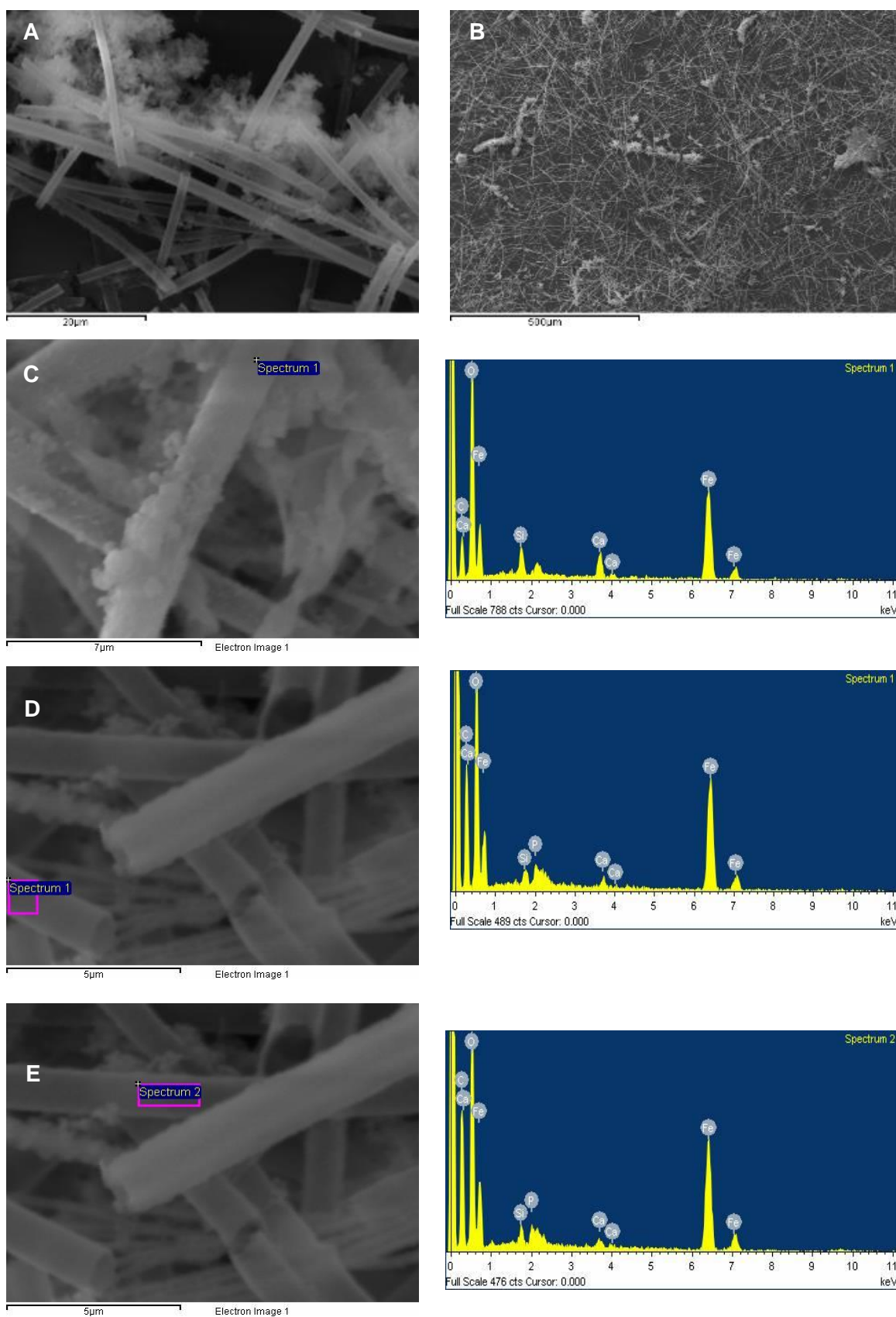
Figure 18 - SEM-EDX characterisation of samples collected from the River Allander. A – E – SEM micrographs with their corresponding EDX spectra.
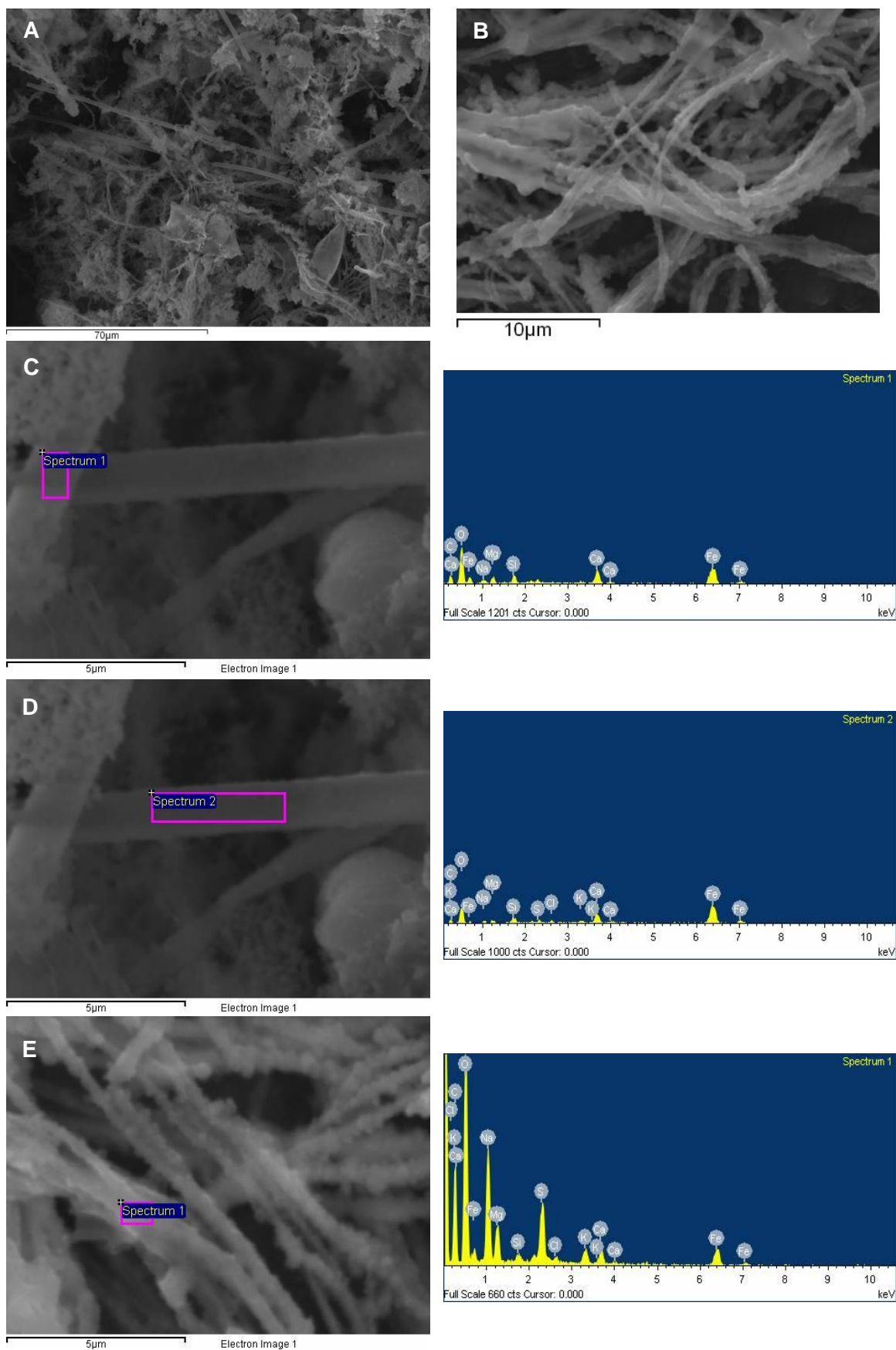
Figure 19 - SEM-EDX characterisation of samples collected from the effluent of a disused coalmine in Allanton. A – E – SEM micrographs with their corresponding EDX spectra.
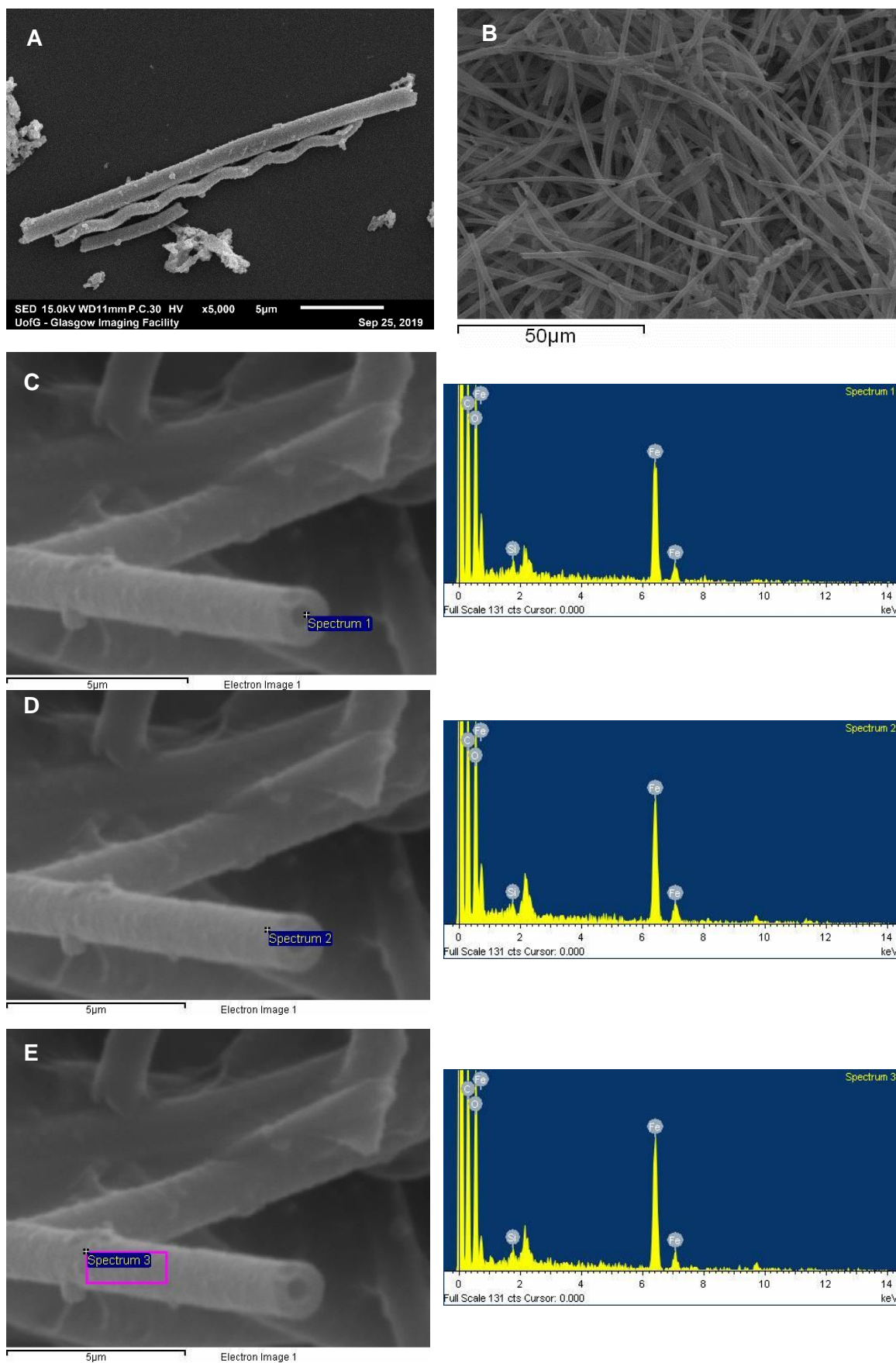
Figure 20 - SEM-EDX characterisation of samples collected from Kelvin Walkway. A – E – SEM micrographs with their corresponding EDX spectra.

The Kelvin Walkway sample was imaged twice. Once using unfixed samples on the School of Chemistry's SEM which has EDX capabilities whilst the other imaging used the Glasgow Imaging Facility (GIF) SEM which does not have EDX capabilities. The GIF samples were fixed and prepared in accordance with the protocol in Chapter 2 Section 2.1 then imaged at a lower energy which produced clearer images.

It can be seen from the previous micrographs that all three sample sites contain a mass of filamentous material and occasionally twisted stalks indicating the presence *L.ochracea* and *G.ferruginea* in these sites. The sample morphology from Allanton appears far less homogeneous than the River Allander and Kelvin Walkway sample sites. Morphologically this sample contains what appears to be mats of *L.ochracea* filaments, far finer filaments and twisted stalks of differing dimensions indicating that there is potentially a greater variety of neutrophilic FeOB in this site. These may include *Gallionella spp.* other than *G.ferruginea* and *Toxothrix spp.* resulting in the formation of intricate mats. This is a positive result as it may allow the isolation of bacteria that can produce a variety of microstructures. The isolation of microstructure forming bacteria will be discussed throughout Chapter 5.

EDX was used to give an indication of the elemental composition of the filaments found in each sample site (Table 11).

Average At %	Kelvin Walkway	Allanton Mine	Allander
Fe	3.8	6.3	5.0
	(2.6 – 4.8)	(5.1 – 6.9)	(4.2 – 6.1)
C	61.4	36.1	51.7
	(54.3 – 69.2)	(31.5 – 42.4)	(43.8 – 61.1)
O	34.5	51.2	42.2
	(27.4 – 40.1)	(40.8 – 57.3)	(37.1 – 48.4)
S	0.2	0.4	-
	(0.1 – 0.3)	(0.1 – 0.6)	
Si	0.1	0.6	0.5
	(0.1 – 0.2)	(0.3 – 0.8)	(0.2 – 0.6)
Na	-	2.4	-
		(1.0 – 2.8)	
K	-	0.2	-
		(0.1 – 0.4)	
Ca	-	1.1	0.5
		(0.6 – 1.5)	(0.2 – 0.7)
Mg	-	1.6	-
		(1.2 – 1.9)	
P	-	-	0.1
			(0.1 – 0.2)
Cl	-	0.1	-
		(0.1 – 0.2)	

Table 11 - Average EDX At % values for each sample. Upper and lower limits given in brackets.

This analysis shows carbon, oxygen and iron to be the dominant elements within the filaments found in the mats. This is consistent with the bulk material containing 2-Fh precipitated on an organic backbone. Trace amounts of other elements such as silicon, sodium, potassium, calcium, phosphorus, magnesium, chlorine and sulfur appear in the samples. Most of these elements are likely from salts in the water or sediment, see Chapter 6 for a discussion on the inorganic species present in sample sites, that have not been completely removed by washing prior to analysis. The silicon seen however is likely associated with the filaments, this may also be true for some of the phosphorus and sulfur seen and would agree with the results of Hashimoto *et al*. who showed that silicates and phosphates are incorporated into the structure of their BIOX samples.[13, 16] The Allanton mine sample showed the highest average At % of iron within the filaments. This is not surprising as coal mines contain large quantities of pyrite, an iron sulfide mineral. When pyrite is oxidised it releases Fe (II) and H_2SO_4 into the environment meaning that any effluent will be rich in dissolved ferrous iron that can be incorporated into the filaments.

The sample collected from Kelvin Walkway appeared to contain fewer contaminants than the other samples both visually and by EDX. The sample site itself is small and contains noticeably less sediment than the other sites meaning there is likely fewer sediment associated impurities than the other sample sites. This sample was also washed centrifugally by pelleting and resuspending in $_dH_2O$ rather than over a filter which may have helped remove more water-soluble impurities.

EDX element mapping was used to show the dispersion of the elements throughout the filaments. A selection of element maps can be seen in Figure 21.
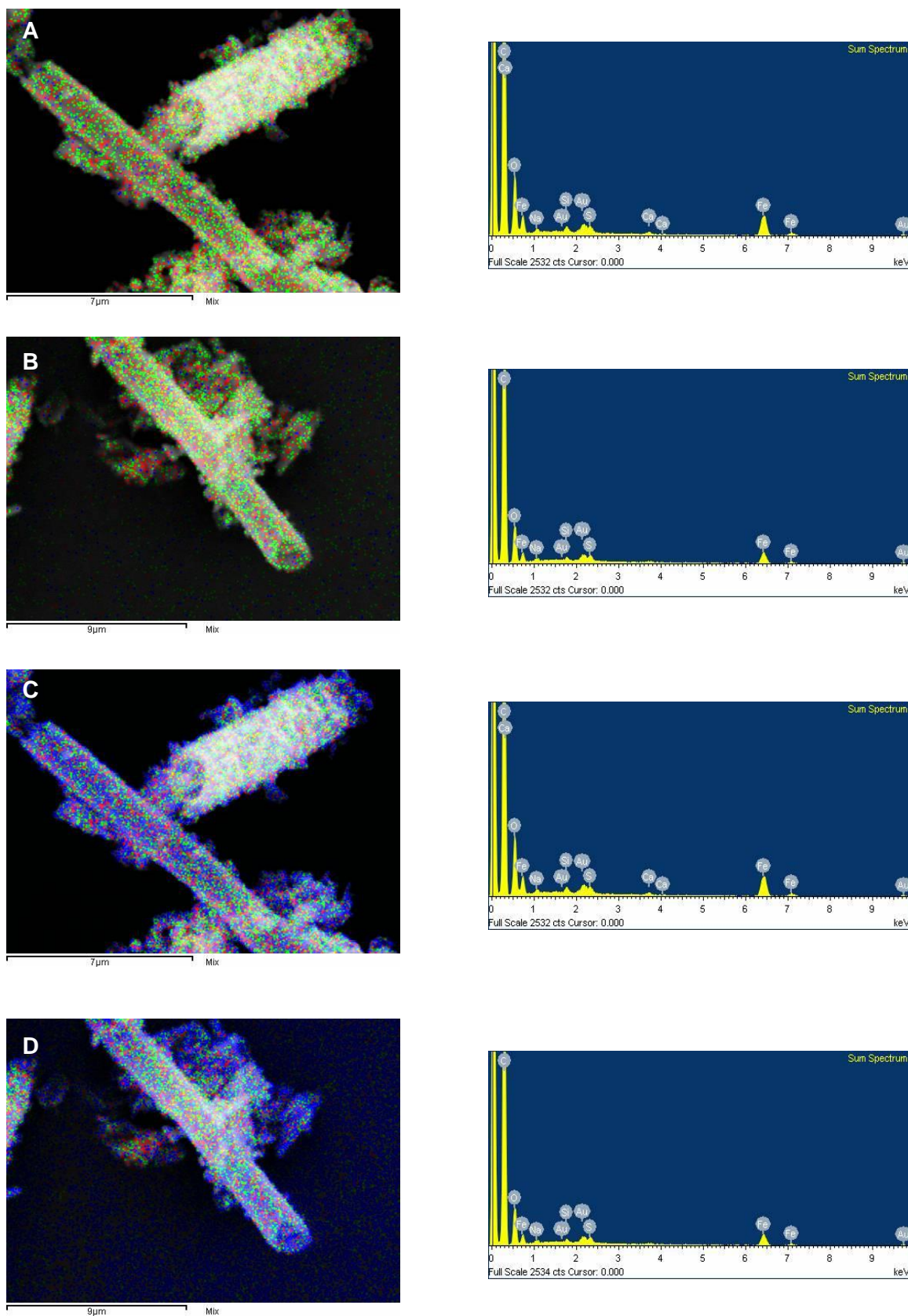
Figure 21 - EDX elemental mapping of filaments. A and B – Red = Iron, Green = Oxygen and Blue = Silicon, C and D – Red = Iron, Green = Oxygen and Blue = Carbon.

From the above element maps it can be seen that iron, oxygen, carbon and silicon are similarly dispersed throughout the filaments and there are no areas where any particular element is more concentrated than the others.

SEM was also used to image the surface film collected from Kelvin Walkway. This surface film is seen at all sample sites, is multicoloured like an oil slick but breaks into discrete fragments, and is typically found in areas surrounding BIOX clusters that indicate the presence of *Leptothrix spp*. Surface film from Kelvin Walkway was collected in accordance with the protocol in Chapter 2 Section 1 and imaged using the GIF SEM meaning that no EDX was carried out.
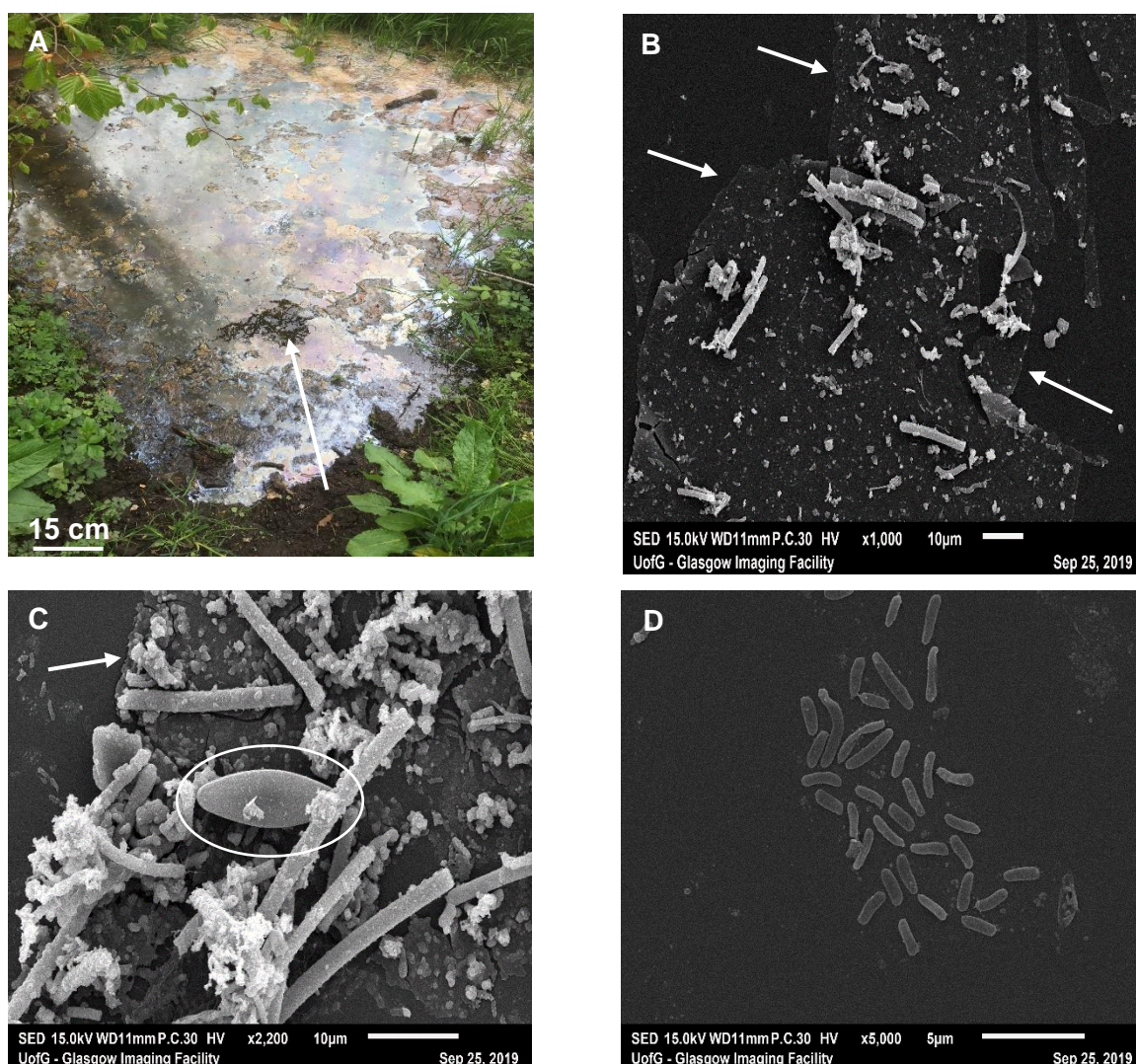
Figure 22 - A - Photograph of surface film (arrow indicates fragmentation) and B - D SEM images of surface film. Images B and C have arrows indicating the edge of the surface film. Image C also has a surface film associated diatom circled. Image D shows a higher magnification image of a cluster of cells on a fragment of surface film.

These SEM images show that the surface film has a variety of materials associated with it. Microtubular filaments can be seen throughout the surface film as can fluffy looking aggregates which are likely clusters of 2-Fh that could be either biogenically formed by FeOB other than *Leptothrix spp.* or abiogenically formed. Canoe shaped diatoms can also be seen associated with the surface film. These are unicellular microalgae with a cell wall made of silica. Diatoms are found in silicon rich water systems throughout the world so it is not surprising that they appear in the same environment as BIOX. Interestingly there is an abundance of cells seen throughout the surface film. This surface film may provide a protected mode of transport for these cells to migrate to different environments. During periods of rain this surface film is disrupted and washed to new areas which the associated bacteria may then be able to colonise. The surface film may also provide protection for sub surface bacteria from UV radiation and may help to keep the environment microoxic.

3.7 Development of a Purification Protocol of FeOB EPS For NMR

To characterise the EPS sheath material associated with BIOX by NMR it must firstly be disrupted and solubilised to yield individual EPS filaments. A variety of reducing agents including dithiothreitol (DTT), tris(2-carboxyethyl)phosphine (TCEP), *β*-mercaptoethanol (BME), sodium borohydride and sodium dithionite were initially added to BIOX samples to check for any visual degradation of material. The chemical structures of these reducing agents can be found in Figure 23 and the visual results of addition can be found in Table 12.
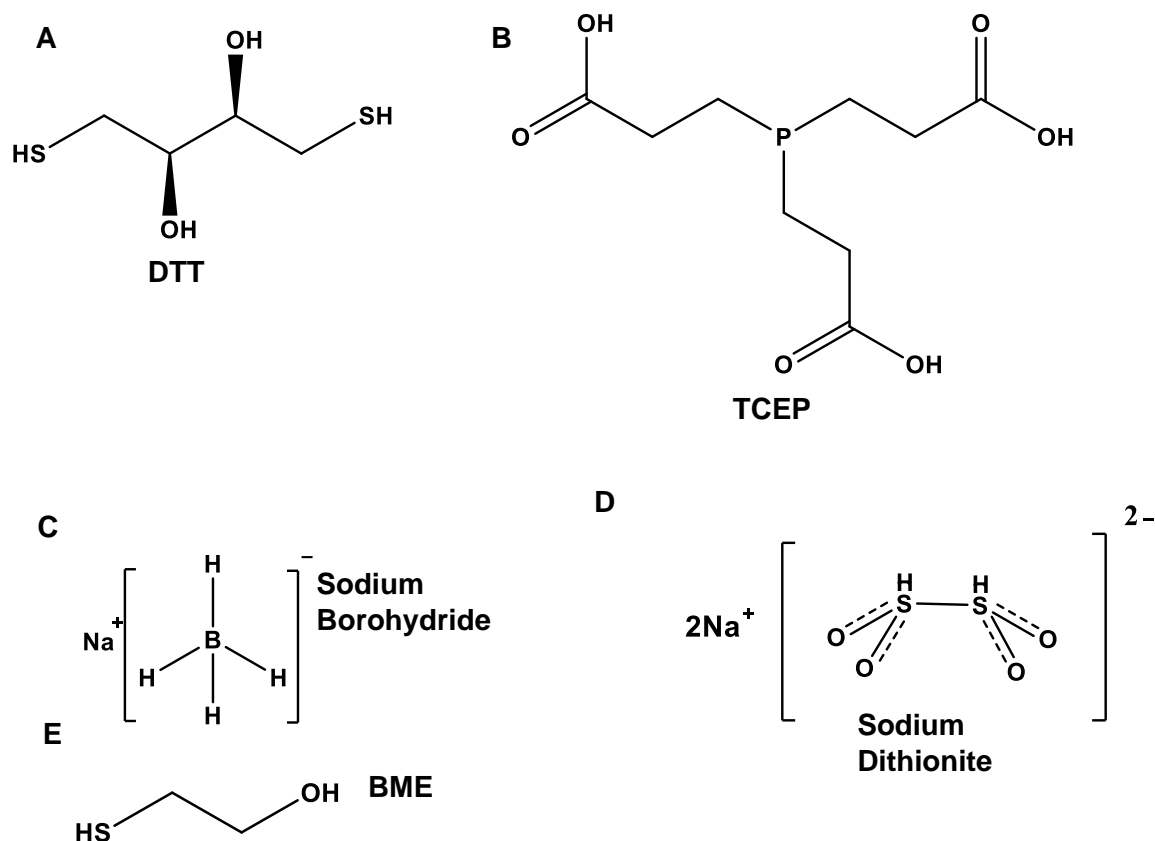
Figure 23 - Structures of reducing agents added to BIOX.

DTT, BME and TCEP were chosen as they are commonly used to reduce disulfide linkages within proteins. Emerson and Ghiorse have extensively studied *L.cholodnii* SP-6 and their research has shown that disulfide linkages provide structural integrity within the organic backbone of the microtubular sheath. They have also shown that the sheath material is resistant to a number of chemical denaturants and lytic enzymes including, sodium dodecyl sulfate (5 %), urea (6 M), guanidine.HCl (6 M), lysozyme (150 μg/ml), trypsin (1000 μg/ml) and hyaluronidase (250 μg/ml). Addition of disulfide cleaving reagents such as DTT (5 mM) however caused rapid degradation of the microtubular sheaths organic backbone.[27] Sodium dithionite was chosen as it is regularly used to solubilise iron oxide complexes (Equation 12) and is particularly effective at dissolving 2-Fh meaning it may be able to remove the 2-Fh from the organic backbone. Sodium dithionite also decomposes into a mixture of sulfur containing species such as hydrogen sulfite and thiosulfate (Equation 13). These species further react to forum polysulfides, hydrogen sulfide and elemental sulfur. This is ideal as these species can be removed by dialysis and do not interfere with ^1H or ^{13}C NMR spectra.[28, 29] Sodium borohydride was chosen as it is a strong general reducing agent.

$$2\ FeOOH + S_2O_4{}^{2-} + 4\ H^+ \longrightarrow 2\ Fe^{2+} + 2\ HSO_3{}^- + 2\ H_2O$$
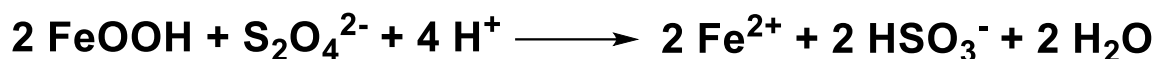
Equation 12 - Sodium dithionite reduction of insoluble ferric species to soluble ferrous species.

$$2\ S_2O_4{}^{2-} + H_2O \longrightarrow 2\ HSO_3{}^- + S_2O_3{}^{2-}$$
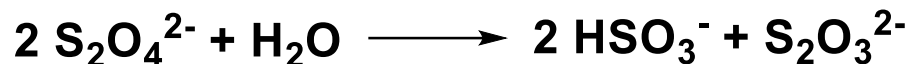
Equation 13 - Initial decomposition of sodium dithionite in H_2O.

Reducing Agent	Reduction Potential (pH 7)	Visual Result
DTT	- 0.33 V	Dark green aggregates formed
BME	- 0.26 V	Dark green aggregates formed
TCEP	- 0.29 V	No visual change
NaBH₄	- 0.75	No visual change
Sodium Dithionite	- 0.66 V	Material appeared to dissolve

Table 12 – Reducing agents added, their reduction potentials and the visual results produced when added.

Addition of TCEP and sodium borohydride yielded no visual change to the material, however addition of thiol containing reducing agents DTT and BME caused the formation of dark green aggregates. This observation has not previously been reported in the literature and is investigated further in Chapter 3 Section 7. Addition of sodium dithionite caused BIOX to dissolve, another observation which has not been reported in the literature. This observation is likely due to the strong reducing power (- 0.66 V) of this reagent and solubility of the iron-sulfate/sulfite species produced. The insoluble ferric iron that may be providing structural integrity is reduced to a soluble ferrous species thereby causing the microtubular sheath to rapidly degrade. This resulted in a clear solution with

only sedimentary material remaining visible. Figure 24 shows micrographs of BIOX filaments pre and post addition of sodium dithionite (25 mM). The filaments in image C post dithionite appear thinner with some appearing translucent when compared with the pre dithionite images. Image D shows a filament that has been incubated with dithionite for over 30 minutes and has started to fragment and dissolve. As only dithionite had the effect of solubilising the 2-Fh and dissolving the microstructure it was decided that a preparative method using this reagent should be developed as there were no literature examples to follow.
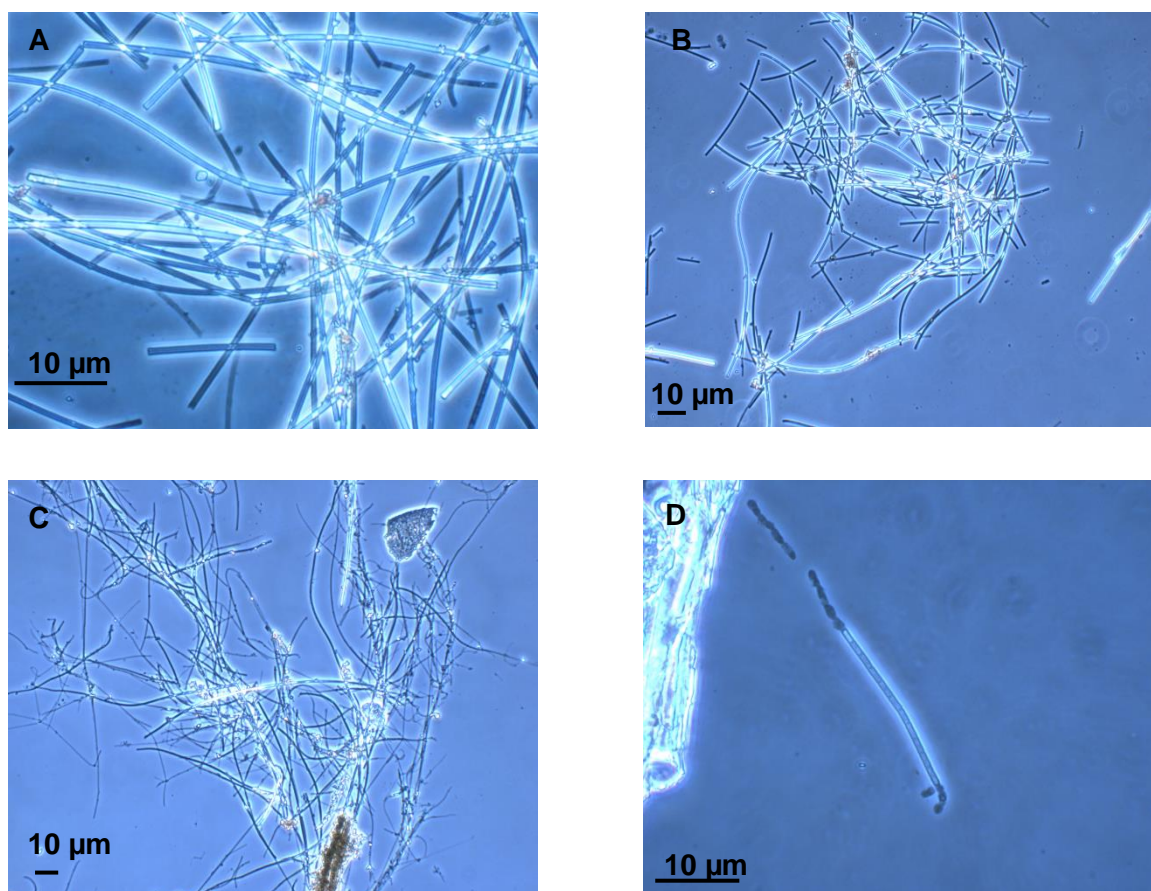
Figure 24 - Phase contrast micrographs of A and B - BIOX filaments pre dithionite (25 mM) addition and B and C - BIOX filaments post dithionite addition.

TLC analysis in a variety of solvent systems was carried out on the dithionite treated material to confirm that organic compounds were in solution (Table 13).

Solvent System	Spot Movement
EtOAc	-
3:1 EtOAc:Heptane	-
Heptane	-
MeOH	-
4:1 MeOH:H_2O	-
5:1:0.25 MeOH:H_2O:HOAc	+
2:1:1 ^{n}BuOH:H_2O:HOAc	+
10:1 AcN:H_2O	-

Table 13 - Solvent systems tried and result.

From these results it can be assumed that the material in solution is highly polar as there is only movement in a mixture of primary alcohol and water with a small amount of acetic acid. This type of solvent mixture appears in literature related to polysaccharide separation.[30, 31] Staining was carried out using potassium permanganate which visualises oxidised functionalities, ninhydrin which visualises amine functionalities and orcinol which visualises saccharides. A selection of these described TLC plates can be seen in Figure 25.
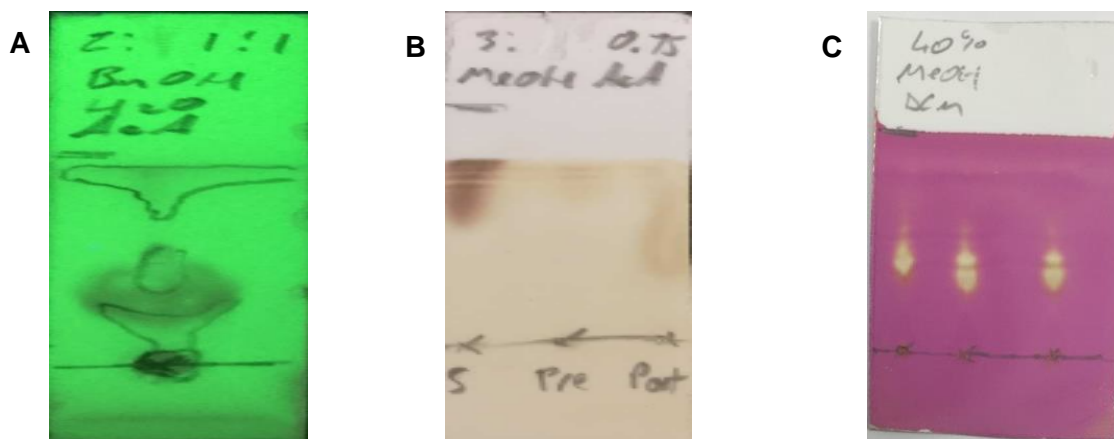
Figure 25 - Images of TLC plates. A (UV-irradiated as not stained) - BuOH:H$_2$O:HOAc mobile phase, B – Orcinol stained plate with MeOH:HOAc mobile phase (S = glucose standard, Pre = pre-reaction, post = post-reaction) and C – KMnO$_4$ stained plate with MeOH:DCM mobile phase. Spots can be seen on all three plates indicating the presence of organics.

Ninhydrin staining was negative indicating no amine functionalities are present. Permanganate and orcinol staining both gave positive results indicating the potential presence of saccharides with oxidisable functionalities however the orcinol staining of this compound was much weaker than the staining of a glucose standard (10 mM) indicating a low concentration of material. Several attempts were made at extracting this compound into a variety of polar protic, polar aprotic and non-polar organic solvents at a range of pH values however the material consistently remained in the aqueous phase. As the compound remained in the aqueous phase a protocol concluding with lyophilisation was developed.

The initial protocol involved collecting and washing BIOX extensively with $_d$H$_2$O and PBS (25 mM) over Whatman 1 filter paper. The filter cake was then washed with sodium dithionite (50 mM, 250 ml) with a visible loss to its orange colour, becoming darker brown and mud-like in appearance. The orange filtrate was collected, concentrated to 10 ml, dialysed against $_d$H$_2$O (1 l) for 18 h and lyophilised yielding an orange/brown solid. Two fundamental problems arose. Firstly, the solubility of the resulting material. As the sample was concentrated a precipitate began to form and the resulting orange/brown solid was found to be poorly soluble in all solvents tried, including H$_2$O, D2O, DMSO and DMSO + LiCl. LiCl was added to the DMSO as this mixture is used to solubilise lignin and other cellulose based polymeric substances.[32, 33] Secondly the resulting NMR spectrum showed broad indistinguishable peaks indicating that dialysis against $_d$H$_2$O was inefficient at completely removing paramagnetic iron from the sample.

To enhance the removal of the iron from the filtrate two approaches were investigated. Firstly, incubating the orange filtrate with a DOWEX™ MARATHON™ C (Na form) cation exchange resin for 18 h on a rotary shaker and secondly by dialysing against a 1 % solution of oxalic acid. A ferrozine assay was carried out, in accordance with the protocol in Chapter 2 Section 3.4, to monitor the total iron concentration throughout the treatments. Total iron concentrations decreased using both methodologies. Concentrations decreased from tens of millimolar to consistently around 10 µM for each 10 ml sample. It was not possible to remove all iron in solution. After this procedure the samples were lyophilised to dryness and the resulting NMR spectra can be seen in Figure 26.
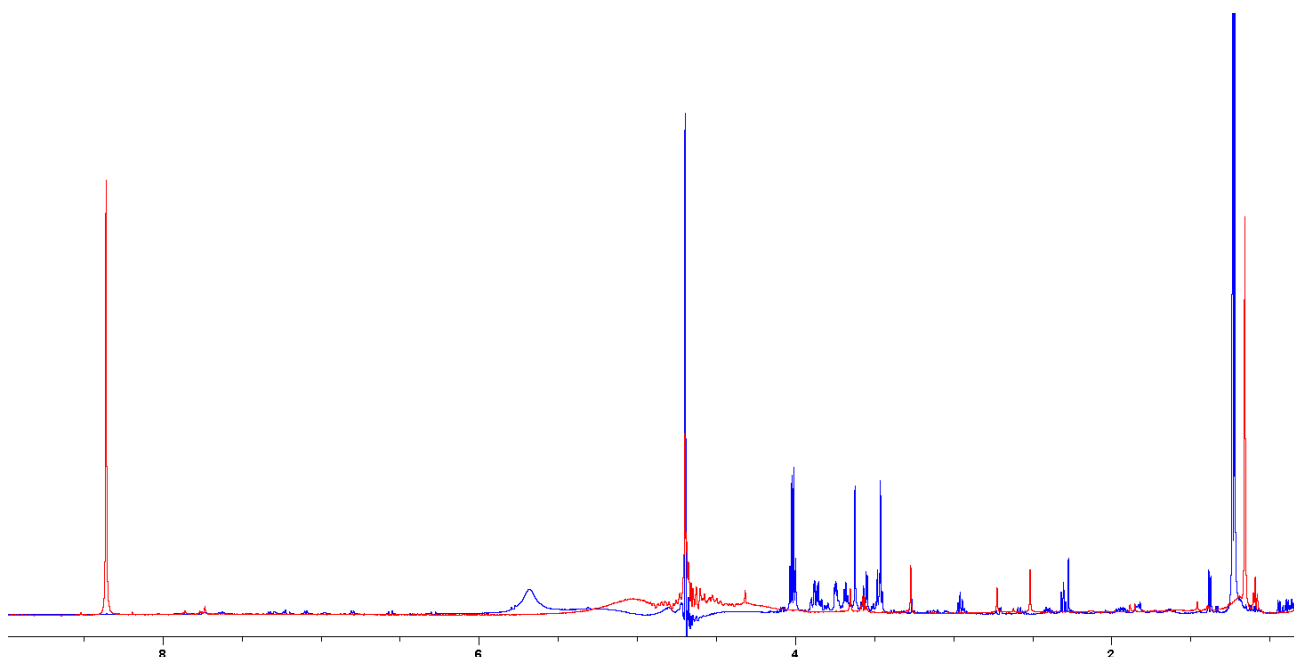
Figure 26 - Overlayed ^{1}H NMR spectra of DOWEX resin and oxalic acid treated samples. Red - DOWEX resin treated sample and Blue – Oxalic acid treated sample.

Both spectra contain sharper signals than the untreated sample (not shown) indicating that the concentration of iron in the samples is adequate, however both spectra appear different from each other suggesting that the procedures are not equivalent. The oxalic acid treated sample shows a higher abundance of intense signals in the 3 – 4 ppm range which is where saccharide signals are typically found. However, as the dialysis membrane is made from regenerated cellulose it is possible that the oxalic acid may have degraded it. This could produce glucose oligomers that would show signals in the 3 – 5 ppm range. The DOWEX cation exchange resin treated sample shows fewer signals. This could indicate that material is being lost on the resin via non-specific interactions. The intense singlet at 8.35 ppm is assigned to formate that has not been sufficiently removed from the resin during pre-treatment washing. The doublet at 1.23 and quartet at 4.02 ppm of the

oxalic acid treated sample are assigned to lactate. It is unclear where lactate contamination could come from. Both spectra also contain broad singlets. The broad singlet at 5 ppm of the DOWEX sample is likely from residual water however the broad singlet in the oxalic acid treated sample is at 5.7 ppm which may be indicative of an amide functionality. These experiments were repeated however the resulting spectra differed to those shown above. This is likely due to the heterogeneous nature of the starting material as it comes from an environmental source rather than a single bacterial culture. 2D HSQC NMR was also carried out on similarly treated samples, however the resulting spectra were inconclusive and provided no insight to any potential coupling that may be occurring.

While researching methods of biofilm characterisation it became apparent that BIOX may be similar in nature to activated sludge flocs found in effluent from wastewater plants as this material is a mixture of bacteria, EPS and inorganic material found within sewage systems and wastewater plants.[34, 35] EPS is typically extracted from activated sludge by incubating with NaOH. It was decided that extracting BIOX first with NaOH may act as a purification step by removing some of the biofilm that holds the mats of BIOX filaments together. This can then be followed with dithionite incubation of the purified filaments.

Extraction with NaOH produced a dark orange/brown solution that contained both Fe (II) and Fe (III) when assayed with ferrozine. This is surprising as at basic pH values iron typically exists as Fe (III). The remaining BIOX was then extracted with dithionite as previous also yielding a dark orange/brown solution however this contained predominately Fe (II) as expected. These were treated with a DOWEX resin to remove the iron, lyophilised and analysed by NMR. The NaOH extract was partially soluble in D_2O. The resulting precipitate was collected by centrifugation and dissolved in NaOH (10 mM) prepared in D_2O for NMR analysis. The resulting spectra can be seen in Figure 27.
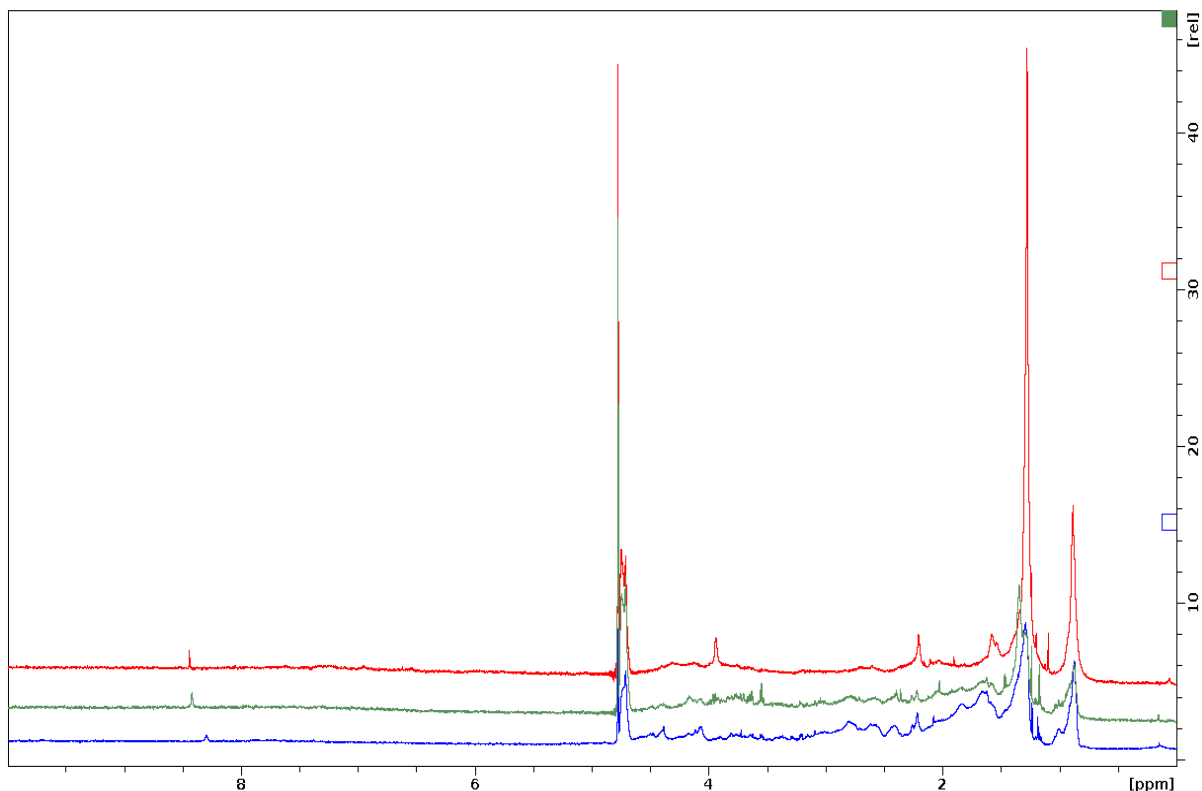
Figure 27 – ^{1}H NMR spectra of EPS extracts. Blue - Dithionite extract, Green - NaOH Extract and Red – NaOH dissolved material.

All three spectra contain broad, poorly resolved signals. This may either be indicative of insufficient paramagnetic iron removal or of potentially polymeric material. As polymeric molecules are larger and tumble more slowly in solution, they can produce broad poorly resolved signals. All three spectra have consistent signals at 4.75 ppm. These are likely from residual water in the samples. There are also signals at 0.8 and 1.4 ppm that are consistent in all three spectra. These lie in the region where greasy aliphatic compounds typically produce signals and have not been assigned. These spectra are inconsistent with previous results and the poor resolution makes assignment challenging.

As the original extraction experiments showed more signals when dialysed against oxalic acid rather than treating with a DOWEX resin a final set of extraction experiments were carried out with oxalic acid. Samples were extracted as previously, initially with NaOH followed by sodium dithionite. The resulting brown filtrates (15 ml) were then dialysed against oxalic acid (1 l, 1 %) for 18 h. The resulting dialysate was light brown while the colour of solution within the membrane was paler. These samples were lyophilised and dissolved in D_2O. There was again insoluble material from the NaOH extraction that could be dissolved in NaOH (10 mM) prepared in D_2O. A blank extraction of H_2O dialysed

against oxalic acid was also carried out. Figure 28 contains the ^{1}H NMR spectra for these extractions.
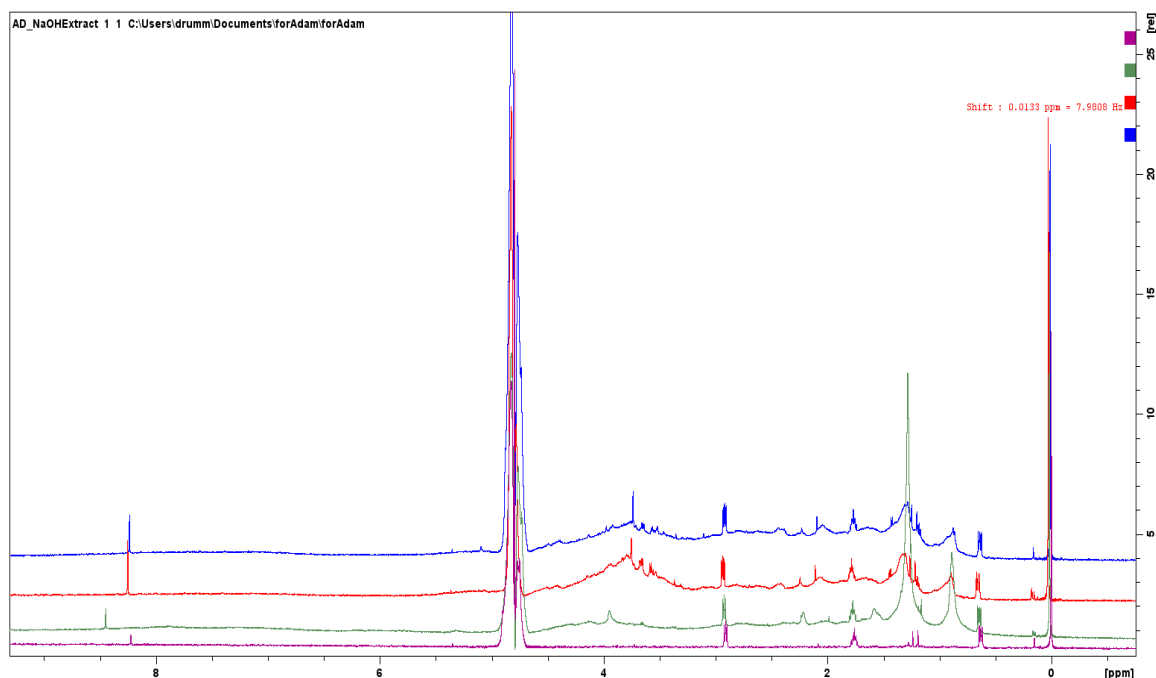
Figure 28 - ^{1}H NMR spectra of EPS extracts. Blue – NaOH extract, Red – sodium dithionite extract, Green – NaOH dissolved material and Magenta – Oxalic acid blank.

The resulting spectra are again inconsistent with previous examples. There appears to be a similar profile between the three extracts, especially the NaOH and sodium dithionite extracts however it is not possible to make any assignments nor comment on the composition of the samples. It is therefore reasonable to say that these extracts contain a complex mixture of organic molecules.

This work has highlighted the difficulties of NMR characterisation of environmental biofilms that are associated with paramagnetic compounds. Takeda *et al.* encountered similar solubility issues when characterising the *L.cholodnii* EPS sheath.[19] To resolve their solubility issues they enzymatically removed all proteinaceous and lipid material from the EPS, denatured the resulting material with TCEP (10 mM) and then either alkylated free sulfhydryls or oxidised them and acetlylated the resulting sulfonates. A similar strategy may need to be applied here, however this would be aided significantly if an axenic culture of *L.ochracea* was produced. Further work is also required at this stage to optimise the extraction protocol and a more thorough sampling protocol must be developed to ensure that each BIOX samples being analysed are as consistent as possible.

3.8 BIOX-Thiol Interaction

Iron-thiol complexes are common in the literature. Examples include iron-sulfur clusters which are essential cofactors of proteins, iron oxide nanoparticle functionalisation by thiol containing ligands and heme iron interactions with thiol containing ligands such as cysteine.[36-38] Coloured complexes of aqueous ferric iron and cysteine have been reported in the literature for over 100 years.[39] Blue complexes have been noted to form under acidic conditions, pH < 1.8, and red complexes noted to form at pH 7.4. These complexes form instantly and degrade rapidly. The blue complex has been attributed to an FeRSH containing compound whereas the red complex has been attributed to an $Fe(RSH)_3$ containing compound whose extinction coefficient and absorption spectrum is analogous to that of ferric isothiocyante.[40] There is also some literature investigating the phase transformation of 2-Fh in the presence of cysteine.[41, 42] This research carried out by Cornell and Schneider has shown that varying the concentration of cysteine promotes the formation of haematite, lepidocrocite and goethite from ferrihydrite at pH values of 6 – 9. At the time of Cornell and Schneider's publication, cysteine was the only known organic ligand to promote the formation of goethite from ferrihydrite, with all other previously studied organic ligands suppressing this transformation.[41, 42] The influence of organic ligands on iron oxide phase transformations has been more recently studied by Colombo *et al.* and Violante *et al.*[43, 44] Both studies have shown that the presence of organic molecules in soils such as humic and fulvic acids promotes the formation of ferrihydrite and inhibits the formation of more crystalline phases such as goethite and haematite. Interestingly, none of the searched literature returned any information on green iron oxide thiol complexes or more specifically green ferrihydrite thiol complexes, nor did any of the literature provide images or descriptions to match what has been described in this study.

3.8.1 Initial Complexation Reaction and Characterisation

As mentioned in 2.3, addition of thiol containing reducing agents namely, DTT and BME to suspensions of BIOX caused the samples aggregated and turned dark green. The Cambridge Structural Database (CSD) was searched for compounds with containing an iron-sulfur bond and dark green colour. This returned only a few results indicating that this is an unusual observation. Further experimentation was required to characterise the chemical nature of this colour change. To investigate this reaction further, a variety of experimental conditions were tried including altering the pH, the thiol used, the solvent and comparing BIOX to different phases of iron oxide. Figure 29 shows a 30 minute time

lapse of the reaction between a suspension of BIOX and DTT followed by a selection of light micrographs pre and post reaction. These images clearly show the colour change of the sample from pale orange to dark green along with the formation of an amorphous looking aggregation adhering to the mats of BIOX sheaths. Identical results were seen upon addition of BME to BIOX suspensions.
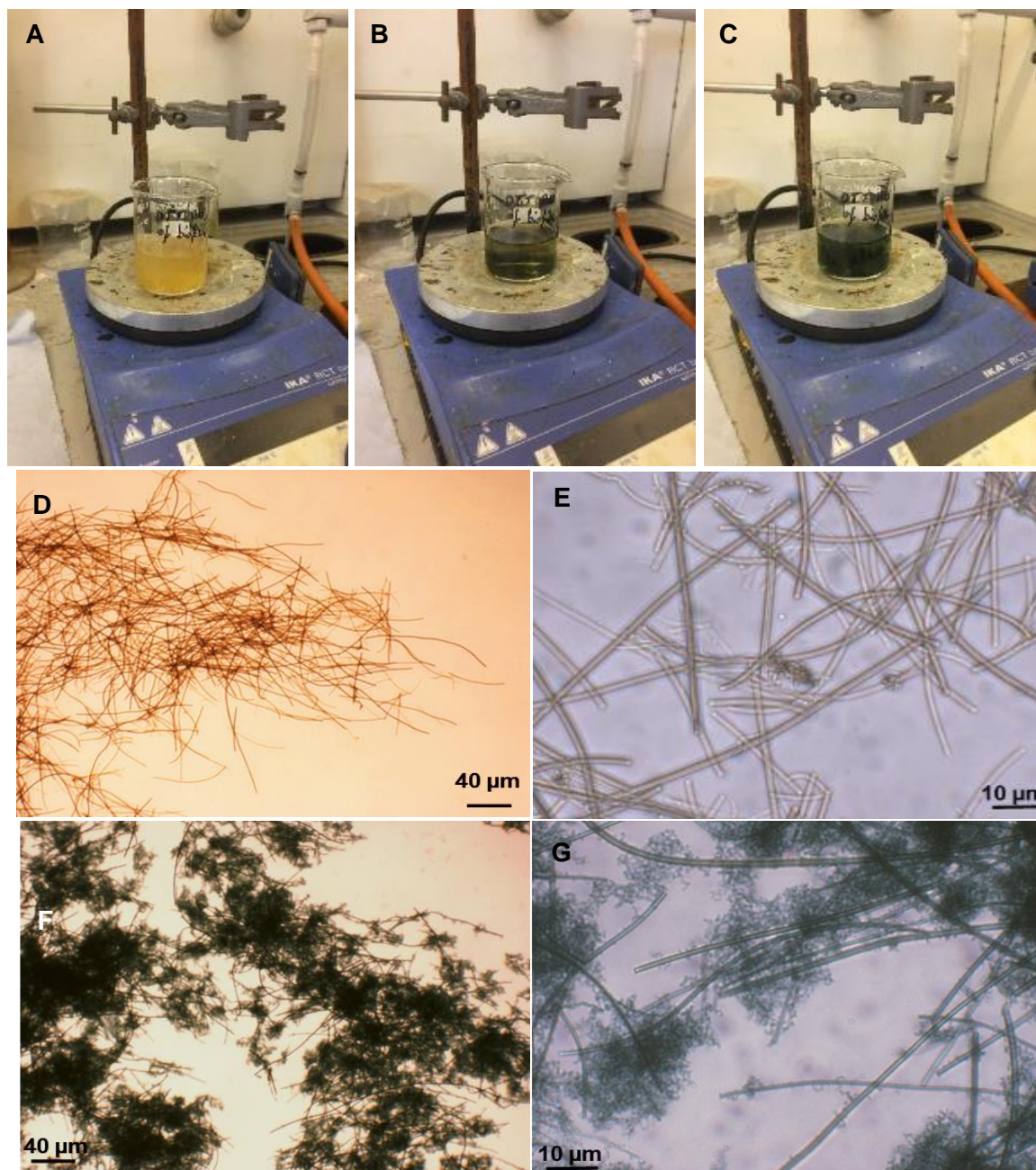
Figure 29 - A – C – BIOX + DTT over 30 mins, D & E – Pre reaction light micrographs and F & G – Post reaction light micrographs.

Dark green aggregated samples were then imaged and analysed by SEM-EDX (Figure 30).
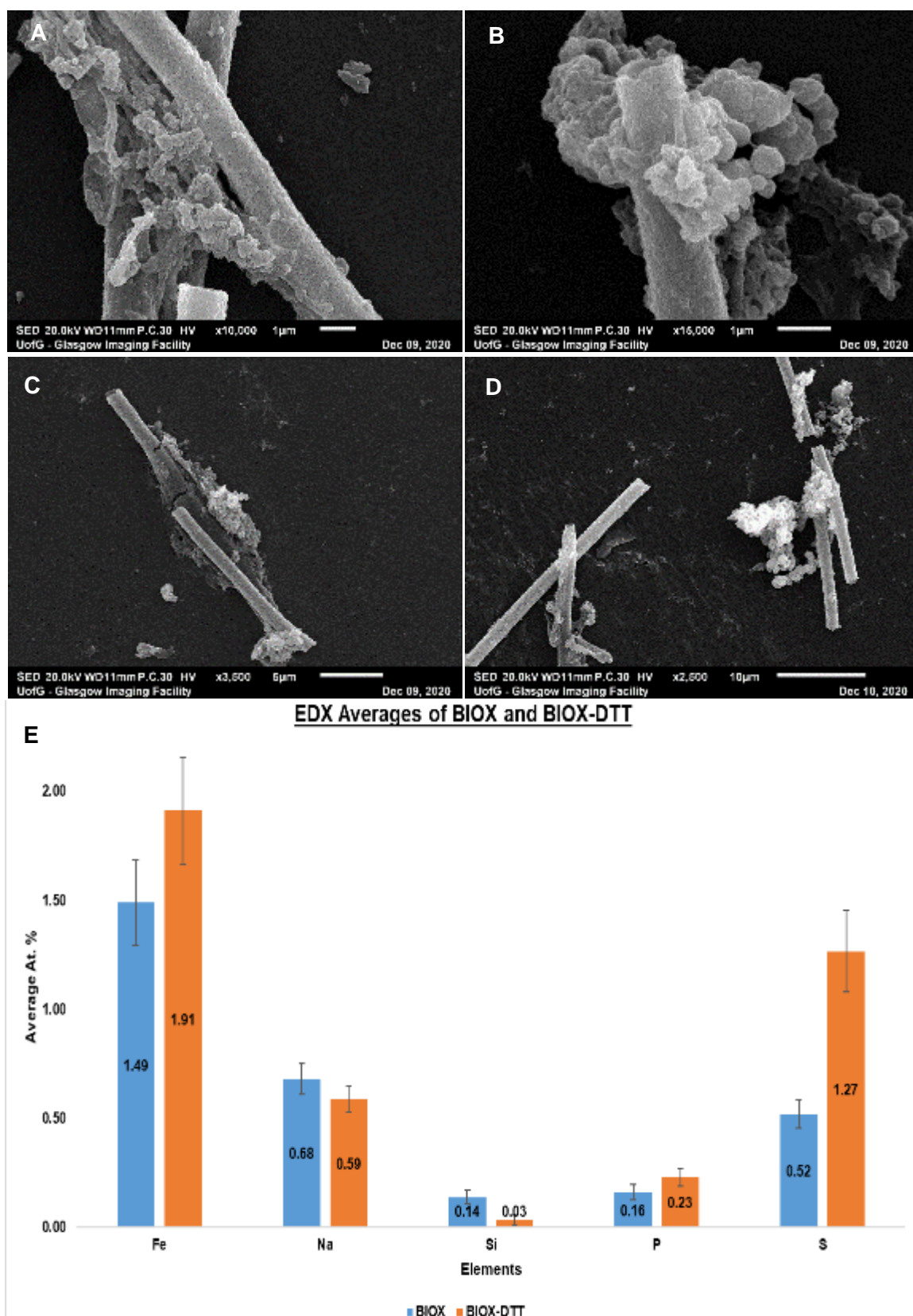
Figure 30 – A – D – SEM images of BIOX-DTT complex at different magnifications and E – A comparison of the EDX analysis of BIOX and BIOX-DTT complex. EDX was carried out at several points to make the data statistically representative. Carbon and oxygen data is not shown.

The SEM images of BIOX-DTT appear different when compared with earlier SEM images of native BIOX. Aggregation can now be seen, and sheaths look to have an aggregated material adhered to them. The EDX analysis has had the carbon and oxygen data removed as the samples are mounted on graphite tape, also EDX has limitations when quantifying lighter elements, making these data potentially less reliable. Figure 30 shows that the proportion of sulfur of BIOX has increased by 144% indicating that the DTT has been incorporated into the material. It is therefore likely that complexation has occurred.

A dilution of the BIOX-DTT material was suitable for spectroscopic analysis. The UV-Vis and FTIR spectra of the green complex are shown in Figure 31.
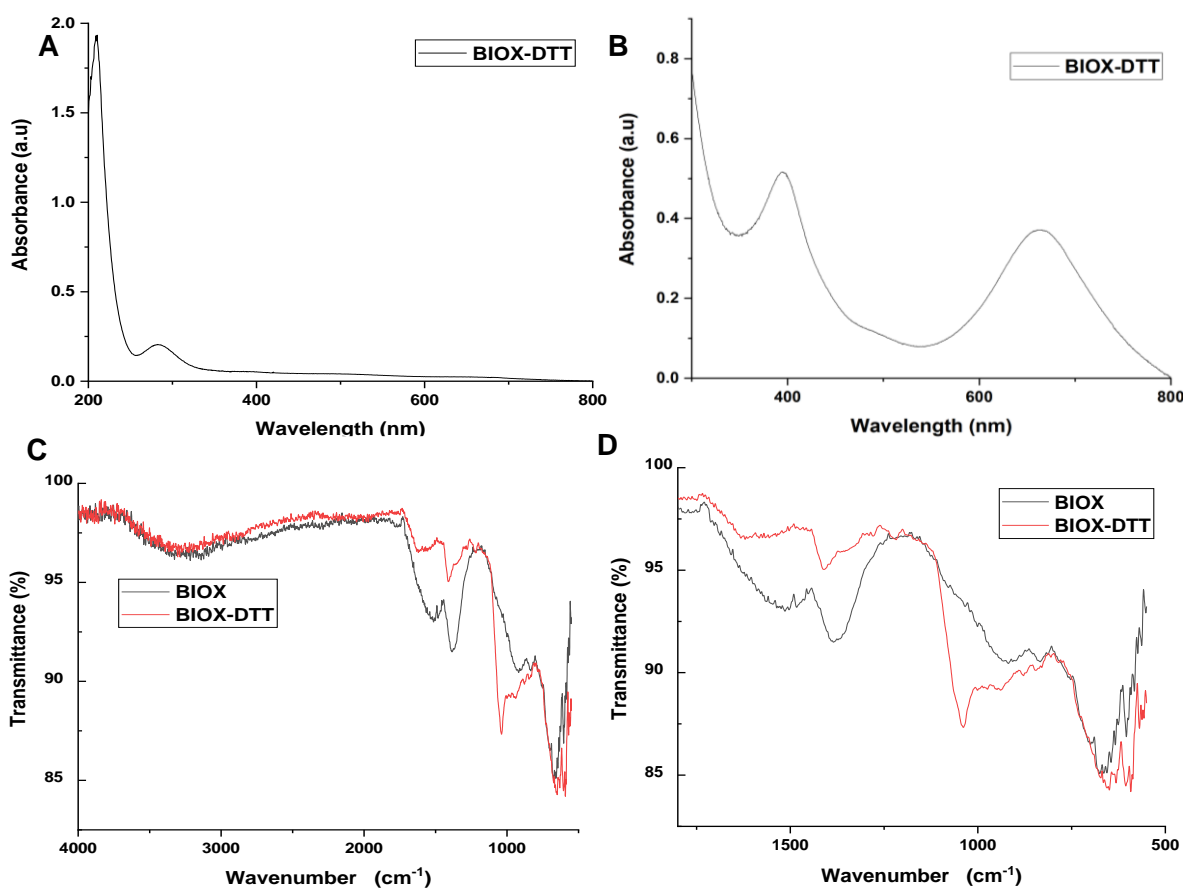
Figure 31 - A- UV-Vis spectrum of BIOX-DTT, B - Expanded UV-Vis spectrum of more concentrated BIOX-DTT, C – overlayed FTIR spectra of BIOX and BIOX-DTT and D – Magnified overlayed FTIR spectra of BIOX and BIOX-DTT.

The A_{284} nm absorption in the UV-Vis spectrum corresponds to the S-S bond that forms when DTT cyclises upon oxidation. The expanded visible spectrum shows absorptions at 395 and 665 nm respectively. These absorptions in the visible region account were initially weak and difficult to discern. A more concentrated sample was prepared to in order to

visualise these absorptions. This absorption spectrum is typical of a dark green material as they tend to absorb at ca. 400 nm and ca. 700 nm.

DTT is known to have a S-H stretching frequency at 2545 cm^{-1} in its FTIR spectrum which is not seen in the BIOX-DTT spectrum indicating that DTT may be bound through deprotonated thiols.[45] BIOX has stretching frequencies from ca.3600 – 3000 cm^{-1} due to surface and bulk OH moieties and a stretch at 625 cm^{-1} due to bulk OH deformations. The broadness of the surface and bulk OH band is another indicator of the 2-line ferrihydrite phase. Ferrihydrite doped with Si shows a stretch at ca.940 cm^{-1} which is seen in the BIOX sample. This is in agreement with the EDX analysis which confirmed the presence of Si.[25] The overlayed BIOX and BIOX-DTT spectra show mostly similar stretches with the main differences being a sharp stretch at 1040 cm^{-1} for the BIOX-DTT sample.

XRD analysis of BIOX-DTT was carried out however the resulting pattern contained only one large amorphous band, from 20 – 40 ° 2θ, that did not provide any further insight to the complexation.

3.8.2 Alternative Reaction Conditions

3.8.2.1 pH Dependence

DTT contains two thiol moieties with pKa values of 9.2 and 10.1 respectively and it is recommended to basify the sample to ca. pH 9 in order to provide optimal reducing power.[46]

The previously discussed results were achieved using neat solutions of DTT which has a pH value of ca. 7.7. This was adjusted to 9.5 and mixed with suspensions of BIOX. The result of this was the formation of a dark red solution with less aggregation than the dark green sample. The sheaths also appear microscopically slighter than both fresh BIOX and pH 7.7 BIOX-DTT (Figure 32). Interestingly when filtered the filtrate remained dark red whereas the filtrate of the neutral reaction is clear indicating that the dark red complex is in solution.

At this higher pH value a greater proportion of the thiols will be deprotonated making the DTT solution more reactive. This may lead to stronger complexation and the formation of a similar type of compound as the previously mentioned red ferric cysteine complex isolated by Page.[40] The UV-Vis spectrum seen in Figure 32 shows the expected oxidised

DTT S-S bond A_{284} nm along with a broad A_{480} nm. Absorbance in this visible region is typical of red iron thiocyanate species.[39, 40] There is also a shoulder at ca. A_{230} nm which is not seen in the spectrum of the green pH 7.7 BIOX-DTT complex. Attempts were made at crystallising solutions of the dark red complex. Firstly the aqueous solution was extracted with ethyl acetate, dichloromethane, chloroform and diethylether however the red complex remained in the aqueous phase. Cetyltrimethylammonium (CTMA) bromide (3 mM) dissolved in chloroform has been used in the literature to extract iron isothiocyanate from the aqueous phase.[47] This methodology was applied here and the red complex was successfully extracted to the aqueous phase. This was left to crystallize by slow evaporation however yielded only a red/brown gum. Further attempts were made to crystallise this material by adding ethyl acetate, heptane and hexane as co-solvents however all attempts yielded only a dark red/brown gum.
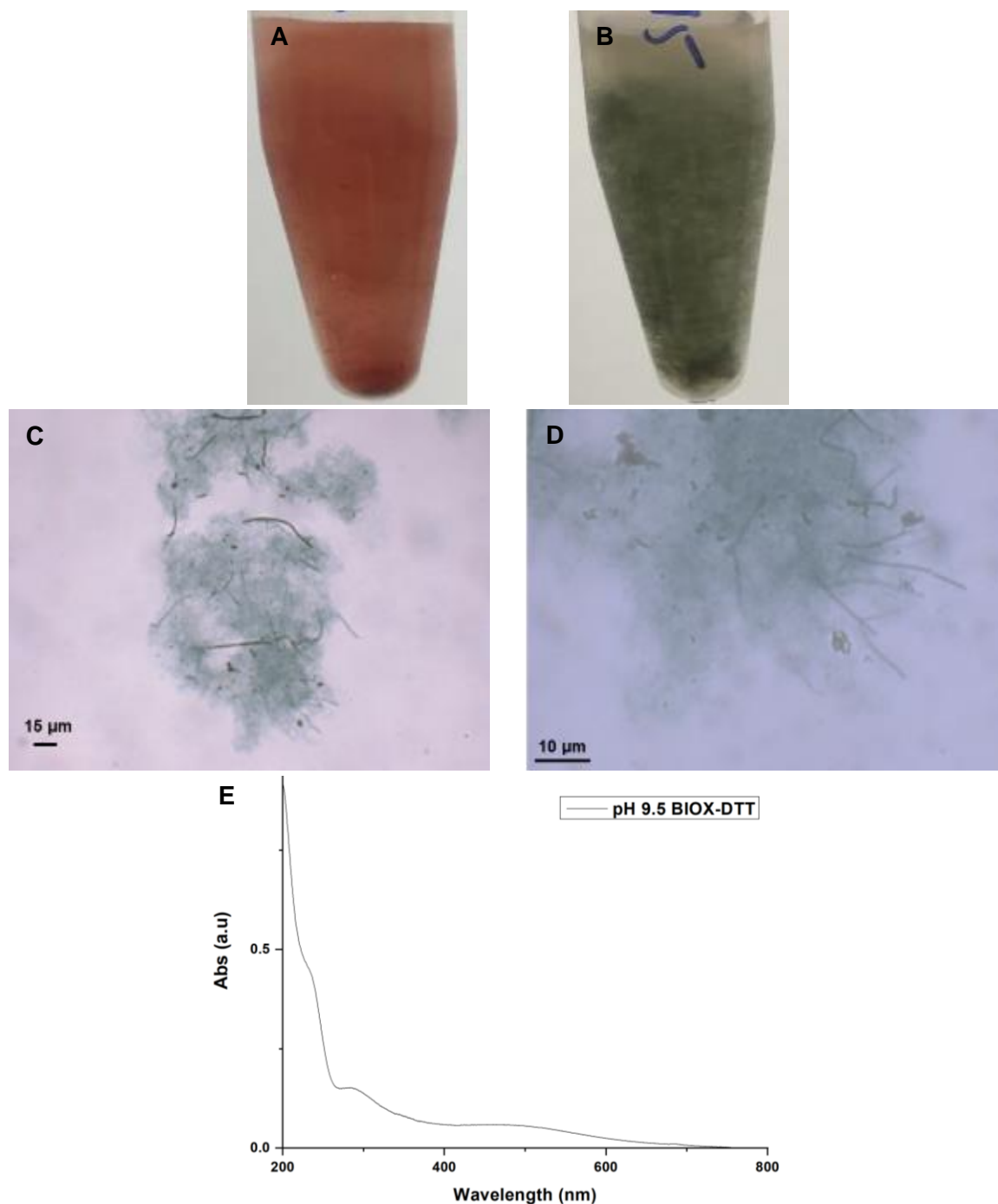
Figure 32 - Images of A - pH 9.5, B pH 7.7 BIOX-DTT, C & D – Light micrographs of pH 9.5 BIOX-DTT and E – UV-Vis spectrum of pH 9.5 BIOX-DTT.

3.8.3 Thiol Dependence

To assess the potential dependency of the formation of green aggregates on the type of thiol used, a selection of chemically distinct thiols (25 mM) (Table 14) were chosen and used to form suspensions of BIOX.

Surprisingly, addition of thiol containing molecules other than DTT or BME failed to produce the instantaneous green aggregation seen previously. This may indicate that steric effects play a role in the complexation. The addition of cysteine eventually yielded a dark green colour however this required a far higher concentration than the 25 mM needed for DTT and was also not rapidly produced. Instead, the sample slowly turned dark green over a period of days and the BIOX sheaths appeared to be structurally damaged and partially dissolved when observed by microscopy.

Initially, it was assumed that steric hinderance of the thiols may be a factor in the colour change. When considering the thiols tested it can be seen that neither primary nor secondary thiols form the dark green coloured complex. Moreover, in the case of the primary thiols we see no differing effect between cysteine and homocysteine which is less sterically hindered.

Considering the structure of both thiols that do spontaneously produce the green colour change, DTT and BME, it can be seen that they both contain a hydroxyl group on the second carbon from the thiol. This result suggests that this is an important group and is the only other group that may contribute to the change. It is plausible that both the thiol and hydroxyl groups may be involved in the complexation of iron within the BIOX.

Thiol	Result
Cysteine	No green aggregates formed.
Homocysteine	No green aggregates formed
Glutathione	No green aggregates formed
Tiopronin	No green aggregates formed
2-Propanethiol	No green aggregates formed
Triethylene glycol mono-11-mercaptoundecyl ether	No green aggregates formed
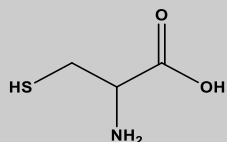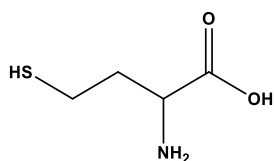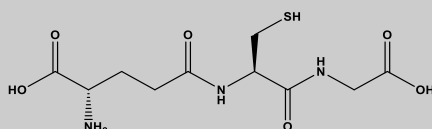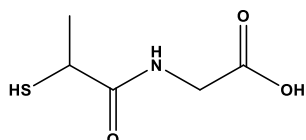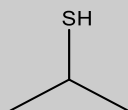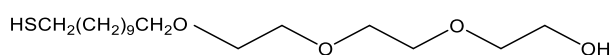
Table 14 - Structures of thiols used and a brief description of visual appearance after addition.

3.8.4 Solvent Dependence

BIOX was dried at 100 °C for three days prior to solvent experiments to remove residual water than may contribute to the reaction. A variety of solvents (Table 15) were used to dissolve DTT and suspend the dried samples.

Solvent	Result
H_2O	
Ethanol	
Methanol	
Acetone	
Acetonitrile	
Chloroform	

Table 15 - Solvents used to dissolve DTT and suspend BIOX and images of the visual change after 25 minutes.

The images in Table 15 show that only aqueous solutions of DTT cause the formation of dark green aggregates after 25 minutes. This formation occurred after 10 minutes which is slower than hydrated BIOX suspensions that typically change colour and aggregate within 2 – 3 minutes of mixing. This may be due to reactive sites being more easily accessible in the hydrated BIOX samples. All DTT solutions in non-aqueous solvents failed to yield a visual change to the material at a neutral pH.

3.8.5 Iron Oxide Dependence

The iron oxides goethite and 2-Fh were prepared synthetically while haematite was purchased. These iron oxides were then subjected to the same reaction conditions as BIOX to investigate whether or not the green aggregation was unique to BIOX. The result of adding DTT to suspensions of these iron oxides is shown in Table 16.

Iron Oxide	Suspension Colour	Appearance after DTT addition
2-Fh	Orange/Brown	Dark green aggregates formed
Goethite (δ-FeOOH)	Yellow/Orange	Unchanged
Haematite (α-Fe$_2$O$_3$)	Dark Red	Unchanged

Table 16 - Colour and appearance of iron oxide suspensions pre- and post-addition of DTT.

Synthetic 2-Fh showed the same visual change as BIOX when combined with DTT while neither goethite nor haematite showed any visual change. 2-Fh is a highly hydroxylated iron mineral containing both bulk and surface hydroxyls. The surface hydroxyl configuration (SHC) influences the surface properties of oxides and oxyhydroxides.[48] The extent of hydroxylation has not yet been confirmed for 2-Fh however it has been hypothesised that bulk hydroxyls occupy vacant Fe(III) sites.[49] Recent modelling studies have shown 2-Fh to have a higher degree of surface hydroxyls that are not hydrogen

bonded to neighbours, and also free Fe Lewis acid sites, when compared with goethite.[50, 51] This may allow an increased reactivity towards the free thiols of DTT. To test this hypothesis a suspension of brucite, a highly hydroxylated mineral comprising magnesium hydroxide, was prepared.[52] Once mixed with DTT the brucite suspension rapidly turned from white to light pink indicating that there may be an interaction occurring between the brucite and the thiols.

These results have shown that addition of DTT and BME to BIOX suspensions cause dark green aggregates to form. This is likely the result of a DTT-BIOX complex forming, with binding occurring through the thiol moieties. This reaction is pH dependent, requiring a circumneutral pH to yield a dark green complex and basic pH to yield a dark red complex, solvent dependent, requiring aqueous suspensions and also iron oxide dependent, requiring 2-Fh phased materials. The amorphous nature of 2-Fh makes it difficult to comment on the exact nature of the complex formed, however this complex formation may indicate the concentration of free iron sites within 2-Fh. Further work is warranted to gain a better understanding of the reaction and what it tells us about iron availability within 2-Fh.

3.8.6 Cambridge Structural Database Comparisons

The Cambridge Structural Database (CSD) was searched for coloured iron-thiol complexes. This search returned 4150 compounds of various colours. Figure 33 shows a pie chart divided according to the colour reported.

Colour of Iron-Thiol Complexes
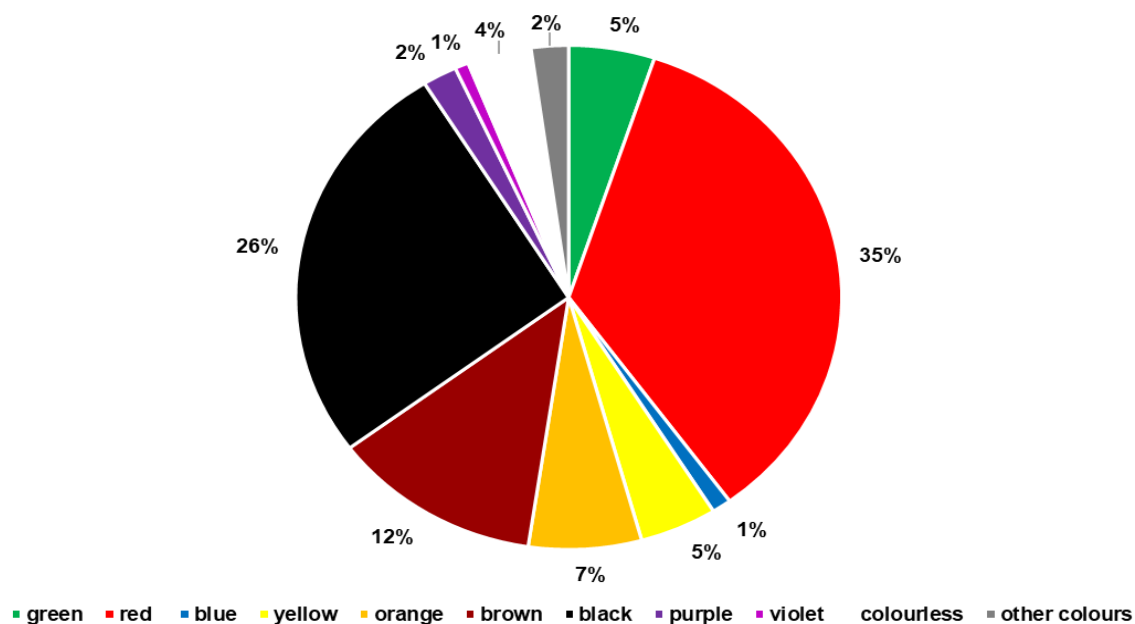
Figure 33 - Pie chart showing the colours of iron-thiol containing complexes found in the CSD.

This data shows that red and black coloured iron-thiol complexes are by far the most common. Green compounds make up only ca. 5 % and dark green compounds make up only ca. 1.2 % indicating that dark green iron-thiol complexes are unusual.

The search was refined to return only the 51 dark green coloured compounds and their crystal structures were checked for similarities that may be responsible for the dark green colour. None of the returned compounds contained only sulfur bonded to the iron centre, instead each compound contained other ligands including nitrogen donors, phosphorus donors, cyano ligands, carbonyl ligands and sandwich complexes. There was also no common geometry between the dark green compounds. Figure 34 shows representative examples from the CSD of regularly occurring crystal structures and ligands seen for the dark green compounds.
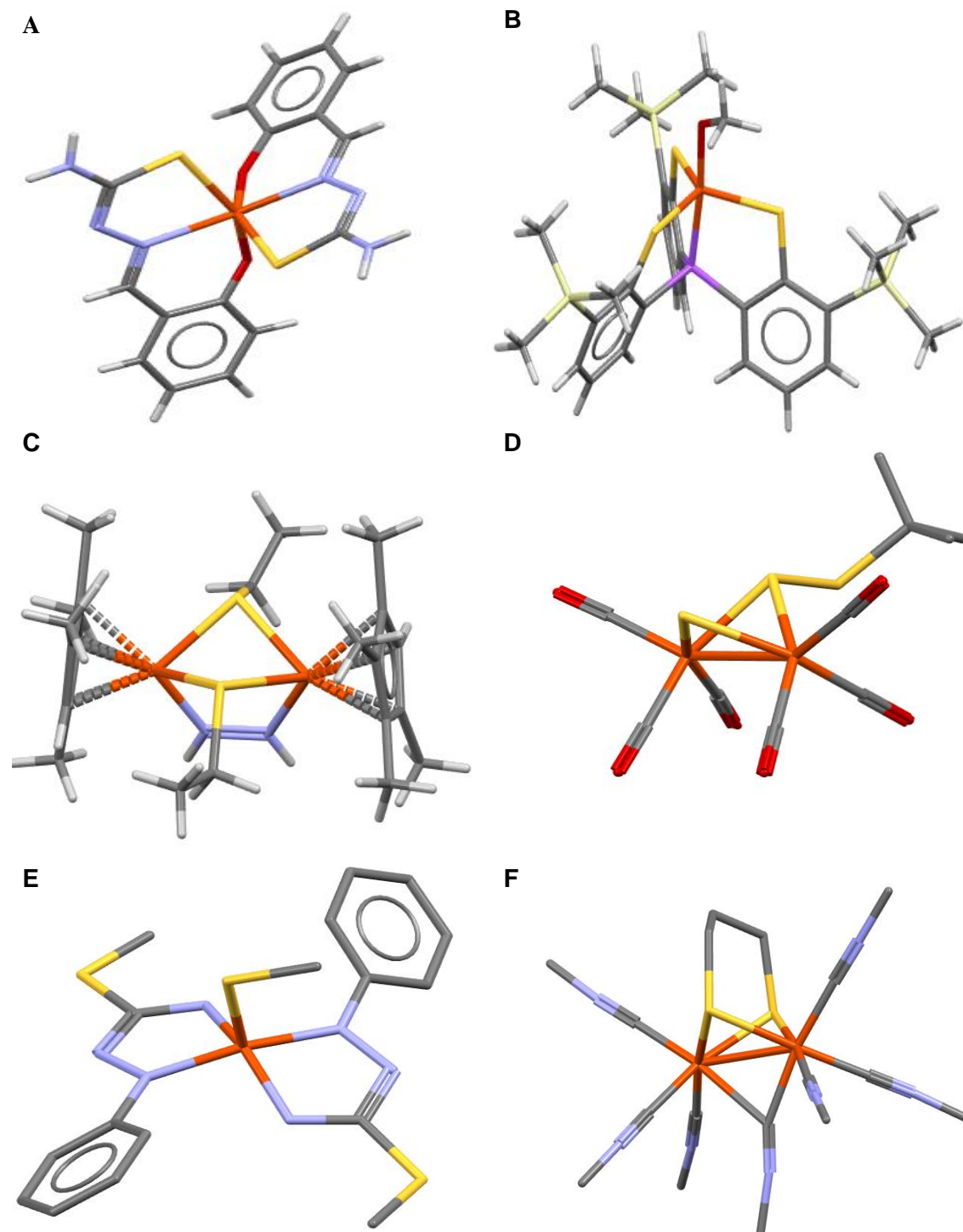
Figure 34 - Crystal structures of dark green compounds containing iron- sulfur bonds returned from searching the CSD. Their CSD codes are as follows: A – EXUYON, B – KAGRES, C - ATIFOZ, D – KADCIB, E – IXUNIY and F - LUMKUZ.[53-58]

As BIOX is composed of 2-Fh, which contains Fe (III) centres octahedrally coordinated with oxygen atoms, the search was further refined to include only dark green compounds containing iron-thiol and iron-oxygen bonds.[25] Interestingly this returned only two compounds - EXUYON and KAGRES, both of which contain Fe (III) centres.[53, 54] The

crystal structures of these are A and B in Figure 34. These structures show that EXUYON is octahedrally coordinated by two tridentate ligands that each contain sulfur, nitrogen and oxygen donors while KAGRES is a distorted trigonal-bipyramid containing four iron-thiol bonds, one iron-phosphorus bond and one iron-oxygen bond via a terminal methoxide ligand. It was decided that EXUYON was a suitable compound to synthesise and study spectrophotometrically as this may provide insight into bonding and transitions present in the BIOX-DTT complex.

Synthesis of EXUYON was carried out under the author's supervision by Ms. Faye Purcell, a final year undergraduate student, in accordance with the protocol in Chapter 2 Section 4, however this did not yield the expected structure. Instead, dark green crystals of a compound denoted MS1 formed. Single crystal XRD analysis was performed, and the structure was solved. This can be seen in Figure 35.
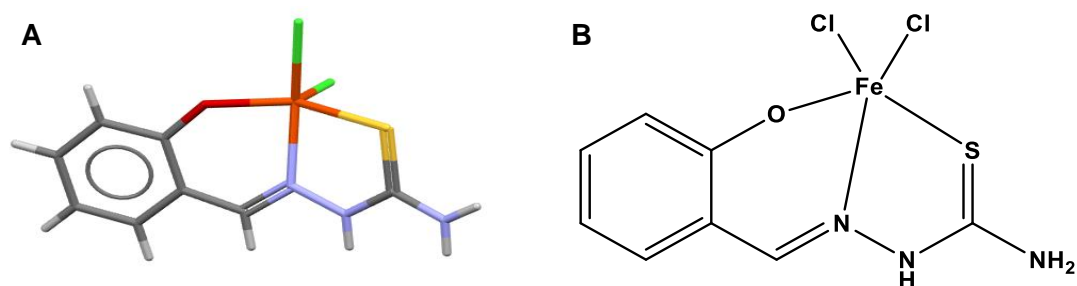
Figure 35 - A - Crystal structure of MS1 and B - Chemical structure of MS1.

MS1 is square pyramidal rather than octahedral with two terminal chloride ligands and one tridentate ligand binding through sulfur, nitrogen and oxygen atoms. This compound has not been reported in CSD however similar compound containing comparable ligands and bond lengths can be found. Bond lengths were compared to known values published by Allen *et al.* and subsequently assigned multiplicity (Table 17).[59]

Bond	Measured bond length (Å)	Assigned bond multiplicity	Bond	Measured bond length (Å)	Assigned bond multiplicity
Fe1-S3	2.391	Single	N3-C6	1.295	Double
Fe1-Cl4	2.220	Single	C5-C6	1.426	Single
Fe1-Cl2	2.296	Single	C5-C9	1.405	Double
Fe1-O1	1.886	Single	C5-C12	1.420	Double
Fe1-N3	2.136	Single	C9-C10	1.371	Double
S3-C13	1.712	Double	C10-C11	1.394	Double
C13-N1	1.319	Single	C11-C4	1.378	Double
C13-N2	1.327	Single	C4-12	1.393	Double
N2-N3	1.384	Single	C12-O1	1.331	Single

Table 17 - Bond distances between non-hydrogen atoms determined from single crystal X-ray diffraction.

The visible spectrum of MS1 was compared with that of BIOX-DTT and the literature spectrum of IXUNIY (Figure 34 – E) as shown in Figure 36.
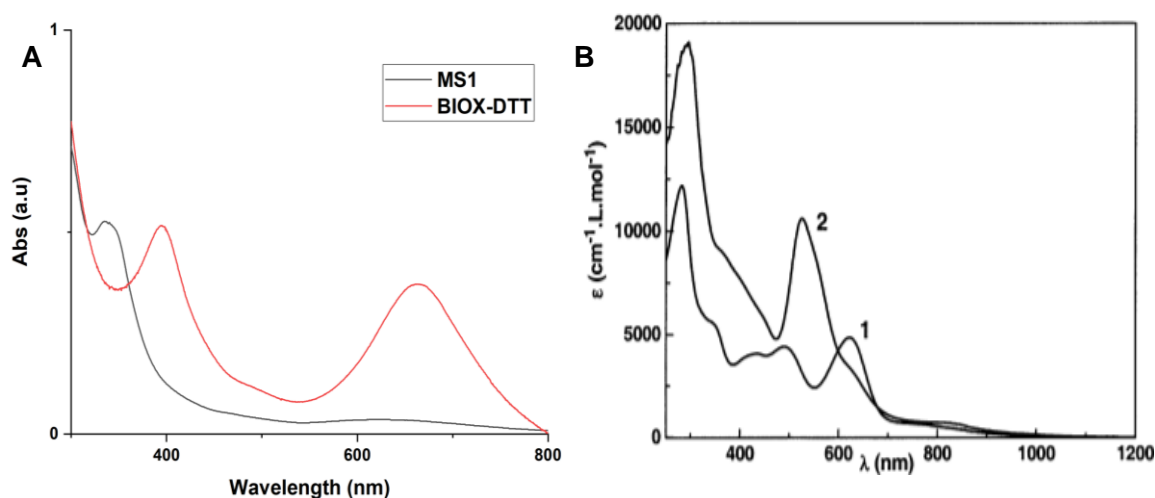
Figure 36 - Visible spectra of A - MS1 and BIOX-DTT and B - IXUNIY (Spectrum 1) from Blanchard *et al.*[55]

MS1 has a broad, weak absorption at ca. 610 nm and a stronger, sharper absorption at 336 nm while BIOX-DTT has a sharper, more intense absorption at 665 nm and a sharp absorption at 395 nm. IXUNIY was chosen as a comparison as it is a dark green iron-thiol containing compound and its visible spectrum was included by Blanchard *et al.*[55] IXUNIY has two sharp absorptions at 620 and 320 nm respectively. This spectrum was recorded in dichloromethane, an aprotic polar solvent, rather than water, a polar protic solvent. This may contribute to the shifted nature of the absorptions when compared with MS1.

Blanchard *et al.* assigned the absorption at 620 nm to a spin and dipole allowed ligand-to-ligand charge transfer between the two planar N,N-coordinated π radical ligands. Similar transitions have been described in other dark green iron complexes containing radical ligands by Ghosh *et al.*[60] Figure 37 shows the structure of the aforementioned dark green radical complexes.[55]
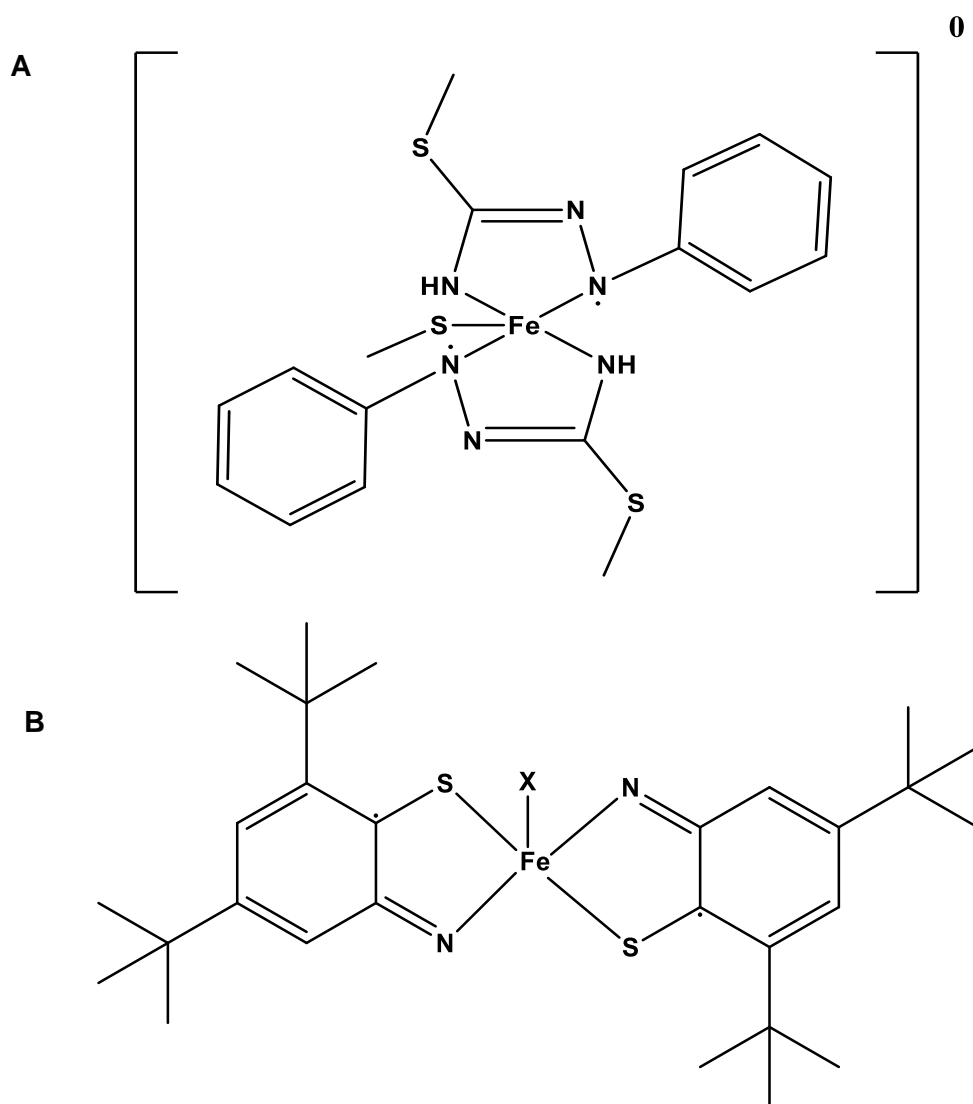
Figure 37 – Diagrams of A - IXUNIY adapted from Blanchard *et al.* showing the radical nitrogen atoms. The iron centre is in the Fe (III) oxidation state and B – Dark green iron complex containing radical ligands from Ghosh *et al.* X = P(OPh)₃ and CN⁻. The iron centre is in the Fe (II) oxidation state.[55, 60]

Both compounds contain aromatic stabilised radicals and have iron centres in different oxidation states, A is Fe (III) and B is Fe (II), indicating that the dark green colour is not influenced by the oxidation state of the iron. It is likely that a similar mechanism is occurring in the synthesised sample MS1 however electron paramagnetic resonance (EPR) spectroscopy and Mössbauer spectroscopy would be required to confirm this.

Regarding the dark green BIOX-DTT complex it is more difficult to say if a radical complex is responsible for the dark green colour and more research is warranted. Charge transfer reactions however are common for iron containing samples and are known to be responsible for causing dark colours.[61] This is true for iron containing minerals rockbridgeite and magnetite which are dark green and black respectively. This suggests that a charge transfer reaction is likely responsible for the intense dark green colour of the BIOX-DTT complex. This charge transfer reaction may either be ligand-metal where charge is transferred from an electron rich ligand to a metal centre or inter-valence charge transfer (IVCT) where charge is transferred between metal centres in different oxidation states. As BIOX contains Fe (III) centres it is possible that a fraction of them may be reduced to Fe (II) when complexed with the electron rich thiol moieties which could then lead to IVCT.

3.9 Conclusions

FeOB such as *Leptothrix spp.* and *Gallionella spp.* are responsible for the production of BIOX with distinct microstructures including hollow filaments and twisted stalks. BIOX were sampled from sites near and around Glasgow. SEM was used to confirm that sampled BIOX contained predominantly filaments likely produced by *L.ochracea,* occasionally twisted stalks produced by *G.ferruginea* and infrequently trichomes produced by *Toxothrix spp.* EDX confirmed that these samples were comprised of mostly iron, oxygen and carbon with low concentrations of inorganics such as silicon and sulfur and element mapping showed that these elements were distributed throughout the filaments. XRD showed that the phase of sampled BIOX was consistently 2-Fh. This confirmed the initial hypothesis for this chapter.

SEM was used to image the surface film collected from Kelvin Walkway and showed that both microbial cells and BIOX filaments were associated with it. There also appeared to be deposits associated with the surface film that are likely also 2-Fh in nature. This surface film may be used by the associated species for a number of reasons including UV light mitigation, oxygen concentration management and as a mean of transport to colonise new areas upon disruption of the film.

A variety of reducing reagents were used to try and solubilise the organic material associated with BIOX to be characterised by NMR. Sodium dithionite was chosen as the best reducing reagent to use and a protocol was developed. TLC of dithionite treated

BIOX confirmed the presence of polar organic material and orcinol staining indicated the presence of a dilute saccharide. NMR characterisation proved challenging due to both the paramagnetic nature of the associated iron and the poor solubility of isolated material. A DOWEX cation exchange resin and oxalic acid were used to reduce the iron concentration however the ^{1}H NMR spectra produced after each of these treatments were inconsistent. The fact that no structural characterisation studies of *L.ochracea* EPS have been attempted is likely due to no axenic culture of *L.ochracea* existing. This means there will always be associated EPS that may belong to other bacteria within the BIOX mat. Also, the heterogeneous nature of environmental sample sites means that there may be variability in composition depending on location, season and other environmental factors.

A protocol involving a NaOH extraction prior to dithionite extraction was then developed to try and purify the BIOX by removing associated EPS. The same problems arose as before and resulting spectra were inconsistent when the experiment was repeated. To properly characterise BIOX in this way a more thorough sampling protocol must be developed to ensure that all BIOX samples are consistent, and the extraction protocol must be optimised to effectively remove iron without losing sample and to improve the solubility of extracted material. This work partially proved the second hypothesis for this chapter in that the reduction of ferric to ferrous iron did allow some characterisation of the organic backbone, however further work is required to fully characterise its structure.

Addition of thiol containing reducing agents DTT and BME caused BIOX suspensions to turn dark green and form aggregates. This result could only be achieved by adding neutral aqueous solutions of the reagents to 2-Fh phased materials. Addition to other phases of iron oxides did not yield any colour change or aggregation. When this reaction was carried out under basic conditions the samples instead turned dark red. FTIR spectroscopy and SEM-EDX indicates that a BIOX-DTT complex is likely forming and that DTT binds through its thiol moiety.

The CSD was searched for coloured iron-thiol compounds. The returned results showed that dark green iron-thiol compounds represent only ca. 1.2 % of coloured iron-thiol compounds. A representative example, EXUYON, was chosen to be synthesised and studied as it contains both iron-thiol and iron-oxygen bonds which are also present in BIOX-DTT. Synthesis was unsuccessful, however a dark green compound, MS1, was synthesised instead. Its crystal structure was solved and its visible absorption profile was compared with that of BIOX-DTT and a literature example, IXUNIY. From this comparison

is it was hypothesised an aromatically stabilised radical complex may have formed and be responsible for the dark green colour. Furthermore, the oxidation state of the iron centre is not responsible for the dark green colour. To confirm this, and the final hypothesis for this chapter, it would be necessary to carry out EPR and Mössbauer spectroscopy on MS1. It is also likely that the dark green colour of the BIOX-DTT complex is due to charge transfer reactions that may either be ligand-metal or IVCT in nature.

3.10 References

1. D. Emerson, E. J. Fleming and J. M. McBeth, *Annual Review of Microbiology*, 2010, **64**, 561-583.
2. C. S. Chan, S. M. McAllister, A. H. Leavitt, B. T. Glazer, S. T. Krepski and D. Emerson, *Frontiers in Microbiology*, 2016, **7**, 796.
3. E. J. Fleming, I. Cetinić, C. S. Chan, D. Whitney King and D. Emerson, *The Isme Journal*, 2013, **8**, 804.
4. R. E. Wheatley, *Journal of Soil Science*, 1988, **39**, 253-264.
5. T. S. Rao, T. N. Sairam, B. Viswanathan and K. V. K. Nair, *Corrosion Science*, 2000, **42**, 1417-1431.
6. N. Kip and J. A. van Veen, *The ISME journal*, 2015, **9**, 542-551.
7. J. Takada, H. Kunoh and T. Kunoh, *Journal of Microbial & Biochemical Technology*, 2015.
8. I. Bica, E. M. Anitas, H. J. Choi and P. Sfirloaga, *Journal of Materials Chemistry C*, 2020, **8**, 6159-6167.
9. V. V. Strykanova, L. B. Gulina, V. P. Tolstoy, E. V. Tolstobrov, D. V. Danilov and I. Skvortsova, *ACS Omega*, 2020, **5**, 15728-15733.
10. T. Ema, Y. Miyazaki, I. Kozuki, T. Sakai, H. Hashimoto and J. Takada, *Green Chemistry*, 2011, **13**, 3187-3195.
11. T. Kunoh, S. Matsumoto, N. Nagaoka, S. Kanashima, K. Hino, T. Uchida, K. Tamura, H. Kunoh and J. Takada, *Scientific Reports*, 2017, **7**, 6498.
12. H. Makita, Y. Nakahara, H. Fukui, Y. Miyanoiri, M. Katahira, H. Seki, M. Takeda and J.-i. Koizumi, *Bioscience, Biotechnology, and Biochemistry*, 2006, **70**, 1265-1268.
13. H. Hashimoto, S. Yokoyama, H. Asaoka, Y. Kusano, Y. Ikeda, M. Seno, J. Takada, T. Fujii, M. Nakanishi and R. Murakami, *Journal of Magnetism and Magnetic Materials*, 2007, **310**, 2405-2407.
14. A. Alharthi, R. A. Blackley, T. H. Flowers, J. S. J. Hargreaves, I. D. Pulford, J. Wigzell and W. Zhou, *Journal of Chemical Technology & Biotechnology*, 2014, **89**, 1317-1323.
15. K. Mandai, M. Hanata, K. Mitsudo, H. Mandai, S. Suga, H. Hashimoto and J. Takada, *Tetrahedron*, 2015, **71**, 9403-9407.
16. H. Hashimoto, A. Itadani, T. Kudoh, S. Fukui, Y. Kuroda, M. Seno, Y. Kusano, Y. Ikeda, Y. Benino, T. Nanba, M. Nakanishi, T. Fujii and J. Takada, *ACS Applied Materials & Interfaces*, 2013, **5**, 5194-5200.
17. I. Suzuki and D. Sahabi, in *Handbook of Metal Biotechnology: Applications for Environmental Conservation and Sustainability* eds. M. Ike, M. Yamashita and S. Soda, Pan Stanford, New York, 1st edn., 2011, ch. 2, pp. 11 - 25.
18. T. Kunoh, M. Nakanishi, Y. Kusano, A. Itadani, K. Ando, S. Matsumoto, K. Tamura, H. Kunoh and J. Takada, *Water Research*, 2017, **122**, 139-147.
19. M. Takeda, K. Kondo, M. Yamada, J.-i. Koizumi, T. Mashima, A. Matsugami and M. Katahira, *International Journal of Biological Macromolecules*, 2010, **46**, 206-211.
20. M. Takeda, H. Makita, K. Ohno, Y. Nakahara and J.-i. Koizumi, *International Journal of Biological Macromolecules*, 2005, **37**, 92-98.
21. M. Takeda, Y. Kawasaki, T. Umezu, S. Shimura, M. Hasegawa and J.-i. Koizumi, *Archives of Microbiology*, 2012, **194**, 667-673.
22. C. S. Chan, S. C. Fakra, D. C. Edwards, D. Emerson and J. F. Banfield, *Geochimica et Cosmochimica Acta*, 2009, **73**, 3807-3818.
23. M. Furutani, T. Suzuki, H. Ishihara, H. Hashimoto, H. Kunoh and J. Takada, *Minerals*, 2011, **1**, 157-166.

24. Y. M. Mos, A. C. Vermeulen, C. J. N. Buisman and J. Weijma, *Geomicrobiology Journal*, 2018, **35**, 511-517.
25. R. M. Cornell and U. Schwertmann, *The Iron Oxides: Structure, Properties, Reactions, Occurences and Uses (First Edition),* 1996.
26. R. Yang, J. Tao, Q. Huang, B. Tie, M. Lei, Y. Yang and H. Du, *Journal of Soils and Sediments*, 2019, **19**, 1319-1327.
27. D. Emerson and W. C. Ghiorse, *Journal of Bacteriology*, 1993, **175**, 7819-7827.
28. A. Langell Marjorie, E. Kadossov, H. Boparai and P. Shea, *Surface and Interface Analysis*, 2009, **41**, 941-950.
29. L. Selwyn and S. Tse, *Studies in Conservation*, 2008, **53**, 61-73.
30. F. Lelchat, S. Cérantola, C. Brandily, S. Colliec-Jouault, A.-C. Baudoux, T. Ojima and C. Boisset, *Carbohydrate Polymers*, 2015, **124**, 347-356.
31. M. Garcia-Vaquero, in *Dietary Fiber: Properties, Recovery, and Applications*, ed. C. M. Galanakis, Academic Press, 2019, pp. 165-197.
32. X. Zhang, T. Yuan, F. Peng, F. Xu and R. Sun, *Separation Science and Technology*, 2010, **45**, 2497-2506.
33. Z. Wang, S. Liu, Y. Matsumoto and S. Kuga, *Cellulose*, 2012, **19**, 393-399.
34. M. Boleij, T. Seviour, L. L. Wong, M. C. M. van Loosdrecht and Y. Lin, *Water Research*, 2019, **164**, 114952.
35. N. Kumar Singh, S. Pandey, R. P. Singh, K. Muzamil Gani, M. Yadav, A. Thanki and T. Kumar, in *Bioremediation of Pollutants*, eds. V. C. Pandey and V. Singh, Elsevier, 2020, pp. 251-273.
36. R. Lill, *Nature*, 2009, **460**, 831-838.
37. L. Maurizi, H. Bisht, F. Bouyer and N. Millot, *Langmuir*, 2009, **25**, 8857-8859.
38. R. Perera, M. Sono, J. A. Sigman, T. D. Pfister, Y. Lu and J. H. Dawson, *Proceedings of the National Academy of Sciences*, 2003, **100**, 3641.
39. A. P. Mathews and S. Walker, *Journal of Biological Chemistry*, 1909, **6**, 299-312.
40. F. M. Page, *Transactions of the Faraday Society*, 1955, **51**, 919-925.
41. R. M. Cornell and W. Schneider, *Polyhedron*, 1989, **8**, 149-155.
42. R. M. Cornell, W. Schneider and R. Giovanoli, *Polyhedron*, 1989, **8**, 2829-2836.
43. C. Colombo, G. Palumbo, V. M. Sellitto, H. G. Cho, C. Amalfitano and P. Adamo, *Journal of Geochemical Exploration*, 2015, **151**, 50-56.
44. A. Violante, E. Barberis, M. Pigna and V. Boero, *Journal of Plant Nutrition*, 2003, **26**, 1889-1908.
45. T. Ismail, M. Qureshi, N. Qureshi, N. Akhtar, Q. Mansoor and M. Ismail, *Tropical Journal of Pharmaceutical Research*, 2016, **15**, 599-603.
46. G. M. Whitesides, J. E. Lilburn and R. P. Szajewski, *The Journal of Organic Chemistry*, 1977, **42**, 332-338.
47. A. Gojmerac Ivšić and B. Tamhina, *Croatica Chemica Acta*, 2003, **76**, 323-328.
48. V. Barrón and J. Torrent, *Journal of Colloid and Interface Science*, 1996, **177**, 407-410.
49. F. M. Michel, V. Barrón, J. Torrent, M. P. Morales, C. J. Serna, J.-F. Boily, Q. Liu, A. Ambrosini, A. C. Cismasu and G. E. Brown, *Proceedings of the National Academy of Sciences*, 2010, **107**, 2787.
50. J.-F. Boily and X. Song, *Communications Chemistry*, 2020, **3**, 79.
51. X. Song and J.-F. Boily, *Physical Chemistry Chemical Physics*, 2012, **14**, 2579-2586.
52. R. L. Frost and J. T. Kloprogge, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 1999, **55**, 2195-2205.
53. J. Patole, S. Padhye, M. S. Moodbidri and N. Shirsat, *European Journal of Medicinal Chemistry*, 2005, **40**, 1052-1055.
54. K.-C. Chang, C.-J. Huang, Y.-H. Chang, Z.-H. Wu, T.-S. Kuo and H.-F. Hsu, *Inorganic Chemistry*, 2016, **55**, 566-572.

55.   S. Blanchard, E. Bill, T. Weyhermüller and K. Wieghardt, *Inorganic Chemistry*, 2004, **43**, 2324-2329.
56.   Y. Chen, L. Liu, Y. Peng, P. Chen, Y. Luo and J. Qu, *Journal of the American Chemical Society*, 2011, **133**, 1147-1149.
57.   X. Wu, K. S. Bose, E. Sinn and B. A. Averill, *Organometallics*, 1989, **8**, 251-253.
58.   J. D. Lawrence, T. B. Rauchfuss and S. R. Wilson, *Inorganic Chemistry*, 2002, **41**, 6193-6195.
59.   F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen and R. Taylor, *Journal of the Chemical Society, Perkin Transactions 2*, 1987, S1-S19.
60.   P. Ghosh, E. Bill, T. Weyhermüller and K. Wieghardt, *Journal of the American Chemical Society*, 2003, **125**, 3967-3979.
61.   S. M. Mattson and G. R. Rossman, *Physics and Chemistry of Minerals*, 1987, **14**, 94-99.

4 Phylogenetic Analysis of BIOX Bacterial Communities

4.1 Hypothesis

Prior to commencing work on this chapter is was hypothesised that firstly, 16S rRNA gene sequences for *Leptothrix spp.*, *Gallionella spp.*, and *Toxothrix spp.* would be found in the datasets for each sample site. It was secondly hypothesised that the bacterial communities of the Allander sample sites would be more closely representative of each other than of those found in other BIOX sites found in the literature. Finally, it was hypothesised that the bacterial communities found within the sample sites will be indicative of the various redox cycles occurring due to their associated microbial metabolisms. For example, the presence of both iron-oxidising and iron-reducing bacteria would show that biogenic iron cycling is likely occurring within the BIOX mats.

4.2 Introduction

Phylogenetics is the study of evolutionary relationships between organisms.[1] This was historically carried out by comparisons of morphology and analysis of fossilised materials, however is now typically carried out via sequencing of genetic material such as DNA, RNA or specific genes. The results of this are then used to construct a phylogeny, also known as a phylogenetic tree, which allows the visualisation of evolutionary relationships and taxonomic hierarchy.[2] The first recorded use of a tree diagram to represent historical relationships between organisms was by Lamarck in 1809 and then more famously by Darwin from 1837 onwards.[3, 4] These trees were included in his notebooks and the only diagram in *On the Origin of Species* was a tree-like diagram.[5] Figure 38 contains a sketch of a tree-like diagram by Darwin and a diagram showing the taxonomic hierarchy of *L.ochracea*.
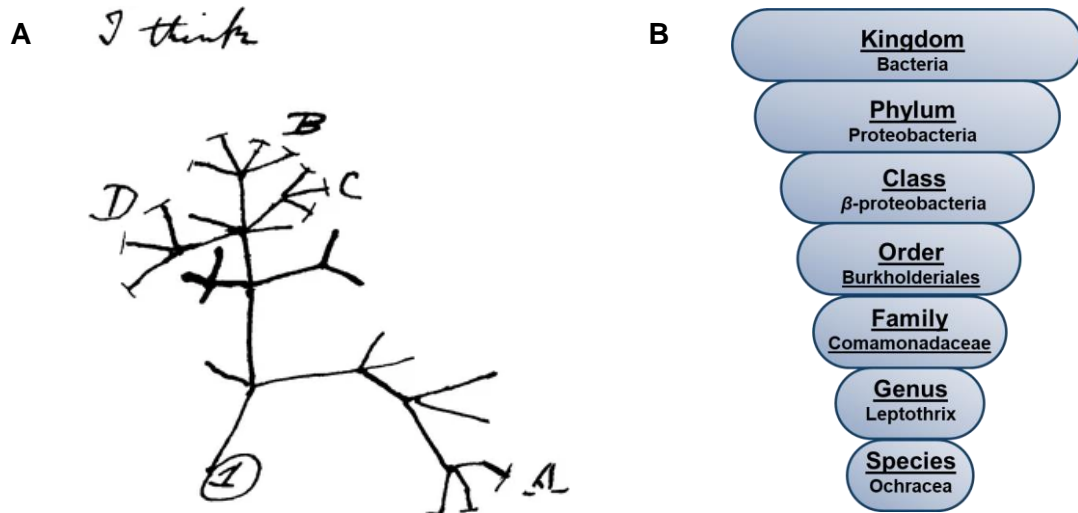
Figure 38 – A – Sketch of a tree-like diagram for Darwin's notebook and B - Taxonomic ranking of _L.ochracea_.[4]

Woese and Fox pioneered the use of the 16S rRNA gene for constructing phylogenies in 1977.[6] Since then it has been utilised in numerous studies and is one of the most important targets to identify environmental and clinical isolates and assign their phylogenetic relationships.[7, 8] The 16S rRNA gene is ca. 1.5 kilobases (kb) in length, is ubiquitous to all microorganisms, evolves slowly and is highly conserved. However, within this gene there are 9 hypervariable regions (V1 – V9) which change rapidly and range from 30 – 100 base pairs (bp) in length. These hypervariable regions combined with the slower changing conserved regions provide a suitable marker for relatedness of organisms. The 16S gene is a component of the 30S small subunit of the prokaryotic bacterial ribosome, which acts as a structural scaffold, and interacts with the 23S rRNA of the 50S subunit thereby allowing the 30S and 50S subunits to bind. A diagram of this can be seen in Figure 39.[9]
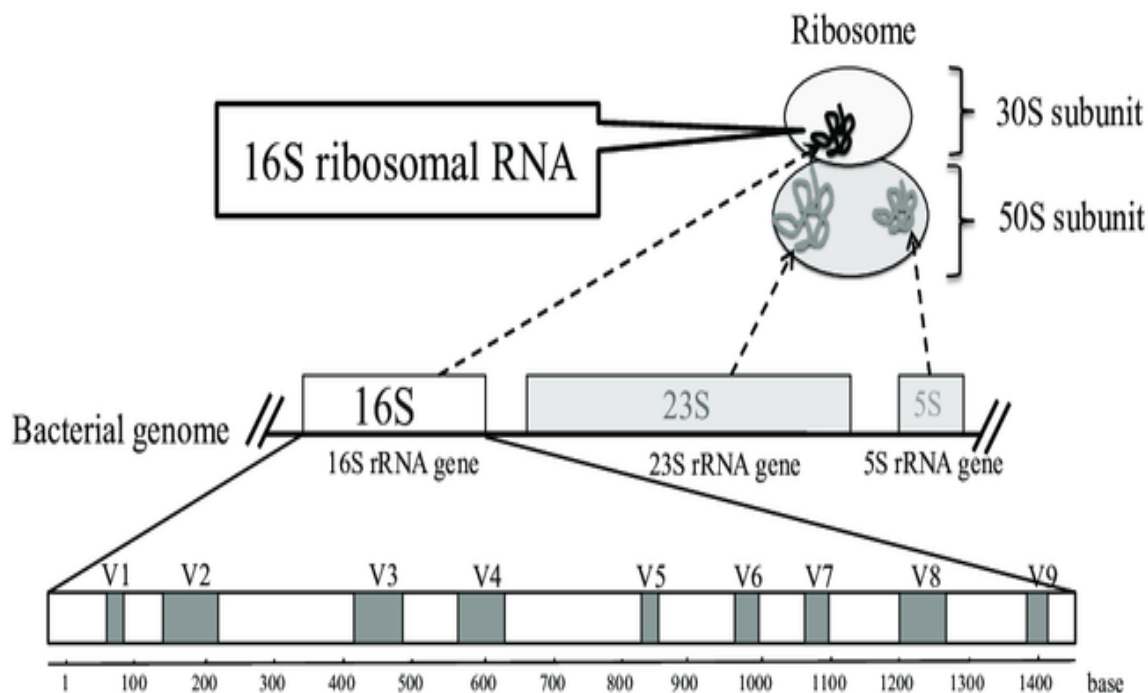
Figure 39 - Diagram showing the 16S rRNA gene with the 9 hypervariable regions highlighted. Adapted from Fukuda _et al._[9]

The highly conserved nature of certain regions of the 16S rRNA gene has allowed the development of universal polymerase chain reaction (PCR) primers to target it.[10] These universal primers can reliably generate the same sections of the 16S rRNA gene across a wide range of taxa. Targeting the highly conserved regions of the 16S rRNA gene is useful for assigning higher ranking taxa such as class or phylum whereas targeting the hypervariable regions allow universal primers to assign phylogeny to the genus level.[11] This is frequently carried out using second generation Illumina sequencing which generates short reads, of < 300 bp, making assigning taxonomic distinction below genus challenging. Technologies such as Oxford Nanopore sequencing can generate much longer reads allowing the entire 1500 bp of the 16S rRNA gene to be sequenced. This technology is becoming more cost effective and popular, however it can generate high sequencing error rates of up to 38 % and has only recently been able to accommodate high throughput sequencing.[12]

The most commonly targeted hypervariable region of the 16S rRNA gene is the V3 – V4 region due to the fact it contains a high amount of base pair heterogeneity and therefore shows considerable discriminating power.[13, 14] Other regions can be targeted including the V1 – V3 region and the V4 – V5 region. The V1 – V3 region can produce reads of up to 750 bp however sequencing of this has been limited to the Roche/454 pyrosequencing methodology which has now become outdated due to sequencing cost and throughput.

The V4 – V5 region is becoming more frequently targeted and the V3 – V4 and V4 – V5 regions have been shown to yield reproducible results when compared with the V1 – V3 region.[15]

4.3 Chapter Aims

There are currently no examples, to the author's knowledge, of comparable studies which investigate the phylogenetic diversity of BIOX sampled from the UK. This chapter aims to isolate genomic DNA from BIOX samples, sequence the 16S rRNA gene via high-throughput Illumina sequencing and use the resulting data to build a phylogenetic profile for each of the sample sites. The 16S gene sequences should contain *L.ochracea* and other FeOB such as *Gallionella spp.* and confirm how similar these are to both environmental and characterised bacteria. The sequence profiles will be compared with published studies from sample sites around the world to assess whether or not unconnected BIOX sample sites contain similar phylogenetic profiles.

4.4 Results and Discussion

4.4.1 Sample Collection and Genomic DNA Extraction

16S rRNA gene sequencing was carried out on three BIOX samples to gain a more thorough understanding of the bacterial ecosystem and to provide insight to the environmental factors that support the organisms in these mats. Based on the microtubular morphology present in the BIOX we can expect the presence of certain bacteria including *Leptothrix spp.*, such as *L.ochracea*, however this sequencing will identify the community of bacteria and their relative abundances. Furthermore, understanding the bacterial ecosystem may allow a better understanding of *L.ochracea* and provide additional information on how to develop conditions in which it can thrive.

Each sample was collected from different BIOX mats within the Allander sample site area using 500 ml Duran bottles. These samples are denoted Allander_1, Allander_2 and Allander_3. These three mats were within 10 m of one another, Allander_1 and Allander_2 within the same body of water, and full descriptions and physicochemical characterisation of them can be found throughout Chapter 6. All three samples contained BIOX filaments indicative of *L.ochracea* however the Allander_2 site appeared visually distinct from the other two as the source of water into the pool contains clumps of BIOX a

short distance (10 cm) from off white material. Images of each sample site highlighting the difference in BIOX can be seen in Figure 40
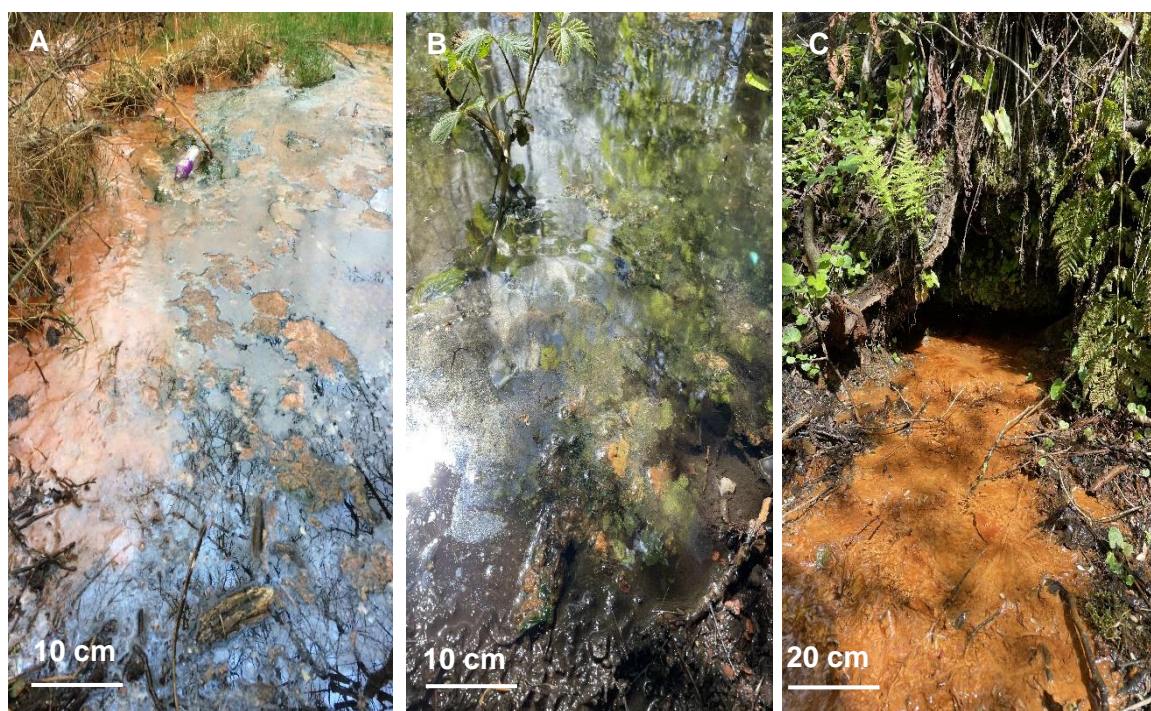
Figure 40 - Photographs of A - Allander_1, B - Allander_2 and C - Allander_3.

Initial BIOX samples collected on the 11/11/19 were stored at 4 °C then had their DNA extracted on the 21/11/19 to be prepared for sequencing. It was subsequently made apparent to the author that storage at 4 °C for 10 days would considerably alter the bacterial profile, therefore making it unrepresentative of the sample sites. To mitigate this, fresh BIOX samples were then collected on the 9/1/20, stored at – 80 °C, and had their genomic DNA extracted on the 11/1/20. DNA was extracted in accordance with the protocol in Chapter 2 Section 7. Briefly, BIOX samples (500 mg) had their DNA extracted via a FastDNA™ SPIN Kit for Soil supplied by MPBio. A workflow for this procedure can be seen in Figure 41.

Figure 41 - Workflow procedure for DNA extraction using a MPBio Fast DNA SPIN Kit for Soil.[16]

Extracted DNA (5 µl) was analysed on a 1 % agarose gel to assess the quality of the sample. High quality samples should contain minimal to no DNA fragmentation meaning that they are suitable for PCR. Figure 42 contains a photograph of the gel confirming that the DNA was suitable to be used for PCR amplification.
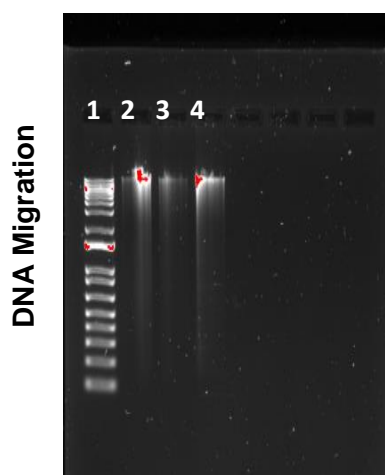
Figure 42 - Photograph of agarose gel of extracted DNA samples. Well 1 – 1000 kb Nucleotide ladder, Well 2 – Allander_1, Well 3 – Allander_2 and Well 4 – Allander_3.

Genomic DNA concentrations were adjusted to 20 ng/µl, a Qubit fluorometer was used to confirm the concentration, and were then sent to GENEWIZ, a company which specialises in genomic services and bioinformatics, for high-throughput Illumina sequencing targeting the V3 - V4 hypervariable regions of the 16S rRNA gene. There the DNA was amplified by PCR using universal primers targeting the 16S rRNA gene to confirm that amplification was possible. The forward primer used was F515 which had a sequence of 'GTGYCAGCMGCCGCGGTAA', while the reverse primer used was R926 which had a sequence of 'CCGYCAATTYMTTTRAGTTT', where the letters ATGC represent the individual bases while YMR represent allowed variation at these positions. Added to these primers are also adapter sequences that are necessary for the Illumina MiSeq sequencing.

4.4.2 16S Sequencing Data and Initial Analysis

The results from the high-throughput Illumina sequencing were processed using the QIIME pipeline.[17] QIIME is an open-source bioinformatics pipeline that is routinely used to process raw DNA sequencing data. 267,118 sequence reads from Allander_1, 227,758 sequence reads Allander_2 and 245,522 sequence reads from Allander_3 were respectively generated. On average these reads were 250 bp in length and had an average GC content of ca. 53%. These correspond to both forward and reverse sequence read pairs for the V3 – V4 hypervariable region of the 16S rRNA gene. The two sequences of each read pair were merged according to overlapping sequences via GENEWIZ's bioinformatics pipeline. If the overlapping sequence was at least 20 bp long, then the merge was deemed successful and the remaining undetermined bases were removed. This produces longer reads of 400 – 500 bp in length. Artefacts are commonly

produced during the sequencing process and as such must be identified, considered and removed for each individual read. These include spurious primer and adapter sequences at the 5' and 3' ends of the sequences and also individual bases with a Q score, a measure of sequence accuracy, of less than 20. The resulting paired sequences containing more than ca. 400 bp were progressed to the next step where sequences were aligned using the UCHIME 'Gold' database to remove chimeric sequences.[18] Chimeric sequences are an artifact introduced from the PCR step where sequences are amplified partially from one 16S gene and then continued from a second 16S gene. The effective sequences produced after this step were deemed ready for analysis and were on average ca. 450 bp in length. Figure 43 shows the distribution of effective sequence lengths generated.

Figure 43 - Effective sequence length distribution generated.

These procedures resulted in 87,952 optimised sequences from Allander_1, 64,820 optimised sequences from Allander_2 and 67,513 optimised sequences from Allander_3 from which operational taxonomic units (OTUs) were generated. OTUs are a definition of taxonomic classification that are assigned to sequence clusters with a > 97 % similarity. Unique sequences were extracted and sequences with only 1 read count were removed. Unique sequences with > 1 read count and > 97 % similarity were then clustered. This generated representative OTU sequences which were compared with the optimised

sequences. Optimised sequences containing a > 97 % sequence similarity with OTU sequences were then considered to belong to that OTU and the corresponding abundance data was subsequently added.

4.4.3 Reprocessing of the 16S Data

OTU clustering has historically been the most common method of assigning the taxonomy of 16S rRNA sequencing data. Current bioinformatics research, however, is replacing OTU clustering with amplicon sequence variants (ASV), also known as exact sequence variants, which are unique DNA sequences recovered from a sequencing dataset that has targeted a marker gene. The most frequently targeted marker gene is the 16S rRNA gene.[19] As they are unique DNA sequence they can show sequence variations down to one nucleotide change which gives greater resolution than the ca. 97 % sequence similarity of OTU clustering. The resulting ASVs are assigned taxonomy via comparisons with known databases such as the Silva database.[20]

To generate ASVs, the raw sequence reads were reprocessed using packages in R (R is an integrated suite of software facilities for data manipulation, calculations, and graphical display) specifically the DADA2 pipeline combined with the Silva nr99 v138 train set of 16S sequences. DADA2 is an open-source R package that is used to correct sequencing errors that arise from Illumina sequencing and has a resolution of one nucleotide base.[21] R-scripts were developed to better treat any errors within the data, as the default position is to remove the erroneous data, and optimised the number of ASVs generated. These R-scripts can be found in Appendix I. This allowed taxonomic assignment of the resulting sequences and yielded 688 ASVs for Allander_1, 937 ASVs for Allander_2 and 1521 ASVs for Allander_3 respectively. Figure 44 contains bar-charts showing the distribution of ASVs for each site versus the number of reads per ASV.
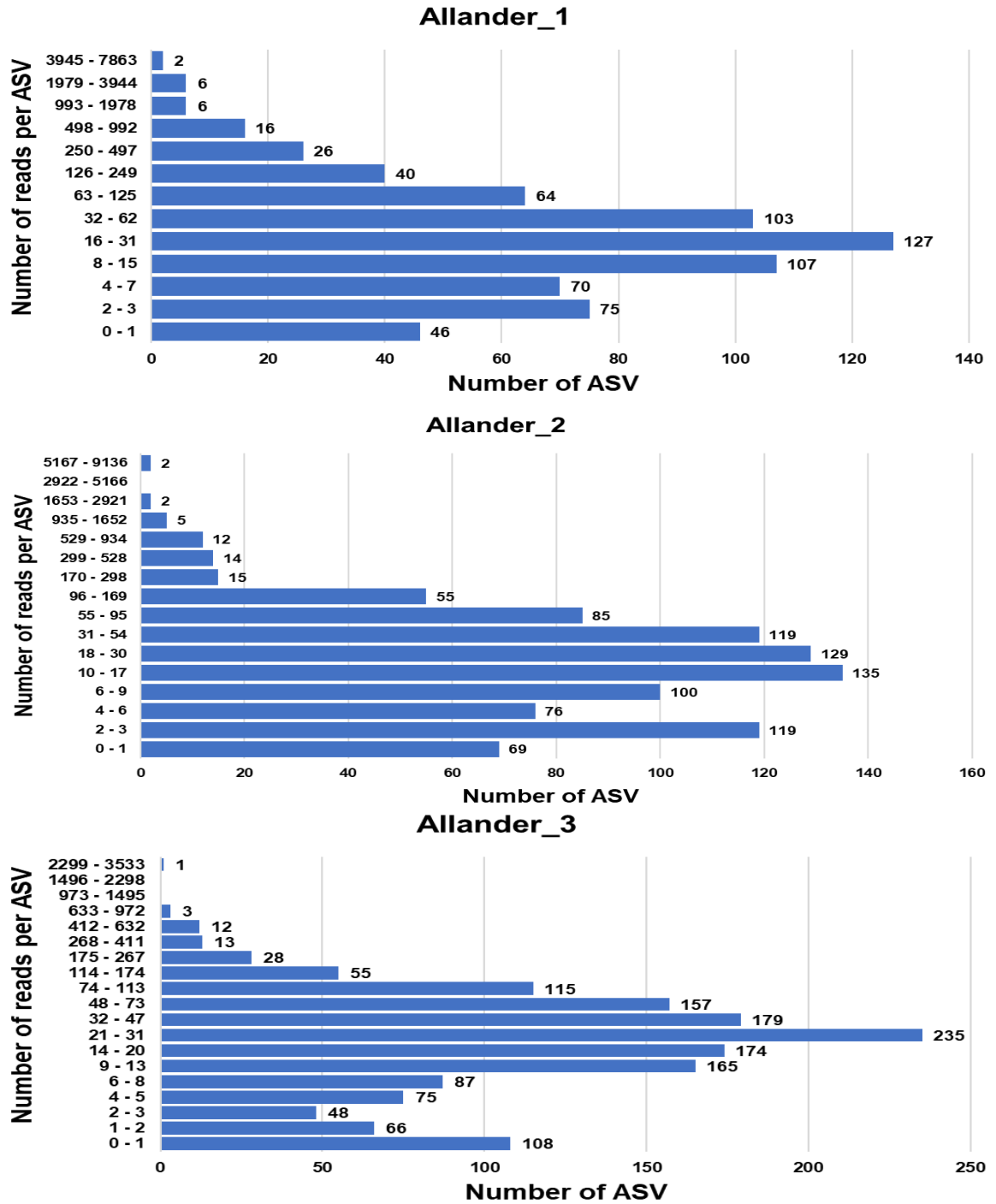
Figure 44 – Bar-charts showing the number of ASVs generated for each site vs. the number of reads per ASV.

This data illustrates that Allander_3 contains substantially more reads than Allander_1 or Allander_2. Furthermore, the number of reads per ASV are most frequently in the range of ca. 20 – 50. It is less common to have multiple hundreds of reads per ASV and particularly rare to have thousands of reads per ASV.

The 10 most abundant bacterial sequences at each site are shown in Table 18. The 10 sequences represent 31%, 38% and 13% of sequence reads from the Allander_1 site, Allander_2 site and Allander_3 site respectively. For a better overview of the distribution of all sequences Figure 45 and Figure 46 contain heatmaps showing fractional abundance, taxonomic overlap and differences between the three datasets.

Allander_1 Abundances	Class	Family	Genus
5658	*γ*-proteobacteria	*Methylomonadaceae*	pLW-20
4414	*γ*-proteobacteria	*Methylomonadaceae*	pLW-20
3073	*γ*-proteobacteria	*Methylomonadaceae*	pLW-20
3028	*β*-proteobacteria	*Comamonadaceae*	*Paucibacter*
2165	*β*-proteobacteria	*Gallionellaceae*	*Gallionella*
2079	*γ*-proteobacteria	*Methylomonadaceae*	pLW-20
1992	*β*-proteobacteria	*Comamonadaceae*	*Rhodoferax*
1724	*β*-proteobacteria	*Gallionellaceae*	*Gallionella*
1684	*γ*-proteobacteria	*Methylomonadaceae*	pLW-20
1187	*β*-proteobacteria	*Gallionellaceae*	*Sideroxydans*

Allander_2 Abundances	Class	Family	*Genus*
7388	*β*-proteobacteria	*Gallionellaceae*	*Gallionella*
5595	Cyanobacteriia	*Phormidiaceae*	*NA*
2345	*β*-proteobacteria	*Gallionellaceae*	*Gallionella*
2316	*γ*-proteobacteria	*Thiotrichaceae*	*Thiothrix*
1439	*β*-proteobacteria	*Gallionellaceae*	*Gallionella*
1330	*β*-proteobacteria	*Comamonadaceae*	*Paucibacter*
1288	*γ*-proteobacteria	*Thiotrichaceae*	*Thiothrix*
1206	*γ*-proteobacteria	*Thiotrichaceae*	*Thiothrix*
983	Campylobacteria	*Sulfurimonadaceae*	*Sulfuricurvum*
923	*β*-proteobacteria	*Sulfuricellaceae*	*Sulfuricella*

Allander_3 Abundances	Class	Family	*Genus*
3417	*β*-proteobacteria	*Gallionellaceae*	*Gallionella*
920	*β*-proteobacteria	*Comamonadaceae*	*Rhodoferax*
773	*β*-proteobacteria	*Comamonadaceae*	*Rhodoferax*
656	*β*-proteobacteria	*Gallionellaceae*	*Sideroxydans*
564	*β*-proteobacteria	*Gallionellaceae*	*Sideroxydans*
563	*β*-proteobacteria	*Gallionellaceae*	*Gallionella*
533	*β*-proteobacteria	*Comamonadaceae*	*Rhodoferax*
533	*β*-proteobacteria	*Gallionellaceae*	*Gallionella*
508	*β*-proteobacteria	*Comamonadaceae*	*Paucibacter*
498	*β*-proteobacteria	*Comamonadaceae*	*Sphaerotilus*

Table 18 - Abundance data (number of sequence reads) for the 10 most abundant bacteria in each sample site.
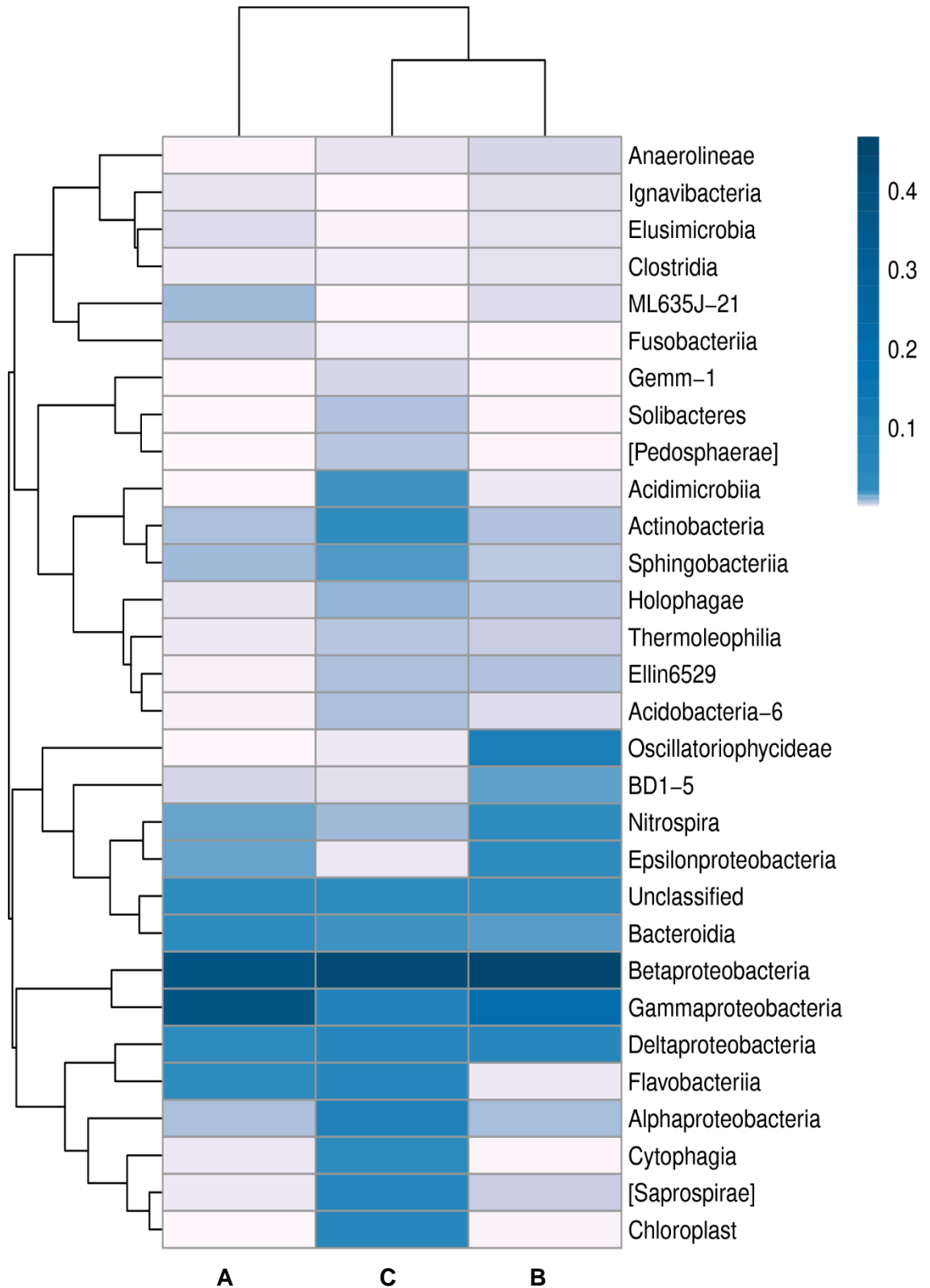
Figure 45 - Heatmap showing the taxonomic class fractional abundance of each site. A - Allander_1, B - Allander_2 and C - Allander_3.
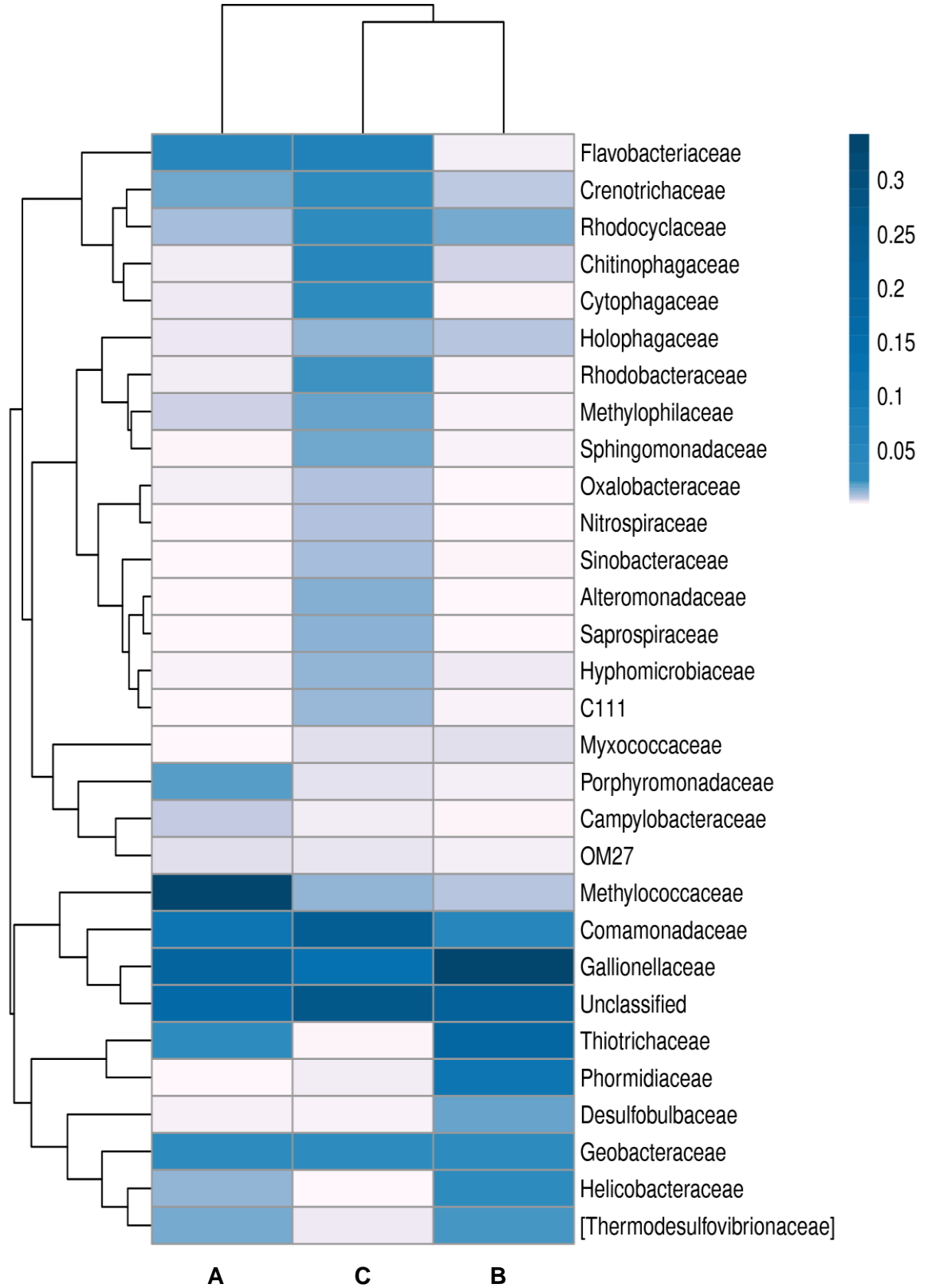
Figure 46 - Heatmap showing taxonomic family fractional abundance of each site. A - Allander_1, B - Allander_2 and C - Allander_3.

Looking in detail at the composition of the three sites we see that there is significant taxonomic overlap between the three sites, however each site has quite a distinct profile. The proteobacteria dominate all three sample sites, comprising 38 %, 27 % and 40 % of sequences from Allander_1, Allander_2 and Allander_3 respectively. Proteobacteria is a phylum level taxonomic distinction that is made up of gram-negative bacteria.[22] This phylum is further broken down into classes, one of which is the *β*-proteobacteria. This is the class to which *L.ochracea* belongs and is the most abundant class across all sites comprising 55 %, 58 % and 46 % of proteobacteria from each site. The *γ*-proteobacteria is the next most abundant class and are more abundant in Allander_1 where they represent 37 % of all proteobacteria while they represent 27 % of all proteobacteria in each of Allander_2 and Allander_3 respectively. At the family level all three sites contain *Comamonadaceae*, to which *L.ochracea* belongs, and *Gallionellaceae*, to which *G.ferruginea* belongs, confirming that all three sample sites contain ferrous rich waters that support the growth of iron-oxidising bacteria (FeOB). 61 % of *γ*-proteobacteria in Allander_1 are *Methylococcaceae*, which are methanotrophic bacteria, while only 34 % of Allander_2 and 28 % of Allander_3 are *Methylococcaceae* indicating that the Allander_1 site is likely richer in methane. The final *γ*-proteobacteria which are prevalent are *Thiothrix*. These are sulfur metabolising bacteria which have three unique sequences in the 10 most abundant Allander_2 sequences. They do not however appear in the Allander_1 10 most abundant sequences and do not appear in the Allander_3 dataset at all. In Allander_2 they represent ca. 30 % of all *γ*-proteobacteria while in Allander_1 they represent ca. 10 % of all *γ*-proteobacteria respectively. This suggests that the Allander_2 site is rich in sulfide and thiosulfate which are reduced to elemental sulfur by these bacteria.

4.4.4 Identification of Key Bacteria

δ-proteobacteria are evenly dispersed across the three sites while *α*-proteobacteria are more common in Allander_2 than Allander_1 or Allander_3. Surprisingly, the 16S data showed that the Allander_1 and Allander_2 sites contain no bacteria that are identified as *Leptothrix* bacteria at the genus level. Allander_3 contains 8 sequences, out of 1521, which can be identified as from the *Leptothrix* genus. These have ca. 95% sequence identities to *Leptothrix mobilis*, *Leptothrix cholodnii*. This result was unexpected as all three sites contain significant amounts of filamentous BIOX that is presumably produced by *L.ochracea*.

The use of either the Silva or RDP reference data sets for assignment of phylogeny from 16S rRNA sequences assumes that suitable reference sequences are present (there are 3 entries in the RDP trainset 18/release 11.5, December 2020). It became apparent that a *L.ochracea* reference was not included and so would not necessarily be identified as a *Leptothrix* genus sequence. The closest related genus to *Leptothrix* found within the three Allander 16S sequence libraries is *Paucibacter*. There is only one unique *Paucibacter* sequence identified in in the Allander_1 and Allander_2 data and four unique sequences in Allander_3, however they all show high abundance. The *Paucibacter* 16S rRNA sequence representing the 4th, 6th and 8th most abundant sequences in the Allander_1, Allander_2 and Allander_3 (Table 18) was input to both BLAST and the RDP database. This returned a 99.23 % sequence identity match, over 388 bases, with *L.ochracea* sequenced by Fleming *et al.*[23] As this *Paucibacter* sequence corresponds with *L.ochracea* and is present in all three Allander sites we can be sure that *L.ochracea* is indeed present in these sample sites and therefore responsible for the microtubular BIOX seen. Figure 47 shows the sequence alignment of this Allander *Paucibacter* sequence with *L.ochracea*.

```
Score = 38260.0
Length of alignment = 388
Sequence ASV_0004.Allander_1_Beta/1-388 (Sequence length = 388)
Sequence     Leptothrix_ochracea/324-711 (Sequence length = 388)

ASV_0004.Allander_1_Beta/1-388 TGGGGAATTTTGGACAATGGACGAAAGTCTGATCCAGCCAT
                               |||||||||||||||||||||||||||||||||||||||||
   Leptothrix_ochracea/324-711 TGGGGAATTTTGGACAATGGACGAAAGTCTGATCCAGCCAT

ASV_0004.Allander_1_Beta/1-388 GCCGCGTGCAGGAAGAAGGCCTTCGGGTTGTAAACTGCTTT
                               |||||||||||||||||||||||||||||||||||||||||
   Leptothrix_ochracea/324-711 GCCGCGTGCAGGAAGAAGGCCTTCGGGTTGTAAACTGCTTT

ASV_0004.Allander_1_Beta/1-388 TGTCAGGGAAGAAATCCTTTGAGCTAATACCTCGGGGGGAT
                               ||||||||||||||||||||||||||||||||||| |||||
   Leptothrix_ochracea/324-711 TGTCAGGGAAGAAATCCTTTGAGCTAATACCTCGGAGGGAT

ASV_0004.Allander_1_Beta/1-388 GACGGTACCTGAAGAATAAGCACCGGCTAACTACGTGCCAG
                               |||||||||||||||||||||||||||||||||||||||||
   Leptothrix_ochracea/324-711 GACGGTACCTGAAGAATAAGCACCGGCTAACTACGTGCCAG

ASV_0004.Allander_1_Beta/1-388 CAGCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATT
                               |||||||||||||||||||||||||||||||||||||||||
   Leptothrix_ochracea/324-711 CAGCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATT

ASV_0004.Allander_1_Beta/1-388 ACTGGGCGTAAAGCGTGCGCAGGCGGTTATACAAGACAGAT
                               ||||||||||||||||||||||||||||||||||||||||
   Leptothrix_ochracea/324-711 ACTGGGCGTAAAGCGTGCGCAGGCGGTTATACAAGACAGAG

ASV_0004.Allander_1_Beta/1-388 GTGAAATCCCCGGGCTCAACCTGGGAACTGCATTTGTGACT
                               |||||||||||||||||||||||||||| ||||||||||
   Leptothrix_ochracea/324-711 GTGAAATCCCCGGGCTCAACCTGGGAACTGCCTTTGTGACT

ASV_0004.Allander_1_Beta/1-388 GTATAGCTAGAGTACGGTAGAGGGGGATGGAATTCCGCGTG
                               |||||||||||||||||||||||||||||||||||||||||
   Leptothrix_ochracea/324-711 GTATAGCTAGAGTACGGTAGAGGGGGATGGAATTCCGCGTG

ASV_0004.Allander_1_Beta/1-388 TAGCAGTGAAATGCGTAGATATGCGGAGGAACACCGATGGC
                               |||||||||||||||||||||||||||||||||||||||||
   Leptothrix_ochracea/324-711 TAGCAGTGAAATGCGTAGATATGCGGAGGAACACCGATGGC

ASV_0004.Allander_1_Beta/1-388 GAAGGCAATCCCCTGGACC
                               |||||||||||||||||||
   Leptothrix_ochracea/324-711 GAAGGCAATCCCCTGGACC

Percentage ID = 99.23
```

Figure 47 – 16S rRNA sequence identity alignment of Allander *Paucibacter* and *L.ochracea*.

This result also highlights the limitation of 16S rRNA sequencing of environmental samples as it is challenging to accurately assign taxonomic distinctions at the species and also sometimes the genus level. To accurately assign bacteria at the species level from environmental samples it is appropriate to combine various techniques and select primers that target genes specific to the bacteria of interest. Fleming *et al.* used primers and probes, shown in Table 19, that targeted the V4 hypervariable region of the 16S rRNA gene to build a pyrosequencing library.

Probe/primer	Sequence (5′–3′)
EUB338	cy3 gct gcc tcc cgt agg agt
NON338	cy3 act cct acg gga ggc agc
BET 42a	cy3 gcc ttc cca ctt cgt tt
GAM42a	Fl gcc ttc cca cat cgt tt
PS-1	cy3 gat tgc tcc tct acc gt
Lepto 175	cy3 atc cac aga tca cat gcg
V4-F	ayt ggg ydt aaa gng
V4-R1	tac nvg ggt atc taa tcc
V4-R2	tac crg ggt htc taa tcc
V4-R3	tac cag agt atc taa ttc
V4-R4	tac dsr ggt mtc taa tcn
27F	agr gtt yga tym tgg ctc ag
907R	ccg tca att cmt ttr agt tt
1492R	tac ggy tac ctt gtt acg act t
gyrB F	kcg caa gcg scc sgg cat gta
gyrB R	ccg tcs acg tcg gcr tcg gtc at

Table 19 - Sequences of primers and probes utilised by Fleming *et al*. to target the V4 hypervariable region of the 16S rRNA gene.

This library was then used to develop a fluorescent *in situ* hybridisation (FISH) probe specific to *L.ochracea* denoted Lepto 175. This probe binds the corresponding region of the 16S rRNA gene thereby allowing *L.ochracea* cells within BIOX mats to be fluorescently labelled, quantified and characterised.[23] A similar methodology should be applied to the Allander sample sites in future.

Another notable sequence is the 2nd most abundant 16S rRNA sequence in the Allander_2 sample. This corresponds to a *Phormidiaceae* spp. which is a filamentous cyanobacterium. The presence of a highly abundant cyanobacterium sequence agrees with visual descriptions of this sample site. The Allander_2 sample site regularly contains green algae like blooms near the water's surface and when viewed by microscopy large filamentous cyanobacteria can be seen. When BIOX mats were treated with sodium dithionite to remove the insoluble ferric compounds these large filamentous cyanobacteria became more apparent. Figure 48 contains phase contrast micrographs of these sodium dithionite exposed filaments and photographs of the Allander_2 sample site with green algal blooms visible.
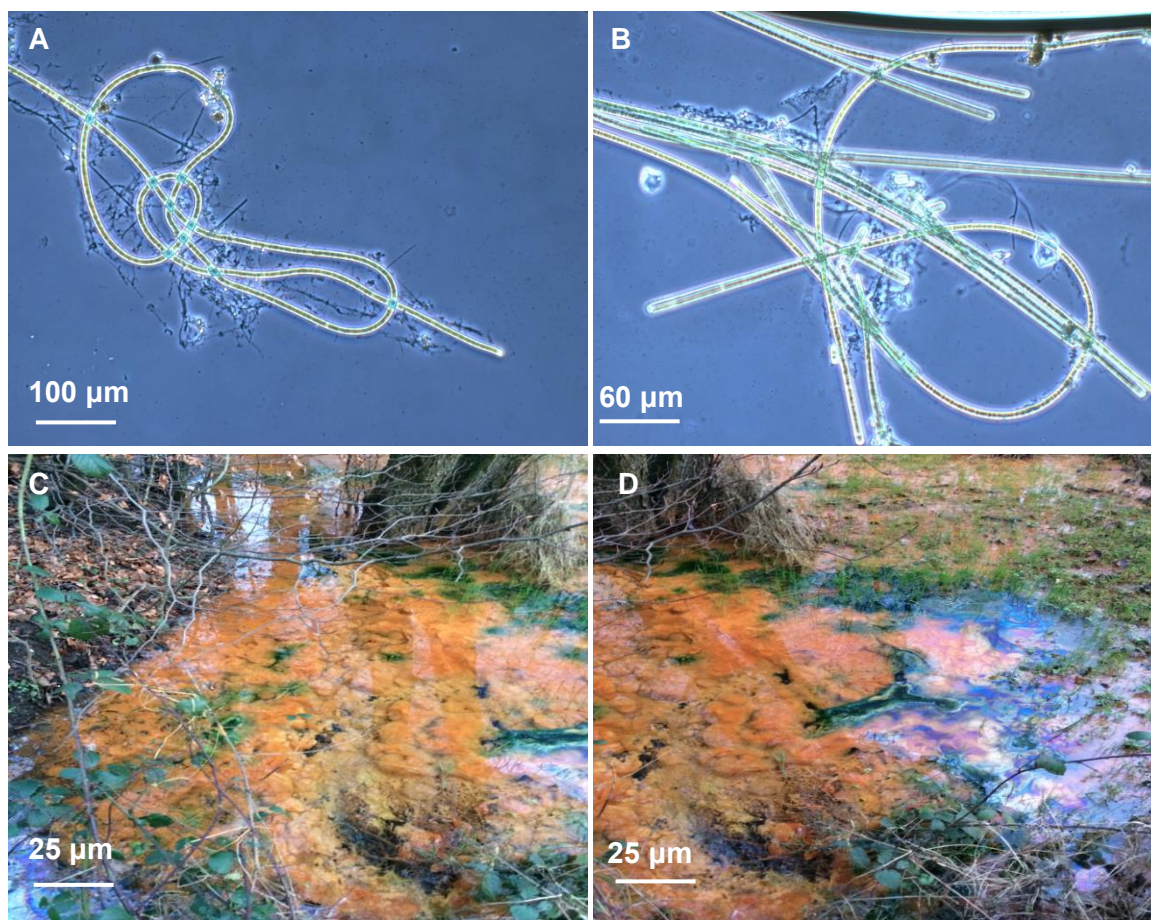
Figure 48 – A and B - Phase contrast micrographs of *Phormidiceae* filaments associated with a sodium dithionite treated BIOX mat. C and D – Photographs of Allander_2 sample site with green algal growths seen.

The cyanobacteria filaments are much larger than *L.ochracea* BIOX filaments, measuring ca. 20 μm in diameter, and extending for millimetres. Current research suggests that cyanobacteria in marine environments may associate themselves with BIOX mats as a means of relieving oxidative stress as reactive oxygen species will react with the ferric iron compounds.[24] This is also hypothesised to be a syntrophic relationship as the cyanobacteria can produce oxygen which may be utilised by bacteria within the mats.[25] This has not been researched in freshwater BIOX mats however the fact that these *Phormidaceae* spp. are regularly associated with these BIOX mats suggests that a similar mechanism may be occurring.

4.4.5 Abundance of Isolated Filamentous Bacteria

The results from Chapter 5 show that filamentous bacteria can be isolated from BIOX samples via a micromanipulator combined with a phase contrast microscope. Four filaments were isolated, two from the Dawsholm sample site and two from Allander_3.

These are denoted Daws_1, Daws_2, Al_1 and Al_2. Figure 49 shows micrographs of the isolated filaments.
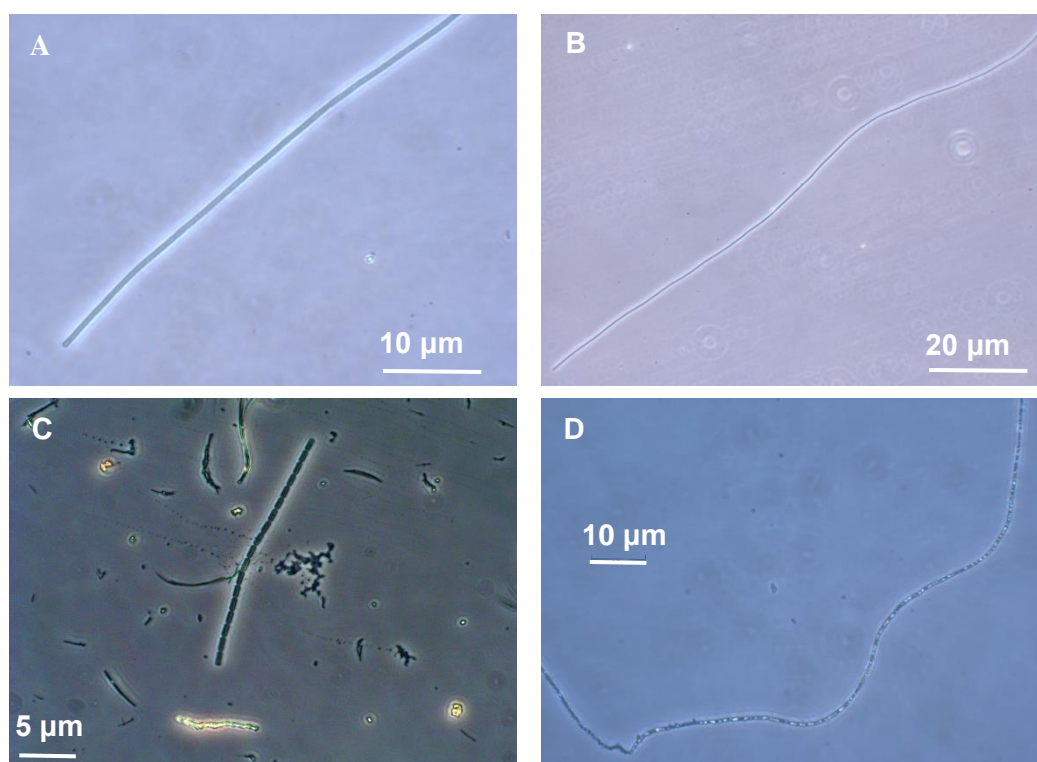
Figure 49 – Phase contrast micrographs of isolated filaments. A – Daws_1, B – Al_1, C – Daws_2 and D – Al_2.

These isolates had their genomic DNA amplified, in accordance with the protocol in Chapter 2 Section 7, and were sent for full genome sequencing. Unfortunately, the Al_2 isolate was not fully sequenced and instead returned a very large number of relatively short reads (ca. 1000 bp in length) that included a proportion of contaminating bacteria sequences. The predominate longer reads correspond to an *α*-proteobacteria from the *Caulobacteraceae* family of bacteria of genus either *Caulobacter or Phenylobacterium.* Unfortunately, no 16S genomic DNA was amplified that could have identified it further. It is possible that contamination from the micromanipulator may have been carried over to the PCR step and subsequently been amplified. It is therefore not sensible to predict the identity of the bacterium Al_2. The other three isolates were sequenced efficiently and their 16S rRNA sequences were searched against the Allander_1, Allander_2 and Allander_3 libraries in an attempt to confirm their presence within the BIOX mats and also to investigate their potential abundance.

As all three of the isolated bacteria are *γ*-proteobacteria a FASTA file containing the 16S rRNA sequences of the 20 most abundant *γ*-proteobacteria from each sample site was

created. Sequences were aligned using Jalview and MEGA-X was used to create a maximum likelihood phylogenetic tree which can be seen in Figure 50.
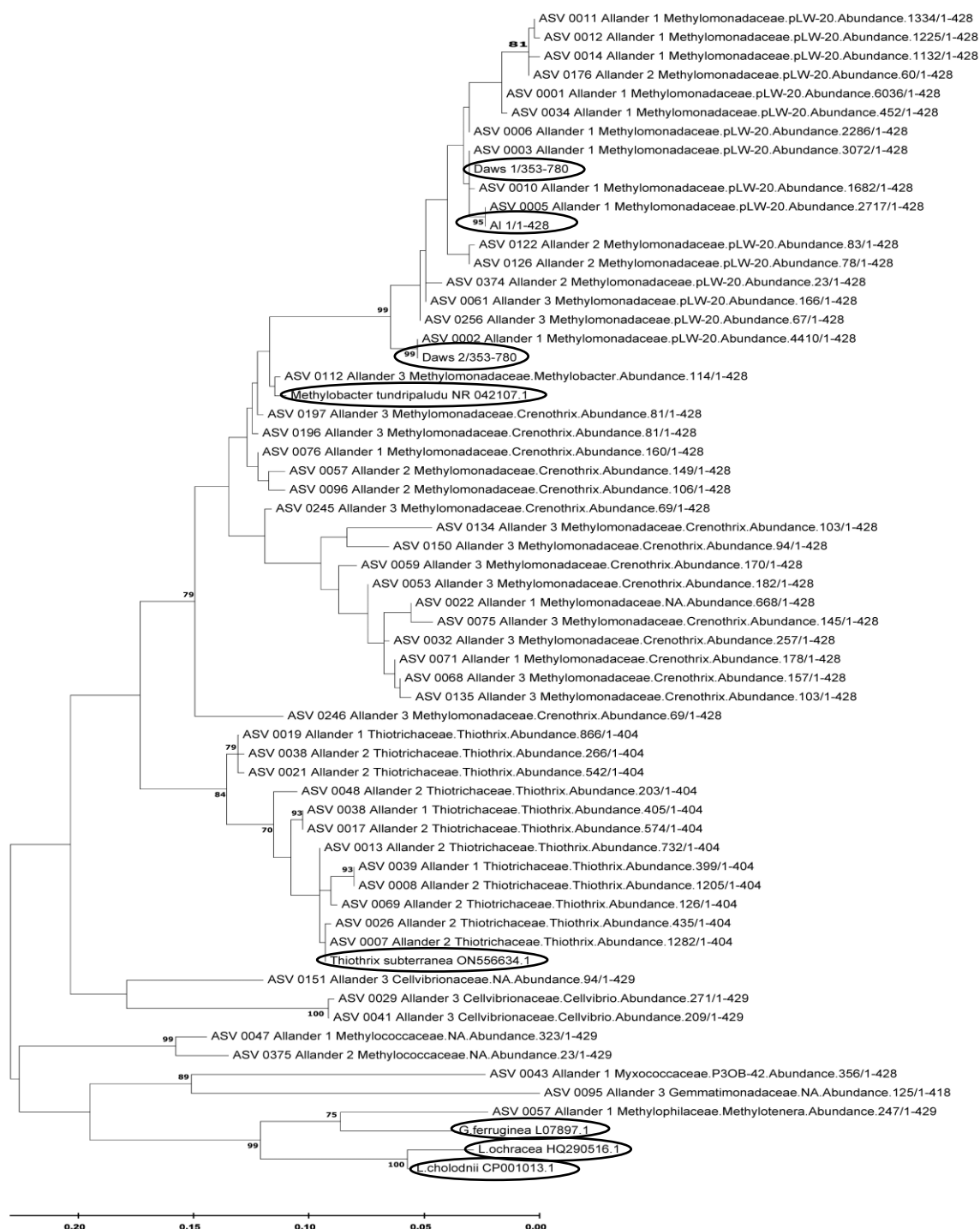
Figure 50 - Phylogenetic maximum likelihood tree of the top 20 *γ*-proteobacteria from each of the three sample sites. Daws_1, Daws_2 and Al_1 (circled) are included to show which bacteria from the library that they are most closely related to. A selection of closely related strains (circled) from the NCBI BLAST database have also been included. Scaler bar = Number of nucleotide substitutions per site, this is the number of nucleotide substitutions divided by the length of the sequence. Bootstrap values >70 are shown.

Sequence percentage identities over 428 bases were calculated using Jalview. Daws_1 and Al_1 share 99.30 % sequence identity while Daws_1 and Daws_2 share 96.26 % and Al_1 and Daws_2 share 95.58 %. These differences in percentage identity are visualised in the tree as Daws_1 and Al_1 are clustered more closely to each other than Daws_2. The lines joining the isolated filamentous bacteria with γ-proteobacteria from the library are flat with no edges at either end indicating that these bacterial sequences are identical. Individual pairwise comparison sequence calculations were carried out using Jalview to confirm this. Table 20 shows each isolated bacteria, their nearest neighbour, their library assignment and their abundance rank in Allander_1.
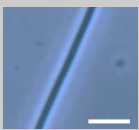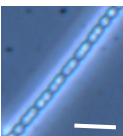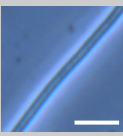
Isolated Bacterium	Nearest Relative (γ-proteobacteria number)	Nearest Relative (Library Family and Genus assignment)	Abundance Rank
Daws_1	Allander_1_ γ-proteobacteria_2	*Methylomonadaceae* pLW-20	3rd
Daws_2	Allander_1_ γ-proteobacteria_3	*Methylomonadaceae* pLW-20	2nd
Al_1	Allander_1_ γ-proteobacteria_4	*Methylomonadaceae* pLW-20	6th

Table 20 - Alignment of each sequenced filamentous bacterium with nearest relative from 16S rRNA library. Scale bars = 3 µm.

The alignment data shows that the 2nd, 3rd and 6th most abundant bacteria in the Allander_1 site have been isolated. This confirms that the micromanipulation was effective at sampling bacteria which were abundant within the BIOX mats, in contrast with the other isolation methods used throughout Chapter 5. All three isolates are from the family *Methylomonadaceae* and are likely different species of the genus pLW-20. pLW-20 have

previously been described in a study by Nercessian *et al.* who were investigating methane metabolising bacteria populations in Lake Washington however they have never been reported as filamentous.[26] No species classification or culturing attempts have been made to the author's knowledge.

It was initially surprising that all three isolates appear only abundantly in the Allander_1 16S library even though they were not isolated from there. This may be due to environmental changes as the samples used to create the library were collected on 6/1/20 while the filamentous bacteria were isolated from Dawsholm samples collected on 29/1/21 and Allander_3 samples collected on 29/3/21. As discussed in Chapter 5 Section 6, the Dawsholm BIOX was initially used to isolate filamentous bacteria as during this period of the study (December/January 20/21) the Allander sample sites were not producing high quality filamentous BIOX. The Dawsholm site contained masses of filamentous bacteria that were continually associated the BIOX. It was not until late March 2021 that the Allander_3 began to show filamentous BIOX again, highlighting the transient nature of these sample sites. Fleming *et al.* reported that BIOX mats collected from the same sample site can have different bacterial profiles depending on the season.[27] This is likely due to environmental conditions such as temperature, rainfall, flow, nutrient availability and pH. The same logic can be applied to the sample sites used for this study and it can therefore be assumed that the isolated bacteria will likely be found in Allander_2 and Allander_3 at other points in the year. Regular 16S rRNA sequencing throughout the year would be required to confirm this and to allow a true comparison between sites. This is also an important factor to consider when making comparisons to other environmental 16S rRNA sequence studies in the literature.

β-proteobacteria were selected as an outgroup for this *γ*-proteobacteria tree. Surprisingly, Allander 1 ASV 0057 is more closely related to *G.ferruginea* than anything else in the dataset. A Jalview pairwise analysis shows them to share a 90.02% sequence identity. This was not investigated further due to time constraints.

4.5 Comparing Sample Site Profiles with Literature Studies

The 16S rRNA sequences of the prevalent bacteria from the Allander sample sites were searched using nucleotide Basic Local Alignment Search Tool (BLAST) against the National Centre for Biotechnology Information (NCBI) Nucleotide Collection and the

Ribosomal Database Project (RDP) databases. Results from these searches could be used to identify other 16S environmental studies from related BIOX sample sites worldwide. The full studies were searched for based on the paper reference or authors in the European Nucleotide Archive (ENA). This is a repository where sequencing data from literature and unpublished studies is stored and can be freely accessed. The study sequences can be downloaded together either as a raw sequence format so that the sequence abundance could be calculated, or more commonly as a collection of unique sequences from the site with no abundance information. Studies related to BIOX sampled from ferrous iron rich groundwater sites were selected for comparison with this study. There are other BIOX phylogeny studies published that relate to sample sites that prove to be quite different in bacterial distribution and environmental conditions from those in this study. These include the Arctic tundra, paddy field soil, hydrothermal vents and deep freshwater lakes and have not been compared here.[28-31]. The studies selected for comparison are found in Germany, America, Canada, Japan and Argentina respectively.[23, 32] No data was available for BIOX samples collected in the greater Glasgow area or indeed the United Kingdom highlighting the novelty of this study. A map of the world labelled with the locations of the aforementioned sample sites can be seen in Figure 51.
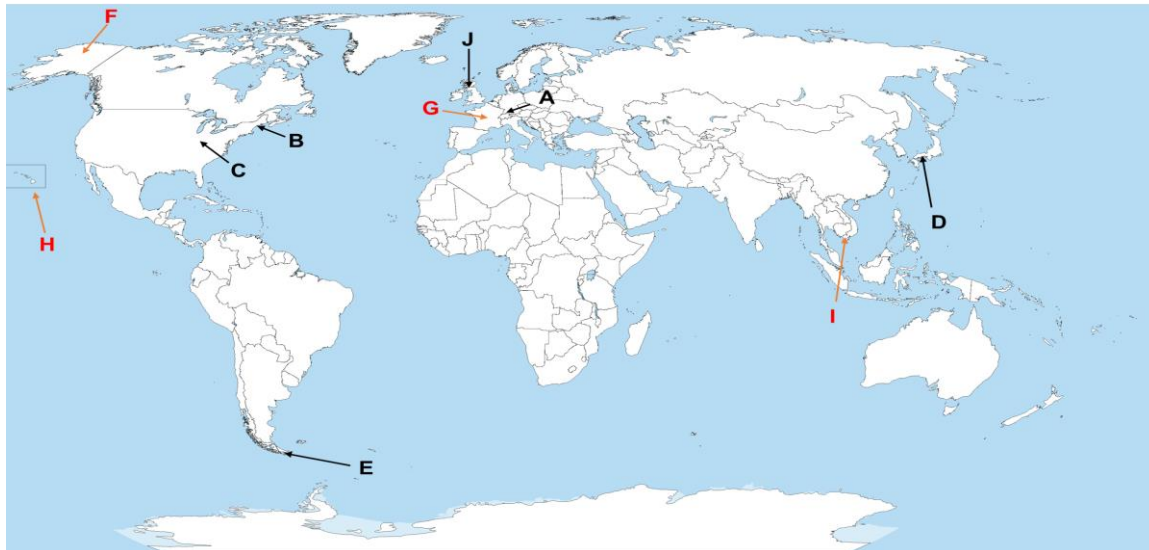
Figure 51 - Map of the world highlighting the location of BIOX sample sites that have been profiled phylogenetically. Arrows coloured black indicate sample sites compared with this study while arrows coloured red indicate BIOX sample sites growing under different environmental conditions than those used for this study. A – Hoffnungsstollen mine (47 °43.8' N, 8' ° 0' E), B – Lakeside Drive (43 ° 51.699′ N, 69 ° 38.929′ W), C- Jackson Creek (39 ° 8.1783' N, 86 ° 30.36' W), D – Budo Pond (34 ° 24.06' N, 132 ° 42.79' E), E- Tierra Del Fuego (54 ° 46.592' S, 67 ° 41.944' W), F – Toolik Lake (68 ° 37.8' N, 149 ° 36' W), G – Lake Pavin (45 ° 29.7508' N, 2 ° 52.2753' E), H - Lō'ihi Seamount (18 ° 55.2' N, 155 ° 16.2' W), I – Mekong Delta (9 °41.88′N, 105 °18.65′ E) and J – Greater Glasgow area.[23, 28-35]

The sequencing data for each site was downloaded and analysed using a combination of R scripts utilising the DADA2 pipeline with the Silva nr99 v138 train set of 16S sequences used for phylogenetic identification. The programs Jalview and MEGA-X were used to align and truncate sequences and for the construction of phylogenetic trees. The γ- and β-bacterial profiles from the Allander sample sites can then be compared with similar studies to look for potential novelty and similarity between the phylogenetic profiles of BIOX samples globally.

4.5.1 Description of Literature Study Sample Sites Based on Location

Figure 52, Figure 53 and Figure 54 show the phylogenetic diversity of the three Allander sites to the order level. These shall be used for comparison with literature studies in this section.
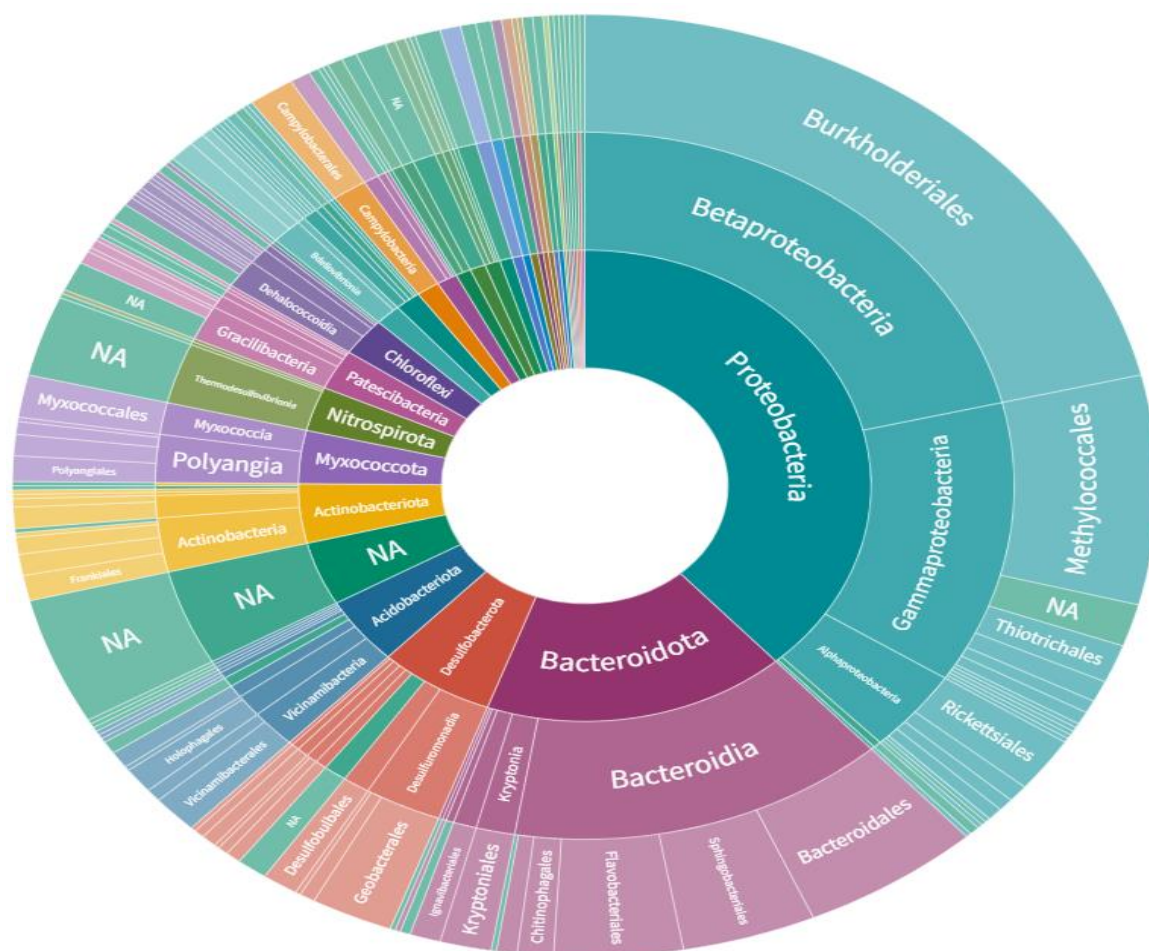
Figure 52 - Phylogenetic diversity of Allander_1 to the order level as a fraction of 688 ASV.
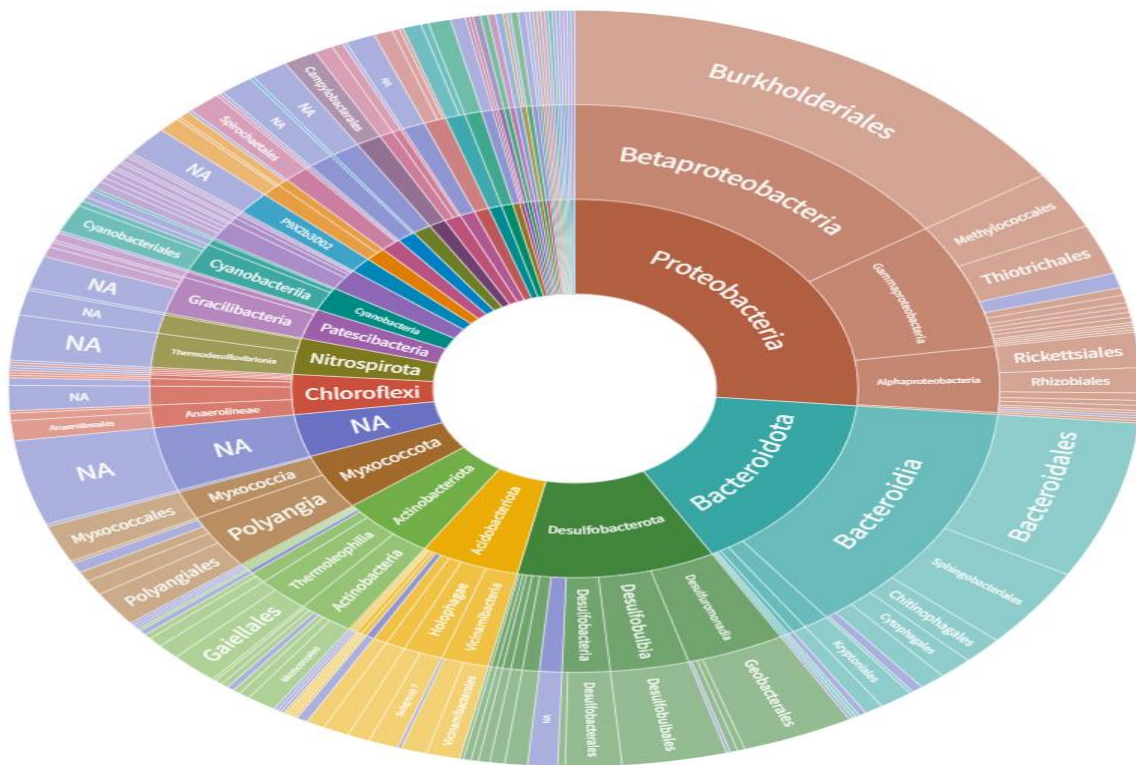
Figure 53 - Phylogenetic diversity of Allander_2 to the order level as a fraction of 937 ASV.
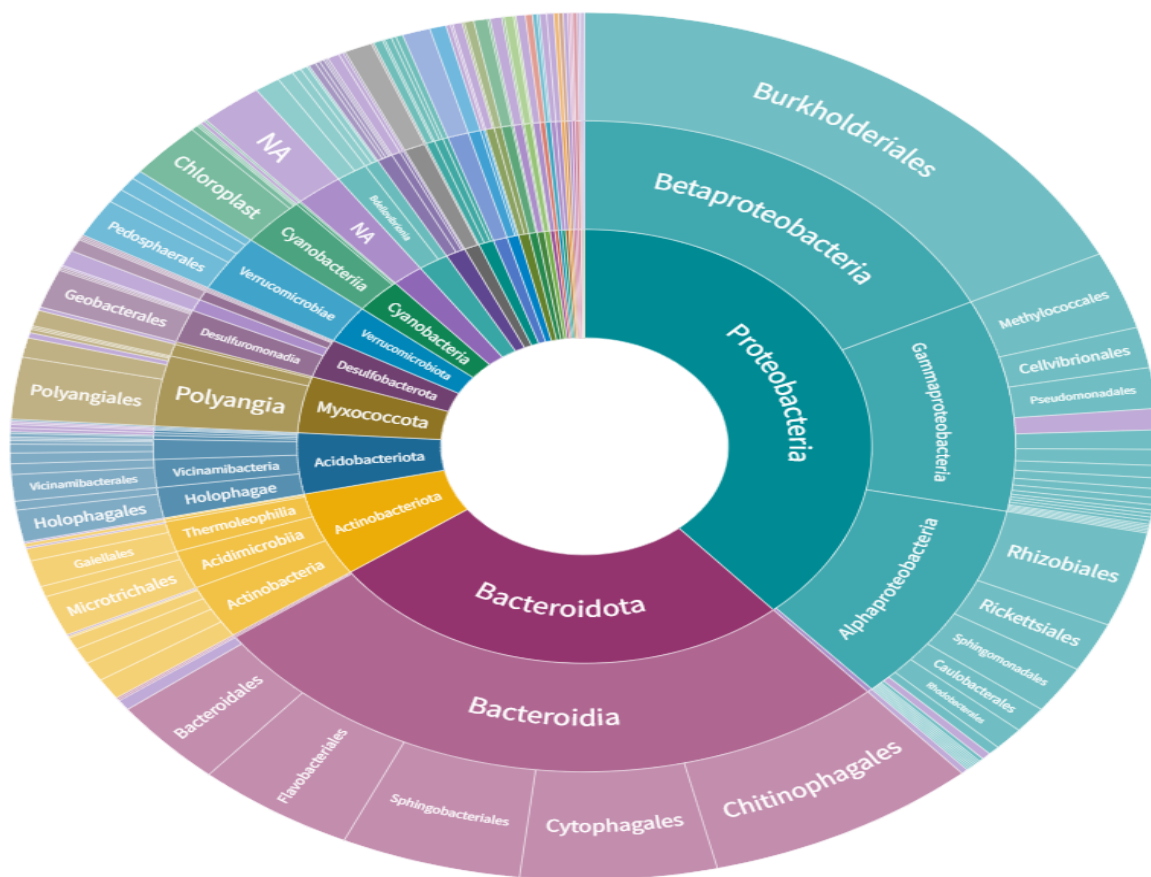
Figure 54 – Phylogenetic diversity of Allander_3 to the order level as a fraction of 1521 ASV.

4.5.1.1 Germany

BIOX samples collected from Hoffnungsstollen mine in the Todtmoos village, in the very south-west of Germany was compared with the Allander samples.[32] The sequencing data is available through the ENA however the study has not been published meaning that the sample site is only briefly described as a microbial mat on a tunnel wall and there is no information regarding the sequencing methodology used. The sequence lengths are ca. 1300 – 1400 bp in length and no abundance data is given. This suggests that high-throughput Illumina sequencing was not used, instead a technology such as Pacific Bioscience RS sequencing may have been used, however it is difficult to confirm. This is a sequencing technology that generates long reads of ca. 1300 bp and is generally used to generate smaller libraries.[36] The Hoffnungsstollen library size here is only 343 ASV.

Fortunately, the Hoffnungsstollen mine is an exhibition mine open to the public and publishes online promotional material. Figure 55 contains a photograph from inside the Hoffnungsstollen mine. Orange deposits can be seen on the tunnel wall which at the very least indicates the type of sample found at this location.

Figure 55 - Photograph from inside the Hoffnungsstollen mine.

Figure 56 contains a diagram showing the phylogenetic diversity of the Hoffnungsstollen mine sample site to the order level.
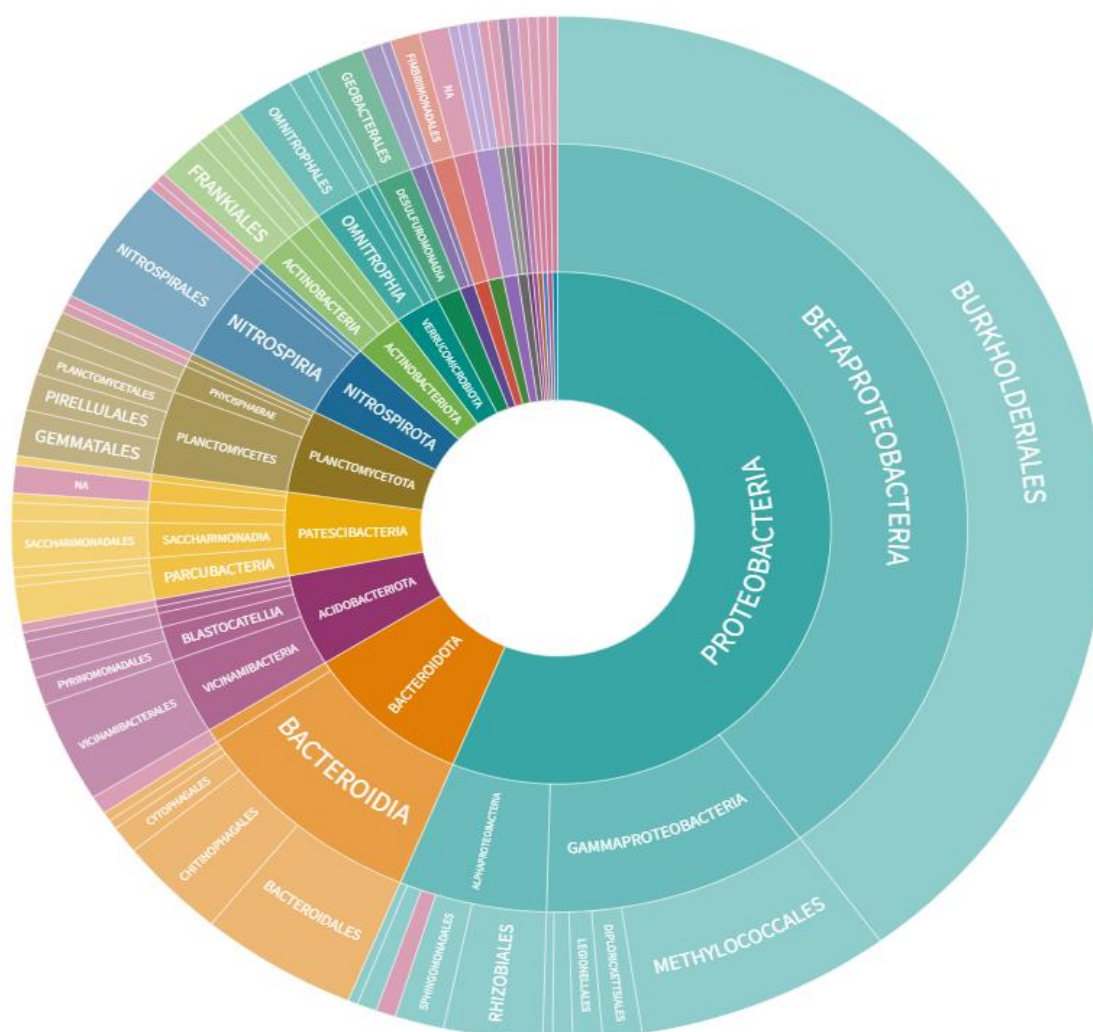
Figure 56 - Phylogenetic diversity of the Hoffnungsstollen mine sample site to the order level as a fraction of 343 ASV.

4.5.1.2 America

BIOX samples collected from Lakeside Drive in Boothbay Harbor, Maine, and Jackson Creek in Bloomington, Indiana were compared with the Allander samples.[23, 33] The Lakeside Drive study was carried out using 454 pyrosequencing technology targeting only the V4 hypervariable region. This is a sequencing methodology that was prominent prior to the advent of high-throughput Illumina sequencing and has now been discontinued.[37] Care must be taken when comparing results achieved from different sequencing technologies due to primer biases and differing error rates. Luo *et al* compared 454 pyrosequencing with Illumina sequencing and found 454 pyrosequencing to recover 14% fewer complete genes. This was attributed to higher sequencing error rates associated with A- and T- rich homopolymers.[37] Luo *et al* also found that Illumina sequencing generally yielded similar assemblies to 454 pyrosequencing, meaning that the following

comparison is not unwarranted, so long as the reader remains vigilant to the fact that different sequencing technologies may yield slightly differing profiles. Another caveat that the reader should be aware of is that Roden *et al* found that comparing clone libraries to high-throughput sequencing data from the same samples yielded differing results.[33] As such, the following comparisons must be treated conservatively, and in future the Allander sites should also be sampled and sequenced in line with the comparative studies for a more representative comparison.

The Jackson Creek study combined two sequencing methodologies to assign their phylogenies. 454 pyrosequencing targeting the V6 and V4 hypervariable regions was used in parallel with Illumina sequencing targeting the V4 and V1 regions.[33] This allowed replication of results and increased the reliability of the data. Taxonomy was assigned using the Silva database, which was used for the author's study, and the RDP database.

The Lakeside Drive site is described as an intermittent stream while the Jackson Creek site is described as a ditch site. Photographs from these sites can be seen in Figure 57 and indicate that the Jackson Creek site is visually similar in nature with the Allander sites while the Lakeside Drive site appears to contain a much larger bloom of BIOX than has typically been seen at the Allander sites.
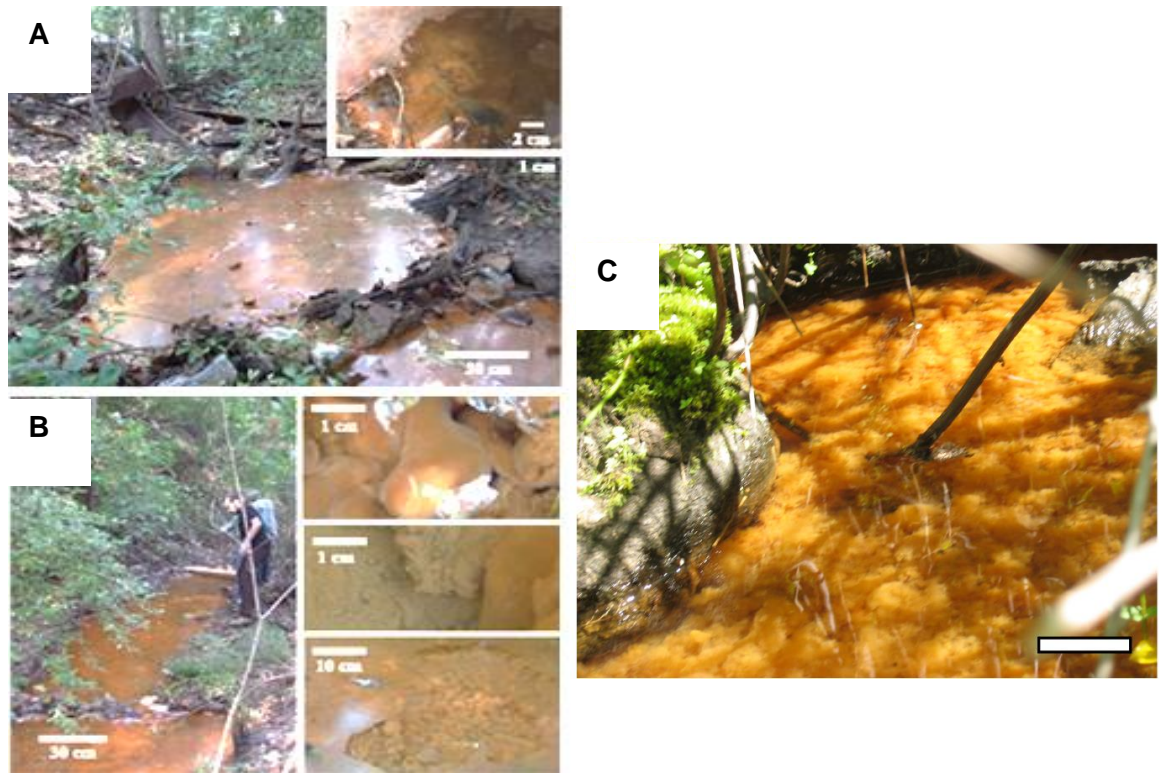
Figure 57 - Photographs of A and B - Jackson Creek sample site and C - Lakeside Drive Sample site (scale bar = 10 cm).[23, 33]

Figure 54 contains a diagram showing the phylogenetic diversity of the Jackson Creek sample site while Figure 59 contains a diagram showing the phylogenetic diversity of the Lakeside Drive sample site.

Figure 58 - Phylogenetic diversity of Jackson Creek (142 ASV) to the order level.

Figure 59 – Phylogenetic diversity of Lakeside Drive (106 ASV) to the order level.

4.5.1.3 Canada

BIOX samples collected from Meilleurs Bay on the south bank of the Ottawa River were compared with Allander samples.[38] This site is described as a groundwater spring, 5 m in length, that is rich in dissolved ferrous iron and flocculent BIOX mats. Samples were collected from three distinct sections at approximately 0 cm, 300 cm and 450 cm respectively. Figure 60 contains photographs of BIOX growing at this sample site. This study used the construction of a clone library combined with Sanger sequencing to produce a library of sequences which were then assigned OTUs against the Greengenes database.[39] Sanger sequencing is an older method of DNA sequencing that is less frequently used today. It has the ability to produce longer DNA reads of > 500 bp however is generally used for small scale studies, not requiring high-throughput, or for confirmation of high-throughput sequencing results.[40] Sanger sequencing does not provide a number of ASV. Figure 61 contains diagrams highlighting the phylogenetic diversity of the three Meilleurs Bay sample sites to the order level.
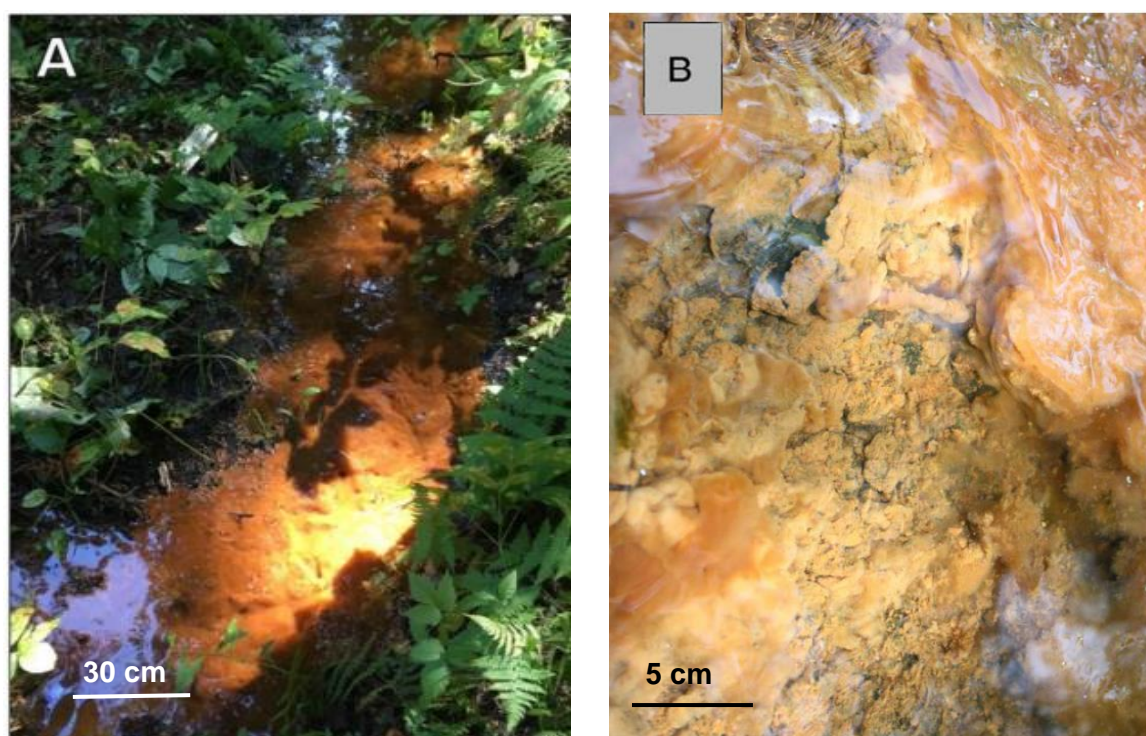
Figure 60 - Photographs of Meilleurs Bay BIOX taken from Edwards.[41]
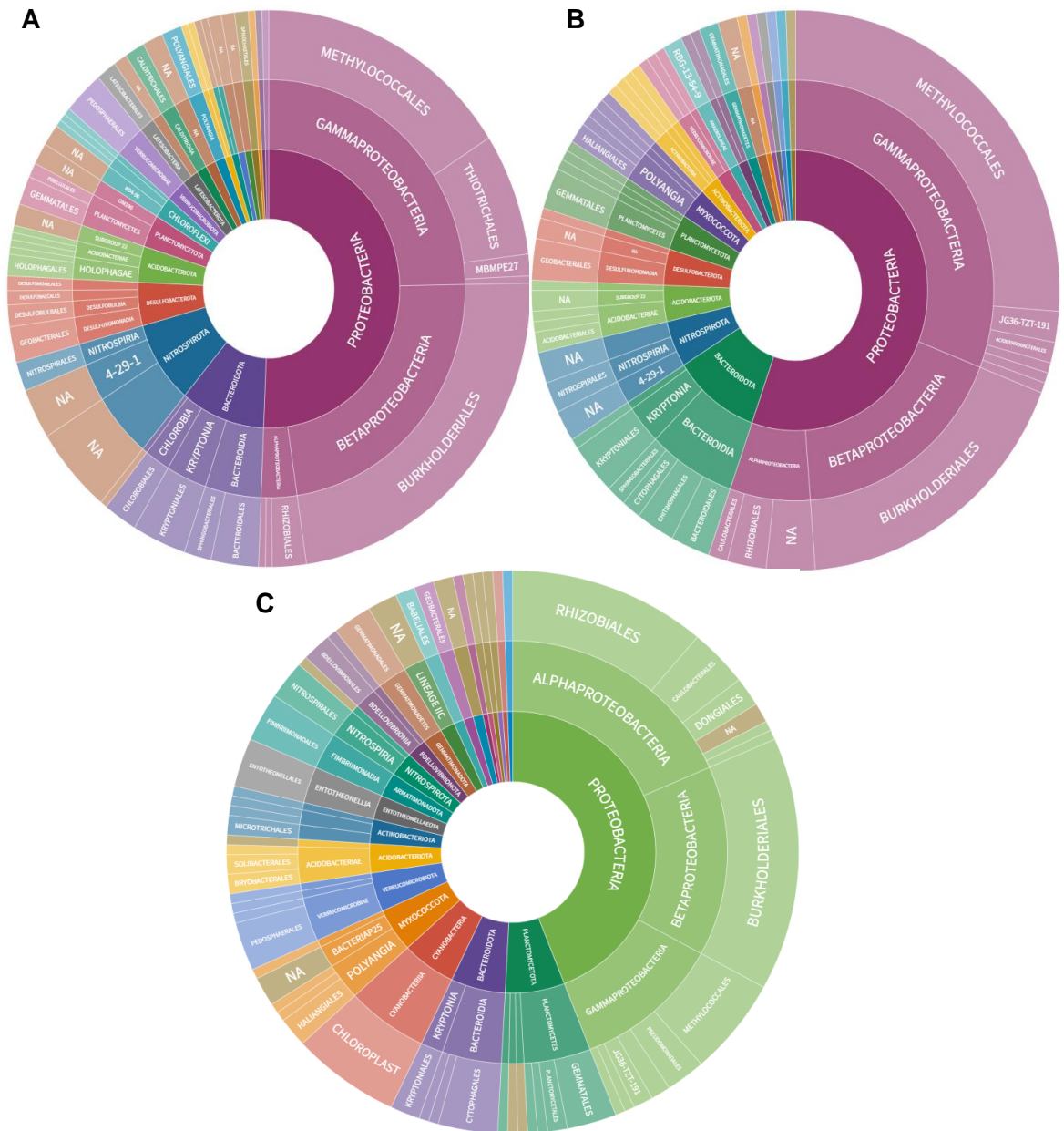
Figure 61 - Phylogenetic diversity of A - Meillerus Bay 0 cm (239 ASV), B - Meillerus Bay 300 cm (138 ASV) and C - Meilleurs Bay 450 cm (179 ASV) sample sites to the order level.

4.5.1.4 Japan

BIOX samples collected from Budo Pond in Hiroshima were compared with Allander samples.[34] This site is described as a groundwater discharge point with organic rich sediment present at the bottom of the pond. BIOX mats are described as 30 cm deep and more abundant during the summer.

Figure 62 - Photograph of BIOX sample in Budo Pond.[34]

This site appears visually quite different from the Allander sites as the BIOX looks less gelatinous and instead is more flocculent in appearance. This may be indicative of a slightly different phylogenetic profile.

The 16S rRNA sequencing at this site involved the construction of a clone library, quantitative PCR and assignment of taxonomy via the Mothur open-source software package.[42] As such, there was no abundance data generated. However, bacteria and archaea sequences are identified that differ from other studies compared here.

Figure 63 contains a diagram showing the phylogenetic diversity of this sample site to the order level.
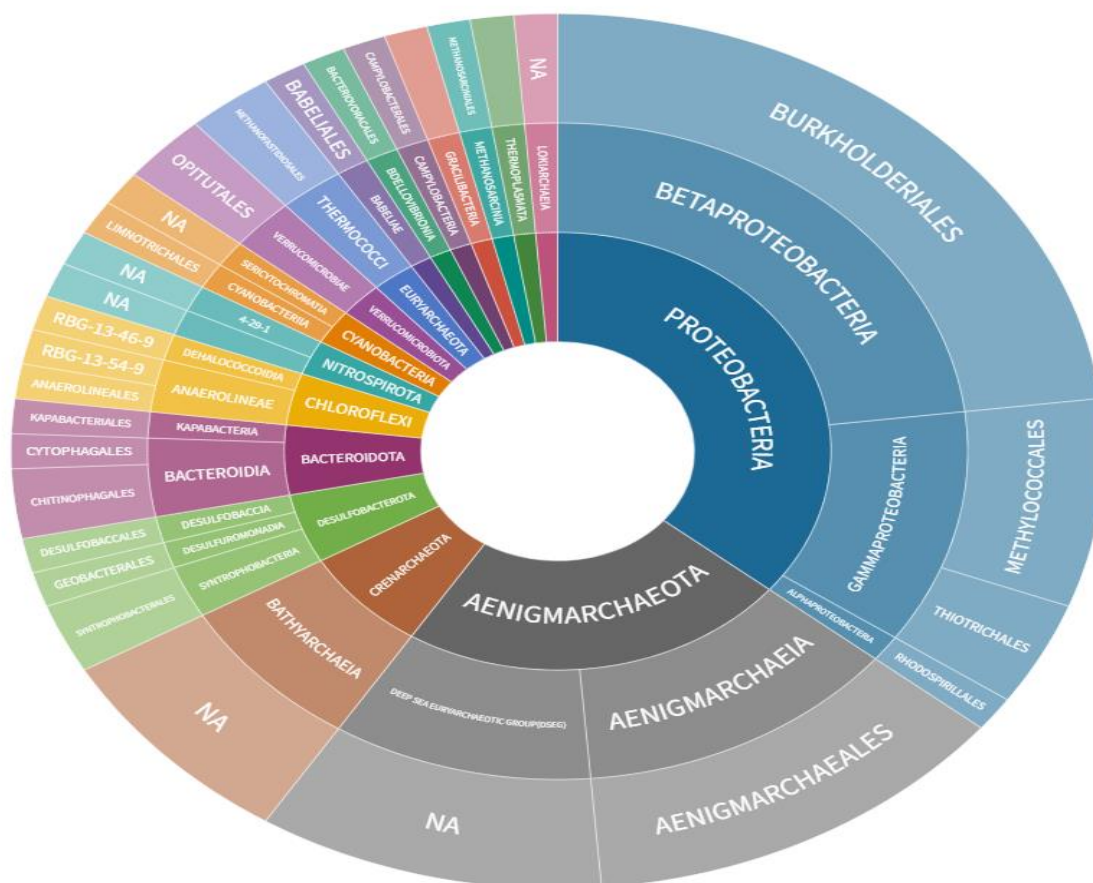
Figure 63 - Phylogenetic diversity of the Budo Pond site (78 ASV) in Hiroshima to the order level.

4.5.1.5 Argentina

BIOX samples were collected from saturated soils, biofilms growing on plants and rocks and flocculent material from ponds in the southern part of Tierra Del Fuego Island, Argentina.[35] These samples were used to isolate multiple bacteria from the *Sphaerotilus-Leptothrix* genus. This study is different from the others mentioned as it investigates bacteria isolated from BIOX rather than constructing an entire phylogenetic profile however it is still relevant as it allows a comparison specifically of the Allander *Leptothrix* bacteria with another sample site. A variety of sample types were taken for this study including BIOX mats, saturated soil, and sediment. Figure 64 shows photographs from this sample site. A diagram highlighting phylogenetic diversity for this site is not included as the study is interested in only the *Sphaerotilus-Leptothrix* genus.
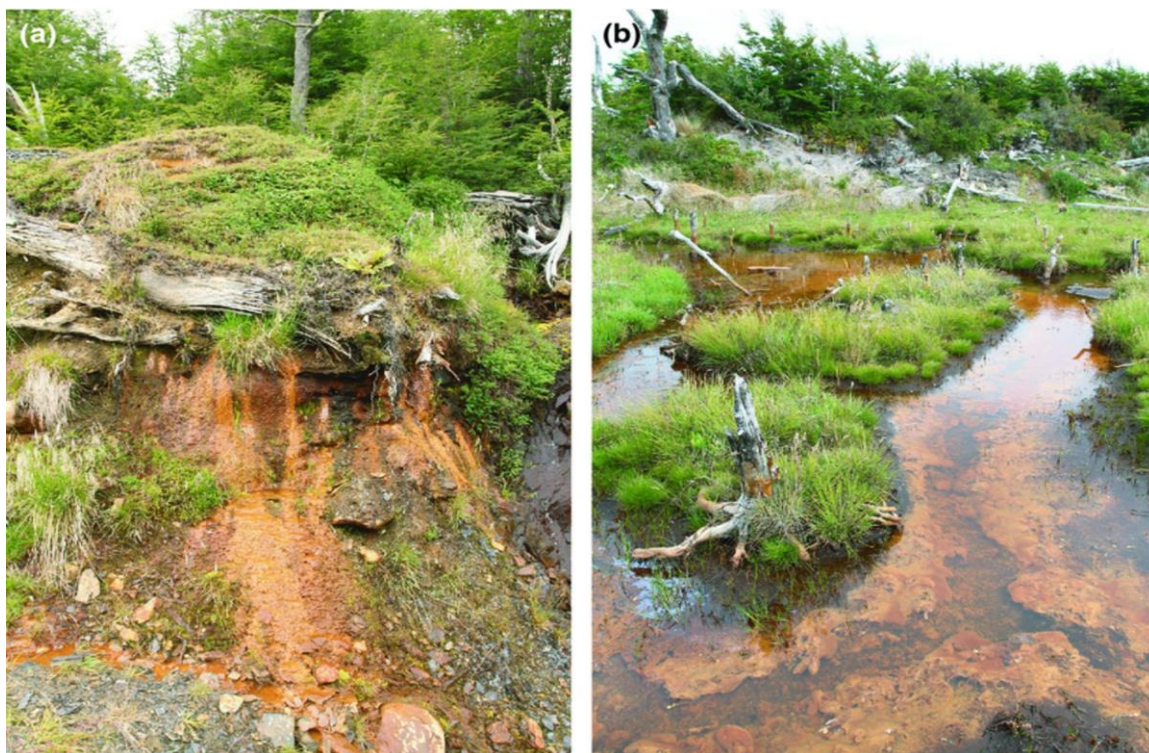
Figure 64 - Photographs from Tierra Del Fuego sample site. A - sediment sampled and B - BIOX mats sampled.[35]

4.5.2 *γ*-proteobacteria: *Methylomonadaceae* pLW-20 Sequences

The most prevalent family of *γ*-proteobacteria found in the Allander sample sites is the *Methylomonadaceae* family of which there are 108 unique sequences found across all three sites. This family contains the pLW-20 genus, which includes the three isolated filamentous bacteria, and has 23 unique sequences in Allander_1, 9 in Allander_2 and 8 in Allander_3 respectively. This suggests that Allander_1 may contain the highest concentration of methane as these bacteria are obligate methanotrophs. The 16S rRNA sequences of the pLW-20 bacteria found in the three Allander were added to a FASTA file containing the 16S rRNA sequences of pLW-20 bacteria found across the relevant studies. These were then truncated to only compare the V3 – V4 hypervariable region and a maximum likelihood phylogenetic tree was constructed. This phylogenetic tree can be seen in Figure 65.
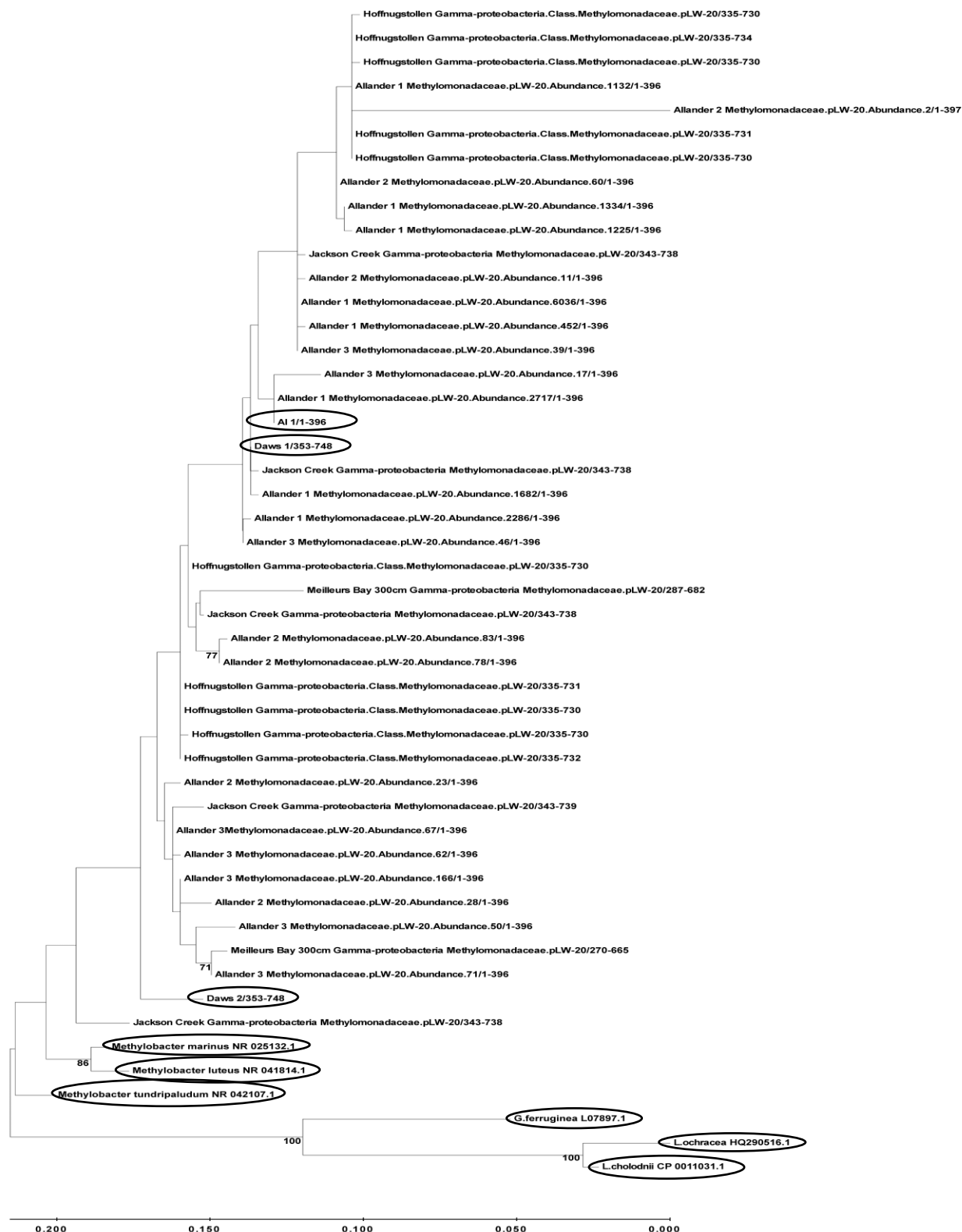
Figure 65 - Maximum likelihood phylogenetic tree of all the pLW-20 16S rRNA sequences found in the Allander datasets compared with the literature studies. Isolates and reference strains from NCBI Blast database have been circled. Scale bar = number of nucleotide substitutions per site, and bootstrap values >70 are show.

This tree shows that pLW-20 genus of bacteria was not reported in the Hiroshima sample however were reported in the other studies. There is overlap between all three Allander sites and the reported literature studies however none of the Allander pLW-20 bacteria share 100 % sequence identity across 404 bases with any of the literature examples indicating that the sites contain related but not identical bacteria. The Daws1 isolate lies in a small cluster of three Allander bacteria and one Jackson Creek pLW-20 bacterium. It shares a 99.75 % identity with the Jackson Creek pLW-20 bacterium and also shares 99.75 % identity with the 6th most abundant pLW-20 bacterium from Allander_1. The Al1 isolate is in another small cluster with two other Allander pLW-20 bacteria, one from Allander_1 and one from Allander_3. It shares a 100 % identity with the 4th most abundant Allander_1 pLW-20 bacterium and a 98.48 % identity with the 8th, and least abundant, pLW-20 bacterium from Allander_3. The Daws2 isolate lies in a larger cluster than the other two isolates and shares a ca. 97 % identity with its three closest neighbours which are all from Allander_3. It is more distantly related to three bacteria from Jackson Creek (95.99 %), Lakeside Drive (95.99 %), and Meilleurs Bay (95.75 %).

$β$-proteobacteria were selected as an outgroup for this $γ$-proteobacteria tree. It can be seen that these appear to be distinct from the rest of the dataset. Three isolated *Methylobacter* sequences selected from the NCBI Blast database were also included as references. These, however, appear to be more closely related to one another than any of the other bacteria found within the dataset.

4.5.3 *β*-proteobacteria

The genera of $β$-proteobacteria with the greatest number of unique 16S rRNA sequences in the Allander datasets are the *Gallionella*, *Rhodoferax* and *Sideroxydans*. The genus *Rhodoferax* contains iron reducing bacteria (FeRB) while the other mentioned genera are all FeOB. Interestingly the *Leptothrix* genus contributes minimally to the phylogenetic profile. These five genera will be compared with the above literature studies to assess whether or not globally distributed BIOX mats contain similar profiles of $β$-proteobacteria.

4.5.3.1 *Leptothrix/Paucibacter*

As the 16S rRNA sequence of a *Paucibacter* bacterium found in all Allander sites has a 99.23 % sequence identity with *L.ochracea* it was decided that all *Leptothrix* and *Paucibacter* sequences from the Allander datasets should be extracted and compared with the literature studies. Allander_1 contained no *Leptothrix* sequences and only one

Paucibacter sequence. Allander_2 also contained no *Leptothrix* sequences however contained two *Paucibacter* sequences. These sample sites share the same body of water and are in close proximity to one another so both sites containing no *Leptothrix* sequences is actually a positive result as it shows reliability of the data. If one site had an abundance of *Leptothrix* whilst the other did not then that would be more difficult to explain. Finally, Allander_3 contained nine *Leptothrix* sequences and four *Paucibacter* sequences. As this site is part of a separate body of water it is not surprising that it contains a different phylogenetic profile. Figure 66 shows a maximum likelihood phylogenetic tree containing all the *Leptothrix* and *Paucibacter* 16S rRNA sequences from the Allander datasets compared with the literature studies. The 16S rRNA sequences of three isolated *Leptothrix* sp, *Leptothrix cholodnii*, *Leptothrix discophora* and *Leptothrix mobilis*, are also included to give a more thorough understanding of the *Leptothrix* and *Paucibacter* bacteria present.
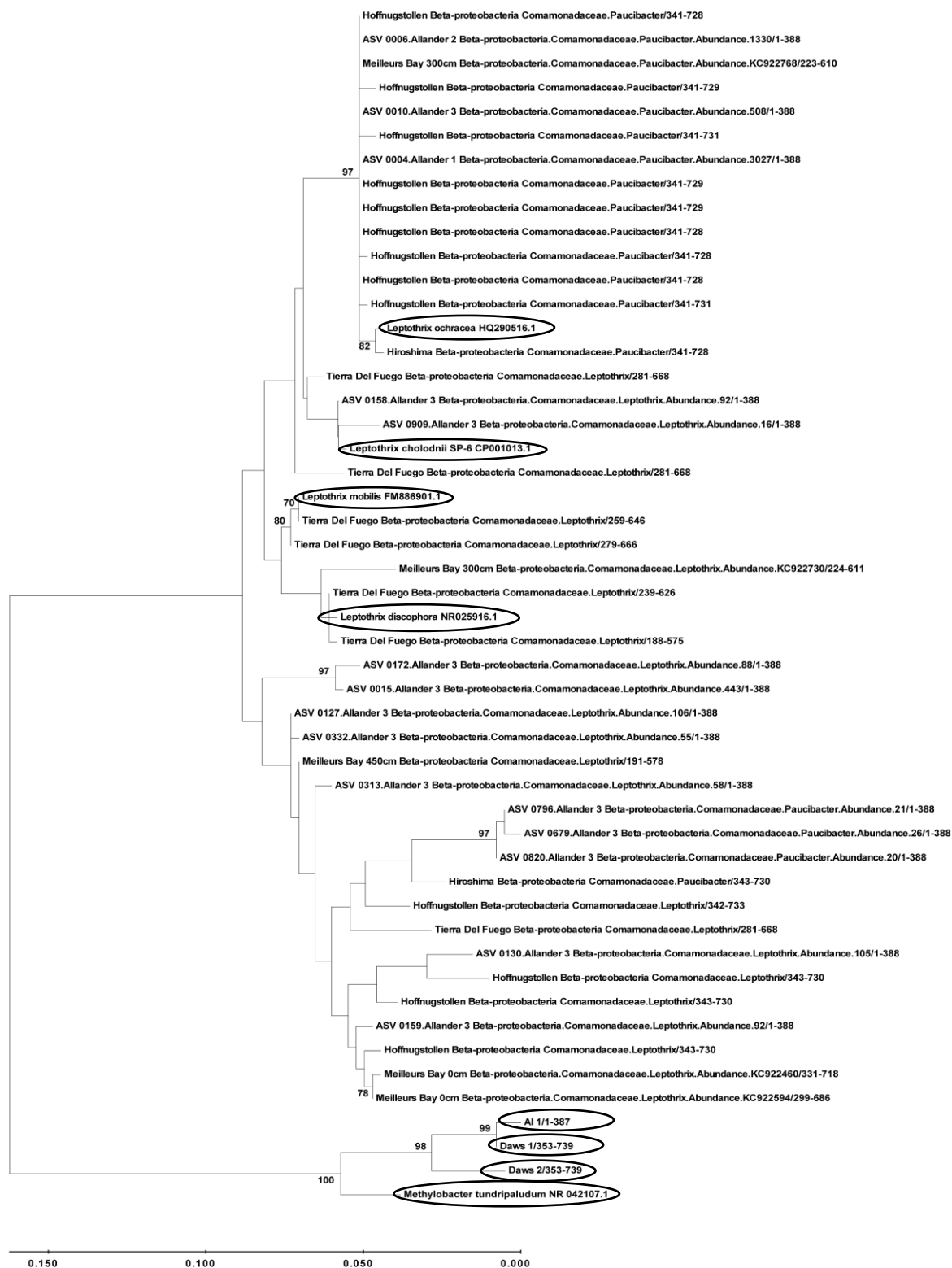
Figure 66 – Maximum likelihood phylogenetic tree of all the *Leptothrix* and *Paucibacter* 16S rRNA sequences found in the Allander datasets compared with the literature studies. Sequences of isolates and reference strains from the NCBI database have been included and circled. Scale bar = number of nucleotide substitutions per site, and bootstrap values >70 are shown.

This phylogenetic tree is similar with the previous *γ*-proteobacteria tree in that there is generally overlap between all sites. Two Allander_3 *Leptothrix* bacteria are closely related to the isolated bacterium *L.cholodnii* where the closest relative has a 99.74 % sequence identity. *L.cholodnii* is a filamentous heterotrophic FeOB, meaning it metabolises organic carbon, that also has the ability to oxidise manganese thereby forming manganese oxide encrusted filaments. It was first isolated from an iron seep by Emerson and Ghiorse in 1992 via repeated streaking of BIOX on manganese supplemented agar plates.[43] As such, a close relative is found here it may be possible to isolate it by utilising a similar methodology. This will be further discussed throughout Chapter 5. The isolates *L.discophora* and *L.mobilis* are found within two small separate clusters. Here it can be seen that they are most closely related to *Leptothrix* bacteria found in the Tierra Del Fuego sample site and do not appear to be closely related to any of the bacteria within the Allander sample sites. *L.ochracea* is found within the cluster at the top of the tree and is most closely related to a bacterium from the Hiroshima sample site, sharing a 99.48 % sequence identity with it. These two bacteria are closely clustered with bacteria predominantly from the Hoffnungsstollen sample site. This cluster also contains the three *L.ochracea*-like Allander *Paucibacter* bacteria. When compared with their nearest neighbours the Allander *Paucibacter* share a 100 % sequence identity with a *Paucibacter* from both the Hoffnungsstollen and Meilleurs Bay sample sites respectively. This cluster shows that the *L.ochracea*-like bacteria at many sites are highly sequence similar, at least at the level of 16S, and contrasts with the diversity of the pLW-20 genus of bacteria for example.

4.5.4 *Gallionella*

Gallionella is a genus of bean shaped chemolithotrophic, potentially mixotrophic, FeOB that can produce twisted stalks of 2-Fh which contribute to the architecture BIOX mats (Figure 67).[44] These stalks are comprised of bundles of thin fibres of EPS, secreted by single cells of *Gallionella*, that are woven together and become encrusted with 2-Fh in a similar manner in which *Leptothrix* filaments become encrusted. Stalks have been reported to measure ca. 0.5 – 3 µm in diameter and hundreds of microns in length. [45] *Gallionella* sp. require microaerophilic conditions for optimal growth and grow in much tighter bands than *L.ochracea*.[46]
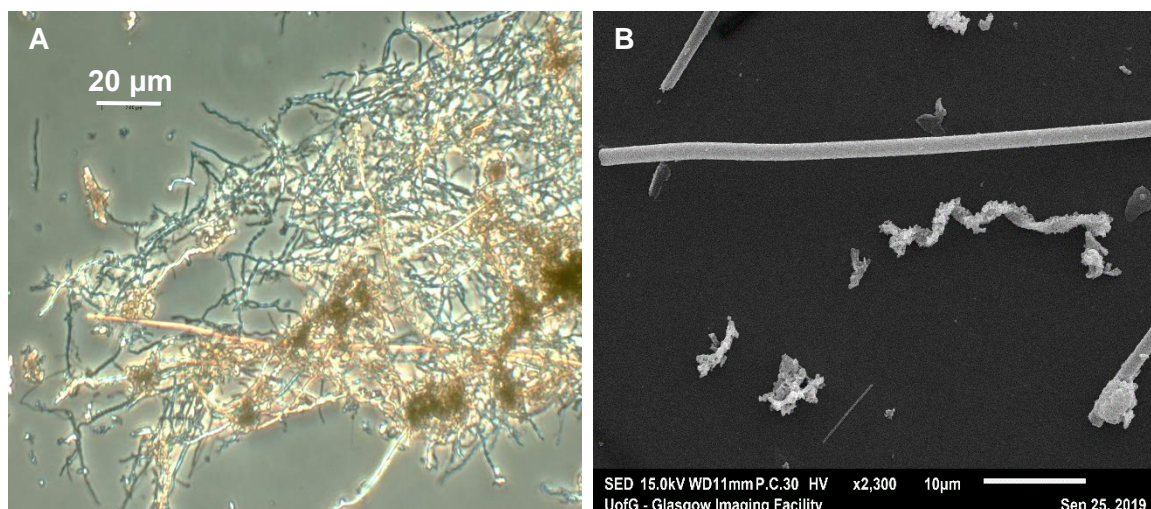
Figure 67 - A - Phase contrast micrograph showing a mat of *Gallionella* sp. twsited stalk and B - SEM micrograph of a *Gallionella* sp. stalk alongside a *L.ochracea* filament.

The Allander datasets were searched for *Gallionella* 16S rRNA sequences with Allander_1 containing 37 unique sequences, Allander_2 containing 33 unique sequences and Allander_3 containing 18 unique sequences. Including all these sequences in the analysis yielded a cluttered and difficult to read maximum likelihood phylogenetic tree. It was decided that only the 10 most abundant sequences from each site should be included as these give the best representation of the *Gallionella* profiles present. These were subsequently extracted and analysed using Jalview along with the 16S rRNA sequences of *Gallionella* sp. from the literature studies. The 16S rRNA sequences of two isolated bacteria, *Gallionella ferruginea* and *Gallionella capsiferriformans ES-2*, were also added to the datasets.[47, 48] Figure 68 shows a maximum likelihood phylogenetic tree of the 10 most abundant 16S rRNA sequences from the Allander datasets compared with the literature studies.
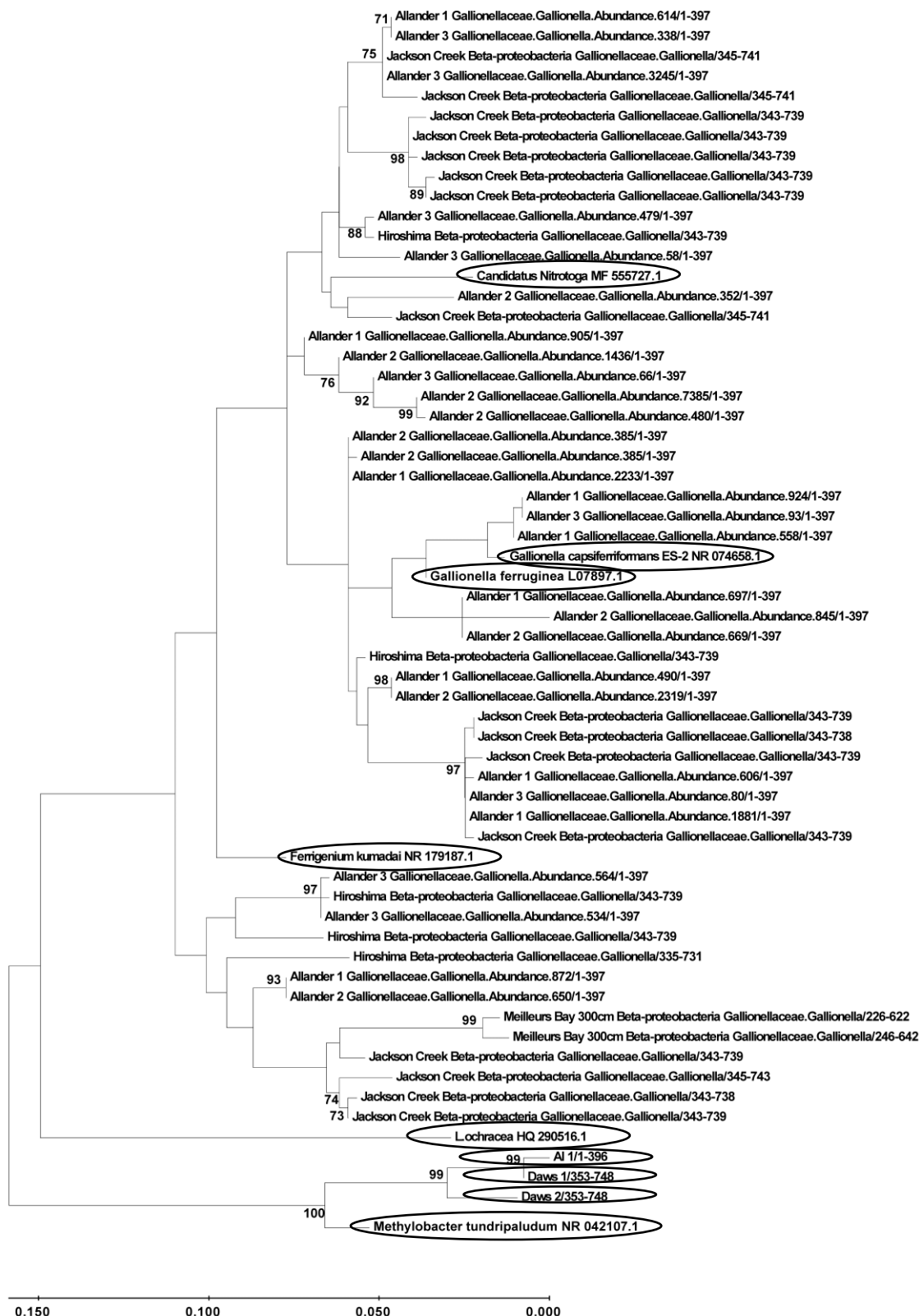
Figure 68 - Maximum likelihood phylogenetic tree of the 10 most abundant *Gallionella* sp. 16S rRNA sequences found in all the Allander datasets compared with the literature studies. Sequences of isolates and reference strains from the NCBI Database have been included and circled. Scale bar = number of nucleotide substitutions per site, bootstrap values >70 are shown.

It can be seen from this tree that there is again overlap between all datasets. The isolates *G.ferruginea* and *G.capsiferriformans* are most closely related to *Gallionella* spp. from the Allander_1 and Allander_3 sample sites. *G.capsiferriformans* shares a 98.55 % sequence identity over 415 bases with an identical Allander_1 and Allander_3 *Gallionella* and a 98.80 % sequence identity with another Allander_1 *Gallionella* respectively. *G.ferruginea* is more distantly related to this cluster however still shares a 97.36 % sequence identity with the Allander_1 and Allander_3 *Gallionella* and a 97.60 % sequence identity with the other Allander_1 *Gallionella*. Interestingly, *G.capsiferriformans* is a non-stalk forming species of *Gallionella* and *G.ferruginea* does not form stalks at pH values less than 6 or if the oxygen concentration is too low. Mats of *Gallionella* stalks are occasionally seen at the Allander sample sites however it is more common to see mats of *Leptothrix* filaments. Fleming *et al.* also hypothesised that different FeOB dominate BIOX mats under different environmental conditions, for example *L.ochracea* was found to be more prevalent than *Gallionella spp.* when dissolved organic carbon (DOC) concentrations were higher while *Gallionella spp.* dominate under tight redox gradients.[27] They also showed that non-stalk forming *Gallionella spp.* can be present year round and may adhere themselves to sheaths of *L.ochracea* to allow favourable positioning access iron and oxygen. This may explain why *Gallionella spp.* 16S rRNA sequences are so prevalent in the Allander datasets yet twisted stalks are generally rare compared with *Leptothrix* filaments. Neither *G.capsiferriformans* or *G.ferruginea* were isolated on solid media, instead they were enriched and isolated from liquid media using iron sulfide as an iron source and Wolfes Minimal Mineral Media (WMMM) (see Chapter 2 Section 6.1 for full composition) as growth medium. *G.capsiferriformans* was isolated using the gradient tube methodology whereby an iron sulfide/agarose plug is added to the bottom of a test tube and overlayed with WMMM media supplemented with minimal agarose to allow the cells buoyancy.[49] This is then sealed, leaving a minimal headspace of air, and CO_2 can be bubbled through to limit the oxygen concentration. The inoculum then grows at the appropriate iron/oxygen gradient. This methodology was employed using inoculum from the Allander sample sites however was unsuccessful at yielding FeOB as discussed in Chapter 5 Section 4.5.

A final notable *Gallionella* result comes from Allander_2. At the time of sampling for this study, the Allander_2 BIOX contained a large amount of *Toxothrix* trichomes that appeared both encrusted and unencrusted. These microstructures were the dominant microstructure found in the mat, being more prevalent than twisted stalks or filamentous sheaths. Figure 69 contains phase contrast micrographs of an encrusted and an unencrusted trichome found in the Allander_2 BIOX. A *Toxothrix spp.* has never been

isolated as an axenic culture and there is comparatively little literature concerning them when compared with bacteria that produce microstructured materials such as *Gallionella spp.* and *Leptothrix spp.*[50] One major observation that has been noted is that they are regularly found in water systems where *Gallionella* stalks are present which is in agreement with the observations of this study.
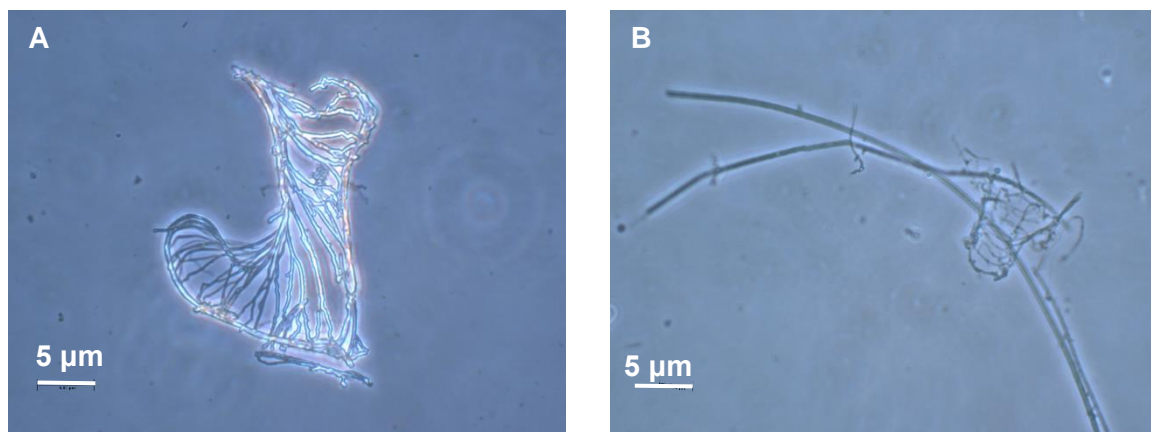
Figure 69 - Phase contrast micrographs of A - Encrusted *Toxothrix* trichome and B - Unencrusted *Toxothrix* trichome growing next to a filamentous sheath.

The appearance of such a high abundance of trichomes and the fact that many are unencrusted indicate that they are likely freshly produced meaning that *Toxothrix spp.* must be actively growing within the mats at a reasonable abundance. Table 18, in Chapter 4 Section 4.2 shows that the first, third and fifth most abundant bacteria in the Allander_2 site are *Gallionella spp.*, with abundances of 7388, 2345 and1439 respectively. These sequences appear both in Allander_1 and Allander_3, however at much lower abundancies. These rankings can be seen in Table 21.

Top three most abundant Allander_2 *Gallionella* (Overall Abundance Rank)	Respective Allander_1 Rank	Respective Allander_3 Rank
1st	109th	493rd
3rd	32nd	390th
5th	35th	264th

Table 21 - Comparison of the three most abundant *Gallionella spp.* in Allander_2 and where they rank in the overall Allander_1 and Allander_3 datasets.

There is currently no sequencing data for any *Toxothrix spp.* in the literature making it impossible to compare any of these sequences with a known *Toxothrix spp.* As the *Toxothrix* trichomes dominated Allander_2 at the time of sampling and there are three highly abundant *Gallionella* sequences present we hypothesise that *Toxothrix* belongs to the *Gallionella* genus of bacteria and is represented by one of these sequences. This hypothesis is consistent with our other datasets as the three abundant *Gallionella* sequences also appear in the other datasets at a much lower abundance which agrees with the low abundance of *Toxothrix* trichomes seen in the other sites. Further work is warranted to confirm this. Confirmation could be confirmed using various methodologies. Firstly, enrichment and isolation of a *Toxothrix spp.* from this sample for whole genome sequencing should be attempted. This should include attempts using both solid and liquid media as well isolation by micromanipulation as discussed throughout Chapter 5. Secondly, FISH probes could be designed using the relevant 16S rRNA sequences. This would allow potential *Toxothrix* cells to be identified via fluorescent microscopy in a similar way to Fleming *et al.* identifying *L.ochracea* cells.[23]

4.5.5 *Sideroxydans*

Sideroxydans is another genus of chemolithotrophic FeOB.[44] They are rod shaped cells which do not typically produce 2-Fh microstructures, however they do produce particulate 2-Fh which they become associated with. This makes them difficult to image by light or phase contrast microscopy, instead requiring fluorescent microscopy combined with DNA staining. They are phylogenetically related to the *Gallionella* genus and an isolated

species, *Sideroxydans lithotrophicus* ES-1, was isolated in the same study as *G.capsiferriformans* ES-2 via the gradient tube methodology.[48] Searching the Allander datasets revealed that Allander_1 contained the greatest number of unique 16S rRNA sequences of *Sideroxydans*, 18, while Allander_2 contained 9 and Allander_3 contained 7. As with the previous examples a maximum likelihood phylogenetic, seen in Figure 70, was constructed to compare the Allander *Sideroxydans* profile with globally relevant BIOX samples. The 16S rRNA sequence of *S.lithotrophicus* was also included in this comparison.
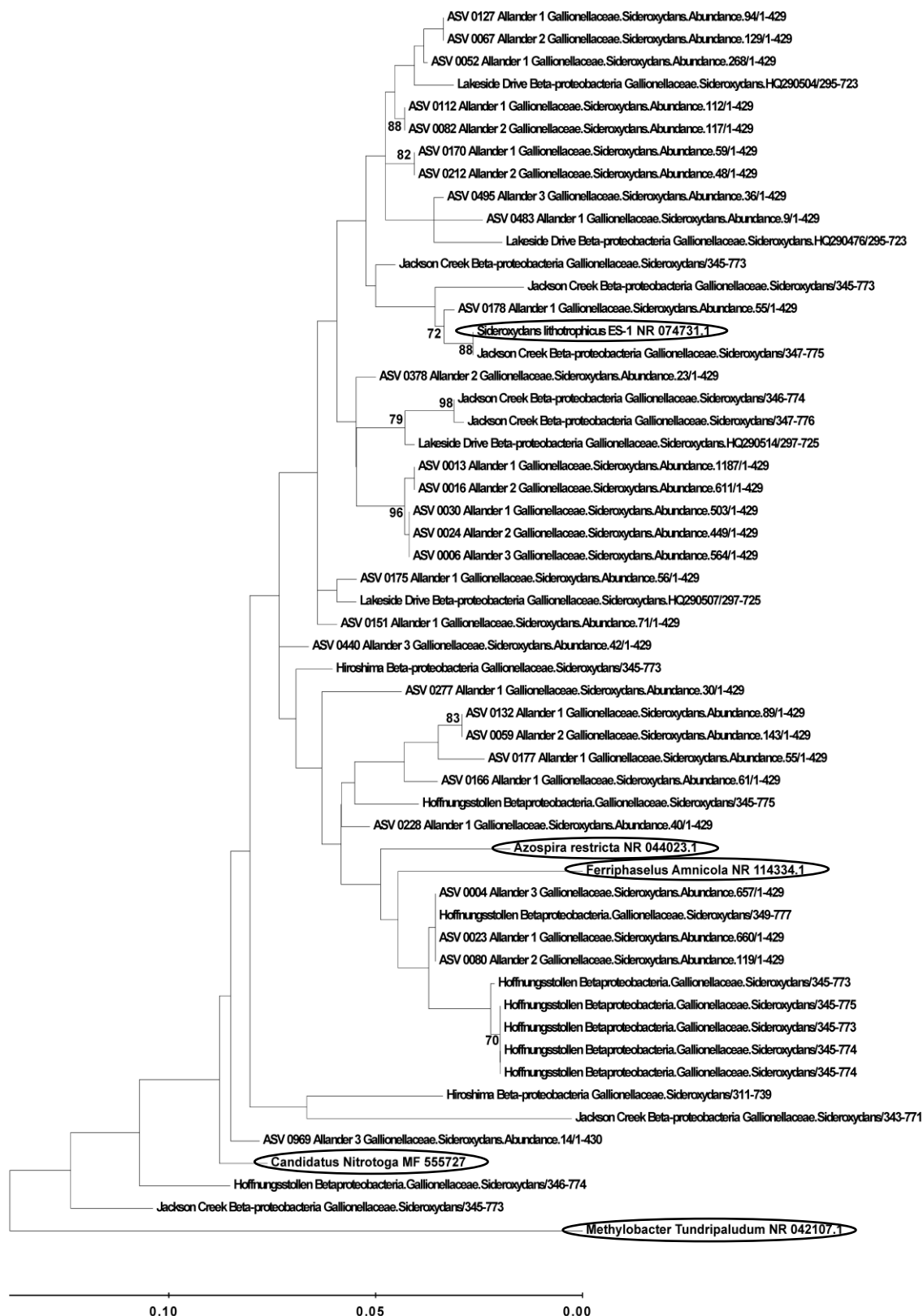
Figure 70 - Maximum likelihood phylogenetic tree of all the *Sideroxydans* 16S rRNA sequences found in the Allander datasets compared with the literature studies. Sequences of reference strains from the NCBI Blast database have been included and circled. Scale bar = number of nucleotide substitutions per site, and bootstrap values >70 are shown.

S.lithotrophicus, has a 100 % sequence identity with a *Sideroxydans* spp. from the Jackson Creek sample site and a 99.30 % sequence identity with an Allander_1 *Sideroxydans* spp. Again, there is overlap between all sites and one of the Hoffnungsstollen sequences has a 100 % sequence identity with a *Sideroxydans* spp. that is found at all three Allander sample sites. In general, however, the Allander *Sideroxydans* spp. appear to be more closely clustered with themselves rather than *Sideroxydans* spp. from other sample sites.

M.tundripuldum has again been selected as an outgroup as it is a γ-gamma proteobacteria and the other sequences belong to β-proteobacteria. As expected, this outgroup appears to be distinct from the other sequences in the tree.

4.5.6 *Rhodoferax*

The final genus of β-proteobacteria from the Allander datasets being compared with the literature studies is the *Rhodoferax* which are described in Chapter 5.4.4. 20 unique sequences from Allander_1, 9 from Allander_2 and 40 from Allander_3 were found to be available for comparison within the datasets. As with the *Gallionella* comparison this is far too many for a concise comparative phylogenetic tree once the literature examples are included. To remedy this, 10 Allander_1, 8 Allander_2 and 11 Allander_3 sequences were selected, based on abundance data and the whole maximum likelihood phylogenetic tree comparison, to be representative of the Allander datasets. These selected sequences were then compared with the literature studies. The 16S rRNA sequences of three literature isolates, *Rhodoferax saidenbacensis*, *Rhodoferax sediminis* and *Rhodoferax ferrireducens*, were included in the comparison as was the 16S rRNA sequence of the *Rhodoferax Allander_3* isolate described in Chapter 5.4.4.[51-53] *R.ferrireducens* was included as it is an FeRB that utilises ferric compounds as an electron acceptor and has been shown to reduce ferric compounds at temperatures as low as 4 °C. If any of the Allander *Rhodoferax* are closely related then they may also have this ability which would indicate iron cycling occurring within the BIOX mats. Figure 71 shows a maximum likelihood phylogenetic tree of the relevant *Rhodoferax* 16S rRNA gene sequences from the Allander datasets compared with the literature examples.

Figure 71 - Maximum likelihood phylogenetic tree of the most abundant *Rhodoferax* 16S rRNA sequences from the Allander datasets compared with the literature studies. Sequences of reference strains have been included and circled. Scale bar = number of nucleotide substitutions per site, and bootstrap values >70 are shown.

Interestingly, there was only one *Rhodoferax* 16S rRNA sequence in the Meilleurs Bay dataset and no available sequences in the Jackson Creek dataset. The Meilleurs Bay example is closely clustered with a *Rhodoferax* spp. from Hoffnungstollen, LakeSide Drive and Allander_3, with which it shares a 99.77 % sequence identity over 440 bases with. *R.saidenbachensis* shares a 100 % sequence identity with another Allander_3 *Rhodoferax* spp., while *R.sediminis* is more distantly related to its nearest Allander neighbours, sharing a 98.83 % sequence identity with an Allander_1 and Allander_2 *Rhodoferax* spp. respectively. The FeRB *R.ferrireducens* is closely clustered with an example from each Allander site and shares a 99.53 % sequence identity with them. This close similarity suggests that *Rhodoferax spp.* within the Allander sample sites can reduce ferric compounds. As BIOX is composed of 2-Fh, a ferric oxyhydroxide, it may be possible for these FeRB to utilise BIOX which in turn would allow a continual supply of ferrous iron within the mats. The *R.Allander_3* isolate shares a 99.77 % sequence identity with its nearest relative, from Allander_1, and a 98.83 % sequence identity with its nearest Allander_3 relative. This can be explained as the BIOX from which it was isolated was sampled 10/11/19 while the BIOX for the 16S rRNA analysis was sampled 6/1/20. This difference may allow a slightly different phylogenetic profile. It is also possible that that the BIOX was not sampled from the exact same location within the sample site. As BIOX mats are heterogeneous structures it has been hypothesised that different areas of the mats may contain different communities due to changes in the microenvironment meaning that the *R.Allander_3* may have been isolated from BIOX that is not 100 % similar with the BIOX sampled for the 16S rRNA analysis.

4.6 Biogeochemical Cycling

The 16S RNA sequences provide an overview of the bacterial communities present at the three Allander sites sampled The relationship between bacteria within environmental samples is important for the distribution of nutrients via biogeochemical cycling.[54] Jorgensen *et al.* have shown that by using 16S rRNA sequencing to build a bacterial community profile it is possible to evaluate the biogeochemical cycling that is occurring within environmental sediments.[55] Table 22 shows a selection of biogeochemical roles and the genera of bacteria found with the Allander BIOX mats that are responsible for them.

Biogeochemical Role	Genus With This Ability
Iron Oxidation	*Leptothrix, Gallionella, Hypomicrobium, Pedomicrobium, Sideroxydans, Dechloromonas, Crenothrix, Ferritrophicum, Aquabacterium*
Iron Reduction	*Geobacter, Bacillus, Rhodoferax*
H_2S Oxidation	*Thiothrix, Beggiatoa, Thiobacillus*
Sulfate Reduction	*Desulfatiferula, Desulfosarcina, Desulfomonile, Desulfovirga, Desulfosporosinus, Acetobacerium, Bacillus*
Methane Oxidation	*Methyolbacter, Methylmicrobium, Methylomonas, Methylovulum*
Ammonification	*Clostridium, Bacillus, Pseudomonas*
Ammonia Oxidation	*Nitrospira*
Anammox	*Pirellula*
Nitrogen Fixation	*Ideonella, Mesorhizobium, Acidothermus,Actinomycetospora*
Denitrification	*Bacilllus, Pierllula, Pseudomonas, Thiobacillus, Klebsiella*
Nitrite Oxidation	*Nitrobacter, Nitrospira, Nitrotoga*

Table 22 - Biogeochemical roles of various genera found within the three BIOX mats. Table adapted from Edwards.[41]

4.6.1 Iron Cycling

FeOB and iron reducing bacteria (FeRB) are both found within the BIOX mats highlighting that cycling between Fe (II) and Fe (III) is potentially occurring and that the product of one reaction is likely utilised by the concomitant bacteria as a reagent. FeOB-FeRB microbial couplings typically occur at the surface-subsurface interface where the microoxic conditions which allow them to outcompete abiotic processes.[44] In addition, the microoxic conditions allow photoferrotrophic (bacteria that using light and Fe (II) to fix CO_2 into biomass) and nitrate reducing FeOB to occupy BIOX mats alongside microaerophiles, which are generally more dominant, such as *Leptothrix spp.*, *Gallionella spp.* and *Sideroxydans spp.* Some of the genera present such as *Aquabacterium spp.* and *Dechloromonas spp.* are nitrate reducers while *Pseudorhodobacter spp.* have been shown to be photoferrotrophs.[56] Equation 14 shows how iron oxidation is coupled with nitrate reduction and Equation 15 shows how phototrophic iron oxidation is coupled with carbon dioxide reduction to yield biomass.

$$10\,Fe(CO)_3 + 2\,NO_3^- + 10\,H_2O \longrightarrow Fe_{10}O_{14}(OH)_2 + 10\,HCO_3^- + N_2 + 8\,H^+$$
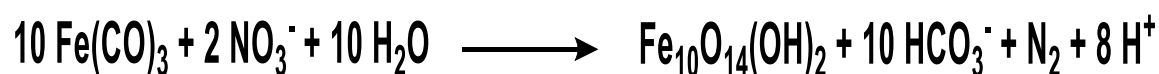
Equation 14 - Iron oxidation coupled with nitrate reduction.[56]

$$4 \text{ Fe (II)} + CO_2 + 11 \text{ } H_2O + \textit{hv} \longrightarrow CH_2O + 4 \text{ Fe(OH)}_3 + 8 \text{ } H^+$$
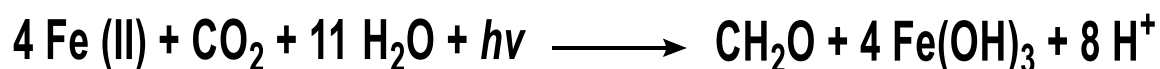
Equation 15 - Phototrophic iron oxidation coupled with CO₂ reduction to give biomass (CH₂O).[56]

4.6.2 Sulfur Cycling

Sulfur cycling also occurs within the BIOX mats in in the form of sulfide oxidation and sulfate reduction. Pyrite is the most abundant sulfide mineral meaning that sulfur cycling is regularly found in environments where iron cycling is present as Pyrite is a source of both elements. In the presence of oxygen pyrite will oxidise to produce sulfate and Fe (II), resulting in the acidification of groundwater as discussed in Chapter 6.5.1.2. The sulfide oxidising bacteria (SOB) found within these BIOX mats are not restricted to pure sulfide oxidation and can oxidise a variety of sulfur species including elemental sulfur, sulfides, sulfates. Enzymes responsible for these processes are encoded by genes including sulfide quinone oxidoreductase (SQR), flavocytochrome c-sulfide dehydrogenase (FCSD) and the sulfur oxidation enzyme (SOX).[57, 58] Figure 72 shows the major inorganic species present as a result of sulfur cycling and the processes that link them.[59]
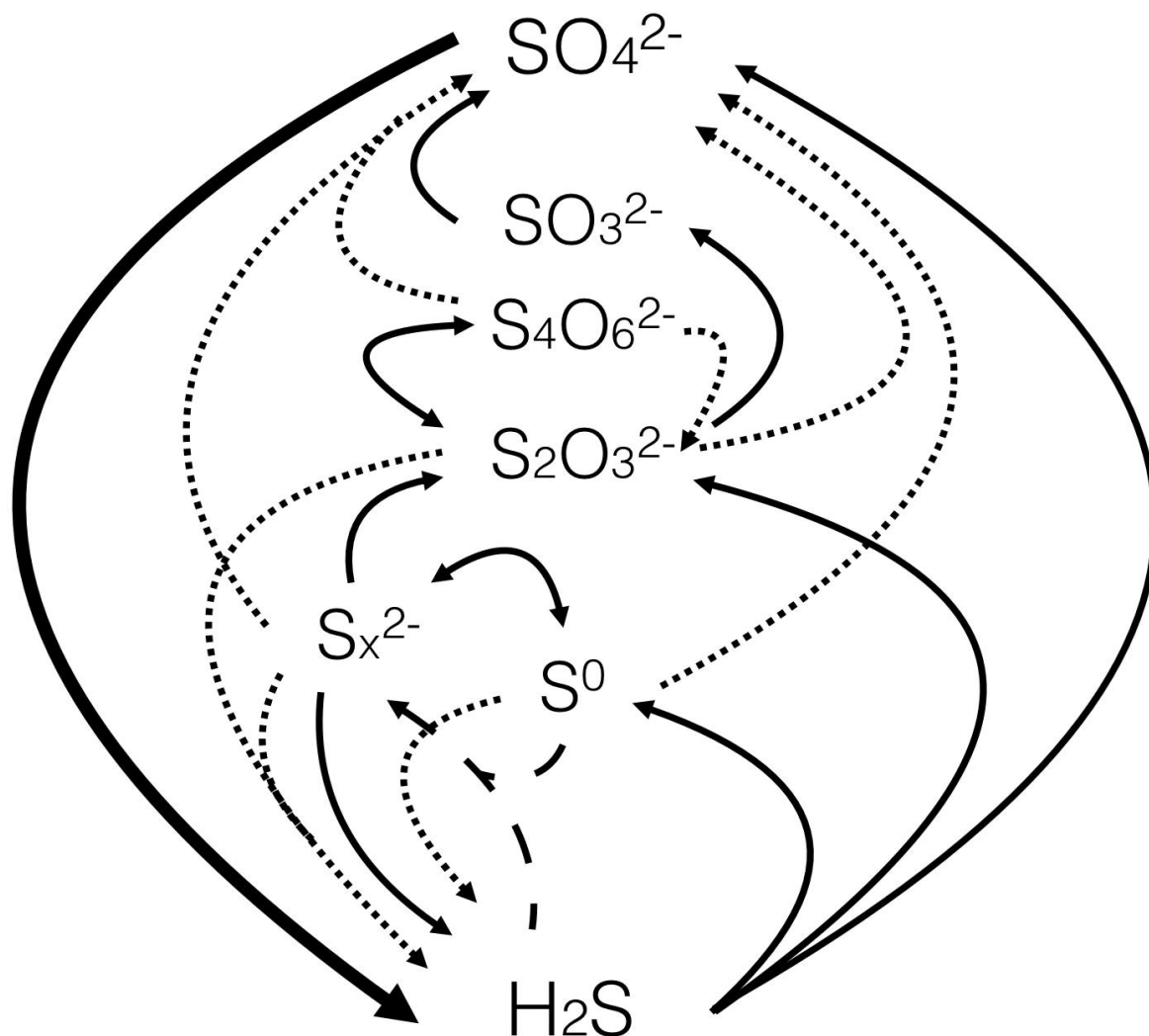
Figure 72 - Major inorganic species present in the sulfur cycle. Thick arrows – sulfate reduction, thin arrows – sulfate oxidation, dashed lines – equilibration to polysulfides and dotted lines – disproportionation reactions. Adapted from Jørgensen *et al.*[59]

SOB such as *Thiothrix* can also be filamentous bacteria and as such are able to deposit elemental sulfur granules within their sheaths via these processes. Figure 73 shows a phase contrast micrograph of a *Thiothrix* filament with sulfur granules highlighted. These oxidised sulfur compounds are then utilised by sulfate reducing bacteria (SRB) much in the same way that the FeOB and FeRB can utilise one another's by-products.
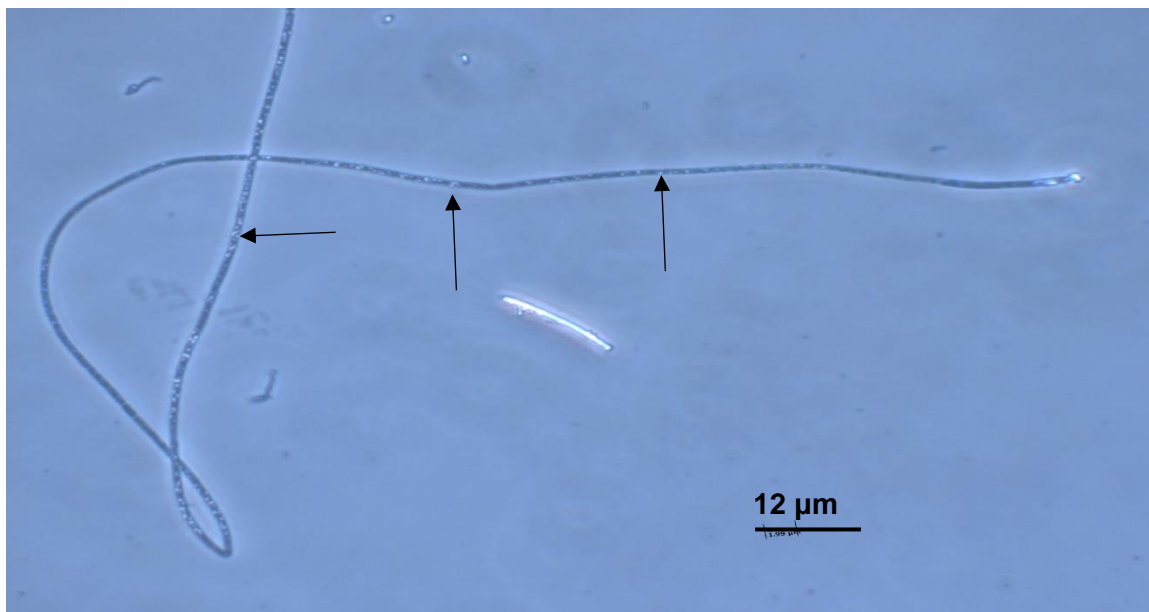
Figure 73 - Phase contrast micrograph of a *Thiothrix* filament with sulfur granules highlighted with arrows.

4.6.3 Methane Cycling

Methanotrophic bacteria are present within the sampled BIOX mats and are most abundant in the Allander_1 site. The biogenic oxidation of methane has been shown to be coupled with other biogenic reactions including iron reduction, sulfate reduction and the reduction of nitrate and nitrite, all of which happen within the mats.[60-62] During these reactions methane is either oxidised to methanol which is then converted to formate and finally to carbon dioxide or to bicarbonate which is then utilised for further biogenic and abiogenic reactions. Figure 74 shows equations for these reactions.

$$1 - \quad CH_4 + SO_4^{2-} \quad \longrightarrow \quad HCO_3^- + HS^- + H_2O$$

$$2 - \quad CH_4 + 4\,NO_3^- \quad \longrightarrow \quad CO_2 + 4\,NO_2^- + 2\,H_2O$$

$$3 - \quad 3\,CH_4 + 8\,NO_2^- + 8\,H^+ \quad \longrightarrow \quad 3\,CO_2 + 4\,N_2 + 10\,H_2O$$

$$4 - \quad CH_4 + 8\,Fe(OH)_3 + 15\,H^+ \quad \longrightarrow \quad HCO_3^- + 8\,Fe^{2+} + 21\,H_2O$$
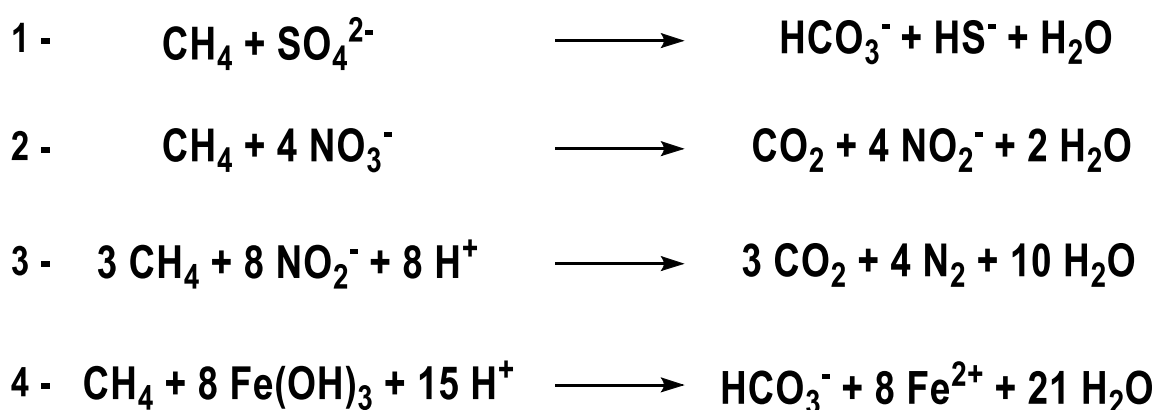
Figure 74 - Equations for the above described redox reactions of methane and 1 - Sulfate, 2 - Nitrate, 3 - Nitrite and 4 – Ferric iron.

The electrons generated here are transferred to the relevant reducing bacteria which then utilise them for reductive mechanisms. A pathway for biogenic methane oxidation can be found in Chapter 5.7.

4.6.4 Nitrogen Cycling

Nitrogenous compounds are essential for life and nitrogen cycling is another important rection which occurs within the BIOX mats. Figure 75 shows a schematic for the bacterial nitrogen cycle.
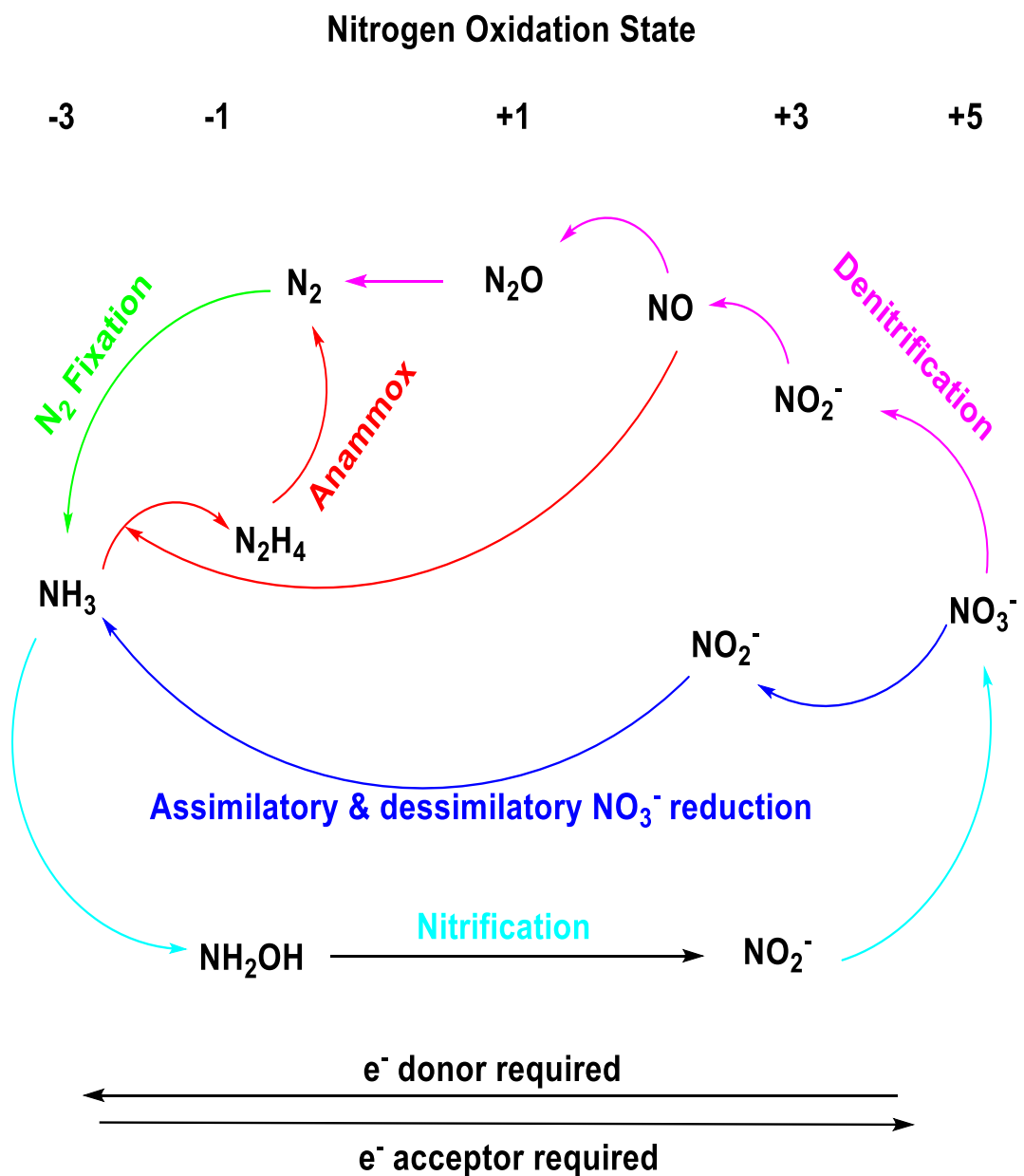
Figure 75 - Schematic representing the bacterial nitrogen cycle. Adapted from Sparancino-Watkins _et al._[63]

Nitrogen fixing bacteria, such as *Mesorhizobium spp.*, convert atmospheric nitrogen to ammonia via nitrogenase enzymes. Corresponding bacteria then convert ammonia to ammonium, nitrate and nitrite by enzymes including ammonia monooxygenase, hydroxylamine dehydrogenase, nitric oxide oxidase and nitrite oxidoreductase. Finally, denitrification is carried out using enzymes such as nitrous oxide reductase.[64, 65]

The nitrogen cycle is not the only method of nitrogen assimilation available to bacteria. Ammonium can be directly assimilated via the AmtB/Amt1 transporters which are ubiquitous throughout bacteria.[66] These are membrane bound transporter proteins that scavenge environmental ammonium ions and have also shown to be involved in ammonium ion sensing.[67] Assimilated ammonium is then typically incorporated into organic molecules via the glutamine synthetase-glutamate synthase cycle, resulting in the production of glutamine and glutamate respectively.[68] Prokaryotes have also evolved the ability to assimilate complex nitrogen containing compounds such as amino acids, xenobiotics and nitrogenous bases. Once these compounds are assimilated they can then be utilised in nitrogen cycling mechanisms. This highlights that the bacteria within the BIOX mats can likely make use of a variety of environmental nitrogenous molecules for nitrogen cycling. It is therefore reasonable to expect that nitrogen compounds in various oxidation are ubiquitous to the microenvironments within the mats.

These results show that BIOX mats house a variety of bacteria with a variety of environmental niches. Biogeochemical cycling of the described compounds highlights just some of the chemistry occurring within a small sample volume. The fact that by-products of one bacterium can be utilised by another indicates that syntrophic relationships likely exist and may also provide insight to why it is so difficult to isolate *L.ochracea* as they thrive in an environment that is in continual flux.

4.7 Conclusions

BIOX samples were collected from three Allander sample sites on 9/1/20 and then subsequently had their genomic DNA extracted. Genomic DNA was amplified via PCR, using universal primers, adjusted to 20 ng/μl and sent for high-throughput Illumina sequencing targeting the V3 – V4 hypervariable regions of the 16S rRNA gene. A library of sequences was created for each sample site to compare their phylogenies. Allander_1 generated 267,118 sequences, Allander_2 generated 227,758 sequences and Allander_3 generated 245,222 sequences. These corresponded to both forward and reverse

sequence read pairs that were on average ca. 250 bp in length and had a GC content of ca. 53 %. Overlapping sequences were merged and artefacts including primer and adapter sequences and chimeric sequences were removed. These procedures resulted in 87,952 sequences from Allander_1, 64,820 sequences from Allander_2 and 67,513 sequences from Allander_3. This data was further refined in R using the DADA2 pipeline combined with the Silva nr99 v138 train set yielding 688, 937 and 1521 ASV for Allander_1 – 3 respectively. These libraries were then used to assess the phylogeny of each sample site, provide insight to the biogeochemical cycling occurring within the mats and were compared with libraries of BIOX samples from literature studies around the globe.

Analysis of the libraries showed that each site had a distinct bacterial profile however the *γ*- and *β*-proteobacteria were the predominant class of bacteria present. The genus *Leptothrix* was comparatively rare when compared with other *β*-proteobacteria such as the *Gallionella*, however an abundant 16S rRNA sequence identified as a *Paucibacter spp.* found in all Allander sites was found to have a very high sequence similarity, 99.23 %, sequence identity with *L.ochracea*. The identification of this sequence lends credence to the first hypothesis for this chapter that *Leptothrix* 16S rRNA sequences should be present. Filamentous methanotrophs of the pLW-20 genus were found at all sites and the Daws_1, Daws_2 and Al_1 isolates were found to have a 100 % sequence identity with highly abundant pLW-20 bacteria from these libraries.

Allander_2 contained a high abundance of *Toxothrix* trichomes when compared with the other sample sites. Analysis of the abundance data showed that three *Gallionella spp.* are in high abundance at this site and that these highly abundant *Gallionella spp.* also appear in Allander_1 and Allander_3 at much lower abundances. It has therefore been hypothesised that one of these sequences may correspond to a *Toxothrix spp.* Further work is required here to confirm this as *Toxothrix* would have to be isolated from the mat for whole genome sequencing. It may also be possible to design a FISH probe using the *Gallionella* 16S rRNA sequences that would allow visual confirmation of *Toxothrix* via fluorescent microscopy. The appearance of *Gallionella spp.* and potentially *Toxothrix spp.* 16S rRNA gene sequences agrees with what was originally hypothesised for this chapter.

Investigating the biogeochemistry suggests that a host of chemical cycles occur within the mats including iron cycling, sulfur cycling, nitrogen cycling and carbon cycling. The products of these cycles are likely utilised by other bacteria within the mats meaning that

there is a continual supply of nutrients allowing the diverse population of bacteria to thrive. This result strengthens the hypothesis that biogenic cycling of elements occurs within the BIOX mats.

Comparing the phylogenetic libraries of the three Allander sample sites with literature studies showed that freshwater BIOX mats tend to have similar genera of bacteria present however the number of unique sequences differs greatly from site to site. This undoubtedly reflects the sequencing approach adopted at the time to characterise the bacterial communities. This is most true for the Jackson Creek and Meillerus Bay sample sites as they contained zero and one *Rhodoferax* spp. respectively while Allander and the other sites contained a large variety. There was general overlap between all sites however it was uncommon for the Allander bacteria to share a 100 % sequence identity with bacteria from the literature studies. From this it can be said that the phylogenetic profiles of the Allander sample sites are more similar to each other than they are to the phylogenetic profiles of the literature studies. This result agrees with what was hypothesised for this chapter in that the Allander sites would be more similar to each other than to literature studies.

4.8 References

1. D. A. Baum, S. D. Smith and S. S. S. Donovan, *Science*, 2005, **310**, 979.
2. T. R. Gregory, *Evolution: Education and Outreach*, 2008, **1**, 121-137.
3. J. B. Lamarck, *Philosophie zoologique*, Paris, 1809.
4. C. Darwin, *Notebook B. p 36, 74.*, 1837.
5. C. Darwin, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, London, U.K, 1859.
6. C. R. Woese and G. E. Fox, *Proceedings of the National Academy of Sciences of the United States of America*, 1977, **74**, 5088-5090.
7. A. Djurhuus, C. J. Closek, R. P. Kelly, K. J. Pitz, R. P. Michisaki, H. A. Starks, K. R. Walz, E. A. Andruszkiewicz, E. Olesin, K. Hubbard, E. Montes, D. Otis, F. E. Muller-Karger, F. P. Chavez, A. B. Boehm and M. Breitbart, *Nature Communications*, 2020, **11**, 254.
8. D. Sune, H. Rydberg, Å. N. Augustinsson, L. Serrander and M. B. Jungeström, *Journal of Microbiological Methods*, 2020, **170**, 105854.
9. K. Fukuda, M. Ogawa, H. Taniguchi and M. Saito, *Journal of UOEH*, 2016, **38**, 223-232.
10. T. Větrovský and P. Baldrian, *PLOS One*, 2013, **8**, e57923-e57923.
11. B. Yang, Y. Wang and P.-Y. Qian, *BMC bioinformatics*, 2016, **17**, 135-135.
12. A. P. Heikema, D. Horst-Kreft, S. A. Boers, R. Jansen, S. D. Hiltemann, W. de Koning, R. Kraaij, M. A. J. de Ridder, C. B. van Houten, L. J. Bont, A. P. Stubbs and J. P. Hays, *Genes*, 2020, **11**, 1105-1122.
13. E. Fadeev, M. G. Cardozo-Mino, J. Z. Rapp, C. Bienhold, I. Salter, V. Salman-Carvalho, M. Molari, H. E. Tegetmeyer, P. L. Buttigieg and A. Boetius, *Frontiers in Microbiology*, 2021, **12**, 283.
14. Y. S. Bukin, Y. P. Galachyants, I. V. Morozov, S. V. Bukin, A. S. Zakharenko and T. I. Zemskaya, *Scientific Data*, 2019, **6**, 190007.
15. F. Teng, S. S. Darveekaran Nair, P. Zhu, S. Li, S. Huang, X. Li, J. Xu and F. Yang, *Scientific Reports*, 2018, **8**, 16321.
16. M. Bio., Fast DNA SPIN Kit For Soil Manual <https://www.mpbio.com/116560000-fastdna-spin-kit-for-soil-samp-cf>).
17. J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld and R. Knight, *Nature Methods*, 2010, **7**, 335-336.
18. R. Edgar, B. Haas, J. Clemente, C. Quince and R. Knight, *Bioinformatics (Oxford, England)*, 2011, **27**, 2194-2200.
19. B. J. Callahan, P. J. McMurdie and S. P. Holmes, *The ISME Journal*, 2017, **11**, 2639-2643.
20. A. E. Pérez-Cobas, L. Gomez-Valero and C. Buchrieser, *Microbial Genomics*, 2020, **6**, mgen000409.
21. B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson and S. P. Holmes, *Nature Methods*, 2016, **13**, 581-583.
22. E. Stackebrandt, R. G. E. Murray and H. G. Truper, *International Journal of Systematic Bacteriology*, 1988, **38**, 321-325.
23. E. J. Fleming, A. E. Langdon, M. Martinez-Garcia, R. Stepanauskas, N. J. Poulton, E. D. P. Masland and D. Emerson, *PLOS One*, 2011, **6**, e17769-e17769.
24. C. N. Brooks and E. K. Field, *mBio*, 2020, **11**, e02720-02720.
25. E. K. Field, S. Kato, A. J. Findlay, D. J. MacDonald, B. K. Chiu, G. W. Luther Iii and C. S. Chan, *Geobiology*, 2016, **14**, 499-508.

26. O. Nercessian, E. Noyes, M. G. Kalyuzhnaya, M. E. Lidstrom and L. Chistoserdova, *Applied and Environmental Microbiology*, 2005, **71**, 6885-6899.
27. E. J. Fleming, I. Cetinić, C. S. Chan, D. Whitney King and D. Emerson, *The Isme Journal*, 2013, **8**, 804.
28. D. Emerson, J. Scott Jarrod, J. Benes, B. Bowden William and F. E. Löffler, *Applied and Environmental Microbiology*, 2015, **81**, 8066-8075.
29. J. S. Berg, D. Jézéquel, A. Duverger, D. Lamy, C. Laberty-Robert and J. Miot, *PLOS ONE*, 2019, **14**, e0212787.
30. C. S. Chan, S. M. McAllister, A. H. Leavitt, B. T. Glazer, S. T. Krepski and D. Emerson, *Frontiers in Microbiology*, 2016, **7**, 796.
31. T. Watanabe, H. Sumida, N. Minh Do, K. Yano, S. Asakawa and M. Kimura, *Soil Science and Plant Nutrition*, 2013, **59**, 337-346.
32. F. Zeitvogel, I. Adaktylou, P. Ingino, G. Schmid, S. Roehler, C. Burkhardt, J. Byrne, M. Halama and D. Emerson, unpublished work.
33. E. Roden, J. McBeth, M. Blothe, E. Percak-Dennett, E. Fleming, R. Holyoke, G. Luther and D. Emerson, *Frontiers in Microbiology*, 2012, **3**, 172.
34. S. Kato, S. Kikuchi, T. Kashiwabara, Y. Takahashi, K. Suzuki, T. Itoh, M. Ohkuma and A. Yamagishi, *Geomicrobiology Journal*, 2012, **29**, 896-905.
35. B. Schmidt, L. A. Sánchez, T. Fretschner, G. Kreps, M. A. Ferrero, F. Siñeriz and U. Szewzyk, *FEMS Microbiology Ecology*, 2014, **90**, 454-466.
36. L. Liu, Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu and M. Law, *Journal of Biomedicine and Biotechnology*, 2012, **2012**, 251364.
37. C. Luo, D. Tsementzi, N. Kyrpides, T. Read and K. T. Konstantinidis, *PLOS ONE*, 2012, **7**, e30087.
38. V. Shirokova, PhD, University of Toronto, 2015.
39. T. DeSantis, H. Philip, N. Larsen, M. Rojas, E. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu and G. Andersen, *Applied and Environmental Microbiology - AEM*, 2006, **72**, 5069-5072.
40. B. E. Slatko, A. F. Gardner and F. M. Ausubel, *Current Protocols in Molecular Biology*, 2018, **122**, e59-e59.
41. B. Edwards, MSc Thesis, University of Toronto, 2018.
42. P. Schloss, S. Westcott, T. Ryabin, J. Hall, M. Hartmann, E. Hollister, R. Lesniewski, B. Oakley, D. Parks, C. Robinson, J. Sahl, B. Stres, G. Thallinger, D. Van Horn and C. Weber, *Applied and Environmental Microbiology*, 2009, **75**, 7537-7541.
43. D. Emerson and W. C. Ghiorse, *Applied and Environmental Microbiology*, 1992, **58**, 4001-4010.
44. D. Emerson, E. J. Fleming and J. M. McBeth, *Annual Review of Microbiology*, 2010, **64**, 561-583.
45. T. Suzuki, H. Hashimoto, N. Matsumoto, M. Furutani, H. Kunoh and J. Takada, *Applied and Environmental Microbiology*, 2011, **77**, 2877-2881.
46. E. J. Fleming, T. Woyke, R. A. Donatello, M. M. M. Kuypers, A. Sczyrba, S. Littmann, D. Emerson and L. Drake Harold, *Applied and Environmental Microbiology*, 2018, **84**, e02239-02217.
47. L. Hallbeck, F. Ståhl and K. Pedersen, *Journal of general microbiology*, 1993, **139**, 1531-1535.
48. D. Emerson and M. Merrill Floyd, *Methods in Enzymology*, Academic Press, 2005, vol. 397, pp. 112-123.
49. D. Emerson and C. Moyer, *Applied and Environmental Microbiology*, 1997, **63**, 4784-4792.
50. P. Hirsch, in *The Prokaryotes: A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications*, eds. A. Balows, H. G. Trüper, M. Dworkin, W. Harder and K.-H. Schleifer, Springer New York, New York, NY, 1992, pp. 4026-4029.

51. R. Kaden, C. Spröer, D. Beyer and P. Krolla-Sidenstein, *International Journal of Systematic and Evolutionary Microbiology*, 2014, **64**, 1186-1193.
52. C.-Z. Jin, Z. Ye, X. Wu, S.-R. Ko, T. Li, F. J. Jin, C.-Y. Ahn, H.-M. Oh, H.-G. Lee and L. Jin, *Microorganisms*, 2020, **8**, 262.
53. K. Finneran, *International Journal of Systematic and Evolutionary Microbiology*, 2003, **53**, 669-673.
54. H. W. Paerl and J. L. Pinckney, *Microbial Ecology*, 1996, **31**, 225-247.
55. S. L. Jorgensen, B. Hannisdal, A. Lanzén, T. Baumberger, K. Flesland, R. Fonseca, L. Øvreås, I. H. Steen, I. H. Thorseth, R. B. Pedersen and C. Schleper, *Proceedings of the National Academy of Sciences*, 2012, **109**, E2846.
56. S. Hedrich, M. Schlömann and D. B. Johnson, *Microbiology*, 2011, **157**, 1551.
57. G. Friedrich Cornelius, D. Rother, F. Bardischewsky, A. Quentmeier and J. Fischer, *Applied and Environmental Microbiology*, 2001, **67**, 2873-2882.
58. A. V. Mardanov, E. V. Gruzdev, D. D. Smolyakov, T. S. Rudenko, A. V. Beletsky, M. V. Gureeva, N. D. Markov, Y. Y. Berestovskaya, N. V. Pimenov, N. V. Ravin and M. Y. Grabovich, *Microorganisms*, 2020, **8**, 2030.
59. B. B. Jørgensen, A. J. Findlay and A. Pellerin, *Frontiers in Microbiology*, 2019, **10**, 849.
60. C. S. Martens and R. A. Berner, *Science*, 1974, **185**, 1167.
61. A. A. Raghoebarsing, A. Pol, K. T. van de Pas-Schoonen, A. J. P. Smolders, K. F. Ettwig, W. I. C. Rijpstra, S. Schouten, J. S. S. Damsté, H. J. M. Op den Camp, M. S. M. Jetten and M. Strous, *Nature*, 2006, **440**, 918-921.
62. E. J. Beal, C. H. House and V. J. Orphan, *Science*, 2009, **325**, 184.
63. C. Sparacino-Watkins, J. F. Stolz and P. Basu, *Chemical Society Reviews*, 2014, **43**, 676-706.
64. L. Y. Stein and G. W. Nicol, *eLS*, 2018, 1-9.
65. L. Schneider, A. Wüst, A. Pomowski, L. Zhang and O. Einsle, 2014, vol. 14, pp. 177-210.
66. G. Coutts, G. Thomas, D. Blakey and M. Merrick, *The EMBO Journal*, 2002, **21**, 536-545.
67. A. Yakunin and P. Hallenbeck, *Journal of Bacteriology*, 2002, **184**, 4081-4088.
68. A. Herrero, E. Flores and J. Imperial, Encylopedia of Micorbiology 4e, Nitrogen Assimilation in Bacteria 2019, 280-300.

5 Isolation of Environmental Bacteria

5.1 Hypothesis

Prior to commencing the work reported in this chapter it was hypothesised that by utilising a variety of bacterial isolation methodologies, *L.ochracea* and other filamentous bacteria could be isolated from BIOX samples. To test this hypothesis would require altering growth conditions such as nutrients present, altering the type of solid media used, utilising non- and semi-solid media and also attempting to mimic environmental growth conditions by including sample site water in isolation protocols.

5.2 Introduction

Iron microbial mats are a mixture of BIOX, bacterial EPS and the bacteria that produce them. They are found in both freshwater and marine environments, where there is a source of dissolved ferrous iron, and contribute to the detoxification of these environments via sorption of pollutants such as heavy metals and phospates.[1-3] The visually striking appearance of these mats has made them targets for microscopic studies for close to 200 years and *Leptothrix spp.* were some of the first bacteria to be discovered by microscopy.[4] Since their discovery there have been many attempts, both successful and unsuccessful, to isolate filamentous bacteria from these environments. There are now many axenic cultures of filamentous bacteria available to study including *Leptothrix cholodnii*, *Leptothrix mobilis, Leptothrix discophora* and *Sphaerotilus natans*.[5] These bacteria were all isolated by inoculating solid agar media with the appropriate biofilm, allowing colonies to grow, picking individual colonies and repeating this process until single axenic cultures were formed.[6, 7] Surprisingly, *Leptothrix ochracea*, the first documented filamentous bacterium and primary producer of BIOX, has never been isolated as an axenic culture. Multiple attempts have been made using both solid agar media and specially designed apparatus that allows flowing ferrous rich water to be continually added to samples (Figure 76), however all have been unsuccessful.[8, 9]
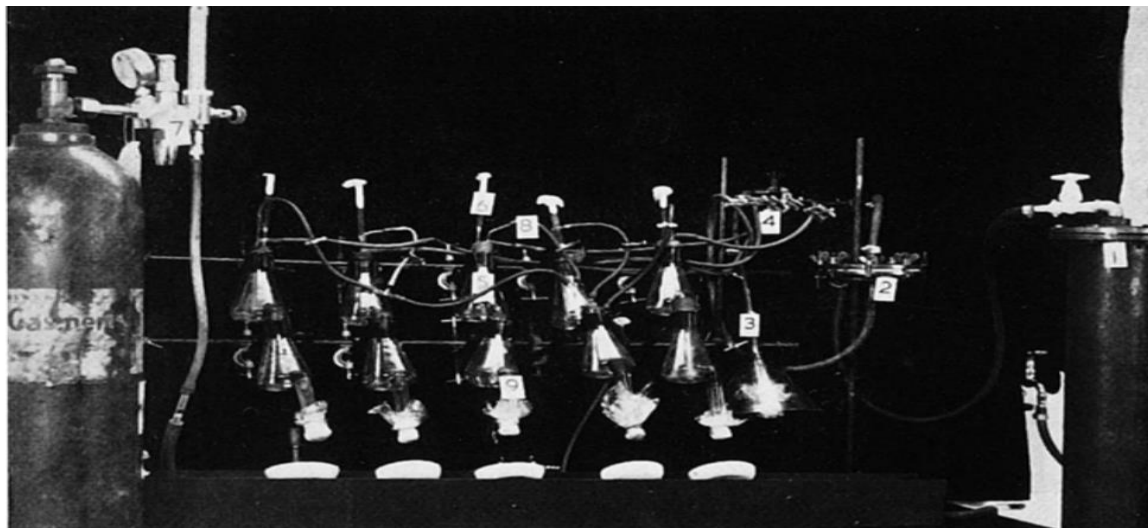
Figure 76 - Photograph of Mulder and van Veen's apparatus used to try and isolate *L.ochracea*. Ferrous rich waters were continually passed through flasks containing BIOX rich soils. Gas cylinders were used to control the oxygen concentration.[10]

Fleming *et al.* have recently published a study investigating the physiology of uncultured *L.ochracea* and have also published a draft genome.[11] To do this they created small growth chambers by gluing microscope slides together, plugging one end with an iron sulfide agarose mixture and overlaying it with 0.2 µm filtered sample site water. This creates opposing iron oxygen gradients and attempts to mimic what would be found in natural environments. Once inoculated with 1:100 BIOX dilutions the cells can then grow within regions of acceptable iron concentration. Figure 77 shows a schematic of their setup and the resulting growth images.
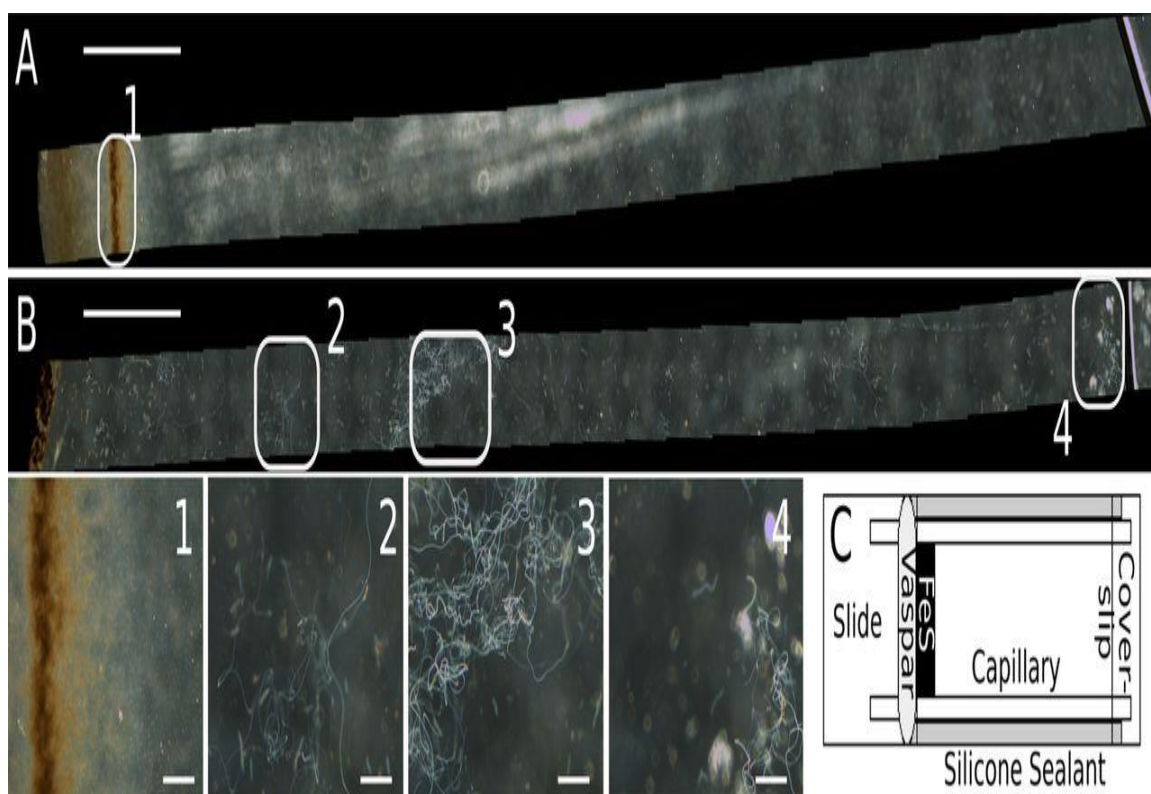
Figure 77 - Images from Fleming *et al*. showing *L.ochracea* microcultures growing in chambers designed to mimic environmental iron oxygen gradients. A and B scalebars are 0.5 mM and 1- 4 scalebars are 20 μm.[11]

From these images it can be seen that Fleming *et al*. were successful at growing *L.ochracea* in a controlled environment however they still could not isolate it as single culture. Interestingly, *L.ochracea* cells grew over a larger area than is typical for iron-oxidising bacteria (FeOB) such as *Gallionella* sp. When grown under these conditions FeOB tend to form tight bands around the oxic-anoxic interface whereas this growth band appears larger and more diffuse. This shows that *L.ochracea* is not confined to tight gradient boundaries and may be able to colonise a greater variety of environmental niches. A fluorescent *in situ* hybridisation (FISH) probe was subsequently designed for *L.ochracea* using 16S rRNA sequence data.[12] This probe allowed for fluorescent cell sorting to be used to isolate *L.ochracea* cellular material for the construction of a draft genome. Analysis of this genome has shown that *L.ochracea's* physiology is likely mixotrophic, meaning that it can derive energy from both the oxidation of inorganic compounds and utilisation of organic material. This indicates a key difference with other FeOB such as *Gallionella* sp., which are lithotrophic and derive energy solely through the oxidation of inorganic substrates, and heterotrophic FeOB, such as the other members of the *Leptothrix* genera, that metabolise organic carbon as an energy source.

The BIOX sites in the greater Glasgow area have not been investigated and so the following work aims to revisit the bacterial isolation methods and in particular growth media and isolation methodologies described for different *Leptothrix* spp. The growth of axenic cultures of bacteria will be invaluable for the characterisation of the EPS and hence BIOX found at these sites. This, and the isolation of filamentous bacterial, will permit the use of Illumina high-throughput sequencing to fully characterise the genomes of theses bacteria within BIOX mats.

5.3 Chapter Aims

To further understand the production of BIOX it is important that *L.ochracea* is isolated as an axenic culture. This chapter describes various approaches used to try and isolate *L.ochracea* and other bacteria from BIOX samples. A variety of microscopic methods including brightfield, darkfield, phase contrast and confocal laser scanning microscopy shall be used to examine the abundance, morphology and live/dead state of bacteria within the BIOX samples. Methodologies including solid media - agar and agar substituted, liquid enrichments, gradient tubes and micromanipulation were developed and used to isolate bacteria from BIOX samples. The genomic DNA from a number of samples was amplified and sequenced via high-throughput Illumina sequencing.

5.4 Results and Discussion

As discussed in Chapter 3 Section 1, current research shows that *L.ochracea* filaments have a variety of material applications making it an important bacteria for study. *L.ochracea* is the dominant FeOB found in BIOX mats and produces far more elegant looking filaments when compared with other FeOB such as *L.cholodnii* (Figure 78). The current lack of an axenic culture however makes it more challenging than studying related species such as *L.cholodnii* or *L.discophora* that grow well on solid and in liquid media supplemented with organics such as pyruvate or glucose. The organic sheath material of *L.cholodnii* and iron-oxidising ability of *L.discophora* have been well characterised due to this fact while the difficulties of studying organics associated with *L.ochracea* produced BIOX are highlighted in Chapter 3 Section 6.[13, 14] The following study employs various methodologies in an attempt to isolate *L.ochracea* from BIOX samples collected from the greater Glasgow area.
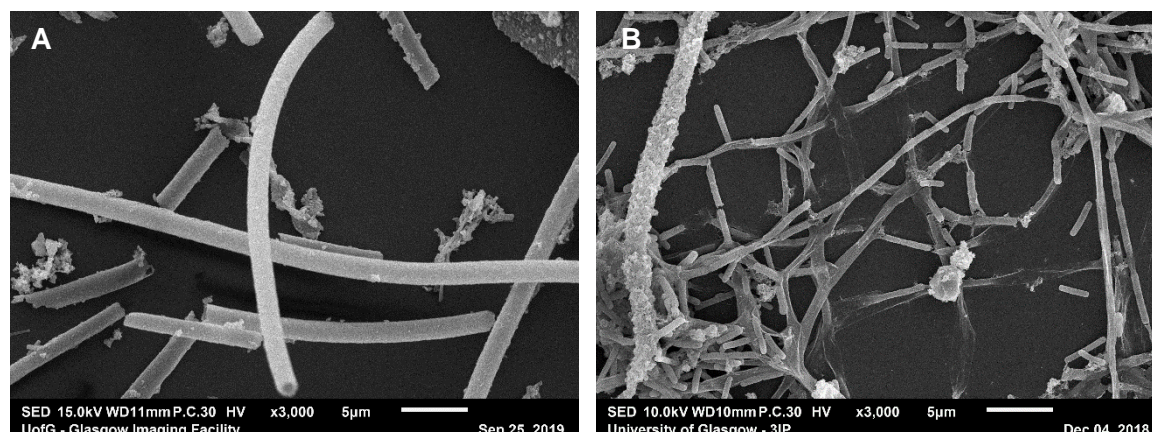
Figure 78 - SEM images at the same magnification of A - *L.ochracea* filaments and B - *L.cholodnii* filaments. These highlight the visual difference in macrostructure between filaments of *L.ochracea* and *L.cholodnii*.

5.5 Growth Experiments

Strains of filament forming *Leptothrix spp.* other than *L.*ochracea have in the past been isolated on solid media.[6, 10, 15] The most recent of these isolation studies that the author is aware of are by Sawayama *et al.* in 2011 and Schmidt *et al.* in 2014.[16, 17] All of these studies initially use solid media supplemented with ferrous iron which allows *Leptothrix spp.* to form irregular shaped colonies with filamentous edges that become orange with iron oxyhydroxide encrustation. These colonies are then used to inoculate liquid media of the same composition minus agar. The compositions of these media can be found in Chapter 2 Section 6.1. *L.cholodnii* was purchased and used for control experiments as it grows well on pyruvate and glucose supplemented agar. Figure 79 shows photographs of colonies formed in the study by Sawayama *et al.* and photographs of filamentous *L.cholodnii* colonies.[16]
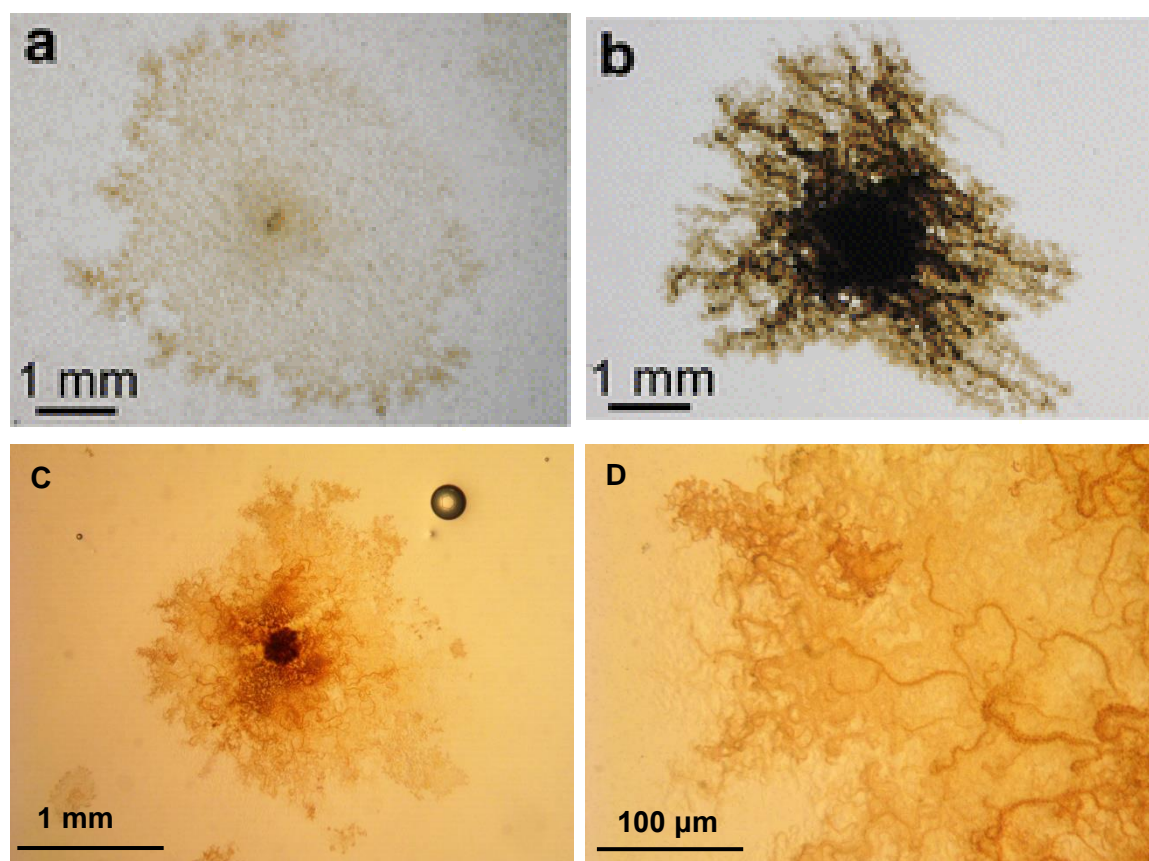
Figure 79 - Photographs of irregular shaped colonies produced by a *Leptothrix* sp. isolated by Sawayama *et al*. A - Unencrusted colony, B - Colony encrusted with manganese, another metal which can be oxidised by *L.cholodnii* and *L.discophora* but not *L.ochracea*, C – *L.cholodnii* colony encrusted with manganese and D – Magnified section of colony C showing filamentous edges.[16]

Brightfield microscopy was initially used to image liquid growths however it became apparent that this technique was not appropriate as the bacterial cells and filamentous material not encrusted with precipitates are transparent. Instead, phase contrast microscopy, which corrects phase changes of light and converts differences in refractive index into contrast, was used to image liquid growths. A comparative brightfield and phase contrast micrograph of some filamentous material can be seen in Figure 80.
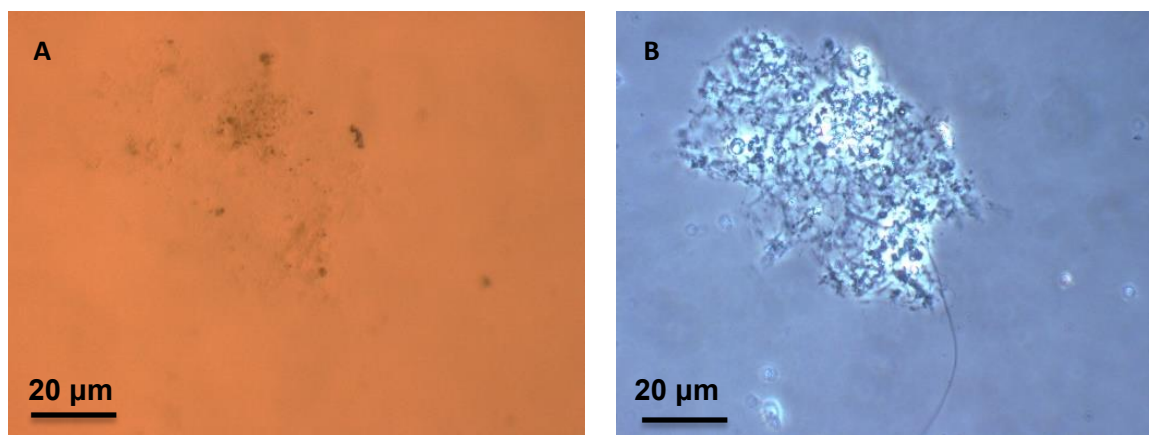
Figure 80 - A - Brightfield micrograph and B - Phase contrast micrograph of a small filamentous cluster highlighting the difficulty of visualising bacterially produced material by microscopy.

5.5.1 Allander Water BIOX Growth

BIOX was collected from sample sites within the vicinity of the Allander Water. Full descriptions and characterisations of these sample sites can be found throughout Chapter 6. For this initial experiment, Allander_1 BIOX (10 ml) was added to SIGP liquid medium (1 l), a ferrous iron rich medium developed by Sawayama *et al.* for growing their *Leptothrix* isolate.[16] Its composition mimics the inorganic composition of groundwater and it contains supplemental glucose and peptone as sources of carbon and nitrogen respectively. The full composition of all media used for this study can be found in Chapter 2 Section 6.1. This was a simple experiment to investigate whether filamentous bacteria would grow or not. After 18 h the medium had turned from pale yellow to cloudy indicating that there was a bloom of bacteria present. This rapid growth of bacteria however were not rod shaped and did not form filaments indicating that non-*Leptothrix* bacteria associated with BIOX easily outgrow *L.ochracea* under these conditions. It is likely that the supplemental glucose in the SIGP medium makes it more nutrient rich than the sample site water thereby promoting the rapid growth of heterotrophic, organic carbon utilising, bacteria.

As the growths in this study require an environmental inoculum it is important to supress the growth of eukaryotes such as cyanobacteria and fungi. Not only are these contaminants but they can also be filamentous which may add some confusion when assessing samples. To mitigate this all growth media is supplemented with cycloheximide (100 mg/ml in 60 % EtOH) at a concentration of 1 ml/l. This has been shown to reduce the growth of eukaryotes in environmental samples used to isolate *Leptothrix spp.* of bacteria.[17] Allander_1 BIOX was diluted 1:10 with 0.2 μm filtered sample site water and then used as inoculum for solid GP and SIGP agar media. Agar plates were prepared in

accordance with the protocol in Chapter 2 Section 6.1 and 20 µl aliquots of the diluent were streaked on each plate inside a laminar flow hood. Plates were then sealed and allowed to grow at room temperature until distinct colonies were visible. Multiple colonies and overlapping clusters had formed after 3 days. All colonies appeared reasonably circular when viewed at 20 × magnification and no indications of filamentous growth or ochreous deposits were seen. Figure 81 shows a selection of micrographs of single colonies formed and image F shows a photograph of a typical agar plate with an abundance of colonies on it. A selection of single colonies were picked and used to inoculate liquid media however no filamentous material was produced.
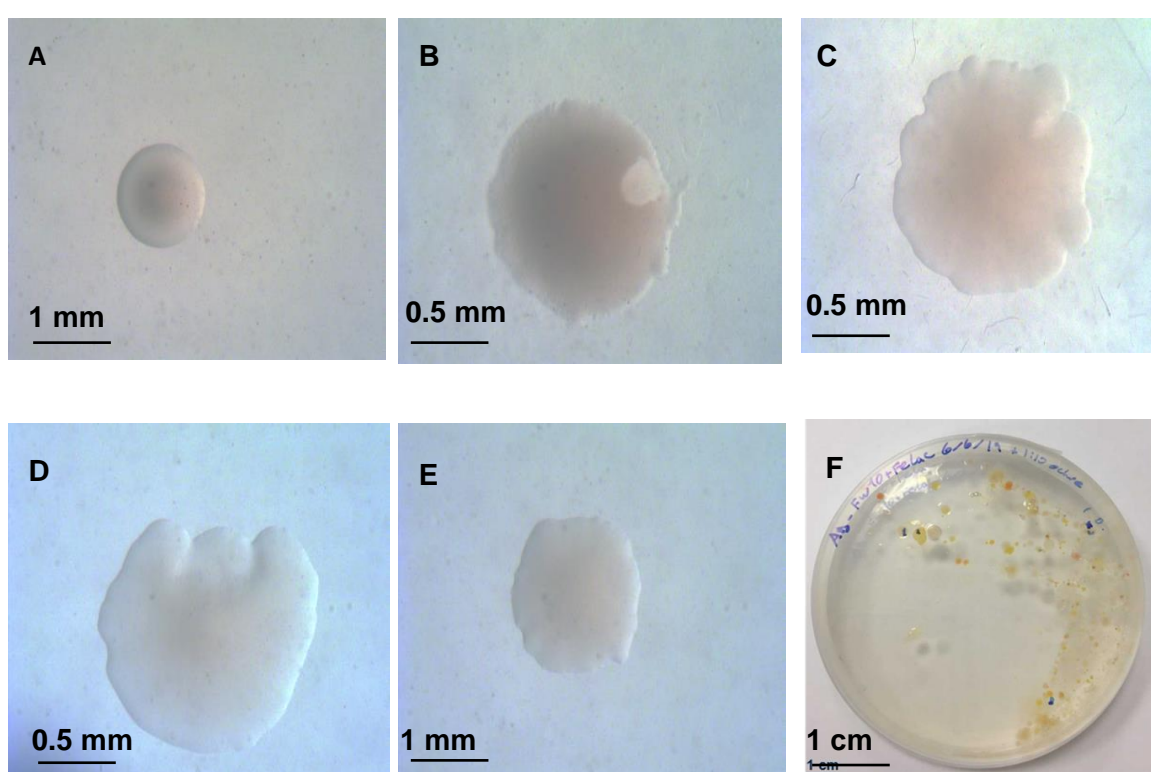
Figure 81 – A- E Micrographs of colonies growing on SIGP media using River Allander inoculum and F – Photograph of typical agar plate containing growing colonies.

5.5.2 Kelvin Walkway Growth

Kelvin Walkway BIOX was used as inoculum for solid GP and SIGP media. The resulting colonies and clusters (Figure 82) were mainly circular however some did have an irregular shape and some contained orange and brown colouration. Addition of these colonies and clusters to liquid media yielded small mats of filamentous growth after 3 days. This growth appeared visually quite different to what is expected for *L.ochracea* BIOX as the filaments are shorter, less than 30 µm in length, and more slender, less than 1 µm in diameter, with no noticeable encrustation. There does however appear to be amorphous looking

aggregation surrounding them which is likely iron oxyhydroxide that has begun to precipitate and become associated with the mat. These filamentous growths are likely not being produced by *L.ochracea* however were still used to inoculate both solid and liquid media. Further filamentous growths could not be obtained which is in agreement with van Veen *et al.* who state that *Leptothrix spp.* can lose their filament producing abilities after repeated isolation and growth experiments.[9]
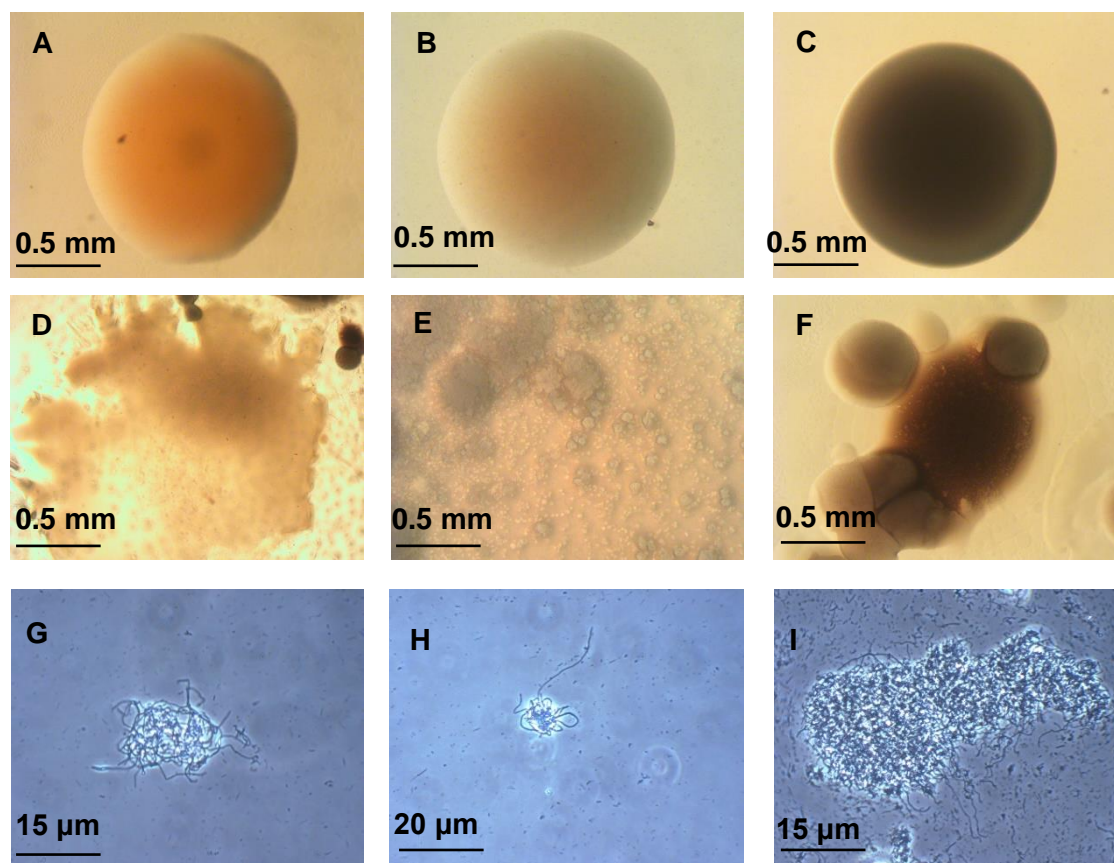
Figure 82 – A – F – Micrographs of colonies and G – I – Phase contrast micrographs of growing liquid samples.

5.5.3 Allanton Mine Effluent Growth

BIOX collected from the effluent of a disused coalmine in Allanton was used as inoculum for GP and SIGP media. Most colonies formed were circular however some were dark orange/brown in colour. These coloured colonies were picked and used to inoculate liquid media resulting in the growth of filamentous material. Micrographs of colonies and filamentous growth can be seen in Figure 83. The colonies in image C yielded the filamentous growth seen in images E and F. This filamentous material appears visually quite different to the BIOX filaments produced by *Leptothrix spp.* Isolated filaments appeared slightly thinner *Leptothrix spp.* and filaments of lengths less than 10 µm were

commonly seen. These filaments did also not appear to be visually encrusted with iron oxyhydroxide precipitates. Repeated growth and isolation of this bacterium did not affect its ability to produce filaments and it was also able to grow and produce filaments in iron free media indicating that this is not a *Leptothrix spp.* SEM imaging of the Allanton site in Chapter 3 Section 5 shows a variety of thinner filaments not seen in other sites so it is not surprising that an unusual filamentous material has been isolated.
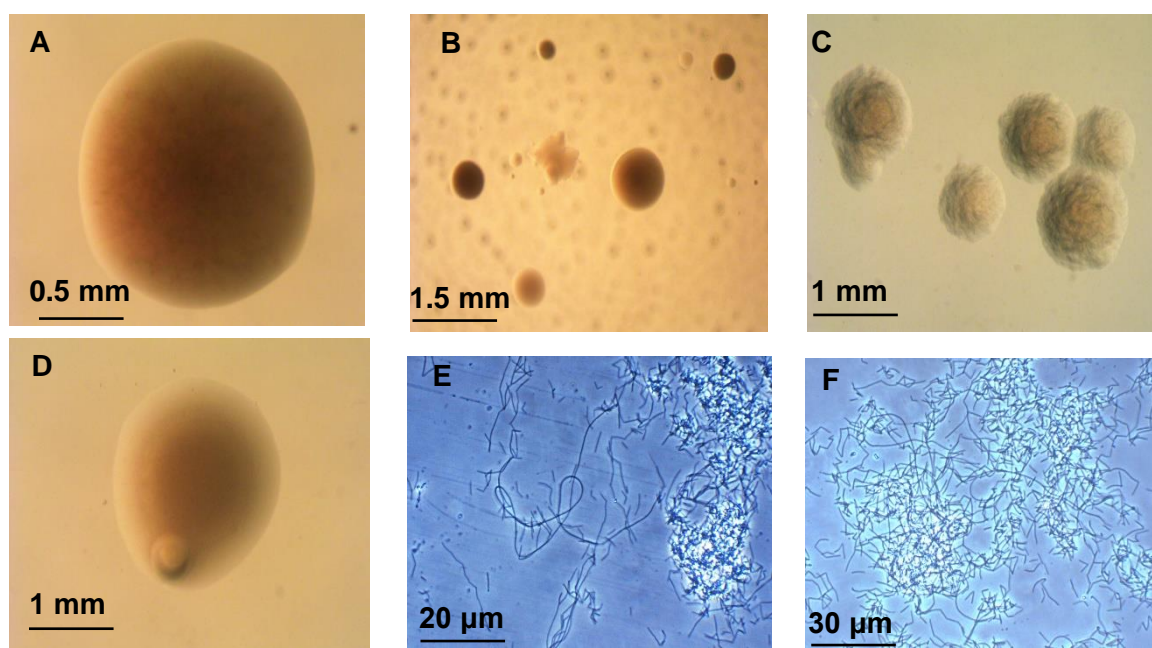
Figure 83 - A - D Micrographs of colonies and E - F - Phase contrast micrographs of liquid growths containing filamentous material grown from colonies in image C.

Isolation of an axenic culture allowed a lawn of bacteria to be produced on solid media which was subsequently sent to MicrobesNG for high-throughput Illumina genome sequencing. This resulted in the bacterium being classified as a potentially novel sub-species of *Arcicella* which is a genus of aerobic bacteria first isolated by Nikitin in 2004 and typically found in environmental water samples.[18] Interestingly there are no reports of any *Arcicella spp.* forming filamentous material while the Allanton isolate forms large mats of filamentous material. Comparing its 16S rRNA sequence to other *Arcicella spp.* using BLASTN at the NCBI (ref) shows that the closest published isolated relative is *Arcicella aurantiaca*, with a 99.36 % similarity, which was isolated from river water in Taiwan.[19] They have similar G+C contents of 35.8 % and 35.1 % respectively for the Allanton isolate and *A.aurantiaca*, these values are typical for *Arcicella spp.*[20] To identify whether the isolate was a new species or not the genome sequence was sent to the Type (Strain) Genome Server (TYGS) this showed that despite significant similarity to *Aricella aurantiaca*, the Allanton bacteria was a new species of *Arcicella*, Table 23 shows the

basic genome details of the two bacteria. The most notable metabolic difference between the two is that the Allanton isolate contains genes for inorganic sulfur assimilation while *A.aurantiaca* does not. This is not surprising as this bacterium was isolated from the effluent of a disused coalmine which will be rich in pyrite, FeS_2, meaning the water is likely rich in a variety of sulfur containing compounds.

Genome	*Arcicella sp acidmine*	*Aricella aurantiaca*
Taxonomy	Bacteria; Bacteroidetes; Cytophagia; Cytophagales; Cytophagaceae; Arcicella; Arcicella sp acidmine	Bacteria; Bacteroidetes; Cytophagia; Cytophagales; Cytophagaceae; Aricella aurantiaca
Size	6,167,621	5,933,091
GC Content	35.8	35.1
Number of Contigs	104	71
Number of Coding Sequences	5373	5317
Number of RNAs	41	42
RAST Subsystem Coverage	26 %	25 %

Table 23 - Genome data from TYGS for *Aricella* bacteria.

5.5.4 Gellan Gum Substituted Agar

'The great plate count anomaly' is a term coined by Staley and Konopka in 1985 that describes the anomalous result that the number of bacteria in an environmental sample is orders of magnitude larger than the number of colonies that form on agar media inoculated with the same sample.[21] It has been estimated that only 1 % of environmental bacteria are isolatable as single colonies on agar plates meaning 99 % of environmental bacteria cannot be isolated and grow as axenic cultures. A number of reasons have been

proposed for this anomaly, these include the difficulty of replicating metastable environmental conditions experimentally, many environmental bacteria requiring oligotrophic conditions meaning typically used nutrient rich medium are inadequate and the potential synergistic relationships that may exist within environmental samples.[22, 23] Bodor *et al*. have made progress at isolating recalcitrant bacteria from environmental samples by staining soil samples with SYBR green DNA stains and using a flow cytometer to isolate cells of size ca. 0.5 µm by fluorescence activated cell sorting. These cells are isolated in wells of a 96 well plate and allowed to grow for 5 months. Access to similar equipment was not available for this study, however studies by Tanaka *et al*., Tamaki *et al*. and Imazaki and Kobori proposed easy to make changes to culturing conditions that can improve culturability of bacteria.[24-26] Tanaka *et al*. have shown that autoclaving agar and phosphate salts within the same vessel can generate hydrogen peroxide within the media which in turn can suppress cell growth while Tamaki *et al*. and Imazaki and Kobori have shown that substituting agar for gellan gum and adding minimal nutrients can increase the number of *β*-proteobacteria isolated. As *Leptothrix spp*. are members of the class *β*-proteobacteria it was decided that these changes should be investigated. BIOX collected from the River Allander site was used to inoculate FW70-gellan and FW70-agar media. FW70 is a minimal media consisting of 0.2 µm filtered sample site water, tryptone, sodium pyruvate and supplemental ferrous iron. The exact composition can be found in Chapter 2 Section 6.1. Both sets of inoculated plates showed an abundance of growth however the FW70-gellan plate showed small irregular shaped orange colonies. These were picked, used to inoculate liquid media, allowed to grow statically at RT for 3 days and streaked on solid media. This methodology was repeated, eventually yielding a plate of small orange colonies produced by rod shaped cells. A selection of images can be seen in Figure 84. This plate was then sent to MicrobesNG for high-throughput Illumina sequencing.
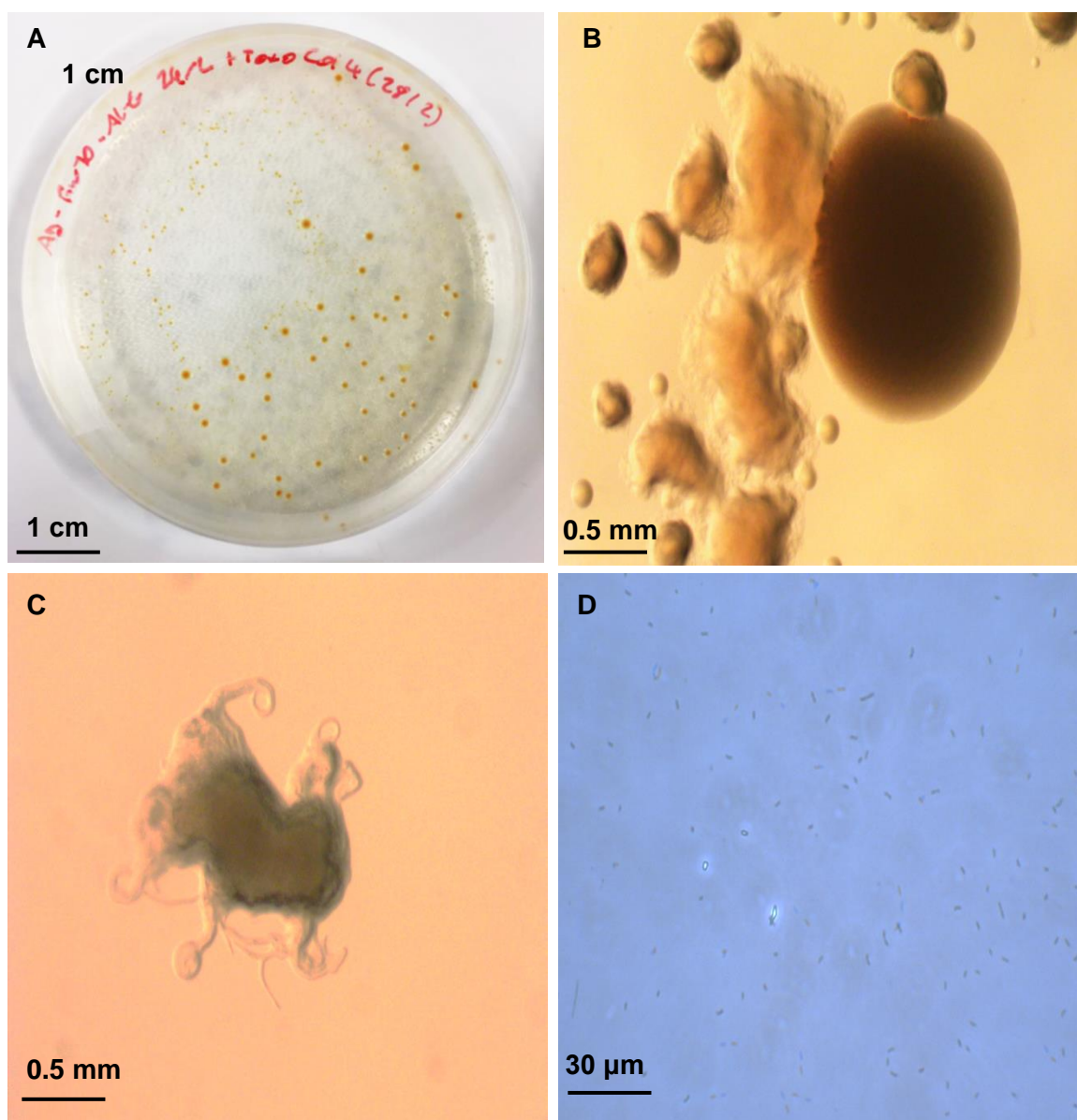
Figure 84 - A - Photograph of FW70-gellan plate of orange single colonies, B - C - Brightfield micrographs of irregular shaped colonies initially formed and D - Phase contrast micrograph of rod shaped cells growing in liquid media.

Surprisingly the resulting data indicated that the sample submitted was not a single species of bacteria, rather the plate contained three distinct bacteria, a *Mesorhizobium spp.*, a *Rhodoferax spp.* and a *Polaromonas spp.* Table 24 shows the TYGS genome data for these bacteria. This result was unexpected as the plate contained only one colony morphology and there appeared to be only rod-shaped cells in the liquid growth. However, all three of these bacteria have rod shaped cells so cell morphology cannot be used as an accurate indicator. Sequence similarity to other sequenced bacteria and nucleotide distribution (G + C content) which is consistent for a given strain allowed the individual genomes to be separated.

Rhodoferax spp. are also *β*-proteobacteria, are typically photoheterotrophic and are consistently found in slow moving water systems and waste effluents.[27] They are capable of iron reduction and have been found in freshwater iron seeps within BIOX mats.[12] *Polaromonas* is another genus of *β*-proteobacteria. They are defined as psychrophiles meaning they thrive at temperatures as low as – 20 °C and have typically been found in glacial environments, particularly at high elevation.[28] The closest related cultured bacterium to this *Polaromonas spp.* is *Rhodoferax saidenbachensis* with a BLAST sequence similarity of 98.77 %. *Mesorhizobium* is a genus of nitrogen fixing *α*-proteobacteria, often found in soils and wastewaters, that can form symbiotic relationships with leguminous plants. They have previously been reported in environments along with *Leptothrix spp.* so it is not surprising that a species has been isolated here.[29, 30] This *Mesorhizobium* isolate shares a 100 % BLAST sequence identity with *Mesorhizobium terrae*, which was isolated from a soil sample in South Korea (unpublished).[31] The genome data from TYGS indicates that both the *Polaramonas* and *Rhodoferax* bacteria are new species while the *Mesorhizobium* bacteria is a subspecies of *Mesorhizobium terrae*.

Genome	*Mesorhizobium terrae sub allander*	Strain *Polaromonas allander*	Strain *Rhodoferax allander*
Taxonomy	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Phyllobacteriaceae; Mesorhizobium; Mesorhizobium terrae sub. allander	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Polaromonas; Polaromonas sp. allander	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Rhodoferax; Rhodoferax allander
Size	5,616,512	4,047,686	3,784,850
GC Content	63.2	54.7	59.9
Number of Contigs	42	73	172
Number of Coding Sequences	5506	4077	3673
Number of RNAs	48	57	33
RAST Subsystem Coverage	26%	27%	28%
Nearest Relative (% Nucleotide Sequence Identity)	*Mesorhizobium terrae* (100)	*Rhodoferax saidenbachensis* (98.77)	*Rhodoferax sp.* K151

Table 24 - Genome data from TYGS for bacteria isolated from Allander.

Even though no *Leptothrix spp.* was isolated, this protocol has confirmed that replacing agar with gellan gum allows the isolation of different bacteria than could be isolated on agar. Optimisation of this protocol may allow further isolation of recalcitrant bacteria.

5.5.5 Enrichment Methods and Gradient Tubes

To obtain cultures of bacteria from the environment requires the growth and isolation within the laboratory. The use of solid media has the advantage that a single bacterium generates a discrete colony that can be picked and thus isolating it from the other bacteria in the sample. However, as described above not all bacteria can be cultured this way therefore it is necessary to consider methods using different types of liquid culture such as enrichment cultures. For FeOB environmental samples this involves using 0.2 μm filtered sample site water supplemented with iron salts, such as ferrous sulfate, ferrous ammonium sulfate, ferrous chloride or ferrous lactate at concentrations of ca. 100 μM, as growth media. BIOX samples (1 ml) were added to growth media (1 l) 1 l and allowed to grow statically at room temperature until appreciable growth was seen. Orange precipitates could be seen after 24 h however observing them via microscopy showed that these precipitates were likely clusters of abiotic iron oxyhydroxide. In an attempt to slow the abiotic oxidation of dissolved ferrous iron the growths were partially degassed and allowed to grow at 4 °C. 4 °C generally slows the growth of bacteria however these samples are collected from groundwater that is routinely below 10 °C indicating that growth at lower temperatures is possible. After three days of incubation there was again orange precipitates seen however this time they had a more fluffy appearance to them which is typical of actively growing BIOX samples. Phase contrast microscopy was then used to confirm that these orange precipitates contained chains of rod-shaped cells. While viewing these chains of cells they appeared to move through the media and within the BIOX indicating that they were alive. These samples continued to grow well at 4 °C over a period of ca. 14 days, after which the biomass did not appear to grow further. Serial dilutions of these enrichments were prepared in an attempt to yield purified samples of difficult to isolate bacteria. Unfortunately, this methodology was unsuccessful and no enrichment that could be sequenced was produced. Figure 85 shows a phase contrast micrograph of chains of cells extending from a BIOX mat.
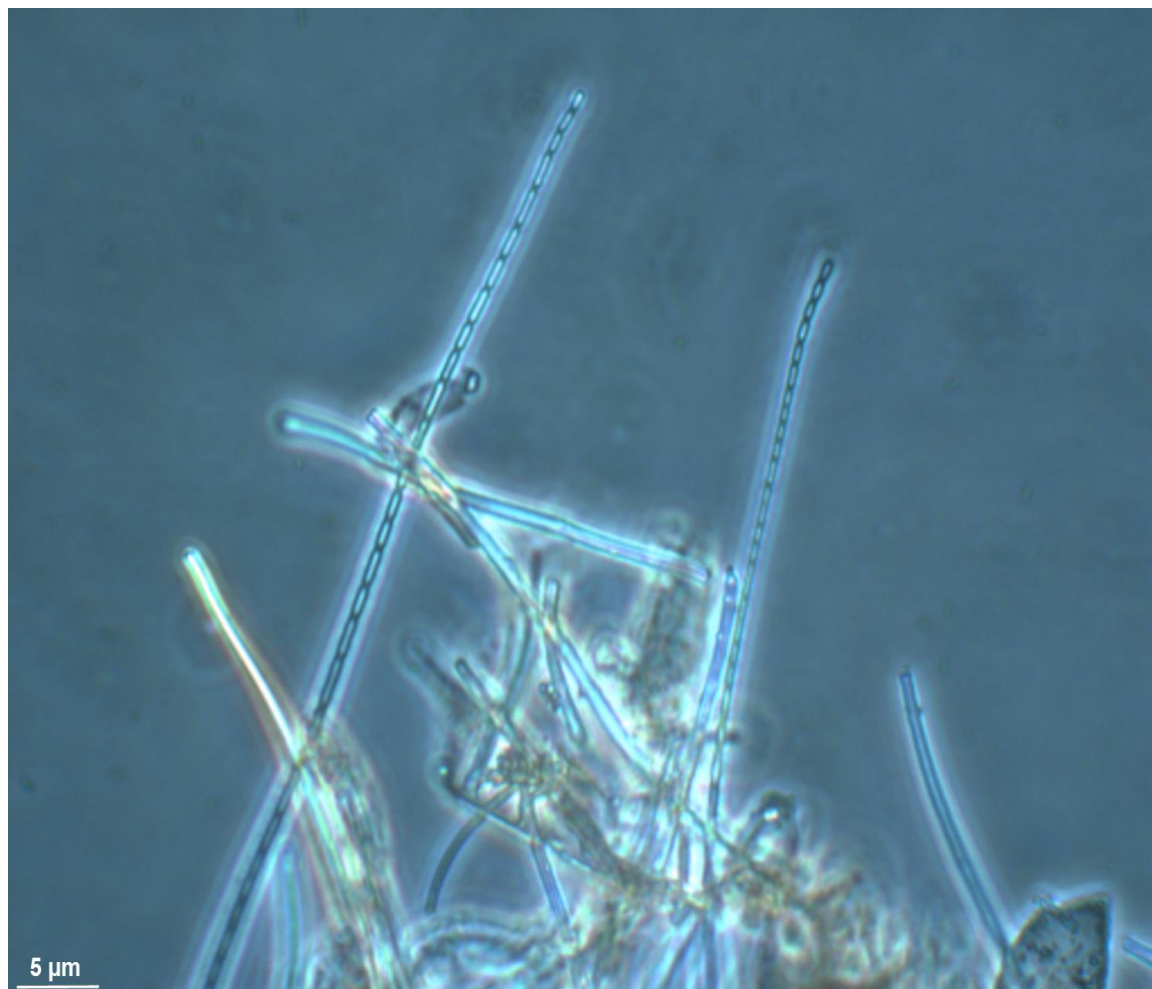
Figure 85 - Phase contrast micrograph of liquid enrichment growth showing chains of rod shaped cells extending from biofilm.

The gradient tube method for isolating FeOB has been pioneered by the Emerson group.[32] This protocol involves suspending BIOX samples in dilute agarose minimal media overlaying a slurry of FeS. The FeS slurry creates a Fe (II) gradient within the media and oxygen diffusing from a small headspace creates an oxic gradient. This attempts to mimic conditions within the environment in which FeOB are frequently found. This methodology was followed repeatedly using fresh BIOX from a number of sites and repeated however these samples did not produce rust-coloured bands indicating the growth of FeOB. Instead blooms of diatoms, a type of microalgae with a siliceous cell wall, were frequently produced. Figure 86 shows photographs of some initial gradient tube setups. This approach, although successful in other labs, was not pursued further.

Figure 86 - Photographs of initial gradient tube set ups.

5.6 Microscopy

5.6.1 Phase Contrast Microscopy

Freshly collected BIOX from each sample site was imaged using phase contrast microscopy. Conventional microscope slides and cover slips were used and sample volumes typically measured 15 µl. A variety of filamentous bacteria were noted using this methodology, phase contrast micrographs of these can be seen in Figure 87, however it became apparent that samples dried out during extended periods of viewing and imaging. This made it difficult to focus on areas of interest for longer than ca. 1 hour. To solve this problem silicon gaskets were adhered to the microscope slides, creating a small pool that allowed sample volumes of 600 µl. These could then be covered with longer coverslips and allowed BIOX mats to be viewed with less disruption to the macrostructure as the coverslip does not completely flatten the sample. Plastic boxes (6 cm × 6 cm × 0.5 cm) had glass beads added to them which were then overlayed with water that would provide a moist atmosphere once sealed with a lid. The covered microscope slides could then be rested on top of the beads and stored at 4 °C meaning the same BIOX sample could be viewed over multiple days. When viewing these samples it was noted that the longer the sample was left undisturbed at room temperature the more likely it was to notice large filamentous bacteria became. After this observation was first noted all further samples were allowed to stand at room temperature for 1 hour prior to imaging. This resulted in large filamentous bacteria regularly extending from and moving through the BIOX mats. These filaments eventually left the mats and were seen to glide through the media.
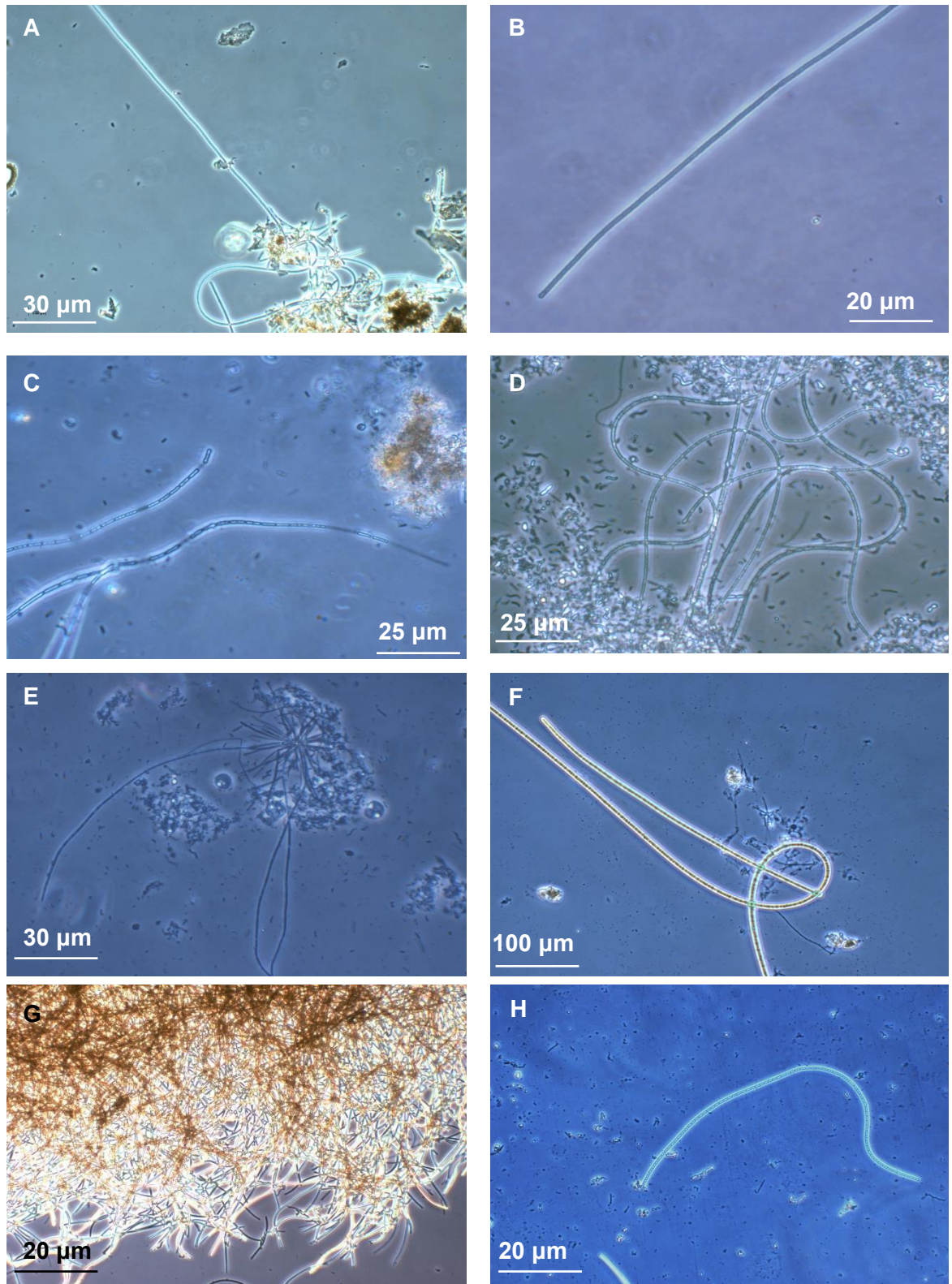
Figure 87 - Phase contrast micrographs of A - Filament extending from BIOX mat, B - Higher magnification image of filament, C – Filament containing rod shaped cells and intracellular spaces, D – Large winding filament containing rod shaped cells with granules seen on the sheath. This could potentially be a *Thiothrix spp.*, E – Rosette with filaments extending from it. This could potentially be another *Thiothrix spp.*, F – Large filamentous cyanobacterium, G – BIOX mat of *L.ochracea* filaments and H – Filament containing large rod shaped cells.

These phase contrast images show that a variety of filamentous morphologies were noted. Image C shows a filament ca. 1- 2 µm in diameter, that does not appear to be encrusted with any ferric material, containing rod-shaped cells and intracellular spaces. This description is consistent with that of *Leptothrix spp.* however this morphology was only seen occasionally. Images D and E show filamentous that potentially belong to *Thiothrix spp.*, a filamentous sulfur oxidising bacterium, that is known to produce filaments coated with sulfur granules and rosettes from which filaments can extend.[33] Image F shows a large filamentous cyanobacterium with diameter ca. 10 µm. Data from Chapter 4 Section 4.2 shows that a cyanobacterium of the family *Phormidiaceae* is highly abundant in the Allander_2 sample site. The cyanobacteria in image F matches with literature images of *Phormidiaceae* filaments. Image G shows a typical *L.ochracea* BIOX mat with filamentous edges while image H shows a filament of larger rod-shaped/bead-shaped cells shown. Images A and B show filaments, of diameter ca. 1 – 2 µm, that contain rod-shaped cells. These filaments were continually seen associated with BIOX mats, more often than the other imaged filaments, and would extend from mats and move through the media once left at room temperature for 1 hour as discussed previously. This filament morphology does not appear to contain ferric encrustation however the dimensions and cell morphology match that of *Leptothrix spp.* As this filament morphology is noted so regularly it was postulated that this may be an actively growing *Leptothrix spp.*

5.6.2 Confocal Laser Scanning Microscopy

Phase contrast microscopy was used to highlight that a variety of filamentous bacteria are present with the BIOX mat samples. In some cases it as possible to view cells within the filaments however it is not possible to accurately describe the material as live or dead. Visualisation of live and dead cells is possible by using confocal laser scanning microscopy (CLSM) combined with nucleic acid staining. To do this, the mats are stained with a LIVE/DEAD ™ *Bac*Light ™ Bacterial Viability Kit from Invitrogen ™. This contains a nucleic acid staining mixture of Syto9 and propidium iodide. Syto9 is a membrane permeable dye which stains the DNA of both live and dead cells while propidium iodide is a membrane impermeable dye which stains the DNA of dead cells that have damaged membranes. These stains have similar absorption profiles and differing emission profiles meaning that excitation from a single source, in this case a 488 nm argon laser, allows the differentiation of live and dead cells. Initial imaging highlighted that a large proportion of the cells were lysing. This is potentially since these bacteria thrive in oligotrophic environments of low ionic strength meaning that addition of high concentration of dye

solutions was found to be toxic. To resolve this the dyes were diluted and a dilution factor of 25 × was found to be optimal at minimising cell lysis. CLSM micrographs were captured of samples using both the microscope slide/gasket set up and regular microscope slides. This should allow visualisation of cells within the BIOX mats as well as within the filaments. Figure 89 shows a CLSM micrograph of a section of a BIOX mat imaged using the gasket set up while Figure 89 and Figure 90 show CLSM micrographs of filamentous material imaged using regular microscope slides.
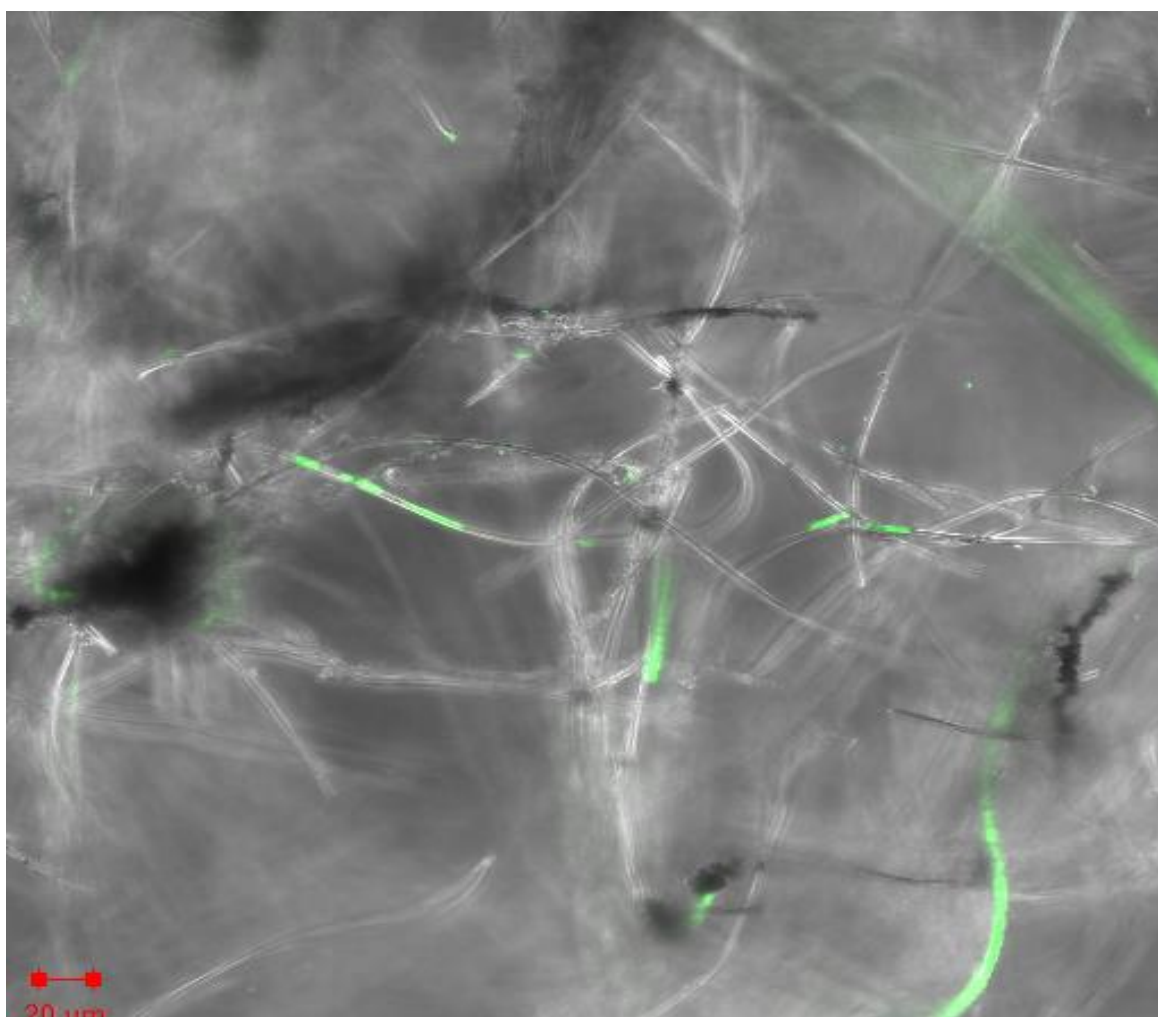
Figure 88 – CLSM micrograph of a section of BIOX mat. Live cells within the mat are coloured green and overlayed on a black and white image of the sample.
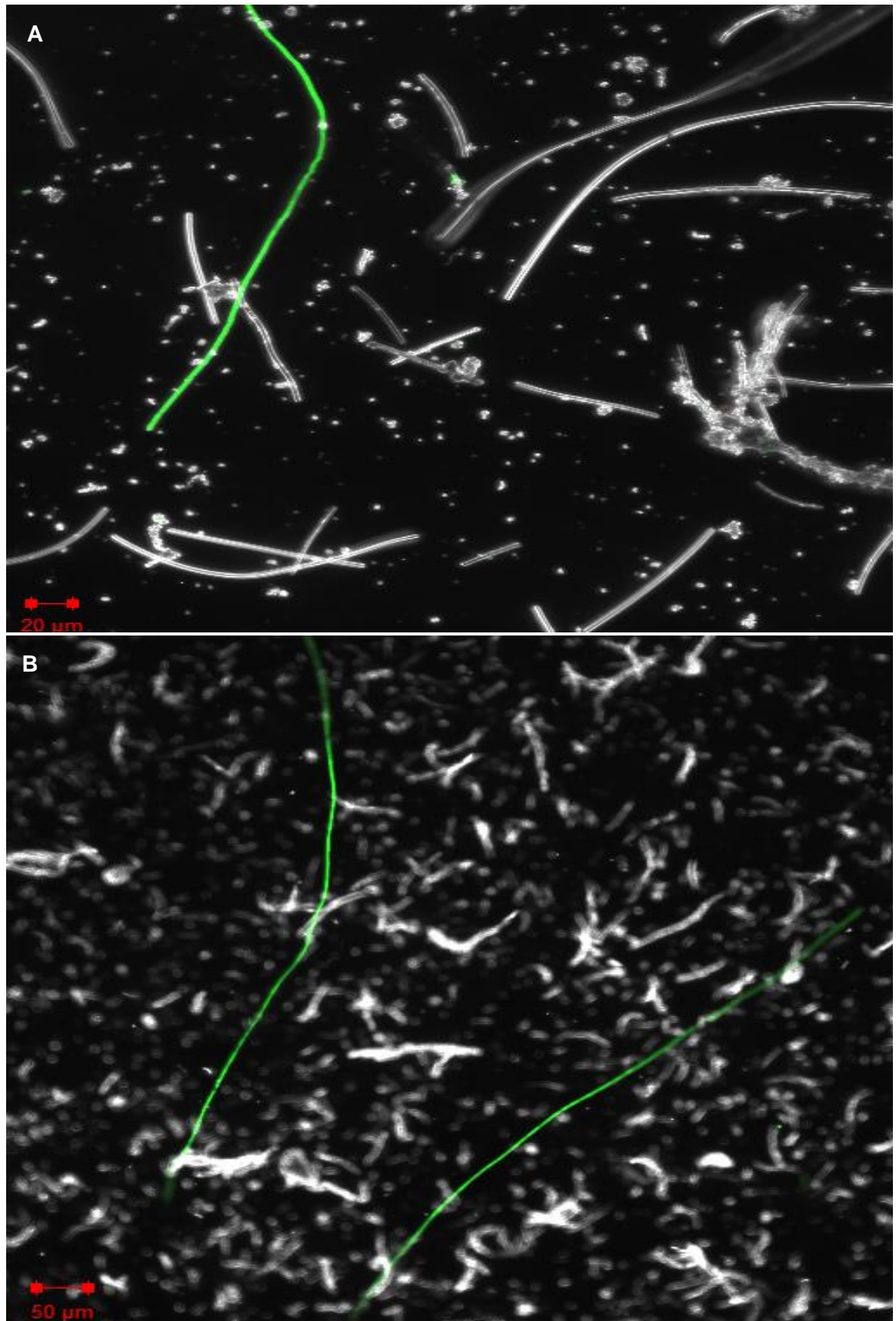
Figure 89 - CLSM micrographs showing live cells (green) within filaments overlayed on black and white micrographs of the sample.
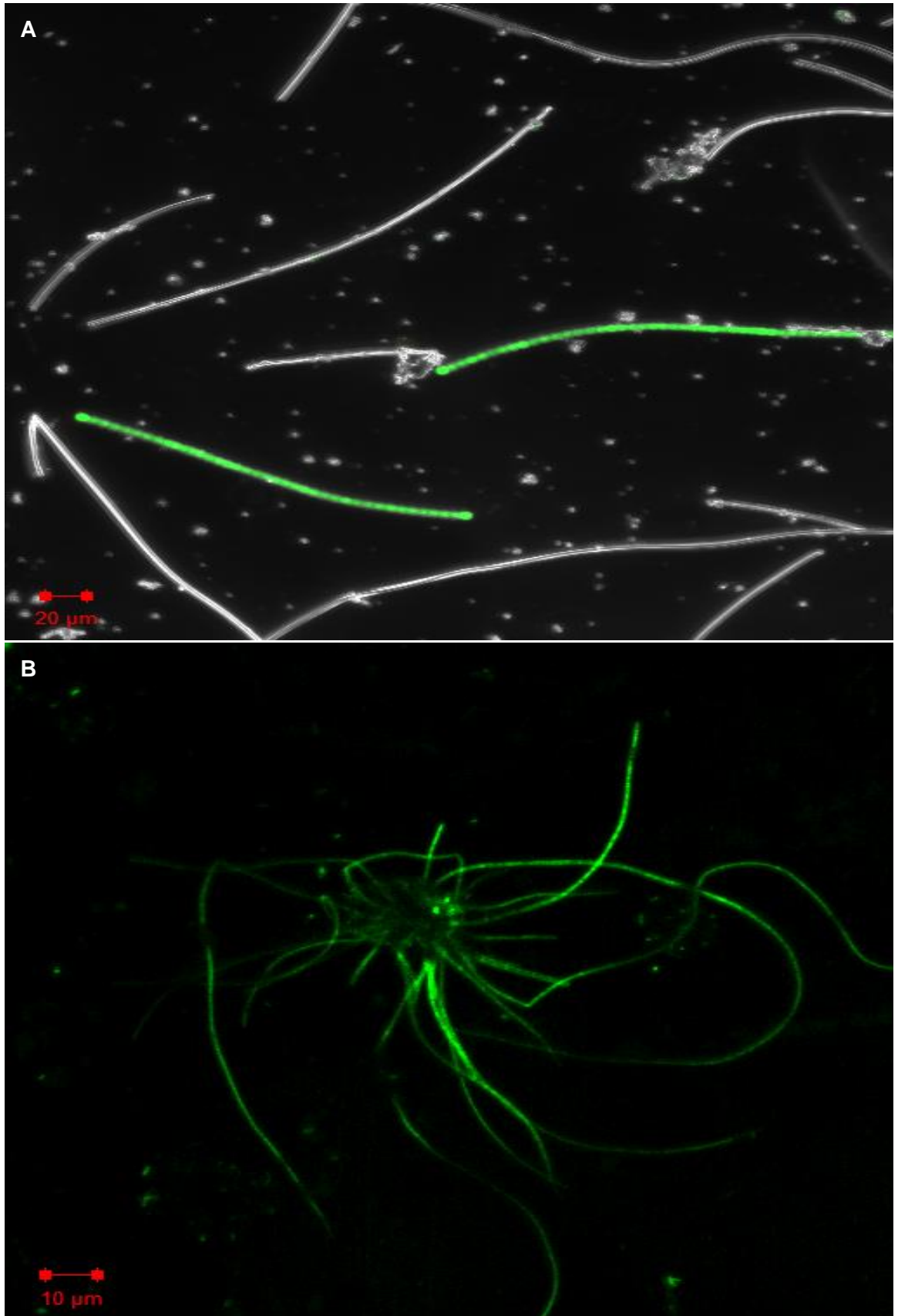
Figure 90 - CLSM micrographs. A - Chains of bead shaped live cells and B - Live cells likely belonging to a *Thiothrix spp.*

These CLSM micrographs show that live cells are present within filaments and that live cells at different depths within the mats can be visualised. Furthermore, it is apparent that only a minority of filaments contain live cells while the rest of the filamentous material is empty. This result is supported by literature examples stating that up to 90 % of filaments are void of cells.[12]

The combination of phase contrast microscopy and CLSM has shown that BIOX mats contain a variety of filamentous bacteria and that these filaments can contain live cells. As particular filament morphologies are regularly seen associated with BIOX mats and share similar dimensions and appearance with those of *Leptothrix spp.*, it was decided that attempts should be made to isolate these. Previous attempts at cultivation and isolation of filamentous bacteria from these samples have focused on producing single colonies on solid media and have generally been unsuccessful indicating that a different methodology is required.

5.7 Micromanipulation

As the filaments of interest are much larger than single bacteria, and frequently seen free floating in media, it was hypothesised that they may be able to be isolated via a micromanipulator. If successful, this may allow the genome sequencing of difficult to culture filamentous bacteria. Micromanipulation has previously been utilised by Pfeffer *et al.*, to isolate filamentous cable bacteria from marine sediment, indicating that this methodology may be applicable here.[34] A micromanipulator was attached to a microscope that had been used to image samples. Images and a description of this set up can be found in Chapter 2 Section 6.3.

Initially, 600 µl of sample site water containing a small amount of BIOX was added to a microscope slide fitted with a silicon gasket. The micromanipulator was then used to try and isolate filamentous bacteria however the high background of satellite bacteria made this difficult. A pre-filtering method was subsequently developed to mitigate this problem. Firstly, BIOX samples were washed through a 500 µm mesh net, using 0.2 µm filtered sample site water, to remove large particulates. The filtered samples were then sequentially washed through 100 µm and 25 µm mesh netting. This methodology was successful at removing particulate matter such as dirt and grit from the samples however it actually concentrated the background of satellite bacteria, making it more difficult to isolate a single filament without also extracting an abundance of contaminating bacteria.

While carrying out this filtration step it was noted that the mesh nettings still had small amounts of BIOX adhered to them. The pre-filtering method was then adjusted to include a step where BIOX is loaded on to mesh netting. Step-by-step photographs can be seen in Figure 91.
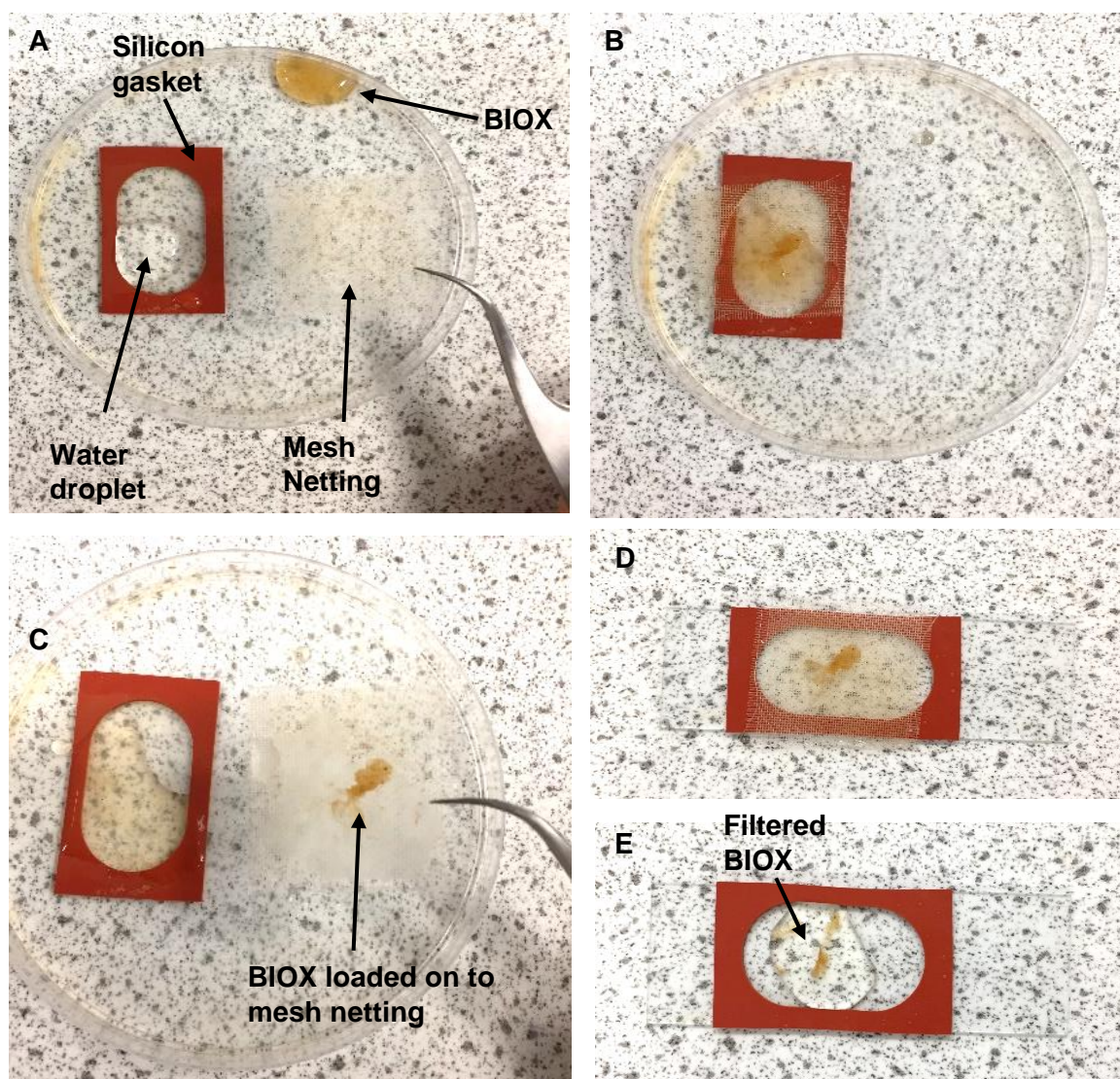
Figure 91 – Step-by-step photographs of the method used to successfully filter BIOX samples by loading them on to mesh netting and washing them. A – BIOX ready to be loaded on to netting. B – BIOX loaded, C – Loaded netting removed, D – Loaded netting added to fresh microscope slide with gasket attached and E – BIOX washed from netting on to microscope slide.

The following is a step-by-step follow along of the above figure. Prior to this all equipment used was sterilised under UV light for 15 minutes and fresh netting was used for each experiment. Gloves were worn throughout this procedure and as this work was carried out during the Corona virus pandemic, facemasks were also worn which may have aided with sterility.

Step A – A water droplet (250 µl) is added to a gasket adhered to a Petri dish. The BIOX to be filtered is also added to the Petri dish but outside the gasket.

Step B – The water droplet is overlayed with the mesh netting and the BIOX sample is gently pipetted on top. Lifting the netting with a pair of tweezers and dabbing it on the water droplet pulls the majority of the material into the water and load a fraction on to the netting.

Step C – BIOX is visible on the netting once it is removed from on top of the gasket.

Step D – The netting is flipped over, so that the loaded BIOX is face down, and used to cover a gasket adhered to a fresh microscope slide.

Step E – Wash the netting with 500 – 600 µl of the desired solvent, typically $_dH_2O$, to remove the filtered BIOX then cover with a coverslip.

This methodology was successful at removing particulate matter and reducing the concentration of background satellite bacteria. A micromanipulator was then used to extract single filaments. To further purify the sample and ensure that no contaminating bacteria were extracted, single filaments were then added to sterilised $_dH_2O$, extracted again, and added to sterile solution of PBS (10 mM). Phase contrast micrographs of isolated filaments can be seen in Figure 92. Please refer to Chapter 2 Section 6.3 for the complete experimental procedure of sampling using the micromanipulator.
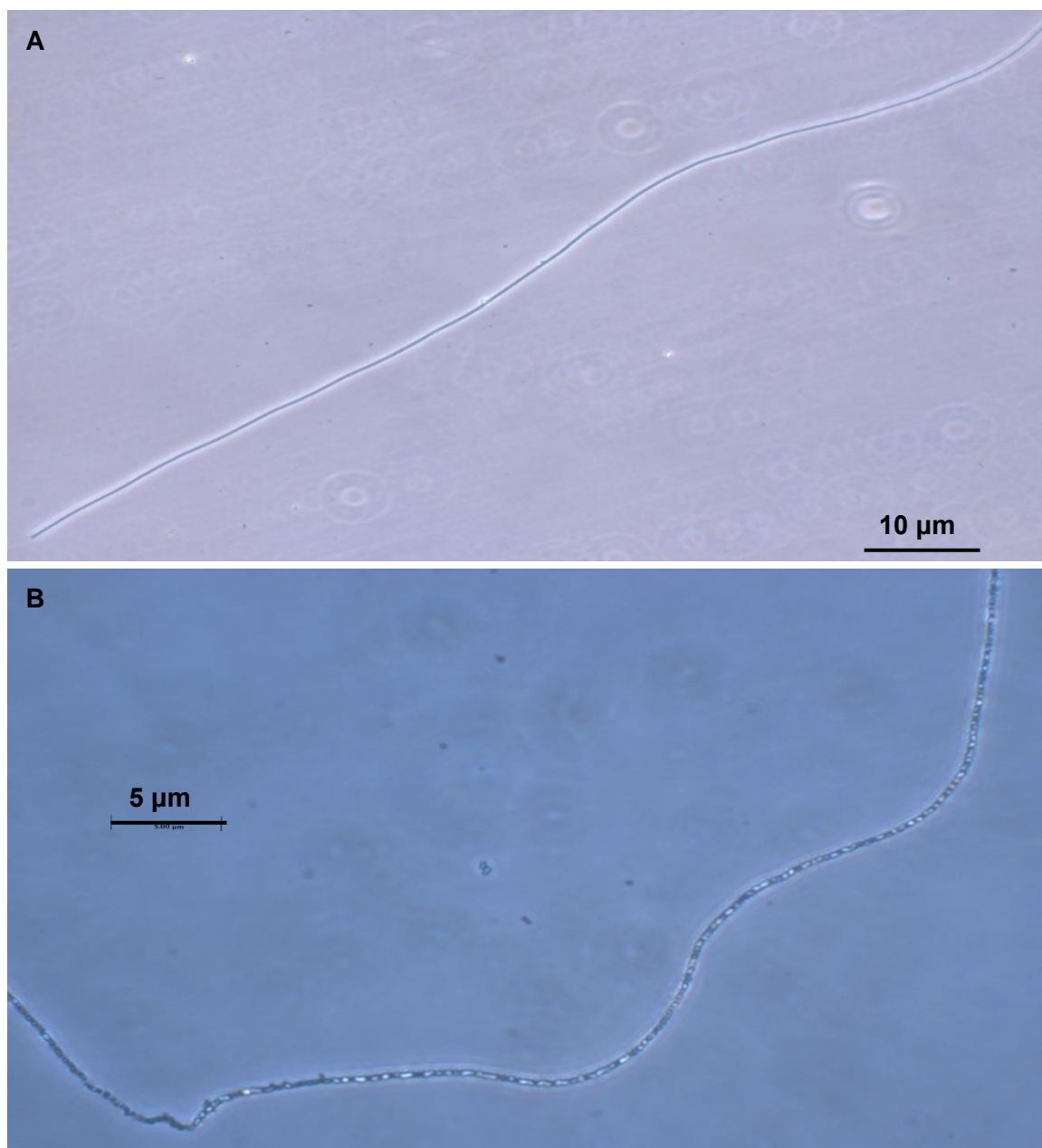
Figure 92 – Phase contrast micrographs of isolated filaments. B is likely a *Thiothrix spp.* due to the granular appearance of the filament.

Four filaments with distinct morphologies were isolated using this methodology. A thick *Leptothrix* like filament and a chain of beaded cells were isolated from BIOX collected from Dawsholm Park while a thinner *Leptothrix* like filament and *Thiothrix* like filament were isolated from Allander_3 BIOX. Descriptions and characterisation of sample sites can be found throughout Chapter 6. Isolated filaments were added to gaskets containing SGP and MSVP media that had been diluted 10 ×. These were left at 4 °C and at room temperature however no growth was noted. No further growth experiments were carried out using this methodology however further work is warranted as the correct media or

conditions may allow the isolation of an axenic culture of filamentous bacteria that cannot be isolated on solid media.

A REPLI-g Single Cell Kit was purchased from Qiagen that allowed the PCR amplification of genomic DNA from single cells. Amplified DNA was purified on a 10 kDa spin column and eluted in Tris.HCl (10 mM, 500 µl, pH 8). A Nanodrop spectrophotometer was used to estimate the quantity and purity of the DNA. Purified DNA was then sent to MicrobesNG for high-throughput Illumina sequencing. A reference for the amplification and purification procedure can be found in Chapter 2 Section 9 and a summary of the isolated bacteria can be found in Table 25.

Sample Location	Isolate Name	Filament Description	Proposed Genus	Nearest Isolated Relative (% similarity)	G + C (Mol%)
Dawsholm	Daws_1	Thick *Leptothrix* like	*Methylomonas*	*Methylobacter tundripaludum* (93.76)[35]	40.93
Dawsholm	Daws_2	Beaded Cells	*Methylomonas*	*Methylomonas methanica* (95.00)[36]	42.35
Allander_ 3	Al_1	Thin *Leptothrix* like	*Methylomonas*	*Methylobacter marinus* (94.63)[37]	40.70
Allander_ 3	Al_2	*Thiothrix* like	*Caulobacter*	-	66.20

Table 25 - Summary of isolated bacteria.

Daws_1 is of the genus *Methylomonas* and its nearest isolated relative is *M.tundripaludum*, a methane oxidising bacterium isolated from wetland soil in Svalbard, with a 16S sequence similarity of 93.73 %.[35] Daws_2 is also of the genus *Methylomonas* and its nearest isolated relative is *M.methanica*, a methane oxidising bacterium isolated from seawater collected off the coast of Penarth, United Kingdom.[36] Daws_2 shares a

95.00 % sequence identity with this bacterium. Al_1 is also of the genus *Methylomonas* and its nearest isolated relative is *M.marinus*, which is also a marine methane oxidising bacterium.[37] Al_1 shares a 94.63 % sequence identity with this bacterium. BLAST searches show that all isolates have relatives of greater similarity than those shown above, however these are uncultured bacteria that have been detected during 16S rRNA studies of environmental samples. Interestingly closest non-isolated relatives of Daws_1, Daws_2 and Al_1 have been found during studies of freshwater BIOX mats.[38, 39] This is in agreement with data from Chapter 4 Section 4 which shows that BIOX mats contain an abundance of methanotrophic γ-proteobacteria. Recent studies have shown the 2-Fh in aquatic and soil environments can enhance methanogenesis via degradation of environmental organics and by biogenic processes.[40, 41] This indicates that BIOX mats may contain a continually replenishing supply of methane, therefore allowing them to be an ideal habitat for methanotrophic bacteria.

The above results show the development of a protocol that combines phase contrast microscopy, CLSM and micromanipulation to successfully isolate single filaments of filamentous bacteria from BIOX mats. Furthermore, it shows that it is then possible to amplify and sequence the isolates genomic DNA. The bacteria currently isolated via this methodology are methanotrophic γ-proteobacteria and not *L.ochracea*, however with further optimisation it may be possible to isolate *L.ochracea* via this protocol.

5.8 Genomic Comparison of Isolated Bacteria

5.8.1 Carbon Utilisation

5.8.1.1 Micromanipulator Isolates

Daws_1, Daws_2 and Al_1 are likely all methanotrophic γ-proteobacteria. Methanotrophs in the γ-proteobacteria class are typically type I methanotrophs meaning that they can utilise methane as a source of energy. All three of these bacteria contain genes encoding for particulate methane monooxygenase (pMMO), a membrane bound enzyme that converts methane to methanol. pMMO has been shown to contain multiple copper centres, both in the Cu (I) and Cu (II) oxidation states, with the functional form of the enzyme being called partly reduced.[42] Methane is converted to methanol via Equation 16.

$$CH_4 + O_2 + NADPH + H^+ \longrightarrow CH_3OH + NADP^+ \, H_2O$$
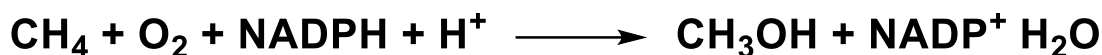
Equation 16 - pMMO conversion of methane to methanol.

Figure 93, adapted from Hanson and Hansnon, shows a potential pathway for the oxidation of methane and assimilation of formaldehyde.[43]
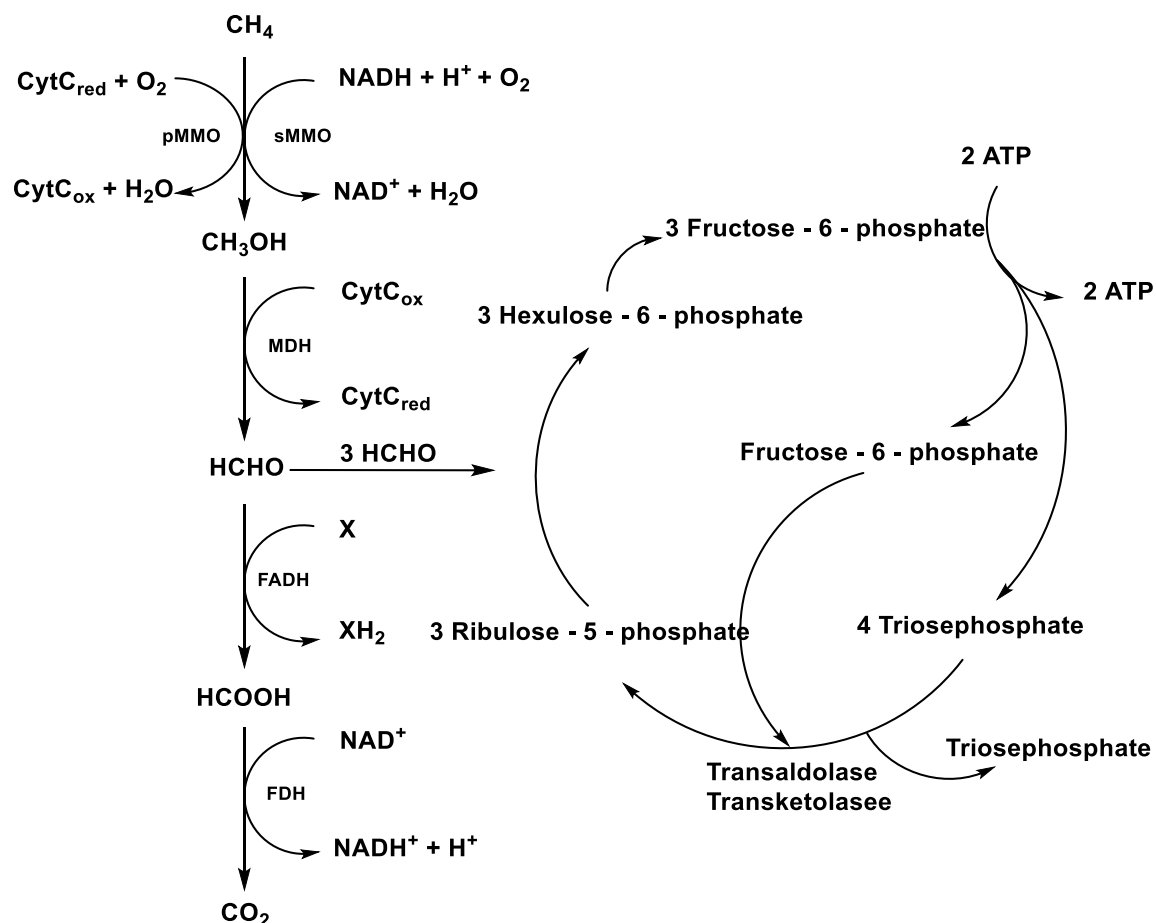
Figure 93 – Methane oxidation and formaldehyde assimilation pathway of type I methanotrophs. pMMO - Particulate methane monooxygenase, sMMO – Soluble methane monooxygenase, MDH – Methanol dehydrogenase, FADH – Formaldehyde dehydrogenase and FDH – Formate dehydrogenase. Adapted from Hanson and Hanson.[43]

Methane is taken up by the cells and enzymatically oxidised by pMMO to methanol, which is then dehydrogenated by methanol dehydrogenase to give formaldehyde. Formaldehyde can then either by converted to formate and then carbon dioxide, by formaldehyde dehydrogenase and formate dehydrogenase respectively, or utilised by the ribulose monophate (RuMP) pathway where it can be converted to biomass.[44] In the case of Daws_1 and Daws_2, both bacteria have a methanol dehydrogenase gene and so may be expected to grow using methanol as a carbon source

These isolated bacteria all contain genes that would allow the metabolism of the disaccharide maltose as well as the monosaccharide mannose. Furthermore, they contain genes allowing glycolysis and gluconeogenesis meaning they can convert glucose to glycogen and *vice versa*. Interestingly, they do not contain genes that encode for membrane bound saccharide transporter proteins meaning that they cannot uptake saccharides from the environment. This suggests that these bacteria are in fact obligate methanotrophs that utilise methane as their sole carbon source.

5.8.1.2 Solid Media Isolates

Strains *R.allander*, *M.allander* and *P.allander* were all isolated on solid media supplemented with organics such as pyruvate and glucose indicating that they are likely heterotrophic. All three of these bacteria contain genes encoding for membrane bound proteins that allow the utilisation of organics such as sugars and acids. This means they can take up these materials as opposed to the micromanipulator isolates which cannot transport organics across their cell membranes. Genes encoding for glycolysis, the overall equation of which can be seen in Equation 17, are found in all three isolates confirming that they can utilise glucose as a source of energy. Strain *M.allander* also contains genes encoding for the tricarboxylic acid (TCA) cycle, another high energy yielding metabolic pathway that also creates an abundance of biosynthetic intermediates. This pathway is not seen in the strains *R.allander* or *M.allander*.

$$C_6H_{12}O_6 + 2\ NAD^+ + 2\ ADP + 2\ P_i \longrightarrow 2\ C_3H_4O_3^- + 2\ NADH + 2\ H^+ + 2\ ATP$$
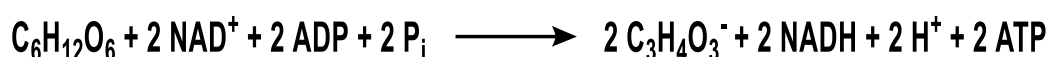
Equation 17 - Overall glycolysis equation. $C_6H_{12}O_6$ = glucose and $C_3H_4O_3^-$ = pyruvate.

All three isolated bacteria also have genes encoding for the metabolism of pyruvate. This would be expected as all three grew well on pyruvate supplemented media. Strain *M.allander* has the greatest number of genes encoding for monosaccharide metabolism. These include xylose, ribose, fructose, gluconate and galactonate. Strain *R.allander* likely has the ability to metabolise mannose, ribose and gluconate while the strain *P.allander* likely has the ability to metabolise mannose, galactonate and gluconate. Strain *M.allander* can likely also utilise amino sugars such as chitin and *N*-acetylglucosamine while the others cannot. Strain *R.allander* is the only isolate that can likely metabolise the di-saccharide maltose, while all three likely have the ability to utilise the alcohol glycerol. This data shows that the three bacteria isolated on solid media are likely more metabolically flexible than the micromanipulator isolates.

5.8.2 Environmental Detoxification

As discussed throughout Chapter 6 Section 5 these sample sites contain high concentrations of iron and manganese along with lower concentrations of other metals and metalloids. These elements themselves can be toxic in sufficient concentrations and can also accelerate the generation of radical oxygen species (ROS), via the Fenton reaction, along with other radical species.[45] Bacteria must therefore be equipped with methods of detoxification to deal with these environmental stresses.

5.8.2.1 Micromanipulator Isolates

Daws_1, Daws_2 and Al_1 all contain genes encoding for ROS detoxifying enzymes including rubrerythrin and superoxide dismutase (SOD) along with glutathione utilisation. They also contain genes encoding proteins for the uptake and detoxification of selenium as well as the metabolism of phosphates. The data in Chapter 6 Section 5.9 shows that selenium and phosphates are present in the water of these sample sites indicating that these pathways may be required to survive. These isolates also contain a variety of genes encoding for heavy metal detoxification. Transenvelope protein complexes that detoxify the periplasm by excretion of heavy metals including cobalt, zinc and cadmium are seen along with proteins for copper resistance.

5.8.2.2 Solid Media Isolates

Strains *R.allander*, *M.allander* and *P.allander* all have similar genes for dealing with ROS as the micromanipulator isolates. These include the utilisation of SOD and ruberythrin along with glutathione. They also have similar genes for heavy metal and metalloid remediation including similar pathways for selenium uptake, cobalt-zinc-cadmium resistance proteins and copper resistance proteins. Interestingly the strains *R.allander* and *M.Allander* also contain genes that encode for mercury detoxifying proteins such as mercuric reductase which reduces toxic Hg (II) to elemental Hg (0). This ability is not seen in the strain *P.allander* or any of the micromanipulator isolates.

5.8.3 EPS Production

5.8.3.1 Micromanipulator Isolates

Daws_1, Daws_2 and AI_1 all contain genes encoding for capsular and extracellular polysaccharide production, specifically dTDP-rhamnose synthesis. Daws_2 contains genes, not found in the other isolates, encoding for further capsular polysaccharide biosynthesis and export indicating that Daws_2 can secrete more complex EPS than the other isolates.

5.8.3.2 Solid Media Isolates

Strains *R.allander*, *M.allander* and *P.allander* all contain the same dTDP-rhamnose synthesis genes as the micromanipulator isolates along with a suite of genes for producing other rhamnose containing polysaccharides. These secreted EPS may contribute to the overall composition of the BIOX mats.

5.9 Conclusions and Future Direction

Attempts were made at isolating *L.ochracea* from BIOX collected from various sample sites. Methodologies included solid plate media, enrichment growths and micromanipulation to isolate filamentous material. *L.ochracea* was not isolated here however progress has been made towards its future isolation.

Solid agar containing media was successful at isolating a filamentous *Arcicella spp.* from the effluent of a disused coal mine. Substituting agar for gellan gum in solid media yielded the successful isolation of three different bacteria from BIOX samples, a *Rhodoferax spp.*, a *Polaromonas spp.*, and a *Mesorhizobium spp.*

Enrichment growths using 0.2 µm filtered sample site water as growth media was used to keep BIOX mats alive and potentially growing however no isolation of bacteria was possible. Gradient tubes were also employed to isolate FeOB, however only blooms of diatoms were produced.

A protocol was developed to isolate single filaments of filamentous bacteria from BIOX mats. Phase contrast microscopy was used to confirm the presence of a variety of filament morphologies within the BIOX samples and that filaments can contain chains of

cells. These filaments and cells ranged in size, however some contained similar dimensions to that of *L.ochracea*. CLSM was used to confirm that filaments could contain live cells and that live cells were dispersed at depths throughout the BIOX mats. A micromanipulator was then used to isolate a number of different filamentous bacteria, including a morphology that was thought could potentially be *L.ochracea* free from 2-Fh encrustation. All isolated bacteria had their total genomic DNA amplified by PCR and were sequenced using high-throughput Illumina sequencing. Results showed three of the isolated bacteria to be methanotrophic *γ*-proteobacteria while the fourth was potentially of the *Caulobacter* genus and that these bacteria have not previously been isolated in the literature. Further optimisation of this protocol is warranted to isolate *L.ochracea* for sequencing and further study in artificial media.

Comparing the genomic data of all isolates showed that the three bacteria isolated on solid media are more metabolically flexible than the micromanipulator isolates. The solid media isolates can utilise a number of different organics as a source of energy whereas the micromanipulator isolates are obligate methanotrophs. All isolated bacteria contain genes encoding for the detoxification of ROS, heavy metals and metalloids and also have the ability to produce a variety of EPS which may contribute to the composition of the BIOX mats.

Importantly, the development of a method to store material on a microscope slide fitted with a silicon gasket has allowed for more extended observation of the samples via phase contrast microscopy. The identification of the main filamentous bacteria from the Allander site and Dawsholm sites as previously uncharacterised methanotrophic *γ*-proteobacteria of differing thicknesses removed them as candidates for *L.ochracea.*

Finally, by observing a slide of filtered BIOX material, stored at 4°C for 12 - 24 hours, it has been possible to observe thinner filamentous bacteria exiting from the associated BIOX and also directly from the BIOX filaments, as seen in Figure 94. These chains of cells are thinner, and in the main shorter, than the Daws_1, Daws_2 and Al_1 methanotrophs. At this point these bacteria are the main target for isolation and characterisation.
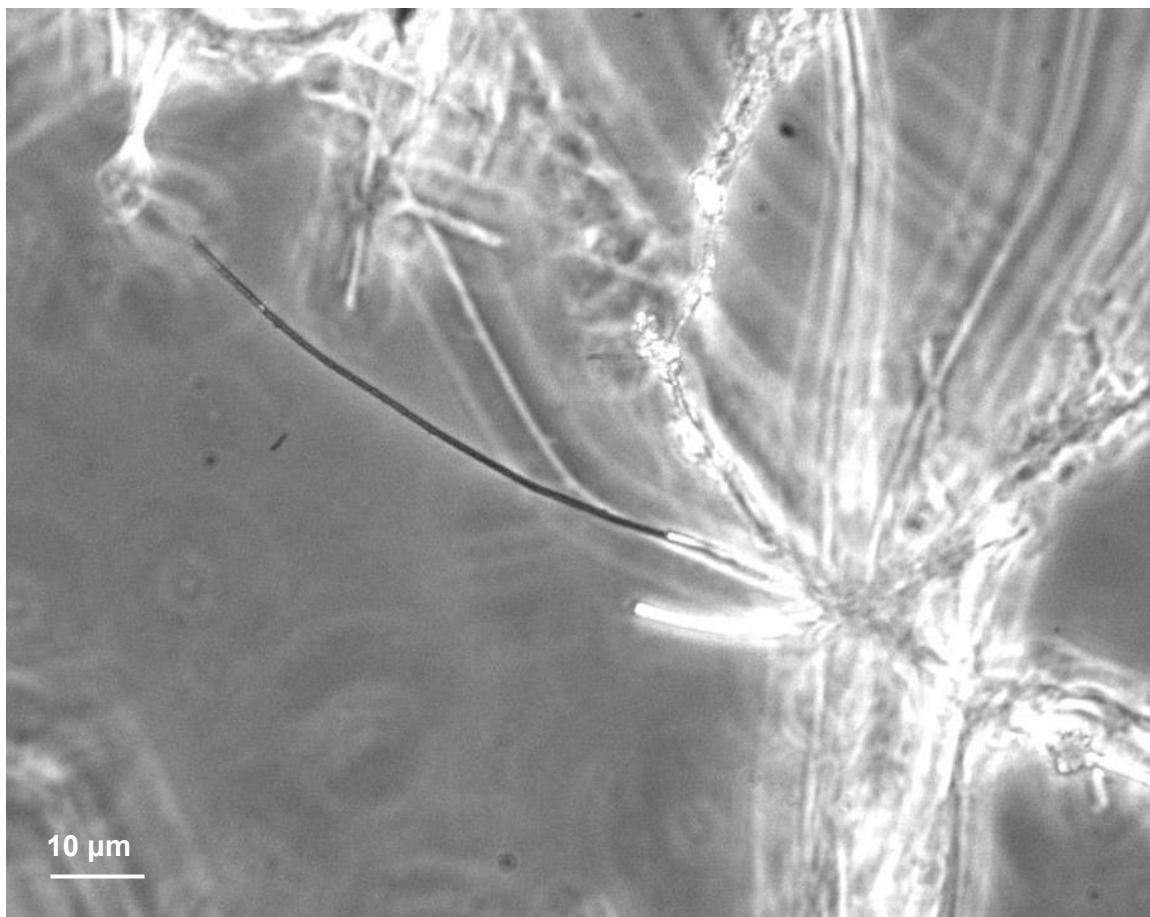
Figure 94 – Phase contrast micrograph of candidate *L.ochracea* bacteria.

L.ochracea was not isolated using any of the above methodologies, however several bacteria – including filamentous bacteria – were isolated using the above methodologies. This partially confirms the initial hypothesis that it would be possible to develop a methodology to isolate filamentous bacteria from BIOX samples. Further work is warranted to optimise the micromanipulation methodology in order to isolate a single filament of *L.ochracea* cells.

5.10 References

1. C. S. Chan, S. M. McAllister, A. H. Leavitt, B. T. Glazer, S. T. Krepski and D. Emerson, *Frontiers in Microbiology*, 2016, **7**, 796.
2. R. Buliauskaitė, P. Wilfert, P. Suresh Kumar, W. W. J. M. de Vet, G.-J. Witkamp, L. Korving and M. C. M. van Loosdrecht, *Environmental Technology*, 2020, **41**, 260-266.
3. L. Castro, M. L. Blázquez, F. González, J. Muñoz and A. Ballester, *Hydrometallurgy*, 2018, **179**, 44-51.
4. Ehrenberg, C. and G., *Annalen der Physik*, 1836, **114**, 213-227.
5. D. Emerson, E. J. Fleming and J. M. McBeth, *Annual Review of Microbiology*, 2010, **64**, 561-583.
6. D. Emerson and W. C. Ghiorse, *Applied and Environmental Microbiology*, 1992, **58**, 4001-4010.
7. K. Liu, H. Chua, W.-h. Lo, H. Lawford and P. H.-F. Yu, *Applied Biochemistry and Biotechnology*, 2002, **98**, 1061-1073.
8. E. G. Mulder and W. L. van Veen, *Royal Dutch Society for Microbiology Journal*, 1963, **29**, 121-153.
9. W. L. van Veen, E. G. Mulder and M. H. Deinema, *Microbiological reviews*, 1978, **42**, 329-356.
10. E. G. Mulder and W. L. van Veen, *Antonie van Leeuwenhoek*, 1963, **29**, 121-153.
11. E. J. Fleming, T. Woyke, R. A. Donatello, M. M. M. Kuypers, A. Sczyrba, S. Littmann, D. Emerson and L. Drake Harold, *Applied and Environmental Microbiology*, 2018, **84**, e02239-02217.
12. E. J. Fleming, A. E. Langdon, M. Martinez-Garcia, R. Stepanauskas, N. J. Poulton, E. D. P. Masland and D. Emerson, *PLOS One*, 2011, **6**, e17769-e17769.
13. M. Takeda, K. Kondo, M. Yamada, J.-i. Koizumi, T. Mashima, A. Matsugami and M. Katahira, *International Journal of Biological Macromolecules*, 2010, **46**, 206-211.
14. P. L. Corstjens, J. P. de Vrind, P. Westbroek and E. W. de Vrind-de Jong, *Applied and Environmental Microbiology*, 1992, **58**, 450-454.
15. S. Spring, P. Kampfer, W. Ludwig and K.-H. Schleifer, *Systematic and Applied Microbiology*, 1996, **19**, 634-643.
16. M. Sawayama, T. Suzuki, H. Hashimoto, T. Kasai, M. Furutani, N. Miyata, H. Kunoh and J. Takada, *Current Microbiology*, 2011, **63**, 173-180.
17. B. Schmidt, L. A. Sánchez, T. Fretschner, G. Kreps, M. A. Ferrero, F. Siñeriz and U. Szewzyk, *FEMS Microbiology Ecology*, 2014, **90**, 454-466.
18. D. Nikitin, *International Journal of Systematic and Evolutionary Microbiology*, 2004, **54**, 681-684.
19. S.-Y. Sheu, C.-S. Yang, M.-H. Chen, A. Ab, Y. Chiu-Chung and W.-M. Chen, *International Journal of Systematic and Evolutionary Microbiology*, 2010, **60**, 2979-2983.
20. W.-R. Abraham, A. J. Macedo, H. Lünsdorf and D. I. Nikitin, *Bergey's Manual of Systematics of Archaea and Bacteria*, 2015, 1-5.
21. J. Staley, *Annual Review of Microbiology*, 1985, **39**, 321-346.
22. A. Bodor, N. Bounedjoum, G. E. Vincze, Á. Erdeiné Kis, K. Laczi, G. Bende, Á. Szilágyi, T. Kovács, K. Perei and G. Rákhely, *Reviews in Environmental Science and Bio/Technology*, 2020, **19**, 1-22.
23. E. J. Stewart, *Journal of Bacteriology*, 2012, **194**, 4151-4160.
24. T. Tanaka, K. Kawasaki, S. Daimon, W. Kitagawa, K. Yamamoto, H. Tamaki, M. Tanaka, H. Nakatsu Cindy, Y. Kamagata and H. Nojiri, *Applied and Environmental Microbiology*, 2014, **80**, 7659-7666.
25. I. Imazaki and Y. Kobori, *Canadian Journal of Microbiology*, 2010, **56**, 333-341.

26. H. Tamaki, Y. Sekiguchi, S. Hanada, K. Nakamura, N. Nomura, M. Matsumura and Y. Kamagata, *Applied and Environmental Microbiology*, 2005, **71**, 2162-2169.
27. J. F. Imhoff, in *The Prokaryotes: Volume 5: Proteobacteria: Alpha and Beta Subclasses*, eds. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer and E. Stackebrandt, Springer New York, New York, NY, 2006, pp. 593-601.
28. J. L. Darcy, R. C. Lynch, A. J. King, M. S. Robeson and S. K. Schmidt, *PLOS ONE*, 2011, **6**, e23742.
29. T. Kaneko, Y. Nakamura, S. Sato, E. Asamizu, T. Kato, S. Sasamoto, A. Watanabe, K. Idesawa, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, Y. Mochizuki, S. Nakayama, N. Nakazaki, S. Shimpo, M. Sugimoto, C. Takeuchi, M. Yamada and S. Tabata, *DNA Research*, 2000, **7**, 331-338.
30. H. Jin, B. Li, X. Peng and L. Chen, *Annals of Microbiology*, 2014, **64**, 689-697.
31. Y. Jung, J., H. Kim, J. and M. Hur, Personal Communication.
32. D. Emerson and C. Moyer, *Applied and Environmental Microbiology*, 1997, **63**, 4784-4792.
33. J. Larkin and D. Shinabarger, *International Journal of Systematic Bacteriology*, 1983, **33**, 841-846.
34. C. Pfeffer, S. Larsen, J. Song, M. Dong, F. Besenbacher, R. L. Meyer, K. U. Kjeldsen, L. Schreiber, Y. A. Gorby, M. Y. El-Naggar, K. M. Leung, A. Schramm, N. Risgaard-Petersen and L. P. Nielsen, *Nature*, 2012, **491**, 218-221.
35. I. Wartiainen, A. G. Hestnes, I. McDonald and M. Svenning, *International Journal of Systematic and Evolutionary Microbiology*, 2006, **56**, 109-113.
36. R. Boden, M. Cunliffe, J. Scanlan, H. Moussard, K. Kits, M. Klotz, M. Jetten, S. Vuilleumier, J. Han, L. Peters, N. Mikhailova, H. Teshima, R. Tapia, N. Kyrpides, N. Ivanova, I. Pagani, J.-F. Cheng, L. Goodwin, C. Han and J. Murrell, *Journal of Bacteriology*, 2011, **193**, 7001-7002.
37. V. A. Romanovskaya and N. L. P. Belkova, V.V.Gladka,G.V.Tashirev,A.B., Personal Communication.
38. S. Kato, C. Chan, T. Itoh and M. Ohkuma, *Applied and Environmental Microbiology*, 2013, **79**, 5283-5290.
39. E. Roden, J. McBeth, M. Blothe, E. Percak-Dennett, E. Fleming, R. Holyoke, G. Luther and D. Emerson, *Frontiers in Microbiology*, 2012, **3**, 172.
40. W. Yan and Y. Zhou, *Science of The Total Environment*, 2019, **688**, 462-469.
41. L. Zhuang, J. Xu, J. Tang and S. Zhou, *Journal of Geophysical Research: Biogeosciences*, 2015, **120**, 876-886.
42. H.-H. T. Nguyen, S. J. Elliott, J. H.-K. Yip and S. I. Chan, *Journal of Biological Chemistry*, 1998, **273**, 7957-7966.
43. R. S. Hanson and T. E. Hanson, *Microbiological Reviews*, 1996, **60**, 439-471.
44. H. He, C. Edlich-Muth, S. N. Lindner and A. Bar-Even, *ACS Synthetic Biology*, 2018, **7**, 1601-1611.
45. C. C. Winterbourn, *Toxicology Letters*, 1995, **82-83**, 969-974.

6 Sample Site Characterisation

6.1 Hypothesis

Prior to commencing the work carried out in this chapter it was hypothesised that firstly, the water found at each sample site would contain higher concentrations of dissolved ferrous iron than is typically found in Scottish waters. It was secondly hypothesised that the waters at each site would have similar physicochemical properties, for example: pH, temperature, conductivity, and redox potential. It was finally hypothesised that the sample site waters would contain dissolved organic carbon and that the chemical composition of the waters would likely be influenced by anthropogenic sources.

6.2 Introduction

What exactly constitutes an ideal habitat for *L.ochracea* is undoubtedly related to its well-known inability to grow reproducibly over an extended length of time *in vitro*. The inability to grow this bacterium in culture means that suitable environmental sites need to be identified and characterised. This is necessary in order to obtain high quality, reproducible BIOX material and better to understand the ecosystem in which this and other FeOB thrive.

Groundwater is located beneath the Earth's surface and fills spaces between soils, sediments and fractures in rocks. The soil, sediment and rocks through which groundwater flows are called aquifers and are typically suitably porous to allow this movement. This area is known as the saturated zone, beneath which is an impenetrable layer of rock, and above which is the unsaturated zone, a layer of soil, sediment and rock containing both water and air. The boundary between the saturated and unsaturated zones is known as the water table. Due to the porous nature of aquifers, groundwater is typically purified as it passes through them. Aquifers are found globally, making groundwater an essential hydrogeological phenomenon as it provides a water source for many springs, seeps and wetlands, all of which can be colonised by iron-oxidising bacteria (FeOB).

Groundwater typically begins as rainwater, falling on land and passing through the soil to the aquifer below, and represents the largest reservoir of freshwater worldwide.[1] It then

slowly flows underground until it reaches a discharge zone at which point it enters water bodies and becomes surface water. During this cycle, various organic and inorganic species are mobilised. There is no overall typical composition of groundwater, rather the composition depends on local environmental and geological factors.[2] A major component that is found in most groundwater however is iron. Iron is the second most abundant metal on Earth, the most abundant element by mass on Earth, and the fourth most abundant element in Earth's crust where it makes up 6.3 % of the elemental composition. The other abundant elements in Earth's crust are oxygen (47 %), silicon (26 %), aluminium (8.1 %), calcium (5.0 %), magnesium (2.9 %), sodium (2.3 %) and potassium (1.5 %).[3] This ubiquity makes iron prevalent in aquatic and sedimentary environments. Iron is also redox active and commonly occurs in either the ferrous, Fe (II), or ferric, Fe (III), oxidation states. Under reducing conditions the ferrous species exists in solution while under oxidising conditions it typically precipitates as ferric oxyhydroxides. This allows geochemical cycling to occur, both biogenically and abiogenically, resulting in iron affecting the mobility, sorption, precipitation, bioavailability and degradation of environmental compounds.[4, 5] Scottish drinking water has a legal limit of 0.2 mg/l iron and groundwater can regularly exceed 30 mg/l.[6] At these high concentrations iron can be termed a cosmetic pollutant as it turns waters orange and turbid, resulting in the clogging of water systems.[7] As iron rich groundwater becomes surface water the concentrations dilute and become appropriate for organisms such as FeOB, including *Leptothrix spp.* and *Gallionella spp.* to utilise. Iron rich groundwaters therefore likely play a key role in creating habitats for FeOB. There are currently few studies to the authors knowledge which have analysed the elemental composition of *Leptothrix spp.* sample sites and none which have drawn comparisons with surrounding geology to comment on potential water sources. As groundwater can supply many of the environments in which FeOB thrive it is therefore essential to understand the composition of the sample site water and the geology of the local aquifers through which it has moved. The following work aims to do this for several sample sites in the Greater Glasgow area, where BIOX is routinely found.

6.3 Chapter Aims

In order to gain a more thorough understanding of the factors contributing to *L.ochracea* growth it is essential to understand the environments in which they thrive. To the author's knowledge there have been no studies investigating environmental factors such as the organic species present in sample sites or the local geological formations and anthropogenic influences that likely contribute to chemical composition of environments in

which *L.ochracea* thrive. This chapter aims to address these topics in relation to sample sites frequented for this study along with other investigating factors including the physicochemical profile and inorganic species present at these sites.

6.4 Results and Discussion

The initial sample site for material was located near the Allander Water at NS 5471 7573 and has been described Alharthi *et al.*[8] The site was not always optimal for collecting samples, especially during the summer months, and so other sites were looked for which might represent more convenient and less complex sampling sites for obtaining BIOX and bacterial samples. The oily surface film and the characteristic look of BIOX meant that potential sites were easily identifiable.

As the study progressed the author identified, or was made aware of from friends and colleagues, many other potential sample sites some of which are described in detail in the next sections. It was important to continually assess new sample sites in the hope that an ideal sample site was located as the quality of BIOX samples differed greatly from site to site.

An ideal sample site should firstly be easily accessible year-round and contain minimal particulate contaminants. These contaminants include vegetation and high concentrations of dirt and sediment. Sample sites should also be sufficiently deep so that sediment on the bed of the site is not disturbed during sampling. There must also be a continual source of ferrous rich water that replenishes the site and could be sampled for growth media applications. This source water must flow slowly so as to not disrupt the structure of the BIOX mats. Finally, there should be a low concentration of background bacteria, algae, diatoms, worms, larva etc associated with the BIOX samples.

Figure 95 shows a variety of potential sample sites which due to proximity or other factors were not used within this study. The sample sites that proved most reliable were Allander_1 – Allander_4, Dawsholm Park, Dougalston Golf Club and Kelvin Walkway and as such will be discussed further throughout this chapter.

Figure 95 - Collage of photographs of potential sample sites brought to the author's attention that were not included in this study. A and B – Lochwinnoch (NS 36897 58994), C and D – Allander Houses (NS 55732 74057), E – Allander Rangers (NS 56256 73645) and F and G – Schwalm-Eder-Kreis, Germany (50 ° 58.8667' N, 9 ° 22.7833' E).

6.5 Sample Site Descriptions

6.5.1 Allander

The Allander Water is a river that flows through Milngavie, a town to the northwest of Glasgow, and feeds into the River Kelvin. During this study, multiple areas surrounding the Allander Water have been found to contain blooms of BIOX. Three of these sites were sampled from regularly and are denoted Allander_1, Allander_2 and Allander_3. These three sites occupy a short stretch of ca. 20 m to the left and righthand side of the Staney Brigg. A fourth site, Allander_4, can be found at the opposite side of the river from Allander_1 and Allander_2, This site was sampled from less frequently as it was only available when the Allander Water's level was low, so no phylogenetic studies were carried out, however some characterisation of the site was performed. An aerial photograph of this area can be seen in Figure 96.

Figure 96 – Aerial photograph showing the area of Allander where the four sites are found. Ordanance Survery grid references are quoted in brackets. Blue – Allander_1 and Allander_2 (NS 5471 7573), Red – Allander_3 (NS 5468 7573) and Green – Allander_4 (NS 5470 7572)

6.5.1.1 Allander_1 and Allander_2

Allander_1 and Allander_2 are two separate blooms of BIOX found at opposite ends of an irregular shaped pond. This pond is ca. 6.5 × 3.5 m with a depth of 9 cm to sediment and a sediment depth of 15 cm. Allander_1 appears to contain stagnant water and tends to

produce BIOX blooms more regularly than Allander_2. Allander_2 occasionally shows bubbling in the water which is likely the main source of the groundwater feeding the pond The BIOX found in Allander_2 is generally less orange, instead appearing browner and dirt like and often has cream coloured deposits close to where the water bubbles up. Furthermore, both Allander_1 and Allander_2 are regularly covered with oil slick like surface film which is frequently associated with *L.ochracea* BIOX mats. Photographs of this area can be seen in Figure 97.

Figure 97 - Photographs of A – C – Allander_1 and D – F – Allander_2

These images highlight the transient nature of the sample sites. Images A and D, taken 6/1/20, show the typical BIOX found when the sites are blooming while B and E, taken 27/8/19, show the sites lacking BIOX. This is commonly seen after periods of rain, when the mats appear to be washed away, and also after periods of warm weather when the water levels are low. Images C and F show Allander_1 and Allander_2 coated in surface film.

6.5.1.2 Allander_3

Allander_3 is a small nook that appears to have a continual source of groundwater and is regularly the best of the three sample sites as a source of actively growing BIOX containing live cells. This sample site is ca. 45 cm in diameter at the mouth and has a depth of ca. 7 cm. A stream (6 m × 1 - 2 m) extends from the mouth and joins the Allander water. This stream is ca. 3 cm deep at its deepest, has a flow rate of ca. 5 m/min and also contains BIOX however the stream associated BIOX becomes more granular and silt like as it approaches the river while the BIOX within the mouth can be thick and gelatinous indicating a greater concentration of EPS. As this site is slightly more sheltered than the others it is affected less by rainy conditions, however high rain can see the Allander Water come and wash up to within 2m of the mouth bringing various debris with it. In summer there can be reduced flow and the width of this site is reduced during these warmer periods. A sulfurous odour at the mouth of the site is more noticeable at this site than either Allander_1 or Allander_2. Photographs of this site, taken 25/5/21, can be seen in Figure 98.

Figure 98 – A – Inside the mouth of Allander_3 showing thick gelatinous BIOX, B – Mouth of Allander_3 extending outwards, C – Allander_3 extending downstream, D – E – Downstream granular BIOX and F – Allander_3 extending into the Allander Water.

6.5.1.3 Allander_4

Allander_4 is located at the opposite bank of the Allander Water from Allander_1 and Allander_2. This site is the most transient of all the Allander sample sites meaning that it regularly contains no BIOX blooms. When BIOX appears, the site measures 1.5 × 0.8 m with a depth of ca. 7 cm making it the smallest of the four sites. This site is much closer to the river making it highly dependent on the water level and rainfall. Photographs of this site can be seen in Figure 99.
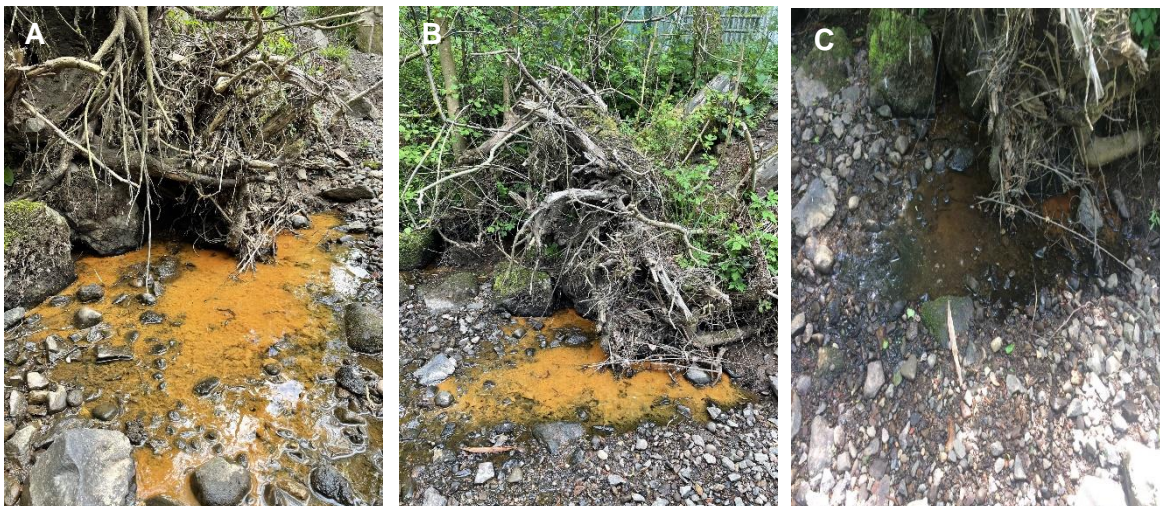
Figure 99 - Photographs of A and B – BIOX bloom at Allander_4 and C – Allander_4 showing only a small puddle and with no BIOX.

Images A and B were taken 25/5/21 while image C was taken 17/6/21. This confirms the ephemeral nature of Allander_4 as over a period of weeks this sample site has transitioned from a BIOX deposit to a small puddle.

6.5.2 Dougalston Golf Club

The Douglaston Golf Club is situated ca. 1 km from the Allander Water sites and was first brought to the author's attention in January 2020. Near the entrance to the golf course there is a ditch connected to a small stream. This site measures ca. 1.2 × 16 m with a depth of 10 cm. This stream appears to be stagnant however when the BIOX is gently disturbed the water flows at ca. 1 m/min. Photographs of this sample site, taken 25/5/21, can be seen in Figure 100 and a map highlighting its proximity to the Allander Water sample sites can be seen in Figure 101.
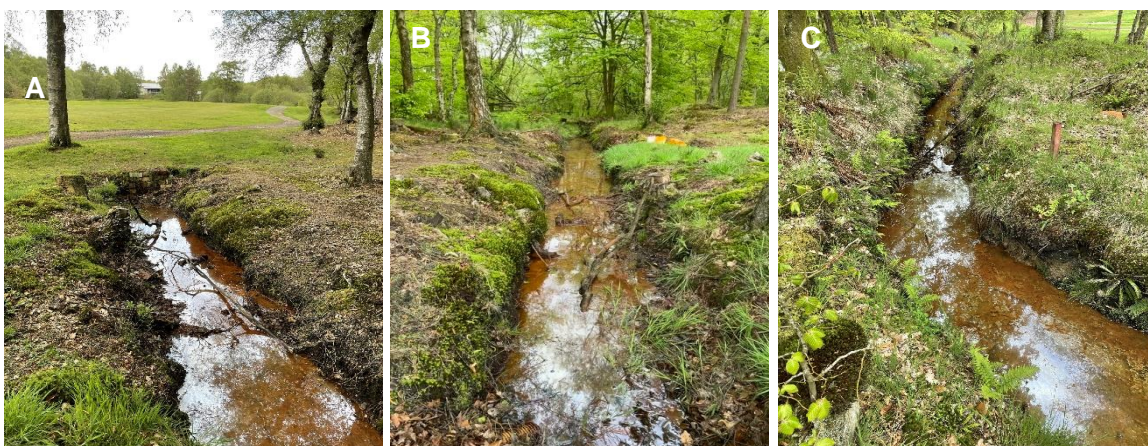
Figure 100 – A- C – Photographs of Dougalston Golf Course sample site.
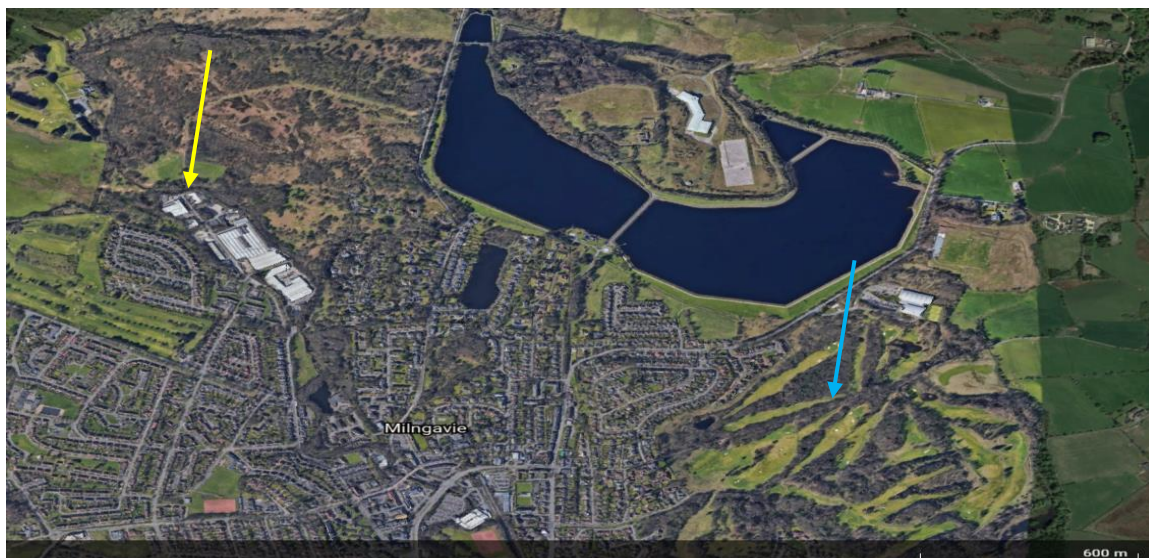
Figure 101 - Map showing the proximity of the Dougalston Golf Course (blue) sample site to the Allander Water sample sites (yellow).

6.5.3 Dawsholm Park

Dawsholm Park, located in the northwest of Glasgow, lies between the University of Glasgow and Milngavie where the previously discussed sites are found. It is ca. 5.3 km from Dougalston Golf Club and ca. 6.3 km the Allander Water. Approximately 60 m from the Dawsholm Road entrance to the park is a ditch containing a small pool of water with a short stream extending from it. This pool measures 1 m × 1.5 m with a depth of 6 cm and the stream measures 0.4 m × 3 m, this made the site ideal for collecting samples. The stream is extremely shallow meaning an accurate depth could not be measured. The author was first made aware of this sample site in September 2020. Samples were collected from here regularly until April 2021 after which the quality of the material began to decrease. As of May 2021 there have been no BIOX blooms noted at this location. Photographs containing BIOX blooms were taken 19/12/20 and photographs without BIOX 25/5/21. These photographs highlight the transient nature of this site can be seen in Figure 102.

Figure 102 – A – C – BIOX blooms in the pool and the stream, D – E – The pool showing no BIOX blooms and the stream appearing almost completely dry and F – The author collecting samples.

6.5.4 Kelvin Walkway

Kelvin Walkway is a 10 mile stretch of path that follows the River Kelvin from Kelvinhaugh to Milngavie. Approximately 50 m between the Botanic Gardens Footbridge and the Queen Margaret Drive bridge is a small 2 cm wide ditch at the foot of a banking. BIOX blooms in this ditch area were first reported to the author in March 2019 and were present periodically for ca. 8 months. This site, although small, was convenient as it was within a 15-minute walking distance from the School of Chemistry. While speaking with members of the public while collecting samples it was made apparent that BIOX blooms have appeared in this ditch area for a number of years and have frequently been reported to Glasgow City Council (GCC) and Scottish Water as a nuisance and potential pollution. GCC and Scottish Water have been carrying out work in the area since the last time BIOX blooms were seen indicating that they may have resolved whatever issue was causing the blooms to appear. The BIOX collected from this site regularly contained long *L.ochracea*

filaments and chains of cells were often noted extending from the mats. These samples were also consistently much cleaner than BIOX collected from other sites as they were free from large contaminants such as vegetation and particulates. Photographs of this sample site, taken 20/3/19, can be seen in Figure 103.

Figure 103 - Photographs of Kelvin Walkway sample site. A - BIOX mat growing in the ditch, B - BIOX growing over a metal drain covering and C - BIOX and surface film deposits on wooden panelling.

6.5.5 Kelvinhall Subway Station

During this study it was made aware to the author that orange material was regularly noted on the tracks of Kelvinhall subway station, which is located on Dumbarton Road near the foot of Byres Road and is a minimum of 10 m beneath the ground. A photograph from this location can be seen in Figure 104.

Figure 104 - Photograph of BIOX blooms on the Kelvinhall Subway track.

This photograph shows that there is a bloom of gelatinous orange material coating the base of the subway track, which comes from a gap within the subway tunnel wall. These blooms are transient in nature, appearing and disappearing at irregular intervals throughout the year. It has been made aware to the author that orange blooms are also occasionally seen on the tracks of Kelvinbridge subway station, however these blooms are less frequent and have not been photographed. This material would be interesting to sample as it likely contains fewer contaminants, such as vegetation and particulates, than the other described sample sites. Strathclyde Partnership for Transport (SPT) was asked if samples could be collected by the author or donated to the School of Chemistry, unfortunately this request was declined. It is therefore not possible to proceed with this site.

6.5.6 Disused Allanton Coal Mine

A final site the was sampled for this study was the effluent of a disused coal mine in Allanton, a village in North Lanarkshire that has a rich history of coal mining. This sample site was only sampled from once as part of a field study conducted with the School of Geographical and Earth Sciences and as such there is no further characterisation of it. A filamentous *Arcicella spp.* bacterium was however isolated from BIOX collected from the initial mine effluent and is further discussed in Chapter 5 Section 4.3. Figure 105 contains a collage of photographs from this sample site. These photographs show a small stream of orange water discharging from the mine and creating a large pool. Before the water reaches this pool it first flows through three man made pools containing high concentrations of calcium carbonate to neutralise any acidic species that may be discharging from the mine. Surprisingly, the measured pH of the initial effluent was ca. 7, which is unusual for coal mine effluents as they are frequently acidic. This suggests that there may be large quantities of calcium carbonate within the mine itself which neutralises any sulfuric acid that forms. Further study of this site is warranted as it is the only sample site which is visibly sourced by the effluent of a disused coal mine.

Figure 105 - Photographs from the Allanton sample site showing effluent leaving the mine and creating a small stream which feeds into a large pool.

6.6 Water Characterisation

6.6.1 Anthropogenic Influences

The historical context of the various sample sites is important as they may be affected by Glasgow's industrial past. Access to online map collections available at the University of Glasgow and through Glasgow City Council permitted the investigation of the site usage and geology over the past 150 years. The Dawsholm park area around the BIOX site is currently a popular space for walks and lies ca. 400 meters from a civic dump. Historic maps accessed through Glasgow City Council GlasgowGis maps seen in Figure 106 and Figure 107 from the 1860's show a series of three reservoirs close to the BIOX site and a paper works close to the river Kelvin which would have used this water. Maps from the 1890's show the addition of the Dawsholm Colliery some 200 meters from the site. The coal mining was gone by 1910 and the land contaminated with oil-shale waste was grassed over and later used for a time as football pitches. In the 1930's the papermill was moved south of the site of interest while the three reservoirs of water remained. In the 1970's on the closing of the papermill the reservoirs now referred to as ponds with the closest pond

to our site drained as used as a refuse tip, Figure 107. The Dawsholm park site is clearly a brown field site that has seen significant industrial activity over the last 150 years.
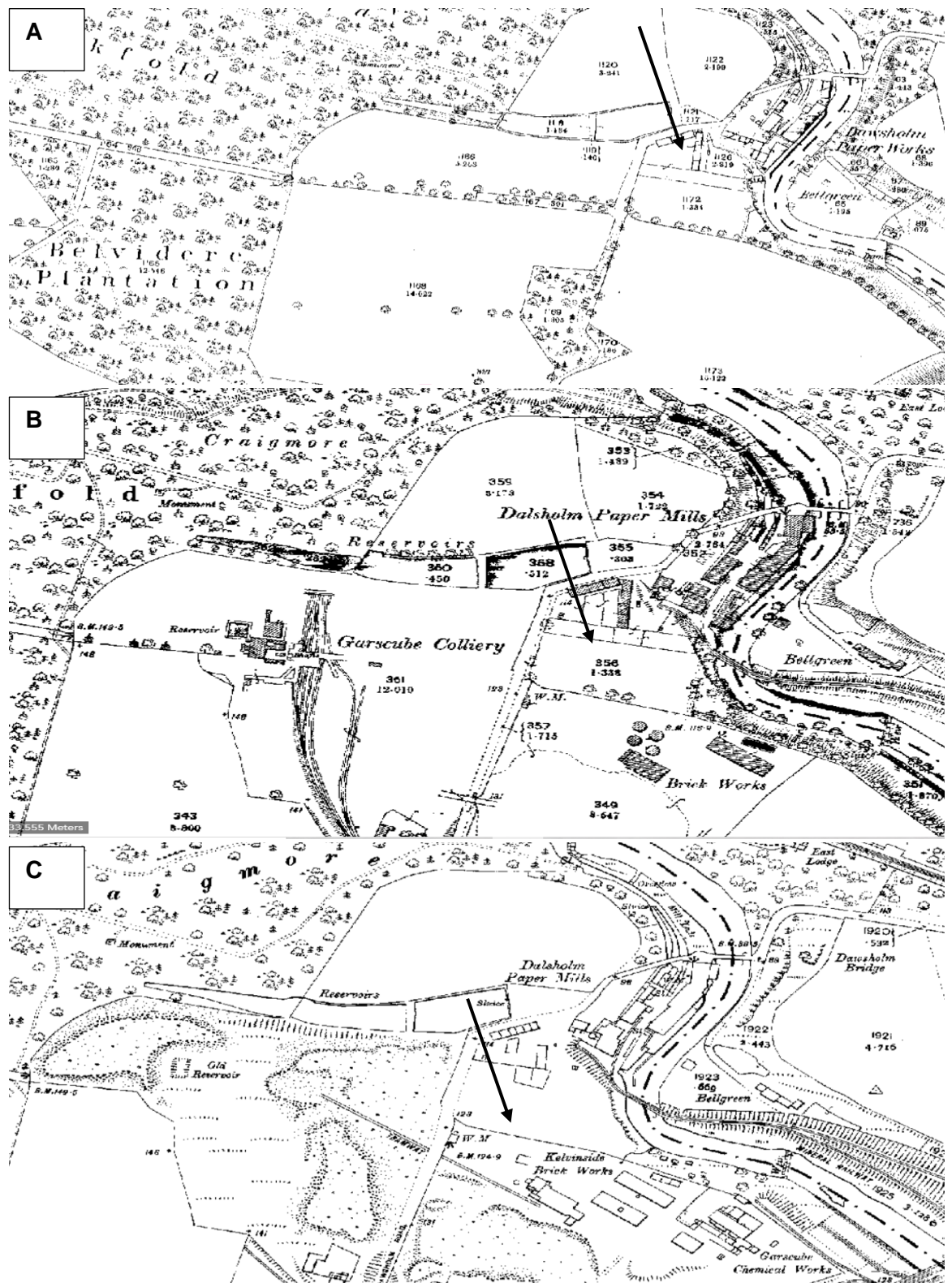
Figure 106 - Maps of Dawsholm Park from A – 1860's, B - 1890's and C - 1910's. An arrow is used to highlight the location of the sample site.
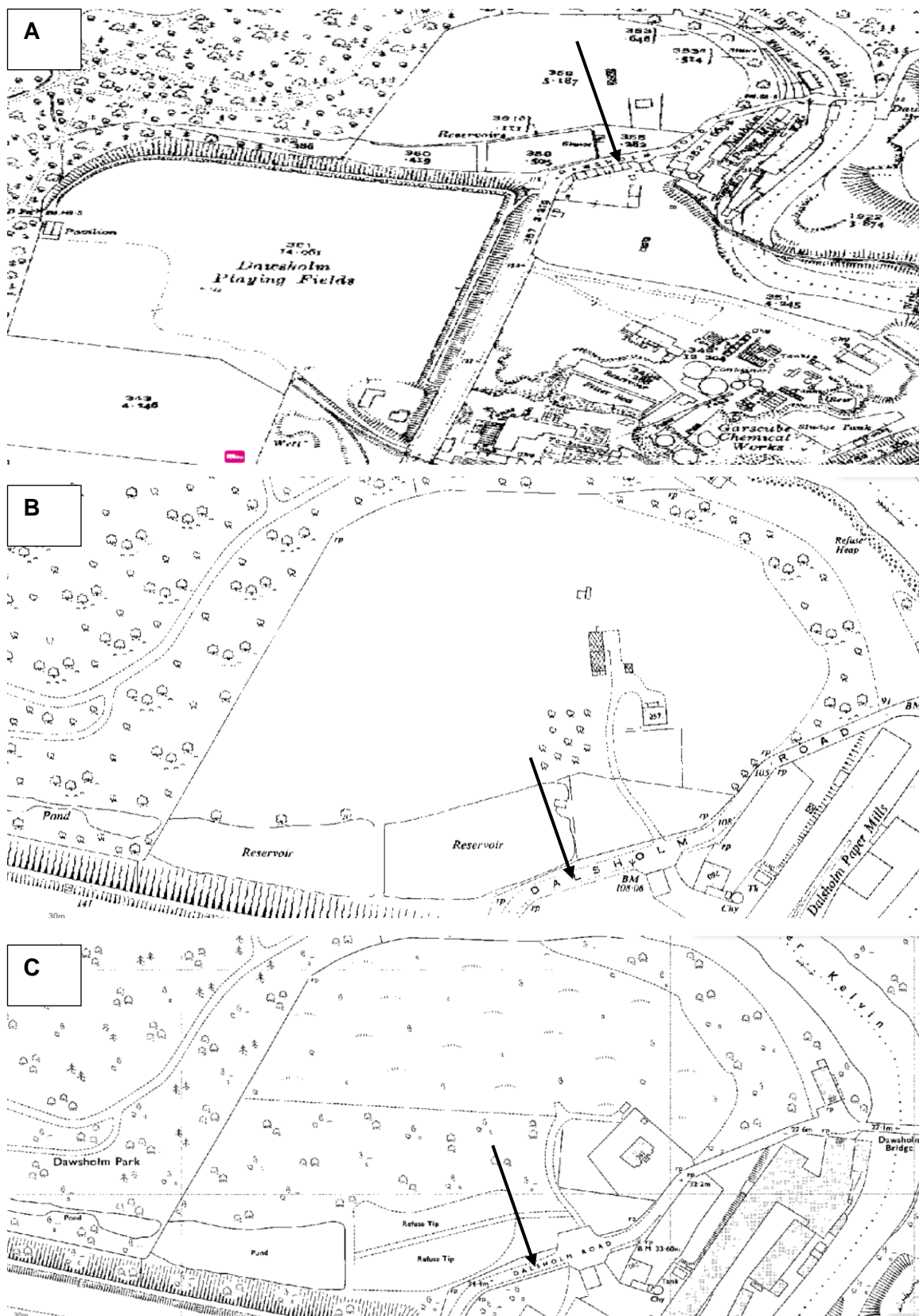
Figure 107 - Maps of Dawsholm Park from A - 1930's, B - 1950's and C - 1970's. B and C show magnified maps to improve the resolution of the image. An arrow is used to highlight the location of the sample site.

The maps of the Allander Water sites are less complete and show no industrial activity. However, whilst collecting samples and speaking with members of the public it was brought to the author's attention that historically the Allander Water may have had a dump located nearby. This is a popular explanation amongst locals for the BIOX deposits and the appearance of the Allander_1 site however it has proved difficult to confirm or discount this. Furthermore, it is worth noting that a dump site can be active for only a small number of years, as is the case with the Dawsholm site, yet have a significant longer-term effect to the ground water present at the site.

6.6.1.1 Leachate

It is reasonable to presume that the ground water from Dawsholm and Allander sites are contaminated by waste from landfill that appears in the sample sites as leachate. Leachate occurs when water, such as rain water, percolates through landfills and other anthropogenic dumps until it reaches surface water.[9] As water percolates through waste, it aids the decomposition processes carried out by bacteria and fungi. These processes generate by-products and create anoxic environments through oxygen usage. In actively decomposing waste the temperature increases and the pH acidifies causing many metal ions that are relatively insoluble at neutral pH dissolving in the developing leachate.

Landfill leachate is routinely a water-based solution of four major contaminants: dissolved organic matter (alcohols, acids, aldehydes, mono-, oligo- and polysaccharides etc.), inorganic species (sulfur species, chloride, iron, aluminium, zinc and ammonia), heavy metals (Pb, Ni, Cu, Hg), and xenobiotic organic compounds such as halogenated organics, (polychlorinated biphenyl (PCBs) and dioxins).[9] Nika *et al*. have shown that complex organics can also be prevalent in landfill leachate.[10] They sampled raw and treated landfill leachate and found 58 complex organic contaminants including 2-OH-benzothiazole in 84% of samples and perfluorooctanoic acid in 68% of samples respectively. Methane and hydrogen are also prevalent in leachate.

6.6.1.2 Coal Mining

Another anthropogenic influence that can affect water composition is coal mining. Glasgow has a rich industrial history that includes mining, ship building, railway engineering, chromium ore processing and steel and iron making.[11] Coal mining was especially prevalent in the west of Glasgow and out towards Milngavie meaning that many sample sites used for this study are close to historical coal mines. From the research of

historic maps it is clear that both Dawsholm and Kelvin walkway are within the extent of mines that are clearly identified. Flooded disused coalmines are the second most common freshwater pollutant in Scotland behind sewage, and are likely the most common freshwater pollutant in areas with a high number of disused mines.[12] Coal mines contain large quantities of exposed pyrite, an iron sulfide mineral, that is prone to oxidation once disused mines flood. This oxidation process generates sulfuric acid which leaches ferrous iron and other toxic metals, such as Pb, Zn, Cu, As and Ni, from surrounding minerals.[13, 14] An equation for this oxidation process can be seen in Equation 18.

$$2\ FeS_{2(s)} + 7\ O_{2(g)} + 2\ H_2O_{(l)} \longrightarrow 2\ Fe^{2+}_{(aq)} + 4\ SO_4^{2-}_{(aq)} + 4\ H^+_{(aq)}$$
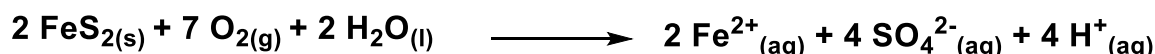
Equation 18 - Oxidation of pyrite resulting in the production of ferrous iron and sulfuric acid.

As oxygen is used during this oxidation process the resulting waters are generally microoxic. This, combined with the acidity of the generated sulfuric acid, stabilises the ferrous iron concentration. The sulfate species released can be converted, either biotically or abiotically, into a variety of sulfur species including sulfites, sulfides and elemental sulfur meaning that mine waters are rich in both dissolved ferrous iron and multiple sulfur species.

Another prevalent chemical species found within coal mine waters is methane. Coal mine related methane is typically generated through biogenic or thermogenic processes.[15] Biogenic coal mine methane is produced by the biological activity of microorganisms via the fermentation of acetate and the reduction of carbon dioxide while thermogenic methane is produced from the degradation of organic matter than was been buried and heated over extended periods of time.[16]

6.6.2 Physicochemical Measurements of Sample Sites

Physicochemical properties were recorded at various times throughout this study thereby permitting the development of assays and sampling methods. To give a consistent picture over several sites, Allander_1 - 4, Dawsholm and Dougalston were systematically measured on the 25/5/21. These measurements included pH, temperature, conductivity, oxidation-reduction potential (ORP), dissolved oxygen (DO), dissolved ferrous iron and total iron concentrations. Both Allander_3 and Dougalston had multiple measurements taken across the length of the sample sites. Selected sites also had their dissolved organic and dissolved inorganic carbon (DOC, DIC) concentrations measured. By taking and comparing these measurements it should be possible to gain a more thorough

understanding of the habitats in which BIOX mats, and *L.ochracea*, flourish. This in turn should allow the development of artificial media and growth conditions which could be employed to enrich and isolate recalcitrant bacteria such was *L.ochracea* and other FeOB The resulting measurements can be found in Table 26, Table 27 and Table 28.

	Allander_1	Allander_2	Allander_4	Dawsholm Pool
pH	6.6	6.7	6.5	6.8
ORP (mV)	-73	-83	-65	-40
DO (mg/l)	1.1	0.75	10.9	1.7
Fe (II) (µmol)	127	110	104	6
Fe_T (µmol)	216	173	399	74
DOC (mg/l)	9.1	6.8	-	8.9
DIC (mg/l)	404	409	-	424
Conductivity (µS/cm)	498	520	505	755
Temp (°C)	10.2	9.9	9.2	11.1

Table 26 - Physicochemical measurements of the Allander_1, Allander_2, Allander_4 and the Dawsholm pool sample sites.

	Allander _3_0 m	Allander_ 3_1 m	Allander_3 _2 m	Allander_ 3_3 m	Allander_ 3_4 m	Allander_ 3_5 m
pH	6.6	6.7	6.8	6.9	6.8	7.1
ORP (mV)	-80	-45	-35	-83	-47	-80
DO (mg/l)	-	-	-	-	-	-
Fe (II) (µmol)	117	93	52	24	16	18
Fe_T (µmol)	233	209	150	153	147	145
DOC (mg/l)	8.1	-	-	-	-	-
DIC	514	-	-	-	-	-
Conductivity (µS/cm)	598	-	-	-	-	-
Temp (°C)	10.5	10.5	10.2	9.7	9.8	10.1

Table 27 - Physicochemical measurements of the Allander_3 sample site beginning with the source of the sample site and then at subsequent 1 m intervals until the River Allander is reached.

	Dougalston Golf Course_4 m	Dougalston Golf Course_6 m	Dougalston Golf Course_8 m	Dougalston Golf Course_16 m
pH	6.5	7.4	7.7	7.6
ORP (mV)	-35	-50	-10	-15
DO (mg/l)	3.5	2.2	3.0	4.1
Fe (II) (μmol)	177	168	175	120
Fe$_T$ (μmol)	296	302	308	254
DOC (mg/l)	9.9	-	-	-
DIC	-	-	-	-
Conductivity (μS/cm)	276	-	-	-
Temp (°C)	11.1	10.8	10.8	10.5

Table 28 - Physicochemical measurements of the Dougalston Golf Course sample site taken at various intervals.

6.6.3 pH, Conductivity and Temperature

Each site has a circumneutral pH as is expected for these environments. The Dougalston site has an average pH value of 7.5 and is the only sample site to have an average pH greater than 7. The pH of the Allander_3 sample site gradually increases from 6.6 at the mouth of the sample site to 7.1 as it approaches the Allander Water. These pH values are typical for Scottish groundwater and are appropriate for neutrophilic FeOB such as *L.ochracea* to thrive.[17, 18] The British Geological Survey note that conductivity values for typical Scottish aquifers range from 200 – 800 μS/cm which agrees with the data recorded

here ranging from 276 – 755 µS/cm. They also note that Scottish groundwater temperatures have an average of 10 °C which is also in agreement with this data.[19]

6.6.4 Oxidation Reduction Potential (ORP)

The ORP of each site was measured as this value affects the stability of ferrous iron which is required by FeOB for growth. A negative ORP value indicates reducing conditions which aids dissolved iron in staying in the Fe (II) oxidation state. All measured sites have adequate negative ORP values. The most negative value being – 83 mV, at Allander_2 and Allander_3_3 m, while the least negative is – 10 mV at Douglaston_8 m. The ORP of Allander_3 fluctuates as the water flows from the mouth to the Allander Water, however both the mouth and the final sampling point have the same ORP value of – 80 mV meaning points throughout that the entire site are adequately reducing to contain ferrous iron.

6.6.5 Dissolved Oxygen (DO)

DO concentration is another contributing factor in altering the ferrous iron concentration and controlling growth of FeOB. Circumneutral river water with an oxygen concentration range of 8 – 2 mg/l is said to be oxic and promotes the rapid oxidation of ferrous iron and growth of aerobes.[20, 21] Microoxic conditions, < ca. 2mg/l oxygen, promotes the growth of microaerophilic FeOB such as *L.ochracea* and *G.ferruginea* as they are able to outcompete abiotic iron processes.[21] DO could not be measured at Allander_3 as the water was insufficiently deep for the probe to be submerged fully however it can be assumed that the DO concentration is sufficiently microoxic given the blooms of actively growing BIOX. The probe was fully submerged at sites where measurements were taken, the cable attached to the probe was marked to keep the depth consistent, and the probe was swirled slowly in a circular pattern to record the concentration. Care was taken to measure at depths the would be representative of where BIOX mats would be found, i.e., as close to the surface as was possible. Allander_4 had a DO concentration of 10.9 mg/l which is almost twice as much as the second highest value of 5.7 mg/l recorded at Dougalston_16 m. This high value is more common in fully saturated river water.[19] The DO concentration of the Allander Water was also measured and was found to be 12 mg/ml. Perhaps the closer proximity of Allander_4 to the Allander Water makes oxygenated water more readily available. Allander_1 and Allander_2 have appropriate concentrations of DO, 1.1 and 0.75 mg/l respectively, to support microaerophilic FeOB. The Dawsholm sample site also has a suitable DO concentration however no blooms of

BIOX at the time of sampling. The Douglaston site had its DO concentration measured at four points throughout the ditch. The concentrations varied from 2.2 mg/l – 4.1 mg/l respectively. These are slightly less than ideal for supporting the growth of microaerophilic FeOB however *L.ochracea* have been shown to be more metabolically flexible than other FeOB and are tolerant of more oxygen rich conditions.[22] This may explain the extensive growth of BIOX mats at this site and potentially at Allander_4. The Allander_4 site was found to contain no BIOX blooms in the weeks after sampling for this study indicating that sampling may have taken place during a transition period and that is why there was both a high DO concentration and BIOX present.

6.6.6 Fe (II) and Fe$_T$

The ferrous iron and total iron concentrations were measured for each site via a ferrozine assay using the method as detailed by Jeitner in Chapter 2 Section 3.4.[23] Calibration was carried out using standard solutions of ferrous ammonium sulfate $((NH_4)_2Fe(SO_4)_2.(H_2O)_6$ with concentrations of 1 mM, 500 µM, 250 µM, 100 µM, 50 µM, 25 µM, 10 µM, 1 µM and a blank $_dH_2O$.

Samples for assays were acidified to pH 3 upon return to the laboratory, to prevent abiotic iron oxidation, and assayed as soon as was possible. Ascorbic acid (1 M) was used to reduce ferric iron for the total iron measurements. Figure 108 contains bar charts highlighting the difference in dissolved iron concentrations between sample sites.
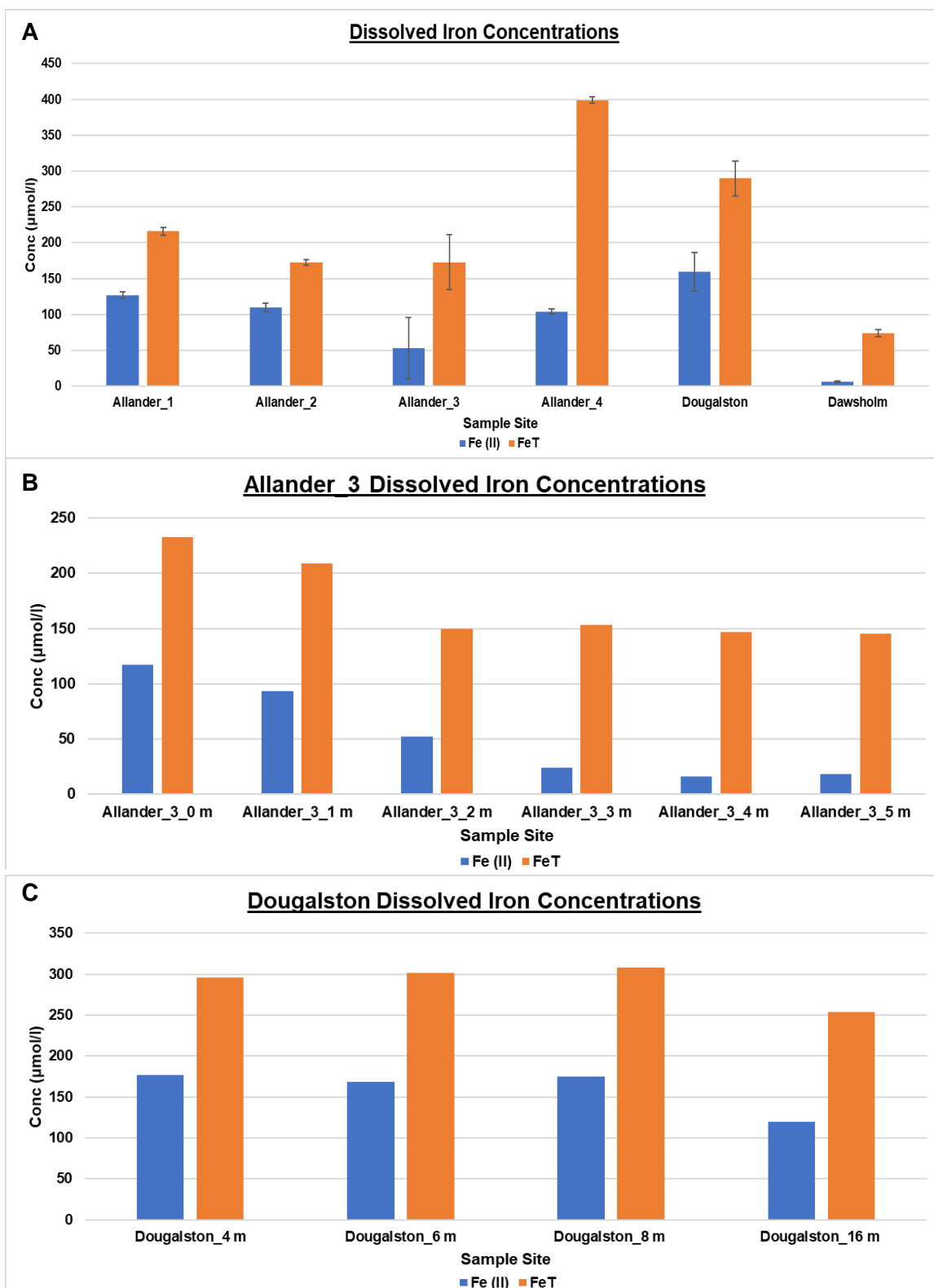
Figure 108 – Dissolved Fe (II) and Fe$_T$ concentrations of A – All sites with Allander_3 and Dougalston being average concentrations of their respective points, B – Allander_3 across all points and C – Dougalston across all points.

Chart A shows the dissolved ferrous iron and total iron concentrations of all the sample sites. The Allander_3 and Dougalston data are averages of their collective concentrations

shown in charts B and C. This data shows that of the Allander sites, at the time of sampling, Allander_1 had the highest concentration of ferrous iron present (127 µM) while the Allander_4 site had the highest total iron concentration present (399 µM). The Dougalston site had the highest ferrous iron concentration of all the sample sites (160 µM) and also had a higher total iron concentration (290 µM) that all sites apart from Allander_4 while the Dawsholm sample site had the lowest ferrous iron (6 µM) and total iron (74 µM) concentrations respectively. The low concentrations at the Dawsholm site were not unexpected as at the time of sampling there were no visible BIOX blooms present. When BIOX blooms were present at Dawsholm, such as on the 8/3/21, the ferrous iron concentration was 150 µM and the total iron concentration was 154 µM.

Chart B shows the ferrous and total iron concentrations for the Allander_3 site, from its source to the Allander Water, measured at six 1 m intervals. This data was averaged for chart A and shows that there is a steady decline in the ferrous iron concentration from 117 µM at the source of Allander_3 to 17 µM at the furthest point from the source. This is not unexpected as the groundwater supplying the source will be less oxic than the subsequent flowing water and should also be continually replenished with dissolved ferrous iron. As the stream flows towards the Allander Water the ferrous iron is utilised by the FeOB present, causing the concentration to decrease. The ORP data for this sample site is negative for the entire length of the stream, meaning that the conditions are appropriate for dissolved iron to remain in the ferrous state if it is not oxidised by biogenic processes.

Chart C shows the ferrous and total iron concentrations for four points across the Dougalston sample site. This data was averaged for chart A and shows that there does not appear to be any correlation with position and concentration as between 4, 6 and 8 m both the ferrous and total iron concentrations are similar. The ferrous iron concentration then drops from 175 µM at this section to ca. 120 µM at the 16 m point. This is a drop in concentration of ca. 32 % which is far less than Allander_3 which drops ca. 85 % across the sample points. This may be due to the slower flow rate seen at the Dougalston sample site or perhaps there is a continual supply of ferrous iron from the stream bed.

The ferrous and total iron concentrations recorded here are consistent with literature studies of BIOX mats containing FeOB.[24, 25] These results will be discussed further in Chapter 6 Section 5.9.

6.6.7 Dissolved Organic Carbon (DOC)

DOC is a complex mixture of various organic compounds found within the environment and represents an important source of bioavailable carbon that can be utilised by heterotrophic bacteria.[26] DOC has been shown to influence processes in aquatic environments, such as binding to extracellular enzymes to regulate degradation processes, and it has been hypothesised that *L.ochracea* may grow more favourably in environments where DOC is available.[24, 27] It was therefore important to measure the DOC content at regularly sampled sample sites. DOC was measured at Allander_1, Allander_2, Allander_3, Dawsholm and Dougalston. At the time of sampling for this part of the study the Allander_4 had dried up and could not be included. Figure 109 contains a bar chart showing the DOC concentrations at these sites.
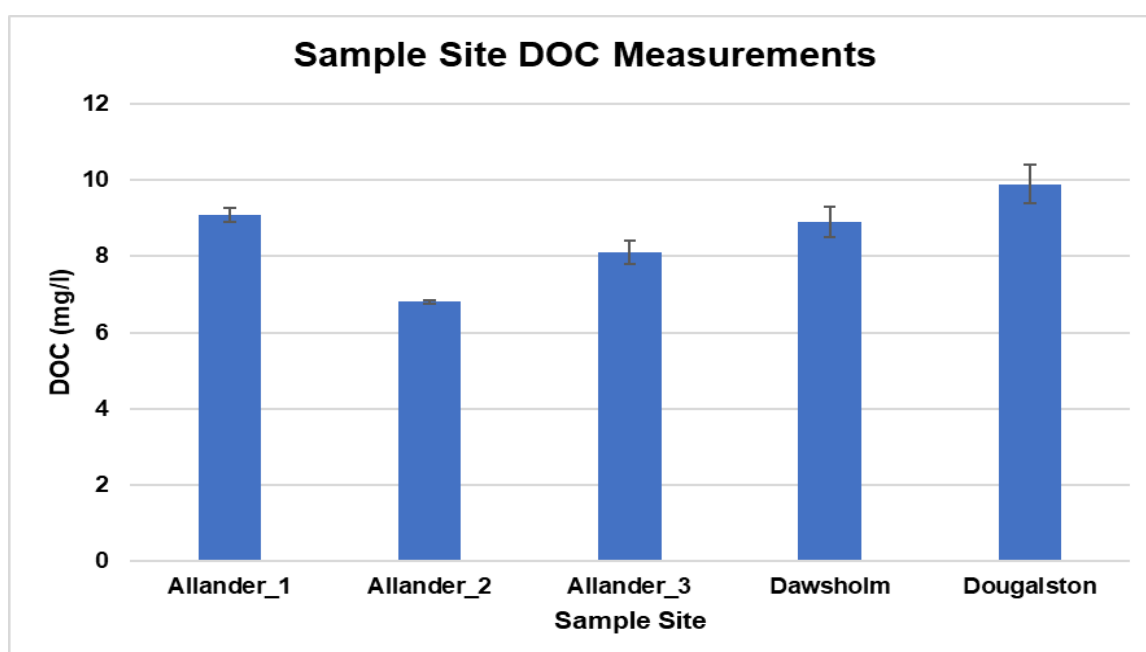
Figure 109 - DOC concentrations of each sample site.

This data shows that at the time of sampling all sites contained similar concentrations of DOC. Dougalston contained the highest concentration of 9.9 mg/ while Allander_2 contained the lowest concentration of 6.8 mg/l. The British Geological Survey have surveyed Scottish groundwater and stated that DOC concentrations are typically a few mg/l and barring anomalous results ca. 12 mg/l is the upper concentration seen.[19] This shows that the collected data are within the range of DOC concentrations found in Scottish groundwater. Fleming *et al.* showed that *L.ochracea* dominates BIOX mats when DOC concentrations are ca. 7 – 8 mg/l while *Gallionellaes* dominate at lower concentrations of ca. 4 – 5 mg/l.[24] This is in agreement with what is seen in this study as the BIOX mats contained predominantly *L.ochracea* filaments at the time of sampling.

6.6.8 Dissolved Inorganic Carbon (DIC)

DIC is comprised of three major aqueous components, carbon dioxide (CO_2), bicarbonate (HCO_3^-) and carbonate (CO_3^{2-}), which play a vital role in the pH buffering of natural water systems.[28, 29] The equilibrium of these components can be seen in Equation 19.

$$H_2O_{(l)} + CO_{2(g)} \rightleftharpoons H_2CO_{3(aq)} \rightleftharpoons H^+_{(aq)} + HCO_3^-_{(aq)} \rightleftharpoons 2H^+_{(aq)} + CO_3^{2-}_{(aq)}$$
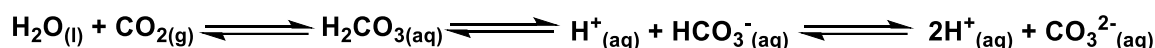
Equation 19 - Equilibrium reaction scheme of DIC species in water.

The formation of DIC is multistep process the involves the weathering of carbonate containing rocks and soils along with the microbial degradation of organic matter.[30] These processes release CO_2 which dissolves in water to form carbonic acid (H_2CO_3), a highly unstable molecule that rapidly dissociates to yield bicarbonate and carbonate ions.[31] It has been thought that DIC was utilised primarily by photosynthetic organisms and autotrophic bacteria, these are bacteria that produce complex organics from CO_2, however research has shown that mixotrophic bacteria can also utilise DIC as a source of energy.[32, 33] Recent research by Fleming *et al.* has been indicated that *L.ochracea* may be a mixotrophic bacterium meaning that DIC concentration may be relevant for growth.[22] Figure 110 contains a bar chart showing the DIC concentration at four sample sites. It was not possible to collect samples from the Dougalston site at the time of sampling.
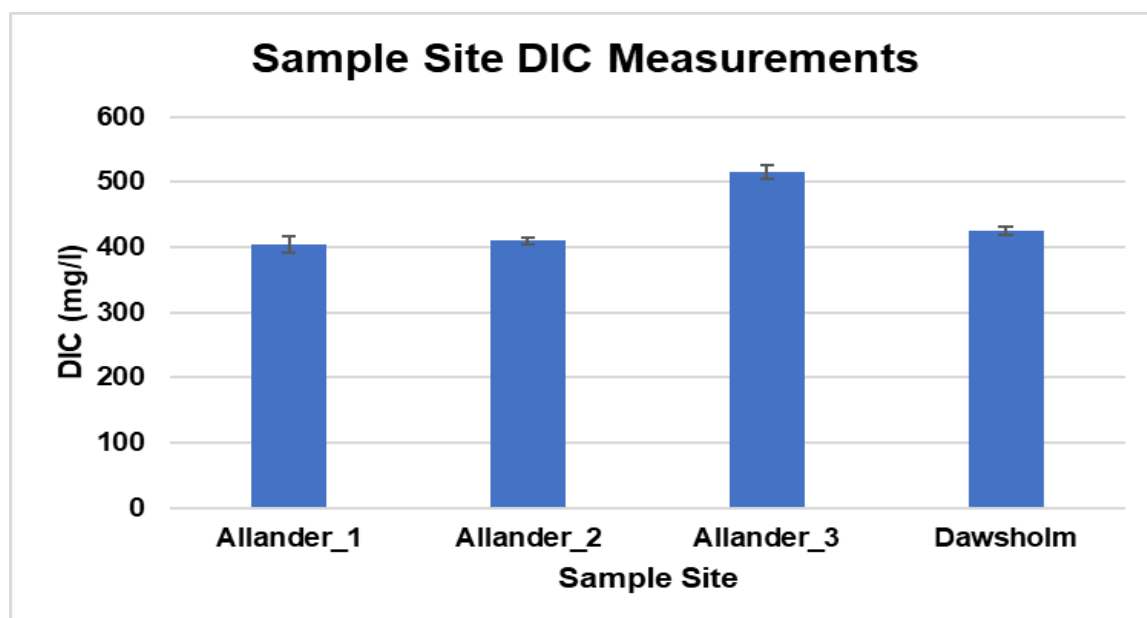
Figure 110 - DIC concentrations of each sample site.

These DIC values can all be classed as high, meaning that the water sampled from these four sites contain an abundance of carbonate minerals and can be described as hard.[29]

Allander_1, Allander_2 and Dawsholm all contain similar concentrations of DIC of ca. 400 mg/l while Allander_3 contains the highest concentration of 514 mg/l. Allander_1 and Allander_2 having similar concentrations is expected as they share the same body of water. The fact that Allander_3 contains considerably more DIC may be indicative of a slightly different geology and increased abundance of carbonate containing minerals such as calcium carbonate when compared with the other sample sites. DIC values have historically been less frequently measured than DOC values meaning that it is difficult to make a comparison with other Scottish waters.[30]

6.6.9 Measurement of Inorganic Species

As the composition of groundwater is dependent on various environmental and geological factors it was appropriate to analyse the inorganic species present within each sample site. This will allow a better understanding of the conditions in which *L.ochracea* thrive meaning that artificial media could potentially be created to mimic this. To do this, water samples collected from Allander_1, Allander_3, Dawsholm and Dougalston were sent to Ivario, a company which carries out water quality testing, for inductively coupled plasma optical emission spectroscopy (ICP-OES) analysis. Water hardness, pH, conductivity and total nitrate and nitrite measurements were also carried out. As Allander_1 and Allander_2 share the same body of water it was decided that only one of these sites should be sampled. Allander_4 was again inaccessible during this period of the study. The resulting data can be found in Table 29

Element/Compound	Allander_1 (12/4/21) Conc (µM)	Allander_3 (28/6/21) Conc (µM)	Dawsholm (12/4/21) Conc (µM)	Dougalston (28/6/21) Conc (µM)
Aluminium	0.7	1.0	0.5	2.4
Ammonium	61.0	105.3	55.4	41.6
Barium	2.0	3.1	2.5	0.4
Beryllium	< 0.1	< 0.1	< 0.1	< 0.1
Boron	22.2	29.6	24.1	2.0
Calcium	1871.4	2270.6	2395.3	424.2
Chloride	239.8	211.6	2708.0	818.1
Chromium	< 0.2	< 0.2	< 0.2	< 0.2
Cobalt	< 0.2	< 0.2	< 0.2	< 0.2
Iron	232.8	429.8	102.1	268.6
Lithium	2.3	2.7	1.4	0.1
Magnesium	234.5	312.7	534.9	164.6
Manganese	10.6	15.7	3.3	34.6
Nickel	0.2	0.2	0.2	0.2
Nitrate	8.1	8.1	21.0	8.1
Nitrite	< 2.2	< 2.2	< 2.2	< 2.2
Orthophosphate	0.3	0.4	0.5	0.1
Potassium	96.4	93.0	169.4	36.5
Silver	< 0.1	< 0.1	< 0.1	< 0.1
Sodium	287.1	300.1	2392.3	869.9
Sulfate	249.9	229.0	333.1	42.7
Zinc	< 0.1	< 0.1	< 0.2	< 0.1
Water Hardness (mg/l)	210	260	290	59
pH	6.7	6.8	7.1	6.5
Conductivity (µS/cm)	452	530	880	237
Total nitrate + nitrite (mg/l)	< 0.05	< 0.05	< 0.05	< 0.05

Table 29 - Concentrations of inorganic species present in the sample sites. As, Cd, Cu, Pb, Se, and U were all present at concentrations < 0.05 µM so do not appear in the table.

This data shows that all sample sites are circumneutral, that no one sample site contains any unique inorganic species and that all sample sites analysed contain the same inorganic species at different concentrations. Dawsholm appears to have the highest total concentration of dissolved ions, especially sodium and chloride, which are an order of magnitude higher than in the Allander sites. This agrees with the conductivity data, which shows that at the time of sampling the Dawsholm sample site had the greatest conductivity of the four measured sites and is consistent with the conductivity data recorded in Chapter 6 Section 5.3.

The most commonly occurring element at each sample site is calcium, which is likely derived from calcium carbonate. As discussed in Chapter 6 Section 5.8, weathering of carbonate containing rocks contributes greatly to the DIC concentration. The DIC concentration of these sites can be described as high meaning it is not surprising that calcium is the dominant inorganic present. Furthermore, calcium and magnesium carbonates contribute to water hardness, with water systems containing high concentrations of these described as hard while water systems with low concentrations are described as soft. Table 30 contains water hardness classifications defined by Scottish Water.[34]

Classification	Ca/Mg Carbonates (mg/l)
Soft	< 50
Moderately Soft	50 – 100
Slightly Hard	100 – 150
Moderately Hard	150 - 200
Hard	200 - 300
Very Hard	> 300

Table 30 - Scottish Water definitions of water hardness.[34]

The waters of Allander_1, Allander_3 and Dawsholm can all be classified as hard as they contain water hardness values of > 200 mg/l while Dougalston is moderately soft as it contains a water hardness value of only 59 mg/l. This is reflected in the calcium and magnesium concentrations where Dougalston contains ca. 4 – 5 × less calcium and 1.5 – 3 × less magnesium than the other sample sites. Another species which can contribute to water hardness is sulfates, in the form of calcium and magnesium sulfates. The concentration of this species is again lowest at the Dougalston site, which agrees with anecdotal evidence that the Dougalston site did not have a sulfureous odour while the other sites did.

The dissolved ferrous and iron concentrations, measured via ferrozine assay, have been discussed in Chapter 6 Section 5.6. The iron concentrations provided by Ivario are total iron concentrations as ICP-OES does not differentiate oxidation states. The iron concentrations recorded here are much higher than those in Chapter 6 Section 5.6. This may indicate the dynamic nature of these sample sites as the samples for ferrozine analysis and ICP-OES analysis were collected months apart. Homoncik *et al.* studied manganese concentrations in Scottish groundwater and showed that high iron concentrations are typically seen in waters that also contain high manganese concentrations.[35] Figure 111 shows a scatter plot from their study showing the relationship between iron and manganese concentrations.
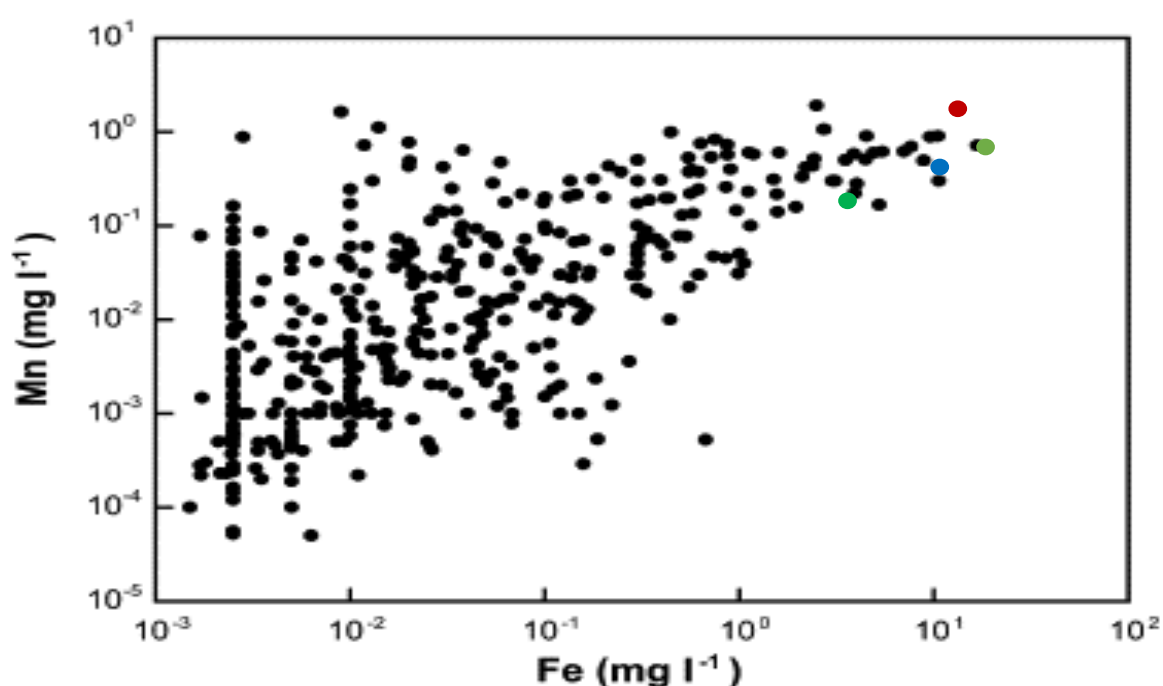
Figure 111 - Scatter plot from a study by Homoncik *et al.* showing the relationship between iron and manganese concentrations in Scottish groundwater.[35] Data from this study has been added as coloured circles. Blue – Allander_1, Orange – Allander_3, Green – Dawsholm and Red – Dougalston.

The highest manganese concentration recorded by Homoncik *et al.*, ca. 20 µM, was found in water containing ca. 70 µM iron while the highest iron concentration recorded, ca. 400 µM, was found in water containing ca. 15 µM manganese. This data is consistent with our ICP-OES data, especially the Allander_3 results which show an iron concentration of ca. 430 µM and manganese concentration of ca. 16 µM. Interestingly, the soft water site at Dougalston contains 34.6 µM manganese which is more than twice the concentration of Allander_3 which has the second highest manganese concentration of 15.7 µM. These differences will be discussed in Chapter 6 Section 6.3.

6.6.10   Characterisation of Organic Species

It is possible that components of DOM are important for the growth of *L.ochracea* however no study, to the author's knowledge, has attempted to characterise the organic species present in the habitats in which they grow. This is likely due to the difficulty of resolving the complexity of the chemicals present within environmental samples. It was decided to attempt analysis using high resolution mass spectrometry of water collected from sample sites where a litre of ground water could be easily sampled. This meant that water was only sampled from two sample sites: Allander_2 and Dawsholm. The full preparation of water samples is described in Chapter 2 Section 3.2. Briefly, water samples (1 l) were collected and promptly acidified to pH 2 to prevent ferric iron precipitation and to protonate side chains so they may bind to the column. Bond Elut PPL cartridges were then activated with three cartridge volumes of methanol then one cartridge volume of HCl (10 mM) and water. Acidified water samples were then concentrated on the column at a flowrate of ca. 10 ml/min. Two cartridge volumes of HCl (10 mM) were eluted to remove salts and the cartridge was dried under vacuum for 5 minutes. Methanol (1 ml) was slowly passed through the column resulting in a dark brown eluent which was wrapped in foil and stored at – 20 °C until needed. This methodology was repeated using $_dH_2O$ as a control, which produced a colourless eluent. Once all samples were concentrated, they were sent to the University of Glasgow Polyomics facility to be analysed by Orbitrap mass spectrometry. Figure 112 shows the resulting total ion chromatograms in positive and negative ion modes and Figure 113 shows a pie chart containing distribution of metabolites assigned using IDEOM software. This attempts to match the exact mass and retention time of a feature to values calculated from known metabolites and environmental compounds in its database of 41,600 compounds. These compounds have been compiled from databases such as Metacyc, Kegg and Human Metabolome Database (HMDB).[36-38]

.
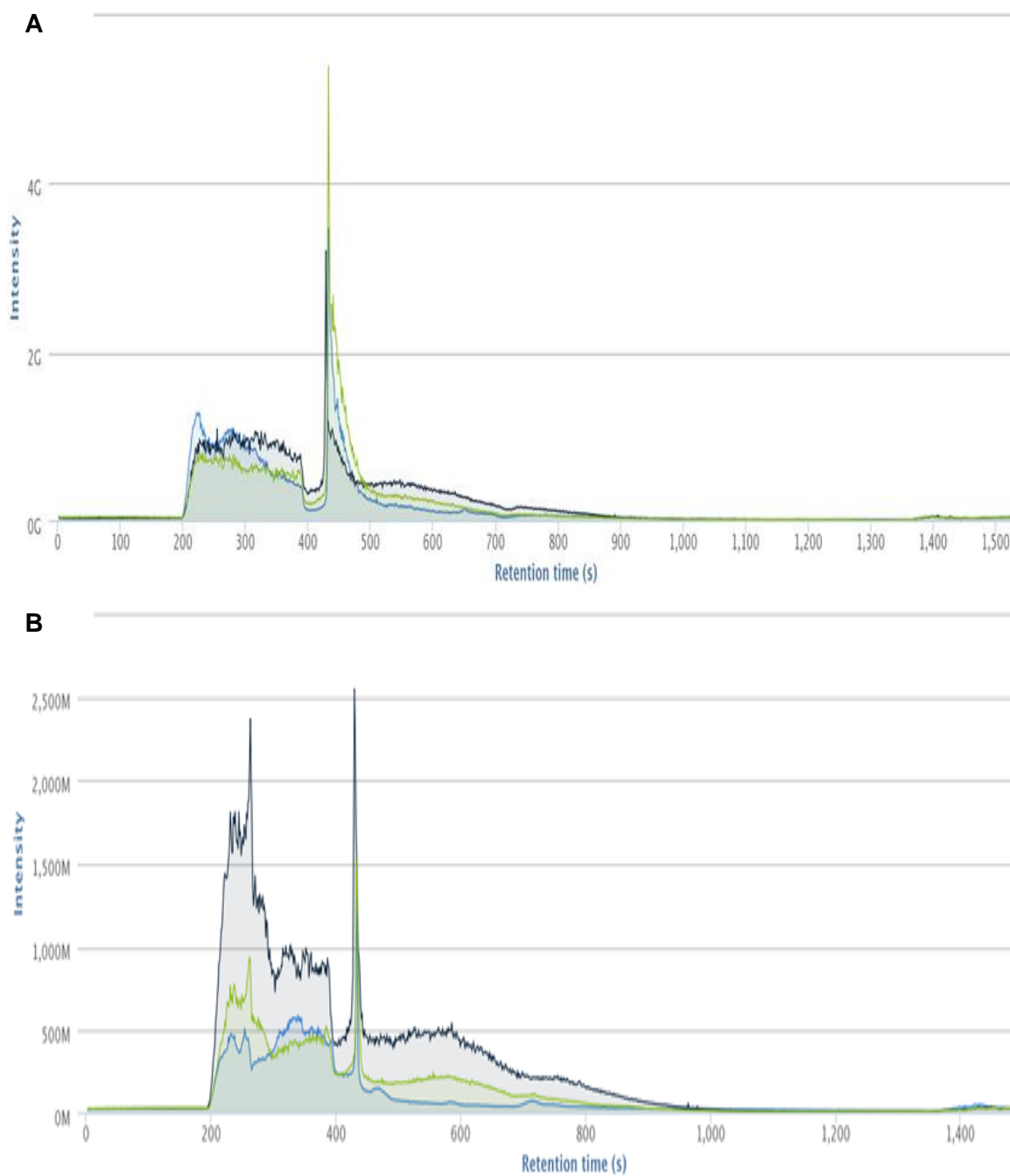
Figure 112 - Total ion chromatograms for: Blue - Blank, Black - Dawsholm and Green - Allander_1. A is in positive ion mode and B is in negative ion mode.
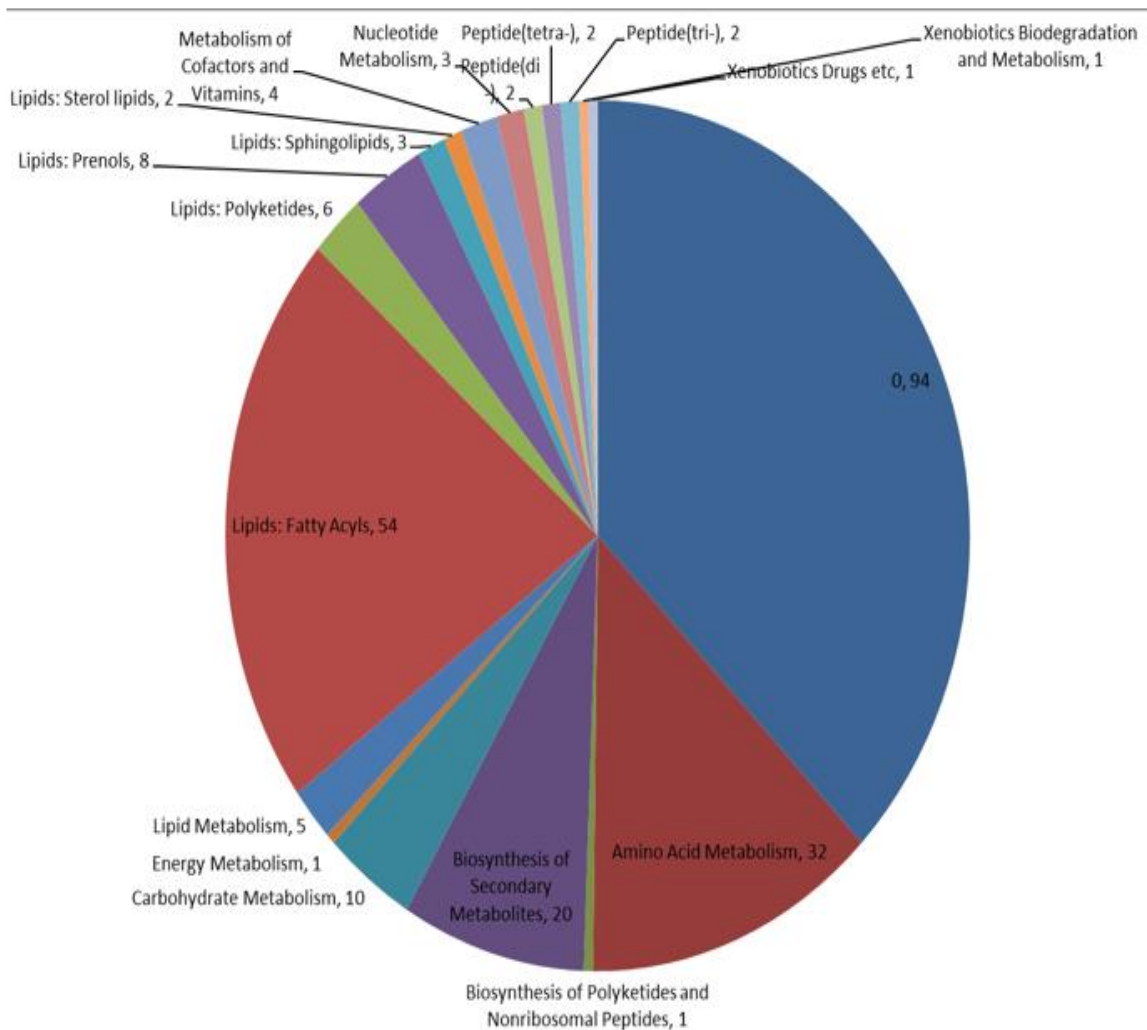
Figure 113 - Pie chart showing the distribution of metabolites annotated in IDEOM according to their class. The area in dark blue represents compounds not assigned to a pathway or class.

Both the Allander_1 and Dawsholm samples produced signals at similar retention times, with the Dawsholm signals being more intense. Unfortunately, the blank sample also produced non-negligible signals and a somewhat similar retention profile. Despite this, a total of 902 organic compounds were putatively identified in the analysis of the samples from the two sites out of a potential 1950 base peak masses recorded. This reflects the sensitivity of the technique, however only 19 compounds were unambiguously identified highlighting the difficulty of identifying components. These 19 compounds were among the 160 standards used to calibrate the system which contained sugars, organic acids, and amino acids amongst other metabolites. This series of 160 standards were run and their retention times used to generate a model for predicting retention times for the whole database of compounds. This model is based on factors such as number of rotatable bonds, positive and negative charges, and log D.[39] Figure 114 shows the agreement between predicted and experimental log retention times.
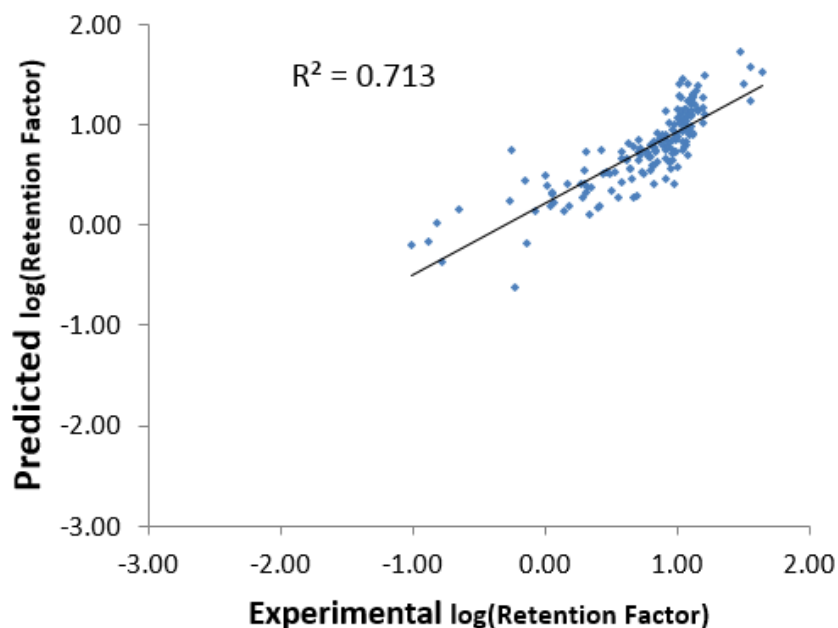
Figure 114 - Agreement between predicted and experimental log retention times.

Of the 902 identified organic compounds 50 were unique to the Dawsholm sample, 10 were unique to the Allander sample and 317 appeared in both profiles. The remaining compounds either produced signal in both sample site profiles and the blank profile or were not identified. As the unique compounds do not appear in the profile of the blank extraction there can be confidence that they are only present within the respective sample site waters. Table 31 shows exact mass data and putative database assignments of these compounds.

This data highlights the potential complexity of the sample sites as masses have been labelled as a variety of different types of compounds including fatty acids (FA), polyketides (PK), flavanones (Fv) and prenol lipids (PR) amongst others.

Sample Site	Exact Mass (g/mol)	Formula	Isomers	Putatative Database Assignment
Allander	138.0350918	C4H10O3S	1	1-butanesulfonate
11/03/2021	139.1361061	C9H17N	1	Pinidine
	226.0589423	C9H10N2O5	1	3-Nitrotyrosine
	226.0958134	C10H14N2O4	2	carbidopa
	266.1646845	C12H27O4P	1	Tributyl phosphate
	272.0719767	C12H16O5S	1	phenyl-1-thio-beta-D-galactopyranoside
	272.1157747	C15H16N2O3	1	indole-3-acetyl-proline
	315.0509709	C24H26N2O16S	1	Prebetanin
	382.2506203	C25H34O3	3	C25-Allenic-apo-aldehyde
	532.3037855	C30H44O8	2	Cucurbitacin J
Dawsholm	242.0941648	C15H14O3	10	Equol
11/03/2021	227.0946192	C14H13NO2	3	N-Acetoxy-4-aminobiphenyl
	314.0618667	C17H12FN2OCl	1	Nuarimol
	255.1257442	C16H17NO2	2	6-(1,2,3,4-Tetrahydro-6-methoxy-2-naphthyl)-2(1H)-pyridone
	298.2148604	C17H30O4	1	[FA (17:1/2:0)] 8E-Heptadecenedioic acid
	316.077472	C20H13N2Cl	1	2-(4-Chlorophenyl)-3-phenyl-3-(2-pyridinyl)acrylonitrile
	218.0983065	C10H18O3S	2	DMB-S-MMP
	232.113974	C11H20O3S	1	2-oxo-10-methylthiodecanoate
	149.1072914	C17H30O4	1	[FA (17:1/2:0)] 8E-Heptadecenedioic acid
	254.057734	C15H10O4	15	[PK] Chrysophanol
	308.0689142	C18H12O5	1	[Fv Methyl(9:1)] 3',4'-Methylenedioxy-[2'',3'':7,8]furanoflavanone
	292.0736573	C18H12O4	2	[Fv] Derriobtusone A
	163.0999233	C10H13NO	3	(R)-2-Methylimino-1-phenylpropan-1-ol
	226.0629634	C14H10O3	2	Anthralin
	269.1414223	C17H19NO2	3	Strobamine
	310.0842265	C18H14O5	1	6-Deoxyjacareubin
	212.0950027	C13H12N2O	2	Harmine
	239.0580824	C14H9NO3	2	Dianthalexin
	147.0895624	C6H13NO3	2	N-hydroxyisoleucine
	228.1150632	C15H16O2	5	[PR] Cacalol
	313.1677823	C19H23NO3	6	Armepavine
	281.1051202	C17H15NO3	3	Annolobine
	297.1005381	C17H15NO4	1	Longifolonine
	222.0679985	C15H10O2	5	2-Phenyl-4-benzopyron
	250.0628413	C8H14N2O5S	2	Glu-Cys
	254.0940005	C16H14O3	5	[Fv Hydroxy,methox] 4'-Hydroxy-2'-methoxychalcone
	329.0722486	C17H13FN3OCl	1	Epoxiconazole
	236.0471911	C7H12N2O5S	1	Asp-Cys
	272.0683597	C15H12O5	26	[Fv] Naringenin
	240.1262217	C15H16N2O	2	Prolyl-2-naphthylamide
	240.0786082	C15H12O3	12	[PK] Chrysophanic acid 9-anthrone
	284.1773879	C19H24O2	7	Androsta-4,9(11)-diene-3,17-dione
	336.1389485	C18H24O4S	1	Steroid O-sulfate
	175.0633738	C10H9NO2	12	Indole-3-acetate
	258.0526783	C14H10O5	5	Gentisin
	234.0352671	C12H10O3S	1	2-(2-Hydroxyphenyl)benzenesulfinate
	212.1201715	C15H16O	2	p-Cumylphenol
	274.2143704	C15H30O4	4	1-Dodecanoyl-sn-glycerol
	294.0892846	C18H14O4	4	[Fv Methoxy(9:1)] 6-Methoxy-[2'',3'':7,8]furanoflavanone
	282.1004963	C16H14N2O3	1	Saphenic acid methyl ester
	276.0785061	C10H16N2O5S	1	Biotinsulfone
	202.0777841	C16H10	1	Pyrene
	282.1617178	C19H22O2	4	4-Prenyldihydropinosylvin
	270.0526564	C15H10O5	24	Pelargonidin
	286.0840641	C16H14O5	38	[Fv] Licodione 2'-methyl ether
	296.1774162	C20H24O2	8	[PR] Crocetindial/ Crocetin dialdehyde
	371.3031566	C21H41NO4	1	Tetradecanoylcarnitine
	244.1210822	C14H16N2O2	3	Cyclo(L-Phe-L-Pro)
	328.2613115	C19H36O4	2	MG(0:0/16:1(9Z)/0:0)
	366.1469139	C22H22O5	8	[Fv] Xanthoangelol C

Table 31 - Exact masses, formulas, number of isomers and putative database assignment of unique compounds from the Allander and Dawsholm samples.

Considering the compounds found at both the Allander and Dawsholm sites, it is encouraging to see the presence of putative compounds that correspond to breakdown components of lignin shown in Figure 115. There was clear presence of leaf litter at both sites and so these would be expected within the organic fraction.

Figure 115 - Various lignin components identified as putative compounds found at both sites studied. The exact masses and the level of confidence of the assignment are shown along with the formular mass and an indication of retention error. The figure was prepared using MarvinView (Chemaxon software).

As described in Chapter 6 Section 4.3, the Dawsholm site is close to historic mining and might expect to be contaminated with pyridine and related methyl derivatives from oil shale waste, however these compounds were not found within the water. The strong sulfurous smell of the site on occasions suggested the presence of thiols that might be too volatile to be detected in this study. Of the compounds unique to Dawsholm, the low molecular weight thiols C_3H_8OS , and $C_3H_6O_3S$ putatively identified as methyl mercaptoethanol and 3-mercaptolactate respectively were identified. Other compounds of note that might be related to coal mining are the presence of polycyclic hydrocarbons with varying degrees of hydroxylation. There is not a great diversity of these compounds, however polycyclic hydrocarbons $C_{16}H_{10}$, and $C_{14}H_{10}O_3$ putatively identified as Pyrene and Anthralin respectively were identified. There are no other polycyclic hydrocarbons identified that are present exclusively in the Allander site or found at both sites.

The exact mass is an extremely powerful discriminating tool for the identification of compounds.[40] The mass spectrometer gives masses accurate to the third or fourth decimal place meaning that the chemical formula identified is almost always correct. It is however extremely challenging to accurately assign these masses to a specific compound through this methodology alone. There can be 20 – 30% errors in the retention times which are in most cases predicted and many of the compounds contain multiple isomeric forms meaning that chemical formula alone is not a unique identifier. An example of this is shown in Figure 116 for the correctly identified formula of $C_{15}H_{14}O_3$ within the database used for the identification. It can be seen from this figure that there are 22 quite different potential molecular structures that can be assigned.
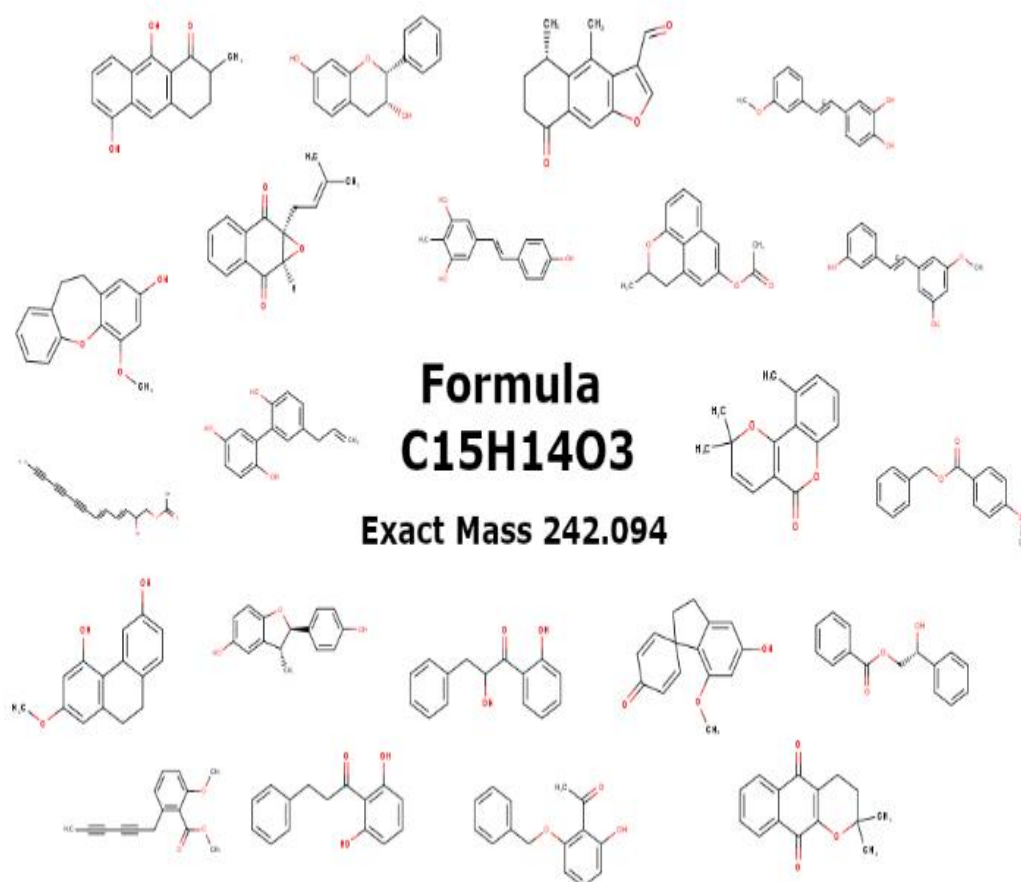
Figure 116 - Potential molecular structures for a compound with formula C$_{15}$H$_{14}$O$_{3}$.

To accurately assign each mass to a chemical structure requires a significant amount of further work utilising multiple standards to match retention times and profiles to individual compounds. However a more promising approach would be in future to compare the metabolic profile of a given source when *L.ochracea* is growing abundantly and conditions when growth is not happening. This would likely eliminate many of the organics as present in both situations as not significant and hopefully would identify a much smaller subset of chemicals for investigation.

This preliminary work has highlighted two potential experimental issues. Firstly, the glassware may not be sufficiently clean meaning that residual organic material has been concentrated on the column and therefore appears in the resulting mass spectra. Secondly LC-MS grade water needs to be used for the blank samples as there may be contaminants in the laboratory water used for the blank sample and glassware washing. The detailed analysis of this experiment should be repeated using fresh glassware that

has been thoroughly acid washed. Literature studies conducting similar experiments have also used furnace blasted glassware however this facility is not accessible within the School of Chemistry.

6.7 Source of Sample Site Water

The potential source of the iron rich ground water where blooms of *L.ochracea* are observed in the greater Glasgow area has not been resolved previously. The industrial past of the city makes it likely that landfill and old mining works contribute to these ground waters. The detailed water analysis performed in the sample sites now allows for a critical assessment of the likely source of the water.

6.7.1 Leachate

Both the Dawsholm site and Allander_1 have been implicated as being close to historic landfill sites. Although the ground water could potentially be leachate the chemical composition measured so for these sites does not support this. The chemical composition of leachate has been shown to be far richer in organics and inorganics than what is seen in the Dawsholm and Allander data. As discussed in Chapter 6 Section 5.7 the DOC concentration of ca. 10 mg/ml at Dawsholm is typical for Scottish groundwater while leachate DOC concentrations are in excess of 1000 mg/l.[9] Furthermore, although the mass spectrometry experiments must be repeated the data produced so far does not show the presence of species such as dioxane which are regularly seen in leachates. Table 32 shows a comparison of the concentrations of selected inorganics present at Dawsholm and Allander_1 compared with typical leachate concentrations from the literature.

Inorganic Species	Dawsholm Conc (mg/l)	Allander_1 Conc (mg/l)	Leachate Literature Conc (mg/l)
Fe	5.7	13	60
Ca	96	75	1000
Mg	13	5.7	500
K	5.1	2.9	60
Na	55	6.6	300
Cl	96	8.5	500
SO_4	32	24	300
NH_3	1	1.1	740

Table 32 - Concentrations of selected inorganic species from Dawsholm compared with literature examples of leachate.[9, 41]

So, despite the evidence of a historic landfill, the data indicates that the water at the Dawsholm and Allander sample sites is highly unlikely to contain leachate. As added proof, leachates are also characterised by their abundance of heavy metals ions, such as cadmium, lead and nickel, whose concentrations in the water sample data in this study are negligible.

6.7.2 Coal Mining

The greater Glasgow area was extensively mined for coal in the 19th century. The Dawsholm sample site in particular is in close proximity to the Garscube coal mine that was mined from 1854-1896. To compare the water samples we used the analysis by Younger who analysed the water composition of 17 Scottish mines.[12] The average concentration of inorganic species in mg/l can be found in Table 33 along with the data from this study in mg/l for comparison. Figure 117 contains box and whisker plots, for

each of the inorganic species, highlighting the distribution of concentrations that were seen.

Inorganic Species	Scottish Mine Average (mg/l)	Allander_1 (mg/l)	Allander_3 (mg/l)	Dawsholm (mg/l)	Dougalston (mg/l)
Fe	54.5	13	24	5.7	15
Mn	3.8	0.58	0.86	0.18	1.9
Ca	176.9	75	91	96	17
Mg	89.6	5.7	7.6	13	4
Na	367.1	6.6	6.9	55	20
K	21.8	2.9	2.8	51	1.1
SO_4	829	24	22	32	4.1
Cl	534.4	8.5	7.5	96	29
Al	1.5	0.019	0.028	0.014	0.066
Zn	0.17	0.0075	0.0085	0.014	0.0091

Table 33 - Average concentrations of inorganic species found in 17 Scottish mines. Data from this study has been provided in mg/l for comparison.[12]
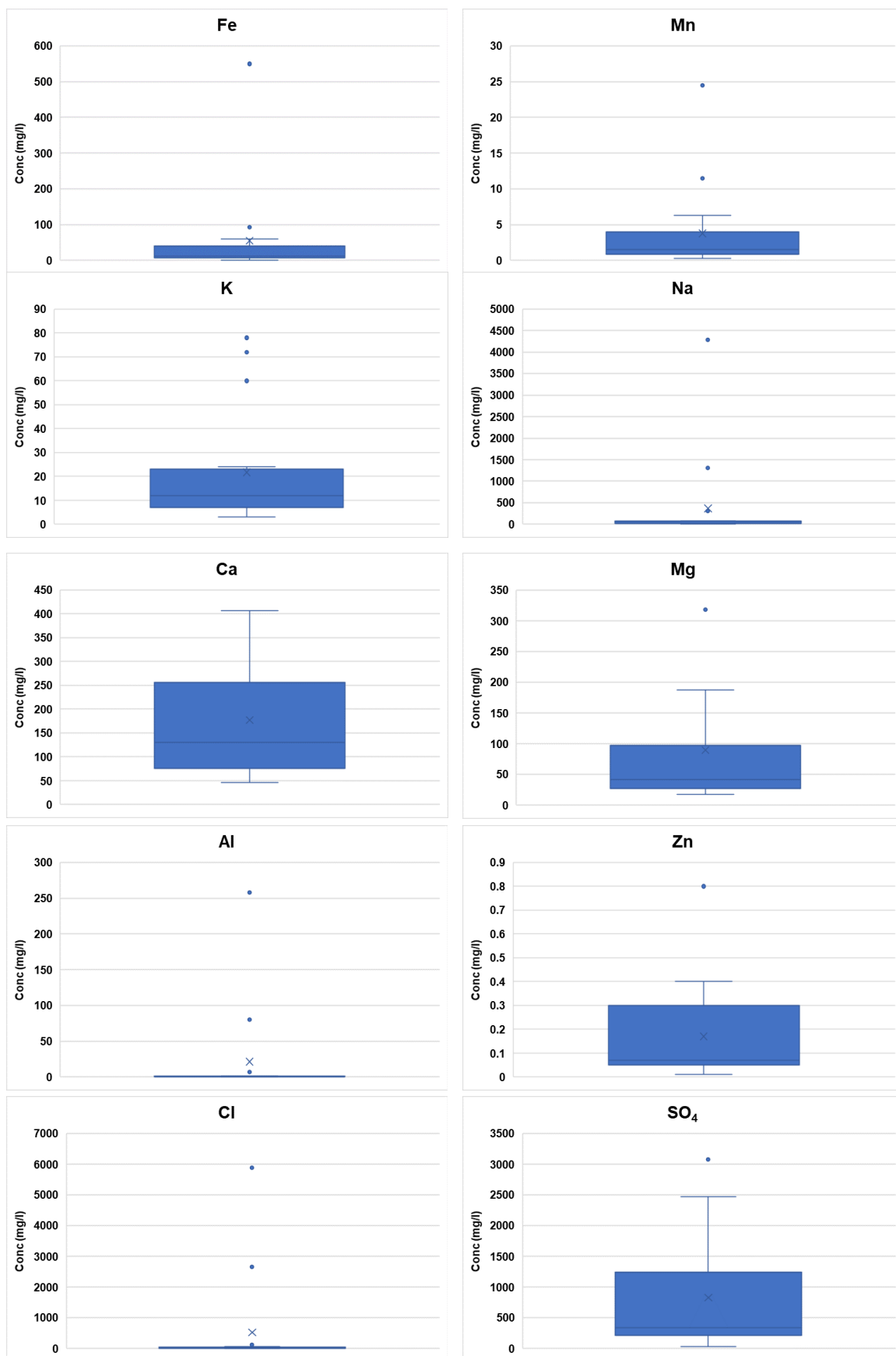
Figure 117 - Box and whisker plots for inorganics of 17 Scottish mines. Data taken from Younger.[12]

The average concentration of inorganic species found within the mine waters are typically multiple times greater than the concentrations found in any of the sample sites. It can be seen however from the box and whisker plots that many species have outlying concentrations which affect this average. Several of the mines analysed were in coastal regions meaning that sea water will have likely contributed to the flooding of these mines resulting in a greater concentration of species including sodium, potassium, calcium and chloride. Considering the whole distribution of Scottish mine water values, it is clear that these are more similar to the sample sites used in this study than the leachate values. Younger also noted that the geological setting of mine water discharge exerts influence over the composition meaning that the concentrations of species can differ greatly from mine to mine.[12] Interestingly, this means that the composition of mine water discharge can be estimated by geological survey maps.

6.7.3 Natural Rock Formations

Digimap is a mapping service available through the University library that allows the user to check the geological features of an area such as bedrock composition. While assessing the areas in which the sample sites are found it was noted that the Allander, Dawsholm and Kelvin sample sites all have bedrock consisting of Limestone Coal formation while the Dougalston site has bedrock consisting of Lawmuir formation. These formations are both defined as Carboniferous and therefore date from the Carboniferous period ca. 358 – 299 million years ago. Carboniferous means "coal bearing" and is used to describe many coal containing bedrocks found throughout the Earth. Limestone, a carbonate containing rock mainly consisting of calcium carbonate, is also a component of Carboniferous bedrock. The Limestone Coal and Lawmuir formations have differing compositions and concentrations of limestone that are layered in a different manner.[42] This likely accounts for the softer water found at the Dougalston site. Figure 118 contains a map provided by the BGS showing the different bedrock aquifers in Scotland. It can be seen that the area surrounding Glasgow, where the sample sites are found, has Carboniferous bedrock that has been extensively mined for coal. This agrees with the Digimap data that the bedrock surrounding the samples sites is Carboniferous. Table 34 shows data from the BGS that highlights the median concentration of certain inorganic species found in Carboniferous bedrock that has been extensively mined for coal.
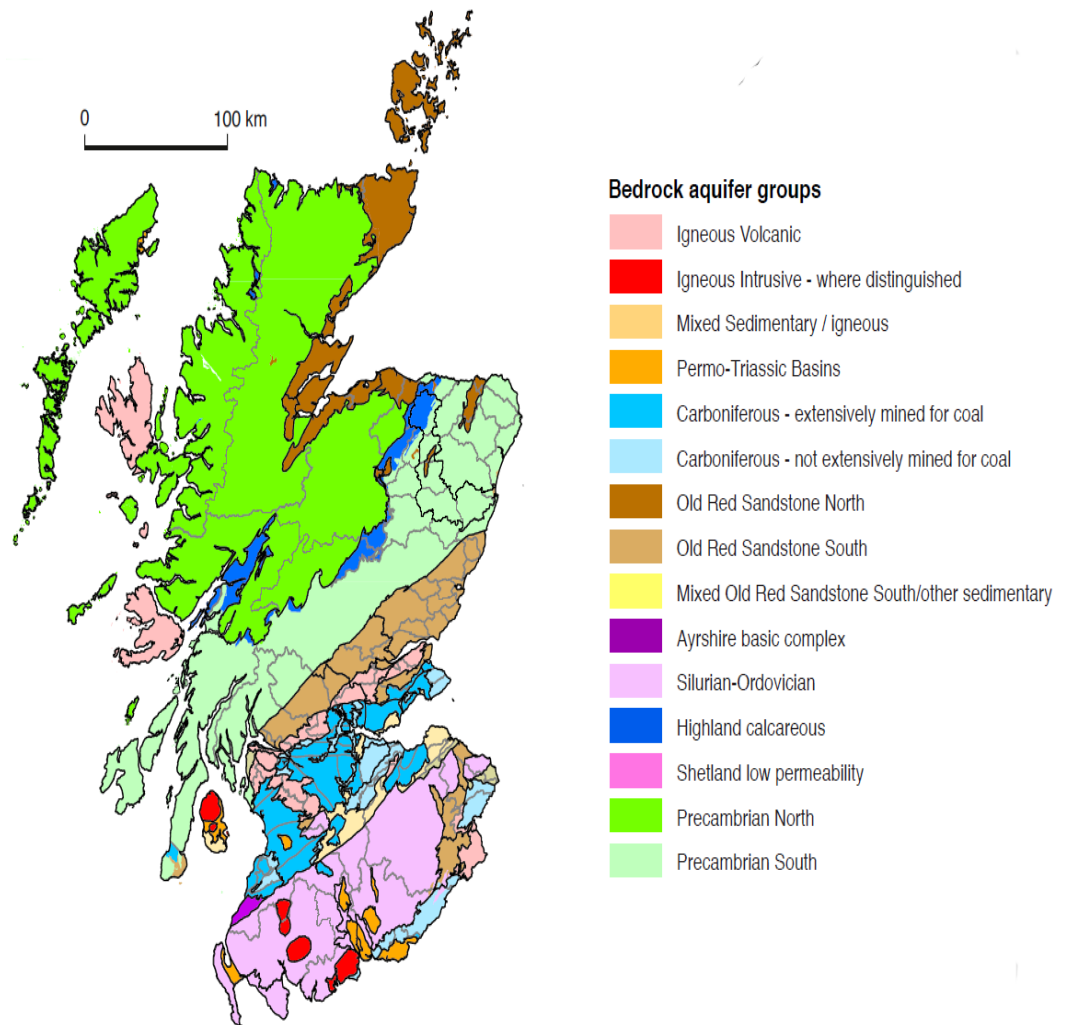
Figure 118 - Map showing bedrock aquifer groups in Scotland.[43]

Element/Compound	Allander_1 Conc (mg/l)	Allander_3 Conc (mg/l)	Dawsholm Conc (mg/l)	Dougalston Conc (mg/l)	Median Carboniferous Aquifers Conc (mg/l)
Calcium	75	91	96	17	71
Chloride	8.5	7.5	96	29	23.4
Iron	13	24	5.7	15	0.675
Magnesium	5.7	7.6	13	4	28.2
Nitrate	0.5	0.5	1.3	0.5	0.13
Potassium	2.9	2.8	5.1	1.1	4.3
Sodium	6.6	6.9	55	20	27.6
Sulfate	24	22	32	4.1	73
Water Hardness	210	290	59	260	324
pH	6.7	7.1	6.5	6.8	7
Conductivity	452	880	237	530	740

Table 34 - Data from the BGS highlighting the median concentration of certain inorganic species in Scottish Carboniferous bedrock that has been extensively mined for coal. This data is compared with the corresponding sample site data.[43]

This data shows that all sample sites from this study have comparable data with that of the BGS. The BGS show that Carboniferous bedrock typically contains very hard water with greater magnesium concentrations than are seen in the sample sites however the calcium data is generally similar, apart from Dougalston which has softer water and a

different type of Carboniferous bedrock. Interestingly, the iron concentrations at all sample sites is far beyond the 90th percentile of the BGS data.

As the hydrogeology of these sites has not been studied in-depth, it is not possible to give a definitive description of the sites. It is however acceptable to use the above data to provide a potential overview of the sites. Given that the concentrations of inorganic species present in the sample sites are consistently low for mine water discharge and are not representative of leachate it can be assumed likely that the waters in these sample sites are not directly sourced by these phenomena. Instead, the sample sites are likely sourced by groundwater that flows through Carboniferous aquifers that have been extensively mined for coal, before discharging in the sample sites. This hypothesis agrees with data from the BGS regarding these bedrock formations and is illustrated in Figure 119.
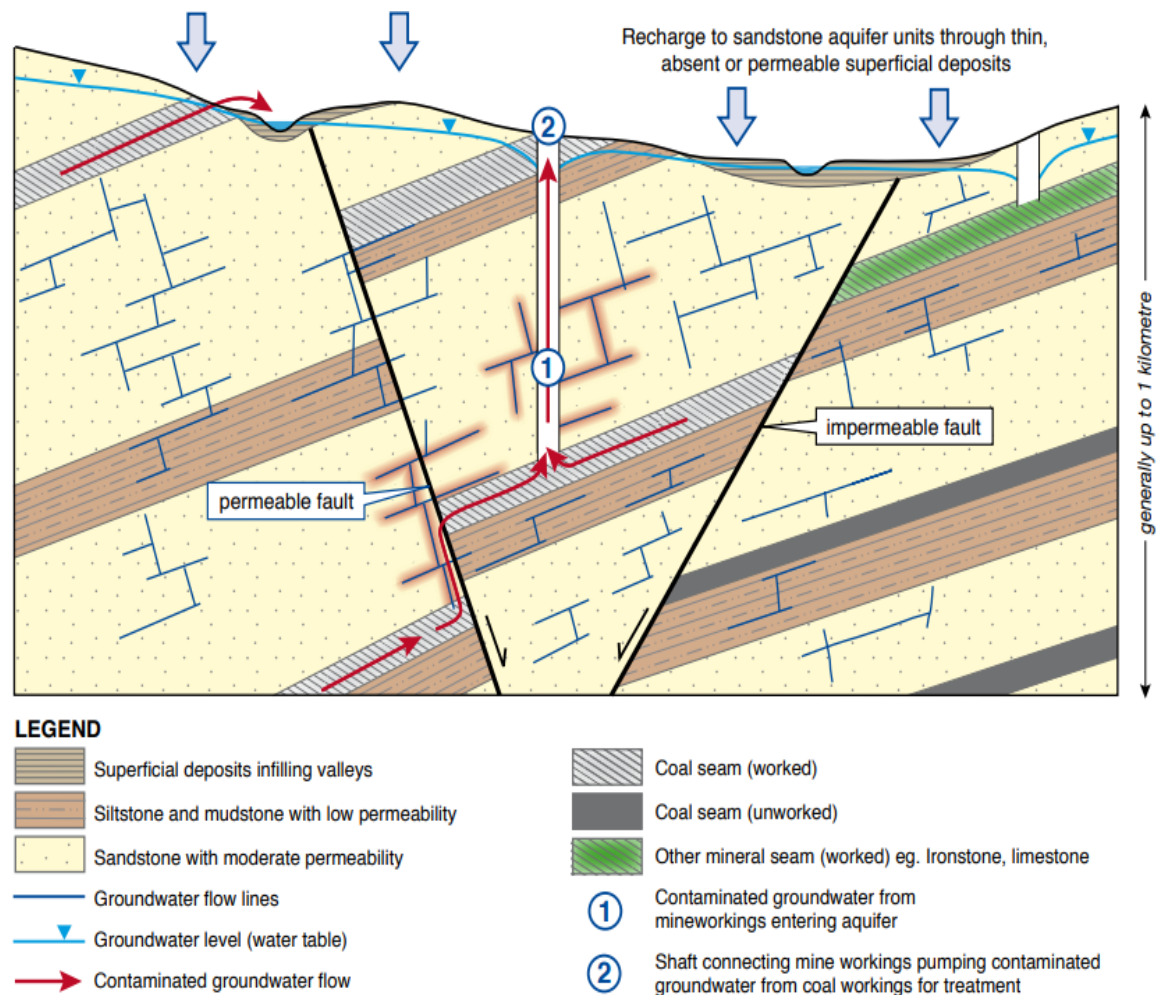
Figure 119 - Schematic cross-section of the hydrogeology of Carboniferous aquifers in Scotland which have been extensively mined for coal.[43]

6.8 Conclusion

Of the sample sites identified during this study, several were characterised, photographed, and compared. Four of the sample sites were on the banks of the Allander Water, one was on the grounds of Dougalston Golf club and the final one was in Dawsholm Park. These sites ranged in size from small pools to streams several meters in length. Frequent photographing of the sites showed them to be transient in nature, with BIOX blooms appearing and disappearing irregularly. BIOX blooms were also noted on Kelvin Walkway, on the tracks of Kelvinhall Subway Station and at a disused coal mine in Allanton however these sites were not characterised in the same manner as the others.

Glasgow has a long industrial past meaning that none of the sites can be considered a green field site. Due to the presence of historical landfill and coal mines near several of the sample sites it was hypothesised that these anthropogenic features may influence the chemical composition of the groundwater present and hence the bacteria present at the sites. To investigate this, a full characterisation of sample sets was recorded. This included physicochemical measurements such as: pH, temperature, ORP, DO, DOC, DIC, conductivity, ferrous iron concentration and total iron concentration. The resulting measurements were generally consistent with values reported for Scottish groundwater. The inorganics present were then analysed by ICP-OES. This measurement was only carried out for Allander_1, Allander_3, Dawsholm and Dougalston. All sample sites analysed contained the same inorganic species however the concentrations differed greatly. Allander_3 contained the greatest concentration of iron while Dougalston was the only site to contain comparatively soft water. The DOC at the various sample sites was in the range of 8 - 10 mg/L which is rather high for ground water in Scotland and seems to have an effect on the bacterial species present.[24] Therefore, a first attempt was made to characterise the organics present in Allander_1 and Dawsholm using Orbitrap mass spectrometry. Despite the results of this experiment being affected by the experimental procedure, due to potential contamination from residual organic material in the glassware or $_d$H$_2$O, usable mass spectrum data was obtained. The data highlights the presence of lignin degradation and flavonoids which would be associated with plant material and leaf litter found at both sites. A small percentage of the mass spectrometry data could be assigned solely to the Dawsholm and Allander samples. This data did not show the appearance of species such as dioxins or xenobiotics which are routinely found in leachates although the Dawsholm site had two polycyclic hydrocarbons not found at the Allander site. Comparing this data with that of Scottish coal mines and landfill leachates

suggests that these anthropogenic features have no significant influence on the chemical composition of the sample site waters.

The geological features of the bedrock aquifers found near to the sample sites was analysed. This shows that all sample sites are surrounded by Carboniferous bed rock that has been extensively mined for coal and that the chemical composition of this bedrock is consistent with that of the sample sites. The Dougalston sample sites contains Lawmuir formations while the others contain Limestone Coal formations. This slight difference in bedrock aquifer likely accounts for the softer water found at the Douglaston sample site. From this it can be stated that the chemical composition of the groundwater at the sample sites is majorly influenced by the surrounding bedrock aquifers.

6.9 References

1. P. Bayer and M. Finkel, *Journal of Contaminant Hydrology*, 2006, **83**, 171-199.
2. D. Nelson, *Groundwater Foundation Annual Meeting*, 2002, 1-8.
3. P. A. Frey and G. H. Reed, *ACS Chemical Biology*, 2012, **7**, 1477-1481.
4. E. Herndon, A. AlBashaireh, D. Singer, T. Roy Chowdhury, B. Gu and D. Graham, *Geochimica et Cosmochimica Acta*, 2017, **207**, 210-231.
5. E. M. Herndon, L. Kinsman-Costello, K. A. Duroe, J. Mills, E. S. Kane, S. D. Sebestyen, A. A. Thompson and S. D. Wullschleger, *Journal of Geophysical Research: Biogeosciences*, 2019, **124**, 227-246.
6. *Journal*, 2020.
7. D. Ityel, *Filtration & Separation*, 2011, **48**, 26-28.
8. A. Alharthi, R. A. Blackley, T. H. Flowers, J. S. J. Hargreaves, I. D. Pulford, J. Wigzell and W. Zhou, *Journal of Chemical Technology & Biotechnology*, 2014, **89**, 1317-1323.
9. P. Kjeldsen, M. A. Barlaz, A. P. Rooker, A. Baun, A. Ledin and T. H. Christensen, *Critical Reviews in Environmental Science and Technology*, 2002, **32**, 297-336.
10. M. C. Nika, K. Ntaiou, K. Elytis, V. S. Thomaidi, G. Gatidou, O. I. Kalantzi, N. S. Thomaidis and A. S. Stasinakis, *Journal of Hazardous Materials*, 2020, **394**, 122493.
11. F. M. Fordyce, S. E. Nice, T. R. Lister, B. É. Ó Dochartaigh, R. Cooper, M. Allen, M. Ingham, C. Gowing, B. P. Vickers and A. Scheib, *Urban Soil Geochemistry of Glasgow - Main Report*, British Geological Survery, 2014.
12. P. L. Younger, *Science of The Total Environment*, 2001, **265**, 309-326.
13. V. P. Evangelou, *Pyrite Oxidation and Its Control*, 1st edn., 1995.
14. F. Doulati Ardejani, B. Jodeiri Shokri, M. Bagheri and E. Soleimani, *Environmental Earth Sciences*, 2010, **61**, 1547-1560.
15. E. Barth-Naftilan, J. Sohng and J. E. Saiers, *Proceedings of the National Academy of Sciences*, 2018, **115**, 6970.
16. J. T. Kulongoski and P. B. McMahon, *npj Climate and Atmospheric Science*, 2019, **2**, 11.
17. D. Emerson, E. J. Fleming and J. M. McBeth, *Annual Review of Microbiology*, 2010, **64**, 561-583.
18. A. M. MacDonald and B. É. Ó Dochartaigh, *Baseline Scotland: an overview of available groundwater chemistry data for Scotland*, British Geological Survey, 2005.
19. A. M. MacDonald, B. É. Ó Dochartaigh and P. L. Smedley, *Baseline Groundwater Chemistry in Scotland's Aquifers*, British Geological Survery, 2017.
20. R. Tyson and T. Pearson, *Geological Society, London, Special Publications*, 1991, **58**, 1-24.
21. A. Kappler, C. Bryce, M. Mansor, U. Lueder, J. M. Byrne and E. D. Swanner, *Nature Reviews Microbiology*, 2021, **19**, 360-374.
22. E. J. Fleming, T. Woyke, R. A. Donatello, M. M. M. Kuypers, A. Sczyrba, S. Littmann, D. Emerson and L. Drake Harold, *Applied and Environmental Microbiology*, 2018, **84**, e02239-02217.
23. T. M. Jeitner, *Analytical Biochemistry*, 2014, **454**, 36-37.
24. E. J. Fleming, I. Cetinić, C. S. Chan, D. Whitney King and D. Emerson, *The Isme Journal*, 2013, **8**, 804.
25. J. A. Rentz, C. Kraiya, G. W. Luther and D. Emerson, *Environmental Science & Technology*, 2007, **41**, 6084-6089.
26. E. M. Thurman, *Organic Geochemistry of Natural Waters*, Springer Netherlands, 1985.

27. F. L. Brailsford, H. C. Glanville, P. N. Golyshin, P. J. Johnes, C. A. Yates and D. L. Jones, *Scientific Reports*, 2019, **9**, 11229.
28. A. Lerman and F. T. Mackenzie, *Carbon in the Geobiosphere - Earth's Outer Shell*, Springer, 1 edn., 2006.
29. J. J. Cole and Y. T. Prairie, in *Reference Module in Earth Systems and Environmental Sciences*, Elsevier, 2014.
30. H. P. Jarvie, S. M. King and C. Neal, *Science of The Total Environment*, 2017, **575**, 496-512.
31. M. H. Moore and R. K. Khanna, *Spectrochimica Acta Part A: Molecular Spectroscopy*, 1991, **47**, 255-262.
32. A. Eiler, *Applied and Environmental Microbiology*, 2006, **72**, 7431-7437.
33. S. C. Maberly and T. V. Madsen, *Aquatic Botany*, 2002, **73**, 1-7.
34. S. Water, *Journal*, 2020.
35. S. C. Homoncik, A. M. MacDonald, K. V. Heal, B. É. Ó Dochartaigh and B. T. Ngwenya, *Science of The Total Environment*, 2010, **408**, 2467-2473.
36. R. Caspi, R. Billington, I. M. Keseler, A. Kothari, M. Krummenacker, P. E. Midford, W. K. Ong, S. Paley, P. Subhraveti and P. D. Karp, *Nucleic Acids Research*, 2020, **48**, D445-D453.
37. M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi and M. Tanabe, *Nucleic Acids Research*, 2016, **44**, D457-D462.
38. D. S. Wishart, D. Tzur, C. Knox, R. Eisner, A. C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M.-A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D. D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G. E. Duggan, G. D. Macinnis, A. M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B. D. Sykes, H. J. Vogel and L. Querengesser, *Nucleic Acids Research*, 2007, **35**, D521-D526.
39. D. J. Creek, A. Jankevics, K. E. V. Burgess, R. Breitling and M. P. Barrett, *Bioinformatics*, 2012, **28**, 1048-1049.
40. K. Clauwaert, S. Vande Casteele, B. Sinnaeve, D. Deforce, W. Lambert, C. Peteghem and J. Van Bocxlaer, *Rapid Communications in Mass Spectrometry : RCM*, 2003, **17**, 1443-1448.
41. T. George and K. Frank, *Handbook of Solid Waste Management, Second Edition*, McGraw-Hill Education, New York, 2nd edn., 2002.
42. M. Dean, M. Browne, C. Waters and J. Powell, *British Geological Survey*, 2011.
43. B. Dochartaigh, É, Ó., A. Macdonald, M., V. Fitzsimons and R. Ward, *Scotland's Aquifers and Groundwater Bodies*, 2015.

7 Conclusions

This study has investigated FeOB and other bacteria found within environmental samples collected primarily from the greater Glasgow area. Aspects including biogenic material characterisation, bacterial population profiling, bacterial isolation and sample site characterisation were investigated through the four main chapters of this study.

Chapter 3 contains a study characterising the BIOX sampled from several different locations. SEM-EDX confirmed the presence of BIOX microstructures consisting primarily of hollow microtubular filaments likely produced by *L.ochracea*. Other microstructures including twisted stalks and trichomes were also noted and are likely produced by *Gallionella spp.* and *Toxothrix spp.* respectively. EDX mapping confirmed the dispersion of iron, oxygen and carbon throughout the structures along with lower concentrations of elements such as silicon and sulfur while XRD showed that the phase of the material was consistent with 2-Fh.

Chapter 3 also aimed to characterise the chemical structure of organic EPS associated with BIOX as understanding its chemical structure may allow a better understanding of how the BIOX microstructures form and allow the development of synthetic mimics. There is currently little information in the literature regarding its chemical structure and as such, attempts to develop an effective protocol which would allow NMR characterisation were carried out. This protocol development utilised a variety of reducing agents to try and dissolve the material as well as NaOH extraction. Ion exchange resins were also utilised to sequester aqueous Fe (II) to mitigate paramagnetic line broadening effects. The resulting NMR spectra indicated that a complex mixture of organics was present however it was not possible to separate and characterise the mixture further at this stage.

During this protocol development it was discovered that the addition of thiol containing reducing agents, DTT and BME, to aqueous suspensions of BIOX or 2-Fh caused them to rapidly turn dark green and aggregate together. This result has not been reported in the literature and was therefore investigated. Reaction conditions including pH, solvent, reducing agent and iron oxide were altered however the green aggregation was found to be unique to the original conditions. Further investigation of this reaction indicated that dark green iron-thiol complexes are unusual, accounting for only ca. 1.2 % of iron-thiol complexes in the CSD, and that a charge transfer complex between the BIOX and DTT is likely forming through the thiol moieties however further study is warranted.

Chapter 4 is critical for this study as it contains a characterisation of the bacterial profile of three Allander sample sites via high-throughput Illumina sequencing of the V3 – V4 region of the 16S rRNA gene. This is important as it provides an insight to the bacterial community profiles at these sites and local biogeochemistry thereby allowing a more thorough understanding of the environment required for *L.ochracea* growth. To the author's knowledge there are currently no studies in the literature investigating the bacterial composition of BIOX mats in the U.K and only a small number of studies globally. Extracted DNA was sequenced by GENEWIZ and cleansing of the sequencing data resulted in libraries containing 688, 937 and 1521 ASV for Allander_1 – 3 respectively. Analysis of these libraries showed that each sample site contained distinct profiles made up predominately of *γ*- and *β*-proteobacteria. Surprisingly, the genus *Leptothrix* was rare when compared with other *β*-proteobacteria however a highly abundant 16S rRNA sequence of a *β*-proteobacterium identified as a *Paucibacter spp.* was found in all three sample sites. This sequence was found to have a 99.23 % sequence identity with *L.ochracea* indicating that they are closely related and that it may be *L.ochracea* in this instance. Filamentous methanotrophic pLW-20 bacteria were found to be abundant in all three sample sites and three of the bacteria isolated in Chapter 5 were found to have a 100 % sequence identity with highly abundant sequences from these libraries.

Allander_2 was found to contain a higher abundance of *Toxothrix* trichomes than the other sample sites. Analysis of the abundance data showed that three *Gallionella spp.* are found in higher abundance at this site than at the others which has led to our hypothesis that one of these sequences may correspond to a *Toxothrix spp.*

To the author's knowledge there is currently no study that compares the phylogenies found in BIOX mats at various sample sites around the globe and as such, these generated libraries were compared with libraries from literature studies of BIOX mats. This comparison indicated that the overall genera present are similar at each site however each site contained several unique sequences however these results are likely highly dependent on the sequencing methodology used. From these results it is clear that the Allander sample sites are more similar to each other than any other site reported in the literature.

Chapter 5 contains a study attempting to isolate *L.ochracea* as an axenic culture and highlights the difficulties associated with this procedure. Multiple isolation methodologies

were employed including solid media, both with agar and gellan gum, liquid enrichments using 0.2 µm filtered sample site water, gradient tubes and micromanipulation. Single colonies of bacteria were isolated on solid media and had their genomes sequenced. These were found to be of the genera *Arcicella, Rhodoferax, Mesorhizobium* and *Polaromonas* respectively and analysis shows that the isolated *Aricella* and *Polaromonas* bacteria likely novel sub-species.

Chapter 5 also contains the development of a protocol utilising a micromanipulator to isolate filamentous bacteria from BIOX samples. To the author's knowledge, micromanipulation has not previously been used in the literature to isolate filamentous *Leptothrix spp*, or any other bacteria, from BIOX mats. Phase contrast microscopy and CLSM was used to confirm the presence of live cells within filamentous material and a micromanipulator was then used to isolate single filaments. Cells within the filaments then had their genomic DNA extracted, amplified by PCR and sequenced. Analysis of the sequencing data showed that the isolated bacteria were in fact methanotrophic *γ*-proteobacteria and not *L.ochracea* or a close relative. This is an interesting result as it indicates that BIOX mats must have a plentiful source of methane for these bacteria to thrive. These isolated bacteria have also not previously been isolated in the literature confirming that micromanipulation can be used to isolate novel filamentous bacteria from BIOX mats. This work has also highlighted the difficulty of assigning filamentous bacteria by appearance as it is now clear that many of the filamentous bacteria frequently seen within the samples are methanotrophic *γ*-proteobacteria and therefore not related to any *Leptothrix spp*.

Chapter 6 contains a study of the sample sites themselves, investigating the physicochemical conditions, chemical composition, bedrock aquifer geology and anthropogenic influences that contribute to them. Historical mapping and discussions with the public allowed for the hypothesis that anthropogenic influences would likely contribute to the chemical composition of sample sites due to the presence of historical landfill and chemical works located near the Allander and Dougalston sample sites respectively. The results from this chapter however highlight that anthropogenic influences likely do not contribute to the chemical composition of the sample sites. This is illustrated by the fact that even though the concentrations of inorganic species differ from site to site they all tend to fall within the expected ranges for Scottish groundwater. The different concentrations at each site instead highlights the effect that the bedrock aquifer likely has on the chemical composition. All sample sites are surrounded by Carboniferous rock that

has been extensively mined for coal. This in turn leads to large deposits of pyritic minerals that when weathered act as a continual source of dissolved ferrous iron and sulfur compounds.

The DIC concentration was measured at chosen sites and found to be within the expected range for Scottish groundwater. Differences in DIC can likely be explained by the bedrock aquifer present due to the variety of possible limestone formations.

The DOC concentration was measured at chosen sites and found to be within the expected range for Scottish groundwater. Attempts were made at characterising the DOC present via Orbitrap mass spectrometry of column concentrated water samples. This generally proved unsuccessful due to the potential contamination of glassware however data from the Dougalston site indicated the presence of species including xenobiotics and dioxins which can be found in leachates.

Overall, this study has combined several areas of research to build an understanding of the bacteria found within BIOX mats and their associated sample sites surrounding the greater Glasgow area. It has also further highlighted the challenges of working with, and trying to isolate, *L.ochracea* which has troubled scientists for two centuries. However, it has also shown that by optimising a micromanipulation protocol, it may be possible to isolate single filaments of *L.ochracea* in the near future.

There is still much research to perform in order to utilise FeOB as a means of producing metal oxide micromaterials in bulk under ambient conditions, and further research is warranted in all areas covered in this thesis. Future researchers may now utilise this work in order to try and isolate *L.ochracea* and comprehensively study its BIOX producing abilities. Once this is fully understood, then using FeOB as a source of metal oxide micromaterials will hopefully one day be achievable.

8 Future Work

Due to time constraints it was unfortunately not possible to thoroughly research all desired areas. The following presents a brief summary of work to be carried out in future.

Chapter 3 contained the development of a protocol to characterise the organic EPS associated with BIOX. As the EPS likely contains polysaccharides it may be possible to dissolve the material into monomer or oligomer units via the utilisation of polysaccharide digesting enzymes. Once subunits have been isolated then further functionalisation such as methylation or N-acetylation may be used to increase hydrophobicity and improve solubility in organic solvents which in turn may allow for a more thorough characterisation.

Chapter 4 contained a study of the bacterial profile of three Allander sample sites. To improve the reliability of this data it is imperative that samples are collected at regular intervals using the same sampling methodology and sequenced to investigate whether the bacterial population is affected by sampling or changes over time. It would also be useful to sequence the samples in parallel via a methodology other than high-throughput Illumina sequencing such as Oxford Nanopore sequencing to compare the compositions of the resulting libraries.

Chapter 5 contained the development of a protocol to isolate filamentous bacteria via micromanipulation. This should be continued to isolate further filaments for sequencing. This protocol could also be used to isolate single filaments and add them to a variety of growth media to investigate growth under controlled conditions.

Chapter 6 contained the characterisation of different sample sites. The mass spectrometry study should be repeated using acid washed or thermally treated glassware to prevent residual organic material contaminating the experiment. LCMS grade water should also be used for the blank experiments to address the possibility that the $_dH_2O$ used was contaminated. It would also be ideal to repeat the extraction process using different resins to extract the maximum amount of organic material as possible. The data gathered from this chapter should then be used to prepare artificial growth media that replicates the chemical composition of the sample sites. This can be combined with the Chapter 5 future micromanipulation work to investigate the growth of filamentous bacteria in controlled media.

Appendix I - R Scripts

Create FASTA File from CSV File

Two different scripts were used to generate FASTA files from data within a CSV file.

Script 1

```
# store the current directory
initial.dir<-getwd()
# change to the working directory
setwd("C:/Users/wd" )
# load the necessary libraries
library(seqinr)
# define the path to .csv files
path <- "C:/sample.csv"
#get sequences input
sample_name <- "sample.csv"
sequences <- read.csv (file = paste0(sample".csv"))
head (sequences)
seqs <- unlist(sequences[,2])
numbers <- seq(1,nrow(sequences),1)
new <- sprintf ("ASV_%04d",numbers)
names <- paste(new, sample_name, sequences[,7], sequences[,8], "Abundance",
sequences[,1], sep=".")
#write fasta file
write.fasta(as.list(seqs),names=names,file.out= paste0(sample_name,".fasta"), open =
"w", nbchar = 60, as.string = TRUE)
# unload the libraries
detach("package:seqinr")
# change back to the original directory
setwd(initial.dir)
```

Script 2

```
# store the current directory

initial.dir<-getwd()

# change to the working directory

setwd("C:/Users/wd" )

# load the necessary libraries

library(seqinr)

# define the path to .csv files

path <- "C:/sample.csv"

#get sequences input

sample_name <- "allander_3_beta"

data_all <- read.csv (file = paste0(sample".csv"))

head (data_all)

#for multiple study need to figure out which column you want e.g. 3 in this example

n <- 3

#and select those sequences which actually have some abundance i.e. > 0

nam2 <- data_all[n] > 0

which_is_pres <- which(nam2)

#get the sequences for those enties

seqsn <- data_all$Sequence[which_is_pres]

#give the fasta entries an informative name

names2 <- paste("Abundance",data_all[which_is_pres,n], colnames(data_all[n]),
data_all$Order[which_is_pres], data_all$Family[which_is_pres],
data_all$Genus[which_is_pres], sep=".")

#write fasta file

write.fasta(as.list(seqsn),names=names2,file.out= paste0(colnames(data_all[n]),".fasta"),
open = "w", nbchar = 60, as.string = TRUE)

# unload the libraries

detach("package:seqinr")

# change back to the original directory

setwd(initial.dir)
```

Create CSV File from FASTA File

```
# store the current directory

initial.dir<-getwd()

# change to the working directory
```

```
setwd("C:/Users/wd")
# load the necessary libraries
library(seqinr)
library(dada2)
# define the path to fasta files
path <- "C:/Users/sample"
seq_file <- file.path(path,"sample.fasta")
#get sequences input
sequences <- read.fasta(file = seq_file, seqtype = "DNA", as.string = TRUE)
name_seq <- substr(attr(sequences,"name"),5,12) # get name of sequences for later only
characters 5-12
proper_seqs <- unlist(sequences)
#get taxonomy of sequences
taxa <- assignTaxonomy(proper_seqs,
"C:/Users/silva_nr99_taxa/silva_nr99_v138_train_set.fa.gz", multithread=FALSE)
taxa.print <- taxa
test_taxa.print <- cbind(Accession = name_seq, Sequence = rownames(taxa.print),
taxa.print) # change sequence rownames to column of sequences
#write .csv file
write.table(unlist(test_taxa.print), file=file.path(path,"taxa.csv") , sep=",",
row.names=FALSE)
# unload the libraries
detach("package:seqinr")
detach("package:dada2")
# change back to the original directory
setwd(initial.dir)
```

Create CSV File from FASTQ File

```
#based on https://benjjneb.github.io/dada2/ITS_workflow.html
# store the current directory
initial.dir<-getwd()
# change to the working directory
setwd("C:/Users/wd" )
# load the necessary libraries
library(dada2)
```

```
library(ShortRead)

library(Biostrings)

# define the path to fasta files

path <- "C:/sample"

# Forward fastq filenames have format: .fastq

fnFs <- sort(list.files(pattern = ".fastq.gz", full.names = TRUE))

# Extract sample names, assuming filenames have format: SAMPLENAME_XXX.fastq

sample.names <- sapply(strsplit(basename(fnFs), ".fastq"), `[`, 1)

#

#fix losses at filtering

setDadaOpt(OMEGA_C = 0)

# Place filtered files in filtered/ subdirectory

fnFs.filtN <- file.path(path ="./filtN", basename(fnFs)) # Put N-filterd files in filtN/ subdirectory

out <- filterAndTrim(fnFs, fnFs.filtN, maxN=5, minLen=60, multithread = FALSE)

#get error model for forward and reverse reads (slow)

errF <- learnErrors(fnFs.filtN, multithread=TRUE)

#plot to check errors sensible

plotErrors(errF, nominalQ=TRUE)

# apply the core sample inference algorithm to the filtered and trimmed sequence data.

dadaFs <- dada(fnFs.filtN, err=errF, multithread=TRUE)

#dadaFs[[1]]

#some useful stuff

seqtab <- makeSequenceTable(dadaFs)

dim(seqtab)

table(nchar(getSequences(seqtab)))

#to deal with identical overlaping sequences of different lengths

seqtab.collapse <- collapseNoMismatch(seqtab, minOverlap = 20, orderBy = "abundance", identicalOnly = FALSE, vec = TRUE, band = -1, verbose = FALSE)

#remove Chimeras

seqtab.nochim <- removeBimeraDenovo(seqtab.collapse, method="consensus", multithread=TRUE, verbose=TRUE)

#track reads through pipeline (assume multiples)

dim(seqtab.nochim)
```

```
getN <- function(x) sum(getUniques(x))

track <- cbind(out, sapply(dadaFs, getN), rowSums(seqtab.collapse),
rowSums(seqtab.nochim))

# If processing a single sample, remove the sapply calls: e.g. replace sapply(dadaFs,
getN) with getN(dadaFs)

colnames(track) <- c("input", "filtered", "denoisedF", "collapsed", "nonchim")

rownames(track) <- sample.names

head(track)

#get taxonomy of sequences

taxa <- assignTaxonomy(seqtab.nochim,
"C:/silva_nr99_taxa/silva_nr99_v138_train_set.fa.gz", tryRC = TRUE,
multithread=FALSE)

taxa.print <- taxa

#collate data and write .csv file

abundance <- as.data.frame(t(seqtab.nochim))

colnames(abundance) <- sample.names

full.print <- cbind(abundance,Sequence = colnames(seqtab.nochim),taxa.print)

write.table(full.print, file=file.path(path,"sample.csv"), sep=",", row.names=FALSE)

# unload the libraries

detach("package:dada2")

# change back to the original directory

setwd(initial.dir)
```

Create CSV File from Forward and Reverse FASTQ Files

```
#based on https://benjjneb.github.io/dada2/tutorial.html

# store the current directory

initial.dir <- getwd()

# change to the working directory

setwd("C:/wd" )

# load the necessary libraries

library(dada2)

# define the path to fastq files

path <- "C:/sample"

# Forward and reverse fastq filenames have format: SAMPLENAME_R1_001.fastq and
SAMPLENAME_R2_001.fastq
```

```
fnFs <- sort(list.files(path, pattern = "_R1_001.fastq.gz", full.names = TRUE))
fnRs <- sort(list.files(path, pattern = "_R2_001.fastq.gz", full.names = TRUE))
# Extract sample names, assuming filenames have format: SAMPLENAME_XXX.fastq
sample.names <- sapply(strsplit(basename(fnFs), "_R"), `[`, 1)
# Place filtered files in filtered/ subdirectory
filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path, "filtered", paste0(sample.names, "_R_filt.fastq.gz"))
names(filtFs) <- sample.names
names(filtRs) <- sample.names
# trim forward and reverse reads as appropriate
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimRight=6,
        maxEE=c(2,3), truncQ=2, rm.phix=TRUE,

                                        compress=TRUE, multithread=FALSE)
head(out)
#fix losses at filtering
setDadaOpt(OMEGA_C = 0)
#get error model for forward and reverse reads (slow)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
#plot to check errors sensible
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)
# apply the core sample inference algorithm to the filtered and trimmed sequence data.
dadaFs <- dada(filtFs, err=errF, multithread=TRUE)
dadaRs <- dada(filtRs, err=errR, multithread=TRUE)
#dadaFs[[1]]
#merge paired reads
mergers <- mergePairs(dadaFs, filtFs, dadaRs, filtRs, maxMismatch = 1, verbose=TRUE)
# Inspect the merger data.frame from the first sample
head(mergers)
#some useful stuff
seqtab <- makeSequenceTable(mergers)
dim(seqtab)
```

```
table(nchar(getSequences(seqtab)))
```

```
#remove Chimeras
```

```
seqtab.nochim <- removeBimeraDenovo(seqtab, method="consensus",
multithread=TRUE, verbose=TRUE)
```

```
dim(seqtab.nochim)
```

```
#track reads through pipeline (assume 1 pair)
```

```
getN <- function(x) sum(getUniques(x))
```

```
track <- cbind(out, getN(dadaFs), getN(dadaRs), getN(mergers),
rowSums(seqtab.nochim))
```

```
colnames(track) <- c("input", "filtered", "denoisedF", "denoisedR", "merged", "nonchim")
```

```
rownames(track) <- sample.names
```

```
head(track)
```

```
#get taxonomy of sequences
```

```
taxa <- assignTaxonomy(seqtab.nochim,
"C:/silva_nr99_taxa/silva_nr99_v138_train_set.fa.gz", multithread=FALSE)
```

```
taxa.print <- taxa
```

```
#collate data and write .csv file
```

```
 full.print <- cbind(Abundance = (seqtab.nochim[1,]), Sequence =
colnames(seqtab.nochim),taxa.print)
```

```
 write.table(full.print, file=file.path(path, paste0(sample.names,".csv")), sep=",",
row.names=FALSE)
```

```
# unload the libraries
```

```
detach("package:dada2")
```

```
# change back to the original directory
```

```
setwd(initial.dir)
```

Appendix II - 16S rRNA Sequences of Isolated Bacteria

This appendix contains the 16S rRNA sequences of all isolated bacteria.

Arcicella

tttacaatggagagtttgatcctggctcaggatgaacgctagcggcaggcctaatacatgcaagtcgaacggtgtagcaatac
atagtggcgcacgggtgcgtaacacgtatgcaacctaccttatactgggggatagcctttggaaacggagattaatacccat
agtacatttgattggcatcaattaaatgttaaagatttatcggtataagctgggcatgcgtccaattagttagttggtgaggtaatg
gctcaccaagactttgattggtaggggaactgagaggttaatcccccacactggcactgagatacgggccagactcctacgg
gaggcagcagtagggaatattgggcaatggaggcaactctgacccagccatgccgcgtgcaggaagaaggcgttatgcgt
tgtaaactgcttttatataggaagaaatagtccttgcgaggaaagttgacggtactatatgaataagcaccggctaactccgtg
ccagcagccgcggtaatacggagggtgcaagcgttgtccggatttattgggtttaaagggtgcgtaggcggtttattaagtcag
tggtgaaagacggtcgctcaacgattgcagtgccattgatactggtagacttgagtgggattgaggtagctggaatggatagt
gtagcggtgaaatgcatagatattatccagaacaccaattgcgtaggcaagttactaagtctcaactgacgctgaggcacga
aagtgtgggtatcaaacaggattagatacectggtagtccacactgtaaacgatgataactaactgttggctttcgagtcagtg
gtacagagaaatcgttaagttatccacctggggagtacgccggcaacggtgaaactcaaaggaattgacgggggtccgca
caagcggtggagcatgtggtttaattcgatgatacgcgaggaaccttacctaggctagaatgtgaaggaatgtatcagaaatg
gtgcagtcagcaatgacctgaaacaaggtgctgcatggctgtcgtcagctcgtgccgtgaggtgttgggttaagtcccgcaac
gagcgcaacccctgttgttagttgccatcaggtaatgctgggaactctaacaagactgcctacgcaagtagagaggaagga
ggggacgacgtcaagtcatcatggcccttacgcctagggcaacacacgtgctacaatggacggtacagcaggtcgctatgt
ggtaacacaatgccaatctctaaagccgttctcagttcggattggggtctgcaactcgaccctatgaagctggaatcgctagta
atcggatatcagctatgatccggtgaatacgttcccggaccttgtacacaccgcccgtcaagccatgaaagtctggtagacct
aaagctggttgtcaaaacaccagttagggtagaacaggtaattggggctaagtcgtaacaaggtagccgtaccggaaggtg
tggctggaacacctcctttttg

Mesorhizobium allander

tttaaacatgagagtttgatcctggctcagaacgaacgctggcggcaggcttaacacatgcaagtcgagcgccccgcaagg
ggagcggcagacgggtgagtaacgcgtgggaatctacccatcactacggaacaactccgggaaactggagctaataccg
tatacgtccttcgggagaaagatttatcggtgatggatgagcccgcgttggattagctagttggtggggtaatggcctaccaag
gcgacgatccatagctggtctgagaggatgatcagccacactgggactgagacacggcccagactcctacgggaggcag
cagtgggggaatattggacaatgggcgcaagcctgatccagccatgccgcgtgagtgatgaaggccctaggggttgtaaagct
ctttcaacggtgaagataatgacggtaaccgtagaagaagccccggctaacttcgtgccagcagccgcggtaatacgaag
ggggctagcgttgttcggatttactgggcgtaaagcgcacgtaggcggattgttaagttaggggtgaaatcccagggctcaac
cctggaactgcctttaatactggcaatctcgagtccggaagaggtgagtggaattccgagtgtagaggtgaaattcgtagatat
tcggaggaacaccagtggcgaaggcggctcactggtccggtactgacgctgaggtgcgaaagcgtggggagcaaacag
gattagataccctggtagtccacgctgtaaacgatggaagctagccgtcggcaagtttacttgtcggtggcgcagctaacgca
ttaagcttcccgcctggggagtacggtcgcaagattaaaactcaaaggaattgacggggggcccgcacaagcggtggagca
tgtggtttaattcgaagcaacgcgcagaaccttaccagcccttgacatcccggtcgcggcctagagagatttaggccttcagtt
cggctggaccggtgacaggtgctgcatggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaa
ccctcgccccttagttgccatcattcagttgggcactctaagggggactgccggtgataagccgagaggaaggtggggatgacg
tcaagtcctcatggcccttacgggctgggctacacacgtgctacaatggtggtgacagtgggcagcgagaccgcgaggtcg
agctaatctccaaaagccatctcagttcggattgcactctgcaactcgagtgcatgaagttggaatcgctagtaatcgcggatc
agcatgccgcggtgaatacgttcccgggccttgtacacaccgcccgtcacaccatgggagttggtttacccgaaggcgctgt
gctaaccgcaaggaggcaggcgaccacggtagggtcagcgactggggtgaagtcgtaacaaggtagccgtaggggaac
ctgcggctggatcacctcctttcta

Rhodoferax allander

gctcagattgaacgctggcggcatgccttacacatgcaagtcgaacggcagcacgggagcaatcctggtggcgagtggcg
aacgggtgagtaatatatcggaacgtgcccagtcgtgggggataacgcagcgaaagctgtgctaataccgcatacgatctat
ggatgaaagcgggggatcgcaagacctcgcgcgattggagcggccgatatcagattagctagttggtggggtaaaagccc
accaaggcgacgatctgtagctggtctgagaggacgaccagccacactggaactgagacacggtccagactcctacggg
aggcagcagtggggaattttggacaatgggcgcaagcctgatccagcaatgccgcgtgcaggatgaaggccttcgggttgt
aaactgcttttgtacggagcgaaacggtctcttctaataaaggggctaatgacggtaccgtaagaataagcaccggctaact
acgtgccagcagccgcggtaatacgtagggtgcaagcgttaatcggaattactgggcgtaaagcgtgcgcaggcggttatat
aagacagatgtgaaatccccgggctcaacctgggacctgcatttgtgactgtatagctagagtacggtagaggggatgga
attccgcgtgtagcagtgaaatgcgtagatatgcggaggaacaccgatggcgaaggcaatcccctggacctgtactgacgc
tcatgcacgaaagcgtggggagcaaacaggattagataccctggtagtccacgccctaaacgatgtcaactggttgttgggt
cttcactgactcagtaacgaagctaacgcgtgaagttgaccgcctggggagtacggccgcaaggttgaaactcaaaggaat
tgacggggacccgcacaagcggtggatgatgtggtttaattcgatgcaacgcgaaaaaccttacccacctttgacatgtacg
gaagtcgctagagatagcttcgtgctcgaaagagagccgtaacacaggtgctgcatggctgtcgtcagctcgtgtcgtgagat
gttgggttaagtcccgcaacgagcgcaacccttgtcattagttgctacattcagttgggcactctaatgagactgccggtgaca
agccggaggaaggtggggatgacgtcaagtcctcatggcccttataggtggggctacacacgtcatacaatggctggtaca
aagggttgccaacccgcgaggggggagctaatcccacaaagccagtcgtagtccggatcgtagtctgcaactcgactacgtg
aagtcggaatcgctagtaatcgtggatcagaatgtcacggtgaatacgttcccgggtcttgtacacaccgcccgtcacaccat
gggagcgggttctgccagaagtagttagcctaaccgcaag

Polaromonas allander

cgaactatagagtttgatcctggctcagattgaacgctggcggcatgccttacacatgcaagtcgaacggcagcacgggag
caatcctggtggcgagtggcgaacgggtgagtaatatatcggaacgtgcccagtcgtgggggataacgtagcgaaagcta
cgctaataccgcatacgatctacggatgaaagcgggggactcgcaagagcctcgcgcgattggagcggccgatatcagat
taggttgttggtgaggtaaaagctcaccaagcctgcgatctgtagctggtctgagaggacgaccagccacactggaactgag
acacggtccagactcctacgggaggcagcagtggggaattttggacaatggacgaaagtctgatccagcaatgccgcgtg
caggacgaaggccttcgggttgtaaactgcttttgtacggaacgaaacggtctgctttaatacagtgggctaatgacggtaccg
taagaataagcaccggctaactacgtgccagcagccgcggtaatacgtagggtgcgagcgttaatcggaattactgggcgt
aaagcgtgcgcaggcggttatataagacagatgtgaaatccccgggctcaacctgggacctgcatttgtgactgtatagctag
agtacggtagaggggatggaattccgcgtgtagcagtgaaatgcgtagatatgcggaggaacaccgatggcgaaggca
atcccctggacctgtactgacgctcatgcacgaaagcgtggggagcaaacaggattagatacctggtagtccacgcccta
aacgatgtcaactggttgttgggtcttcactgactcagtaacgaagctaacgcgtgaagttgaccgcctggggagtacggccg
caaggttgaaactcaaaggaattgacggggacccgcacaagcggtggatgatgtggtttaattcgatgcaacgcgaaaaa
ccttacccacctttgacatgtacggaattcgccagagatggcttagtgctcgaaagagaaccgtaacacaggtgctgcatggc
tgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaacccttgtcattagttgctacatttagttgggcact
ctaatgagactgccggtgacaaaccggaggaaggtggggatgacgtcaagtcctcatggcccttataggtggggctacaca
cgtcatacaatggctggtacaaagggttgccaacccgcgaggggggagctaatcccataaaaccagtcgtagtccggatcgc
agtctgcaactcgactgcgtgaagtcggaatcgctagtaatcgtggatcagaatgtcacggtgaatacgttcccgggtcttgta
cacaccgcccgtcacaccatgggagcgggttctgccagaagtagttagcctaaccgcaaggagggcgattaccacggca
gggttcgtgactggggtgaagtcgtaacaaggtagccgtatcggaaggtgcggctggatcacctcctttct

Methylomonas allander

tttaaactgaagagtttgatcatggctcagattgaacgctggcggtatgcttaacacatgcaagtcgaacggtaacagctcttcg
gaggctgacgagtggcggacgggtgagtaatgcgtaggaatctgcctagtagtgggggataacgtggggaaactcacgct
aataccgcatacgccctacggggggaaaacggggacgccttcgggcctcgcgcgcattagatgagcctacgtcagattagcttgtt
ggtgaggtaacggctcaccaaggcgacgatctgtagctggtctgagaggatgatcagccacactggaactgagacacggt

ccagactcctacgggaggcagcagtggggaatattggacaatgggcgcaagcctgatccagcaataccgcgtgtgtgaag
aaggccttagggttgtaaagcactttcaattgggaggaataacctagcgatcaatacccgctagcttgacattacctttagaaga
agcaccggctaactccgtgccagcagccgcggtaatacggagggtgcgagcgttaatcggaattactgggcgtaaagagt
gcgtaggcggtcgtttaagtcagatgtgaaagccccgggctcaacctgggaacggcatttgatactggacgactagagtttgg
gagaggcaagtggaatttcaggtgtagcggtgaaatgcgtagagatctgaaggaacaccagtggcgaaggcgacttgctg
gcctaaaactgacgctgaggcacgaaagcatgggtagcaaacaggattagataccctggtagtccatgccgtaaacgatgt
caactaaccgttgggtctctttgagacttagtggtggagctaacgtattaagttgaccgcctggggagtacggccgcaaggcta
aaactcaaatgaattgacggggggcccgcacaagcggtggagcatgtggtttaattcgatgcaacgcgaagaaccttaccta
cccttgacatccagagaatctgttagagatagtagagtgccttcgggaactctgagacaggtgctgcatggctgtcgtcagctc
gtgtcgtgagatgttgggttaagtcccgtaacgagcgcaacccttatccttagttgccagcgggttatgccgggaactctaggg
agactgccgatgataaatcggaggaaggtggggacgacgtcaagtcatcatggcccttatgggtagggctacacacgtgct
acaatggtcggtacagagggctgcaaaactgcgaagtcaagcgaatcccagaaagccgatcttagtccggattggagtctg
caactcgactccatgaagtcggaatcgctagtaatcgcgaatcagaatgtcgcggtgaatacgttcccgggccttgtacacac
cgcccgtcacaccatgggagtgggttgcaaaagaagtgggtagtctaaccttcgggagggcgctcaccactttgtgattcatg
actggggtgaagtcgtaacaaggtagccctaggggaacctggggctggatcacctccttaca

Methylobacter allander

taaaactgaagagtttgatcatggctcagattgaacgctggcggtatgcttaacacatgcaagtcgaacggtaacagctcttcgg
aggctgacgagtggcggacgggtgagtaatgcgtaggaatctgcctaatagtgggggataacttagggaaacttaagctaa
taccgcatacgccctacggggggaaagcggggggatcgcaagacctcgcgctattagatgagcctacgtcagattagctagtt
ggtggggtaaaagcccaccaaggcgacgatctgtaactggtctgagaggatgatcagtcacactggaactgagacacggt
ccagactcctacgggaggcagcagtggggaatattggacaatgggcgcaagcctgatccagcaataccgcgtgtgtgaag
aaggccttagggttgtaaagcactttcaattgggaggaatacctgttggttaatacccgacagattgacattacctttagaagaa
gcaccggctaactccgtgccagcagccgcggtaatacggagggtgcgagcgttaatcggaattactgggcgtaaagagtg
cgtaggcggttatttaagtcagatgtgaaagccctgggcttaacctgggaacggcatttgatactgaataactggagtttggga
gaggtaagtggaatttcaggtgtagcggtgaaatgcgtagatatctgaaggaacaccagtggcgaaggcgacttactggcct
aaaactgacgctgaggcacgaaagcatgggtagcaaacaggattagataccctggtagtccatgccgtaaacgatgtcaa
ctaaccgttgggcttctttggaagcttagtggtggagctaacgtattaagttgaccgcctggggagtacggtcgcaagactaaa
actcaaatgaattgacggggggcccgcacaagcggtggagcatgtggtttaattcgatgcaacgcgaagaaccttacctacc
cttgacatccagagaatctgttagagatagtagagtgccttcgggagctctgagacaggtgctgcatggctgtcgtcagctcgt
gtcgtgagatgttgggttaagtcccgtaacgagcgcaacccttatccttagttgccatcaggtcatgctgggaactctagggag
actgccgctgataaagcggaggaaggtggggacgacgtcaagtcatcatggcccttatgggtagggctacacacgtgctac
aatggtcggtacagagggctgcaaaactgcgaagtcaagcaaatcccacaaagccgatcttagtccggattggagtctgca
actcgactccatgaagtcggaatcgctagtaatcgcgaatcagaacgtcgcggtgaatacgttcccgggccttgtacacacc
gcccgtcacaccatgggagtgggttgcaaaagaagtaggtagtctaacccgcaagggagggcgcttaccactttgtgattca
tgactggggtgaagtcgtaacaaggtagccctaggggaacctggggctggatcacctccttaca

Methylobacter dawsholm

attaaactgaagagtttgatcatggctcagattgaacgctggcggtatgcttaacacatgcaagtcgaacggtaacagctcttc
ggaggctgacgagtggcggacgggtgagtaatgcgtaggaatctgcctaatagtgggggataacttagggaaacttaagct
aataccgcatacgccctacggggggaaagcggggggatcgcaagacctcgcgctattagatgagcctacgtcagattagcta
gttggtagggtaaaagcctaccaaggcgacgatctgtaactggtctgagaggatgatcagtcacactggaactgagacacg
gtccagactcctacgggaggcagcagtggggaatattggacaatgggcgcaagcctgatccagcaataccgcgtgtgtga
agaaggccttagggttgtaaagcactttcaattgggaggaatacctattggtcaatacccgatagattgacattacctttagaag
aagcaccggctaactccgtgccagcagccgcggtaatacggagggtgcgagcgttaatcggaattactgggcgtaaagag
tgcgtaggcggttatttaagtcagatgtgaaagccctgggcttaacctgggaacggcatttgatactgaataactggagtttggg
agaggtaagtggaatttcaggtgtagcggtgaaatgcgtagatatctgaaggaacaccagtggcgaaggcgacttactggc
ctaaaactgacgctgaggcacgaaagcatgggtagcaaacaggattagataccctggtagtccatgccgtaaacgatgtca
actaaccgttgggcttctttggaagcttagtggtggagctaacgtattaagttgaccgcctggggagtacggtcgcaagactaa

aactcaaatgaattgacgggggtccgcacaagcggtggagcatgtggtttaattcgatgcaacgcgaagaaccttacctacc
cttgacatccagagaatctgttagagatagtagagtgccttcgggagctttgagacaggtgctgcatggctgtcgtcagctcgtg
tcgtgagatgttgggttaagtcccgtaacgagcgcaacccttatccttagttgccatcaggtcatgctgggaactctagggaga
ctgccgctgataaagcggaggaaggtggggacgacgtcaagtcatcatggcccttatgggtagggctacacacgtgctaca
atggtcggtacagagggctgcaaaactgcgaagtcaagcaaatcccacaaagccgatcttagtccggattggagtctgcaa
ctcgactccatgaagtcggaatcgctagtaatcgcgaatcagaacgtcgcggtgaatacgttcccgggccttgtacacaccg
cccgtcacaccatgggagtgggttgcaaaagaagtaggtagtctaacccgcaagggagggcgcttaccactttgtgattcat
gactggggtgaagtcgtaacaaggtagccctaggggaacctggggctggatcacctccttaca

Appendix III - Allander Illumina Sequencing Data

This appendix contains the 16S rRNA sequences and abundance data of the ten most abundant bacteria found in each sample. The following forty most abundant bacteria then have their abundance data included.

Allander_1

>ASV_0001.Allander_1.Methylomonadaceae.pLW-20.Abundance.6036

TGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCAATACCGCGTGTGTGAAG
AAGGCCTTAGGGTTGTAAAGCACTTTCAATTGGGAGGAATACCTATTCGCTAATACC
GAGTAGATTGACATTACCTTTAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGC
GGTAATACGGAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGAGTGCGTAG
GCGGTTATTTAAGTCAGATGTGAAAGCCCTGGGCTTAACCTGGGAACGGCATTTGAT
ACTGAATAACTGGAGTTTGGGAGAGGTAAGTGGAATTTCAGGTGTAGCGGTGAAATG
CGTAGATATCTGAAGGAACACCAGTGGCGAAGGCGACTTACTGGCCTAAAACTGAC
GCTGAGGCACGAAAGCATGGGTAGCAAACAGG

>ASV_0002.Allander_1_Methylomonadaceae.pLW-20.Abundance.4410

TGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCAATACCGCGTGTGTGAAG
AAGGCCTTAGGGTTGTAAAGCACTTTCAATTGGGAGGAATACCTAGCGATCAATACC
CGCTAGCTTGACATTACCTTTAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGC
GGTAATACGGAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGAGTGCGTAG
GCGGTCGTTTAAGTCAGATGTGAAAGCCCCGGGCTCAACCTGGGAACGGCATTTGA
TACTGGACGACTAGAGTTTGGGAGAGGCAAGTGGAATTTCAGGTGTAGCGGTGAAAT
GCGTAGAGATCTGAAGGAACACCAGTGGCGAAGGCGACTTGCTGGCCTAAAACTGA
CGCTGAGGCACGAAAGCATGGGTAGCAAACAGG

>ASV_0003.Allander_1_Methylomonadaceae.pLW-20.Abundance.3072

TGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCAATACCGCGTGTGTGAAG
AAGGCCTTAGGGTTGTAAAGCACTTTCAATTGGGAGGAATACCTATTGGTCAATACC
CGATAGATTGACATTACCTTTAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGC
GGTAATACGGAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGAGTGCGTAG
GCGGTTATTTAAGTCAGATGTGAAAGCCCTGGGCTTAACCTGGGAACGGCATTTGAT
ACTGAATAACTGGAGTTTGGGAGAGGTAAGTGGAATTTCAGGTGTAGCGGTGAAATG
CGTAGATATCTGAAGGAACACCAGTGGCGAAGGCGACTTACTGGCCTAAAACTGAC
GCTGAGGCACGAAAGCATGGGTAGCAAACAGG

>ASV_0004.Allander_1_Comamonadaceae.Paucibacter.Abundance.3027

TGGGGAATTTTGGACAATGGACGAAAGTCTGATCCAGCCATGCCGCGTGCAGGAAG
AAGGCCTTCGGGTTGTAAACTGCTTTTGTCAGGGAAGAAATCCTTTGAGCTAATACCT
CGGGGGGGATGACGGTACCTGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCG

CGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCA
GGCGGTTATACAAGACAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTTGT
GACTGTATAGCTAGAGTACGGTAGAGGGGGATGGAATTCCGCGTGTAGCAGTGAAA
TGCGTAGATATGCGGAGGAACACCGATGGCGAAGGCAATCCCCTGGACCTGTACTG
ACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG

>ASV_0005.Allander_1_Methylomonadaceae.pLW-20.Abundance.2717

TGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCAATACCGCGTGTGTGAAG
AAGGCCTTAGGGTTGTAAAGCACTTTCAATTGGGAGGAATACCTGTTGGTTAATACC
CGACAGATTGACATTACCTTTAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGC
GGTAATACGGAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGAGTGCGTAG
GCGGTTATTTAAGTCAGATGTGAAAGCCCTGGGCTTAACCTGGGAACGGCATTTGAT
ACTGAATAACTGGAGTTTGGGAGAGGTAAGTGGAATTTCAGGTGTAGCGGTGAAATG
CGTAGATATCTGAAGGAACACCAGTGGCGAAGGCGACTTACTGGCCTAAAACTGAC
GCTGAGGCACGAAAGCATGGGTAGCAAACAGG

>ASV_0006.Allander_1_Methylomonadaceae.pLW-20.Abundance.2286

TGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCAATACCGCGTGTGTGAAG
AAGGCCTTAGGGTTGTAAAGCACTTTCAATTGGGAGGAATACCTATCGGTCAATACC
CGGTAGATTGACATTACCTTTAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGC
GGTAATACGGAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGAGTGCGTAG
GCGGTTATTTAAGTCAGATGTGAAAGCCCTGGGCTTAACCTGGGAACGGCATTTGAT
ACTGAATAACTGGAGTTTGGGAGAGGTAAGTGGAATTTCAGGTGTAGCGGTGAAATG
CGTAGATATCTGAAGGAACACCAGTGGCGAAGGCGACTTACTGGCCTAAAACTGAC
GCTGAGGCACGAAAGCATGGGTAGCAAACAGG

>ASV_0007.Allander_1_Gallionellaceae.Gallionella.Abundance.2233

TGGGGAATTTTGGACAATGGGCGAAAGCCTGATCCAGCCATACCGCGTGAGTGAAG
AAGGCCTTCGGGTTGTAAAGCTCTTTCAGCCGGAAAGAAATCGCACTTCTTAATACG
AAGTGTGGATGACGGTACCGGAAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGC
AGGCGGTTTTGTAAGCCAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTTG
GAACTGTAAGACTAGAGTATAGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGAAA
TGCGTAGATATGTGGAGGAATACCAATGGCGAAGGCAGCCCCCTGGGTTAATACTG
ACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG

>ASV_0008.Allander_1_Comamonadaceae.Rhodoferax.Abundance.1991

TGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGCAATGCCGCGTGCAGGATG
AAGGCCTTCGGGTTGTAAACTGCTTTTGTACGGAGCGAAAAGGTTCTCTCTAATAAA
GGGGGCTCATGACGGTACCGTAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGC
AGGCGGTGATGTAAGACAGATGTGAAATCCCCGGGCTCAACCTGGGACCTGCATTT
GTGACTGCATTGCTAGAGTACGGTAGAGGGGGATGGAATTCCGCGTGTAGCAGTGA
AATGCGTAGATATGCGGAGGAACACCGATGGCGAAGGCAATCCCCTGGACCTGTAC
TGACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG

>ASV_0009.Allander_1_Gallionellaceae.Gallionella.Abundance.1881

TGGGGAATTTTGGACAATGGGCGAAAGCCTGATCCAGCCATACCGCGTGAGTGAAG
AAGGCCTTCGGGTTGTAAAGCTCTTTCAGCCGGAAAGAAATCGCGTTGGTTAATATC
CAGCGTGGATGACGGTACCGGAAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCC
GCGGTAATACG

TAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCGGTTTGA
TAAGCCAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTTGGAACTGTCAGA
CTAGAGTATAGCAGAGGGGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGA
TGTGGAGGAATACCAATGGCGAAGGCAGCCCCCTGGGTTAATACTGACGCTCATGC
ACGAAAGCGTGGGGAGCAAACAGG

>ASV_0010.Allander_1_Methylomonadaceae.pLW-20.Abundance.1682

TGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCAATACCGCGTGTGTGAAG
AAGGCCTTAGGGTTGTAAAGCACTTTCAATTGGGAGGAATACCTATTGGTCAATACC
CGATAGATTGACATTACCTTTAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGC
GGTAATACGGAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGAGTGCGTAG
GCGGTTATTTAAGTCAGATGTGAAAGCCCTGGGCTTAACCTGGGAACGGCATTTGAT
ACTGAATAACTGGAGTTTGGGAGAGGTAAGTGGAATTTCAGGTGTAGCGGTGAAATG
CGTAGATATCTGAAGGAACACCAGTGGCGAAGGCGACTTACTGGTCTAAAACTGACG
CTGAGGCACGAAAGCATGGGTAGCAAACAGG

>ASV_0011.Allander_1_Methylomonadaceae.pLW-20.Abundance.1334

>ASV_0012.Allander_1_Methylomonadaceae.pLW-20.Abundance.1225

>ASV_0013.Allander_1_Gallionellaceae.Sideroxydans.Abundance.1187

>ASV_0014.Allander_1_Methylomonadaceae.pLW-20.Abundance.1132

>ASV_0015.Allander_1_Gallionellaceae.Gallionella.Abundance.924

>ASV_0016.Allander_1_Gallionellaceae.Gallionella.Abundance.905

>ASV_0017.Allander_1_Flavobacteriaceae.Flavobacterium.Abundance.900

>ASV_0018.Allander_1_Gallionellaceae.Gallionella.Abundance.872

>ASV_0019.Allander_1_Thiotrichaceae.Thiothrix.Abundance.866

>ASV_0020.Allander_1_Gallionellaceae.Gallionella.Abundance.697

>ASV_0021.Allander_1_Comamonadaceae.Rhodoferax.Abundance.680

>ASV_0022.Allander_1_Methylomonadaceae.NA.Abundance.668

>ASV_0023.Allander_1_Gallionellaceae.Sideroxydans.Abundance.660

>ASV_0024.Allander_1_Geobacteraceae.Geobacter.Abundance.641

>ASV_0025.Allander_1_Sulfurimonadaceae.Sulfuricurvum.Abundance.636

>ASV_0026.Allander_1_Gallionellaceae.Gallionella.Abundance.614

>ASV_0027.Allander_1_Gallionellaceae.Gallionella.Abundance.606

>ASV_0028.Allander_1_Gallionellaceae.Gallionella.Abundance.558

>ASV_0029.Allander_1_NA.NA.Abundance.554

>ASV_0030.Allander_1_Gallionellaceae.Sideroxydans.Abundance.503

>ASV_0031.Allander_1_Paludibacteraceae.Paludibacter.Abundance.495

>ASV_0032.Allander_1_Gallionellaceae.Gallionella.Abundance.490

>ASV_0033.Allander_1_Gallionellaceae.NA.Abundance.463

>ASV_0034.Allander_1_Methylomonadaceae.pLW-20.Abundance.452

>ASV_0035.Allander_1_Gallionellaceae.Gallionella.Abundance.415

>ASV_0036.Allander_1_Paludibacteraceae.Paludibacter.Abundance.413

>ASV_0037.Allander_1_Methylomonadaceae.pLW-20.Abundance.408

>ASV_0038.Allander_1_Thiotrichaceae.Thiothrix.Abundance.405

>ASV_0039.Allander_1_Thiotrichaceae.Thiothrix.Abundance.399

>ASV_0040.Allander_1_KD3-93.NA.Abundance.397

>ASV_0041.Allander_1_Flavobacteriaceae.Flavobacterium.Abundance.378

>ASV_0042.Allander_1_Gallionellaceae.NA.Abundance.361

>ASV_0043.Allander_1_Myxococcaceae.P3OB-42.Abundance.356

>ASV_0044.Allander_1_Flavobacteriaceae.Flavobacterium.Abundance.338

>ASV_0045.Allander_1_Arcobacteraceae.Pseudarcobacter.Abundance.336

>ASV_0046.Allander_1_Comamonadaceae.Rhodoferax.Abundance.326

>ASV_0047.Allander_1_Methylococcaceae.NA.Abundance.323

>ASV_0048.Allander_1_Flavobacteriaceae.Flavobacterium.Abundance.317

>ASV_0049.Allander_1_Rhodocyclaceae.Sulfuritalea.Abundance.300

>ASV_0050.Allander_1_Gallionellaceae.Gallionella.Abundance.291

Allander_2

>ASV_0001.Allander_2_Gallionellaceae.Gallionella.Abundance.7385

TGGGGAATTTTTGGACAATGGGCGAAAGCCTGATCCAGCCATACCGCGTGAGTGAAG
AAGGCCTTCGGGTTGTAAAGCTCTTTCAGCCGGAAAGAAATCCTGCTTACTAATACTA
AGCGGGGATGACGGTACCGGAAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGC
AGGCGGCATGATAAGCCAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTT
GGAACTGTCAAGCTAGAGTGTAGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGA
AATGCGTAGAGATGTGGAGGAATACCGATGGCGAAGGCAGCCCCCTGGGCTAACAC
TGACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG

>ASV_0002.Allander_2_Phormidiaceae.NA.Abundance.5596

TGGGGAATTTTCCGCAATGGGCGCAAGCCTGACGGAGCAAGACCGCGTGAGGGAG
GAAGGTTCTTGGATTGTAAACCTCTTTTCTCTGGGAAGAACACAATGACGGTACCAG
AGGAATAAGCATCGGCTAACTCCGTGCCAGCAGCCGCGGTAAGACGGAGGATGCAA
GCGTTATCCGGAATGATTGGGCGTAAAGCGTCCGCAGGTGGTTTTTCAAGTCTGCTG
TTAAAGACCGGGGCTTAACTCCGGGCAGGCAGTGGAAACTGAAAGACTAGAGTATG
GTAGGGGCAGAGGGAATTCCTGGTGTAGCGGTGAAATGCGTAGAGCTCAGGAAGAA
CATCGGTGGCGAAGGCGCTCTGCTAGGCCATAACTGACACTCAGGGACGAAAGCTA
GGGGAGCGAATGGG

>ASV_0003.Allander_2_Thiotrichaceae.Thiothrix.Abundance.2319

TCGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCAATACCGCGTGTGTGAAG
AAGGCCTGCGGGTTGTAAAGCACTTTCAGTTGGGAAGATAATGACGTTACCAACAGA
AGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCG
TTAATCGGAATTACTGGGCGTAAAGCGTGCGTAGGCGGCCTGTTAAGTCAGATGTGA
AAGCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAGGCTAGAATTTAGGA
GAGGGGAGTGGAATTTCCGGTGTAGCGGTGAAATGCATAGAGATCGGAAGGAACAT
CAGTGGCGAAGGCGGCTCCCTGGACTAAAATTGACGCTGAGGCACGAAAGCGTGG
GTAGCAAACAGG

>ASV_0004.Allander_2_Gallionellaceae.Gallionella.Abundance.2319

TGGGGAATTTTGGACAATGGGCGAAAGCCTGATCCAGCCATACCGCGTGAGTGAAG
AAGGCCTTCGGGTTGTAAAGCTCTTTCAGCCGGAAAGAAATCGCATTCCTTAATACG
GGATGTGGATGACGGTACCGGAAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGC
AGGCGGTTTTGTAAGCCAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTTG
GAACTGTAAGACTAGAGTATAGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGAAA
TGCGTAGATATGTGGAGGAATACCAATGGCGAAGGCAGCCCCCTGGGTTAATACTG
ACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG

>ASV_0005.Allander_2_Gallionellaceae.Gallionella.Abundance.1436

TGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGCCATACCGCGTGAGTGAAG
AAGGCCTTCGGGTTGTAAAGCTCTTTCAGCCGGAAAGAAATCCTACTTACTAATACTA
AGTGGGGATGACGGTACCGGAAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGC
AGGCGGTTTTATAAGCCAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTTG
GAACTGTAAGACTAGAGTGTAGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGAA
ATGCGTAGAGATGTGGAGGAATACCGATGGCGAAGGCAGCCCCCTGGGCTAACACT
GACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG

>ASV_0006.Allander_2_Comamonadaceae.Paucibacter.Abundance.1330

TGGGGAATTTTGGACAATGGACGAAAGTCTGATCCAGCCATGCCGCGTGCAGGAAG
AAGGCCTTCGGGTTGTAAACTGCTTTTGTCAGGGAAGAAATCCTTTGAGCTAATACCT
CGGGGGGGATGACGGTACCTGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCA
GGCGGTTATACAAGACAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTTGT
GACTGTATAGCTAGAGTACGGTAGAGGGGGATGGAATTCCGCGTGTAGCAGTGAAA
TGCGTAGATATGCGGAGGAACACCGATGGCGAAGGCAATCCCCTGGACCTGTACTG
ACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG

>ASV_0007.Allander_2_Thiotrichaceae.Thiothrix.Abundance.1282

TCGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCAATACCGCGTGTGTGAAG
AAGGCCTGCGGGTTGTAAAGCACTTTCAGTTGGGAAGATAATGACGTTACCAACAGA
AGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCG
TTAATCGGAATTACTGGGCGTAAAGCGTGCGTAGGCGGTTTTTTAAGTCAGATGTGA
AATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGGAGACTAGAGTGTGGGA
GAGGAGAGTGGAATTTCCGGTGTAGCGGTGAAATGCATAGAGATCGGAAGGAACAT
CAGTGGCGAAGGCGACTCTCTGGACCAACACTGACGCTGAGGCACGAAAGCGTGG
GTAGCAAACAGG

>ASV_0008.Allander_2_Thiotrichaceae.Thiothrix.Abundance.1205

TCGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCAATACCGCGTGTGTGAAG
AAGGCCTGCGGGTTGTAAAGCACTTTCAGTTGGGAAGATAATGACGTTACCAACAGA
AGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCG
TTAATCGGAATTACTGGGCGTAAAGCGTGCGTAGGCGGTTATTTAAGTCAGATGTGA
AATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGGTAACTAGAGTGTGGGA
GAGGAGAGTGGAATTTCCGGTGTAGCGGTGAAATGCATAGAGATCGGAAGGAACAC

CAGTGGCGAAGGCGGCTCTCTGGACCAACACTGACGCTGAGGCACGAAAGCGTGG
GTAGCAAACAGG

>ASV_0009.Allander_2_Sulfurimonadaceae.Sulfuricurvum.Abundance.980

TGAGGAATATTGCACAATGGAGGAAACTCTGATGCAGCAACGCCGCGTGGAGGATG
ACGCATTTCGGTGTGTAAACTCCTTTTATTAGGGAAGATAATGACGGTACCTAATGAA
TAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTT
ACTCGGAATCACTGGGCGTAAAGGGTGCGTAGGCTGGCTTCTAAGTCAGATGTGAA
ATCCAATGGCTTAACCATTGAACTGCATTTGAAACTGGGAGCCTAGAGTTCAGAAGG
GGCAGATGGAATTAGTGGTGTAGGGGTAAAATCCGTAGATATCACTAGGAATATCAA
AAGCGAAGGCGATCTGCTGGGATGATACTGACGCTGAGGCACGAAAGCGTGGGGA
GCAAACAGG

>ASV_0010.Allander_2_Sulfuricellaceae.Sulfuricella.Abundance.921

TGGGGAATTTTGGACAATGGGGGGAAACCCTGATCCAGCCATTCCGCGTGAGTGAAG
AAGGCCTTCGGGTTGTAAAGCTCTTTCAGGAGGGAAGAAACGGTTGCAGCTAATACC
TGCGACTAATGACGGTACCTCAAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGC
AGGCGGTTTTGTAAGTCAGATGTGAAAGCCCCGGGCTCAACCTGGGAACTGCGTTT
GAAACTACAAGGCTAGAGTGTAGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGA
AATGCGTAGAGATGTGGAGGAATACCGATGGCGAAGGCAGCCCCCTGGGTTAACAC
TGACGCTCATGTACGAAAGCGTGGGGAGCAAACAGG

>ASV_0011.Allander_2_Gallionellaceae.Gallionella.Abundance.845

>ASV_0012.Allander_2_Gallionellaceae.NA.Abundance.816

>ASV_0013.Allander_2_Thiotrichaceae.Thiothrix.Abundance.732

>ASV_0014.Allander_2_Gallionellaceae.Gallionella.Abundance.669

>ASV_0015.Allander_2_Gallionellaceae.Gallionella.Abundance.650

>ASV_0016.Allander_2_Gallionellaceae.Sideroxydans.Abundance.611

>ASV_0017.Allander_2_Thiotrichaceae.Thiothrix.Abundance.574

>ASV_0018.Allander_2_Gallionellaceae.Candidatus Nitrotoga.Abundance.559

>ASV_0019.Allander_2_Gallionellaceae.NA.Abundance.557

>ASV_0020.Allander_2_Sulfuricellaceae.Sulfuricella.Abundance.552

>ASV_0021.Allander_2_Thiotrichaceae.Thiothrix.Abundance.542

>ASV_0022.Allander_2_Gallionellaceae.Gallionella.Abundance.480

>ASV_0023.Allander_2_Phormidiaceae.NA.Abundance.461

>ASV_0024.Allander_2_Gallionellaceae.Sideroxydans.Abundance.449

>ASV_0025.Allander_2_NA.NA.Abundance.448

>ASV_0026.Allander_2_Thiotrichaceae.Thiothrix.Abundance.435

>ASV_0027.Allander_2_Gallionellaceae.Gallionella.Abundance.385

>ASV_0028.Allander_2_Gallionellaceae.Gallionella.Abundance.385

>ASV_0029.Allander_2_Rhodocyclaceae.Sulfuritalea.Abundance.385

>ASV_0030.Allander_2_NA.NA.Abundance.372

>ASV_0031.Allander_2_Phormidiaceae.NA.Abundance.369

>ASV_0032.Allander_2_Gallionellaceae.Gallionella.Abundance.352

>ASV_0033.Allander_2_Hydrogenophilaceae.NA.Abundance.337

>ASV_0034.Allander_2_NA.NA.Abundance.319

>ASV_0035.Allander_2_NA.NA.Abundance.313

>ASV_0036.Allander_2_Gallionellaceae.Gallionella.Abundance.287

>ASV_0037.Allander_2_Comamonadaceae.Rhodoferax.Abundance.286

>ASV_0038.Allander_2_Thiotrichaceae.Thiothrix.Abundance.266

>ASV_0039.Allander_2_Gallionellaceae.Gallionella.Abundance.263

>ASV_0040.Allander_2_NA.NA.Abundance.261

>ASV_0041.Allander_2_NA.NA.Abundance.253

>ASV_0042.Allander_2_NA.NA.Abundance.238

>ASV_0043.Allander_2_Geobacteraceae.Geobacter.Abundance.238

>ASV_0044.Allander_2_Sulfurimonadaceae.Sulfuricurvum.Abundance.235

>ASV_0045.Allander_2_Comamonadaceae.Rhodoferax.Abundance.217

>ASV_0046.Allander_2_Desulfobulbaceae.NA.Abundance.214

>ASV_0047.Allander_2_Gallionellaceae.Gallionella.Abundance.213

>ASV_0048.Allander_2_Thiotrichaceae.Thiothrix.Abundance.203

>ASV_0049.Allander_2_Phormidiaceae.NA.Abundance.197

>ASV_0050.Allander_2_Gallionellaceae.Gallionella.Abundance.174

>ASV_0051.Allander_2_Sulfurimonadaceae.Sulfurimonas.Abundance.160

Allander_3

>ASV_0001.Allander_3_Gallionellaceae.Gallionella.Abundance.3245

```
TGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGCCATACCGCGTGAGTGAAG
AAGGCCTTCGGGTTGTAAAGCTCTTTCAGCCGGAAAGAAATCGCGTTTATTAATACTA
AGCGTGGATGACGGTACCGGAAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGC
AGGCCGTTTTCTAAGACAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTTG
TGACTGGAAGACTAGAGTGTAGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGAA
ATGCGTAGAGATGTGGAGGAATACCGATGGCGAAGGCAGCCCCCTGGGCTAACACT
GACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG
```

>ASV_0002.Allander_3_Comamonadaceae.Rhodoferax.Abundance.782

```
TGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGCAATGCCGCGTGCAGGATG
AAGGCCTTCGGGTTGTAAACTGCTTTTGTACGGAACGAAACGGTCCTTTCTAATATAA
GGGGCTAATGACGGTACCGTAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCA
GGCGGTTATATAAGACAGATGTGAAATCCCCGGGCTCAACCTGGGACCTGCATTTGT
GACTGTATAGCTAGAGTACGGTAGAGGGGGATGGAATTCCGCGTGTAGCAGTGAAA
TGCGTAGATATGCGGAGGAACACCGATGGCGAAGGCAATCCCCTGGACCTGTACTG
ACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG
```

>ASV_0003.Allander_3_Comamonadaceae.Rhodoferax.Abundance.663

```
TGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGCAATGCCGCGTGCAGGATG
AAGGCCTTCGGGTTGTAAACTGCTTTTGTACGGAACGAAACGGTGAGCTCTAATAAA
GTTTGCTAATGACGGTACCGTAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCA
GGCGGTTATATAAGACAGATGTGAAATCCCCGGGCTCAACCTGGGACCTGCATTTGT
GACTGTATAGCTAGAGTACGGTAGAGGGGGATGGAATTCCGCGTGTAGCAGTGAAA
TGCGTAGATATGCGGAGGAAC
```

ACCGATGGCGAAGGCAATCCCCTGGACCTGTACTGACGCTCATGCACGAAAGCGTG
GGGAGCAAACAGG

>ASV_0004.Allander_3_Gallionellaceae.Sideroxydans.Abundance.657

TGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGAAG
AAGGCCTTCGGGTTGTAAAGCTCTTTCAGCTGGAAAAAAATCGCTTCACCTAATACGT
GGAGTGGATGATGGTACCAGCAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCA
GGCGGTTTTGTAAGACAGACGTGAAATCCCCGGGCTTAACCTGGGAACTGCGTTTGT
GACTGCAAGGCTAGAGTATGGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGAAA
TGCGTAGATATGTGGAGGAATACCGATGGCGAAGGCAGCCCCCTGGGTCAATACTG
ACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG

>ASV_0005.Allander_3_Gallionellaceae.Gallionella.Abundance.564

TGGGGAATTTTGGACAATGGGCGAAAGCCTGATCCAGCCATACCGCGTGAGTGAAG
AAGGCCTTCGGGTTGTAAAGCTCTTTCAGACGGAAAGAAAAGGTTTCGGATAACACC
TGGAACTGATGACGGTACCGTAAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGC
AGGCGGTTTTTTAAGCCAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTTG
GAACTGGAAGACTAGAGTATAGCAGAGGGAGGTAGAATTCCACGTGTAGCAGTGAA
ATGCGTAGATATGTGGAGGAATACCGATGGCGAAGGCAGCCTCCTGGGTTAATACT
GACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG

>ASV_0006.Allander_3_Gallionellaceae.Sideroxydans.Abundance.564

TGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGAAG
AAGGCCTTCGGGTTGTAAAGCTCTTTCAGCTGGAAAAAAATCGTACTTATTAATACTA
AGTATGGATGATGGTACCAGCAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCA
GGCGGTTTTGTAAGACAGACGTGAAATCCCCGGGCTTAACCTGGGAACTGCGTTTGT
GACTGCAAGGCTAGAGTATGGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGAAA
TGCGTAGATATGTGGAGGAATACCGATGGCGAAGGCAGCCCCCTGGGTCAATACTG
ACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG

>ASV_0007.Allander_3_Comamonadaceae.NA.Abundance.538

TGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGCCATACCGCGTGAGTGAAG
AAGGCCTTCGGGTTGTAAAGCTCTTTCAGCCGGAAAGAAATCGCGTTTATTAATACTA
AGCGTGGATGACGGTACCGGAAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGC
AGGCCGGTTATATAAGACAGATGTGAAATCCCCGGGCTCAACCTGGGACCTGCATTT
GTGACTGTATAGCTAGAGTACGGTAGAGGGGGATGGAATTCCGCGTGTAGCAGTGA
AATGCGTAGATATGCGGAGGAACACCGATGGCGAAGGCAATCCCCTGGACCTGTAC
TGACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG

>ASV_0008.Allander_3_Gallionellaceae.Gallionella.Abundance.534

TGGGGAATTTTGGACAATGGGCGAAAGCCTGATCCAGCCATACCGCGTGAGTGAAG
AAGGCCTTCGGGTTGTAAAGCTCTTTCAGACGGAAAGAAAAGGTTTCGGATAATACC
TGGAACTGATGACGGTACCGTAAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGC
AGGCGGTTTTTTAAGCCAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTTG
GAACTGGAAGACTAGAGTATAGCAGAGGGAGGTAGAATTCCACGTGTAGCAGTGAA
ATGCGTAGATATGTGGAGGAATACCGATGGCGAAGGCAGCCTCCTGGGTTAATACT
GACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG

>ASV_0009.Allander_3_Comamonadaceae.Rhodoferax.Abundance.513

TGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGCAATGCCGCGTGCAGGATG
AAGGCCTTCGGGTTGTAAACTGCTTTTGTACGGAACGAAACGGCGAGCTCTAATAAA
GCTTGCTAATGACGGTACCGTAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGC
AGGCGGTTATATAAGACAGATGTGAAATCCCCGGGCTCAACCTGGGACCTGCATTTG
TGACTGTATAGCTAGAGTACGGTAGAGGGGGATGGAATTCCGCGTGTAGCAGTGAA
ATGCGTAGATATGCGGAGGAACACCGATGGCGAAGGCAATCCCCTGGACCTGTACT
GACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG

>ASV_0010.Allander_3_Comamonadaceae.Paucibacter.Abundance.508

TGGGGAATTTTGGACAATGGACGAAAGTCTGATCCAGCCATGCCGCGTGCAGGAAG
AAGGCCTTCGGGTTGTAAACTGCTTTTGTCAGGGAAGAAATCCTTTGAGCTAATACCT
CGGGGG

GATGACGGTACCTGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAA
TACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCGGT
TATACAAGACAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTTGTGACTGT
ATAGCTAGAGTACGGTAGAGGGGGATGGAATTCCGCGTGTAGCAGTGAAATGCGTA
GATATGCGGAGGAACACCGATGGCGAAGGCAATCCCCTGGACCTGTACTGACGCTC
ATGCACGAAAGCGTGGGGAGCAAACAGG

>ASV_0011.Allander_3_Comamonadaceae.Rhodoferax.Abundance.497

>ASV_0012.Allander_3_Gallionellaceae.Gallionella.Abundance.479

>ASV_0013.Allander_3_Comamonadaceae.Sphaerotilus.Abundance.463

>ASV_0014.Allander_3_Comamonadaceae.Rhodoferax.Abundance.451

>ASV_0015.Allander_3_Comamonadaceae.Leptothrix.Abundance.443

>ASV_0016.Allander_3_NA.NA.Abundance.427

>ASV_0017.Allander_3_NA.NA.Abundance.393

>ASV_0018.Allander_3_Comamonadaceae.Rhodoferax.Abundance.386

>ASV_0019.Allander_3_Comamonadaceae.Rhodoferax.Abundance.372

>ASV_0020.Allander_3_Spirosomaceae.Lacihabitans.Abundance.353

>ASV_0021.Allander_3_Gallionellaceae.Gallionella.Abundance.338

>ASV_0022.Allander_3_Flavobacteriaceae.Flavobacterium.Abundance.320

>ASV_0023.Allander_3_Rhodocyclaceae.Uliginosibacterium.Abundance.313

>ASV_0024.Allander_3_Comamonadaceae.Rhodoferax.Abundance.289

>ASV_0025.Allander_3_NA.NA.Abundance.288

>ASV_0026.Allander_3_Chitinophagaceae.Ferruginibacter.Abundance.284

>ASV_0027.Allander_3_Comamonadaceae.Sphaerotilus.Abundance.276

>ASV_0028.Allander_3_Flavobacteriaceae.Flavobacterium.Abundance.271

>ASV_0029.Allander_3_Cellvibrionaceae.Cellvibrio.Abundance.271

>ASV_0030.Allander_3_Xanthobacteraceae.Bradyrhizobium.Abundance.267

>ASV_0031.Allander_3_Sphingomonadaceae.Novosphingobium.Abundance.261

>ASV_0032.Allander_3_Methylomonadaceae.Crenothrix.Abundance.257

>ASV_0033.Allander_3_Comamonadaceae.Rhodoferax.Abundance.257

>ASV_0034.Allander_3_Mitochondria.NA.Abundance.249

>ASV_0035.Allander_3_Comamonadaceae.Rhodoferax.Abundance.234

>ASV_0036.Allander_3_Rhodobacteraceae.Rhodobacter.Abundance.230

>ASV_0037.Allander_3_Comamonadaceae.Rhodoferax.Abundance.228

>ASV_0038.Allander_3_Flavobacteriaceae.Flavobacterium.Abundance.223

>ASV_0039.Allander_3_Comamonadaceae.Rhodoferax.Abundance.218

>ASV_0040.Allander_3_Rhodocyclaceae.Ferribacterium.Abundance.210

>ASV_0041.Allander_3_Cellvibrionaceae.Cellvibrio.Abundance.209

>ASV_0042.Allander_3_Hyphomonadaceae.Hirschia.Abundance.205

>ASV_0043.Allander_3_Spirosomaceae.Flectobacillus.Abundance.201

>ASV_0044.Allander_3_Comamonadaceae.Rhodoferax.Abundance.201

>ASV_0045.Allander_3_Chitinophagaceae.Ferruginibacter.Abundance.200

>ASV_0046.Allander_3_Gaiellaceae.Gaiella.Abundance.198

>ASV_0047.Allander_3_Comamonadaceae.Rhodoferax.Abundance.192

>ASV_0048.Allander_3_NA.NA.Abundance.192

>ASV_0049.Allander_3_Sphingomonadaceae.Novosphingobium.Abundance.190

>ASV_0050.Allander_3_Flavobacteriaceae.Flavobacterium.Abundance.189

Appendix IV - Full List of Metabolites Identified During Orbitrap Mass Spectrometry Experiment

This appendix contains all 902 putative compounds that were assigned individually to the Allander and Dawsholm samples as well as to both samples.

Sample Site	Putative metabolite	Formula	Exact Mass	Isomers
AL	Hordatine B	C29H40N8O5	290.15595	1
AL	tricarballylate	C6H8O6	176.03216	15
AL	1,3-Diaminopropane	C3H10N2	74.084422	2
AL	[SP (14:0)] 1-deoxy-tetradecasphing-13-enine	C14H29NO	227.22497	2
AL	Dimethyl citraconate	C7H10O4	158.05795	5
AL	[FA methyl(14:0)] 12-methyl-tetradecanoic acid	C15H30O2	242.22479	9
AL	Pentanoate	C5H10O2	102.0681	5
AL	pyrrolidine	C4H9N	71.073527	2
AL	minoxidil	C9H15N5O	209.12768	1
AL	di-n-Undecylamine	C22H47N	325.37072	1
AL	Trimethylaminoacetone	C6H13NO	115.09974	2
AL	Met-Gly-Ser	C10H19N3O5S	293.10412	2
AL	Putrescine	C4H12N2	88.100031	1
AL	Glycine	C2H5NO2	75.032041	3
AL	Tetraethylammonium	C8H19N	129.1518	2

AL	1-butanesulfonate	C4H10O3S	138.035 09	1
AL	Pinidine	C9H17N	139.136 11	1
AL	2-Oxoglutarate	C5H6O5	146.021 58	7
AL	L-Dehydroascorbate	C6H6O6	174.016 49	3
AL	Asp-Pro-Arg	C15H26N6O6	193.095 19	1
AL	3-Nitrotyrosine	C9H10N2O5	226.058 94	1
AL	carbidopa	C10H14N2O4	226.095 81	2
AL	Tributyl phosphate	C12H27O4P	266.164 68	1
AL	Adenosine	C10H13N5O4	267.096 51	3
AL	phenyl-1-thio-β-D-galactopyranoside	C12H16O5S	272.071 98	1
AL	indole-3-acetyl-proline	C15H16N2O3	272.115 77	1
AL	Prebetanin	C24H26N2O1 6S	315.050 97	1
AL	C25-Allenic-apo-aldehyde	C25H34O3	382.250 62	3
AL	[ST hydroxy,methyl(4:0)] (22E)-(8S)-3-hydroxy-22-methyl-9,10-seco-1,3,5(10),22-cholestatetraen-9-one	C28H42O2	410.318 62	3
AL	Cucurbitacin J	C30H44O8	532.303 79	2
AL'DA WS	[PR] Etretinate	C23H30O3	354.219 58	1
AL'DA WS	3,3',4'5-Tetrahydroxystilbene	C14H12O4	244.073 6	7
AL'DA WS	[FA (11:2)] 4,10-undecadiynal	C11H14O	162.104 55	1
AL'DA WS	Asp-Cys-Gly	C9H15N3O6S	293.068 81	1
AL'DA WS	Pisatin	C17H14O6	314.079 93	16

AL'DA WS	2-amino-4-methylphenol	C7H9NO	123.068 47	4
AL'DA WS	[PK] Visnagin	C13H10O4	230.057 76	7
AL'DA WS	2-Phenylacetamide	C8H9NO	135.068 47	4
AL'DA WS	D-Cathinone	C9H11NO	149.084 12	1
AL'DA WS	Pseudopelletierine	C9H15NO	153.115 38	1
AL'DA WS	SR 12813	C24H42O7P2	252.120 28	1
AL'DA WS	4-Methylumbelliferyl acetate	C12H10O4	218.057 89	5
AL'DA WS	[FA methyl(18:2)] methyl 9,10-epoxy-12,15-octadecadienoate	C19H32O3	308.235 39	3
AL'DA WS	Farrerol	C17H16O5	300.100 11	30
AL'DA WS	(R)-2-Methylimino-1-phenylpropan-1-ol	C10H13NO	163.099 79	3
AL'DA WS	Homoferreirin	C17H16O6	158.047 38	21
AL'DA WS	Mearsine	C9H13NO	151.099 74	8
AL'DA WS	Phenylethanolamine	C8H11NO	137.084 1	7
AL'DA WS	Phyllanthin	C24H34O6	209.117 89	3
AL'DA WS	Delphinidin	C15H10O7	302.042 99	15
AL'DA WS	UDP-3-ketoglucose	C15H22N2O1 7P2	282.019 99	1
AL'DA WS	Indolepyruvate	C11H9NO3	203.058 36	2
AL'DA WS	Hordenine	C10H15NO	165.115 42	5
AL'DA WS	Kynurenate	C10H7NO3	189.042 59	6
AL'DA WS	[Fv hydroxy,methox] 2',6'-Dihydroxy-4'-methoxydihydrochalcone	C16H16O4	272.104 97	14

AL'DA WS	Eriodictyol	C15H12O6	288.063 69	26
AL'DA WS	Hygrine	C8H15NO	141.115 44	9
AL'DA WS	Tropinone	C8H13NO	139.099 73	2
AL'DA WS	Tremetone	C13H14O2	202.099 37	1
AL'DA WS	4,4'-Sulfonyldiphenol	C12H10O4S	250.030 12	1
AL'DA WS	Phenmetrazine	C11H15NO	177.115 4	2
AL'DA WS	[FA (12:5)] 3,5,7,9,11-dodecapentaenoic acid	C12H14O2	190.099 5	6
AL'DA WS	Indoleacrylicacid	C11H9NO2	187.063 37	1
AL'DA WS	Carpacin	C11H12O3	192.078 82	6
AL'DA WS	Peonidin	C16H12O6	300.063 89	26
AL'DA WS	Metazocine	C15H21NO	231.162 35	1
AL'DA WS	[Fv] Luteolin	C15H10O6	286.047 27	16
AL'DA WS	Coniferyl aldehyde	C10H10O3	178.063 09	11
AL'DA WS	3-Methylindolepyruvate	C12H11NO3	217.073 9	2
AL'DA WS	4-trans-(N,N-dimethylamino)cinnamaldehyde	C11H13NO	175.099 76	1
AL'DA WS	Methyl cinnamate	C10H10O2	162.068 08	11
AL'DA WS	Glu-Met-Thr-Asp	C18H30N4O1 0S	247.084 4	2
AL'DA WS	Norlichexanthone	C14H10O5	258.052 79	5
AL'DA WS	Robustine	C12H9NO3	215.058 27	1
AL'DA WS	N-Methylpelletierine	C9H17NO	155.131 08	1

AL'DA WS	N-Hydroxy-4-aminobiphenyl	C12H11NO	185.0841	2
AL'DA WS	1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridine N-oxide	C12H15NO	189.11537	1
AL'DA WS	D-Phenylalanine	C9H11NO2	165.07914	7
AL'DA WS	[FA (14:5)] 5,7,9,11,13-tetradecapentaenoic acid	C14H18O2	218.13081	1
AL'DA WS	ferroxamine	C25H45FeN6O8	306.12919	1
AL'DA WS	Nicosulfuron	C15H18N6O6S	410.10159	1
AL'DA WS	Benzyl 2-methyl-3-oxobutanoate	C12H14O3	206.09439	4
AL'DA WS	[FA hydroxy(20:3)] 8,9-dihydroxy-5Z,11Z,14Z-eicosatrienoic acid	C20H34O4	338.24608	11
AL'DA WS	7-Ethoxycoumarin	C11H10O3	190.06309	3
AL'DA WS	10-Deoxymethynolide	C17H28O4	296.19891	2
AL'DA WS	Dihydromonacolin L	C19H30O3	306.21947	12
AL'DA WS	Costunolide	C15H20O2	232.14648	10
AL'DA WS	Darlingine	C13H17NO2	219.12597	1
AL'DA WS	[FA hydroxy(12:1)] 12-hydroxy-10-dodecenoic acid	C12H22O3	214.15717	10
AL'DA WS	N-Benzoyl-4-hydroxyanthranilate	C14H11NO4	257.06867	1
AL'DA WS	[PR] Citronellyl acetate	C12H22O2	198.16213	17
AL'DA WS	(-)-Salsolinol	C10H13NO2	179.09462	5
AL'DA WS	Precocene 2	C13H16O3	220.11004	3
AL'DA WS	[FA (14:2)] 5,8-tetradecadienoic acid	C14H24O2	224.17776	20
AL'DA WS	Leiokinine A	C14H17NO2	231.12596	2

AL'DA WS	[FA (11:1)] 10-undecenoic acid	C11H20O2	184.146 42	17
AL'DA WS	Glu-Phe-Trp	C25H28N4O6	240.100 03	1
AL'DA WS	[PR] alpha-Santonin	C15H18O3	246.125 6	11
AL'DA WS	5,6-Dihydroxyindole	C8H7NO2	149.047 74	8
AL'DA WS	[FA methyl(12:1)] 2-methyl-2-dodecenoic acid	C13H24O2	212.177 84	6
AL'DA WS	(1S,2R,4S)-(-)-Bornyl acetate	C12H20O2	196.146 42	29
AL'DA WS	[FA (11:2)] 2E,4E-undecadienoic acid	C11H18O2	182.130 81	12
AL'DA WS	(-)-Arctigenin	C21H24O6	372.157 66	4
AL'DA WS	Acronycidine	C15H15NO5	289.094 82	1
AL'DA WS	2-butyl-4-hydroxy-5-methyl-3(2H)-furanone	C9H14O3	170.094 41	2
AL'DA WS	Peonidin	C16H12O6	300.063 25	26
AL'DA WS	Fabianine	C14H21NO	219.162 35	2
AL'DA WS	[FA oxo,hydroxy(2:0)] 9-oxo-11R,15S-dihydroxy-5Z,13E-prostadienoic acid	C20H32O5	352.225 19	32
AL'DA WS	Coniferyl alcohol	C10H12O3	180.078 76	7
AL'DA WS	[FA (13:2)] 3E,5E-tridecadienoic acid	C13H22O2	210.162 13	12
AL'DA WS	Kokusaginine	C14H13NO4	259.084 34	3
AL'DA WS	Benzoyldehydro-2,3-dihydroxy-benzone	C13H13NO3	231.089 54	1
AL'DA WS	5-Hydroxytryptophol	C10H11NO2	177.079	3
AL'DA WS	Haplopine	C13H11NO4	245.068 73	1
AL'DA WS	19-Oxoandrost-4-ene-3,17-dione	C19H24O3	300.172 5	13

AL'DA WS	12-trans-Hydroxy juvenile hormone III	C16H26O4	282.183 13	1
AL'DA WS	Indole-3-ethanol	C10H11NO	161.084 08	2
AL'DA WS	Kavapyrone	C14H16O3	232.109 96	3
AL'DA WS	[FA oxo(5:2/5:0/4:0)] (1S,2S)-3-oxo-2-pentyl-cyclopentanebutanoic acid	C14H24O3	240.172 71	2
AL'DA WS	5-Hydroxyindoleacetate	C10H9NO3	191.058 24	5
AL'DA WS	[FA (12:4)] 2E,4E,8Z,10E-dodecatetraenoic acid	C12H16O2	192.115 17	5
AL'DA WS	Bellendine	C12H15NO2	205.110 41	1
AL'DA WS	[FA (15:0)] 3-pentadecynoic acid	C15H26O2	238.193 41	9
AL'DA WS	(+/-)-6-Hydroxy-3-oxo-alpha-ionone	C13H18O3	222.125 68	2
AL'DA WS	hispidin	C13H10O5	246.052 64	4
AL'DA WS	3-Methyloxindole	C9H9NO	147.068 42	3
AL'DA WS	Benzoyltropein	C15H19NO2	245.141 4	1
AL'DA WS	Scoparone	C11H10O4	206.057 97	6
AL'DA WS	6-Prenylnaringenin	C20H20O5	340.131 62	33
AL'DA WS	1,4-dicarboxynaphthalene	C12H8O4	216.042 4	4
AL'DA WS	Oblongolide	C14H20O2	220.146 5	2
AL'DA WS	Bufuralol	C16H23NO2	261.172 78	5
AL'DA WS	1,3-Dimethyl-8-isoquinolinol	C11H11NO	173.084 16	2
AL'DA WS	Cys-Cys-Cys	C9H17N3O4S 3	327.037 77	1
AL'DA WS	[FA amino(16:0)] 2R-aminohexadecanoic acid	C16H33NO2	271.251 04	4

AL'DA WS	Dopamine	C8H11NO2	153.078 99	5
AL'DA WS	5-Methoxytryptophol	C11H13NO2	191.094 63	1
AL'DA WS	5-O-Methylembelin	C18H28O4	308.199 08	1
AL'DA WS	[Fv] 4-Methoxylonchocarpin	C21H20O4	336.136 32	13
AL'DA WS	[FA] Methyl jasmonate	C13H20O3	224.141 42	6
AL'DA WS	2,6-Dioxo-6-phenylhexanoate	C12H12O4	220.073 67	2
AL'DA WS	[6]-Gingerol	C17H26O4	294.183 58	3
AL'DA WS	[FA hydroxy,oxo(7:0/2:0)] 4-hydroxy-2-oxo-Heptanedioic acid	C7H10O6	190.047 89	4
AL'DA WS	4,4'-Dihydroxy-3,5-dimethoxydihydrostilbene	C16H18O4	274.120 47	2
AL'DA WS	Angustibalin	C17H20O5	304.131 32	3
AL'DA WS	[Fv] Kaempferol	C15H10O6	286.048 12	16
AL'DA WS	[FA oxo(12:3)] 12-oxo-5E,8E,10Z-dodecatrienoic acid	C12H16O3	208.110 09	7
AL'DA WS	Columbianetin	C14H14O4	246.089 2	4
AL'DA WS	[PR] Juvenile hormone III	C16H26O3	266.188 31	7
AL'DA WS	Riccionidin A	C15H8O6	284.032 39	3
AL'DA WS	Octadecanamide	C18H37NO	283.287 38	1
AL'DA WS	N-Acetylvanilalanine	C12H15NO5	253.094 98	1
AL'DA WS	2,2-Dimethyl-8-prenylchromene 6-carboxylic acid	C17H20O3	272.141 32	4
AL'DA WS	pyridine 3-aldehyde adenine dinucleotide	C22H26N6O15P2	338.046 94	1
AL'DA WS	7-Ethoxycoumarin	C11H10O3	190.062 98	3

AL'DA WS	[Fv] O-Methylodoratol	C18H20O5	316.131 39	8
AL'DA WS	4-Heptyloxybenzoic acid	C14H20O3	236.141 35	2
AL'DA WS	Arborinine	C16H15NO4	285.099 96	1
AL'DA WS	[FA hydroxy(9:1)] 4-hydroxy-2-nonenal	C9H16O2	156.115 12	13
AL'DA WS	demethylsuberosin	C14H14O3	230.094 18	9
AL'DA WS	2-Heptyl-4-hydroxyquinoline-N-oxide	C16H21NO2	259.156 96	4
AL'DA WS	Fagaramide	C14H17NO3	247.120 77	1
AL'DA WS	[FA hydroxy(18:0/2:0)] 8-hydroxy-13,17-octadecadiene-9,11-diynoic acid	C18H24O3	288.172 58	9
AL'DA WS	3-Methoxy-4-hydroxyphenylacetaldehyde	C9H10O3	166.063 02	17
AL'DA WS	Eugenol methyl ether	C11H14O2	178.099 49	5
AL'DA WS	Maculine	C13H9NO4	243.053 07	1
AL'DA WS	(10S)-Juvenile hormone III acid diol	C15H26O4	270.183 23	1
AL'DA WS	7,8-Dihydroxykynurenate	C10H7NO5	221.032 62	1
AL'DA WS	[FA oxo(18:3)] 4-oxo-9Z,11E,13E-octadecatrienoic acid	C18H28O3	292.204 2	15
AL'DA WS	Leptodactylone	C11H10O5	222.052 97	6
AL'DA WS	1,4-Dihydroxy-2-naphthoate	C11H8O4	204.042 39	5
AL'DA WS	(+-)-Carnegine	C13H19NO2	221.141 63	2
AL'DA WS	[Fv] Glabrachalcone	C23H24O6	396.158 29	10
AL'DA WS	Zearalenone	C18H22O5	318.146 93	1
AL'DA WS	Pralidoxime	C7H8N2O	136.063 68	3

AL'DA WS	Podolide	C19H22O5	330.147 02	4
AL'DA WS	Indolebutyric acid	C12H13NO2	203.094 75	2
AL'DA WS	6-Hydroxytremetone	C13H14O3	218.094 26	3
AL'DA WS	Anisomycin	C14H19NO4	265.131 34	2
AL'DA WS	3-Amino-3-(4-hydroxyphenyl)propanoate	C9H11NO3	181.074 02	11
AL'DA WS	Ibuprofenacylglucuronide	C19H26O8	382.163 05	1
AL'DA WS	15-Keto-prostaglandinE2	C20H30O5	350.209 64	18
AL'DA WS	11-Ketoetiocholanolone	C19H28O3	304.204 14	25
AL'DA WS	Didrovaltratum	C22H32O8	212.105 03	2
AL'DA WS	L-1-Pyrroline-3-hydroxy-5-carboxylate	C5H7NO3	129.042 65	6
AL'DA WS	[ST Dioxo(3:0)] 3,12-Dioxochola-1,4,9(11)-trien-24-oic Acid	C24H30O4	382.214 6	2
AL'DA WS	sodium dodecyl sulfate	C12H26O4S	266.155 34	1
AL'DA WS	2,3-dioctanoylglyceramide	C19H35NO5	357.251 34	1
AL'DA WS	[ST hydroxy(3:0)] 3,16alpha-dihydroxy-estra-1,3,5(10)-trien-17-one	C18H22O3	286.156 56	7
AL'DA WS	Xanthurenic acid	C10H7NO4	205.037 55	2
AL'DA WS	Trans-2-3-4-Trimethoxycinnamate	C12H14O5	238.084 29	2
AL'DA WS	[FA hydroxy(17:2)] 7-hydroxy-10E,16-heptadecadien-8-ynoic acid	C17H26O3	278.188 17	4
AL'DA WS	Epinephrineglucuronide	C15H21NO9	359.121 42	1
AL'DA WS	O-methylandrocymbine	C22H27NO5	385.189 33	1
AL'DA WS	L-Tyrosine methyl ester	C10H13NO3	195.089 65	6

AL'DA WS	Lophophorine	C13H17NO3	235.120 79	1
AL'DA WS	dihydroartemisinic acid hydroperoxide	C15H24O4	268.167 58	1
AL'DA WS	3-Methoxy-4-hydroxyphenylglycolglucoronide	C15H20O10	360.105 99	1
AL'DA WS	2-Methoxyestradiol-17beta	C19H26O3	302.188 14	16
AL'DA WS	Shiromodiol diacetate	C19H30O5	338.209 88	1
AL'DA WS	[FA oxo(8:1)] 5-oxo-7-octenoic acid	C8H12O3	156.078 74	2
AL'DA WS	Aristolochic acid	C17H11NO7	341.053 43	1
AL'DA WS	(9Z)-Tetradecenoic acid	C14H26O2	226.193 45	12
AL'DA WS	Scopolamine	C17H21NO4	303.147 12	5
AL'DA WS	Acetylpseudotropine	C10H17NO2	183.126	1
AL'DA WS	Suberic acid	C8H14O4	174.089 26	3
AL'DA WS	[FA methyl,hydroxy,oxo(5:2/8:0)] methyl 8-[2-(2-formyl-vinyl)-3-hydroxy-5-oxo-cyclopentyl]-octanoate	C17H26O5	310.178 25	2
AL'DA WS	[FA oxo(5:1/5:0/4:0)] (1R,2R)-3-oxo-2-(2'Z-pentenyl)-cyclopentanebutanoic acid	C14H22O3	238.157 09	3
AL'DA WS	[Fv Trihydroxy,trimethy] 2',4',4''-Trihydroxy-3',6'',6''-trimethylpyrano[2'',3'':6',5']dihydrochalcone	C21H24O5	356.162 96	11
AL'DA WS	(3R)-3-Isopropenyl-6-oxoheptanoate	C10H15O3	183.102 14	2
AL'DA WS	hexane-1,3,4,6-tetracarboxylate	C10H14O8	262.068 87	1
AL'DA WS	Arctiopicrin	C19H26O6	350.173 24	3
AL'DA WS	Prostaglandin A2	C20H30O4	334.214 87	32
AL'DA WS	1-(4-Methoxyphenyl)-3-(4-morpholinyl)-1-propanone	C14H19NO3	249.136 39	2
AL'DA WS	10-Oxodecanoate	C10H18O3	186.125 71	19

AL'DA WS	4-(2-Aminophenyl)-2,4-dioxobutanoate	C10H9NO4	207.053 23	6
AL'DA WS	3"-hydroxy-geranylhydroquinone	C16H22O3	262.156 9	2
AL'DA WS	Glaucarubolone	C20H26O8	394.163 05	3
AL'DA WS	Gigantine	C13H19NO3	237.136 46	3
AL'DA WS	Biflorin	C16H18O9	354.095 46	2
AL'DA WS	[FA oxo,hydroxy(18:3)] 12-oxo-14,18-dihydroxy-9Z,13E,15Z-octadecatrienoic acid	C18H28O5	324.193 93	1
AL'DA WS	Zinnimidine	C15H19NO3	261.136 41	3
AL'DA WS	(R)-Vicianin	C19H25NO10	427.147 64	2
AL'DA WS	Gabaculine	C7H9NO2	139.063 32	3
AL'DA WS	[FA methoxy,hydroxy(18:2)] 8-methoxy-13-hydroxy-9,11-octadecadienoic acid	C19H34O4	326.245 7	2
AL'DA WS	Vernoflexin	C20H24O4	328.167 81	4
AL'DA WS	Leukotriene B4	C20H32O4	336.230 39	48
AL'DA WS	Belladine	C19H25NO3	315.183 64	1
AL'DA WS	Asp-Leu-Asp-Tyr	C23H32N4O1 0	262.105 33	4
AL'DA WS	[FA (12:4/2:0)] 2E,4E,8E,10E-Dodecatetraenedioic acid	C12H14O4	222.089 36	8
AL'DA WS	[PR] Hemigossypol	C15H16O4	260.104 81	5
AL'DA WS	Scopolin	C16H20O9	356.111 09	3
AL'DA WS	2,5-diamino-6-(5-phospho-D-ribitylamino)pyrimidin-4(3H)-one	C9H18N5O8P	355.090 02	1
AL'DA WS	3-Hydroxysebacicacid	C10H18O5	218.115 52	2
AL'DA WS	Phaseolic acid	C13H12O8	296.053 45	1

AL'DA WS	2-Hydroxylauroylcarnitine	C19H37NO5	359.267 04	1
AL'DA WS	parabanate	C3H2N2O3	114.006 6	1
AL'DA WS	[FA oxo(4:0)] 11-oxo-5Z,9,12E,14E-prostatetraenoic acid	C20H28O3	316.203 88	32
AL'DA WS	1,2-dioctanoyl-1-amino-2,3-propanediol	C19H37NO4	343.272 07	2
AL'DA WS	Calanolide A	C22H26O5	370.178 8	2
AL'DA WS	[PR] Novaxenicins A	C20H28O5	348.193 96	8
AL'DA WS	O-Feruloylquinate	C17H20O9	368.111 05	1
AL'DA WS	Trichodermin	C17H24O4	292.167 7	2
AL'DA WS	[FA (18:1/3:0)] 15-octadecene-9,11,13-triynoic acid	C15H18O2	230.130 62	6
AL'DA WS	1-Isomangostin	C24H26O6	410.173 81	2
AL'DA WS	Byakangelicin	C17H18O7	334.105 7	1
AL'DA WS	Euphorbia factor Ti2	C32H42O7	269.147 32	1
AL'DA WS	Syringetin	C17H14O8	346.069 44	4
AL'DA WS	1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide	C9H15N4O8P	338.063 17	1
AL'DA WS	Metalaxyl	C15H21NO4	279.146 95	1
AL'DA WS	Cys-Gly-Arg	C11H22N6O4 S	334.142	1
AL'DA WS	[FA methyl(5:1/5:2/8:0)] methyl 8-[3,5-epidioxy-2-(3-hydroperoxy-1-pentenyl)-cyclopentyl]-octanoate	C22H36O6	396.251 3	2
AL'DA WS	[ST Trioxo(2:0)] 3,7,12-Trioxochola-1,4-dien-24-oic Acid	C24H30O5	398.209 41	2
AL'DA WS	Acronycidine	C15H15NO5	289.095 02	1
AL'DA WS	[FA (8:1/5:2/7:0)] 5-hydroperoxy-7-[3,5-epidioxy-2-(2-octenyl)-cyclopentyl]-6-heptenoic acid	C19H30O6	354.204 56	1

AL'DA WS	5,6,7-Trimethoxycoumarin	C12H12O5	236.068 42	2
AL'DA WS	factor γ-F420-1	C24H29N4O15P	322.067 62	1
AL'DA WS	2,6-Dioxo-6-phenylhexanoate	C12H12O4	220.073 61	2
AL'DA WS	[FA (10:1/2:0)] 2E-Decenedioic acid	C10H16O4	200.105	7
AL'DA WS	alpha-Zearalenol	C18H24O5	320.162 62	3
AL'DA WS	[FA trihydroxy(2:0)] 9S,11,15S-trihydroxy-2,3-dinor-thromboxa-5Z,13E-dien-1-oic acid	C18H30O6	342.204 27	3
AL'DA WS	Bendiocarb	C11H13NO4	223.084 46	2
AL'DA WS	N(alpha)-Benzyloxycarbonyl-L-leucine	C14H19NO4	265.131 34	2
AL'DA WS	[FA (20:0)] N-eicosanoyl-ethanolamine	C22H45NO2	355.345 06	1
AL'DA WS	Decylubiquinol	C19H32O4	324.230 08	3
AL'DA WS	sn-glycero-3-Phospho-1-inositol	C9H19O11P	334.067 33	4
AL'DA WS	2-isocapryloyl-3R-hydroxymethyl-γ-butyrolactone	C13H22O4	242.152 04	1
AL'DA WS	[Fv] Kushenol D	C27H32O6	452.220 07	4
AL'DA WS	[FA (7:0/2:0)] Heptanedioic acid	C7H12O4	160.073 63	4
AL'DA WS	Mangiferin	C19H18O11	422.085 29	3
AL'DA WS	2,5-Dioxopentanoate	C5H6O4	130.026 68	7
AL'DA WS	sinapaldehyde glucoside	C17H22O9	370.126 66	1
AL'DA WS	L-Ornithine	C5H12N2O2	132.089 88	6
AL'DA WS	(1R,6R)-6-Hydroxy-2-succinylcyclohexa-2,4-diene-1-carboxylate	C11H12O6	240.063 6	1
AL'DA WS	Scoparone	C11H10O4	206.057 98	6

AL'DA WS	Columbianetin	C14H14O4	246.089 11	4
AL'DA WS	N,N-Dimethylaniline	C8H11N	121.089 15	6
AL'DA WS	Fraxetin	C10H8O5	208.037 28	1
AL'DA WS	Samaderin A	C18H18O6	330.110 82	24
AL'DA WS	Domoic acid	C15H21NO6	311.136 86	1
AL'DA WS	[FA oxo,hydroxy(4:0)] 9-oxo-15S-hydroxy-5Z,10Z,13E,17Z-prostatetraenoic acid	C20H28O4	332.199 2	20
AL'DA WS	1'-Acetoxychavicol acetate	C13H14O4	234.089 16	1
AL'DA WS	Methylisoeugenol	C11H14O2	178.099 5	5
AL'DA WS	aloenin	C19H22O10	410.121 7	1
AL'DA WS	L-365260	C24H22N4O2	398.173 45	1
AL'DA WS	Flavonol 3-O-D-xylosylglucoside	C26H28O12	266.079 11	3
AL'DA WS	[FA hydroxy(6:0)] 4-hydroxy-hexanoic acid	C6H12O3	132.078 67	14
AL'DA WS	7,8-diketopelargonate	C9H14O4	186.089 33	2
AL'DA WS	UDP-2,4-diacetamido-2,4,6-trideoxy-β-L-altropyranose	C19H30N4O16P2	316.056 67	1
AL'DA WS	5,6-Dihydroxy-3-methyl-2-oxo-1,2,5,6-tetrahydroquinoline	C10H11NO3	193.074 09	10
AL'DA WS	Ac-Tyr-OEt	C13H17NO4	251.115 64	1
AL'DA WS	(-)-jasmonoyl-L-valine	C17H27NO4	309.193 92	2
AL'DA WS	Canin	C15H18O5	278.115 38	6
AL'DA WS	Inulicin	C17H24O5	308.162 68	2
AL'DA WS	Chaparrin	C20H28O7	380.183 95	1

AL'DA WS	2-Succinylbenzoate	C11H10O5	222.052 91	6
AL'DA WS	Cerulenin	C12H17NO3	223.120 88	2
AL'DA WS	dTDP-L-olivose	C16H26N2O1 4P2	266.042 73	5
AL'DA WS	Trichothecin	C19H24O5	332.162 71	6
AL'DA WS	4-Coumaroylshikimate	C16H16O7	320.089 8	3
AL'DA WS	3'-dAMP	C10H14N5O6 P	331.069	2
AL'DA WS	5-phosphoribosyl-4,5-aminoimidazole	C8H15N4O7P	310.067 84	1
AL'DA WS	17beta-Hydroxy-4-oxa-5alpha-estr-1-en-3-one acetate	C19H26O4	318.182 96	4
AL'DA WS	N-Glucosylnicotinate	C12H15NO7	285.084 69	2
AL'DA WS	Carnosol	C20H26O4	330.183 49	4
AL'DA WS	Nigakihemiacetal A	C22H34O7	410.230 84	2
AL'DA WS	(10S)-Juvenile hormone III acid diol	C15H26O4	270.183 31	1
AL'DA WS	Xanthoxin	C15H22O3	250.157	9
AL'DA WS	Diethofencarb	C14H21NO4	267.146 87	2
AL'DA WS	5-Hydroxyferulate	C10H10O5	210.052 98	2
AL'DA WS	[Fv] Protofarrerol	C17H18O6	318.110 68	3
AL'DA WS	(+)-Gallocatechin	C15H14O7	306.073 21	3
AL'DA WS	Indolelactate	C11H11NO3	205.073 99	5
AL'DA WS	Taxifolin	C15H12O7	304.059 1	8
AL'DA WS	[FA (20:2)] 15S-hydroperoxy-11Z,13E-eicosadienoic acid	C20H36O4	340.261 38	5

AL'DA WS	[FA hydroxy(20:4/2:0)] 5S,12R-dihydroxy-6Z,8E,10E,14Z-eicosatetraene-1,20-dioic acid	C20H30O6	366.204 53	7
AL'DA WS	Flossonol	C13H16O3	220.110 07	3
AL'DA WS	Tyr-OEt	C11H15NO3	209.105 29	3
AL'DA WS	Triangularine	C18H25NO5	335.173 17	3
AL'DA WS	3-(4-Hydroxyphenyl)lactate	C9H10O4	182.058 07	13
AL'DA WS	Chrysophanol 8-O-beta-D-glucoside	C21H20O9	416.111 47	7
AL'DA WS	Salicin	C13H18O7	286.105 36	1
AL'DA WS	Erythratidine	C19H25NO4	331.178 37	5
AL'DA WS	(+)-2,7-Dideoxypancratistatin	C14H15NO6	293.089 64	1
AL'DA WS	2,2',3-trihydroxy-3'-methoxy-5,5'-dicarboxybiphenyl	C15H12O8	320.053 52	5
AL'DA WS	Salidroside	C14H20O7	300.121 12	1
AL'DA WS	Taxiphyllin	C14H17NO7	311.100 39	4
AL'DA WS	[Fv] Antiarone K	C22H26O7	402.168 13	1
AL'DA WS	(+)-Gallocatechin	C15H14O7	306.074 16	3
AL'DA WS	dehypoxanthine futalosine	C14H16O7	296.088 87	1
AL'DA WS	trans-2,3-Dihydroxycinnamate	C9H8O4	180.042 39	11
AL'DA WS	Harzianopyridone	C14H19NO5	281.126 18	1
AL'DA WS	Vanilpyruvicacid	C10H10O5	210.052 96	2
AL'DA WS	1-Peroxyferolide	C17H22O7	338.136 96	1
AL'DA WS	Epinephrineglucuronide	C15H21NO9	359.121 42	1

AL'DA WS	[SP (17:0)] heptadecasphing-4-enine	C17H35NO2	285.266 65	3
AL'DA WS	Apigenin 7-O-[beta-D-apiosyl-(1->2)-beta-D-glucoside]	C26H28O14	282.074 04	6
AL'DA WS	[FA methyl,hydroxy(2:0)] 9-methylene-11R,15S-dihydroxy-5Z,13E-prostadienoic acid	C21H34O4	350.245 84	11
AL'DA WS	Damascenine	C10H13NO3	195.089 64	6
AL'DA WS	Antiarol	C9H12O4	184.073 64	6
AL'DA WS	Oblongolide	C14H20O2	220.146 34	2
AL'DA WS	5-Aminopentanoate	C5H11NO2	117.078 97	16
AL'DA WS	4-(beta-D-Glucosyloxy)benzoate	C13H16O8	300.084 75	1
AL'DA WS	Intermedine	C15H25NO5	299.173 18	5
AL'DA WS	3-pyridinecarboxaldehyde adenine dinucleotide	C21H26N6O14P2	324.048 47	1
AL'DA WS	[FA (12:5)] 3,5,7,9,11-dodecapentaenoic acid	C12H14O2	190.099 39	6
AL'DA WS	p-Coumaroyl quinic acid	C16H18O8	338.100 6	4
AL'DA WS	Arbutin	C12H16O7	272.089 77	2
AL'DA WS	3-Methyl-2-oxobutanoic acid	C5H8O3	116.047 37	9
AL'DA WS	Ala-Phe-Phe-Thr	C25H32N4O6	242.115 4	2
AL'DA WS	(R)-3-((R)-3-Hydroxybutanoyloxy)butanoate	C8H14O5	190.084 28	2
AL'DA WS	Prenyl caffeate	C14H16O4	248.104 72	2
AL'DA WS	p-Cresolglucuronide	C13H16O7	284.089 71	1
AL'DA WS	[SP (14:0)] 1-deoxy-tetradecasphinganine	C14H31NO	229.240 58	1
AL'DA WS	Tetraneurin E	C17H24O6	324.157 67	1

AL'DA WS	Ac-Tyr-OEt	C13H17NO4	251.115 65	1
AL'DA WS	[FA hydroxy(10:0/2:0)] 10-hydroxy-2E,8E-Decadiene-4,6-diynoic acid	C10H12O3	180.078 78	7
AL'DA WS	4,5-dihydro-5,5-dimethyl-4-(3-oxobutyl)furan-2(3H)-one	C10H16O3	184.110 04	9
AL'DA WS	3,4-Dihydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione	C19H24O4	316.167 5	6
AL'DA WS	3-(3,4-Dihydroxyphenyl)pyruvate	C9H8O5	196.037 31	3
AL'DA WS	(3E)-4-(2-Carboxyphenyl)-2-oxobut-3-enoate	C11H8O5	220.037 21	4
AL'DA WS	Sinapate	C11H12O5	224.068 56	2
AL'DA WS	[PK] 5-O-Methylvisamminol	C16H18O5	290.115 2	2
AL'DA WS	N-D-Glucosylarylamine	C12H17NO5	255.110 98	1
AL'DA WS	(4E)-2-Oxohexenoic acid	C6H8O3	128.047 39	9
AL'DA WS	Pinolidoxin	C18H26O6	338.173 38	2
AL'DA WS	nornorcapsaicin	C16H23NO3	277.167 68	1
AL'DA WS	4-O-beta-D-Glucosyl-4-hydroxycinnamate	C15H18O8	326.100 48	7
AL'DA WS	1-(3,4-dimethoxylphenyl)ethane-1,2-diol	C10H14O4	198.089 34	5
AL'DA WS	[FA hydroxy,dioxo(4:0/2:0)] 11R-hydroxy-9,15-dioxo-2,3,4,5-tetranor-prostan-1,20-dioic acid	C16H22O7	326.136 34	2
AL'DA WS	2-Furoate	C5H4O3	112.016 06	4
AL'DA WS	Ethyl cinnamate	C11H12O2	176.083 76	8
AL'DA WS	Polhovolide	C23H32O8	436.209 8	1
AL'DA WS	Sesartemin	C23H26O8	430.163 1	2
AL'DA WS	Decylubiquinone	C19H30O4	322.214 46	2

AL'DA WS	Tyramineglucuronide	C14H19NO7	313.116 09	1
AL'DA WS	α-oxo-1-carboxy-4-tetrahydrothiopyranpropanoate S-oxide	C9H12O6S	248.035 75	1
AL'DA WS	[Fv] Heteroflavanone B	C24H28O7	428.183 65	2
AL'DA WS	Glu-Glu-Ile-Tyr	C25H36N4O1 0	276.121 07	2
AL'DA WS	1alpha,5alpha-Dimercaptoandrostane-3alpha,17beta-diol	C19H32O2S2	356.183 89	1
AL'DA WS	Hippurate	C9H9NO3	179.058 41	6
AL'DA WS	Ailanthone	C20H24O7	376.152 58	4
AL'DA WS	Chlorogenate	C16H18O9	354.095 44	2
AL'DA WS	Microhelenin C	C20H26O5	346.178 32	8
AL'DA WS	(2S)-2-Isopropylmalate	C7H12O5	176.068 6	7
AL'DA WS	Hexadecanoic acid	C16H32O2	256.240 4	16
AL'DA WS	N-Acetyl-L-phenylalanine	C11H13NO3	207.089 59	6
AL'DA WS	2-Hydroxy-3-(4-hydroxyphenyl)propenoate	C9H8O4	180.042 44	11
AL'DA WS	4,5-Dioxopentanoate	C5H6O4	130.026 68	7
AL'DA WS	[FA oxo,hydroxy(5:1/5:0)] (1S,2R)-3-oxo-2-(5'-hydroxy-2'Z-pentenyl)-cyclopentaneacetic acid	C12H18O4	226.120 76	4
AL'DA WS	2,2',3,3'-tetrahydroxy-5,5'-dicarboxybiphenyl	C14H10O8	306.037 96	2
AL'DA WS	[FA methyl,oxo,hydroxy(2:0)] methyl 9-oxo-11R-hydroxy-15R-acetoxy-5Z,13E-prostadienoate	C23H36O6	408.251 44	2
AL'DA WS	L-Lysine	C6H14N2O2	146.105 55	8
AL'DA WS	3-Hydroxydodecanedioicacid	C12H22O5	246.146 81	1
AL'DA WS	Glaucarubinone	C25H34O10	494.215 11	2

AL'DA WS	ubiquinol-1	C14H20O4	252.136 27	1
AL'DA WS	Compactin diol lactone	C18H26O4	306.183 14	2
AL'DA WS	Furfural diethyl acetal	C9H14O3	170.094 44	2
AL'DA WS	Maculosin	C14H16N2O3	260.115 43	1
AL'DA WS	cis-1,2-Dihydroxy-4-methylcyclohexa-3,5-diene-1-carboxylate	C8H10O4	170.057 95	6
AL'DA WS	Formyl-5-hydroxykynurenamine	C10H12N2O3	208.084 79	2
AL'DA WS	Cavinine	C18H21NO6	347.136 75	3
AL'DA WS	[Fv] Dihydrokanakugiol	C19H22O6	346.142	12
AL'DA WS	sinapaldehyde glucoside	C17H22O9	370.126 73	1
AL'DA WS	Malonate	C3H4O4	104.010 99	3
AL'DA WS	Sarracine	C18H27NO5	337.188 69	3
AL'DA WS	N-(3-oxooctanoyl)-L-homoserine	C12H21NO5	259.141 88	1
AL'DA WS	2-oxosuberate	C8H12O5	188.068 53	1
AL'DA WS	Tetraneurin A	C17H22O6	322.141 92	2
AL'DA WS	[PR] Parthenin	C15H18O4	262.120 31	10
AL'DA WS	Genipin	C11H14O5	226.084 15	1
AL'DA WS	Glutathionylaminopropylcadaverine	C18H36N6O5 S	448.246 21	2
AL'DA WS	Harzianopyridone	C14H19NO5	281.126 29	1
AL'DA WS	Oxalate	C2H2O4	89.9953 77	1
AL'DA WS	omega-Carboxy-trinor-leukotriene B4	C18H24O6	336.157 6	1

AL'DA WS	[PR ethyl,hydrox] gamma-carboxyethyl-hydroxychroman beta-D-glucoside	C21H30O9	426.189 35	1
AL'DA WS	14-Dihydroxycornestin	C16H20O6	308.126 28	2
AL'DA WS	Gentiopicrin	C16H20O9	356.111 09	3
AL'DA WS	[FA (21:0/2:0)] Heneicosanedioic acid	C21H40O4	356.292 66	6
AL'DA WS	Asn-Phe-Gln-Tyr	C27H34N6O8	285.121 12	1
AL'DA WS	Perillyl alcohol	C10H16O	152.120 13	38
AL'DA WS	[FA dioxo(10:0)] 3,6-dioxo-decanoic acid	C10H16O4	200.105 04	7
AL'DA WS	[FA (20:0)] N-(11Z-eicosaenoyl)-ethanolamine	C22H43NO2	353.329 22	1
AL'DA WS	Eriolangin	C20H28O6	364.189 05	10
AL'DA WS	Kievitone hydrate	C20H22O7	374.137	15
AL'DA WS	HMDBOA-Glc	C16H21NO10	387.116 38	1
AL'DA WS	[Fv] Chalconaringenin 2'-xyloside	C20H20O9	404.111 26	2
AL'DA WS	(-)-Sedamine	C14H21NO	219.162 37	2
AL'DA WS	[Fv methoxy,methyl(6:0)] 2,3,4,5,2',6'-Hexamethoxy-3',4'-methylenedioxychalcone	C22H24O9	432.142 37	8
AL'DA WS	L-Proline	C5H9NO2	115.063 35	4
AL'DA WS	[FA] Methyl jasmonate	C13H20O3	224.141 23	6
AL'DA WS	D-Tyrosine	C9H11NO3	181.074	11
AL'DA WS	ubiquinol-2	C19H28O4	320.198 86	4
AL'DA WS	3-Hydroxy-5, 8-tetradecadiencarnitine	C21H37NO5	383.266 94	1
AL'DA WS	Piceid	C20H22O8	390.132 06	2

AL'DA WS	ethyl-(2R)-methyl-(3S)-hydroxybutanoate	C7H14O3	146.094 36	8
AL'DA WS	N-Acetylvanilalanine	C12H15NO5	253.094 93	1
AL'DA WS	[Fv] Exiguaflavanone M	C25H30O7	442.199 38	2
AL'DA WS	2-Octenoylcarnitine	C15H27NO4	285.193 81	1
AL'DA WS	Citrate	C6H8O7	192.027 23	12
AL'DA WS	6-Keto-decanoylcarnitine	C17H31NO5	329.220 15	1
AL'DA WS	Alatolide	C19H26O6	350.173 42	3
AL'DA WS	[FA dioxo,hydroxy(4:0/2:0)] 9,15-dioxo-11R-hydroxy-2,3,4,5-tetranor-prostan-1,20-dioic acid	C16H24O7	328.152 48	3
AL'DA WS	3-Isopropylbut-3-enoic acid	C7H12O2	128.083 77	23
AL'DA WS	Butenylcarnitine	C11H19NO4	229.131 44	1
AL'DA WS	[PR] 1,2-Dihydrosantonin	C15H20O3	248.141 02	20
AL'DA WS	Lotaustralin	C11H19NO6	261.121 12	4
AL'DA WS	1alpha,5alpha-Epidithio-17a-oxa-D-homoandrostan-3,17-dione	C19H26O3S2	366.131 87	1
AL'DA WS	N-Methylmescaline	C12H19NO3	225.136 49	3
AL'DA WS	2S-Hydroxytetradecanoic acid	C14H28O3	244.203 93	12
AL'DA WS	[FA hydroxy(20:4)] 15S-hydroxy-5Z,8Z,11Z,13E-eicosatetraenoic acid	C20H32O3	320.235 36	53
AL'DA WS	[FA (10:0/2:0)] Decanedioic acid	C10H18O4	202.120 68	2
AL'DA WS	N,N-Dihydroxy-L-tyrosine	C9H11NO5	213.063 91	2
AL'DA WS	[FA methyl,hydroxy,oxo(5:2/4:0)] methyl 4-[2-(2-formyl-vinyl)-3-hydroxy-5-oxo-cyclopentyl]-butanoate	C13H18O5	254.115 6	1
AL'DA WS	[FA methyl(14:0/2:0)] 3-methyl-tetradecanedioic acid	C15H28O4	272.198 57	3

AL'DA WS	Tetraneurin E	C17H24O6	324.157 64	1
AL'DA WS	N-Pyruvoyl-5-methoxy-3-hydroxyanthranilate	C11H11NO6	253.058 94	1
AL'DA WS	Ecgonine methyl ester	C10H17NO3	199.120 93	4
AL'DA WS	2-oxopimelate	C7H10O5	174.052 87	6
AL'DA WS	[FA hydroxy(11:2/11:2)] 2R,9R-dihydroxy-3S,4S,7S,8S-diepoxy-5E,10-undecadien-1-ol	C11H16O5	228.099 99	1
AL'DA WS	Caproylcholine	C11H23NO2	201.172 95	2
AL'DA WS	alpha-Ribazole	C14H18N2O4	278.126 58	3
AL'DA WS	12-Hydroxydodecanoic acid	C12H24O3	216.172 63	11
AL'DA WS	Lariciresinol	C20H24O6	360.157 74	12
AL'DA WS	1-Peroxyferolide	C17H22O7	338.137 08	1
AL'DA WS	Rhipocephalin	C21H28O6	376.188 92	2
AL'DA WS	N-Acetyl-D-fucosamine	C8H15NO5	205.095 14	3
AL'DA WS	[PR] Alliacol A	C15H20O4	264.136 03	29
AL'DA WS	Decanoylcholine	C15H31NO2	257.235 4	1
AL'DA WS	[FA hydroxy(11:0)] 2-hydroxy-undecanoic acid	C11H22O3	202.157 08	6
AL'DA WS	3-Dimethylallyl-4-hydroxymandelic acid	C13H16O4	236.104 77	1
AL'DA WS	[FA amino(16:0)] 2S-aminohexadecanoic acid	C16H33NO2	271.250 96	4
AL'DA WS	Gln-Leu-Trp-Pro	C27H38N6O6	271.141 86	2
AL'DA WS	Glu-Leu-Trp-Gln	C27H38N6O8	287.136 7	2
AL'DA WS	4'-Cinnamoylmussatioside	C34H44O16	354.131 92	1

AL'DA WS	NG,NG-Dimethyl-L-arginine	C8H18N4O2	202.143 21	3
AL'DA WS	[FA (8:1/5:2/7:0)] 5-hydroperoxy-7-[3,5-epidioxy-2-(2-octenyl)-cyclopentyl]-6-heptenoic acid	C19H30O6	354.204 72	1
AL'DA WS	aloenin	C19H22O10	410.121 76	1
AL'DA WS	[PR] (-)-Limonene	C10H16	136.125 25	30
AL'DA WS	[Fv Hydroxy,methoxy(7:0)] 6'-Hydroxy-2,3,4,5,2',3',4'-heptamethoxychalcone	C22H26O9	434.157 96	8
AL'DA WS	hydrogen iodide	HI	127.912 35	1
AL'DA WS	Buddledin A	C17H24O3	276.172 32	2
AL'DA WS	Cys-Leu-Trp-Trp	C31H38N6O5 S	303.131 75	2
AL'DA WS	Sulfate	H2O4S	97.9673 84	1
AL'DA WS	[Fv methoxy(6:0/9:1)] 5,6,7,3',4',5'-Hexamethoxyflavanone	C21H24O8	404.147 38	2
AL'DA WS	[PR] Vomitoxin	C15H20O6	296.126 25	2
AL'DA WS	Vernoflexuoside	C21H28O8	408.178 77	1
AL'DA WS	[FA methyl,oxo(4:0/4:0)] methyl 4R,12S-diacetoxy-9-oxo-5Z,7E,10Z,13Z-prostatetraenoate-cyclo[8,12]	C25H34O7	446.230 49	7
AL'DA WS	Sinapyl alcohol	C11H14O4	210.089 39	4
AL'DA WS	[FA (12:4)] 2E,4E,8Z,10E-dodecatetraenoic acid	C12H16O2	192.115 17	5
AL'DA WS	2-Hydroxy-6-ketononatrienedioate	C9H8O6	212.032 26	3
AL'DA WS	1'-Acetoxyeugenol acetate	C14H16O5	264.099 84	1
AL'DA WS	[FA (7:0)] heptanoic acid	C7H14O2	130.099 44	14
AL'DA WS	N-Ethylmaleimide	C6H7NO2	125.047 69	4
AL'DA WS	2-Isopropylmaleate	C7H10O4	158.057 98	5

AL'DA WS	(R)-3-Hydroxy-3-methyl-2-oxopentanoate	C6H10O4	146.057 98	16
AL'DA WS	Codonopsine	C14H21NO4	267.146 91	2
AL'DA WS	[FA (10:0)] O-decanoyl-R-carnitine	C17H33NO4	315.240 91	2
AL'DA WS	L-Hypoglycin	C7H11NO2	141.079 05	3
AL'DA WS	Rosmarinine	C18H27NO6	353.183 88	1
AL'DA WS	[FA dioxo(8:0)] 4,7-dioxo-octanoic acid	C8H12O4	172.073 71	4
AL'DA WS	Echimidine	C20H31NO7	397.209 74	2
AL'DA WS	Valtratum	C22H30O8	422.194 59	1
AL'DA WS	Monomethyl sulfate	CH4O4S	111.983 04	2
AL'DA WS	O-Butanoylcarnitine	C11H21NO4	231.147 06	3
AL'DA WS	dTDP-glucose	C16H26N2O16P2	282.037 69	3
AL'DA WS	Amabiline	C15H25NO4	283.178 21	2
AL'DA WS	3-Hydroxy-cis-5-tetradecenoylcarnitine	C21H39NO5	385.282 33	1
AL'DA WS	[FA (6:0)] O-hexanoyl-R-carnitine	C13H25NO4	259.178 23	2
AL'DA WS	Longifolonine	C17H15NO4	297.099 45	1
AL'DA WS	Anacrotine	C18H25NO6	351.167 83	6
AL'DA WS	[PR] Tretinoin/All-Trans Retinoic Acid	C20H28O2	300.209 16	25
AL'DA WS	(S)-Methylmalonate semialdehyde	C4H6O3	102.031 7	8
AL'DA WS	[FA trihydroxy(2:0)] 9S,11R,15S-trihydroxy-2,3-dinor-5Z,13E-prostadienoic acid-cyclo[8S,12R]	C18H30O5	326.209 31	4
AL'DA WS	[Fv Hydroxy,trimethox] 4-Hydroxy-2',4',6'-trimethoxydihydrochalcone 4-O-glucoside	C24H30O10	478.183 9	1

AL'DA WS	Erioflorin acetate	C21H26O7	390.168 3	1
AL'DA WS	Dodecanamide	C12H25NO	199.193 69	1
AL'DA WS	Euparotin acetate	C22H26O8	418.163 13	3
AL'DA WS	[FA methyl,oxo(5:0/5:0)] methyl 4R,12S-diacetoxy-9-oxo-5Z,7E,10Z,13Z,17Z-prostapentaenoate-cyclo[8,12]	C25H32O7	444.214 94	1
AL'DA WS	ubiquinol-1	C14H20O4	252.136 03	1
AL'DA WS	Cardiospermin	C11H17NO7	275.100 4	3
AL'DA WS	Linamarin	C10H17NO6	247.105 49	2
AL'DA WS	[FA methyl,hydroxy,oxo(18:0)] methyl 9,12-dihydroxy-13-oxo-10-octadecenoate	C19H34O5	342.240 56	2
AL'DA WS	Dihydromyricetin	C15H12O8	320.053 59	5
AL'DA WS	2,3,5-Trihydroxytoluene	C7H8O3	140.047 39	8
AL'DA WS	[FA (11:0)] undecanoic acid	C11H22O2	186.162 07	12
AL'DA WS	Senkirkine	C19H27NO6	365.183 43	1
AL'DA WS	Phenylpropanoate	C9H10O2	150.068 11	10
AL'DA WS	Petasitenine	C19H27NO7	381.178 45	1
AL'DA WS	Hydroxypropionylcarnitine	C10H19NO5	233.126 32	1
AL'DA WS	cis-(homo)3aconitate	C9H12O6	216.063 6	1
AL'DA WS	3,4-Dihydro-7-methoxy-2-methylene-3-oxo-2H-1,4-benzoxazine-5- carboxylic acid	C11H9NO5	235.048 18	1
AL'DA WS	2-Hexenoylcarnitine	C13H23NO4	257.162 66	1
AL'DA WS	1-(4-Hydroxyphenyl)-2-aminoethanol	C8H11NO2	153.078 99	5
AL'DA WS	Xanthoxin	C15H22O3	250.156 78	9

AL'DA WS	3-sulfopropanoate	C3H6O5S	153.9937	1
AL'DA WS	2-Hydroxydecanedioicacid	C10H18O5	218.11557	2
AL'DA WS	N'-Phosphoguanidinoethyl methyl phosphate	C4H13N3O7P2	277.02236	1
AL'DA WS	4-Hydroxyphenylethanol	C8H10O2	138.06809	6
AL'DA WS	Furcatin	C20H28O10	214.08448	1
AL'DA WS	Tutin	C15H18O6	294.11068	1
AL'DA WS	Decanoic acid	C10H20O2	172.14646	9
AL'DA WS	Ibuprofenacylglucuronide	C19H26O8	382.16321	1
AL'DA WS	2-Acetolactate	C5H8O4	132.0423	16
AL'DA WS	[PR] (1S,4R)-1-Hydroxy-2-oxolimonene	C10H16O2	168.11507	60
AL'DA WS	Arnicolide A	C17H22O5	306.14671	14
AL'DA WS	[ST (2:0)] estra-1,3,5(10)-triene-3,17alpha-diol 3-D-glucuronide	C24H32O8	448.21002	6
AL'DA WS	1alpha,5alpha-Epidithio-17a-oxa-D-homoandrostan-3,17-dione	C19H26O3S2	366.13179	1
AL'DA WS	Hydroxybutyrylcarnitine	C11H21NO5	247.1419	1
AL'DA WS	Orthophosphate	H3O4P	97.977082	1
AL'DA WS	trans-Cinnamoyl beta-D-glucoside	C15H18O7	310.10568	3
AL'DA WS	N-(octanoyl)-L-homoserine	C12H23NO4	245.16265	4
AL'DA WS	Aspidinol	C12H16O4	224.10484	3
AL'DA WS	3,4,6-trihydroxy-cis-cinnamate	C9H8O5	196.03732	3
AL'DA WS	N-(2,3-Dihydroxybenzoyl)-L-serine	C10H11NO6	241.05855	2

AL'DA WS	Nummularine F	C23H32N4O4	214.120 63	1
AL'DA WS	all-trans-Retinoyl-beta-glucuronide	C26H36O8	238.120 48	1
AL'DA WS	(-)-Jasmonic acid	C12H18O3	210.125 67	5
AL'DA WS	9-Decenoylcarnitine	C17H31NO4	313.225 29	1
AL'DA WS	O-Propanoylcarnitine	C10H19NO4	217.131 48	1
AL'DA WS	[FA hydroxy(11:2/11:2)] 2R,9R-dihydroxy-3S,4S,7S,8S-diepoxy-5E,10-undecadien-1-ol	C11H16O5	228.100 01	1
AL'DA WS	Actinodaphnine	C18H17NO4	311.114 97	3
AL'DA WS	cis-(homo)2aconitate	C8H10O6	202.047 9	1
AL'DA WS	Uplandicine	C17H27NO7	357.178 53	1
AL'DA WS	Monocerin	C16H20O6	308.126 41	2
AL'DA WS	Browniine	C25H41NO7	467.288 44	3
AL'DA WS	3-4-5-Trimethoxycinnamicacid	C12H14O5	238.084 31	2
AL'DA WS	[SP methyl(16:0)] 15-methyl-hexadecasphing-4E-enine	C17H35NO2	285.266 62	3
AL'DA WS	3-Dimethylallyl-4-hydroxymandelic acid	C13H16O4	236.104 75	1
AL'DA WS	3-Methyl-1-(2,4,6-trihydroxyphenyl)butan-1-one	C11H14O4	210.089 29	4
AL'DA WS	Deoxycholicacid3-glucuronide	C30H48O10	284.162 18	2
AL'DA WS	[PR trihydroxy(2:0)] (-)-8alpha,15-diacetoxy-4beta,9alpha,14beta-trihydroxy-3(16)-fusicoccene	C24H38O7	438.261 79	1
AL'DA WS	Asp-Phe-Pro-Tyr	C27H32N4O8	270.110 56	1
AL'DA WS	Nona-2,6-dienal	C9H14O	138.104 49	8
AL'DA WS	(E)-2-Butenyl-4-methyl-threonine	C9H17NO3	187.120 92	3

AL'DA WS	[FA hydroxy,oxo,methyl(2:0)] 9S,15S-dihydroxy-11-oxo-15-methyl-5Z,13E-prostadienoic acid	C21H34O5	366.240 67	9
AL'DA WS	Tyr-Gly-Tyr-Tyr	C29H32N4O8	282.110 48	1
AL'DA WS	[SP hydrox] 6-hydroxysphing-4E-enine	C18H37NO3	315.277 39	3
AL'DA WS	[6]-Gingerol	C17H26O4	294.183 08	3
AL'DA WS	2-oxosuberate	C8H12O5	188.068 63	1
AL'DA WS	[PR] 9-Hydroxy-helminthosporol	C15H24O3	252.172 41	7
AL'DA WS	Didrovaltratum	C22H32O8	212.104 89	2
AL'DA WS	4-hydroxy-5-methyl-2-propyl-3(2H)-furanone	C8H12O3	156.078 74	2
AL'DA WS	[PR] Flakinin A	C15H20O5	280.130 91	6
AL'DA WS	Phthalate	C8H6O4	166.026 67	7
AL'DA WS	L-Metanephrine	C10H15NO3	197.105 33	3
AL'DA WS	[SP amino,tetramethyl(4:0/18:0/3:0)] 2S-amino-5,9,13,17-tetramethyl-8E,16-octadecadiene-1,3R,14-triol	C22H43NO3	369.324 14	1
AL'DA WS	3-Methoxy-4-hydroxyphenylethyleneglycol	C9H12O4	184.073 64	6
AL'DA WS	Retronecine	C8H13NO2	155.094 71	4
AL'DA WS	[SP (14:0)] tetradecasphing-4E-enine	C14H29NO2	243.219 78	2
AL'DA WS	Sinapate	C11H12O5	224.068 63	2
AL'DA WS	[FA amino(10:0)] 10-amino-decanoic acid	C10H21NO2	187.157 26	5
AL'DA WS	[FA (3:2/20:4)] N-propyl-5Z,8Z,11Z,14Z-eicosatetraenoyl amine	C23H39NO	345.303 17	3
AL'DA WS	4-Hydroxybenzoate	C7H6O3	138.031 75	7
AL'DA WS	N-Succinyl-L-glutamate 5-semialdehyde	C9H13NO6	231.074 4	2

AL'DA WS	[FA amino(13:0)] 13-amino-tridecanoic acid	C13H27NO2	229.204 22	5
AL'DA WS	(+)-7-Isojasmonic acid	C12H18O3	210.125 67	5
AL'DA WS	N-D-Glucosylarylamine	C12H17NO5	255.110 65	1
AL'DA WS	Mycosporine	C11H19NO6	261.121 23	4
AL'DA WS	Hydroxypropionylcarnitine	C10H19NO5	233.126 31	1
AL'DA WS	pentane-1,3,4,5-tetracarboxylate	C9H12O8	248.053 22	1
AL'DA WS	sorbate	C6H8O2	112.052 39	17
AL'DA WS	3Hydroxycoumarin	C9H6O3	162.031 78	3
AL'DA WS	Perillic acid	C10H14O2	166.099 44	17
AL'DA WS	N-(3-Oxooctanoyl)homoserine lactone	C12H19NO4	241.131 41	2
AL'DA WS	Naringenin 7-O-beta-D-glucoside	C21H22O10	434.121 68	27
AL'DA WS	8-Methoxykynurenate	C11H9NO4	219.053 42	2
AL'DA WS	8-Amino-7-oxononanoate	C9H17NO3	187.120 92	3
AL'DA WS	12alpha-Hydroxyamoorstatin	C28H36O10	266.115 56	1
AL'DA WS	Salicin	C13H18O7	286.105 48	1
AL'DA WS	Nivalenol	C15H20O7	312.121 35	3
AL'DA WS	L-Octanoylcarnitine	C15H29NO4	287.209 52	1
AL'DA WS	L-Adrenaline	C9H13NO3	183.089 62	3
AL'DA WS	Dodecanoic acid	C12H24O2	200.177 78	10
AL'DA WS	[FA (7:2)] 2,4-heptadienal	C7H10O	110.073 17	2

AL'DA WS	N(alpha)-t-Butoxycarbonyl-L-leucine	C11H21NO4	231.1471	3
AL'DA WS	[FA oxo(6:0)] N-(3-oxo-hexanoyl)-homoserine lactone	C10H15NO4	213.10009	2
AL'DA WS	Rugosal	C15H22O4	266.15173	5
AL'DA WS	[FA methyl,hydroxy,oxo(5:2/8:0)] methyl 8-[2-(2-formyl-vinyl)-3-hydroxy-5-oxo-cyclopentyl]-octanoate	C17H26O5	310.17804	2
AL'DA WS	(S)-Cheilanthifoline	C19H19NO4	325.13069	7
AL'DA WS	Sinapyl alcohol	C11H14O4	210.08925	4
AL'DA WS	p-Cresolglucuronide	C13H16O7	284.08974	1
AL'DA WS	4-(2-Amino-3-hydroxyphenyl)-2,4-dioxobutanoate	C10H9NO5	223.0481	6
AL'DA WS	Phenylgalactoside	C12H16O6	256.09494	5
AL'DA WS	Serratine	C16H25NO3	279.18327	3
AL'DA WS	L-Arogenate	C10H13NO5	227.07934	1
AL'DA WS	[PR] 4-Ketoalloxanthin	C40H50O3	289.18868	1
AL'DA WS	Dopaquinone	C9H9NO4	195.05329	9
AL'DA WS	Glycyphyllin	C21H24O9	420.14237	6
AL'DA WS	4-Propylphenol	C9H12O	136.08885	4
AL'DA WS	N-(3-oxohexanoyl)-L-homoserine	C10H17NO5	231.11067	3
AL'DA WS	[FA (8:0)] octanoic acid	C8H16O2	144.11506	11
AL'DA WS	8-keto-7-aminoperlagonate	C9H15NO3	185.10528	5
AL'DA WS	Salidroside	C14H20O7	300.12116	1
AL'DA WS	Hordatine B	C29H40N8O5	290.15594	1

AL'DA WS	(2S)-2-{[1-(R)-Carboxyethyl]amino}pentanoate	C8H15NO4	189.100 13	5
AL'DA WS	Succinate	C4H6O4	118.026 65	7
AL'DA WS	futalosine	C19H16N4O7	412.101 03	1
AL'DA WS	2-Amino-3-phosphonopropanoate	C3H8NO5P	169.013 88	4
AL'DA WS	3,6-Nonadienal	C9H14O	138.104 49	8
AL'DA WS	Betaine	C5H11NO2	117.078 97	16
AL'DA WS	dihydroartemisinic acid hydroperoxide	C15H24O4	268.167 33	1
AL'DA WS	N,N-Dihydroxy-L-tyrosine	C9H11NO5	213.063 77	2
AL'DA WS	Deidaclin	C12H17NO6	271.105 45	1
AL'DA WS	Pancratistatin	C14H15NO8	325.079 96	1
AL'DA WS	[PR] Iridotrial	C10H14O3	182.094 4	8
AL'DA WS	CAI-1	C13H26O2	214.193 48	15
AL'DA WS	4-(beta-D-Glucosyloxy)benzoate	C13H16O8	300.084 88	1
AL'DA WS	Luciduline	C13H21NO	207.162 33	1
AL'DA WS	Tenuazonic acid	C10H15NO3	197.105 31	3
AL'DA WS	5-(2'-Formylethyl)-4,6-dihydroxypicolinate	C9H9NO5	211.048 16	5
AL'DA WS	Phenylethyl alcohol	C8H10O	122.073 15	12
AL'DA WS	L-Arginine	C6H14N4O2	174.111 75	2
AL'DA WS	Genipin	C11H14O5	226.084 34	1
AL'DA WS	[FA amino(14:0)] 2-amino-tetradecanoic acid	C14H29NO2	243.219 8	2

AL'DA WS	[FA (9:2)] 2,6-nonadienoic acid	C9H14O2	154.099 38	11
AL'DA WS	Graphinone	C16H24O5	296.162 33	1
AL'DA WS	Indicine	C15H25NO5	299.173 29	5
AL'DA WS	3-Hydroxyanthranilate	C7H7NO3	153.042 7	15
AL'DA WS	Ascorbate	C6H8O6	176.032 15	15
AL'DA WS	L-Glutamate	C5H9NO4	147.053 15	14
AL'DA WS	Lancerin	C19H18O10	406.090 37	1
AL'DA WS	2-n-nonyl-4-hydroxyquinoline-N-oxide	C19H27NO2	301.203 96	1
AL'DA WS	Lophocerine	C15H23NO2	249.172 79	1
AL'DA WS	Crotanecine	C8H13NO3	171.089 56	2
AL'DA WS	Pimelylcarnitine	C14H25NO6	303.168 15	1
AL'DA WS	N-(3-oxooctanoyl)-L-homoserine	C12H21NO5	259.141 91	1
AL'DA WS	alpha-N-Phenylacetyl-L-glutamine	C13H16N2O4	264.110 89	4
AL'DA WS	cis-Aconitate	C6H6O6	174.016 5	3
AL'DA WS	Uplandicine	C17H27NO7	357.178 5	1
AL'DA WS	Thromboxane B2	C20H34O6	370.235 54	6
AL'DA WS	[SP hydroxy,methyl,methyl(18:2/2:0)] 2S-(hydroxymethyl)-4S-(12-methyloctadecyl)azetidin-3R-ol	C23H47NO2	369.360 65	1
AL'DA WS	1,3,5-trimethoxybenzene	C9H12O3	168.078 68	3
AL'DA WS	[FA methyl(18:2)] methyl 9-hydroperoxy-10,12-epidioxy-13,15-octadecadienoate	C19H32O6	356.219 72	14
AL'DA WS	Sulfoacetate	C2H4O5S	139.978 01	1

AL'DA WS	Westiellamide	C27H42N6O6	273.157 48	3
AL'DA WS	3-Isopropylcatechol	C9H12O2	152.083 75	4
AL'DA WS	3-hydroxy-3-carboxy-4,5-cyclopropylhex-5-enoate	C8H10O5	186.052 93	2
AL'DA WS	Secologanate	C16H22O10	374.121 58	3
AL'DA WS	Succinyl proline	C9H13NO5	215.079 42	1
AL'DA WS	3-Oxopropanoate	C3H4O3	88.0160 43	3
AL'DA WS	N-Acetyl-L-leucine	C8H15NO3	173.105 35	5
AL'DA WS	aeruginosin 98-b	C29H46N6O9 S	327.152 96	1
AL'DA WS	Artecanin	C15H18O5	278.115 44	6
AL'DA WS	dTDP	C10H16N2O1 1P2	402.022 66	1
AL'DA WS	Hexadecasphinganine	C16H35NO2	273.266 64	1
AL'DA WS	Fasoracetam	C10H16N2O2	196.121 24	2
AL'DA WS	Terminaline	C23H41NO2	363.313 54	1
AL'DA WS	2,5-Dihydroxybenzoate	C7H6O4	154.026 68	10
AL'DA WS	(+-)-Carnegine	C13H19NO2	221.141 61	2
AL'DA WS	Stipitatate	C8H6O5	182.021 7	5
AL'DA WS	3-(3,4-Dihydroxyphenyl)lactate	C9H10O5	198.052 92	5
AL'DA WS	Prostaglandin G2	C20H32O6	368.219 9	14
AL'DA WS	3,4-Dihydroxyphenylethyleneglycol	C8H10O4	170.057 95	6
AL'DA WS	demethyl-phosphinothricin	C4H10NO4P	167.034 55	1

AL'DA WS	[FA hydroxy(10:0)] N-(3S-hydroxydecanoyl)-L-serine	C13H25NO5	275.173 16	2
AL'DA WS	[ST Trihydroxy,ox] 3alpha,7beta,12alpha-Trihydroxy-6-oxo-5alpha-cholan-24-oic Acid	C24H38O6	422.267 2	5
AL'DA WS	Ser-Tyr	C12H16N2O5	268.106 07	1
AL'DA WS	[SP amino,trihydroxy,hydroxy,methyl,oxo(20:0)] 2S-amino-3R,4R,5S-trihydroxy-2-(hydroxymethyl)-14-oxo-eicos-6E-enoic acid	C21H39NO7	417.272 62	1
AL'DA WS	nornorcapsaicin	C16H23NO3	277.167 64	1
AL'DA WS	9-Oxononanoic acid	C9H16O3	172.110 03	12
AL'DA WS	Nummularine F	C23H32N4O4	214.120 59	1
AL'DA WS	Maleylpyruvate	C7H6O6	186.016 55	11
AL'DA WS	Catalposide	C22H26O12	482.142 52	1
AL'DA WS	beta-Eucaine	C15H21NO2	247.157 13	3
AL'DA WS	[FA methoxy(16:1)] 2-methoxy-5Z-hexadecenoic acid	C17H32O3	284.234 59	7
AL'DA WS	all-trans-tridecaprenyl diphosphate	C65H108O7P2	531.377 37	1
AL'DA WS	L-Phenylalanine	C9H11NO2	165.079 05	7
AL'DA WS	2-5-Furandicarboxylicacid	C6H4O5	156.006 02	1
AL'DA WS	o-topolin	C12H13N5O	243.111 85	3
AL'DA WS	Hexanoic acid	C6H12O2	116.083 74	16
AL'DA WS	[SP (2:0)] sphinga-4E,14Z-dienine	C18H35NO2	297.266 76	9
AL'DA WS	N-Butyryl-L-homoserine lactone	C8H13NO3	171.089 54	2
AL'DA WS	(+/-)-5-[(tert-Butylamino)-2'-hydroxypropoxy]-3,4-dihydro-1(2H)- naphthalenone	C17H25NO3	291.183 28	7
AL'DA WS	4-Guanidinobutanoate	C5H11N3O2	145.085 16	3

AL'DA WS	L-Leucine	C6H13NO2	131.094 65	12
AL'DA WS	N-Caffeoylputrescine	C13H18N2O3	250.131 54	1
AL'DA WS	3,4-Dihydroxy-L-phenylalanine	C9H11NO4	197.068 92	2
AL'DA WS	N'-Phosphoguanidinoethyl methyl phosphate	C4H13N3O7P 2	277.022 41	1
AL'DA WS	2-Dehydro-3-deoxy-L-rhamnonate	C6H10O5	162.052 95	25
AL'DA WS	[FA amino(14:0)] 2-amino-tetradecanoic acid	C14H29NO2	243.219 8	2
AL'DA WS	[FA hydroxy(7:1/2:0)] 2,4-dihydroxy-2-heptenedioc acid	C7H10O6	190.047 85	4
AL'DA WS	Brugine	C12H19NO2S 2	273.086 37	1
AL'DA WS	Abscisic acid glucose ester	C22H32O8	212.105 06	2
AL'DA WS	[FA methyl(20:4)] N-methyl-5Z,8Z,11Z,14Z-eicosatetraenoyl amine	C21H35NO	317.271 81	3
AL'DA WS	Harmalol	C12H10N2O	198.079 4	4
AL'DA WS	Shikimate	C7H10O5	174.052 86	6
AL'DA WS	(R)-2-Hydroxyglutarate	C5H8O5	148.037 28	18
AL'DA WS	3-Hydroxy-2-methylpyridine-4,5-dicarboxylate	C8H7NO5	197.032 59	3
AL'DA WS	chelidamate	C7H5NO5	183.016 99	1
AL'DA WS	Phenacetin	C10H13NO2	179.094 65	5
AL'DA WS	3-(2-propenoic acid)-4,6-hydroxy cyclohexa-2,5-dienone	C9H8O5	196.037 34	3
AL'DA WS	saccharin	C7H5NO3S	182.999 22	1
AL'DA WS	Pyridoxine	C8H11NO3	169.073 99	4
AL'DA WS	Xanthurenic acid	C10H7NO4	205.037 79	2

AL'DA WS	tetrahomocitrate	C10H16O7	248.089 73	1
AL'DA WS	AminoDHQ	C7H11NO5	189.063 69	4
AL'DA WS	(2S)-2-Isopropyl-3-oxosuccinate	C7H10O5	174.052 88	6
AL'DA WS	4-Hydroxybutanoic acid	C4H8O3	104.047 39	13
AL'DA WS	beta-Carboline	C11H8N2	168.068 8	1
DAWS	Dimethisterone	C23H32O2	340.240 46	2
DAWS	3,4',5-Trihydroxystilbene	C14H12O3	228.078 72	8
DAWS	Formononetin	C16H12O4	268.073 68	6
DAWS	Equol	C15H14O3	242.094 16	10
DAWS	N-Acetoxy-4-aminobiphenyl	C14H13NO2	227.094 62	3
DAWS	Nuarimol	C17H12FN2O Cl	314.061 87	1
DAWS	6-(1,2,3,4-Tetrahydro-6-methoxy-2-naphthyl)-2(1H)-pyridone	C16H17NO2	255.125 74	2
DAWS	[FA (17:1/2:0)] 8E-Heptadecenedioic acid	C17H30O4	298.214 86	1
DAWS	2-(4-Chlorophenyl)-3-phenyl-3-(2-pyridinyl)acrylonitrile	C20H13N2Cl	316.077 47	1
DAWS	DMB-S-MMP	C10H18O3S	218.098 31	2
DAWS	2-oxo-10-methylthiodecanoate	C11H20O3S	232.113 97	1
DAWS	[FA (17:1/2:0)] 8E-Heptadecenedioic acid	C17H30O4	149.107 29	1
DAWS	[PK] Chrysophanol	C15H10O4	254.057 73	15
DAWS	[Fv Methyl(9:1)] 3',4'-Methylenedioxy-[2'',3'':7,8]furanoflavanone	C18H12O5	308.068 91	1
DAWS	[Fv] Derriobtusone A	C18H12O4	292.073 66	2

DAWS	(R)-2-Methylimino-1-phenylpropan-1-ol	C10H13NO	163.099 92	3
DAWS	Anthralin	C14H10O3	226.062 96	2
DAWS	Strobamine	C17H19NO2	269.141 42	3
DAWS	6-Deoxyjacareubin	C18H14O5	310.084 23	1
DAWS	Harmine	C13H12N2O	212.095	2
DAWS	Dianthalexin	C14H9NO3	239.058 08	2
DAWS	N-hydroxyisoleucine	C6H13NO3	147.089 56	2
DAWS	[PR] Cacalol	C15H16O2	228.115 06	5
DAWS	Armepavine	C19H23NO3	313.167 78	6
DAWS	[FA oxo(13:0)] 2-oxo-tridecanoic acid	C13H24O3	228.172 83	4
DAWS	Annolobine	C17H15NO3	281.105 12	3
DAWS	Longifolonine	C17H15NO4	297.100 54	1
DAWS	2-Phenyl-4-benzopyron	C15H10O2	222.068	5
DAWS	Glu-Cys	C8H14N2O5S	250.062 84	2
DAWS	[Fv Hydroxy,methox] 4'-Hydroxy-2'-methoxychalcone	C16H14O3	254.094	5
DAWS	Epoxiconazole	C17H13FN3O Cl	329.072 25	1
DAWS	Asp-Cys	C7H12N2O5S	236.047 19	1
DAWS	[Fv] Naringenin	C15H12O5	272.068 36	26
DAWS	Prolyl-2-naphthylamide	C15H16N2O	240.126 22	2
DAWS	[PK] Chrysophanic acid 9-anthrone	C15H12O3	240.078 61	12

DAWS	Androsta-4,9(11)-diene-3,17-dione	C19H24O2	284.177 39	7
DAWS	Steroid O-sulfate	C18H24O4S	336.138 95	1
DAWS	Indole-3-acetate	C10H9NO2	175.063 37	12
DAWS	Gentisin	C14H10O5	258.052 68	5
DAWS	2-(2-Hydroxyphenyl)benzenesulfinate	C12H10O3S	234.035 27	1
DAWS	p-Cumylphenol	C15H16O	212.120 17	2
DAWS	1-Dodecanoyl-sn-glycerol	C15H30O4	274.214 37	4
DAWS	[Fv Methoxy(9:1)] 6-Methoxy-[2",3":7,8]furanoflavanone	C18H14O4	294.089 28	4
DAWS	Saphenic acid methyl ester	C16H14N2O3	282.100 5	1
DAWS	Biotinsulfone	C10H16N2O5 S	276.078 51	1
DAWS	Pyrene	C16H10	202.077 78	1
DAWS	4-Prenyldihydropinosylvin	C19H22O2	282.161 72	4
DAWS	Pelargonidin	C15H10O5	270.052 66	24
DAWS	[Fv] Licodione 2'-methyl ether	C16H14O5	286.084 06	38
DAWS	[PR] Crocetindial/ Crocetin dialdehyde	C20H24O2	296.177 42	8
DAWS	Tetradecanoylcarnitine	C21H41NO4	371.303 16	1
DAWS	Cyclo(L-Phe-L-Pro)	C14H16N2O2	244.121 08	3
DAWS	MG(0:0/16:1(9Z)/0:0)	C19H36O4	328.261 31	2
DAWS	[Fv] Xanthoangelol C	C22H22O5	366.146 91	8
DAWS	D-Sorbitol	C6H14O6	182.079 26	6

DAWS	Volkenin	C12H17NO7	287.10085	1
DAWS	(S)-Malate	C4H6O5	134.0216	4
DAWS	4-(2-Phenylethyl)phenol	C14H14O	198.10462	1
DAWS	Pterostilbene	C16H16O3	256.10982	9
DAWS	L-Serine	C3H7NO3	105.04265	3
DAWS	Palmiticamide	C16H33NO	255.25613	1
DAWS	gamma-Fagarine	C13H11NO3	229.07387	1
DAWS	Diethyl adipate	C10H18O4	202.12066	2
DAWS	Elaeokanine C	C12H21NO2	211.15728	1
DAWS	CPP	C9H9O3Cl	200.02423	3
DAWS	Indole-3-acetaldehyde	C10H9NO	159.06846	4
DAWS	[FA (11:0/2:0)] Undecanedioic acid	C11H20O4	216.13624	1
DAWS	[FA trihydroxy(18:0)] 9,10,18-trihydroxy-octadecanoic acid	C18H36O5	332.2567	4
DAWS	4,4'-Sulfonyldiphenol	C12H10O4S	250.03009	1
DAWS	Benzyl (2R,3S)-2-methyl-3-hydroxybutanoate	C12H16O3	208.11003	7
DAWS	Triethanolamine	C6H15NO3	149.10525	1
DAWS	Urocanate	C6H6N2O2	138.04296	3
DAWS	[FA (14:0/2:0)] Tetradecanedioic acid	C14H26O4	258.18331	6
DAWS	3-(Dimethylamino)propyl benzoate	C12H17NO2	207.12602	2
DAWS	myristic amide	C14H29NO	227.22499	2

DAWS	[FA (18:2)] 9S-hydroperoxy-10E,12Z-octadecadienoic acid	C18H32O4	312.230 43	39
DAWS	[FA (11:0/2:0)] Undecanedioic acid	C11H20O4	216.136 26	1
DAWS	[FA hydroxy(18:1)] 9,10-dihydroxy-12Z-octadecenoic acid	C18H34O4	314.246 12	12
DAWS	1,8-Diazacyclotetradecane-2,9-dione	C12H22N2O2	226.168 07	1
DAWS	Traumatic acid	C12H20O4	228.136 35	3
DAWS	Isogentisin	C14H10O5	258.052 6	5
DAWS	2-Oxooctadecanoic acid	C18H34O3	298.251 03	46
DAWS	(-)-Pellotine	C13H19NO3	237.136 41	3
DAWS	N(6)-(Octanoyl)lysine	C14H28N2O3	272.209 77	1
DAWS	[PR] Perillyl aldehyde	C10H14O	150.104 47	26
DAWS	Decanoylcholine	C15H31NO2	257.235 32	1
DAWS	[FA oxo(12:1)] 12-oxo-10E-dodecenoic acid	C12H20O3	212.141 42	9
DAWS	2-sec-Butylphenyl N-methylcarbamate	C12H17NO2	207.125 98	2
DAWS	L-Citrulline	C6H13N3O3	175.095 78	3
DAWS	13-Hydroxylupanine	C15H24N2O2	264.183 7	5
DAWS	[FA oxo(15:0)] 4-oxo-pentadecanoic acid	C15H28O3	256.204 04	5
DAWS	Darlingine	C13H17NO2	219.125 95	1
DAWS	[ST (3:0)] (5Z,7E)-(1S,3R,24R)-22-oxa-9,10-seco-5,7,10(19)-cholestatriene-1,3,24-triol	C26H42O4	418.308 08	10
DAWS	16-hydroxypalmitate	C16H32O3	272.235 29	18
DAWS	(10S)-Juvenile hormone III diol	C16H28O4	284.198 82	1

DAWS	[FA hydroxy(11:0)] 11-hydroxy-undecanoic acid	C11H22O3	202.1571	6
DAWS	Sulcatone	C8H14O	126.10447	6
DAWS	Cochlearine	C15H19NO3	261.13638	3
DAWS	3-Mercaptolactate	C3H6O3S	122.00381	1
DAWS	17beta-Nitro-5alpha-androstane	C19H31NO2	305.2356	1
DAWS	1,6,6-Trimethyl-2,7-dioxabicyclo[3.2.2]nonan-3-one	C10H16O3	184.11001	9
DAWS	[FA hydroxy(16:0/2:0)] 9-hydroxy-hexadecan-1,16-dioic acid	C16H30O5	302.20967	2
DAWS	Diethyl (2R,3R)-2-methyl-3-hydroxysuccinate	C9H16O5	204.09997	2
DAWS	[FA trihydrox] 9S,11R,15S-trihydroxy-2,3-dinor-13E-prostaenoic acid-cyclo[8S,12R]	C18H32O5	328.22534	6
DAWS	[FA hydroxy(18:0)] 2S-hydroxy-octadecanoic acid	C18H36O3	300.26664	27
DAWS	Pantothenol	C9H19NO4	205.13149	1
DAWS	10,16-Dihydroxyhexadecanoic acid	C16H32O4	288.22982	6
DAWS	2-Methoxyhexadecanoic acid	C17H34O3	286.25088	4
DAWS	(R)-Lactate	C3H6O3	90.031754	6
DAWS	[FA oxo(19:0)] 10-oxo-nonadecanoic acid	C19H36O3	312.26644	4
DAWS	3-Oxododecanoic acid	C12H22O3	214.15708	10
DAWS	Tributyrin	C15H26O6	302.17324	1
DAWS	[FA amino(9:0)] 9-amino-nonanoic acid	C9H19NO2	173.14167	4
DAWS	[FA hydroxy(7:0)] 2-hydroxy-heptanoic acid	C7H14O3	146.09438	8
DAWS	[FA (16:0/2:0)] Hexadecanedioic acid	C16H30O4	286.2145	5

DAWS	[FA (11:0/2:0)] Undecanedioic acid	C11H20O4	216.136 29	1
DAWS	[FA hydroxy(15:0)] 2-hydroxy-pentadecanoic acid	C15H30O3	258.219 71	10
DAWS	Ethyl (R)-3-hydroxyhexanoate	C8H16O3	160.110 02	14
DAWS	Octadecanoic acid	C18H36O2	284.271 62	12
DAWS	methylmercaptoethanol	C3H8OS	92.0293 82	2
DAWS	10-Hydroxydecanoic acid	C10H20O3	188.141 36	13
DAWS	[FA hydroxy(9:0)] 2-hydroxy-nonanoic acid	C9H18O3	174.125 68	8
DAWS	[SP (14:0)] 1-deoxy-tetradecasphinganine	C14H31NO	229.240 59	1
DAWS	[FA amino(12:0)] 12-amino-dodecanoic acid	C12H25NO2	215.188 61	2
DAWS	3-Hydroxytetradecanedioicacid	C14H26O5	274.178 17	1
DAWS	2-Octenedioicacid	C8H12O4	172.073 69	4
DAWS	Apiole	C12H14O4	222.089 31	8
DAWS	Ethyl 3-oxohexanoate	C8H14O3	158.094 34	9
DAWS	[FA dimethyl(13:0)] 2,5-dimethyl-2E-tridecenoic acid	C15H28O2	240.209 12	5
DAWS	[SP hydroxy,hydroxy,methyl(10:2/2:0)] 6R-(8-hydroxydecyl)-2R-(hydroxymethyl)-piperidin-3R-ol	C16H33NO3	287.245 89	1
DAWS	[FA trihydroxy(18:1)] 9S,12S,13S-trihydroxy-10E-octadecenoic acid	C18H34O5	330.241 01	10
DAWS	Leucyl-leucine	C12H24N2O3	244.178 62	4
DAWS	Botrydial	C17H26O5	310.178 3	2
DAWS	(S)-3-Methyl-2-oxopentanoic acid	C6H10O3	130.063 09	17
DAWS	[FA (18:3)] 13S-hydroperoxy-9Z,11E,14Z-octadecatrienoic acid	C18H30O4	310.214 38	19

DAWS	trans-Cinnamate	C9H8O2	148.052 48	7
DAWS	Nonanoic acid	C9H18O2	158.130 72	11
DAWS	Prostaglandin F1alpha	C20H36O5	356.256 64	6
DAWS	[FA hydroxy(13:0)] 2-hydroxy-tridecanoic acid	C13H26O3	230.188 31	6
DAWS	6-Hydroxynicotinate	C6H5NO3	139.026 98	7
DAWS	Hexylamine	C6H15N	101.120 46	3
DAWS	4-8dimethylnonanoylcarnitine	C18H35NO4	329.256 57	2
DAWS	[FA oxo(16:0)] 3-oxo-hexadecanoic acid	C16H30O3	270.219 65	19
DAWS	[FA oxo(5:2/5:0/6:0)] (1R,2R)-3-oxo-2-pentyl-cyclopentanehexanoic acid	C16H28O3	268.203 66	2
DAWS	trans-4-Hydroxycyclohexanecarboxylate	C7H12O3	144.078 69	8
DAWS	Phenethylamineglucuronide	C14H19NO6	297.121 47	1
DAWS	[FA amino(11:0)] 11-amino-undecanoic acid	C11H23NO2	201.172 95	2
DAWS	Carpacin	C11H12O3	192.078 71	6
DAWS	Benzoate	C7H6O2	122.036 79	6
DAWS	Trimethylaminoacetone	C6H13NO	115.099 73	2
DAWS	5-(2'-Formylethyl)-4,6-dihydroxypicolinate	C9H9NO5	211.048 22	5
DAWS	(S)-2-Aceto-2-hydroxybutanoate	C6H10O4	146.057 99	16
DAWS	Dodecanamide	C12H25NO	199.193 7	1
DAWS	[FA oxo(8:0)] 3-oxo-octanoic acid	C8H14O3	158.094 36	9
DAWS	Methanesulfonic acid	CH4O3S	95.9879 33	1

DAWS	Chlorate	HO3Cl	83.9614 33	1
DAWS	[FA (16:0)] N-hexadecanoyl-taurine	C18H37NO4S	363.244 01	1
DAWS	Glycerol	C3H8O3	92.0473 83	1
DAWS	omega-Cyclohexylundecanoic acid	C17H32O2	268.240 36	18
DAWS	Isocitrate	C6H8O7	192.027 26	12
DAWS	[PR] (+)-3-longipinen-5-one	C16H24	108.093 92	1
DAWS	[FA (5:1)] 2-(2-cyclopentenyl)-ethanoic acid	C6H8O2	112.052 37	17