Microbial Lignin Degradation in the Environment

Bruce Young Steel

Doctor of Philosophy (PhD)

University of East Anglia, Norwich, UK

School of Biological Sciences

April 2023

Abstract

Lignocellulosic biomass has attracted attention as a renewable industrial feedstock due to its abundance in nature and wide range of potential applications. Current conversion processes, such as in biofuel production, are complicated by the presence of lignin within lignocellulose. Research has identified many environmental microorganisms capable of degrading lignin, which could potentially be exploited for use in bioprocessing, however, little is known of how microorganisms, specifically bacteria, contribute towards the degradation of lignin. Therefore, the aim of this project was to explore lignin degradation in a wide range of environments and to identify the main microbial players, including key enzymes and biochemical pathways that contribute to the degradation of lignin in these environments. This was done via biochemical and molecular biological approaches, such as stable isotope probing using ^{13}C-labelled lignin, rRNA gene amplicon sequencing, and metagenomics. In this study, we investigated microbial lignin degradation in a wide range of oxic and anoxic environments, including soil, herbivore animal faeces, compost and lake sediment. Metagenomic sequencing of the environmental samples identified distinct microbial communities with the genetic potential to activate diverse lignin degradative mechanisms. Lignin-degrading bacterial isolates from these environments and their phylogenetic characterization and ligninolytic enzyme activity analysis revealed varying utilization of oxidative enzymes by different isolates. Furthermore, gas chromatography-mass spectrometry analysis of lignin oxidised phenols from bacterial isolate cultures showed considerable variation compared to control cultures, suggesting specific and diverse lignin breakdown. These results provide new insights into microbial degradation of recalcitrant plant material in the environment.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Meaning
2,4-DCP	2,4-dichlorophenol
2,6-DMP	2,6-dimethoxyphenol
ABTS	2,2’-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid
ASV	Amplicon sequencing variant
BLAST	Basic local alignment search tool
bp	Base pairs
BSA	Bovine serum albumin
BSTFA	bis(trimethylsilyl)trifluoroacetamide
C54	Compost sample 5-4
C55	Compost sample 5-5
C61	Compost sample 6-1
CAZyme	Carbohydrate-active enzyme
cDNA	Complementary deoxyribonucleic acid
CsCl	Cesium chloride
CsTFA	Cesium trifluoroacetate
CLIMB	Cloud Infrastructure for Microbial Bioinformatics
CS	Compost soil
CsTFA	Caesium trifluoroacetate
CTAB	hexadecyltrimethylammonium bromide
DNA	Deoxyribonucleic Acid
dsDNA	Double-stranded deoxyribonucleic acid
DyP	Dye-decolourising peroxidase
EDTA	Ethylenediaminetetraacetic acid
EF	Elephant faeces
EF1	Elephant faeces sample 1
EF2	Elephant faeces sample 2
EtoAc	Ethyl Acetate
GB	Gradient buffer
GC-MS	Gas chromatography-mass spectrometry
HCl	Hydrochloric acid
HMM	Hidden markov models
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidase
HS	Hickling sediment
HS(c)	Hickling sediment sample from broad centre

HS(s)	Hickling sediment sample from broad shore
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Lysogeny broth
LCA	Lowest common ancestor
LDA	Lignin-degrading auxiliary enzymes
LDPC	lignin-derived phenolic compound
LiP	Lignin peroxidase
LMCO	Laccase-like multicopper oxidase
LME	Lignin-modifying oxidative enzymes
MAG	Metagenome-assembled genome
MnP	Manganese-dependent peroxidase
MSM	Mineral salt medium
NCBI	National Centre for Biotechnology Information
NGS	Next generation sequencing
OD_{600}	Optical density at 600nm
OG	Olduvai Gorge
OUT	Operational taxonomic unit
PCoA	Principle coordinate analysis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PLFA	Phospholipid fatty acid
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic Acid
RNase	Ribonuclease
rRNA	Ribosomal ribonucleic acid
SDS	sodium dodecyl sulphate
SIP	Stable isotope probing
SS	Savanna soil
SSU	Small subunit
TAE	Tris-acetate-EDTA
TCA	Tricarboxylic acid
TIC	Total ion chromatogram
TM	Termite mound
VP	Versatile peroxidase
WGS	Whole genome sequencing
YEMS	Yeast extract mineral salt
ZS	Zebra-grazed soil

Acknowledgments

I would firstly like to thank all the members of my previous and current supervisory teams for their unparalleled support and advice throughout the duration of my PhD project. Thanks to my primary supervisor, Prof. Jonathon Todd, who always provided fantastic supervision and support throughout my project and always made himself available to give me advice (despite being extremely busy). Thanks to my secondary supervisor, Dr Julia de Rezende, who was there at the darkest of times and always reminded me of my self-worth and ability. I would not have submitted this thesis if it was not for Julia. And thanks to Prof. Theodore Henry for his wise words and support during the writing of this thesis, who was also always available to provide positive encouragement and advice.

I would also like to thank my colleagues at Heriot Watt, University of East Anglia and the Lyell Centre, who provided me with encouragement and advice throughout my project. Thanks to Dr Clayton Magill for his constant positivity and support throughout my project, who was always there to give guidance and reassurance, regardless of how busy he was. Thanks to Dr Xiang Shi (my lab buddy), with whom I spent many of the days/nights in the lab with, for his advice both inside and outside the lab. Thanks to Dr Lavinia Stancampiano for teaching me the biochemistry methods in my project and always being the level-headed one. Thanks to Dr Julianne Bischoff, who helped me countless times in the lab and was more of a friend than technician. Thanks to Prof. Gabriela Medero for guiding me through the toughest of times.

Without the constant support from my friends and family, this PhD would not have been possible. Thanks, especially, to my partner, Emma-Lee Nicholls, who was there when I started my PhD and stuck with me until the end. Her daily support and positivity got me through the best and worst of times during my project, and I could not have done it without her. She was constantly there to listen to me and support me, no matter what. I would also like to thank my parents, Brian and Sharon Steel, for their continued love and support throughout my life and education. They have supported me through every aspect of life and I am eternally grateful to them.

Finally, I would like to thank my industrial partner Recircle, for their financial contributions and placement opportunity within this project. I would like to thank EnvEast/Aries for the DTP studentship opportunity and the training provided throughout my PhD. I would also like to thank the colleagues who contributed samples towards this project - Dr Ryan Pereira, Dr Andrew Sweetman and Dr Thomas Aspray.

1 Chapter 1: Introduction

1.1 Lignin and Lignocellulose – Structure

The plant cell is protected by a strong cell wall composed of a complex network of carbon-based polymers which contribute towards the robust nature of these cells. The primary molecular components of the plant cell wall (up to 90% composition) are polysaccharides - cellulose, hemicellulose and pectin - followed by the aromatic heteropolymer lignin, amongst other structural molecules (Bugg *et al*., 2011b; de Vries *et al*., 2001; Kameshwar and Qin, 2016). Cellulose is the most abundant polysaccharide present within the cell wall, comprising 40-50% of plant material (dry weight), and exists in the form of organized, linear molecules called microfibrils. Cellulose appears as long β(1,4)-linked chains of glucose subunits, known as cellobiose, attached through hydrogen bonds to form these crystalline microfibrils, with amorphous regions spread through the fibrils (Kameshwar and Qin, 2016; Martínez *et al*., 2005; Souza, 2013). Amongst the network of cellulose microfibrils and lignin polymeric matrix are the hemicelluloses, or collectively hemicellulose, the second most abundant polysaccharides in the plant cell wall. The hemicelluloses encompass a range of heterogeneous sugar polymers, distinguishable by the primary polysaccharides present within their molecular backbone structure (Souza, 2013). Hemicellulose is generally composed of hexoses, pentoses and sugar acids (most notably, glucuronic acid), which act as bonding agents between cellulose and lignin via cross-linkages, inciting overall robustness within the plant cell wall matrix (Malherbe and Cloete, 2002; Laureano-Pérez *et al.*, 2005; Kameshwar and Qin, 2016).

The second most abundant molecule within the cell wall is lignin – a highly complex and heterogeneous aromatic polymer. Lignin consists of three recurring phenylpropanoid polymer subunits (p-hydroxyphenyl, guaiacyl and syringyl). These subunits are derived from hydroxycinnamoyl alcohol monomers (also known as monolignols) - p-coumaryl, coniferyl and sinapyl alcohols, respectively - which are either linked via carbon-carbon bonds or ether (carbon-oxygen-carbon) bonds (see **Figure 1** below) (Martínez *et al*., 2005; Ruiz-Dueñas and Martínez, 2009; Kameshwar and Qin, 2016; Datta *et al*., 2017). The types and abundance of monolignols present within lignin, as well as the types of bonds between these subunits, can vary between different plant species, with non-woody plant and grass lignin being mostly p-hydroxyphenyl-dominant and softwood (coniferous) lignin being mostly guaiacyl-dominant (Bugg *et al.*, 2011a; Mathews *et al.,* 2015; Janusz *et al.*, 2017). Lignin plays a fundamental role in the cell wall, providing strength to the network with its structural complexity and strong carbon bonds. The high-

molecular weight, strong carbon bonds, water insolubility and three-dimensional nature of the lignin polymer also make it extremely recalcitrant and therefore able to withstand a range of chemical and biological degradation mechanisms (Malherbe and Cloete, 2002; Vikman *et al*., 2002; Martínez *et al*., 2005; Souza, 2013; Kai *et al*., 2016;)

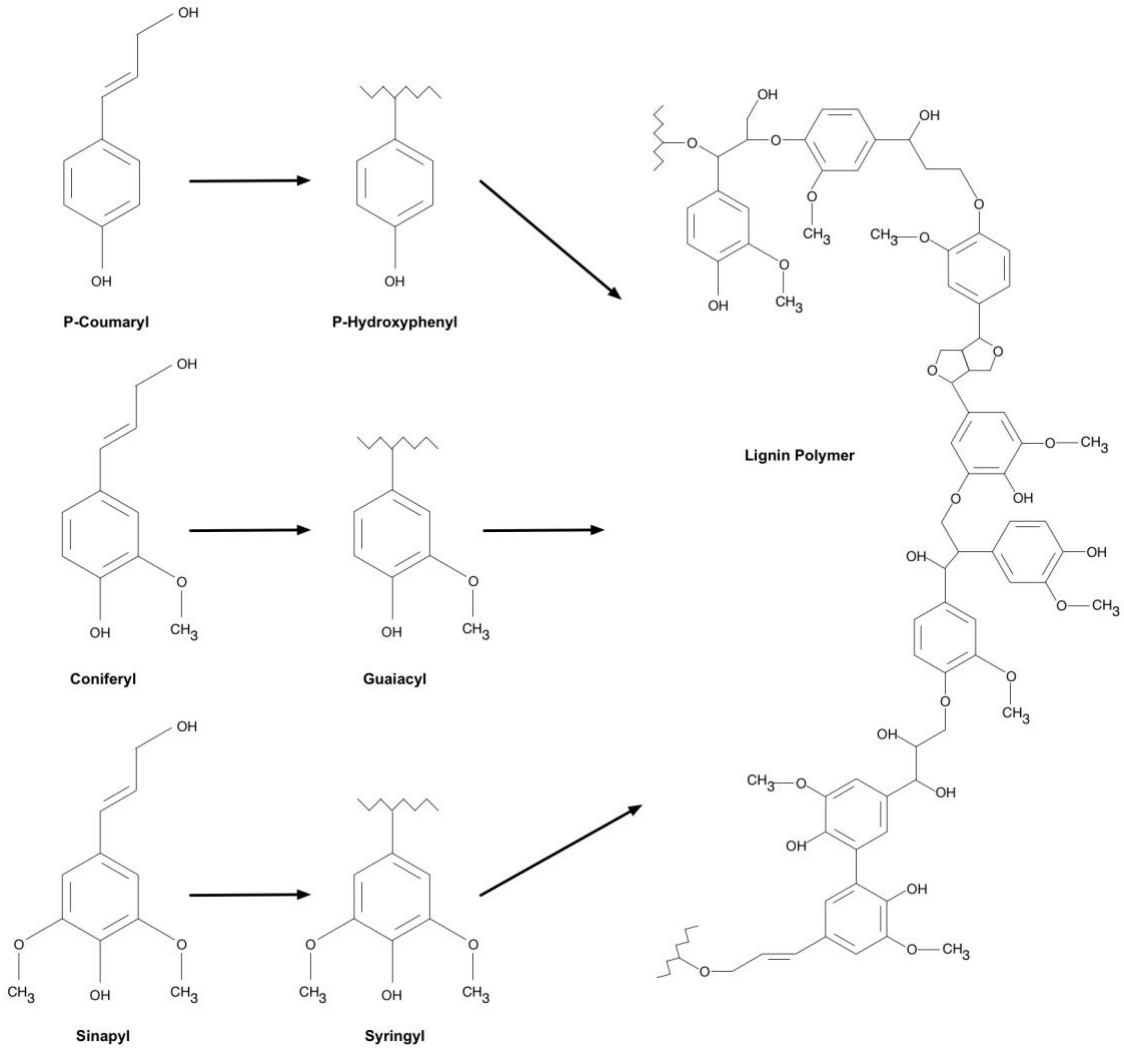

Figure 1 *– Structure and composition of lignin within the plant cell wall. Structure and composition can vary between plants. Adapted from Kai et al., 2016. and Bugg et al., 2011b.*

All of the components within the plant cell wall interact with each other through various chemical bonds to form a network. Lignin interacts closely with cellulose, with the aid of hemicellulose, through cross-linkages and strong covalent bonds to create an amorphous polymeric matrix known as lignocellulose (Ahmad *et al.*, 2011; Bugg *et al*., 2011b; Souza, 2013; Laureano-pérez, 2014) (**Figure 2**). Lignocellulose (or plant biomass) is the most abundant organic polymer on the planet (Kumar *et al.,* 2008; Mathews *et al.,* 2015). Composition of lignocellulose varies between different plant types. Softwoods contain approximately 40% cellulose, 30% hemicellulose and 30%

lignin, whereas hardwoods contain approximately 50% cellulose, 30% hemicellulose and 20% lignin. Due to the global affluence of lignocellulosic material, as well as the increasing drive for renewable feedstocks, there is a strong demand to identify potential uses and develop methods of utilizing lignocellulosic biomass, aside from the 'traditional' ways in which this biomass is processed and used, such as in the pulp and paper industry.

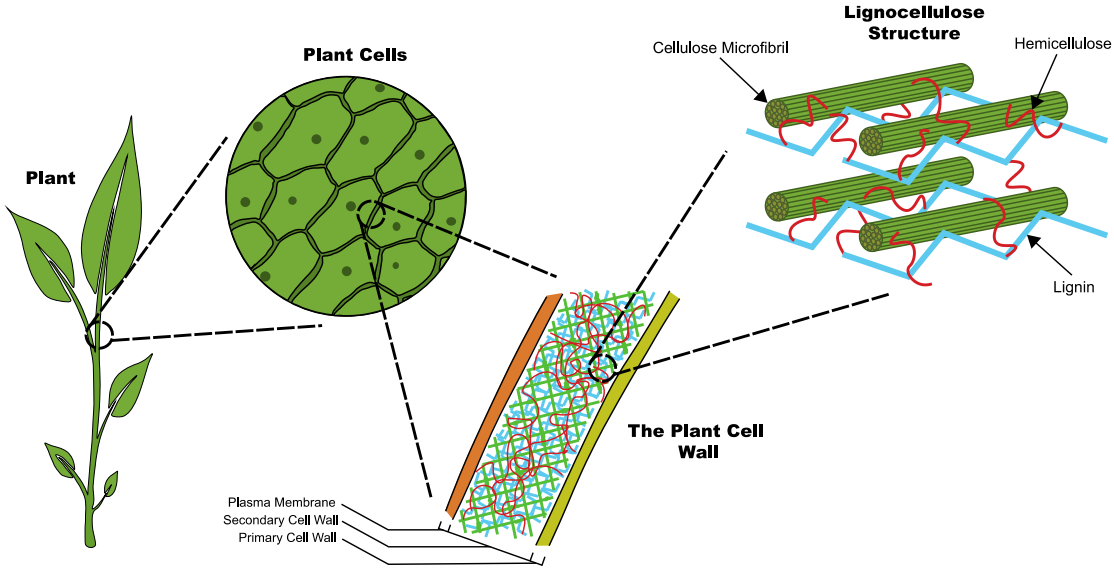

Figure 2 *- Lignocellulose structure within the plant cell wall. Modified from Adapa et al. (2009).*

1.2 Lignocellulose Utilization

1.2.1 Present

In addition to its 'original' use as fuel in log fires, lignocellulosic biomass, or more specifically the sugars contained within lignocellulose (cellulose and hemicelluloses), have been used in the paper and chemical industries for decades, and more recently in small-scale biofuel industries. Research has begun to reveal an expanding variety of applications for lignocellulose, some of which involve lignin utilization, which was previously thought of as a waste by-product. This is because the presence of lignin usually restricts biomass processing, due to its strong nature and close association with polysaccharides within the lignocellulose molecule. These characteristics, along with a variety of other factors, such as hemicellulose content, often inhibit enzyme and chemical activity during lignocellulose conversion processes, for example preventing access to and release of the polysaccharides during bioethanol production (Malherbe and Cloete, 2002; Sun *et al*., 2016). Lignin can also decrease the quality of the final product, for example in paper production, where high lignin levels within the final paper product result

in decreased strength and brightness, and therefore quality (Kumar *et al*., 2008; Mathews *et al*., 2015; Thakur and Thakur, 2015). It is for these reasons that current lignocellulose conversion processes involve an initial pretreatment stage to remove lignin and hemicellulose from the biomass. The methods involved in pretreatment are very harsh, commonly requiring powerful machinery and strong chemicals in order to successfully remove these molecules. Lignocellulose pretreatment and the processes which follow this, however, result in substantial underutilization of the lignocellulose molecule, creating a number of waste streams. These waste streams contain cellulose, hemicellulose and lignin – all of which can be converted into valuable products – therefore, more complete utilization is a strong focus of current research (Mathews *et al*., 2015).

Cellulose, lignin and hemicellulose each have a variety of properties which make them attractive to industry. Lignin is extremely rigid, highly resilient to biodegradation and has high thermal stability due to its complex molecular structure and strong carbon bonds, and can be converted into a range of valuable products. Depending on the processes used, lignin-derived compounds can be used as stabilizing agents in foam fire extinguishers, binding agents in glass wool insulation, and pigment which reinforces rubber during compounding, to name just a few examples (Agrawal *et al*., 2014; Thakur and Thakur, 2015; Guerriero *et al*., 2016; Kai *et al*., 2016;). Cellulose and hemicelluloses (primarily xylan) are the most abundantly available polysaccharides on Earth and have applications in a wide range of industries, such as paper, bioethanol and textile production, as well as in the production of various biochemicals. For example, cellulose can be converted to fructose sweetener for food, and, similar to xylan, has additional applications in adhesive and plastic production. (Thakur and Thakur, 2015).

Feedstocks for industrial lignocellulose conversion come from a variety of different sources, however, softwood and hardwood lignocellulose are much more challenging to degrade in these processes compared to agricultural lignin waste (e.g. wheat straw) due to their higher lignin content and reduced porousness (accessible surface area) (Palonen *et al*., 2004; Janusz *et al*., 2017). The three main types of conversion processes to turn lignocellulosic biomass components into usable products are mechanical, chemical and thermal processing, followed by biological processing which will be the focus of this review (Bridgwater, 2006). In order to break the hydrogen bonds and strong covalent bonds between the molecules within lignocellulose and remove the lignin-hemicellulose complex, most processing methods require initial physical reduction. It is difficult to effectively access and use the polysaccharides contained within lignocellulose without removing lignin, and whilst removing the cellulose fibers from the lignocellulose, the

hemicellulose and lignin molecules are either degraded or completely lost, resulting in the afore mentioned waste streams (Gabrielii *et al*., 2000; Martínez *et al*., 2005; Kameshwar and Qin, 2016).

Generally, mechanical pretreatment is used to reduce the raw plant biomass, breaking the large lignocellulose polymer into smaller particles. These particles have a larger surface area, making them more accessible to the agents used in later chemical pretreatment (Mathews *et al*., 2015). The most commonly used chemicals for removing lignin and hemicellulose are sodium sulfide and sodium hydroxide, which break the polymer apart in the traditional Kraft pulping process (see **Figure 3** below) (Zakzeski *et al*., 2010). The material can be exposed to high temperatures (thermal pretreatment), making the molecular constituents water-soluble and susceptible to hydrolysis. Acid or alkali pretreatments can also be used to precipitate or solubilize lignin, respectively, after which the mixture can be screened to collect the cellulose fibers (or other desired polysaccharides). The cellulose fibers are subsequently washed and bleached to remove any residual, undesirable molecules, and the resulting 'black liquor' (containing the lignin collected from screening) is normally concentrated by evaporators and finally burned in recovery boilers to produce energy (Zakzeski *et al.*, 2010; Mathews *et al.,* 2015). Black liquor primarily contains lignin and approximately 60% of the polysaccharides which were originally present within the lignocellulose biomass, therefore resulting in significant underutilization of the biomass. Wastewater from washing and bleaching the cellulose contains a variety of chemicals which, when disposed of, are harmful to plants and animals in the surrounding ecosystems, having a negative impact to the environment. Furthermore, carrying out these processing methods is economically challenging as the mechanical and thermal stages require high energy input and the chemicals involved can be expensive, with some pulp and paper plants having dedicated chemical production facilities on-site (Mathews *et al.,* 2015). There is therefore an abundance of industrial research being carried out, tasked with addressing these issues.

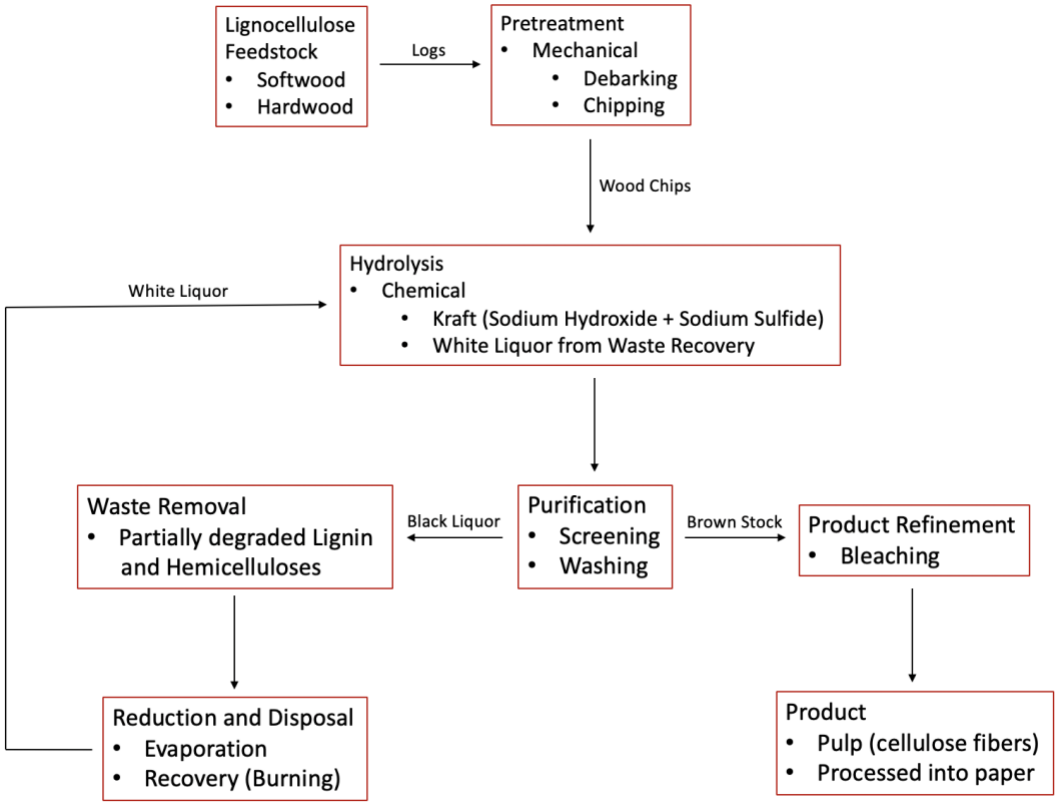

Figure 3 *- Typical (Kraft) Pulping Process. Adapted from Mathews et al., 2015.*

1.2.2 Biofuels

The rapidly deteriorating state of Earth's environments and ecosystems due to climate change, accelerated by the ever-increasing consumption of fossil fuels, has become one of the most important challenges of science and society. It has been a strong driving force for the development of sustainable and renewable energy alternatives, such as biofuels – fuels produced from biomass (Bridgwater, 2006; Kumar *et al*., 2009; Kameshwar and Qin, 2016). Lignocellulose, being the most abundant planetary source of polysaccharides, has become the feedstock-of-choice for the production of these fuels, with different types of lignocellulosic biomass being used in different processes, and the most common feedstock being derived from dedicated 'energy crops', such as sugarcane bagasse and corn stalks (Souza, 2013). Bioethanol (fuel produced from fermenting plant-based sugars) has attracted the most attention and currently has the largest market-value as a fuel additive and, more recently, 'neat' fuel (i.e. using bioethanol as fuel alone). This is due to its renewability, lower carbon emissions compared to gasoline, and the ability to be mixed with fossil fuels, reducing environmental impact (Kumar *et al*., 2008; Sun *et al*., 2016) Bioethanol, and less commonly, biopropanol and biobutanol are widely produced and used as a fuel additive

in the USA (from corn) and Brazil (from sugarcane), and some European countries (Galbe and Zacchi, 2007; Kumar *et al*., 2009; Chen and Fu, 2016). Bioethanol can be produced from starch sugars, however, cellulose is more appealing due to its abundance, despite the requirement for more severe conditions for conversion of these materials (Sun *et al*., 2016).

The most common lignocellulose feedstock for bioethanol production is agricultural waste from crops grown purposely for fuel production. Waste such as wheat straw, corn stalks and sugarcane bagasse from these high-cellulose, low-lignin crops are easier to process than plants with higher lignin content (Kumar *et al*., 2009; Janusz *et al*., 2017). These crops are renewable feedstocks, however, there is growing concerns over the sustainability of using them. It has been argued that the exponential growth of Earth's population and the positively correlated demand for food production means agricultural plants, and the land used to grow these crops, are better used for agriculture. Lack of sustainability seems to be the main impeding factor holding back successful food crop biomass utilization, which is why increasing research has been carried out to investigate alternative sources of lignocellulosic biomass (Sun *et al*., 2016). Alternative lignocellulose sources, such as forestry (softwood/hardwood) debris and industrial waste (such as concentrated black liquor from Kraft pulping), are less common sources of lignocellulose for biofuel production due to the high lignin content and resulting process challenges involved in converting these materials. These sources would however significantly increase the availability of lignocellulose biomass, with a lesser effect on human agriculture, and perhaps reduce the costs involved in producing biofuels (Wan and Li, 2010; Mathews *et al*., 2015).

A range of biofuel production processes exist such as: mechanical processing used to produce rapeseed oil; thermal processing (pyrolysis and gasification) used to produce bio-oil and gas fuel; and biological processing (digestion and fermentation) used to produce biogas and bioethanol (Bridgwater, 2006). Similar to the conversion process of biomass to pulp, the typical bioethanol production process begins with a pretreatment stage, in order to expose and release the polysaccharides encased within the hemicellulose-lignin barrier and break the crystalline cellulosic fibers (Kumar *et al*., 2009; Zakzeski *et al*., 2010). These pretreatment methods are the most expensive, and can be similar to the methods used in pulp and paper production discussed earlier, comprising mechanical reduction, strong chemicals and high temperatures, as well as physicochemical (steam and ammonia fiber explosion) and electrical treatments (Kumar *et al*., 2009; Mathews *et al*., 2015). A common example is steam explosion, in which the biomass is treated with steam under high temperature and pressure, after which the

pressure is rapidly decreased, causing the material to decompress explosively. Release of acetic acids during decompression partially hydrolyses hemicellulose, and the high temperature partially depolymerizes lignin (Kumar *et al*., 2009). By exposing and partially degrading the cellulose, the sugars can then be subjected to hydrolysis which breaks the bonds holding the cellulose-hemicellulose-lignin molecule together and fragments the cellulose into glucose monomers, paving the way for microbial fermentation and conversion into ethanol. Hydrolysis methods involve acid or alkaline treatments (dilute or concentrated) or cellulase enzyme (primarily derived from fungi and bacteria) treatment which fragment the molecule, solubilizing hemicellulose and precipitating lignin (Mathews *et al*., 2015).

Generally, alkaline pretreatment is the preferred method as it requires milder process conditions (temperature and pH) and reduced polysaccharide degradation compared to acid pretreatment (Kumar *et al*., 2009). Cellulase enzyme hydrolysis can also be carried out in much milder pH and temperature conditions (pH 5, 50°C) as these enzymes are commonly produced by microorganisms which inhabit, for example, the stomachs of animals such as cows (Kumar *et al*., 2009; Sun and Cheng, 2002). As well as mild process conditions (low temperature and agitation, flexible pH, etc.), enzyme hydrolysis is beneficial in that the enzymes are renewable and sustainable, given that they are naturally produced by fungi and bacteria. The disadvantages involved with enzyme hydrolysis are that the high recalcitrance of lignocellulose can reduce enzyme efficiency by preventing access to the cellulose. The process time is significantly higher compared to chemical and physiochemical treatments, which can make enzyme hydrolysis unattractive to industry (Laureano-Pérez *et al.*, 2005; Kumar *et al.*, 2009).

Biofuel production, specifically bioethanol, has become well established due to a high R&D focus in this area, however there are still limitations preventing widespread use of these fuels. The processes involved in producing biofuels can be complex, normally involving several different pretreatment steps before hydrolysis and fermentation, in order to retrieve the highest fuel yield possible from the feedstock. Furthermore, these processes vary in effectiveness between different lignocellulosic materials, due to composition (e.g. hemicellulose and lignin content, cellulose crystallinity, etc.) and enzyme/chemical specificity. Upstream pretreatment processes can produce tertiary by-products which inhibit or reduce the activity of downstream hydrolysis agents (e.g. phenolic compounds and weak acids produced by steam explosion) therefore requiring additional washing of the lysate before hydrolysis (Kumar *et al*., 2009; Sun *et al*., 2016). Like the process of pulp and paper production, lignocellulose is heavily underutilized in biofuel production as many of the methods partially degrade or destroy hemicellulose

and lignin. Both hemicellulose and lignin have potential for production of high-value chemicals, for example, the production of aromatic monomers and other intermediates from lignin degradation can be used to synthesize other fuels and surfactants (Agrawal *et al.,* 2014; Ahmed *et al.,* 2016). In addition to these limitations, there is also the afore mentioned concern of feedstock sustainability, with the majority of raw materials coming from dedicated 'energy crops' and agricultural waste which is arguably better used in food and animal feed production, as opposed to biofuel production. The issues concerning, not only biofuel production, but overall industrial conversion of lignocellulose, open up many fields for further research to simplify and improve these processes, with a lot of research results pointing towards the use of microorganisms for both upstream and downstream processing of these materials.

1.3 Microbial Lignin Degradation

For an industrial process to be feasible, the methods involved must be cost-effective, energy-efficient and scalable to industrial volumes. Current chemical, physical and thermal methods of lignocellulose conversion have high energy and economical costs, as well as being harmful to the environment (Mathews *et al*., 2015). Biological degradation processes are therefore a promising option for the future, as they have the potential to merge efficiency, cost-effectiveness and environmental benefits, through the use of microorganisms which naturally use the lignocellulose as a carbon source through enzymatic degradation (Kumar *et al*., 2009; Ahmed *et al*., 2016; Tian *et al*., 2016). As mentioned previously, the current conversion of lignocellulose into commercial products requires vigorous upstream process treatments such as strong toxic chemicals and high energy-dependent equipment in order to remove lignin. However, the use of microorganisms and their enzymes may present alternative methods of dealing with the recalcitrance of the polymer, improving or possibly replacing current process steps.

Some current lignocellulose bioconversion processes already incorporate microorganisms and their enzymes, using them in downstream stages (after lignin has been removed), for final conversion of extracted materials into usable products (e.g. fermentation in bioethanol production). These processes however still require pretreatment stages to remove lignin, which can be challenging for current commercial microorganisms/enzymes, hindering upstream bioprocessing. With this being said, there has been an increase in research regarding microorganisms in the natural environment which are capable of effectively degrading the lignin molecule and their potential uses in lignocellulose conversion (Ruiz-Dueñas and Martínez, 2009). The ability of microorganisms to degrade plant biomass in the natural environment depends on a

number of factors. These factors include whether or not the specific species is able to degrade lignin (determined by enzyme capabilities), as well as the type and abundance of lignin within the biomass, with lower percentages of lignin (such as in marine plants) being easier for microorganisms to degrade (Malherbe and Cloete, 2002; Ahmad *et al*., 2010; Souza, 2013). Furthermore, it has been shown that some microorganisms are more capable than others, and that 'key' degraders vary between different environments. Research on ligninolytic microorganisms has primarily been focused on microbes derived from soil, and more recently from niche environments, such as from wood-feeding insects (Bugg *et al*., 2016).

1.3.1 Fungal Lignin Degradation

The majority of research on microbial lignin degradation has been focused on fungi, specifically the white-rot and brown-rot fungi (Basidiomycetes), which are known for being highly effective lignin degraders in that they are able to degrade lignin at a higher rate compared to most of their bacterial counterparts (Vikman *et al*., 2002; Janusz *et al*., 2017). In addition to white-rot species are the soft-rot and stain fungi, which are less efficient but still able to degrade lignin (Martínez *et al*., 2005). Microbial lignin degradation is a highly complex process involving multiple enzymatic pathways which target specific chemical structures and bonds within the lignin polymer. Degradation occurs through the action of extracellular free radicals and/or a range of different intracellular/extracellular oxidative enzymes dedicated to breaking down lignin, working synergistically, much like enzymes involved in cellulose degradation (Guerriero *et al*., 2016; Kameshwar and Qin, 2016). These enzymes and free radicals are used as markers for microbial lignin degradation.

Brown-rot fungi degrade lignocellulose via non-enzymatic free radical production as opposed to white-rot fungi, which use a range of enzymes, in turn creating free radicals leading to further degradation (Datta *et al*., 2017). A study carried out by Kerem *et al.* (1999) found that brown-rot fungi break down the aromatic rings of lignin monomers by producing iron ions and hydrogen peroxide, and subsequently hydroxyl radicals (Fenton reaction), which are responsible for lignin hydrolysis and partial oxidation (Kerem *et al*., 1999; Hrdlicka *et al*., 2014). Whereas, some white-rot fungi, such as *Rigidoporous lignosus* and *Fomes sclerodermeus*, produce phenol oxidase (laccase) and peroxidase enzymes, which degrade lignin through oxidation, as well as produce hydroxyl radical by-products, further degrading lignin (Papinutti *et al.,* 2003; Souza, 2013; Datta *et al.*, 2017). For this reason, white-rot fungi are able to degrade lignocellulose to a greater extent and much faster than brown-rot fungi. However, recent studies suggest that some

brown-rot fungi are able to produce extracellular enzymes when cultured in certain conditions, suggesting that ligninolytic capability is species-dependent and the use of different degradation mechanisms may be dependent on environmental factors (Lee *et al*., 2004; Martínez *et al*., 2005). These ligninolytic enzymes are only produced by capable microbes as a response to carbon and nitrogen limitation, allowing some fungi to be selective in their carbon source (Kumar *et al*., 2009).

Studies on biological pretreatment of lignocellulose have primarily involved the use of these white and brown-rot fungi, due to their superior degradative capabilities, however efficient degradation by these processes requires significantly more time than regular processing methods (Kumar *et al*., 2009) (**Table 1**). An example being the use of white-rot fungi species (*Ceriporiopsis subvermispora and Cyathus stercoreus*) to degrade Bermuda grass lignocellulose, taking six weeks for 32% and 77% degradation (change in dry weight) to occur, respectively (Akin *et al*., 1995). Similarly and more recently, Hermosilla *et al.* (2017) used white-rot fungi (*Ganoderma lobatum*) to degrade wheat straw, achieving 50% lignin degradation and 18.5% dry weight change in 40 days through increased nitrate availability and addition of enzyme inducers (Hermosilla *et al*., 2017). The recalcitrance of lignin makes degradation difficult for even the most efficient ligninolytic fungi, therefore current biological treatments must be used in combination with conventional pretreatment methods to compensate for this. Common examples being thermal and chemical pretreatment processes. The high temperatures and strong chemicals involved in these methods can however have significant impacts in culture conditions. Taniguchi *et al.* (2010) combined the methods of using basidiomycetes (*Pleurotus ostreatus*) with steam explosion on rice straw biomass, resulting in 33% cellulose degradation and approximately 20% lignin degradation in 36 days (Taniguchi *et al*., 2010). Wan and Li (2010) however showed that degradation rate can be significantly increased through process optimization, achieving approximately 30% lignin and 94% cellulose degradation in 18 days using *Ceriporiopsis subvermispora* (Wan and Li, 2010). This study, in addition to the study carried out by Hermosilla *et al.*, illustrated the high selectivity of fungi with regards to culture conditions, but also highlighted that this characteristic can be exploited to achieve more efficient degradation of lignocellulose, without the need for pretreatment. Lignocellulose bioconversion processes require further development if these methods are to compete with current processes, which produce superior yields of cellulose in shorter time periods.

Table 1 *- Examples of studies investigating fungal lignin-degrading microorganisms*

Microorganism (Species)	**Substrate**	**Degradation (dry weight change)**	**Time-Scale**	**Source**
Ceriporiopsis subvermispora, *Cyathus stercoreus*	Bermuda grass lignocellulose	32%, 37%	6 weeks	Akin *et al., 1995*
Ganoderma lobatum	Wheat straw	18.5% (50% lignin)	40 days	Hermosilla *et al*., 2019
Pleurotus ostreatus	Rice straw biomass	33% cellulose, 20% lignin	36 days	Taniguchi *et al*., 2010
Ceriporiopsis subvermispora	Rice straw biomass	94% cellulose, 30% lignin	18 days	Wan and Li, 2010

1.3.2 Ligninolytic Enzymes Used by Microorganisms

The enzymes used by microorganism to degrade lignin are either lignin-modifying oxidative enzymes (LME), which directly interact with lignin, or lignin-degrading auxiliary enzymes (LDA), which aid in the degradation process (Kameshwar and Qin, 2016). The most common LMEs are laccase (or phenol oxidase) enzymes, which are only able to specifically target and degrade phenolic subunits of the lignin polymer (10% of total molecule) through direct oxidation. This is attributable to their low oxidation-reduction potential, however this can be increased by the addition of synthetic mediators, which has gained increased interest in industrial applications (Martínez *et al*., 2005). For example, laccases have been observed to oxidise non-phenolic lignin compounds in the presence of mediators such as Remazol Blue in a study carried out by Bourbonnais and Paice (1990). Many microorganisms produce their own variants of laccase and expression can be intracellular or extracellular, depending on the species, with bacterial laccases commonly being expressed intracellularly. Some intracellular laccases will only be expressed when the cell in exposed to specific environmental conditions, similar to some extracellular laccases which will only function in specific conditions, whereas others are more versatile (Santhanam *et al*., 2011; Huang *et al*., 2013).

Some microorganisms also produce a variety of peroxidase enzymes (also LMEs). Peroxidases are found throughout nature - from bacteria to plants to humans - and their

function can vary between organism and species. Just as peroxidases are used to build the subunits of lignin for cell wall biosynthesis in plant cells, peroxidases are used to break these subunits apart by bacteria and fungi (Passardi *et al*., 2005; Colpa *et al*., 2014). Examples include: general manganese-dependent peroxidase (MnP), which can target the phenolic or, through mediation, non-phenolic subunits of lignin; lignin peroxidase (LiP), which targets the more abundant non-phenolic subunits; and versatile peroxidases (VP), which are able to break down both phenolic and non-phenolic subunits due to higher redox potential (Bugg *et al*., 2011a; Guerriero *et al*., 2016; Datta *et al*., 2017). VP peroxidases are a hybridization of MnP-LiP peroxidases, combining the activity of both (Guerriero *et al*., 2016). More recently, dye-decolourising peroxidase (DyP) has been identified as an additional lignin-degrading peroxidase superfamily. Like LiP, DyP acts by targeting and breaking bonds within the abundant non-phenolic subunits of lignin (Guerriero *et al*., 2016; Datta *et al*., 2017). Despite the previous belief that DyP peroxidases were exclusive to bacteria, they have also been identified in other microorganisms, such as DyP type-D enzymes found in fungi (Van Bloois *et al*., 2010; Brown and Chang, 2014; Datta *et al*., 2017).

All of these enzymes (MnP, LiP, VP and DyP) oxidize and degrade lignin by breaking different bonds between the subunits and cleaving the aromatic rings within the subunits, through various electron transfer reactions and production of free radical byproducts (see **Figure 4** below). DyP peroxidases, however, are also able to hydrolyze the lignin molecule, making it of special interest to industry (Datta *et al*., 2017; Guerriero *et al*., 2016). Microorganisms use one or several of these enzymatic pathways to degrade lignin, in combination with a range of various other auxiliary enzymes (LDAs). Examples of LDAs include glyoxyl oxidase and cellobiose dehydrogenase/oxidoreductase, which aid the degradation process, usually producing hydrogen peroxide for peroxidase enzymes (Guerriero *et al*., 2016; Kameshwar and Qin, 2016; Janusz *et al*., 2017).

It is generally understood that aerobic conditions are required for the functioning of the afore mentioned oxidative enzymes, with these enzymes requiring oxygen (e.g. laccase) or molecules containing oxygen (e.g. peroxidase via H_2O_2) to function. Furthermore, oxygen (or reactive oxygen species) is required by microorganisms to break the aromatic rings of the lignin-derived subunits following degradation (Fuchs, 2008; Janusz *et al*., 2017). Anaerobic degradation of lignin is poorly understood at the time of writing this thesis, however, it is thought that anaerobic microorganisms use alternative enzymatic pathways which do not require oxygen (Li *et al*., 2009). For example, *Sphingomonas paucimobilis* SYK-6 and *Rhodococcus* sp. RHA1 are known to use β-aryl etherase and

the β-ketoadipate pathway, respectively, to degrade lignin and phenolic compounds (DeAngelis *et al*., 2013).

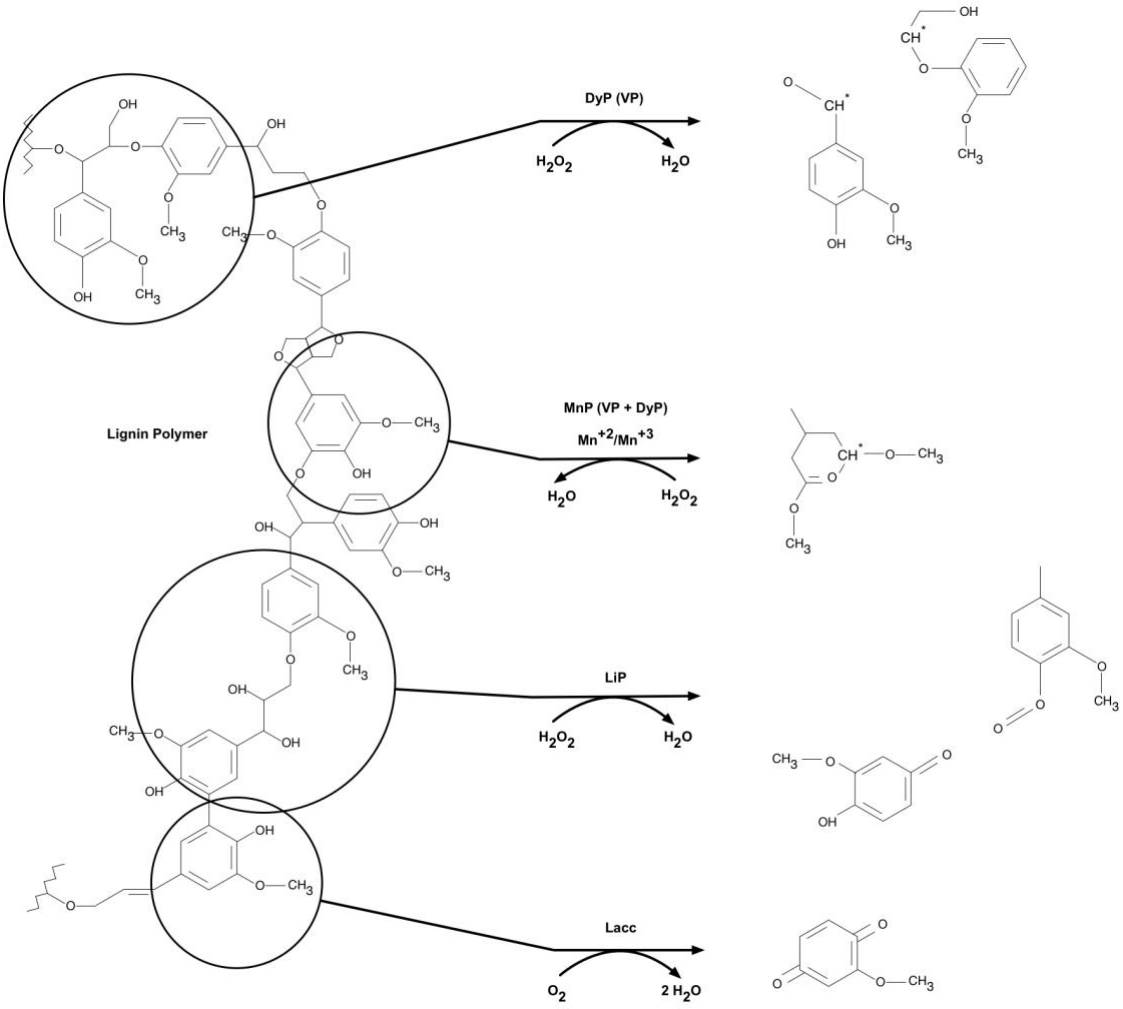

Figure 4 *- Target areas within lignin which different enzymes target and the resulting degradation products. Adapted from Bugg et al., 2011b. and Janusz et al., 2017.*

1.4 Bacterial Lignin Degradation

Many of these enzymes are also naturally produced by species of soil bacteria which regularly encounter lignocellulose in their natural habitat. Despite the fast-degradative capabilities of fungi, attributed to the higher oxidative capabilities of their enzymes, bacteria have received more attention with regards to industrial use because of their inherently robust nature. Bacteria, and their enzymes, are able to function in more challenging conditions (temperature, pH, oxygen saturation, etc.) compared to fungi and fungal enzymes. In addition to these characteristics, bacteria also proliferate much faster than fungi (Santhanam *et al*., 2011; Tian *et al*., 2016). Whilst many genera of bacteria

are capable of degrading cellulose (through production of cellulase), they lack the enzymatic ability to degrade the protective hemicellulose-lignin barrier, and so only a select number of bacteria are able to use the components of lignocellulose as sources of carbon (Souza, 2013). In addition to bacteria, limited research on archaea has revealed that some species may also be able to produce ligninolytic enzymes. So far, 5 laccase genes and 11 DyP genes have been identified in the genomes of several archaeal species. Whilst Archaea have been shown to express cellulase, ligninolytic enzyme expression is yet to be assayed (Graham *et al*., 2011; Colpa *et al*., 2014; Tian *et al*., 2014).

Lignin-degrading bacteria are most commonly found under the genera Actinomycetes, Firmicutes, as well as α-, β- and γ-Proteobacteria (Bugg *et al*., 2016, 2011). Research carried out by Ramachandra *et al.* (1989) and Spiker *et al.* (1992) demonstrated that, similar to brown-rot fungi, various species of bacteria, such as the actinomycete *Streptomyces viridosporus,* were able to partially degrade the lignin molecule, oxidizing phenolic subunits through limited enzyme production, but were unable to completely degrade the entire molecule (Ramachandra *et al*., 1988; Spiker *et al*., 1992). *S. viridosporus* was shown to produce peroxidase enzymes, through dependency on hydrogen peroxide availability for degradation. Purified peroxidase enzymes were shown to be active against only phenolic lignin monomers, explaining the partial degradation. A recent study carried out by Ahmad *et al.* (2014) identified *Rhodococcus jostii* (an Actinomycete) and *Pseudomonas putida* (γ-Proteobacteria) as lignocellulose degraders through the use of fluorescently-labeled lignin assays, however degradation activity was much weaker than that of the white-rot fungus. Furthermore, this study indicated that *R. jostii* and *P. putida* were using extracellular enzymes to break down lignin, most likely laccases, based on the result that degradation occurred in the absence of hydrogen peroxide which is required for peroxidase enzyme function. *Rhodococcus jostii* has also been shown to encode the gene required to produce the highly effective lignin-degrading DyP peroxidase enzymes through genomic analysis, in addition to the laccase enzymes, suggesting that some bacteria have a variety of encoded enzymes at their disposal and can use these in combination to degrade lignocellulose, much like fungi (Ahmad *et al*., 2011; Souza, 2013).

Building on the findings of Ramachandra *et al.* and Spiker *et al.*, further characterization of *Streptomyces viridosporus* has revealed that it in fact produces a wide variety of lignin-degrading enzymes, such as laccases, lignin peroxidases and DyP peroxidases (Asina *et al*., 2017; Lambertz *et al*., 2016). Furthermore, several studies have revealed that some bacterial species (such as *Streptomyces*), instead of breaking down lignin to

access cellulose, use lignin as a carbon source (Asina *et al*., 2017; Brown and Chang, 2014). For example, Bianchetti *et al.* (2013) investigated an irregular combination enzyme, isolated from *Streptomyces sp.* found in the guts of wood-boring wasps that was able to simultaneously bind lignin and break open aromatic rings, possibly enabling the *Streptomyces sp.* to feed on the lignin carbon (Bianchetti *et al*., 2013). These studies present just a few examples of bacterial lignin degraders, their enzymes and the advantages of using bacteria. These studies also exemplify the knowledge obtained from in-depth studies on known lignin degraders, but also highlight a number of areas which require more research. As of yet, the abundance of studies focused on analyzing community lignin degradation in the natural environment is limited. Pure culture studies provide information on how a specific species interacts with lignin but does not explain how these species fit into the larger scheme of environmental lignin degradation. It is for this reason that community studies on environmental samples are also important, to identify key lignin degrading organisms and determine how these organisms contribute to lignin degradation in the natural environment.

1.4.1 Single-Species vs. Community Studies for Industrial Applications

Further research into bacterial communities from different environmental backgrounds will allow comparative analysis to be carried out in order to highlight diversity and variation in the mechanisms involved in lignin degradation between environments. Soil temperature, pH and depth (oxygen availability), as well as lignin content and distribution within the soil, are just some of the factors which could elicit differences between bacterial degradation mechanisms. In addition to these factors, externally expressed enzymes may only be produced in specific conditions and have condition-dependent functionality. Evidence from community studies, such as environmental sample analysis, can lead to further research on individual species, or highly-effective defined microbial consortia. This research provides opportunities to uncover novel lignin-degrading species and enzymes, and display the mechanisms of how each species degrades the lignin molecule, as well as present the potential of a consortium/organism for industrial use. Opposed to the concept of using a single organism for industrial lignin degradation, natural lignin degradation, like almost every microbial catabolic process in the environment, is carried out as a community, much like the gut flora of herbivore animals and wood feeding insects (Ahmed *et al*., 2016).

Chandra *et al.* (2007) determined the lignin degradation potential (through Kraft lignin decolourisation) of three bacterial species found in paper mill waste (effluent sludge), as

well as a consortium of the three species together. The individual bacterial cultures (*Paenibacillus sp.*, *Aneurinibacillus aneurinilyticus* and *Bacillus sp.*) were found to degrade lignin by 30%, 33% and 37%, respectively, whereas the consortium resulted in 40% degradation (Chandra *et al*., 2007). A study carried out by Wang *et al.* (2013) investigated a bacterial consortium from a pond sludge sample which was able to degrade more than 50% of lignin and 40% hemicellulose from reed biomass at 30 °C in 15 days, however only 2% of the cellulose was degraded from the biomass (Wang *et al*., 2013). In support of this, Lv *et al.* (2014) achieved 96% lignin degradation efficiency in 7 days with a microbial consortium containing *Pseudomonas putida* with sucrose supplementation, with the conclusion being that higher lignin degradation efficiency can be achieved following initial cell growth with sucrose, however it is unclear how degradation efficiency compares to percentage mass degradation measured in similar studies (Lv *et al*., 2014).

Conversely, recent research carried out by Blánquez *et al* (2017) suggests that efficient degradation can be achieved using single species which are highly potent in lignin degradation. This study revealed laccase SilA produced by *Streptomyces ipomoeae* as a significant enzyme involved in lignin degradation, by comparing degradation with a knock-out mutant. Despite the low redox potential and known conditional specificity of these enzymes, Blánquez *et al* also presented laccases as being highly robust, with high functionality in a wide range of temperatures and pH levels. This study achieved 45% lignin degradation in 7 days, with significantly higher lignin degradation rate compared to a knock-out mutant (Blánquez *et al*., 2017). Additionally, a study carried out by Shi *et al.* (2013) achieved 44.4% lignin degradation in 7 days with a *Cupriavidus basilensis* culture and lignin being the primary carbon source (Shi *et al*., 2013). These findings are superior to that of the research carried out by Hermosilla *et al.* and Wan and Li *et al.* (mentioned above), who achieved 50% lignin degradation in 40 days and 30% lignin degradation in 18 days, respectively, using single white-rot fungi species, despite process optimization.

It is perhaps possible that the single species used in these studies were key lignin degrading bacteria in their natural environments, which would explain their high affinity to lignin. From the presented research, it is difficult to make a conclusion on whether singular bacterial species or bacterial consortiums are superior for industrial lignin degradation, highlighting that more research is required in this area. Using bacterial consortia for industrial lignin degradation is attractive due to higher degradation potential and the concept of simplifying industrial processes, without the need for species isolation. However, these studies show that, regardless of the number of species used,

bacteria can be just as, if not more effective than fungi with regards to lignin degradation and are able to use lignin as a primary carbon source, with the optimization of culture conditions and nutrient supplementation. Generally, it can be extremely difficult to use data from pure culture or select consortia studies to predict how a substrate is degraded in the natural environment. As lignin degradation research is still in its infancy, studies combining and comparing pure culture degradation and consortia degradation are limited, compared to the growing abundance of individual consortia and pure culture studies. However, the results obtained from these studies complement each other, and are valuable to understanding and predicting how these processes naturally function in the environment.

1.4.2 Key Lignin Degrading Microorganisms

Each environment is different in terms of microbial diversity and functioning towards substrate catabolism. The study of any microbial catabolic process begins with investigating the most capable organisms, and the mechanisms they use to interact with the substrate. These initial studies provide a foundation for further study on how these organisms fit into the overall catabolic process, and can highlight factors which impact the natural process. Lu *et al* (2014) presented a direct correlation between *Streptomyces* laccase-like multicopper oxidase (LMCO) gene presence, lignocellulose material availability and degradation rate in compost through the use of PCR primers targeted to these genes and surveying the presence of these genes during compost incubation. The study showed that initial high volumes of lignocellulose material resulted in proposed upregulation of *lmco* genes in the *Streptomyces* sp. which, in turn, resulted in increased degradation rates. Comparing the obtained *Streptomyces sp. lmco* sequences with the sequences of related organisms on Genbank (using BLAST), Lu *et al.* found that the sequences were closest to that of *Streptomyces griseus, Streptomyces ipomoeae and Streptomyces violaceusniger*, which have also been known to produce extracellular peroxidases (Brown and Chang, 2014; Lu *et al*., 2014). Woo *et al.* (2014) investigated a tropical forest soil sample to explore the differences in enzyme activity between the various bacteria present within the sample. The study found a variety of *Gordonia sp.* which were highly effective lignin degraders due to high phenol oxidase and peroxidase activity but were inefficient cellulose degraders due to suppressed β-D-glucosidase and cellobiohydrolase activity. Conversely, the study found a range of *Aquitalea sp.* which had the opposite phenotypes (high β-D-glucosidase and cellobiohydrolase, low phenol oxidase and peroxidase) and were therefore more effective at degrading cellulose (Woo *et al*., 2014).

This research provides further evidence for species-specific degradation mechanisms but also bring attention to the effects of soil lignin content on bacterial communities within the natural environment and the effects of diversity within these environments, with certain organisms being 'specialized' in lignin degradation. The same observations have been made using select species of rot fungi, with certain species being more attracted to either polysaccharide carbon or lignin carbon within the lignocellulose polymer (Kumar *et al*., 2009). Due to strong competition within microbial environments, certain strains have become accustomed to using particular carbon sources from lignocellulose and only produce the complementary enzymes because of this. These studies also highlight that different soil environments have different key lignin degrading bacterial species, with *Streptomyces sp.* appearing to be the primary lignin degrader in the compost sample used by Lu *et al.* and *Gordonia sp.* in the forest soil sample used by Woo *et al.*

1.5 Linking Microbial Genomics to Lignin Degradation

The majority of these studies have focused on in-depth biochemical analysis of enzymes and their complementary pathways, however advancement of genome sequencing technologies has allowed research to be carried out on the genetics behind these enzymes. Using these methods to investigate microbial lignin degradation allows identification and characterization of key ligninolytic microbes, at species-level, and reveals how these microbes contribute to lignin degradation in the natural environment, at community-level. Methods such as whole genome sequencing can provide information on the enzymatic potential of individual known lignin degraders, through investigating genes encoding these enzymes. More complex methods, such as metagenomics, can do this on a larger scale, generating genomic data of entire microbial communities, providing an overview of microbial diversity and functioning in the natural environment.

Techniques such as high-throughput 16S rRNA gene amplicon sequencing have become standard methods within biology due to these advancements. 16S rRNA gene sequencing allows a researcher to screen samples with high microbial diversity for identification and characterization of all microorganisms within that sample by comparing these genome sequences with that of similar organisms in genome databases (Woo *et al*., 2008). This approach is based on the knowledge that small subunit ribosomal DNA/RNA is highly conserved within microorganism's due to non-evolution (i.e. unlikely to be different between species) and therefore sequencing this DNA can lead to accurate identification of a microorganism (Kameshwar and Qin, 2016). Metagenomics/transcriptomics are methods which can also be used to identify and analyze the genomic/transcriptomic material of all organisms within an environmental

sample. The process typically involves DNA/RNA extraction, sequencing, assembly and annotation of the metagenome/metatranscriptome enabling a researcher to conduct in-depth analysis of a whole microbial community, investigating the relationship between genetics and cellular function (Thomas *et al*., 2012). These methods are used in combination with advanced computing tools, programmed to handle genomic data, allowing detailed analysis of sequenced genomes which can significantly contribute to the previously described methods.

Although metagenomics is an extremely powerful tool, it can sometimes be challenging to assay high-diversity samples using this method. Microorganisms (and therefore their DNA) commonly appear in varying abundance in environmental samples, and the most abundant organisms are easier to detect compared to less abundant organisms (Coyotzi *et al*., 2016). An alternative method of identification and genotype-phenotype relationship analysis combines nucleic acid stable isotope probing (SIP) with metagenomics, using isotope-labelled substrates to identify microorganisms which are actively incorporating the particular substrate into their biomass (DNA or RNA). This method is based on the concept that a microorganism will degrade certain substrates (such as lignin) and incorporate molecules from the substrate into their cellular components, such as DNA, which can be extracted from the cells and analyzed (via centrifugation and separation of the heavier labelled DNA) (See **Figure 5** below) (Neufeld *et al*., 2007; Coyotzi *et al*., 2016). Using this method in combination with metagenomics allows targeted analysis of highly diverse samples, eliminating the afore mentioned disadvantage of metagenomics and permitting detection of less-abundant and potentially important species, which would otherwise be overlooked (Coyotzi *et al*., 2016). Furthermore, DNA SIP eliminates the need for culture enrichment, specifically targeting desired species from communities.

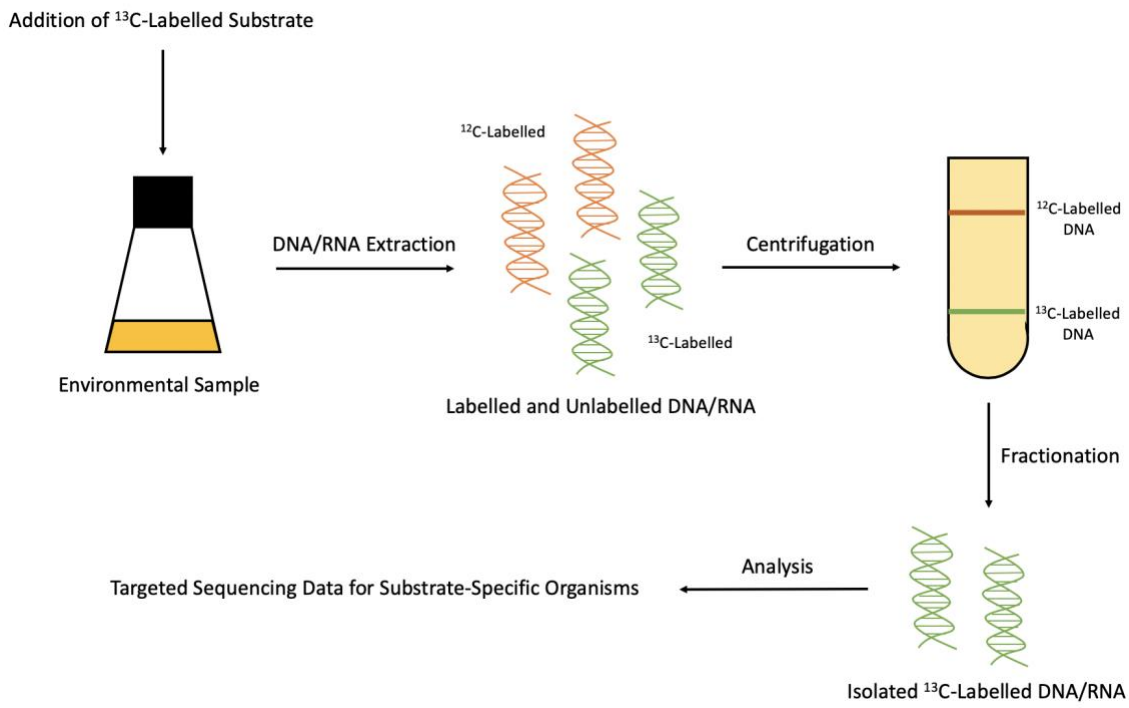

Figure 5 *- DNA/RNA stable isotope probing. Adapted from Coyotzi et al., 2016.*

1.5.1 Genome Sequencing: From Single Species to Whole Communities

Recently, a number of studies have been carried out regarding the use of these genomic technologies to investigate lignin degradation within environmental samples. Prabhakaran *et al.* (2015) used Illumina sequencing technology to sequence the whole genome of *Pseudomonas sp.* and *Rhizobium sp.* isolated from a degrading wood sample and used a range of bioinformatic tools to assemble the genome and predict gene location and identity (Hyatt *et al*., 2010). This study found that both genomes contained genes coding for a range of enzymes and other proteins involved in lignin degradation, including laccases, dye-decolourising peroxidase, chloroperoxidase and other auxiliary enzymes (Prabhakaran *et al*., 2015). Further genomic analysis and culture studies revealed that the same *Rhizobium sp.* also produces lignin peroxidase and the auxiliary enzymes glyoxyl oxidase and glycolate oxidase, which provide peroxide for peroxidase function (Jackson *et al*., 2017). The bacteria used in these studies were from pure cultures, in conditions optimized to the microorganisms needs, sometimes using pre-prepared lignin substrates and derivatives. These studies produce highly significant findings with regards to the ligninolytic enzymes used by these specific bacteria, however, as microorganisms degrade lignin as a community in the natural environment, pure culture studies provide only part of the story. Microorganisms that degrade lignin

may do so differently in the natural environment, using different enzymes in different conditions. It is therefore vital to explore microbial communities from a variety of environmental samples, in addition to specific species, to reveal how each organism interacts with the lignin molecule in different conditions.

An appropriate example would be the study carried out by DeAngelis *et al.* (2011) which investigated the diversity and lignin degradation mechanisms of microflora within tropical forest soil over time, using lignin-bound beads, quantitative polymerase chain reaction (qPCR), 16S DNA sequencing analysis and enzyme activity assays. “Bio-traps” containing lignin-bound and unbound beads were buried in the soil and samples were taken from 1-30 weeks. Results from this study identified the bacterial phyla Actinobacteria, Proteobacteria, Acidobacteria and Firmicutes as the primary lignin degraders, with abnormally low fungi presence. It was hypothesized that, due to the anaerobic nature of the environmental samples and the preference of fungi to aerobic conditions, fungal activity was suppressed, allowing the more robust bacterial lignin degraders to gain a competitive advantage. Over time, the abundance of microorganisms within the lignin samples seemed to change, with 12 total species being present in week one and approximately 70 total species present in week 30. Furthermore, the diversity of bacteria also evolved over time, with select species of *Rhodobacterium* (α-Proteobacteria), *Bacillus* (Firmicutes) and Actinobacteria being highly present in week 1, and several α-Proteobacteria, β-Proteobacteria, Acidobacteria and Chloroflexi species being highly present in week 30 (DeAngelis *et al*., 2011). Oxidative enzyme (laccase and peroxidase) assay results from the lignin-bound beads showed initial high activity which gradually decreased throughout the 30-week period. With the combination of high iron concentration and low oxygen availability in the soil, it was proposed that non-oxidative mechanisms were also being used by the community.

These findings highlight that certain species play specialized roles in lignin degradation, and as the molecule is broken down by key degraders, perhaps opportunistic organisms with less-evolved degradation methods are then able to access products from this degradation, and carbon moves down the chain. This is further supported by the observed decrease of laccase and peroxidase activity over time, which are enzymes previously shown to be produced by several Proteobacteria, Actinobacteria and Firmicutes, present in the initial sample but absent in the final sample. A study carried out by Pold *et al.* (2015) used similar methods to instead explore the effects of climate change on ligninolytic bacteria within forest soil in Massachusetts. Like the previous study, the results identified a range of known and unknown lignin degrading bacteria

having high affinity for the lignin-bound beads, with bacteria being significantly more abundant compared to fungal species.

Combining methods such as 16S rRNA gene sequencing and whole genome sequencing with metagenomic methods, enables researchers to carry out culture-dependent and culture-independent investigations to reveal community functioning and how individual microbes function in the community. Furthermore, the introduction of nucleic acid stable isotope probing can help to narrow the scope of these analyses, providing higher accuracy in the data obtained. A number of metagenomics studies have been carried out on the ligninolytic microorganisms which inhabit the intestinal tracts of wood-feeding insects. Li *et al.* (2013) investigated the microbial community inhabiting the intestines of a species of lower termite (*Coptotermes gestroi*) using metagenomic analysis and high-throughput sequencing technology (Illumina). 80% of the total genomic material was derived from bacteria, belonging to a variety of phyla, including Spirochaetes, Proteobacteria, Bacteroidetes, Firmicutes and Actinomycetes. This study was however targeted towards genes involved in carbohydrate degradation (cellulose and hemicellulose), and not lignin degradation (Li *et al*., 2013). The guts of wood-feeding insects provide a unique environment which selects for a niche microbial community, primarily composed of highly capable lignocellulose degraders, which could have potential for industrial applications. An example of a soil metagenomics study is Wang *et al*. (2016) who used a combination of 16S rRNA gene amplicon sequencing and metagenomics analysis to investigate the ligninolytic potential of an enriched bacterial consortia derived from compost. Results from both analyses confirmed the consortia to be bacteria dominant, with Actinomycetes, Proteobacteria and Firmicutes being the most abundant phyla. Through screening the metagenome for enzymes involved in lignocellulose degradation, enzymatic profiles were constructed for each species, highlighting Actinomycetes as the most capable (Wang *et al*., 2016). These findings were similar to results from Lu *et al.,* mentioned previously, which identified the Actinomycete *Gordonia* as the primary degrader in compost, suggesting that the thermophilic Actinomycetes are better suited to high temperatures within compost environments.

Whilst DNA SIP studies investigating microbial lignocellulose degradation are rare, Darjany *et al.* 2014 used DNA SIP with ^{13}C-labelled lignocellulose to identify ligninolytic bacteria within salt marsh sediment samples. This study identified species of *Kangiella, Desulfosarcina* and *Spirochaeta* as the most prominent bacterial species involved in lignocellulose breakdown in this environment. Whilst species of *Spirochaeta* have previously been shown to degrade cellobiose, *Kangiella, Desulfosarcina* and *Spirochaeta* have not previously been shown to degrade lignin. However, Darjany *et al.*

noted that, as previously mentioned, it was likely that some of the species identified in the lignin-labelled sample were in fact using by-products of lignin degradation being carried out by more capable microbes. Similar methods were employed by Pepe-Ranney *et al.* (2015) using ^{13}C-labelled cellulose and xylose instead of lignocellulose. Results obtained in this study were similar to those obtained by Darjany *et al*. and DeAngelis *et al.*, with cellulose and xylose degradation being primarily carried out by specialized bacteria and that the relative abundance of these bacteria changed over time. Like lignin degradation, Pepe-Ranney *et al.* proposed that the carbon from labelled xylose moved down “through different trophic levels within the soil bacterial community” (Pepe-Ranney *et al*., 2015). More recently, Wilhelm *et al*. (2018) carried out DNA-SIP and shotgun metagenomics with a variety of soil samples (organic and mineral layers) using ^{13}C-labelled cellulose, hemicellulose and lignin. The most prominent lignin degrading microorganisms were bacteria from the families *Comamonadaceae* (Betaproteobacteria) and *Caulobacteraceae* (Alphaproteobacteria). In this study, most microorganisms were found to only actively degrade one substrate, for example fungi showing specific affinity to the cellulose substrate, whereas several *Caulobacter* species were able to degrade more than one of the substrates. Furthermore, several operation taxonomic units (OTUs) derived from *Caulobacteraceae* enriched on ^{13}C-lignin were from unknown organisms. The focus of DNA SIP on microbes able to degrade plant-derived substrates helps to provide an oversight of environmental carbon-cycling, whilst providing opportunities to identify and characterize undiscovered species which are less abundant within communities, which would otherwise be missed by other investigatory methods.

1.6 Summary

This review displays ample evidence for the potential of using microorganisms in lignocellulose treatment for production of high-value products. From the literature, it is clear that the factors limiting successful lignocellulose utilization are the recalcitrance of lignin and the resulting problems concerning the methods of breaking this molecule during processing. Microorganisms naturally occurring in soil environments have been shown to be capable of degrading lignin more than or as efficiently as current processing methods, with a significantly reduced impact on the environment. These organisms could potentially provide means of reducing the costs and optimizing current processing methods by providing alternative feedstock options (plants with higher lignin content) and eliminating the need for biomass pretreatment, respectively.

Studies so far have shown that there are specialized microorganisms present within the environment which are capable of degrading lignin using species-specific mechanisms.

Furthermore, these studies have shown that bacteria play an important role in the lignin degrading community. However, if microorganisms, and more specifically bacteria, are to be successfully used in industry for the processing of lignocellulosic biomass, more research is required. Studies regarding key lignin-degrading microbes in the environment, the enzymes used to facilitate this breakdown and the genes encoding these enzymes, will help to further characterize and potentially discover new, more capable species. Furthermore, research regarding how microorganisms collaboratively degrade lignin as a community could provide a deeper insight into the role of each species in this process. This knowledge can be gained through the use of complex genomic techniques to explore the genetics behind ligninolytic enzyme production of prolific lignin degrading microbes, and their role in the overall degradation of lignin within the environmental community. From there, further studies focused on specific key lignin degrading species/lineages could reveal new methods for industrial use.

1.6.1 Project Aim

The primary aim of this project was to determine the key microorganisms and enzyme pathways involved in lignin degradation within a range of diverse environments (detailed in **Methods and Materials Section 2.2**), in order to discover novel and efficient candidates for industrial application. With a focus on bacteria, the activity of microbial lignin degradation will be analysed in environments such as different types of soil and sediment (savanna soil and Hickling Broad lake sediment) as well as several highly specific environments (such as compost and the gastrointestinal tract of elephant). These environments are very different with regards to conditions (such as, pH, ambient temperature and lignin content) and are therefore likely to contain diverse communities of microorganisms with ligninolytic potential.

1.6.2 Project Objectives

Objective 1: Screening of environmental samples and identification of microorganisms involved in lignin degradation using stable isotope probing:

Using DNA/RNA-SIP, samples from different environments will be incubated with ^{13}C – labelled lignin, ^{12}C (unlabelled) lignin or $^{13}CO_2$ (secondary labelling control) to label the nucleic acid material of active lignin degraders. A $^{13}CO_2$ control will be included in this experiment to determine the background labelling of microbial communities which have incorporated ^{13}C into their genome from $^{13}CO_2$ as opposed to that from the labelled lignin. Samples of these incubations will be taken periodically and will undergo DNA/RNA

extraction, density gradient ultracentrifugation and fractionation to isolate the genetic material of lignin degrading microorganisms throughout the incubation period. The DNA/cDNA will then undergo qPCR analysis with primers targeting microbial SSU rRNA genes and functional marker genes of lignin degradation (such as oxidative enzyme genes) followed by microbial SSU rRNA gene amplicon sequencing and metagenomics. These analyses will primarily determine the identity and functional mechanisms of the key active lignin degrading organisms (bacteria, fungi and archaea), as well as how these characteristics differ between each environment.

Objective 2: Investigating the identity of and mechanisms being employed by lignin degraders using microbial SSU rRNA and functional gene PCR and qPCR, amplicon sequencing and metagenomics/transcriptomics:

DNA/RNA extracts from environmental samples will undergo quantitative polymerase chain reaction (qPCR) analysis with primers targeting both microbial SSU rRNA and lignin-active enzyme-encoding functional genes, in order to reveal presence/abundance of microbial communities and functionality. Nucleic acids from environmental samples will also undergo various sequencing methodologies followed by sequence alignment, annotation and analysis using bioinformatic computational tools. Microbial SSU rRNA gene amplicon sequencing and analysis will allow reconstruction of microbial community profiles within each environmental sample, elucidating presence and abundance of microbial taxa and community diversity. Analysis of the metagenome will allow identification in addition to the screening for genes encoding oxidative enzymes (such as laccases and peroxidases), and therefore the enzymatic potential of the entire microbial community within each environmental sample. Our study will make use of the carbohydrate-active enzymes (CAZy) database (Drula *et al*., 2022) for identification of lignocellulose-degrading genes and will use methods similar to that of Wilhelm *et al*. (2018), mentioned above. The CAZy database contains genes coding for microbial enzymes that degrade or modify glycosidic bonds, which encompasses lignin-degrading enzymes. Information regarding the potential functional mechanisms for the microorganisms present in the environment will be obtained. This data will be combined with the information obtained from the lignin SIP results to provide further insight into the key microbial lignin degraders and their mechanisms in each environment.

Objective 3: Identification and characterisation of key lignin degrading species, isolated from environmental samples through enrichment methods, and their lignin-degrading mechanisms:

In addition to the culture-independent methods mentioned above, lignin degrading microorganisms will also be enriched and isolated from the environmental samples using alkali lignin as a substrate in conditions similar to natural environmental conditions. These isolates will encompass species representative of members of the lignin-degrading community within each environment. Resulting isolates which display high lignin degrading potential and enzyme activity (from enzyme assays – **Objective 4** - and GC-MS analysis - **Objective 5** - respectively) will also be selected to undergo DNA extraction followed by whole genome sequencing (WGS) to confirm phylogenetic placement and genomic functional potential of each isolate. This data will provide detailed information on selected isolates, such as gene clusters responsible for oxidative enzyme and auxiliary enzyme expression/regulation. Isolate whole genome sequences will be compared with the metagenomic data in order to determine the abundance and diversity of isolate enzyme pathways for lignin degradation.

Objective 4: Determination of isolate and enzyme functionality using functional cloning and lignin-degrading enzyme activity assays:

New isolates and enzymes of interest, identified in the enrichment incubations, will be analysed through the use of functional cloning and oxidative enzyme activity assays to confirm and analyse functionality. Functional cloning is carried out to determine whether or not the oxidative enzyme genes identified within the metagenomic data are functional.

Furthermore, the lignin degrading species and enzymes isolated in this project will undergo laccase and lignin peroxidase enzyme assays with industrial CASE partner, Recircle Ltd. A number of assays derived from literature will be used to test for the activity of multiple different enzymes, from laccase and general peroxidase (protocol from CASE partner) to specific enzymes such as lignin peroxidase and manganese-dependent peroxidase. Enzymes and isolates obtained in this project will undergo these analyses to determine activity and industrial suitability. Furthermore, using the CASE partners collection of lignin degrading organisms, these activities can be compared with known degraders to determine industrial capability. Furthermore, combination with GC-MS analysis of lignin breakdown products from isolates (see **Objective 5**) will allow further clarification of degradation mechanisms.

Objective 5: Determination of isolate lignin breakdown mechanisms and resulting aromatic compounds (with industrial relevance) through advanced biochemical analysis:

Additional analyses on isolates, and potentially SIP enrichments, will be carried out in collaboration with Dr Clayton Magill (Lyell Research Fellow, The Lyell Centre, HWU) and Lavinia Stancampiano (PhD Student, The Lyell Centre, HWU). In order to determine microbial activity, or more specifically the resulting breakdown products derived from the lignin polymer throughout culturing with alkali lignin, low molecular weight aromatic compounds will be extracted from isolate cultures, will be derivatized and will undergo Gas chromatography-mass spectrometry (GC-MS) analysis. Total ion chromatogram data obtained from the GC-MS will undergo analysis through identification and quantification of peaks representing aromatic oxidation products. By analysing the changes in monomeric/dimeric composition of the lignin polymer throughout the incubations, the extent to which isolates break down the lignin polymer will be revealed, whilst also potentially providing evidence for the enzymatic mechanisms being used by these isolates, i.e. presence/abundance of specific monomeric/dimeric oxidation products may be linked to activity of specific enzyme(s). Growth medium plus lignin (no inoculum) will be used as a standard (T=0) sample. Furthermore, the same methodologies will be used in order to analyse samples taken throughout environmental sample SIP incubation with $^{12}C/^{13}C$-lignin in order to identify the lignin breakdown products in an environmental context and by potentially highlighting specific labelled breakdown products identified in the respective isolate experiments. Observed lignin-degrading mechanisms used by isolates will be confirmed by construction of oxidative enzyme gene profiles from the whole genome sequencing data.

1.7 Project Objective Flowchart

A flowchart summarizing the objectives and therefore the processes and work involved in achieving these objectives can be found overleaf (**Figure 6**).

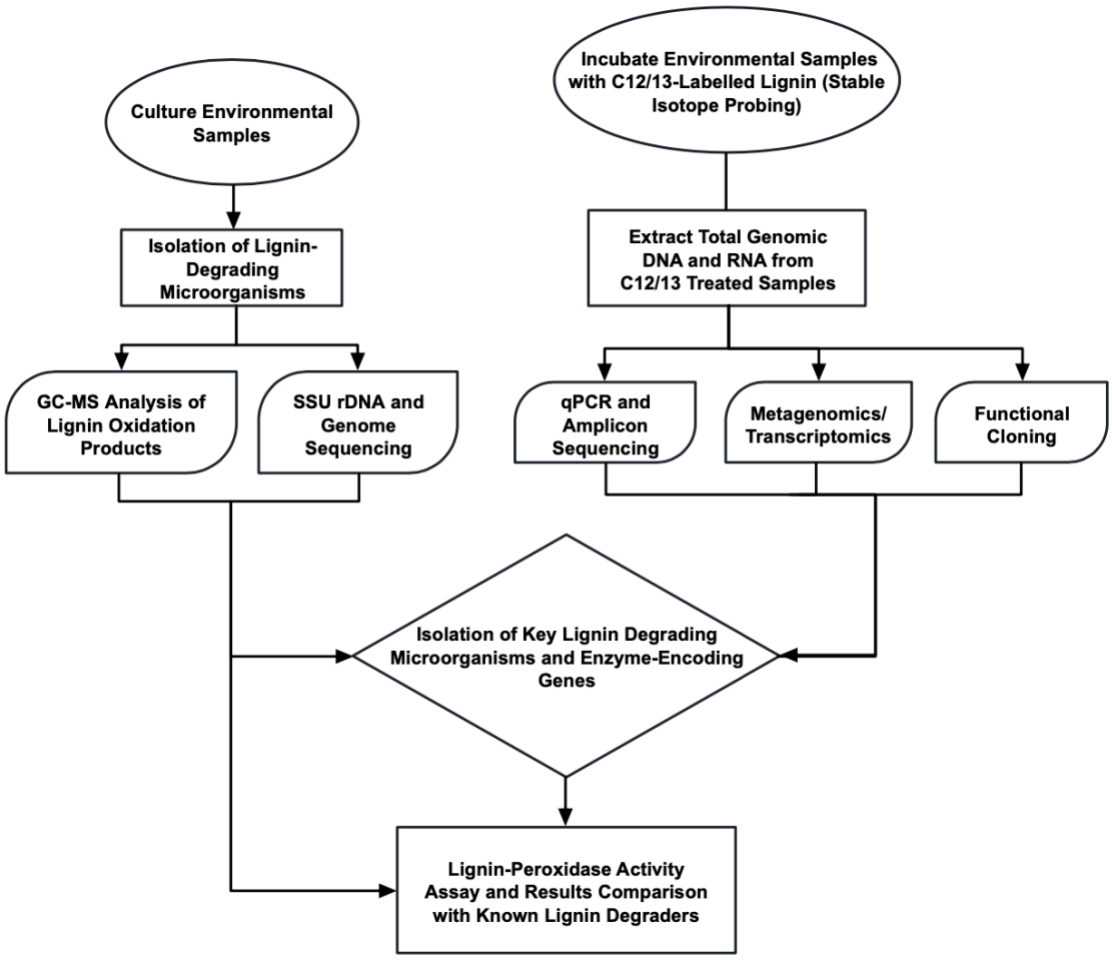

Figure 6 *– Flowchart of project objectives, showing culture-dependent and culture-independent methodologies used.*

2 Chapter 2: Methods and Materials

2.1 Chemicals and Reagents

All analytical grade and molecular biology grade chemicals and reagents used in this study were obtained from Sigma-Aldrich (Missouri, USA) (including MilliQ distilled water for media preparation), Fisher Scientific (Loughborough, UK), ThermoFisher (Massachusetts, USA), Promega (Southampton, UK) and Quiagen (Germany), unless otherwise stated in text. Gases were obtained from BOS (UK). Ultracentrifuge and the accompanying rotor and tubes were obtained from Beckman Coulter (California, USA).

2.2 Environmental Samples

The environmental samples used in this project are detailed in **Table 2** below. The catalogue is composed of three soil samples from Olduvai Gorge, Tanzania (obtained by Dr Clayton Magill, Lyell Fellow), two elephant faeces samples from Bristol Zoo, UK (obtained by Dr Ryan Pereira, Lyell Fellow), two Hickling Broad sediment samples from Norwich, UK (obtained by Peter Rivera, UEA PhD Student) and three compost samples (provided by co-supervisor Dr Thomas Aspray, Solidsense Ltd). These samples underwent analysis via culture-dependent and culture-independent approaches. From this point forward, the environmental samples will be referred to by their sample ID codes detailed in **Table 2**.

***Table 2** - Catalogue of environmental samples used in this project.*

Sample	Sample ID Code	Sample Type	Location	Sample Year
Savanna Soil	SS	Surface (O horizon) soil - from under vegetation	Olduvai Gorge, Tanzania	Apr-14
Zebra Grazed-Soil	ZS	Soil/sediment from area grazed by African Zebra	Olduvai Gorge, Tanzania	Apr-14
Termite Mound Soil	TM	Soil from area surrounding a Termite Mound	Olduvai Gorge, Tanzania	Jun-14
Elephant 1 Faeces	EF1	Faecal matter from African Elephant 1	Bristol Zoo, UK	Oct-16
Elephant 2 Faeces	EF2	Faecal matter from African Elephant 2	Bristol Zoo, UK	Oct-16
Sediment (centre)	HS(c)	Sediment from the centre of Hickling Broad	Hickling, UK	May-18
Sediment (shore)	HS(s)	Sediment from the shore of Hickling Broad	Hickling, UK	May-18
Compost 5-4	C54	Compost formed from green and food waste – mature – maintained in-vessel	Glasgow, UK	May-18
Compost 5-5	C55	Compost formed from green and food waste – sanitised – maintained in vessel	Glasgow, UK	May-18
Compost 6-1	C61	Compost formed from green waste – mature – maintained ex-situ	Glasgow, UK	Jun-18

Furthermore, gut and bore-hole samples were also obtained from marine wood-boring molluscs (*Xylophaga dorsalis*) within wood from Osterfjorden, Norway (near Bergen) as part of a collaborative project with Robert Harbour (PhD student) and Professor Andrew Sweetman (detailed in **Section 1.2.5.**). Whilst these samples underwent analysis via culture-independent methodologies, the results of this work is not included in this thesis.

2.2.1 Olduvai Gorge Soil Sample Descriptions

Three soil samples, taken from distinctly different environments within Olduvai Gorge, Tanzania, Africa, were used in this study. The samples, taken in April 2014 and June 2014, were surface soil (organic horizon) from under vegetation (SS; savanna soil), soil taken from within close proximity to a modern termite mound (TM; termite mound soil) and from a watering hole commonly grazed by Zebra (ZS; zebra-grazed soil) in Olduvai Gorge. These samples were provided by Dr Clayton Magill of The Lyell Centre, Heriot Watt University, Edinburgh, UK. Samples were taken and stored in plastic sampling bags at room temperature until further analysis.

The savanna soil sample (SS) was collected in April 2014 from Olduvai Gorge, Tanzania (Coordinates: S 3º12'16.5"E 35º16'30.5"; 1904 m.a.s.l. (metres above sea level)). The sample was collected from a "contemporary ground-surface" (O/A horizon [1-3 cm of topsoil]) that featured local vegetation, such as *Croton macrostachys* and *Calotropis procera* ('Sodom Apple'). The area also featured abundant forbs and low-growing grasses (species unknown). This sample was one of ten samples collected from 75m along a ~200 m transect. The zebra-grazed soil sample (ZS) was also collected in April 2014 from Olduvai Gorge, Tanzania (Coordinates: S 3º12'10.7"E 35º16'42.2"; 1932 m.a.s.l.).The sample was collected from a "contemporary ground-surface" (O/A horizon [1-3 cm of topsoil]) featuring grassland that had 'heavy grazing' (i.e., mastication-cut blades of high-growing grasses [species unknown]) from zebras (confirmation via visual inspection). This sample was the last of the transect samples, collected from 200m along a ~200m transect. The termite-mound soil sample (TM) was collected in June 2014 from Olduvai Gorge, Tanzania (Coordinates: S2º59'11.0"E 35º22'36.7"; 1386 m.a.s.l.).The sample was representative of a modern (i.e. contemporary mound that is unoccupied or 'dead') termite nest (a.k.a. termitarium) from an unknown species of the genus *Macrotermes*.

2.2.2 Elephant Faeces Sample Descriptions

Two elephant faeces samples (taken October 2016), derived from two different elephants (EF1 and EF2) based at Bristol Zoo, Bristol, UK, were used in this study. Samples were stored in plastic sampling bags and stored at 4ºC in a fridge until further analysis.

2.2.3 Hickling Broad Sediment Sample Descriptions

Two lake sediment samples, derived from the centre (HS(c)) and shore (HS(s)) of the Hickling Broad, Norwich, UK, were used in this study. Using sterile plastic sampling tubes, samples were taken from surface sediment (from centre and shore of broad) and stored at 4 ºC until further analysis.

2.2.4 Compost Soil Sample Descriptions

Three compost samples, each varying in either maturity or composition, were used in this study. Compost 5-4 (C54) and compost 5-5 (C55), both sampled in May 2018, were derived from green and food waste and were maintained in-vessel (actively aerated composting chambers), however C54 was composted to maturity whereas C55 only

reached the sanitization stage. Compost sanitisation is regarded as the stage at which temperatures peak within the compost due to accelerated microbial degradation of the organic matter contained within, leading to the destruction of pathogenic organisms and therefore sanitisation of the compost (Mengistu *et al*., 2017). Compost 6-1 (C61), on the other hand, was a mature compost derived from green waste only and was maintained ex-situ (windrows or heaps maintained on concrete pads and turned with a bespoke windrow turner or manually with a shovel). Samples were stored in plastic sampling bags at room temperature until further analysis.

2.3 Cultivation of Microorganisms

2.3.1 Growth Media

Microorganisms in this study were routinely isolated and cultured in mineral salt medium (MSM) (Jimenez *et al*, 2013) and Lysogeny Broth (LB) Medium (Bertani, 1951), respectively. MSM is a minimal medium (containing no carbon source), whereas LB is a nutrient-rich medium. This specific MSM medium was chosen as it was a minimal medium (containing no carbon source) and was previously used by Jimenez *et al* (2013) to grow plant-degrading microorganisms. LB medium was used as it is a medium routinely used to grow microorganisms. Within the oxidative enzyme assay and GC-MS experiments carried out in this study, yeast extract mineral salt (YEMS) medium (adapted from Tuncer *et al*., 2004 by industrial partner) was used to culture microorganisms, which is also a nutrient-rich medium.

MSM, along with the associated trace element and vitamin solutions, were prepared according to the protocol detailed by Jimenez *et al* (2013), with minor modifications. MSM was composed of 7 g/l Na_2HPO_4; 2 g/l K_2HPO_4; 1 g/l $(NH_4)SO_4$; 0.1 g/l $Ca(NO_3)_2$; 0.2 g/l $MgCl_2$. MSM trace element solution was composed of 2.5 g/l ethylenediaminetetraacetic acid (EDTA); 1.5 g/l $FeSO_4$; 0.025 g/l $CoCl_2$; 0.025 g/l $ZnSO_4$; 0.015 g/l $MnCl_2$; 0.015 $NaMoO_4$; 0.01 g/l $NiCl_2$; 0.02 g/l H_3Bo_3; 0.005 g/l $CuCl_2$. MSM vitamin solution was composed of 0.1 g/l Ca-panthothenate; 0.1 g/l cyanocobalamine (vitamin B12); 0.1 g/l nicotinic acid; 0.1 g/l pyridoxal (pyridoxal hydrochloride); 0.1 g/l riboflavin; 0.1 g/l thiamin (thiamine hydrochloride); 0.001 g/l biotin; 0.1 g/l folic acid.

LB medium (Miller) was composed of 10 g/l typtone, 5 g/l yeast extract and 10 g/l NaCl. LB medium had no accompanying (e.g. trace/vitamin) solutions.

YEMS medium was composed of 6 g/l yeast extract; 0.1 g/l $(NH_4)_2SO_4$; 0.3 g/l NaCl; 0.1 g/l $MgSO_4$. YEMS mediums accompanying trace element solution was composed of 1 g/l $FeSO_4$; 0.9 g/l $ZnSO_4$; 0.2 g/l $MnSO_4$.

Media were prepared in glassware sterilised via autoclave and acid washing (10% HCl bath). pH of the MSM and LB media was adjusted to the required level (see **Table 3** below) using a pH meter (Fisherbrand Accumet XL150 Benchtop pH Meter) and hydrochloric acid (HCl) or sodium hydroxide (NaOH). Once prepared, all growth media were sterilised via autoclave. Trace elements/vitamin solutions, along with carbon sources, were added to media following sterilisation during cultivation procedures.

2.3.2 Enrichment and Isolation of Lignin-Degrading Microorganisms

Environmental samples (listed in **Table 2** above) underwent enrichment with Alkali Lignin (Sigma Aldrich, product number: 370959) for time periods depending on activity of the given sample. Enrichment culture incubation information detailed in **Table 3** below. These enrichment conditions (incubation temperature and pH) were chosen to match the natural conditions of each environment, respectively, as closely as possible within a laboratory environment. Enrichment cultures were prepared in sterile (autoclaved and acid washed) 100 ml serum vials with homogenized environmental sample (50 g/l) and MSM with trace element and vitamin solution (1% of each). Alkali lignin was used as the carbon source in all enrichment cultures (10 g/l), with replicates also being prepared with glucose as an additional carbon source (10 g/l). Specifically, cultures contained 20 ml growth medium (19.96 ml MSM, 20 µl trace element solution and 20 µl vitamin solution) with 0.2 g lignin and 1 g environmental sample, with 200 µl of glucose stock (1 g/ml) being added to glucose replicates. Furthermore, anoxic replicates were also prepared for the elephant faeces, compost and Hickling sediment sample enrichments by flushing serum vials with N_2/CO_2. Once prepared and inoculated, enrichment cultures were sealed with rubber stoppers and crimp caps, followed by incubation according to conditions stated below (Incubator: Grant-bio Orbital Shaker – Incubator ES-80).

Table 3 *- Environmental sample enrichment details including enrichment temperature, pH and agitation conditions, as well as additional replicates and enrichment time.*

	Enrichment Conditions			Additional Replicates		Enrichment time	
Sample Set	**Temperature (°C)**	**pH**	**Agitation (rpm)**	**Glucose replicates**	**Oxic/Anoxic replicates**	**Total**	**Subcultures**
Olduvai Gorge	25	7.5	100	Yes	No	12 weeks	Biweekly
Elephant Faeces	37	6.5	100	Yes	Yes	4 weeks	Weekly
Compost	50	7	100	Yes	Yes	4 weeks	Weekly
Hickling Sediment	25	7	100	Yes	Yes	4 weeks	Weekly

Subcultures were prepared at time-points throughout the enrichment process (weekly/biweekly) by transferring 200 µl of the original enrichment culture into new growth medium and substrate (prepared according to original enrichment cultures). Subcultures were transferred to new medium and substrate in a similar fashion weekly. After the enrichment period had ended, glycerol stocks were prepared for the enrichment cultures. After subcultures had been transferred several times, they were spread onto solid, selective growth medium (comprised of MSM, lignin and agar) and nutrient-rich medium (comprised of LB and agar) in petri dishes according to the spread plate technique. Plates were incubated at temperatures respective to their enrichment conditions (detailed in **Table 3** above). Morphologically different (size, colour, texture) colonies were picked from the plate with a sterile toothpick and streaked onto new agar plates. Biomass was taken from the colony streak plates for preparation of 13-streak plates to check purity. Despite being isolated with lignin as sole carbon source, isolates were re-streaked onto MSM-lignin-agar plates to confirm lignin utilization capability.

2.3.3 Determination of Bacterial Isolate Optimal Growth Conditions

Growth experiments were carried out with bacterial isolates in order to determine optimal growth conditions. Starter cultures were prepared for Olduvai Gorge (OG) and elephant faeces (EF) isolates with YEMS medium (plus trace element solution) which were aseptically inoculated with isolate biomass from LB-agar plates. Once grown, sterile serum vials containing 29 ml YEMS medium (and trace element solution) (pH 7) with 0.05 g/L alkali lignin were inoculated with 1 ml isolate starter culture to a starting OD_{600} of 0.4. For the OG isolates, replicates were incubated at 25 °C or 37 °C over a period of

3 days (150 rpm agitation). A lower concentration of lignin was used in this experiment compared to the enrichment cultures to avoid the effects of lignin interfering with spectrophotometer absorbance readings. For the EF isolates, replicates were incubated at 25 °C, 37 °C and 45 °C over a period of 5 days (150 rpm agitation). During culture incubation, daily samples were taken for OD_{600} measurements with a spectrophotometer (Jenway Visible Spectrophotometer). In both experiments, YEMS medium with lignin was used as a 'blank' for the spectrophotometer, to which replicates were compared. Negative controls were included in both experiments which were replicates containing YEMS medium and lignin (similar concentrations to test replicates) with no inoculum (i.e. uninoculated controls). All culture preparation work and subsequent sampling was carried out under a laminar flow hood to ensure sterile working conditions therefore limiting culture contamination.

2.4 Screening of Microorganisms for Lignin-Degrading Activity

2.4.1 Oxidative Enzyme Activity Analysis of Isolates

Oxidative enzyme assays were carried out during placement with my industrial CASE partner Recircle Ltd.. These assays involved the testing of culture supernatant for the presence of oxidative enzymes, whereby substrates specific to certain enzymes were added to culture supernatant, along with enzyme cofactors. The oxidation of these substrates by the enzymes present within the isolate cultures caused a colour change which could be detected using a spectrophotometer. A 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid (ABTS) assay (Murugesan et al., 2007), a 2,4-dichlorophenol (2,4-DCP) assay (Antonopoulos et al., 2001) (modified by CASE partner Recircle Ltd.) and a 2,6-dimethoxyphenol (2,6-DMP) assay (Salvachua *et al*., 2015) were used to determine the oxidative enzyme activity of environmental sample isolates. The ABTS assay was used to measure laccase activity, the 2,4-DCP assay used to measure general peroxidase activity and the 2,6-DMP assay, a three-stage assay, used to measure laccase, general peroxidase and manganese-dependent peroxidase, in consecutive order.

2.4.1.1 Enzyme Assay Culture Preparation

Isolates were prepared in cultures containing 50 ml YEMS medium (plus 1 µl/ml accompanying trace element solution) with lignin (0.05 g/l conc.) and MSM medium (plus

1 µl/ml accompanying trace element and vitamin solutions) with lignin and glucose (0.05 g/l and 10 g/l, respectively). All cultures were inoculated from starter cultures (initial OD_{600} 0.4). The results presented in this thesis are derived from two enzyme assay experiments (experiments/rounds 5 and 6). In experiment (round) 5, all isolate cultures were incubated at 25 ºC for 10 days, whereas in experiment (round) 6, the isolate cultures were incubated at their respective enrichment temperatures (Olduvai Gorge isolates at 25ºC, elephant faeces isolates at 37ºC) for 13 days. In both experiments, daily samples were taken for enzyme assay analysis. A *Rhodococcus* sp., provided by industrial partner Recircle Ltd., was used as a positive control for the 2,4-DCP assay (due to evidence of peroxidase activity in past trials carried out by Recircle Ltd.) and was also tested in the ABTS and 2,6-DMP assays. All culture preparation work and subsequent sampling was carried out under a laminar flow hood to ensure sterile working conditions therefore limiting culture contamination.

2.4.1.2 Oxidative Enzyme Assays

All assay reaction mixtures (detailed below) were prepared in 96-well plates and isolate culture supernatant was added to wells in duplicate. Culture supernatant was obtained by centrifuging a culture aliquot (removing cell mass) and removing liquid from the aliquot. Assays were carried out using a spectrophotometric 96-well microplate reader (Labtech LT-4500) and over a set time limit (2-10 minutes), with absorbance measurements being taken every minute. Absorbance change (ΔAbs) over time (t), along with dilution factor was used to calculate enzyme activity (U/L) according to the formula specified in **Figure 7**. Premixed solutions of reaction components were tested and used for ease of 96-well plate preparation, and reaction mixtures were added to plate wells with a multi-channel pipette. Premixed reaction mixtures contained buffer, enzyme substrate (e.g. 2,4-DCP) and co-substrates (e.g. H_2O_2), if required. One unit (1U) of enzyme activity is defined as the quantity of enzyme required to produce 1 µmol of reaction product per minute (under the reaction conditions used).

$$U/mg_{solid} = \frac{\frac{\Delta Abs}{t} \times V \times h \times \varepsilon}{u}$$

***Figure 7** - Enzyme activity formula provided by Sigma-Aldrich with ABTS as substrate. ΔAbs = change in absorbance; t = time (mins); V = total reaction volume; h = light path (cm^{-1}); u = volume of enzyme; ε = extinction coefficient of substrate (ABTS/2,4-DCP/2,6-DMP).*

Laccase enzyme (from *Trametes versicolor*) (0.02 mg/ml, final activity 0.01 U/ml) (Sigma-Aldrich) was used as an enzyme positive control in the ABTS and 2,6-DMP (stage 1) assays and peroxidase (from horseradish) (HRP) (0.002 mg/ml, final activity 0.03 U/ml) (Sigma-Aldrich) was used in the 2,4-DCP and 2,6-DMP (stage 2) assays. An uninoculated negative control culture was also prepared for each medium/substrate combination.

2.4.1.2.1 ABTS Enzyme Assay

The ABTS assay, used to measure laccase activity, was carried out in (final conc.) 100 mM sodium acetate buffer (pH 5) and 10 mM ABTS with enzyme assay culture supernatant and absorbance change (at 420 nm) was measured at room temperature (ε = 36,000 M^{-1} cm^{-1}). 4 µl of culture supernatant was added to the reaction mixture (200 µl total volume).

2.4.1.2.2 2,4-DCP Enzyme Assay

The 2,4-DCP assay, used to measure peroxidase activity, was carried out in (final conc.) 100 mM Tris-HCl buffer (pH 8), 16 mM 4-aminoantipyrine, 25 mM 2,4-DCP and 50 mM hydrogen peroxide with enzyme assay culture supernatant and absorbance change (at 510 nm) was measured at room temperature (ε = 6,710 M^{-1} cm^{-1}). 40 µl of culture supernatant was added to the reaction mixture (200 µl total volume).

2.4.1.2.3 2,6-DMP Enzyme Assay

The 2,6-DMP assay, used to measure laccase, peroxidase and manganese-dependent peroxidase activity (performed in 3 stages within the same plate), was carried out in (final conc.) 0.1 mM sodium malonate buffer (pH 7), 125 mM 2,6-DMP and enzyme assay culture supernatant, followed by addition of 5 mM hydrogen peroxide and then 5 mM manganese sulfate. An overview of the stages involved in this assay are shown in **Figure 8** below. Absorbance change (at 469 nm) was measured at room temperature (ε = 55,000 M^{-1} cm^{-1}). 20 µl of culture supernatant was added to the reaction mixture (200 µl total volume). Activity data from the laccase stage of this assay was subtracted from peroxidase stage data, and the peroxidase stage data from the manganese-dependent peroxidase data.

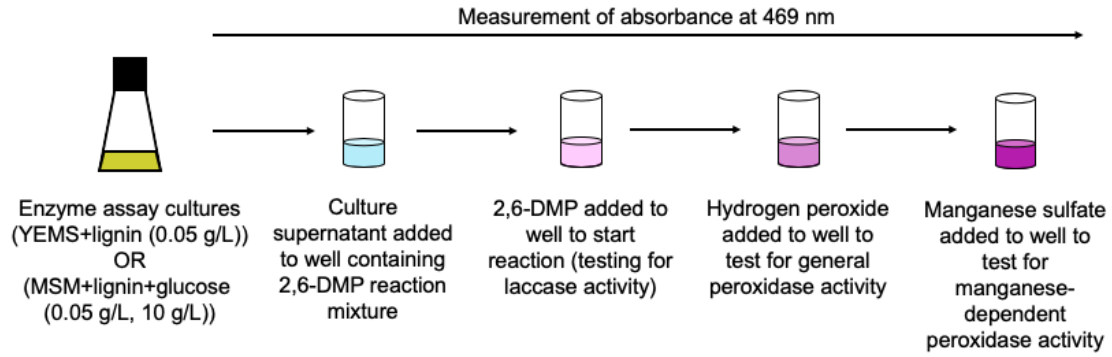

Figure 8 *- Simplified overview of 2,6-DMP assay method.*

2.4.2 Gas Chromatography-Mass Spectrometry (GC-MS) Analysis of Lignin-Derived Phenolic Compounds in Isolate Cultures

2.4.2.1 Isolate Culture Preparation

Isolate cultures were prepared with YEMS medium (described above) and lignin (0.1 g/l). All cultures were inoculated from starter cultures (initial OD_{600} 0.4). The isolate cultures were incubated at their respective enrichment temperatures (25 ºC for Olduvai Gorge isolates, 37 ºC for elephant faeces isolates) for three weeks, with daily samples being taken for enzyme assay analysis. A *Rhodococcus* sp., provided by industrial partner Recircle Ltd., was also tested in this experiment. All culture preparation work and subsequent sampling was carried out under a laminar flow hood to ensure sterile working conditions therefore limiting culture contamination.

2.4.2.1.1 Ethyl Acetate (EtoAc) Extraction of Lignin-Derived Phenolic Compounds from Culture Supernatant

2 ml of isolate culture sample was centrifuged at 16,000 rpm for 5 minutes to obtain supernatant. 1 ml culture supernatant was transferred into a 5 ml glass vial and approximately 1 ml ethyl acetate (EtoAc) was added to the supernatant. The glass vial containing the mixture was then sealed and vortexed. The top solvent layer was transferred to a new 5 ml vial using a glass Pasteur pipette. This process (addition of EtoAc and transfer of the solvent layer to the glass vial) was repeated two more times. EtoAc extracted lignin-derived phenolic compounds (LDPC) preparations were flushed with nitrogen gas and heated using a heating block (to 25 °C) to evaporate EtoAc. Approximately 100 µl of hexane was added to dried LDPCs using a graduated glass

syringe, followed by 25 µl of bis(trimethylsilyl)trifluoroacetamide (BSTFA) (with a BSTFA-designated glass syringe) and the vial was heated to 60 °C in an oven for 15 minutes to derivatize. 1 ml glass vials were prepared with glass inserts and the total derivatized LDPC preparation was transferred to the insert with a glass Pasteur pipette. Derivatized LDPC preparations were evaporated again following the method detailed above and, once dry, 100 µl of hexane was added to vial.

2.4.2.2 Gas Chromatography-Mass Spectrometry (GC-MS) Analysis of Supernatant EtoAc Extracts

Vials containing LDPC preparations were placed onto a GC-MS autosampler and run on a DB-5 5%(phenyl)- methyloplysiloxane capillary column inside a Trace 1300 Gas Chromatograph with ISQ LT Single Quadrupole Mass Spectrometer. 1 µl of derivatized extracts was injected into the GC-MS injector port and conditions were as follows: split ratio – 15:1; injector – 200 °C; flow rate – 1.5 ml/min; oven ramp – 100 °C (1 min), (4 °C/min), 280 °C (15 min); detection mode – m/z 50-650 (full scan). Total ion chromatogram (TIC) data was observed and analysed using Chromeleon chromatogram analysis software (Thermo Fisher Scientific). Phenolic product peaks were verified via matching retention times within TIC data to previous literature and investigating major ions, and peak area data was pulled from verified peaks to estimate product abundance.

2.5 Nucleic Acid Extraction and Manipulation

2.5.1 Extraction of Total Nucleic Acids from Environmental Samples

Total nucleic acids were extracted from environmental samples and pure cultures using phenol-chloroform phase extraction methods involving either 6% hexadecyltrimethylammonium bromide (CTAB) or 3% sodium dodecyl sulphate (SDS) solutions, based on methods described by Griffiths *et al* (2000) and Burgmann *et al* (2003), respectively. Both extraction methods were carried out on environmental samples to determine the best method for the given sample (i.e. method which produces best quality and highest concentration of DNA). SDS extracts of Olduvai Gorge soil and Hickling sediment samples, and CTAB extracts of elephant faeces and compost soil samples were selected for further analysis based on these factors.

2.5.1.1 CTAB Nucleic Acid Extraction Protocol

0.5 g of environmental sample was added to a MPBio Lysing Matrix E bead tube, followed by 500 µl of 6% CTAB extraction buffer (final concentrations: 6% CTAB; 0.35 M NaCl; 120 mM potassium phosphate in water – pH 8) and 500 µl of phenol-chloroform-isoamyl alcohol (25:24:1, pH 8). Tubes containing sample were homogenized in a bead beater (FastPrep-24 5G, MPBio) for 45 s at 6 m/s, and put on ice. Tubes were centrifuged at full speed (16,000 x *g*) for 5 min at 4°C. The upper aqueous phase was extracted from the tube and transferred to a new 2 ml tube. An equal volume (approx. 700 µl) of phenol-chloroform-isoamyl alcohol was added to the aqueous phase, the tube was vortexed briefly and centrifuged at full speed for (16,000 x *g*) for 5 min at 4°C. The upper aqueous phase was extracted from the tube and transferred to a new 2 ml tube. An equal volume (approx. 700 µl) of chloroform-isoamyl alcohol was added to the aqueous phase, the tube was vortexed briefly and centrifuged at full speed (16,000 x *g*) for 5 min at 4°C. The upper aqueous phase was extracted from the tube and transferred to a new 2 ml tube. 2 volumes (approx. 1 ml) of 30% PEG-precipitation buffer (final concentrations: 30% PEG; 14.6% NaCl in water) was added to the aqueous phase and mixed by inversion. Nucleic acids were precipitated for 2 h at room temperature. After precipitation, precipitated nucleic acids were centrifuged at full speed (16,000 x *g*) for 10 min at 4°C. The supernatant was discarded and the pellet was washed with 500 µl ice-cold 70% (vol/vol) ethanol. The tube was then centrifuged again at full speed (16,000 x *g*) for 10 min at 4°C. Ethanol was discarded from the tube and nucleic acid pellets were air dried (approx. 5 min). Dry nucleic acid pellets were resuspended in 100 µl nuclease-free water.

2.5.1.2 SDS Nucleic Acid Extraction Protocol

0.5 g of environmental sample was added to MPBio Lysing Matrix E bead tube, followed by 1 ml 3% SDS extraction buffer (final concentrations: 3% SDS; 0.2 M sodium phosphate buffer (pH 8); 1 M NaCl; 0.1 M EDTA (pH 8) in water – filter sterilised). Tubes containing sample were homogenized in a bead beater (FastPrep-24 5G, MPBio) for 45 s at 6 m/s, and put on ice. Tubes were then centrifuged at full speed (16,000 x *g*) for 5 min at 4°C. The upper aqueous phase was extracted from the tube and transferred to a new 2 ml tube. An equal volume (approx. 700 µl) of phenol-chloroform-isoamyl alcohol was added to the aqueous phase, the tube was vortexed briefly and then centrifuged at full speed for (16,000 x *g*) for 5 min at 4°C. The upper aqueous phase was extracted from the tube and transferred to a new 2 ml tube. An equal volume (approx. 700 µl) of chloroform-isoamyl alcohol was added to the aqueous phase, the tube was vortexed briefly and centrifuged at full speed (16,000 x *g*) for 5 min at 4°C. The upper aqueous

phase was extracted from the tube and transferred to a new 2 ml tube. 1 ml of 20% PEG-precipitation buffer (final concentrations: 20% PEG; 14.6% NaCl in water) was added to the aqueous phase and mixed by inversion. Nucleic acids were precipitated for 1 h at room temperature. After precipitation, precipitated nucleic acids were centrifuged at full speed (16,000 x *g*) for 30 min at 20°C. Supernatant was discarded and pellet was washed with 800 µl ice-cold 75% (vol/vol) ethanol. The tube was then centrifuged at full speed (16,000 x *g*) for 10 min at 4°C. The ethanol was discarded and the nucleic acid pellets were air dried (approx. 5 min). Dry nucleic acid pellets were resuspended in 100 µl of nuclease-free water.

2.5.2 DNA Purification

Environmental sample nucleic acid extracts (e.g. Olduvai Gorge sil samples) which were observed to have a “brown” hue (presumably due to humic acid content) underwent purification using Zymo Research OneStep PCR Inhibitor Removal Kit.

2.5.3 Nucleic Acid Manipulation

Agarose gel electrophoresis analysis (1.5% agarose gel prepared in 1x tris-acetate-EDTA (TAE) buffer) was used to determine presence and quality of extracted nucleic acids (gels viewed on UVP BioDoc-It^2 Imager). Nucleic acids were quantified for further analysis using a Qubit 4 Fluorometer using manufacturer reagents and operating procedure.

2.6 Polymerase Chain Reaction (PCR) and Gel Electrophoresis

2.6.1 Reaction Mixtures and Protocols

2.6.1.1 Bacterial 16S rRNA Gene Amplification

Bacterial 16S rRNA gene polymerase chain reaction (PCR) was carried out with PCR master-mix (25 µl reaction volume) containing (final concentrations) DreamTaq Green Buffer 1x (Invitrogen), 0.4 µg/µl bovine serum albumin (BSA, Sigma Aldrich), 0.2 mM dNTP mix, 0.2 µM of both 8F and 1392R 16S primers (Amann *et al.*, 1995) (8F: 5’-AGAGTTTGATCMTGGCTCAG-3’; 5’-1392R: ACGGGCGGTGTGTACA-3’) and DreamTaq polymerase enzyme (Invitrogen) in nuclease-free PCR water (Ambion).

For colony PCR, isolate colonies were picked into PCR tubes containing PCR mastermix using sterile toothpicks and bacterial 16S PCR protocol was carried out (25 cycles) (94°C for 5 minutes; 94°C for 30 seconds; 55°C for 45 seconds; 72°C for 1 minute; 72°C for 5 minutes; 4°C for remaining time) in an Applied biosystems (Thermo Fisher Scientific) Verti 96 well Thermal Cycler. PCR products underwent agarose gel electrophoresis (1.5% agarose gel prepared in 1x TAE buffer) (gels viewed on UVP BioDoc-It^2 Imager).

2.6.1.2 Eukaryotic 18S rRNA Gene Amplification

Eukaryotic 18S rRNA gene PCR was carried out with PCR master-mix (25 µl reaction volume) containing (final concentration) DreamTaq Green Buffer 1x; 0.2 mM dNTP mix; 0.2 µM of both Euk_1391f and EukBr general eukaryotic 18S primers (Amaral-Zettler *et al*., 2009) (Euk_1391f: 5'-GTACACACCGCCCGTC-3'; EukBr: 5'-TGATCCTTCTGCAGGTTCACCTAC-3') and DreamTaq polymerase enzyme in nuclease-free PCR water. PCR reaction conditions were as follows: initial denaturation at 94 ºC for 3 min followed by 35 cycles of 94 ºC for 45 secs, 55 ºC for 1 min and 72 ºC for 1.5 mins followed by final elongation of 72 ºC for 10 mins.

2.6.1.3 *dyp* Functional Gene Amplification

PCR was carried out with reaction mixtures similar to those used for bacterial 16S rRNA gene amplification (**Section 2.6.1.1** above), but with primers F-DYPR (5'-GAY CTG TGC TTY GAR CTS GC- 3′) and R-DYPR (5′- ASC CGA TRA ART ASG TGC C-3′) (Tian et al., 2016a). The *dyp* gene encodes DyP-type peroxidase enzyme which, as stated in the introduction of this thesis, have been shown to be a potent family of lignin-degrading enzymes. PCR reaction conditions were as follows: initial denaturation at 94 ºC for 10 min followed by 30 cycles of 94 ºC for 30 secs, 60 ºC for 40 secs and 72 ºC for 45 secs followed by final elongation of 72 ºC for 10 mins. Conditions adapted from Tian et al (2016a) – initial denature time (3 min to 10 min) and elongation cycle time (60 secs to 45 secs) – to lyse cells and reduce non-specific amplification, respectively. The expected *dyp* gene product size was 300-400 bp in length.

2.6.1.4 *lmco* Functional Gene Amplification

PCR was carried out with reaction mixtures similar to those used for bacterial 16S rRNA gene amplification (**Section 2.6.1.1** above), but with primers Cu1AF (5'-ACMWCBGTYCAYTGGCAYGG-3' and Cu2R (5' -GRCTGTGGTACCAGAANGTNCC-

3') (Kellner et al., 2008). The *lmco* gene encodes laccase-like multicopper oxidase which, like the DyP peroxidases, are key enzymes involved in microbial lignin degradation. PCR conditions were as follows: Denaturation at 94 ºC for 5 min followed by 35 cycles of 94 ºC for 30 secs, 50 ºC for 30 secs and 72 ºC for 35 secs followed by final elongation of 72 ºC for 5 mins. Conditions adapted from Kellner et al (2008) – increased denature time (3 mins to 5 mins) and decreased elongation time (80 secs to 35 secs) – to lyse cells and reduce non-specific amplification, respectively. The expected *lmco* gene product size was ~142 bp in length.

2.7 Nucleic Acid Sequencing: Sample Preparation and Next Generation Sequencing (NGS) Technologies

2.7.1 Sanger Sequencing of PCR Amplification Products and Analysis

2.7.1.1 Sample Preparation for Sanger Sequencing

Before being sent for sequencing, PCR products were first purified using the NucleoSpin Gel and PCR Clean-Up kit (Macherey-Nagel) (according to manufacturers protocol).

Clean PCR products which were to be sent for sequencing were prepared according to instructions detailed by Source Bioscience (Nottingham, UK) (https://www.sourcebioscience.com/genomics/sanger-sequencing/sample-requirements/). PCR amplifications products were diluted to 10 ng/µl concentration and primers were diluted to 3.2 pmol/µl concentration in nuclease-free water.

2.7.1.2 PCR Amplification Product Sanger Sequencing

All PCR amplification products which were selected for sequencing were sent to Source Bioscience for Sanger sequencing. Sequencing data was analysed in MEGAX sequence editing software (Kumer *et al*., 2018) and similar sequences of closest neighbours (with accession numbers) obtained from Silva rRNA database (https://www.arb-silva.de) were used to generate neighbour-joining phylogenetic trees using MEGA7 software, with 500 bootstrap replicates.

Functional gene (*dyp*/*lmco*) PCR amplification products were, like 16S/18S genes, analysed using MEGAX, however similar sequences of closest neighbours were

obtained from Blastn searches against the NCBI GenBank nucleotide database (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.7.2 Metagenomic and Metatranscriptomic Sequencing of Environmental Sample Nucleic Acid Extracts and Analysis

2.7.2.1 Metagenome Data Processing

Environmental sample nucleic acid extracts underwent RNase digestion and gDNA purification (Zymo DNA clean-up kit), followed by metagenomic sequencing (Illumina NovaSeq 6000, paired-end reads, 150 bp length) (Edinburgh Genomics, Edinburgh, UK). Furthermore, nucleic acid extracts from anoxic elephant faeces (E1-12-A1 and E2-12-A1) ^{12}C SIP incubation samples underwent RNA purification, reverse transcription into double stranded cDNA and were sent for metatranscriptomic sequencing (both cDNA samples pooled as individual samples were below nucleic acid concentration requirements).

Metagenomic sequencing data was processed using bioinformatic tools on an instance within the Cloud Infrastructure for Microbial Bioinformatics (CLIMB) (www.climb.ac.uk) server which is dedicated to the processing of large bioinformatic datasets. Metagenomic raw reads were first checked for quality using FastQC (v0.11.5) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), followed by trimming, joining of paired ends and removal of adapter sequences and low quality reads using Trimmomatic (v0.39) (http://www.usadellab.org/cms/index.php?page=trimmomatic). Trimmed raw reads were assembled using MegaHit (v1.2.9) (Li *et al*., 2015). Quality of metagenomic assemblies was checked using MetaQUAST (v5.0.2) (http://quast.sourceforge.net/metaquast). Annotation of assembled metagenomes was carried out using PROKKA (v1.14.6) (Seeman, 2014).

In order to bin metagenomes into high-quality metagenome-assembled genomes (MAGs), a DAS_Tool (Sieber *et al*., 2018) consensus binning pipeline was used, with input bins derived from CONCOCT (v1.1.0) (Alneberg *et al*., 2014), Maxbin2 (v2.2.4) (Wu *et al*., 2015), Metabat2 (v1.7) (Kang *et al*., 2015) and Binsanity (v0.4.4) (Graham *et al*., 2017). All binning tools were run with minimum contig length >2000 bp. Bins/MAGs were analysed (to determine bin completeness, contamination and taxonomic assignment, amongst other characteristics) using CheckM (v1.1.2) (Parks *et al*., 2016).

2.7.2.2 Metagenome Taxonomic Profiling

Metagenome taxonomic profiles were generated using the Kraken2 taxonomic sequence classification system (Wood et al., 2019) using trimmed paired read data (obtained via Trimmomatic). Metagenomic k-mers were queried against a database containing NCBI taxonomic information, including bacterial, archaeal and viral Refseq genomes. K-mers were mapped to that of the lowest common ancestor (LCA) of the genomes within the database. As the full kraken2 database is very large, a smaller pre-formatted database was used for taxonomic assignment (16 GB PlusPFP database – standard database including Protazoa, Fungi and Plant references). Relative abundances of taxonomic groups were calculated relative to total classified reads.

2.7.2.3 Metagenome CAZy Gene Analysis

A bioinformatic pipeline for identification and quantification of genes encoding Carbohydrate-Active enZYmes (CAZymes) was set up according to the methodologies used by Wilhelm et al. 2018, to functionally analyse the metagenome data. The CAZy bioinformatic pipeline (CAZy_utils, available from: https://github.com/nielshanson/CAZy_utils) involved the use of the CAZyme online database (http://www.cazy.org), which was converted into a DIAMOND (Buchfink *et al*., 2021) blastx searchable format, against which our metagenomes were searched in order to identify and quantify lignin-active genes and microorganisms responsible for their presence. The whole CAZy database was downloaded (encompassing glycoside hydrolases, glycosyltransferases, polysaccharide lyases, carbohydrate esterases and auxiliary enzymes) and loaded into an SQLite database which was then used to download the CAZy RefSeq sequences from NCBI (via the NCBI API), creating a fasta file compatible with DIAMOND. The RefSeq sequences were then formatted into a DIAMOND database and metagenomic assemblies were blasted against the database with DIAMOND Blastx, with a minimum e-value of 0.0001 (same e-value used by Wilhelm et al., 2018). The DIAMOND Blastx results were then fed into MEGAN6 (community edition) (Huson et al., 2016) which maps NCBI-nr accessions to taxonomic and functional classes (mapping file provided by MEGAN6) allowing taxonomic and functional annotation. The EC and InterPro2Go databases (included in MEGAN6) were used for extracting the lignocellulose degrading enzyme hits.

Aligned reads from CAZy laccase and DyP peroxidase gene assignments within the environmental sample metagenome assembly contigs were extracted from MEGAN6 and were searched against the NCBI GenBank nucleotide database using blastn

searches. The assigned laccase and *dyp* sequences were aligned with similar/neighbouring sequences obtained from blastn searches and neighbour-joining phylogenetic trees were constructed from these alignments using MEGAX software.

2.7.3 Microbial SSU rRNA Gene Amplicon Sequencing of Environmental Sample Nucleic Acid Extracts and Analysis

DNA extracts from the environmental samples were sent for high-throughput microbial 16S and Eukaryotic 18S rRNA gene amplicon sequencing (Illumina MiSeq, Nu-Omics, Northumbria University, Newcastle, UK) using 16S V4 primers (515F: 5'-GTGCCAGCMGCCGCGGTAA-3', 806R: 5'-GGACTACHVGGGTWTCTAAT-3') (Kozich *et al*. 2013) and the Earth Microbiome 18S primers (Euk1391f: 5'-GTACACACCGCCCGTC-3', EukBr: 5'-TGATCCTTCTGCAGGTTCACCTAC-3') (Thompson *et al*., 2017), respectively.

The microbial 16S/18S rRNA gene amplicon sequencing data obtained from environmental samples was processed and analysed using the QIIME 2 pipeline (Bolyen *et al*., 2019). QIIME 2 was installed onto the CLIMB instance within a Conda environment. Briefly, the 16S data was processed by importing into a QIIME 2 artefact, followed by demultiplexing and trimming, denoising and clustering (>97%) into amplicon sequence variants (ASV's) with DADA2. The DADA2 pipeline is a plugin which carries out multiple steps in one command, such as: trims sequences; filters Illumina phiX reads and removes/corrects 'noisy' reads; removes sequence replicates (identical sequences); clusters similar sequences into amplicon sequence variants (ASV); checks for chimeric sequences; and joins paired-end reads (if dealing with paired-end data). The outputs of this pipeline are the feature table (sequence frequencies within samples) and representative sequence data (map of ASV identifiers to representative sequences) which can be used for an array of different downstream analyses. Taxonomy was assigned to the representative sequence data using a Silva classifier (silva-132-99-515-806-nb-classifier), to determine identity and abundance of microorganisms at different taxonomic levels.

Beta diversity analysis was carried out on the (demultiplexed, trimmed and denoised) bacterial 16S rRNA gene amplicon sequencing data. This analysis was carried out as part of the QIIME 2 pipeline and the weighted and unweighted Unifrac distance matrixes were used for PCoA plotting.

2.8 Quantitative Polymerase Chain Reaction (qPCR)

All qPCR work was carried out using an Applied biosystems (Thermo Fisher Scientific) QuantStudio 3 Real-Time PCR System.

2.8.1 Bacterial 16S rRNA Gene qPCR Amplification

The bacterial 16S rRNA gene was amplified using primers Ba519f (5'-CAGCMGCCGCGGTAANWC - 3′) and Ba907r (5′- CCGTCAATTCMTTTRAGTT -3′) (Lane, 1991). PCR reactions were carried out in 25 µl total volume, containing 12.5 µl SYBR Green Jumpstart Taq Readymix (2x), 7 µl nuclease-free water, 4 µl magnesium chloride ($MgCl_2$) solution (25mM), 0.25 µl BSA (20 µg/µl), 0.125 µl of each primer and 1 µl of DNA/cDNA. All reactions carried out in duplicate in a 96-well plate. A standard for bacterial 16S rRNA gene quantification was prepared from a PCR amplification product obtained from *E. coli*. The qPCR protocol was as follows: initial denature at 96 ºC for 10 min followed by 40 cycles of 96 ºC for 30 sec, 52 ºC for 30 sec and 72 ºC for 1 min; followed by a brief denature step of 94 ºC for 15 secs and then 100 cycles of 75-95 ºC (+0.2 ºC per cycle).

2.8.2 Dyp Functional Gene qPCR Amplification

In order to obtain a standard for qPCR analysis using the *dyp* primer set, PCR with primers targeting the *dyp* functional gene was carried out with all Olduvai Gorge and elephant faeces isolates (PCR methodology detailed in **Section 2.6.1.3** above). *dyp* PCR products successfully obtained from isolates were sent for Sanger sequencing by Source BioScience. Sequences were searched against the NCBI GenBank nucleotide database using blastn searches. Isolate *dyp* sequences were aligned with similar/neighbouring sequences obtained from blastn searches and a neighbour-joining phylogenetic tree was constructed from this alignment using MEGA7 software.

dyp PCR products from isolates were tested as standards for qPCR (qPCR conditions detailed below) in a dilution series of 10^7-10^1 copy. Based on these results, a product derived from isolate OG-T3A2 was selected for future *dyp* qPCR runs due to high amplification (through high ΔRn values) and consistent melt curve results. The coefficient of determination (R^2) value of 0.999 (maximum $R^2 = 1$) suggests that data fit the standard curve of OG-T3A2 well. More information, including the results of this test, can be found in the **Appendix section 7.2**.

dyp qPCR conditions used in this experiment: Initial denaturation at 94 ºC for 3 min followed by 30 cycles of 94 °C for 30 secs, 60 °C for 40 secs and 72 °C for 30 secs followed by a brief denature step of 94 °C for 15 secs and then 100 cycles of 75-95 °C (+0.2 °C per cycle) (conditions adapted from Tian et al (2016)).

2.9 Nucleic Acid Stable Isotope Probing (SIP) of Environmental Samples with ^{13}C-Labelled Lignin

2.9.1 Preparation of SIP Microcosms

Environmental samples underwent SIP incubation in conditions similar to natural environmental conditions (in dark) with ^{13}C-labelled lignin, ^{12}C-lignin (natural abundance control), $^{13}CO_2$ (secondary labelling control) and no substrate control replicates. Microcosms were prepared in sterile 37 ml vials with 3 g environmental sample, 3 mg labelled/unlabelled lignin and 2.5/3 ml sterile MilliQ water (5% $^{13}CO_2$ injected into vial via syringe (1.5 ml)). 0.5 g samples were taken from the microcosms half-way through incubation ($T_{0.5}$) (x1 sample per incubation) and after incubation (T_1) (whole microcosm sampled). Samples which required long incubation time (Olduvai Gorge samples) were aired weekly to refresh oxygen (all incubations except $^{13}CO_2$ replicates). Samples were flash-frozen in liquid nitrogen directly after collection and stored in -80 °C freezer.

The Olduvai Gorge soil samples underwent SIP incubation at 25°C, 150 rpm for 28 days with three ^{13}C- lignin, two ^{12}C-lignin and one $^{13}CO_2$ control replicates per sample (SS, ZS and TM). Samples were taken from the OG sample incubations after 14 days (T0.5) and at the end of incubation at 28 days (T_1). The elephant faeces samples underwent incubation at 35 °C, 150 rpm for 14 days in both oxic and anoxic conditions (anoxic incubations flushed with nitrogen/CO_2) with two ^{13}C-lignin, two ^{12}C-lignin, one $^{13}CO_2$ control and one no-substrate control per sample (EF1 and EF2). The compost soil samples underwent incubation at 50 °C, 150 rpm for 14 days (oxic and anoxic conditions) with two ^{13}C-lignin, two ^{12}C-lignin, one $^{13}CO_2$ control and one no-substrate control per sample (5-4, 5-5 and 6-1). For the elephant faeces and compost incubations, samples were taken after 7 days ($T_{0.5}$) and at the end of incubation (T_1).

T_1 samples taken from SIP microcosms underwent nucleic acid extraction (according to methodology detailed above in **Section 2.5.1**) followed by the gradient ultracentrifugation and fractionation methodologies detailed below.

2.9.2 DNA SIP: Cesium Chloride (CsCl) Gradient Ultracentrifugation and Fractionation

4.5 µg of DNA (quantified using Qubit fluorometer) was loaded into CsCl (1.85 g/ml) solution and gradient buffer (GB) (final conc. in distilled water: 1 mM EDTA, 0.1 M Tris-HCl (pH 8.0), 0.1 M KCl) in Beckman & Coulter Quick Seal polypropylene tubes. Density was adjusted to 1.4029 (+/- 0.0002) using CsCl/GB solutions (measurements taken using a refractometer). CsCl DNA gradients underwent ultracentrifugation (VTI 65.2 vertical rotor) at 45,000 rpm for minimum 36 hours (20 °C) and gradients were fractionated into 12x 450 µl aliquots. The density of each fraction was measured to calculate CsCl concentration within each fraction. DNA fractions were precipitated using 30% polyethylene glycol (PEG) solution, with addition of 20 µg glycogen (Sigma) (to aid in pellet visualisation).

2.9.3 RNA Purification and Reverse Transcription into cDNA

The total nucleic extracts from SIP incubation samples underwent DNase digestion (RNase-free DNase set, Qiagen) followed by RNA purification (Qiagen RNeasy Mini Kit) to isolate RNA. Bacterial 16S rRNA gene PCR was carried out on RNA preparations in order to determine the presence of residual DNA within the extracts before use, and RNA quantification was carried out via Qubit fluorometer (RNA HS kit) to determine RNA concentration for further use (e.g. gradient loading).

RNA was reverse transcribed into cDNA in an initial reaction containing RNA fraction, Rnase-free water and hexanucleotide mix (1:50) (reaction carried out at 75 ºC for 5 mins in a thermal cycler), followed by a second reaction containing dNTP mix (10 mM), RNase inhibitor (Fisher), M-MLV reverse transcriptase (Promega) and M-MLV buffer (5x) (carried out at 37 ºC for 1 hr).

2.9.4 RNA SIP: Cesium Trifluoro Acetate (CsTFA) Gradient Ultracentrifugation and Fractionation

Up to 500 ng RNA was loaded into 2 g/ml (final conc.) caesium trifluoro acetate (CsTFA) solution (with 3.59% formamide) and gradient buffer (see composition above) (density adjusted to 1.3726 using a refractometer) in Beckman&Coulter Quick Seal polypropylene tubes. CsTFA RNA gradients underwent ultracentrifugation (VTI 65.2 vertical rotor) at

39,000 rpm for 65 hours (20 ºC) and gradients were fractionated into 12x 450 µl aliquots. The density of each fraction was measured using a refractometer to calculate CsTFA concentration within each fraction. RNA fractions were precipitated using sodium acetate (3M, pH 5.2), glycogen (20 µg) and 96% ice-cold ethanol.

3 Chapter 3: Isolation and Characterisation of Lignin-Degrading Bacterial Isolates from Environmental Samples

3.1 Introduction

This results chapter will address **Objectives 3, 4 and 5** stated in **Introduction Section 1.6.2**, which encompass the identification and functional characterisation of key lignin-degrading microorganisms from environmental samples (detailed in **Methods and Materials Section 2.2**). More specifically, the culture-dependent methods used to obtain the results presented in this chapter were employed to characterise novel lignin-degrading microorganisms and determine their potential for use in industrial lignocellulose processing (through enzyme activity and production of high-value compounds).

As discussed extensively in the **Introduction Chapter**, the use of microorganisms which degrade lignin through specialized oxidative enzyme pathways have been shown to be a promising alternative to current lignocellulose pre-treatment processes (Kumar *et al*., 2009; Ahmed *et al*., 2016; Tian *et al*., 2016). This has led to targeted exploration of various environments enriched in lignin-degrading microorganisms. Specifically, environments of interest encompass those in which lignocellulosic material is abundant (such as composts) and are more 'niche' compared to environments such as soil, for example the gastrointestinal tract of termites (Malherbe and Cloete, 2002; Bugg *et al*., 2011a; Zhou *et al*., 2017). The exploration of these environments could potentially provide highly capable lignin degrading species, which are better suited to the harsh conditions of industrial processing. Whilst the highly-efficient Basidiomycetes (white/brown-rot fungi) have been the primary focus of lignin-degradation studies, bacteria have been shown to be as effective, with the additional benefits of being better suited to industrial use due to higher proliferation rate and flexibility in terms of growth conditions (Vikman *et al*., 2002; Taylor *et al*., 2012; Janusz *et al*., 2017). Many studies have been carried out involving the isolation and characterisation of lignin-degrading microorganisms, primarily from soil environments (Bandounas *et* al., 2011; Taylor *et al*., 2012; Huang *et al*., 2013; Tian *et al*., 2016). Bacteria have been shown to employ oxidative enzymes such as laccase and peroxidase to degrade the lignin polymer, as well as the potent DyP-type peroxidases (Guerriero *et al*., 2016; Datta *et al*., 2017).

Despite this work, research regarding the use of microorganisms in the bioprocessing of lignocellulosic material is still in its infancy and is not yet at a level to compete with current processing methods (with regards to cost and effectiveness). Therefore, more research is required in order to discover novel lignin degrading candidates with industrial potential. Specifically, the identification of new and, potentially, more capable lignin-degrading species which are able to proliferate in industrial conditions and produce several lignin-active enzymes (such as laccases and peroxidases).

Therefore, in this study, previously underexplored environments such as compost, termite mound soil and elephant faeces were used, as well as soil and sediment samples, in order to obtain and characterise novel ligninolytic species. For the purposes of this study, novel isolates are defined as those that are previously unidentified as lignin-degrading species. The objectives of this study were achieved by, firstly, isolating and identifying lignin-degrading microorganisms using substrate-restricted enrichment culture methods and bacterial 16S rRNA gene sequencing, followed by phylogenetic analysis (**Objective 3**). This allowed the initial identification of isolates as previously known or unknown lignin-degrading species. The functionality (or lignin-degrading potential) of these isolates was then explored via oxidative enzyme-encoding functional gene screening and their capacity to produce oxidative enzymes and break down the lignin polymer via *in vitro* oxidative enzyme assays and biochemical analysis (**Objectives 4 and 5**, respectively). Oxidative enzyme assays provided information on the enzymatic capabilities of the isolates under various culturing conditions, therefore providing evidence for industrial potential. Biochemical (GC-MS) analysis of the lignin-derived compounds produced by isolates provided information on their lignin-degrading capability, as well as the identification of industrially relevant compounds produced. Due to the industrial advantage of bacterial species of fungal species discussed above, this study focussed on the isolation of bacterial degraders.

3.2 Results

3.2.1 Microbial Isolation and Initial Characterisation

In order to obtain lignin-degrading microorganisms from environmental samples, enrichment cultures were prepared with environmental samples in MSM media with either lignin as sole carbon source or lignin and glucose (medium composition detailed in **Method and Materials Chapter**), from which subcultures were prepared at time-points throughout (illustrated in **Figure 9**). The MSM medium used throughout this study, as mentioned in **Methods and Materials Section 2.3.1**, is a minimal medium with no

carbon source (other than that added by the researcher), making it useful for isolating substrate-specific microorganisms. The Olduvai Gorge samples underwent enrichment for 12 weeks, whereas the elephant faeces and compost samples underwent enrichment for 4 weeks (due to their observed increased activity). Lignin and glucose substrates were added at 10 g/L concentration, unless otherwise stated. As the alkali lignin substrate made enrichment cultures completely opaque, this prevented the use of spectrophotometry, therefore preventing absorption measurements to be taken and made the tracking of microbial proliferation and extent of lignin degradation difficult. Lignin degradation could, however, be visually tracked through the change of colour within the cultures (dark brown to light brown). Microbial growth could have been determined by various alternative methods (such as measurement of CO_2 production or colony counting on plates), however cell density was determined to not be necessary for this experiment.

Following colony selection and purification (via 13-streak plating), bacterial isolates were successfully obtained from the Olduvai Gorge, elephant faeces and compost sample enrichment cultures/subcultures (Olduvai Gorge isolates detailed in **Table 4**, elephant faeces isolates detailed in **Table 5**, compost isolates detailed in **Table 6**). Colonies which looked morphologically distinct from one another were selected, in order to prevent the isolation of replicate species. Isolates obtained from the Olduvai Gorge soil and elephant faeces samples were assigned unique IDs which include isolate ID and isolation time point in enrichment timeline. For example, T1A2 describes isolate A2, isolated at the 1-week timepoint during enrichment (T1). The compost isolates, on the other hand, due to the low number of isolates obtained, were simply named C1 and C2.

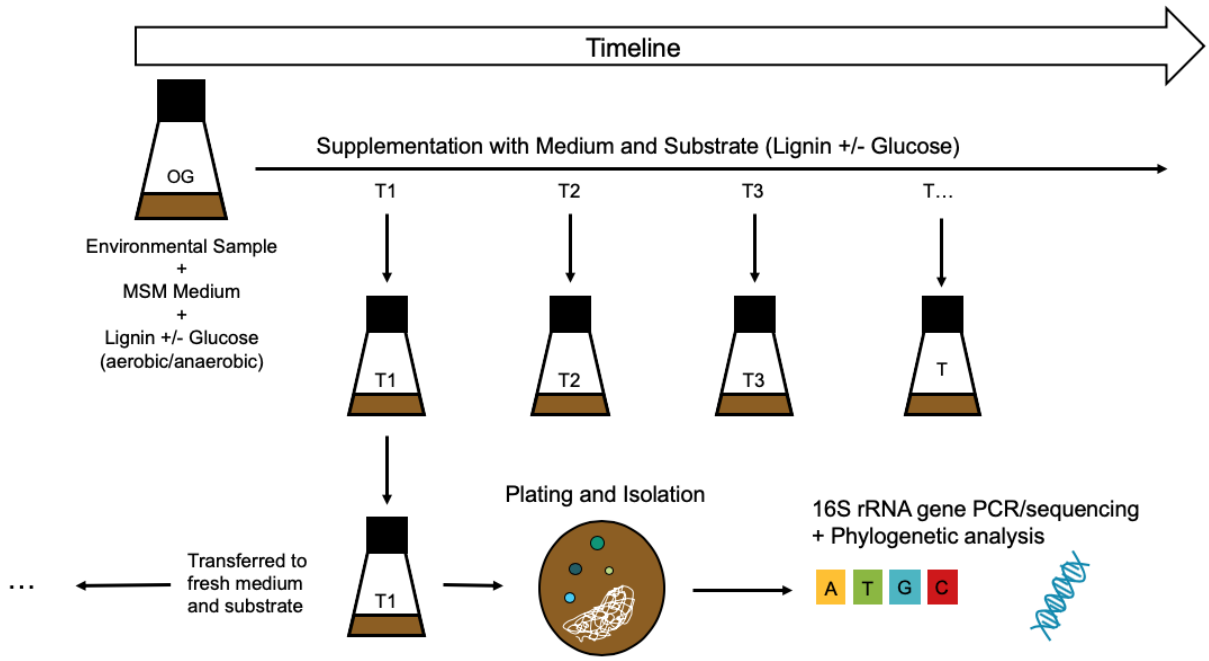

Figure 9 *– Schematic overview of enrichment process workflow. Enrichment culture incubation and subculture preparation with purification steps taken to obtain lignin-degrading isolates and 16S rRNA gene sequencing data.*

Table 4 *– Bacterial isolates obtained from Olduvai Gorge soil sample enrichment cultures from week 1 (T1), week 3 (T3) and week 5 (T5) time-points, including culture conditions and observed bacterial colony morphology. Isolates highlighted in red were selected for further study based on similarity to other isolates, as discussed below.*

			Replicate Culture Conditions		Colony Morphology				
Time-Point	**Isolate**	**Source**	**Carbon Source(s)**	**Oxygen Availability**	**Colour**	**Shape/Consistency**	**Identity of Closest Relative (Top BLAST Hit)**	**Sequence Identity [%]**	**Query Cover [%]**
T1	T1A1	Zebra-grazed	Lignin + Glucose	Oxic	Cream/clear	Round, regular	-	-	-
	T1A2	Zebra-grazed	Lignin + Glucose	Oxic	Cream/white	Round, regular	*Enterobacter sp.*	100	100
	T1A3	Zebra-grazed	Lignin + Glucose	Oxic	Cream/white	Round, regular	*Enterobacter amnigenus*	99	100
	T1B1	Termite-mound	Lignin + Glucose	Oxic	Cream/white	Round, regular	-	-	-
	T1B2	Termite-mound	Lignin + Glucose	Oxic	Cream/clear	Round, regular	*Ochrobactrum anthropi*	100	100
	T1B3	Termite-mound	Lignin + Glucose	Oxic	Cream/clear	Round, regular	*Ochrobactrum anthropi*	99	100
	T1C1	Termite-mound	Lignin + Glucose	Oxic	Cream/white	Round, regular	*Ochrobactrum sp.*	100	100
	T1D1	Savanna	Lignin	Oxic	White	Round, regular	*Ochrobactrum sp./Agrobacterium sp.*	99	100
	T1D2	Savanna	Lignin	Oxic	White/clear	Round, regular	*Pseudomonas sp.*	100	100
	T1D3	Savanna	Lignin	Oxic	White/clear	Round, regular	-	-	-
T3	T3A1	Savanna	Lignin	Oxic	White/cream	Round, regular	-	-	-
	T3A2	Savanna	Lignin	Oxic	White/cream	Round, regular	*Pseudomonas sp.*	99	100
	T3B1	Zebra-grazed	Lignin	Oxic	White/cream	Round, regular	*Stenotrophomonas maltophilia*	99	100
	T3B2	Zebra-grazed	Lignin	Oxic	White/cream	Round, regular	-	-	-
	T3C1	Termite-mound	Lignin	Oxic	White/cream	Round, regular	-	-	-
	T3C2	Termite-mound	Lignin	Oxic	White/cream	Round, regular	*Enterobacter hormaechai*	100	100
	T3F1	Termite-mound	Lignin + Glucose	Oxic	White/cream	Round, regular	-	-	-
	T3F2	Termite-mound	Lignin + Glucose	Oxic	White/cream	Round, regular	*Enterobacter hormaechai*	100	100
T5	T5A1	Savanna	Lignin	Oxic	White/cream	Round, regular	*Burkholderia sp.*	94	100
	T5A2	Savanna	Lignin	Oxic	White/cream	Round, irregular	*Pseudomonas sp.*	100	100
	T5B1	Savanna	Lignin + Glucose	Oxic	Yellow/cream	Round, regular	*Pseudescherichia vulgaris*	99	100
	T5C1	Zebra-grazed	Lignin	Oxic	White/cream	Round, regular	*Stenotrophonmonas sp.*	99	100
	T5C2	Zebra-grazed	Lignin	Oxic	White/cream	Round, irregular	*Pseudomonas sp.*	100	100
	T5D1	Termite-mound	Lignin	Oxic	White/cream	Round, irregular	*Enterobacter sp.*	99	100
	T5D2	Termite-mound	Lignin	Oxic	Clear	Irregular	*Pseudomonas sp.*	98	100
	T5D3	Termite-mound	Lignin	Oxic	White/cream	Round, regular	*Pantoea sp./Enterobacter sp.*	100	100
	T5E1	Termite-mound	Lignin	Oxic	White/cream	Round, regular	-	-	-
	T5E2	Termite-mound	Lignin	Oxic	White/cream	Irregular	*Stenotrophomonas sp.*	100	100
	T5F1	Termite-mound	Lignin + Glucose	Oxic	White/cream	Round, regular	*Enterobacter hormaechei*	100	100
	T5F2	Termite-mound	Lignin + Glucose	Oxic	White/cream	Round, regular	-	-	-

Table 5 *- Bacterial isolates obtained from elephant faeces (1 and 2) sample enrichment cultures from week 1 (T1), week 2 (T2), week 3 (T3) and week 4 (T4) time-points, including culture conditions and observed bacterial colony morphology. Isolates highlighted in red were selected for further study based on similarity to other isolates, as discussed below.*

			Replicate Culture Conditions		Colony Morphology				
Time-Point	**Isolate**	**Source**	**Carbon Source(s)**	**Oxygen Availability**	**Colour**	**Shape/Consistency**	**Identity of Closest Relative (Top BLAST Hit)**	**Sequence Identity [%]**	**Query Cover [%]**
T1	T1A1	Elephant 1	Lignin + Glucose	Anoxic	White/clear	Irregular	*Bacillus coagulans*	100	99
	T1A2	Elephant 1	Lignin + Glucose	Anoxic	White/clear	Irregular	*Bacillus coagulans*	99	100
	T1B1	Elephant 1	Lignin	Oxic	Yellow	Round/regular	*Pseudomonas stutzeri*	99	100
	T1B2	Elephant 1	Lignin	Oxic	Orange	Irregular	*Pseudomonas stutzeri*	100	100
	T1C1	Elephant 1	Lignin + Glucose	Oxic	Yellow	Round Regular	*Micrococcus sp.*	99	100
T2	T2A1	Elephant 1	Lignin	Oxic	Yellow/orange	Round/regular	*Pseudomonas stutzeri*	100	100
	T2A2	Elephant 1	Lignin	Oxic	Yellow	Round/regular	*Pseudomonas stutzeri*	99	100
T3	T3A1	Elephant 1	Lignin + Glucose	Anoxic	White/clear	Irregular	*Bacillus coagulans*	100	100
	T3A2	Elephant 1	Lignin + Glucose	Anoxic	White/clear	Irregular	-	-	-
	T3B1	Elephant 2	Lignin	Anoxic	Cream	Irregular	*Brevundimonas sp.*	99	100
	T3B2	Elephant 2	Lignin	Anoxic	Cream	Irregular	-	-	-
	T3C1	Elephant 1	Lignin	Anoxic	Clear	Regular	-	-	-
T4	T4A1	Elephant 1	Lignin	Oxic	Yellow	Round/regular	*Paenibacillus xylanilyticus*	100	100
	T4A2	Elephant 1	Lignin	Oxic	Yellow	Round/regular	-	-	-
	T4B1	Elephant 1	Lignin + Glucose	Oxic	Cream	Irregular (slimy)	*Paenibacillus xylanilyticus*	99	100
	T4B2	Elephant 1	Lignin + Glucose	Oxic	Cream	Irregular (slimy)	*Paenibacillus xylanilyticus*	99	100
	T4C1	Elephant 1	Lignin	Anoxic	Clear	Regular	-	-	-

Table 6 *- Bacterial isolates obtained from compost (C54, C55, C61) sample enrichment cultures from week 1 (T1) time-point, including culture conditions and observed bacterial colony morphology. Isolates highlighted in red were selected for further study based on similarity to other isolates, as discussed below.*

			Replicate Culture Conditions		Colony Morphology				
Time-Point	**Isolate**	**Source**	**Carbon Source(s)**	**Oxygen Availability**	**Colour**	**Shape/Consistency**	**Identity of Closest Relative (Top BLAST Hit)**	**Sequence Identity [%]**	**query cover [%]**
T1	C1	Compost C61	Lignin	Anoxic	Pink/beige	Irregular/spreading	Lysinibacillus sp.	100	100
	C2	Compost C61	Lignin + Glucose	Anoxic	White/cream	Regular/spreading	Paenibacillus sp.	100	99

From **Tables 4-6** shown above, generally, it can be seen that a large number of bacterial isolates were obtained from the environmental samples (forty-nine isolates in total). Thirty isolates were obtained from the Olduvai Gorge soils (eight from savanna soil, seven from zebra-grazed soil and fifteen from termite-mound soil), seventeen isolates from elephant faeces (fifteen from elephant faeces 1, two from elephant faeces 2) and two isolates from the compost samples (both from compost sample C61). Of the forty-nine total isolates obtained, twenty were obtained from lignin+glucose cultures, whereas twenty-nine were obtained from cultures containing lignin as sole carbon source. All isolates obtained from the Olduvai Gorge soils were aerobic microorganisms (as no anoxic cultures were prepared for these enrichments), however, eight bacterial isolates were obtained from anoxic elephant faeces cultures and both compost isolates were obtained from anoxic cultures. Whilst these isolates were obtained from anoxic culturing conditions, tests were not carried out to determine whether their anaerobic nature was facultative or not. From the wide variety of colonies present on the enrichment spread plates, an effort was made to select as diverse a range as possible to obtain different microorganisms from these samples, hence the diversity observed between colony morphologies within the isolate catalogue. As the colony morphologies alone were not evidence enough to determine similarity/dissimilarity in identity, further analysis was carried out to confirm this (involving SSU rRNA gene sequencing and phylogenetic analysis). **Tables 4-6** above also show the identities of closest relatives, sequence identity and query cover obtained from searching the bacterial 16S rRNA gene sequences against the NCBI basic local alignment search tool (BLAST) database (https://blast.ncbi.nlm.nih.gov), which is discussed below.

Whilst twenty of the isolates were obtained from cultures containing glucose as an additional carbon source, all isolates (following purification) were streaked onto plates containing MSM medium (plus trace element and vitamin solutions), agar and alkali lignin to determine their ability to grow with lignin as sole carbon source. Furthermore, all isolates were inoculated into liquid cultures containing minimal medium and lignin, and spread onto MSM-lignin-agar plates to confirm this. From these tests, it was determined that all isolates were capable of growth with lignin only and were confirmed lignin degraders, although it cannot be excluded that an alternative component of the minimal growth medium could have been used instead to support growth. Examples of isolate images are shown below in **Figure 10**.

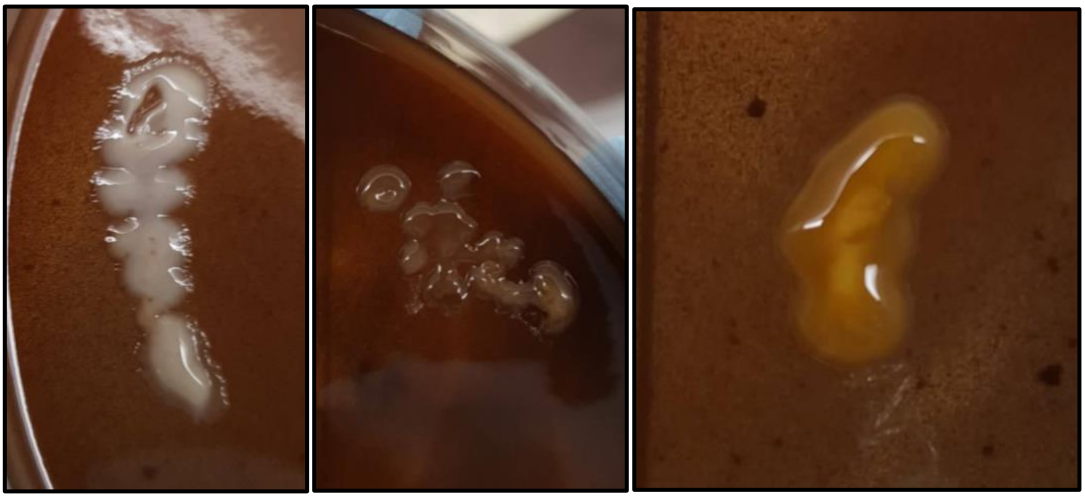

Figure 10 *– Examples of images of various isolates growing on lignin as sole carbon source.*

3.2.2 Isolate Identification and Functional Screening

Bacterial 16S rRNA gene colony PCR followed by gel electrophoresis was carried out with all isolates to obtain products for Sanger sequencing (Source Bioscience). This was done to identify the bacterial isolates and their phylogenetic placement, and to identify isolates of the same species. From this analysis, the catalogue of forty-nine isolates could be narrowed down to specific isolates representing the same species, which would be carried forward for further study. Bacterial 16S rRNA gene sequencing data was analysed in MEGA7 sequence editing software and similar sequences of closest neighbours (with accession numbers) obtained from Silva rRNA database (https://www.arb-silva.de) were used to generate a neighbour-joining phylogenetic tree using MEGA7 software, with 500 bootstrap replicates (**Figure 11** – isolates selected for further study highlighted in red). It should be noted here that identification via microbial SSU rRNA gene sequencing provides only a tentative identification, and whole genome sequencing is required in order to confirm accurate placement of these isolates and to give full insight into their metabolic potential. Exploration of isolate metabolic potential would include determination of central metabolic pathways that possibly channel intermediates into the tricarboxylic acid (TCA) cycle, alongside elucidating alternative primary lignin-degrading enzymes.

Following on from phylogenetic analysis of the isolates and identification of isolates of the same species, functional gene PCR screening was carried out with select isolates (representatives of the same species) in order to determine the presence of genes required for expression of the lignin-active enzymes laccase (*lmco*) and dye-decolorising peroxidase (DyP) (*dyp*) (important enzymes involved in bacterial lignin degradation, as discussed in the **Introduction Chapter** of this thesis). Gene presence, and therefore

functional potential, was determined by successful amplification of the functional genes from each isolate. An example gel image of *dyp* functional gene products obtained from isolates can be found in **Figure 12** (in **Section 3.2.2.3** of this chapter) and the results obtained from all isolates can be found in **Figure 11** below (*lmco* and *dyp* functional gene presence symbolised by yellow and blue dots, respectively).

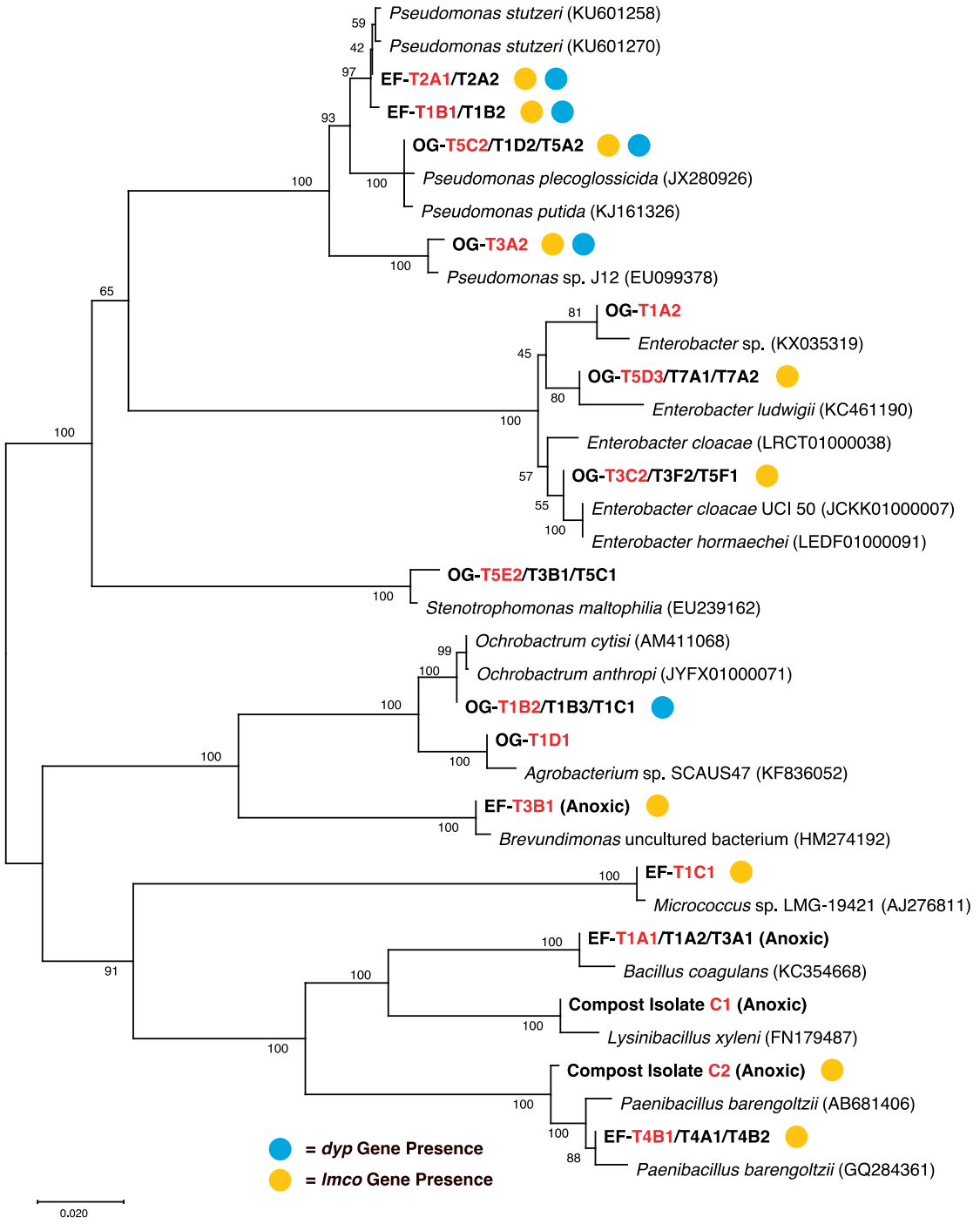

Figure 11 *- Olduvai Gorge, elephant faeces and compost isolate bacterial 16S rRNA gene phylogenetic neighbour-joining tree with neighbouring sequences obtained from SILVA - Tree generated using MEGAX with bootstrap analysis of 500 replicates. All microorganisms isolated in this project in bold, isolates currently used for further analysis highlighted in red. Orange circles indicate lmco gene presence and blue dots indicate dyp gene presence.*

The results obtained from phylogenetic analysis of the isolates 16S rRNA genes (as well as the results obtained from BLAST) revealed that the isolates obtained from all sample sets fall into the genera of *Pseudomonas* (Gammaproteobacteria), *Enterobacter* (Gammaproteobacteria), *Ochrobactrum/Agrobacterium* (Alphaproteobacteria), *Stenotrophomonas* (Gammaproteobacteria), *Bacillus* (Firmicutes), *Micrococcus* (Actinobacteria), *Paenibacillus* (Firmicutes), *Burkholderia* (Betaproteobacteria), *Brevundimonas* (Alphaproteobacteria), *Lysinibacillus* (Bacilli) and *Pseudescherichia* (Gammaproteobacteria). Phylogenetic similarity was judged based on the similarity/difference of the isolate sequences when aligned, and their proximity to one another within the phylogenetic tree (taking into account bootstrap values). Using this judgement, several of the isolates were found to likely be different/similar to one another. However, it is also possible that some sequences obtained were from different 16S operons from the same species, which was not investigated. This information would have provided more accurate information on isolate similarity.

From the phylogenetic tree in **Figure 11**, it can be seen that *Pseudomonas* spp. were isolated from both the OG and EF sample sets. The EF *Pseudomonas* spp., however, were phylogenetically similar compared to the OG *Pseudomonas* spp. which were phylogenetically different from one another and from the EF *Pseudomonas* species. *Enterobacter* spp. and *Ochrobactrum*/*Agrobacterium* spp. were isolated from the OG samples only and were phylogenetically clustered together, however each unique species was phylogenetically different from one another. A species of *Stenotrophomonas* was also isolated from the OG sample set, which was phylogenetically different from any other isolate obtained, much like the *Brevundimonas*, *Micrococcus* and *Bacillus* isolates obtained from the EF sample sets. Two phylogenetically different species of *Paenibacillus* were isolated from the CS and EF samples. In addition to the *Paenibacillus* sp., a species of *Lysinibacillus* was also obtained from the CS samples.

From the initial catalogue of forty-nine isolates, many were found to be identical species (confirmed by 100% identity over ~1300 nucleotides of the isolate bacterial 16S rRNA genes), for example OG isolates T5C2, T1D2 and T5A2 all being the same *Pseudomonas* sp. Therefore a single isolate representing each species obtained was selected for further analysis (highlighted in red in **Figure 11**), narrowing the isolate catalogue down to sixteen unique isolates.

3.2.2.1 Isolate Selection for Further Analysis

The isolates detailed below in **Tables 7, 8 and 9** were those selected for further study from the Olduvai Gorge, elephant faeces and compost samples, respectively. The percentage identity and query cover of the isolate 16S genes compared to neighbouring sequences obtained via BLAST are also shown in these tables, with each showing both identity and query cover percentages greater that 99%. The oxygen dependency of these isolates was not tested exhaustively and, whilst isolates grew in the presence or absence of oxygen, some isolates may have been facultative anaerobes. Furthermore, the anaerobic isolates obtained from elephant faeces and compost samples (shown in **Tables 8 and 9**) could also have been facultative anaerobes.

Table 7 *- Olduvai Gorge Isolates – Isolate ID, 16S rRNA gene top blast hit (including sequence identity and query cover), enrichment source and oxygen dependency. Isolates from zebra-grazed soil are highlighted in green, termite mound soil highlighted in orange and savanna soil highlighted in blue.*

Isolate ID	Source	Oxygen Dependency	16S rRNA Gene Top BLAST Hit	Identity [%]	Query Cover [%]	# of Identical Isolates
OG-T1A2	Zebra-Grazed Soil	Aerobic	*Enterobacter sp./Klebsiella sp.*	100	100	0
OG-T1B2	Termite Mound Soil	Aerobic	*Ochrobactrum sp.*	100	100	2
OG-T1D1	Savanna Soil	Aerobic	*Ochrobactrum sp.*	99	100	0
OG-T3A2	Savanna Soil	Aerobic	*Pseudomonas sp.*	99	100	0
OG-T3C2	Termite Mound Soil	Aerobic	*Enterobacter sp./Klebsiella sp.*	100	100	2
OG-T5C2	Zebra-Grazed Soil	Aerobic	*Pseudomonas sp.*	100	100	2
OG-T5D3	Termite Mound Soil	Aerobic	*Pantoea sp./Enterobacter sp.*	100	100	2
OG-T5E2	Termite Mound Soil	Aerobic	*Stenotrophomonas sp./Pseudomonas sp.*	100	100	2

Table 8 *- Elephant Faeces Isolates – Isolate ID, 16S rRNA gene top blast hit (including sequence identity and query cover), enrichment source and oxygen dependency. Isolates from elephant 1 faeces sample are highlighted in yellow and elephant 2 faeces highlighted in grey.*

Isolate ID	Source	Oxygen Dependency	16S rRNA Gene Top BLAST Hit	Identity [%]	Query Cover [%]	# of Identical Isolates
EF-T1A1	Elephant 1	Anaerobic	*Bacillus coagulans*	100	99	2
EF-T1B1	Elephant 1	Aerobic	*Pseudomonas stutzeri*	100	100	1
EF-T2A1	Elephant 1	Aerobic	*Pseudomonas stutzeri*	100	100	1
EF-T1C1	Elephant 1	Aerobic	*Micrococcus sp.*	99	100	0
EF-T3B1	Elephant 2	Anaerobic	*Brevundimonas sp.*	100	100	0
EF-T4B1	Elephant 1	Aerobic	*Paenibacillus barengoltzii*	99	100	2

Table 9 *- Compost Isolates – Isolate ID, 16S rRNA gene top blast hit (including sequence identity and query cover), enrichment source and oxygen dependency. Isolates derived from compost sample C61 (highlighted in aquamarine blue).*

Isolate ID	Source	Oxygen Dependency	16S rRNA Gene Top BLAST Hit	Identity [%]	Query Cover [%]	# of Identical Isolates
C1	Compost C61	Anaerobic	*Lysinibacillus sp.*	100	100	0
C2	Compost C61	Anaerobic	*Paenibacillus sp.*	100	100	0

3.2.2.2 Presence/Absence of Isolates Within Their Respective Environments

Species of the *Pseudomonas* and *Ochrobactrum* genera, like those obtained in this study, have been isolated from or identified within various soil samples in previous studies exploring lignin/lignocellulose degradation (Tian *et al*., 2016; Jimenez *et al*., 2014; Taylor *et al*., 2014; Bugg *et al*., 2016). The same can be said for members of the *Enterobacter* and *Stenotrophomonas* genera, similar to those isolated in this study (Manter *et al*., 2011; DeAngelis *et al*., 2011; Tian *et al*., 2016). Furthermore, strains of *Ochrobactrum* (OG-T1B2) and *Enterobacter* (OG-T3C2) obtained in this study were isolated from termite mound soil, species of which have previously been identified within samples derived from termite gut (Bugg *et al*., 2011; Wenzel *et al*., 2002; Zhou *et al*., 2017; Chew *et al*., 2018).

From the isolates obtained from the EF samples, two were identified as being of the *Pseudomonas* genus. A study carried out by Jakeer *et al* (2020) found that, through

metagenomic analysis of elephant faeces, *Pseudomonas*, *Brevundimonas*, *Bacillus* were prevalent genera within the elephant gut microbiome. Furthermore, other metagenomic studies of the elephant gut microbiome have revealed the prevalence of *Pseudomonas* as well as presence of *Bacillus* and *Paenibacillus* within these environments (Zhang *et* al., 2019). The evidence of *Pseudomonas* predominance within the elephant gut was further supported by a metagenome study carried out by Ilmberger *et al* (2014). The *Paenibacillceae* family has additionally been identified within the fecal and gut microbiomes of herbivorous animals such as the yak (Gong *et al*., 2020) and panda (Zhan *et al*., 2020), respectively. So far, to the best of the authors knowledge, species of *Micrococcus* have not yet been identified within the elephant gut/fecal microbiomes.

Of the compost isolates obtained within this study, several studies have identified *Paenibacillus* (Ueda and Kurosawa, 2015; Vaz-Moreira *et al*., 2007) and *Lysinibacillus* (Wang *et al*., 2016; Hayat *et al*., 2014) to be present within compost environments previously.

3.2.2.3 Isolate Functional Potential to Produce Lignin-Active Enzymes

An example gel image of *dyp* functional gene PCR amplification products can be seen in **Figure 12**. Functional gene screening results (indicated by yellow (*lmco* gene) and blue (*dyp* gene) dots in **Figure 11**) show that all *Pseudomonas* spp. obtained (OG-T3A2, OG-T5C2, EF-T1B1 and EF-T2A1) have the genetic potential to express both laccase and DyP, whereas species of *Enterobacter* (OG-T3C2 and OG-T5D3), *Brevundimonas* (EF-T3B1), *Micrococcus* (EF-T1C1) and *Paenibacillus* (EF-T4B1 and C2) have only the laccase gene. The OG isolate *Ochrobactrum* sp. (OG-T1B2), conversely, has only the gene required for DyP expression. From these results, it was also observed that isolates OG-T5E2 (*Stenotrophomonas* sp.), EF-T1A1 (*Bacillus* sp.) and C1 (*Lysinibacillus* sp.) lacked the presence of both genes. The presence of these genes indicates the genetic potential of these microorganisms to produce laccase and DyP-type peroxidases, however do not suggest that they are capable of expressing them. Additionally, the lack of these genes within some isolates does not indicate a lack of lignin-degrading potential, as lignin can be broken down by a variety of alternative mechanisms. Whilst several of the isolates in this study showed no evidence for the presence of genes responsible for laccase or DyP-peroxidase, they were cultured with lignin as sole carbon source and were therefore capable of using this substrate, suggesting the use of alternative mechanisms.

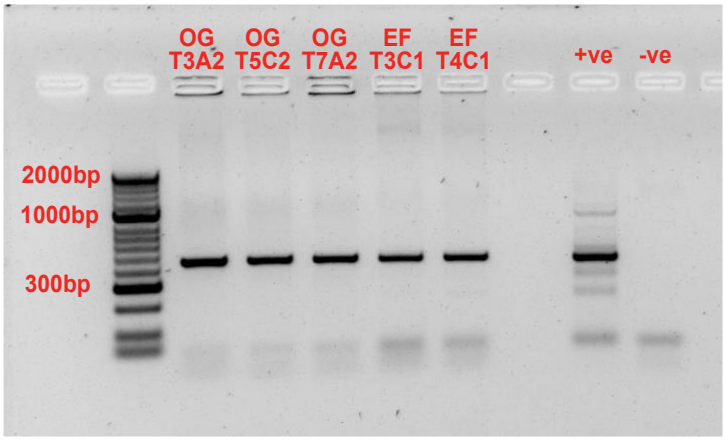

Figure 12 *– Example agarose gel electrophoresis image of dyp functional gene colony PCR amplification products obtained from Olduvai Gorge and Elephant Faeces isolates OG T3A2, OG T5C2, OG T7A2, EF T3C1 and EF T4C1. Positive control = previous dyp PCR product from EF T4C1; negative control = water. DNA ladder = 50 bp Hyperladder. Expected product size: 300-400 bp.*

All microorganisms isolated in this study are from genera which contain species previously shown to actively degrade lignin (Chandra *et al*., 2007; Tian *et al*., 2014; Tian *et al*., 2016b; Bugg *et al*., 2016). Several species of *Pseudomonas* have been shown to express a number of different lignin-active oxidative enzymes, such as the well-known lignin degrader *Pseudomonas putida*, capable of expressing laccase and DyP-B peroxidases, supporting the functional gene screening results of this study (Xu *et* al., 2018; Tian *et al*., 2014).

Ochrobactrum sp., previously cultivated from termite gut, has previously been shown to break down lignin-derived compounds (Tian *et al*., 2016). Xu *et al* (2018) also showed, through spectrophotometric enzyme assays (similar to those used in this project), that *Ochrobactrum tritici* was capable of producing low levels of laccase and peroxidase (lignin peroxidase and manganese-dependent peroxidase), however, to the best of the authors knowledge, no evidence has yet been found for *Ochrobactrum* sp. to have the gene coding for DyP peroxidase.

Several *Enterobacter* species have been shown to be potent lignin degraders, more prevalent examples being *Enterobacter lignolyticus* and *Enterobacter soli* (Manter *et al*., 2011; Bugg *et al*., 2016). Mongkolthanaruk *et al* (2012) has also shown the capability of an *Enterobacter* sp. to express laccase in the presence of guaiacol (a phenolic component of lignin). Furthermore, *Enterobacter lignolyticus* has also been shown to degrade lignin in anoxic conditions (Manter *et al*., 2011) and through the production of peroxidase enzymes, such as the DyP (type B) peroxidases (DeAngelis *et al*., 2013).

Strains of *Stenotrophomonas maltophilia* (similar to isolates T3B1, T5C1 and T5E2) have also been shown to degrade lignin (Tian *et al*., 2014; Tian *et al*., 2016; Lambertz *et al*., 2016). In addition, *Stenotrophomonas* species have been shown to metabolise lignin-like dyes through the production of high-redox potential enzymes such as lignin peroxidase (Tian *et al*., 2016).

Species of *Bacillus* (like isolate EF-T1A1) have previously been found to be responsible for production of pectin lyase in the elephant gut microbiome (Jakeer *et al*., 2020) and have been shown to degrade lignin (Chandra *et al*., 2007). Furthermore, several species of *Bacillus*, such as *B. licheniformis*, *B. pumilus* and *B. subtilis*, have been shown to produce/encode several lignin-active enzymes, such as laccase and DyP-type peroxidase (Bugg, Rahmanpour and Rashid, 2016; Datta *et al*., 2017; Colpa, Fraaije and Van Bloois, 2014).

Species of the *Micrococcus* genus (like EF-T1C1) have been shown to break down Kraft lignin (Bugg *et al*., 2016) and have been shown to produce laccase (Joseph *et al*., 2014), similar to the results observed in this study. However, *Micrococcus* has not previously been described to be active in lignin degradation within the elephant, to the best of the authors knowledge.

Paenibacillus has been shown to participate in cellulose degradation within the panda gut (Zhan *et al*., 2020) and has been shown to be capable of lignin degradation (Chandra *et al*., 2007). A study carried out by Eastman *et al* (2014), involving genomic sequencing of *Paenibacillus polymyxa*, found that this species contained genes required for both laccase and DyP-type peroxidase expression, however only the laccase gene was observed for the *Paenibacillus* isolate in this study.

Our study also displayed that the *Brevundimonas* sp. isolated from EF samples harboured the gene responsible for laccase production. These results are reflected by those obtained by the study carried out by Poomima and Velan (2018) which demonstrated the ability of a *Brevundimonas* sp. to produce laccase.

Lysinibacillus sp., isolated from compost in this study, has previously been shown to produce laccase and peroxidase enzymes, (Chantarasiri and Boontanom, 2017) and more specifically have been shown to produce DyP-type peroxidases (Granja-Travez *et al*., 2020).

From the results obtained in this project and the information provided above, the OG *Ochrobactrum* sp. isolate harbouring the gene responsible for production of DyP peroxidase and the EF *Micrococcus* sp. isolate being a lignin degrader within the EF environment are, to the best of the authors knowledge, novel information.

3.2.3 Determination of Isolate Optimal Growth Temperatures

In order to determine the optimum growth conditions of isolates and, therefore, the conditions to be used in further experiments, growth experiments were carried out in cultures composed of YEMS medium with lignin as sole carbon source. These growth experiments were, however, being carried out during and after further experimentation (GC-MS analysis and enzyme assays), therefore the optimal growth conditions of isolates were not determined until some experiments had been completed. As stated in **Methods and Materials Section 2.3.3**, lignin was added at 0.05 g/L concentration, as opposed to the higher 10 g/L concentration added to the enrichment cultures, to avoid the effect of lignin interfering with spectrophotometer absorbance readings. The OG isolate growth experiment was carried out with cultures incubated at 25 ºC and 37 ºC for 3 days, whereas the EF isolate growth experiment was carried out with cultures incubated at 25 ºC, 37 ºC and 45 ºC for 5 days. These growth conditions were selected for testing as these were the conditions that were planned to be used in the enzyme assay and GC-MS experiments. Lignin was included in these growth experiments (at similar concentrations to those planned to be used in further experimentation) to determine growth with this substrate present and, therefore, provide growth patterns to be expected in future experiments. Cultures grown at the higher temperature of 45 ºC were included for the EF isolates as some of these microorganisms may have been better adapted to higher temperatures (potential extremophiles), however the core temperature of elephants has previously been found to be approximately 37 ºC (Weissenbock *et al*., 2010). Isolate EF-T4B1 was not included in the EF isolate growth experiment as the starter culture did not achieve a suitable cell density for use. Growth experiments were not carried out with compost isolates due to lack of time. The results of the OG and EF isolate growth experiments can be found in **Figure 13** and **Figure 14**, respectively.

The results of the OG isolate growth experiment (**Figure 13**) show that all isolates except OG-T3A2 (*Pseudomonas* sp.) grew optimally at 25 ºC as opposed to 37 ºC, indicated by the higher yield achieved at 25 ºC. At 25 ºC, most isolates achieved OD_{600} values between 2-3, whereas OG-T5E2 achieved a yield of OD_{600} 1.5. The results observed for the uninoculated negative controls show that no growth occurred within these tests,

confirming no contamination within this experiment. These results indicated that the incubation temperature of 25 ºC was to be used for these isolates in the enzyme assay experiments, therefore ensuring higher enzyme activity measurements due to higher proliferation.

The results of the EF isolate growth experiment (**Figure 14**) show that all isolates have a preference for growth at 25 ºC except EF-T1C1 (*Micrococcus* sp.), which showed no preference between the two incubation temperatures. In the 25 ºC cultures, EF isolates grew to peak cell density and remained at that density for the duration of the experiment. In the 37 ºC cultures, EF isolates appeared to peak in cell density, after which density began to decline. Almost all 45 ºC cultures did not grow at all, except EF-T1A1, which achieved peak OD_{600} of 0.2. At 25 ºC, isolates EF-T1A1 and EF-T3B1 both reached peak cell density (OD_{600}) of 2.8 and 3, respectively, whereas EF-T1B1, EF-T1C1 and EF-T2A1 achieved lower yield with OD_{600} values around 1.75. These results confirmed that all EF isolates had growth optima at 25 ºC.

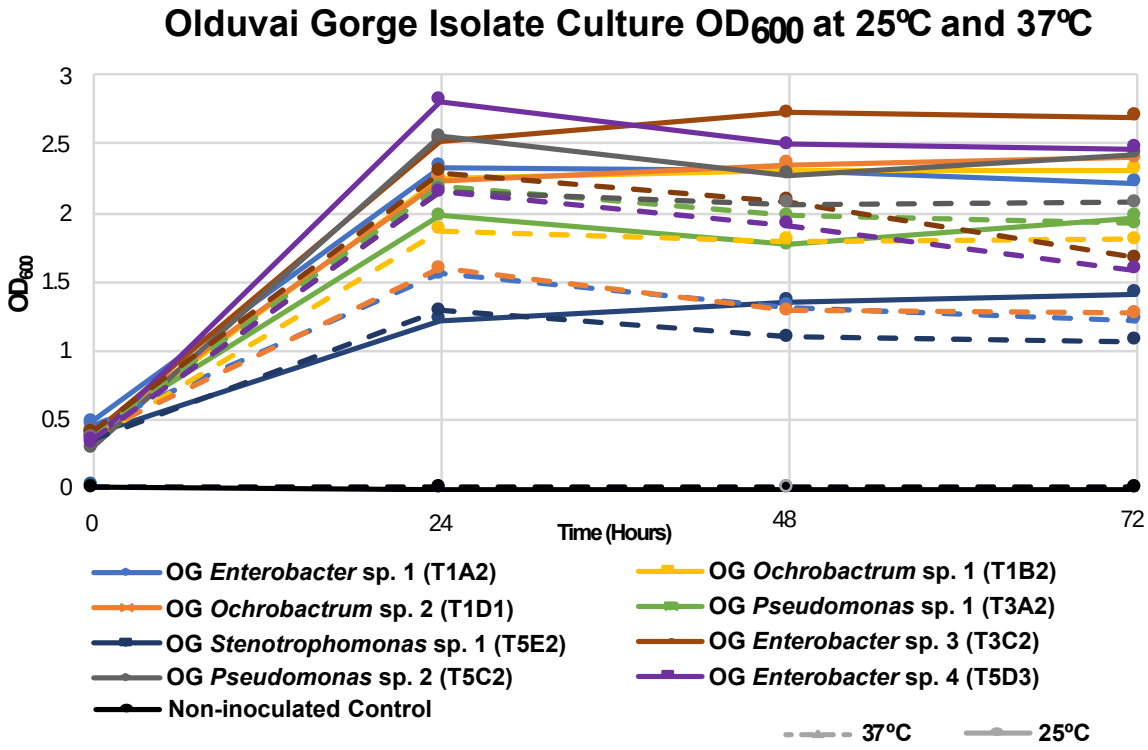

Figure 13 *- Olduvai Gorge isolate growth (determined by OD_{600} measurements) at 25 ºC and 37 ºC in YEMS medium with lignin as additional carbon source. 25 ºC cultures represented by solid lines, 37 ºC cultures represented by dashed lines.*

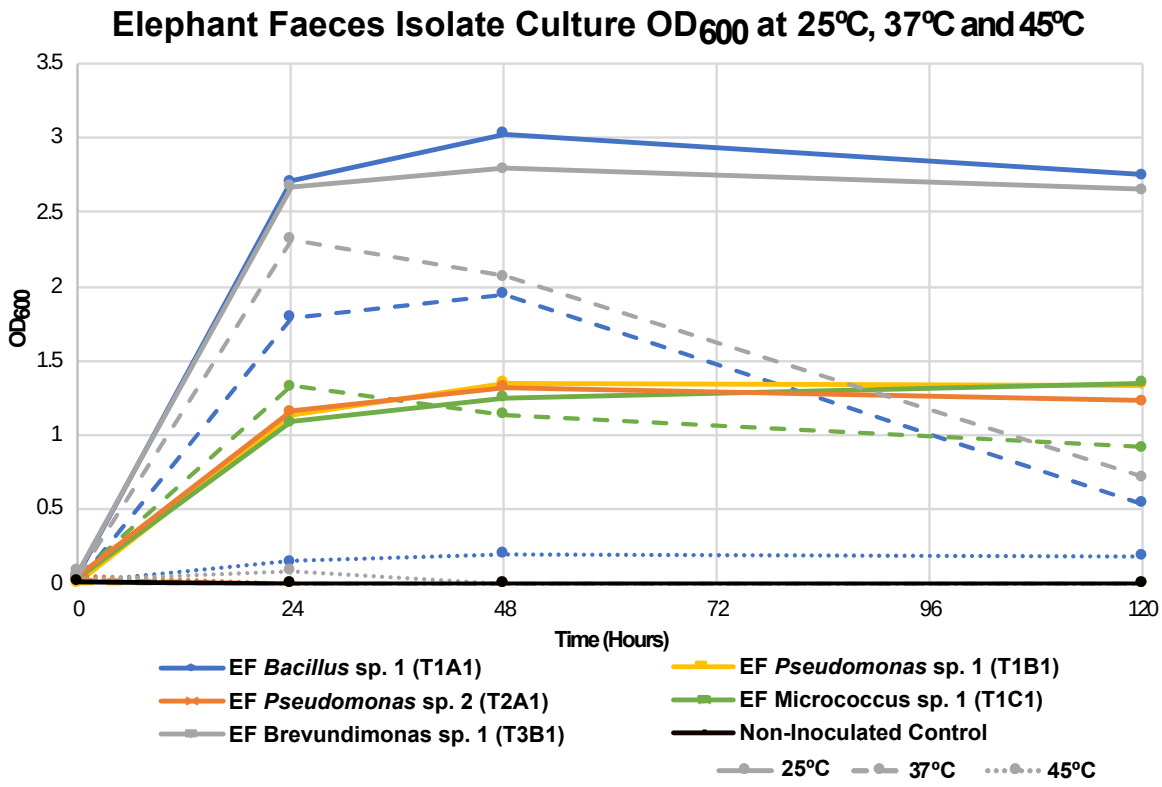

Figure 14 *– Elephant faeces isolate growth (determined by OD_{600} measurements) at 25 ºC, 37 ºC and 45 ºC in YEMS medium with lignin. 25 ºC cultures represented by solid lines, 37 ºC cultures represented by dashed lines and 45 ºC cultures represented by dotted lines. Only isolate EF-T1A1 grew within the 45 ºC cultures. Some data points overlapped with the negative control (OD_{600} = 0) and are therefore not visible in the figure.*

3.2.4 Fungal Isolates Obtained from Olduvai Gorge Enrichments

In addition to the bacterial isolates described above, several fungal isolates obtained from the Olduvai Gorge enrichment cultures (lignin and lignin/glucose replicates) were selected for initial characterisation. Whilst seven fungal isolates were obtained from the OG samples (F1-7), only two amplification products were successfully obtained via eukaryotic 18S rRNA gene colony PCR (F2, derived from zebra-grazed soil, and F5, derived from termite mound soil). Similar to the bacterial 16S rRNA gene amplification products, these isolates underwent phylogenetic analysis via Sanger sequencing of amplification products and sequence analysis with MEGAX sequence editing software. Similar sequences of closest neighbours obtained from Blastn searches against the NCBI GenBank nucleotide database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) were used to construct the phylogenetic tree of fungal isolates shown in **Figure 15** below.

The results of phylogenetic analysis of the two fungal isolates show that the eukaryotic 18S sequence of F2 most closely aligned with *Trichoderma koningii* whereas F5 most closely aligned with a species of *Penicillium*. *Trichoderma* spp. and *Penicillium* spp. both

encompass species which are capable of degrading lignin. Bohacz and Korniłłowicz-Kowalska (2020), via spectrophotometric enzyme assays and HPLC, determined that several *Trichoderma* species isolated from compost were able to break down post-industrial lignin into monolignols such as vanillic acid and syringic acid through the use of peroxidase enzymes. In addition, *Penicillium simplicissimum* has been shown to degrade natural lignin and utilise lignin model compounds as growth substrate via lignin peroxidase and laccase (Yu *et al.,* 2005), as well as alkali lignin via lignin peroxidase and manganese peroxidase (Liu *et al*, 2014). Due to time limitations and the focus of this study being primarily on bacterial candidates (due to their industrial potential), the fungal isolates obtained from the OG sample sets were not analysed further.

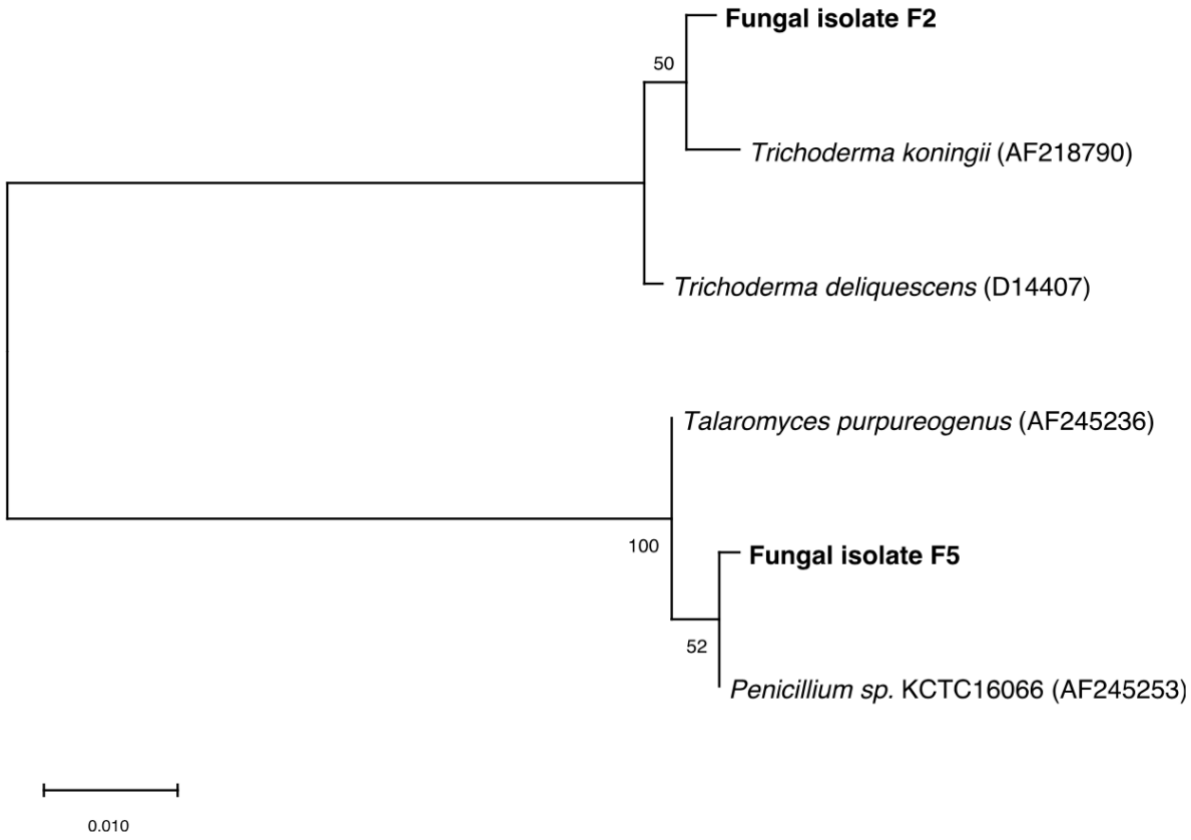

Figure 15 - Olduvai Gorge fungal isolates F2 (culture T1 ZS-MSM-L) and F5 (culture T13 TM-M9-L+G) eukaryotic 18S rRNA gene phylogenetic neighbor-joining tree with neighboring sequences obtained from SILVA (using the alignment, classification and tree service) - Tree generated using MEGAX with bootstrap analysis of 500 replicates. Microorganisms isolated in this project in bold.

3.2.5 Determination of Isolate Oxidative Enzyme Production

Olduvai Gorge (OG) and Elephant Faeces (EF) isolates (specified in **Tables 7 and 8** above) underwent six enzyme assay experiments with industrial partner Recircle Ltd., throughout which the enzyme assay methodologies were altered/improved (exact protocols can be found in **Methods and Materials Section 2.4.1**). Presented here are

the results of the final two experiments which were described as “rounds” (i.e. the final two experiments were described as “round-5” and “round-6”) of experimentation. In both experiments, cell-free supernatant from the isolate-lignin cultures was tested with an ABTS assay (Murugesan *et al*., 2007) and 2,4-DCP assay (Antonopoulos *et al*., 2001). A 2,6-DMP assay was also used (Salvachua *et al*., 2015), however, due to lack of positive control data, no viable conclusions could be made from the results of this assay. The ABTS assay was used to measure laccase activity and the 2,4-DCP assay was used to measure general peroxidase activity. The 2,6-DMP assay, which involved the use of manganese as a co-factor in the assay, would have provided information on manganese-dependent peroxidase activity of the isolates if results had been successfully obtained. The 2,4-DCP assay, however, provides information only on general peroxidase activity.

Isolates were cultured in replicates containing 50 ml YEMS medium with lignin (0.05 g/L conc.) and MSM medium with lignin and glucose (0.05 g/L and 10 g/L, respectively). As mentioned in **Methods and Materials Section 2.3.1**, YEMS medium is nutrient-rich whereas MSM is a minimal-medium. YEMS was used in this experiment to maximise cell proliferation and therefore maximise enzyme production, whereas MSM was used to determine enzyme production in more restricted conditions. The anaerobic isolates EF-T1A1 (*Bacillus* sp.) and EF-T3B1 (*Brevundimonas* sp.) were cultured and underwent enzyme analysis in anoxic conditions. A schematic overview of the enzyme assay work can be found in **Figure 16**.

The oxidative enzyme assays (as well as the GC-MS analysis detailed in the following section) were carried out before the growth experiments, therefore the growth optima of the isolates was not known previous to this. In the round-5 experiment, all isolates (OG and EF) were cultured at 25 °C, whereas, in round-6, OG isolates were cultured at 25 °C and EF isolates were cultured at 37 °C. As discussed later in this section, the enzyme activity observed from replicates grown at 37 ºC (in round-6) were significantly lower than that observed from the 25 ºC replicates (in round-5), likely due to 25 ºC being the better growth temperature for all isolates. Despite this, the enzyme assay experiments were planned to be carried out with the isolates at various temperatures regardless, as enzyme activity may have been unaffected by growth temperature.

Laccase enzyme (from *Trametes versicolor*; 0.02 mg/ml, Sigma-Aldrich) was used as an enzyme positive control in the ABTS assay and peroxidase (from horseradish, HRP; 0.002 mg/ml, Sigma-Aldrich) was used in the 2,4-DCP assay. A *Rhodococcus* sp., provided by industrial partner Recircle Ltd., was used as a positive control for the 2,4-DCP assay (due to previous evidence of peroxidase production) and was also tested in

the ABTS assays. An uninoculated negative control culture was also prepared for each medium/substrate combination.

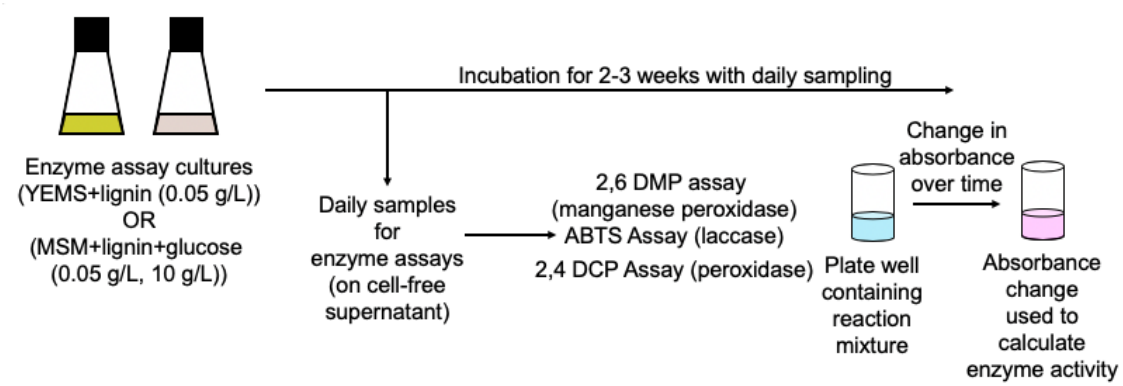

Figure 16 *– Schematic overview of oxidative enzyme assay workflow carried out in both Round 5 and Round 6. YEMS = nutrient-rich medium, MSM = minimal medium.*

In round-5 (experiment 5), all isolate assay cultures were incubated at 25 °C and enzyme assays were carried out for 10 days. Isolates OG-T3A2 (*Pseudomonas* sp. 1) and OG-T3B1 (*Stenotrophomonas* sp.) were not run in round-5 as these isolates had not been confirmed via bacterial 16S rRNA gene sequencing and EF-T4B1 (*Paenibacillus* sp.) was not run due to limited growth in starter culture. Round-6 was similar to round-5, however the OG isolate cultures were incubated at 25 °C and EF isolate cultures at 37 °C (respective to their original enrichment temperatures) and enzyme assays were carried out for 13-days.

Isolates were cultured in YEMS medium (50 ml) before starting the enzyme assays, and inoculum was washed with MSM before inoculating MSM replicate incubations at OD_{600} 0.4 concentration of biomass. All enzyme assays were carried out at room temperature in a spectrophotometric 96-well plate reader (Labtech LT-4500), and enzyme activity of isolates was calculated according to methods detailed in **Methods and Materials Section 2.4.1.2** of this thesis.

3.2.5.1 Measurement of Isolate Laccase Production

The results obtained from both round 5 and 6 of the ABTS assay are shown in **Figure 17**. In both rounds, the laccase enzyme positive control started with an enzyme activity of approximately 10 U/L, which slowly decreased towards 8 U/L throughout the durations of both experiments. For clarification, the laccase activity observed within the figure below is based on daily measurements of laccase enzyme (prepared before the experiment and stored in a fridge) which was added to reaction mixture (much like the

isolate culture supernatant). The uninoculated and water negative controls showed no activity. Results obtained from OG and EF isolates in round 5 seemed to show that several select OG isolates displayed limited levels of laccase activity after ~168h of incubation in YEMS medium. However, this could not be repeated in round 6, possibly indicating that the results were false and no true laccase activity was observed. The hypothesis that these initial results were false was further supported by no laccase activity being observed in biochemical experiments conducted in the following section.

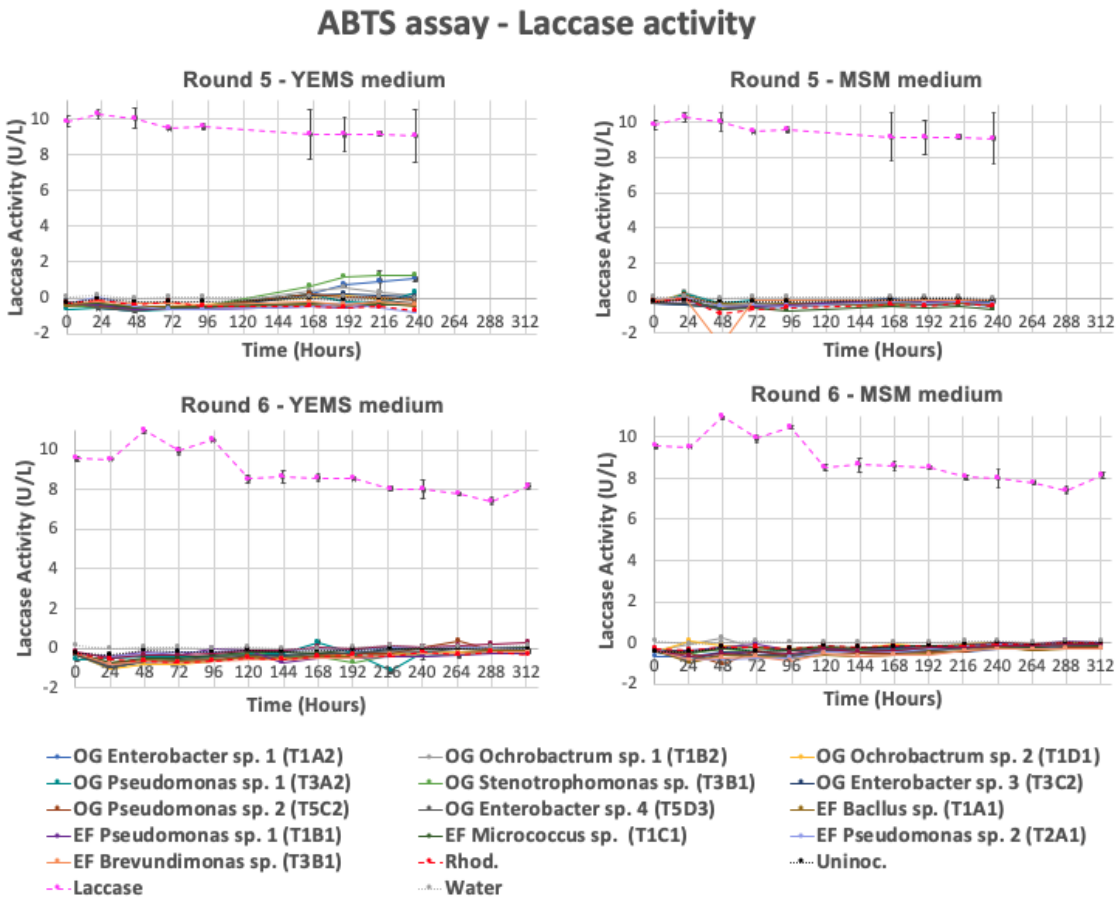

Figure 17 *- Round 5 and 6 ABTS assay results – Laccase activity from Olduvai Gorge and Elephant Faeces Isolates in YEMS and MSM media. Laccase enzyme positive control results in pink solid/dashed line and water negative control results in grey dotted line. YEMS = nutrient-rich medium, MSM = minimal medium.*

Previous studies have reported species of *Pseudomonas putida* (KT2440), *Enterobacter lignolyticus* and *Rhodococcus jostii*, like the isolates obtained in this study, to produce low levels of laccase (6 U/L, 2 U/L and 1 U/L, respectively) using the 2,6-DMP assay (Salvachua *et al*, 2015). However, it could be that these species are functionally different from the isolates used in this study. Furthermore, Xu *et al* (2018) found that a different species of *P. putida* (NX-1) and a species of *Ochrobactrum tritici* were able to produce maximum laccase activity of 100 U/L and 635.9 U/L, respectively. It is possible that Xu *et al* (2018) grew these microorganisms in more favourable conditions compared to

those used in this study and, whilst it is mentioned that Xu *et al* used 30 ºC throughout the study (with 1 g/L kraft lignin), the author does not specify the conditions used when growing the microorganisms for the enzyme assays.

However, the *Pseudomonas* and *Ochrobactrum* species isolated in this study are likely not identical to those used by Xu *et al*., thus one might expect them to behave differently. Despite previous evidence of species within the genera of the isolates presented in this study producing laccases, our isolates did not show any evidence of laccase production in all 6 experiments. As positive enzyme activity was observed for the laccase enzyme positive control, it was confirmed that the enzyme assay itself was working properly. The lack of enzyme activity observed for the isolates could be due to a number of potential factors, for example: the culturing conditions being insufficient for laccase production or a lack of requirement for laccase production due to production of other enzymes (such as peroxidases being produced, shown in **Figure 18** below). Many of these isolates were shown, however, to have the *lmco* functional gene coding for laccases, in the functional gene screening results (shown in **Figure 11**). Therefore, lack of laccase production observed in the enzyme assays could have been due to either of the reasons stated above, or some other factor which is not obvious.

3.2.5.2 Measurement of Isolate Peroxidase Production

The results obtained from the 2,4-DCP assays are displayed in **Figure 18**. In round-6, the HRP enzyme positive control showed stable enzyme activity at ~40 U/L throughout the experiment whereas, in round 5, the HRP control showed stable enzyme activity around 20 U/L. Both the uninoculated and water negative controls showed little to no activity as expected.

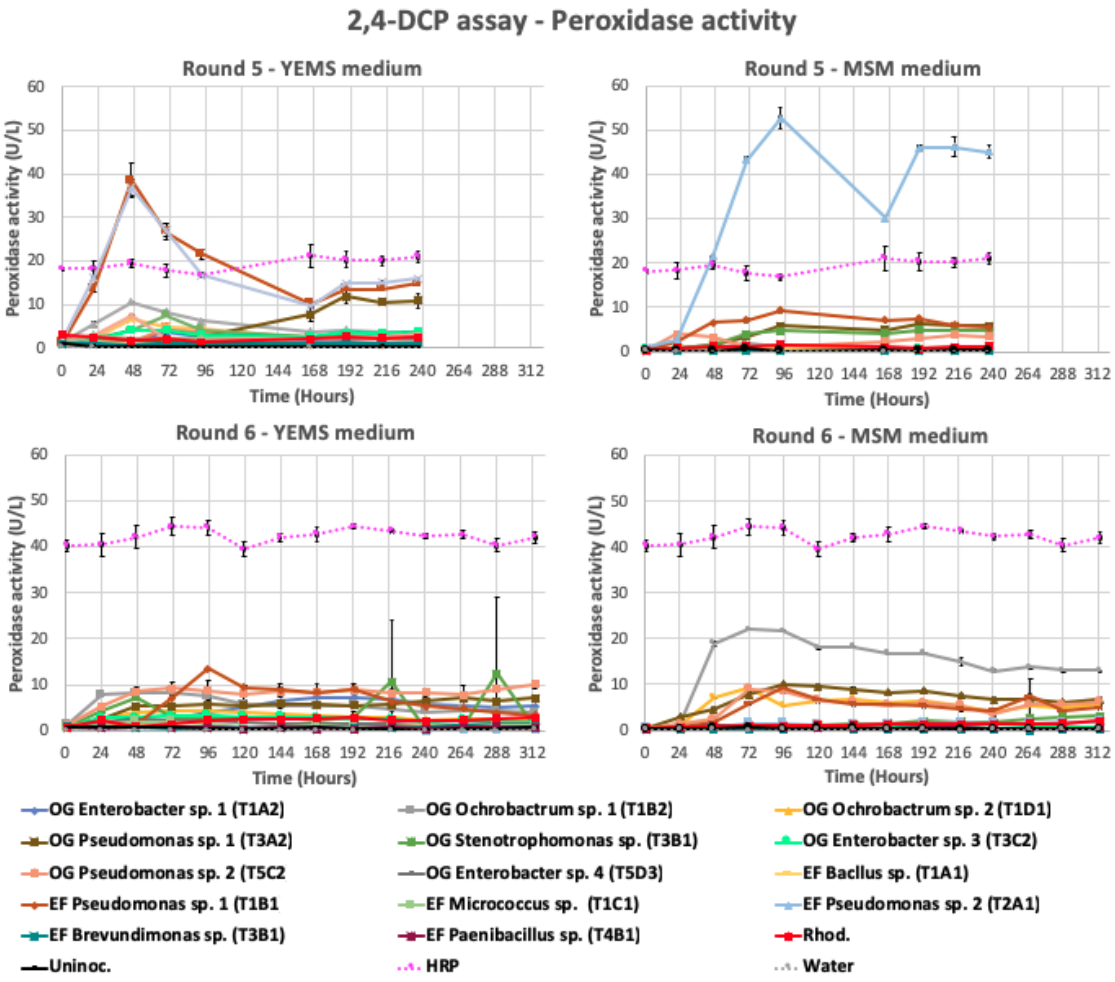

Figure 18 *- Round 5 (solid) and 6 (dashed) 2,4-DCP Assay Results – Peroxidase activity from Olduvai Gorge and Elephant Faeces Isolates in YEMS (top) and MSM (bottom) media. HRP enzyme positive control results in pink solid/dashed line and water negative control results in grey dotted line. YEMS = nutrient-rich medium, MSM = minimal medium.*

Results obtained from round 5 and 6 of the 2,4-DCP assay revealed that almost all isolates expressed signs of peroxidase activity. Furthermore, most isolates expressed similar peroxidase activity patterns, with a high initial peak within 2-3 days of incubation, followed by sustained production of enzyme and a gradual decrease in activity after time. The most active isolates in round 5 were EF-T1B1 and EF-T2A1 (both *Pseudomonas* spp.), followed by OG-T1B2 and OG-T1D1 (both *Ochrobactrum* spp.) and OG-T5C2 (*Pseudomonas* sp.). Interestingly, despite most isolates showing lower enzyme activity in MSM medium in round-5, the *Pseudomonas* isolates expressed relatively high activity in MSM medium (minimal medium), with isolate EF-T2A1 expressing higher peroxidase activity in MSM than in YEMS medium (nutrient-rich medium). The maximum enzyme activities achieved by each isolate are summarised below in **Table 10**.

As in round-5, many of the Olduvai Gorge and elephant faeces isolates expressed high levels of peroxidase activity in round-6. Most isolates followed a similar activity pattern to round-5, either expressing an initial peak in peroxidase production followed by a gradual decrease over time, or an increase to peak activity and stabilization of activity for the duration of the experiment. As shown in **Figure 18** and **Table 10**, in YEMS medium, EF-T1B1 (*Pseudomonas* sp.) was the most active, followed by OG-T5C2 (*Pseudomonas* sp.), OG-T1B2 (*Ochrobactrum* sp.), OG-T3B1 (*Stenotrophomonas* sp.) and OG-T1A2 (*Enterobacter* sp.). In MSM medium, the most active isolates were OG-T1B2 (*Ochrobactrum* sp.), followed by OG-T3A2 (*Pseudomonas* sp.), EF-T1B1 (*Pseudomonas* sp.), OG-T5C2 (*Pseudomonas* sp.) and OG-T1D1 (*Ochrobactrum* sp.). Like the high activity observed for EF-T2A1 in MSM medium in round-5, OG-T1B2 also showed higher enzyme activity in MSM medium in round-6. It could be that the high activity observed for some isolates in MSM suggests that these species prefer more minimal media. In both medium configurations, EF-T1A1 (*Bacillus* sp.) and EF-T3B1 (*Brevundimonas* sp.) were the poorest performing isolates, with peroxidase activity remaining below 1.5 U/L throughout both experiments, with the same results being observed for isolate EF-T4B1 (*Paenibacillus* sp.) in round-6. The low results observed for EF-T1A1 and EF-T3B1 can likely be explained by their anaerobic nature and therefore slower proliferation rates. Also, it is noteworthy that there were various peaks with large error bars within the OG-T3B1 YEMS replicate observed in **Figure 18**, which were data points likely impacted by bubbles being present within the reaction well of the 96-well plate, therefore having altered the absorbance readings. These bubbles were difficult to avoid due to the plate being prepared quickly to avoid loss of activity, as the reactions would occur very quickly.

Many of the isolates achieved similar peroxidase activity patterns between round-5 and 6, showing replicability, however some did not, the most notable being OG-T1A2 (*Enterobacter* sp.), OG-T5C2 (*Pseudomonas* sp.), OG-T5D3 (*Enterobacter* sp.), EF-T1B1 and EF-T2A1 (both *Pseudomonas* sp.). Despite the EF isolates being incubated at 37 ºC in round-6 compared to 25 ºC in round 5, the enzyme activity of EF-T1A1, EF-T1C1 and EF-T3B1 remained similar, however, in round 6, both EF *Pseudomonas* spp. (the most active isolates in round 5) performed poorly, with EF-T1B1 having greatly reduced enzyme activity and EF-T2A1 expressing little/no activity. Lower growth yield was observed for these cultures compared to previous enzyme assay experiments, indicating that the increased incubation temperature (25 ºC to 37 ºC) in round 6 had a negative impact on growth and enzyme activity of these isolates, similar to the results observed in the growth experiments. As explained above, the growth experiments (**Section 3.2.3**) performed with these isolates were carried out during and after the

oxidative enzyme assays, therefore it was unknown at the time that the EF isolates grew more effectively at 25 °C compared to 37 °C. Previous studies have shown the optimum growth conditions of some *Pseudomonas* spp. to be around 25-30 °C (Young *et al*., 1977), which may explain the lack of activity/growth in this experiment at 37 °C. Slightly reduced activity observed from EF-T3B1 (anoxic *Brevundimonas* sp.) and EF-T1C1 (*Micrococcus* sp.) in this round also provides evidence for these impacts of growth temperature. In support of this, the results of growth experiments carried out in this PhD project (shown above) illustrated that EF-T1B1 and EF-T2A1 grew only at 25 ºC. In these growth experiments, it was also shown that EF-T1C1 had no preference between 25 ºC and 37 ºC (explaining the limited variation in activity between round 5 and 6), but EF-T1A1 and EF-T3B1 showed a strong preference for 25 ºC. Whilst these isolates may have had a preference for 25 ºC, the activity from these isolates in the enzyme assays was so low (almost no activity) that the preference was unobservable.

The maximum peroxidase activities of OG-T3A2, OG-T5C2, EF-T1B1 and EF-T2A1 (all *Pseudomonas* spp.) (summarized in **Table 10** below) were between 9.25-52.57 U/L, making them the most active peroxidase-producing isolates in this study, closely followed by OG-T1B2 and OG-T1D1 (both *Ochrobactrum* spp.). Whilst the results of OG-T3A2 and OG-T5C2 (11.87 U/L and 9.25 U/L, respectively) are relatively similar to that of the *Pseudomonas putida* and *Pseudomonas fluorescens* strains used by Salvachua (2015) (~6 U/L and ~4.5 U/L, respectively), the EF *Pseudomonas* spp. (EF-T1B1 and EF-T2A1) were significantly higher (38.53 U/L and 52.57 U/L, respectively). Salvachua *et al* (2015) also tested a strain of *Enterobacter lignolyticus* (same genus as the *Enterobacter* spp. isolated in this study) which showed no evidence of manganese-independent peroxidase activity, however it showed manganese-dependent peroxidase activity of ~2 U/L, whereas the manganese-independent peroxidase activity demonstrated by our *Enterobacter* spp. (OG-T1A2 and OG-T3C2) ranged between 2.46-7.39 U/L. Therefore, the peroxidase used by the *Enterobacter lignolyticus* strain reported by Salvachua and colleagues requires manganese to function, whereas the peroxidase used by our *Enterobacter* sp. does not.

Whilst the enzyme activity results of this study were relatively similar (if not higher) than those achieved by Salvachua *et al* (2015), Salvachua *et al* used a different enzyme assay (using 2,6-DMP as substrate, lower H_2O_2 concentration and different buffer) as well as different microbial culturing conditions. The different cultivation and enzyme assay conditions used by other studies therefore make comparisons of our results to other datasets extremely difficult, as also stated by Salvachua *et al* (2015). These difficulties are exemplified by studies such as those carried out by Xu *et al* (2018) and Kumar *et al*

(2015). Xu *et al* (2018), using an enzyme assay with veratryl alcohol as substrate, found that wild-type strains of *Ochrobactrum tritici* and *Pseudomonas putida* expressed maximum peroxidase activity of 3500 U/L and 6500 U/L, respectively, which are significantly higher values compared to those obtained in our experiment and by Salvachua *et al* (2015). Kumar *et al* (2015) found similarly high peroxidase activity (2750 U/L) from a species of *Pandoreae* using an enzyme assay incorporating Azure B as substrate. Whilst these differences in activity are most likely related to differences in bacterial species being used and the sensitivity of assay being used (including co-factor H_2O_2 concentration), the vast array of different enzyme assay and cultivation methodologies used by other studies makes it difficult to determine the contributing factors to differences and, more importantly, difficult to determine the most capable microorganisms. Despite the difficulties in comparing the results of our study to that of similar studies, the results of the enzyme assay experiments show that all of the isolates tested were capable of producing peroxidase enzymes, with the most active isolates being the EF *Pseudomonas* species.

Table 10 *– Table of maximum enzyme activity achieved from Olduvai Gorge and elephant faeces isolates in enzyme assay round 5 and 6 experiments, including peak enzyme activity observed, time of peak activity, medium in which best enzyme activity was observed and round in which best enzyme activity was observed. YEMS = nutrient-rich medium, MSM = minimal medium. All peak enzyme activity values were achieved at 25 ºC.*

Isolate	Round	Best Medium	Peak Enzyme Production (U/L)	Time of Peak (h)
OG *Enterobacter* sp. 1 (T1A2)	6	YEMS	7.39	168
OG *Ochrobactrum* sp. 1 (T1B2)	6	MSM	22.03	71
OG *Ochrobactrum* sp. 2 (T1D1)	6	MSM	9.25	71
OG *Pseudomonas* sp. 1 (T3A2)	5	YEMS	11.87	189
OG *Stenotrophomonas* sp. (T3B1)	5	YEMS	7.56	69
OG *Enterobacter* sp. 3 (T3C2)	5	YEMS	4.24	69
OG *Pseudomonas* sp. 2 (T5C2)	6	YEMS	9.25	71
OG *Enterobacter* sp. 4 (T5D3)	6	YEMS	2.46	47
EF *Bacllus* sp. (anox) (T1A1)	6	YEMS	1.38	24
EF *Pseudomonas* sp. 1 (T1B1)	5	YEMS	38.53	46
EF *Micrococcus* sp. (T1C1)	5	YEMS	3.02	21
EF *Pseudomonas* sp. 2 (T2A1)	5	MSM	52.57	93
EF *Brevundimonas* sp. (anox) (T3B1)	5	YEMS	1.18	189
EF *Paenibacillus* sp. (T4B1)	6	MSM	1.00	47

3.2.6 Investigation of Isolate Lignin Breakdown Products and Mechanisms

In order to determine the types of catabolic mechanisms being employed and extent of catabolism owing to isolate breakdown of polymeric lignin, an experiment was carried out involving GC-MS analysis of lignin-derived phenolic compounds produced by isolates. This work was also carried out to identify the industrial relevance of lignin breakdown products produced by the isolates. This experiment involved all of the Olduvai Gorge and elephant faeces isolates, which were grown at their respective enrichment temperatures (OG: 25 ºC; EF: 37 ºC) in YEMS medium with lignin (0.1 g/L instead of 0.05 g/L used in enzyme assay cultures), and oxidative enzyme assays (ABTS and 2,4-DCP assays, discussed above) were run in parallel, to link enzymatic activity with phenolic compound presence. A negative control was also prepared containing YEMS medium and lignin (without inoculum), including an additional MilliQ water control. Cultures were incubated for 3 weeks, during which daily samples were taken for ABTS and 2,4-DCP enzyme assay analysis (to determine timepoints at which enzyme activity peaked and regressed) and for later lignin-derived phenol extraction. An overview of the work involved in this experiment is provided in **Figure 19**.

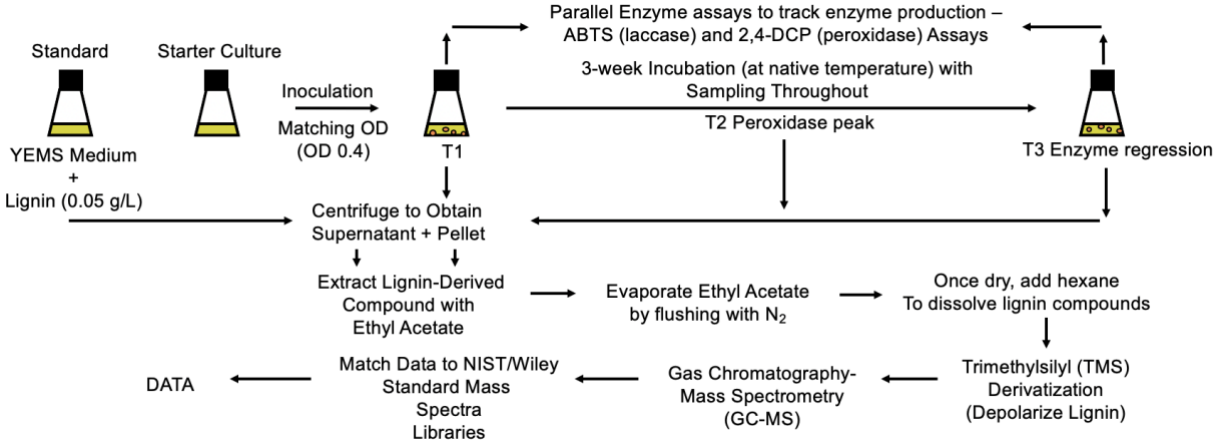

Figure 19 *- Schematic overview of workflow used to carry out GC-MS analysis of microbial lignin-derived phenolic compounds. Figure encompasses isolate cultivation, control (standard) culture preparation, enzyme assay analysis, phenol extraction from cell supernatant and GC-MS analysis.*

From the oxidative enzyme assay results (carried out using the same methodology as the enzyme assays in the previous section), no laccase enzyme activity was observed in the ABTS assay (**Figure 20**) for any of the isolates tested, much like the results shown for the ABTS assay in the previous section (shown above in **Figure 17**). Furthermore, similar to previous results, the laccase positive control maintained enzyme activity

around 8 U/L and slowly decreased over time. The positive results observed in the laccase positive control, as well as the lack of activity observed in the uninoculated and water controls, showed that the ABTS assay functioned successfully.

Despite the lack of laccase activity observed, the results of the 2,4-DCP enzyme assay (Olduvai Gorge isolates in **Figure 21**, elephant faeces isolates in **Figure 22**), again, showed peroxidase activity present in all isolates. The activity of the HRP positive control ranged between ~14-19 U/L, which was similar to the HRP activity observed in the round-5 enzyme assay results, suggesting better replication between rounds compared to the laccase control in the laccase assay. Most isolates also showed similar peroxidase production patterns to previous results, with peak production in all isolates occurring by the day-10 (240 hours) time-point and enzyme activity stabilising by day-21 (504 hours). The most active isolates (highest peak activity) were OG-T3A2 (*Pseudomonas* sp.), followed by EF-T1B1 (*Pseudomonas* sp.), OG-T1B2 (*Ochrobactrum* sp.) and OG-T5C2 (*Pseudomonas* sp.). *Enterobacter* spp. (OG-T1A2, OG-T3C2 and OG-T5D3) also produced relatively high peroxidase activity, however activity started to increase after 1-2 days of delay, whereas the *Pseudomonas* spp./*Ochrobactrum* spp. immediately expressed peroxidase after addition of lignin. Like the previous experiments, isolates EF-T1A1 (*Bacillus* sp.), EF-T3B1 (*Brevundimonas* sp.) and EF-T4B1 (*Paenibacillus* sp.) expressed low levels of peroxidase activity, perhaps due to their anoxic nature and therefore slower proliferation rates. As mentioned previously, however, these isolates may have instead been facultative anaerobes and therefore simply did not express peroxidase enzyme. According to the Bacdive database (Reimer *et al*., 2022), most *Brevundimonas* isolates are listed as aerobes and few *Paenibacillus* are listed as anaerobes. Therefore, it is more likely that these isolates were instead facultative anaerobes and did not produce peroxidase enzymes. The EF *Pseudomonas* spp., which were observed as the most active isolates in previous enzyme assay experiments, showed lower activity in this experiment when grown at 37 ºC, again demonstrating the impacts of growth temperature on enzyme activity.

According to the enzyme activity patterns of isolates, samples from day-1 (24 hours, T1), day-13 (312 hours, T2) and day-21 (504 hours, T3) were selected for analysis as these timepoints correlated with initial inoculation, post-enzyme peak production and post-sustained enzyme production, respectively, for most isolates. Whilst additional time points were also selected for analysis, the data from these time points have been intentionally left out for simplicity. Samples underwent ethyl acetate (EtoAc) extraction to isolate lignin-derived phenolic compounds, followed by derivatisation (depolarisation) with bis(trimethylsilyl)trifluoroacetamide (BSTFA) and GC-MS analysis.

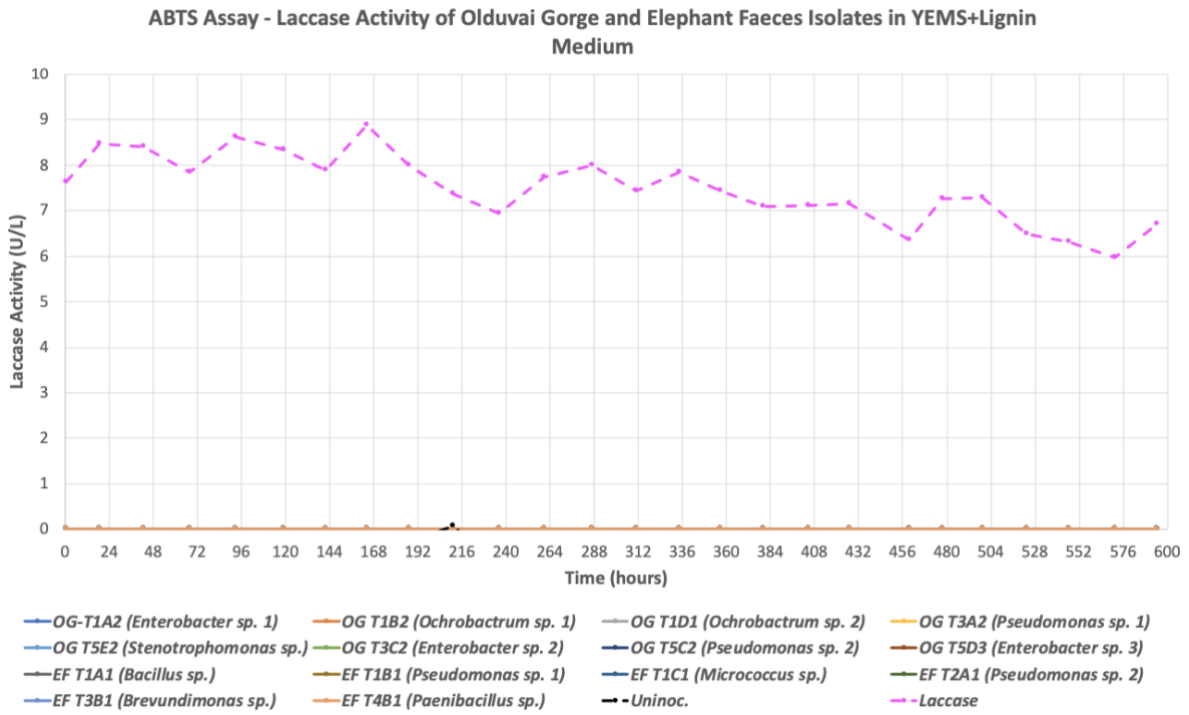

Figure 20 *- ABTS Assay Results – Laccase activity from Olduvai Gorge and Elephant Faeces Isolates in YEMS+lignin medium. Laccase enzyme positive control results in pink dashed line and uninoculated culture negative control results in black dashed line. All data-points for the isolate cultures had activity values <0 U/L, and can therefore not be seen in the figure.*

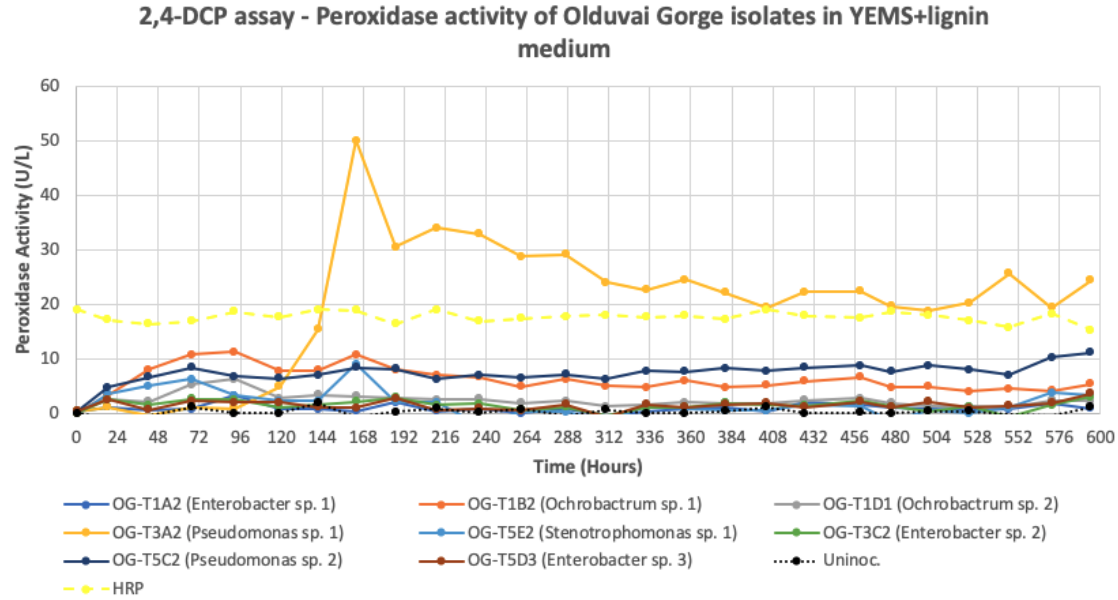

Figure 21 *- 2,4-DCP assay results – Peroxidase activity from Olduvai Gorge and elephant faeces isolates in YEMS+lignin medium (incubated at 25 ºC). HRP enzyme positive control results in yellow dashed line and uninoculated culture negative control results in black dotted line.*

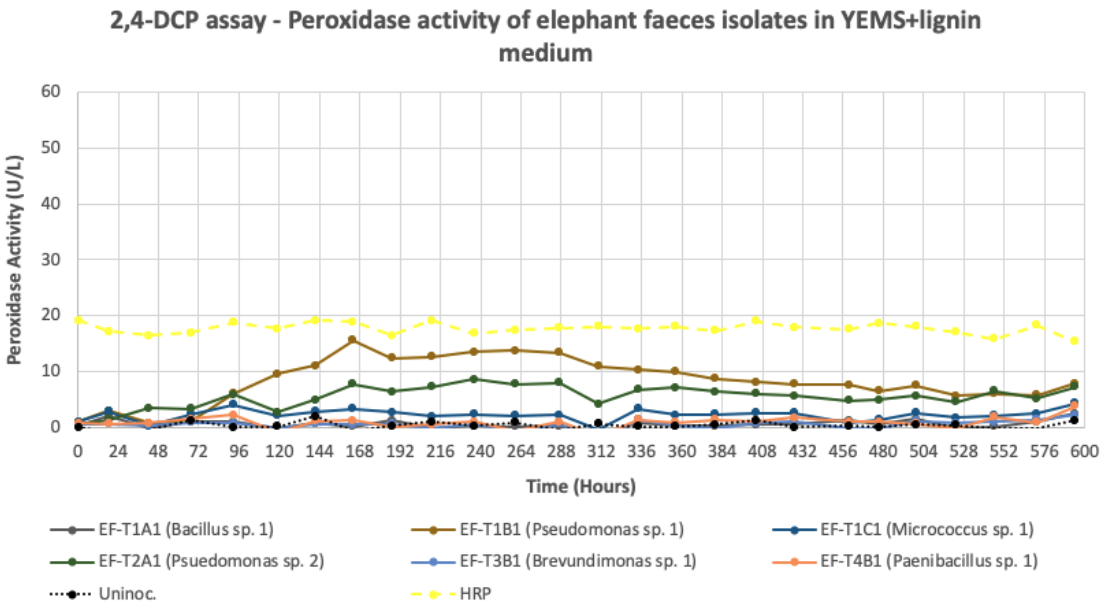

Figure 22 *- ,4-DCP Assay Results – Peroxidase activity from Elephant Faeces Isolates in YEMS+lignin medium (incubated 37 ºC). HRP enzyme positive control results in yellow dashed line and uninoculated culture negative control results in black dotted line.*

3.2.6.1 Lignin-Derived Phenolic Products from Isolates

Phenolic product peaks were verified through matching retention times and major ions within the total ion chromatogram (TIC) data to previous literature (Kaiser and Benner, 2012) and peak area data was pulled from verified peaks to estimate product abundance. Product abundance was calculated relative to total peak area of all products within TIC data for each time-point, shown in **Figure 23**. Product quantification replication between independent GC-MS runs and GC-MS methodologies is not completely accurate, however these quantities can provide important information on product ratios within GC-MS runs. It is also worth noting that the lignin-derived phenolic compounds that were searched for in this study are a small portion of potential products that can be produced from lignin degradation. The product catalogue used in our study was limited by the availability of GC-MS retention time and major ion data used by Kaiser and Benner (2012), however this catalogue includes some of the more common breakdown products, with a number of them being industrially relevant. Illustrations of the chemical structures of the phenolic compounds searched for within this experiment can be found in **Figures 24 and 25**, for reference.

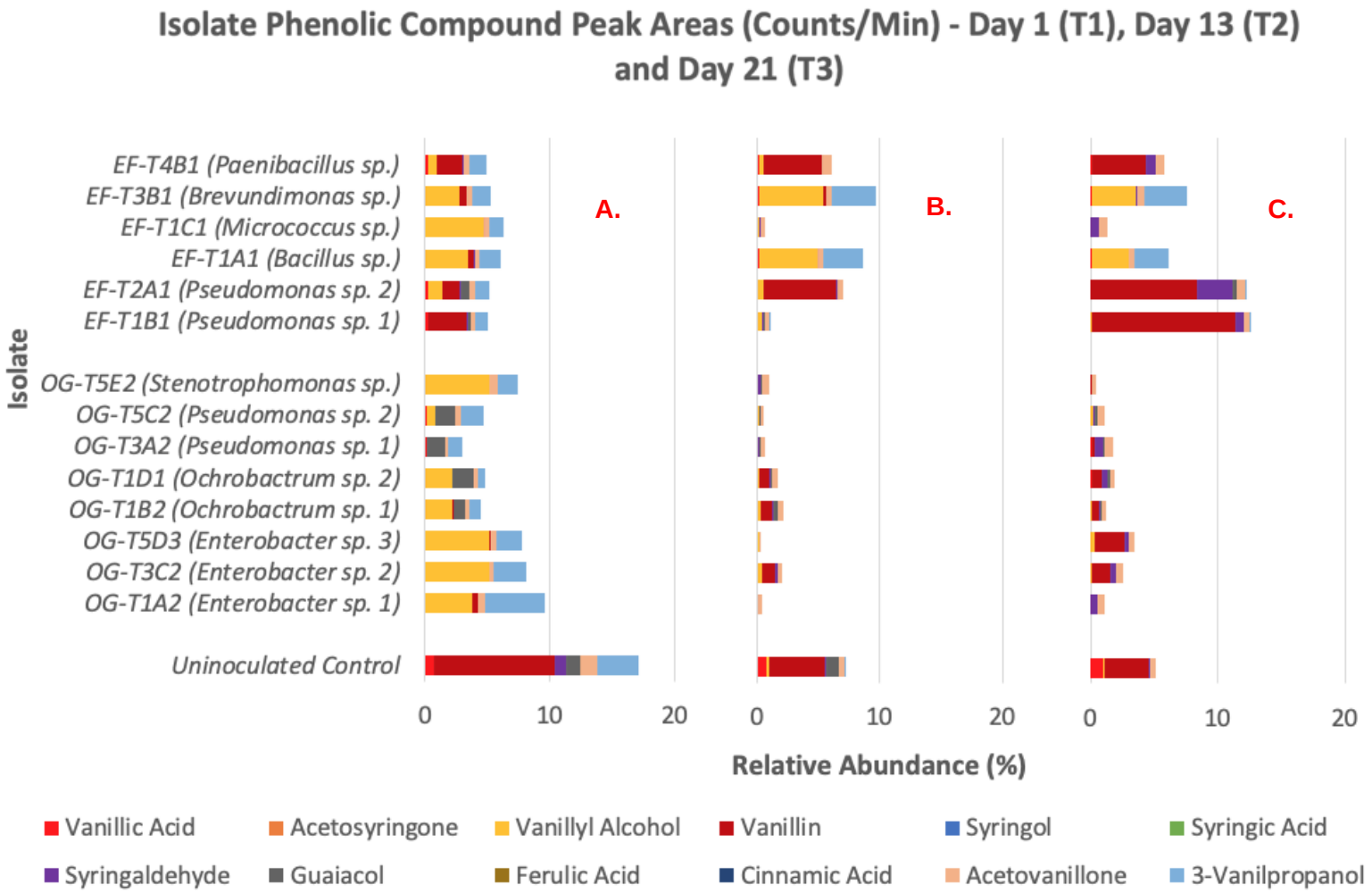

Figure 23 *- A. Day-1, B. Day-13 and C. Day-21 Isolate phenolic compound GC-MS relative abundances – comparisons of relative abundances of lignin-derived phenolic compounds between uninoculated control cultures and Olduvai gorge/elephant faeces isolate cultures.*

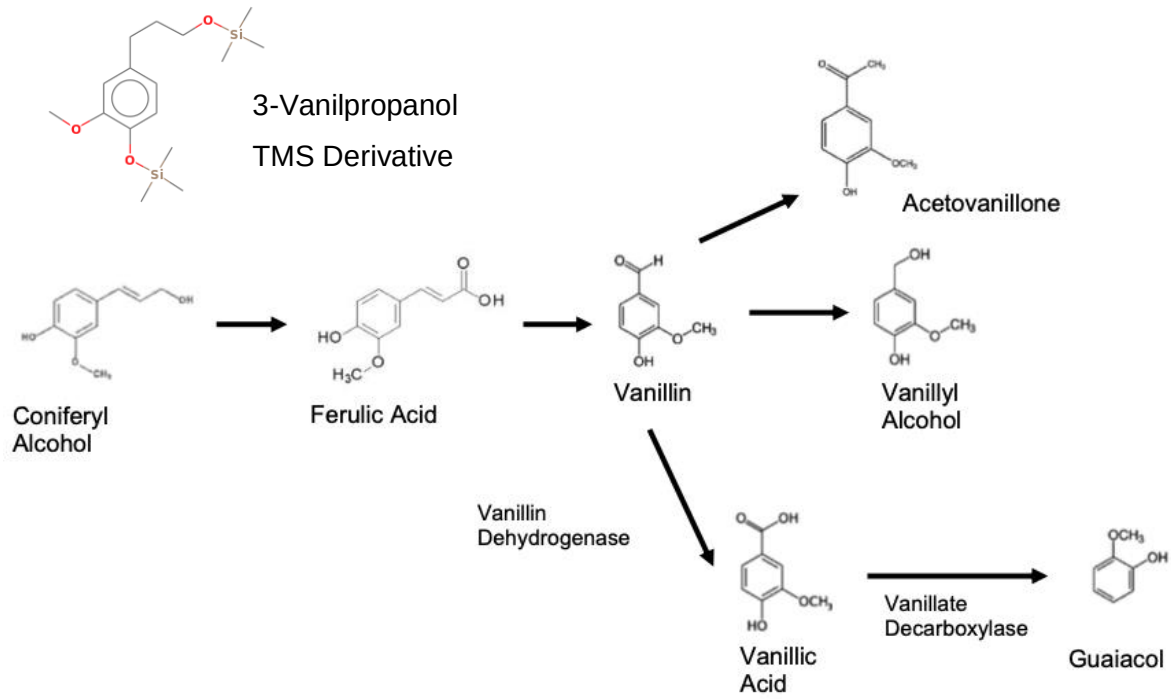

Figure 24 *- Bacterial metabolic pathway for the degradation of vanillin – Degradation of coniferyl alcohol and ferulic acid into vanillin, and subsequent breakdown products of the vanillin monomer. Adapted from the eLIGNIN database (http://www.elignindatabase.com/entries/pathways/P00005.php). Includes image of 3-vanilpropanol, taken from https://webbook.nist.gov.*

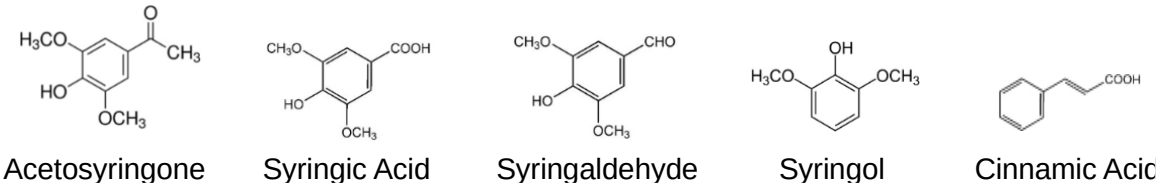

Figure 25 *- Illustrations of additional lignin-derived phenolic compounds found within the GC-MS results, adapted from images taken from https://us.vwr.com.*

Analysis of GC-MS peak area data showed stark differences between the uninoculated lignin control and the isolates already on day-1 (**Figure 23, A.)**. The uninoculated lignin control contained vanillin and 3-vanilpropanol, with relatively lower abundances of vanillic acid, syringaldehyde, guaiacol, and acetovanillone (structures of which are shown in **Figure 24 and 25**). These results in particular illustrate that the alkali lignin substrate used in this PhD project contains several lignin-derived phenolic products, in addition to polymeric lignin. Compared to the uninoculated control, generally, the isolate cultures showed significantly less vanillin and more vanillyl alcohol, with an increase in guaiacol observed within the OG *Pseudomonas* and *Ochrobactrum* cultures. As the differences between the uninoculated control and isolate cultures were so significant,

these changes may have been due to microbial action, with increases in phenolic compounds resulting from degradation of larger compounds (or polymeric lignin) and decreases in phenolic compounds resulting from further metabolism. The less enzymatically active *Enterobacter* sp., *Micrococcus* sp. and *Bacillus* sp. cultures (see **Figures 21 and 22**) contained more phenolic compounds indicating slower phenolic degradation, whereas the opposite results were observed for the more enzymatically active *Pseudomonas* spp./*Ochrobactrum* spp. cultures. Specifically, higher relative abundance of vanillyl alcohol and 3-vanilpropanol was observed in the *Enterobacter* sp., *Micrococcus* sp. and *Bacillus* sp. cultures, whereas generally lower vanillyl alcohol was observed within the *Pseudomonas* spp./*Ochrobactrum* spp. cultures, in addition to higher guaiacol within the OG isolates and higher vanillin in EF isolates.

Compared to the uninoculated control, less degraded (larger) products such as vanillic acid, vanillin and acetovanillone decreased in abundance within the isolate culture, more significantly for vanillin abundance, suggesting all isolates had begun metabolising non-polymeric lignin phenols immediately upon inoculation. The vanillin metabolic pathway, illustrated in **Figure 24**, shows the breakdown of coniferyl alcohol into vanillin and subsequently into acetovanillone, vanillyl alcohol and vanillic acid. Vanillic acid is then broken down into guaiacol (elignindatabase.com, 2021). The decrease in vanillin and increase in vanillyl alcohol within almost all isolate cultures (compared to the uninoculated control) provides evidence for microbial conversion of this lignin-derived compound. Furthermore, the increase in guaiacol within the OG *Pseudomonas* and *Ochrobactrum* spp. suggest potential conversion of any vanillic acid (observed within the uninoculated control) into guaiacol by these microorganisms.

Upon analysing the day-13 (**Figure 23, B.**) and day-21 (**Figure 23, C.**) GC-MS peak area data, generally, it was observed that phenolic products and their abundances continued to change within all isolate cultures. It was, however, also observed that product relative abundances changed within the negative control over time, with decreasing 3-vanilpropanol, increasing vanillic acid and decreasing vanillin between the day-1 and day-21 time-points. As 3-vanilpropanol remained present within the anoxic isolate cultures, but decreased within the control culture, it may be that 3-vanilpropanol oxidised abiotically over time in the uninoculated control, which was done under oxic conditions. The slight increase in vanillic acid and decrease in vanillin throughout time can also be explained by this, as vanillic acid is an oxidation product of vanillin (**Figure 24**). The results observed for EF-T1A1 (*Bacillus* sp.) and EF-T3B1 (*Brevundimonas* sp.) show little change with regards to product presence, however the abundance of vanillyl alcohol and 3-vanilpropanol appears to increase over time. As these isolates were thought to be

anaerobes and were grown in anoxic conditions, it is possible that these isolates were slow-growing and therefore less active than the other isolates (explaining the lack of change in product presence). With this being said, vanillic acid appears at the day-13 time-point and vanillin appears within the cultures by the day-21 time-point, suggesting low lignin metabolic activity.

Within the other isolate cultures, generally, it can be seen that vanillyl alcohol and 3-vanilpropanol decrease significantly at day-13, and abundances of vanillin begin to increase for isolates OG-T1D1, OG-T1B2 (both *Ochrobactrum* spp.) and OG-T3C2 (*Enterobacter* sp.), and more so within the EF-T2A1 (*Pseudomonas* sp.) and EF-T4B1 (*Paenibacillus* sp.) cultures, along with acetovanillone abundances. At the day-21 time-point, this pattern of increase for vanillin continues (with more significant increases in EF-T1B1 (*Pseudomonas* sp.) and OG-T5D3 (*Enterobacter* sp.) along with an increase in syringaldehyde in both EF-T1B1 and EF-T2A1 (both *Pseudomonas* spp.), EF-T4B1 (*Paenibacillus* sp.) and several OG isolate cultures. From these results, significant changes in phenolic product presence and abundance can be observed in the OG *Pseudomonas*, *Ochrobactrum*, *Enterobacter*, *Stenotrophomonas* spp. and EF *Micrococcus* and *Paenibacillus* spp. cultures. Furthermore, EF-T1B1 and EF-T2A1 (both *Pseudomonas* spp.) seem to be the most active in lignin metabolism, with the most significant changes in phenolic product presence and abundance throughout incubation compared to the uninoculated control (as well as being the most active isolates in previous enzyme assay analysis).

As vanillin generally increases over time in most cultures, this would suggest that larger components of the lignin polymer are being metabolised between the day-1 and day-21 time-points, for example ferulic acid, however no peaks representing ferulic acid were found in the GC-MS TIC data (according to vanillin breakdown pathway displayed in **Figure 24**). Furthermore, the increases in acetovanillone at day-21 suggests that the abundant vanillin present at day-13 is being broken down, which is most significantly displayed by the EF *Pseudomonas* spp.. The increases in syringaldehyde abundances towards the end of incubation also indicate that the isolates begin to degrade larger lignin subunits, as syringaldehyde is a breakdown product derived from sinapic acid (derived from sinapyl alcohol – a primary component of lignin; elignindatabase.com, 2021).

Looking closer at the results of the more active EF isolates, the EF *Pseudomonas* spp. cultures can be seen to generally contain lower product abundance compared to the control at day-1, suggesting metabolism of the phenolic products within the alkali lignin substrate. However, differences between the two isolates can be seen on day-13 with

EF-T1B1 containing significantly less products than EF-T2A1. Despite this, by day-21 both isolate cultures contain similar product abundances (except EF-T2A1 has higher syringaldehyde and lower vanillin abundance).

Very similar results to EF-T2A1 can also be observed within EF-T4B1 (*Paenibacillus* sp.), however EF-T1C1 (*Micrococcus* sp.) showed results more akin to the OG isolates. EF-T1B1 was the most active with regards to peroxidase production, and enzyme activity on day-13 can be seen to be higher compared to EF-T2A1. Taking the differences in product abundances and enzyme activity into consideration for isolate EF-T1B1, it could be possible that the significant decrease in general product abundance at day-13 and higher vanillin abundance at day-21 is due to higher peroxidase activity, as peroxidases act on both polymeric lignin and lignin-derived products. Furthermore, the increase in vanillin (in the EF-T1B1, EF-T2A1 and EF-T4B1 isolate cultures) could also be indicative of peroxidase activity from these isolates as peroxidases such as DyPB have been shown to oxidatively cleave C_α -C_β linkages within β-aryl ether lignin subunits creating vanillin (Bugg *et al*., 2016). A study carried out by Bugg et al (2011a) found that bacterial breakdown pathways are extremely varied between bacterial species and highlighted that vanillin and vanillic acid utilisation is common amongst bacteria via the vanillin dehydrogenase pathway. Research regarding bacterial lignin breakdown products is limited, however some studies have been carried out with select bacterial species, such as those involving *Bacillus* sp. (Raj et al., 2007), *Pandoraea* sp. (Kumar et al., 2015) and *Novosphingobium* sp. (Chen et al., 2012).

The OG isolates all showed similar changes in product abundance patterns over the course of incubation, however some differences can be observed between the *Enterobacter*/*Stenotrophomonas* and the *Pseudomonas*/*Ochrobactrum* sp. isolates. For example, the *Enterobacter* and *Stenotrophomonas* isolate cultures contained higher vanillyl alcohol and 3-vanilpropanol abundance on day-1 compared to the *Pseudomonas* and *Ochrobactrum* cultures, which instead contained higher guaiacol. On day-13, all isolates had distinctly lower product abundances compared to the previous time-point, with no significant differences between isolates except similar product patterns in the OG-T3C2 (*Enterobacter* sp.), OG-T1B2 and OG-T1D1 (both *Ochrobactrum* spp.) cultures (with higher vanillin content). The same can be said for the day-21 time-point, in which most isolate cultures can be seen to contain vanillin and syringaldehyde, both products providing evidence for potential lignin and lignin-derived product degradation.

From these results, the significant differences (general decrease in phenolic compounds) observed between the uninoculated and isolate cultures throughout the experiment

suggests microbial metabolism of phenolic lignin-derived compounds. Furthermore, the general increases in vanillyl alcohol and decreases in vanillin within isolate cultures suggests the use of the microbial vanillin degradation pathway (shown in **Figure 24**). The higher phenolic compound abundance, as well as the presence of syringaldehyde, observed within the EF *Pseudomonas* spp. cultures indicates that these isolates were the most active degraders (by releasing more phenolic compounds and potentially breaking down polymeric lignin). The lack of change in the presence of phenolic compounds within the anaerobic isolate cultures (EF-T1A1 and EF-T3B1) over time therefore indicates a lack of lignin-degrading activity, however differences were observed between these cultures and the uninoculated control, indicating the initial use of the vanillin degradation pathway.

3.3 Conclusions

In summary, a varied catalogue of lignin-degrading microorganisms were enriched and isolated from Olduvai Gorge soils, elephant faeces and compost using minimal medium and lignin as sole carbon source. The majority of these isolates were shown to have the genetic potential to produce the lignin-oxidising enzymes laccase or DyP-type peroxidase (via the *lmco* and *dyp* genes, respectively), with the *Pseudomonas* isolates having genes for both. Despite the presence of the *lmco* gene in many of the isolates, no laccase activity was observed within any of the isolates tested in the ABTS oxidative enzyme assay. However, in the 2,4-DCP enzyme assay, almost all isolates demonstrated evidence for peroxidase production, with the best performing isolates being the EF *Pseudomonas* spp. followed by OG *Ochrobactrum* and *Pseudomonas* spp.. Whilst the peroxidase activity of EF *Pseudomonas* spp. was significantly higher than those of a similar study (Salvachua *et al*., 2015), the differences in cultivation and enzyme assay methodologies used by other studies made it difficult for comparisons to be made between datasets. Therefore, it was difficult to determine whether or not the peroxidase activity of the EF *Pseudomonas* spp. was higher than that of previously reported species. Despite this, the key point to be made about these results was that activity was observed and the EF *Pseudomonas* spp. displayed significantly high activity.

Through GC-MS analysis of isolate lignin breakdown products, all isolates were shown to degrade lignin-derived phenolic compounds, except both EF anoxic isolates, which showed little activity likely due to their anaerobic growth and therefore slow proliferation rates (perhaps due to them being facultative anaerobes). The EF *Pseudomonas* spp., which showed high peroxidase production within the enzyme assays, also appeared to be the most active lignin degraders within the GC-MS data, through production of vanillin

and syringaldehyde. Increases in vanillin and syringaldehyde suggested conversion of β-aryl ether lignin subunits and sinapyl alcohol subunits, respectively, both present within polymeric lignin. The evidence provided by both the enzyme assay and GC-MS analysis of these isolates suggests that the *Pseudomonas* spp. isolated from the elephant faeces samples are key lignin degrading microorganisms within these environments and show potential for use in industrial processes. Despite the EF *Pseudomonas* spp. isolates being the most active of those obtained, the other isolates were enriched from these environments using lignin as sole carbon source and each isolate demonstrated relatively high peroxidase activity and significant changes in lignin-derived metabolic products when incubated with lignin. This, therefore, provides evidence to suggest that these isolates, too, are active contributors to lignin degradation within their respective environments.

3.4 Limitation and Future Work

With this project being an ambitious one and time being restricted (not to mention the impacts of the COVID-19 pandemic), much of the work originally planned for this study was not possible within the allotted timeframe. For example, the compost and fungal isolates (which were shown to be known lignin-degrading microorganisms) did not undergo the planned oxidative enzyme assays and GC-MS analysis. Furthermore, time did not allow for isolation of lignin-degrading microorganisms from the Hickling sediment samples.

The growth experiment data (**Section 3.2.3**) was obtained whilst the enzyme assay experiments were being performed (as detailed in **Section 3.2.5**) and, as mentioned in this section, it was found that 25 ºC was the optimum growth temperature for oxidative enzyme production for all isolates. As the isolates were obtained at certain enrichment temperatures (OG isolates at 25 ºC and EF isolates at 37 ºC), it was assumed that these temperatures would be the optimum temperatures for growth and enzyme production, hence these experiments being carried out later in the project. In hindsight, these experiments would ideally have been carried out before the enzyme assays and GC-MS analysis, providing information on optimum conditions for isolate growth earlier. The growth experiments being carried out alongside the enzyme assays introduced limitations on time available for these experiments, therefore there were limitations on what could be tested. If these experiments were to be repeated, the growth conditions would have included a wider range of temperatures and isolates would have been tested in duplicate or triplicate, providing a more robust dataset. The growth experiments would also have been carried out for longer durations to match the length of time the oxidative

enzyme assays were carried out providing full growth curves for the isolates. Furthermore, similar to the enzyme assays, the growth experiments would have included replicates which were grown in MSM media to provide information on optimum growth conditions in this medium alongside the YEMS medium.

In the enzyme assays, more time would have allowed further testing of optimum enzyme production cultivation conditions (with more pH and temperature replicates) for each of the sixteen isolates, as well as extraction and testing of the enzymes produced, providing a better understanding of the enzymatic kinetics (optimum functionality) and therefore industrial potential. This work, however, would likely have been time-consuming and would not have been possible within the PhD time-frame. In addition, a large portion of the work carried out on the enzyme assays was spent optimising an additional 2,6-DMP assay (Salvachua *et al*., 2015), which would have provided information on manganese-dependent and -independent peroxidase enzymes, the data from which was not shown in this thesis due to a lack of reliable positive control results. With more time, this enzyme assay could have been fully optimised and used to determine activity of these enzymes from the isolates, providing a deeper understanding of their enzymatic and therefore industrial potential. Additionally, more time would have allowed isolates to be tested in replicate (duplicate or triplicate) to make the enzyme assay data more robust. For example, the ABTS/2,4-DCP/2,6-DMP assays would have been carried out with isolates which had been cultured with lignin in triplicate cultures. Assays would have been carried out with the supernatant from these triplicate cultures, leading to more dependable data.

Within the GC-MS experimentation carried out on the isolates, these experiments would have been repeated to account for the growth optima of EF isolates (25 ºC instead of the 37 ºC used), which would likely have provided better data from these isolates (as shown in the enzyme assay experiments). Moreover, several additional samples were extracted from time-points throughout this experiment (for day-4, day-9 and day-18) which were also to be analysed via GC-MS, however time did not permit this work to be carried out. The additional time-points would likely have provided more information on how the phenolic compounds observed in the data presented changed throughout the incubation time, and therefore would have potentially provided more information on the metabolic mechanisms of the isolates.

With more time, whole genome sequencing (WGS) would have been carried out with isolates in order to obtain more accurate identifications and obtain a deeper understanding of their potential lignin-degrading mechanisms via exploration of their lignin degrading genes. With accurate identities, the isolates obtained in this study could

have been better compared to known/unknown lignin degrading microorganisms. Furthermore, the genetic information provided by WGS would have allowed full metabolic maps to be constructed for degradation of lignin/lignin-derived compounds for each isolate. In addition to WGS, functional cloning was also initially planned to be used to determine the functionality of the oxidative enzyme genes identified within this study. However, due to the aforementioned restrictions, time did not permit this work to be carried out.

3.5 Final Words

This study presents the utility of using culture-dependent methodologies for the enrichment and isolation of industrially-relevant substrate-specific microorganisms. Despite the aforementioned limitations, the results presented in this chapter provide interesting insights into the identity and functionality of several key lignin-degrading species, which could have potential for industrial use. These isolates were obtained from environments which are suitable habitats for specialized lignin-degraders and are therefore of high industrial relevance. The microorganisms isolated in this project have been stored as glycerol stocks for future exploration and analysis. Future analysis would likely involve a deeper exploration of the enzymatic capabilities of these isolates, incorporating more assays targeting lignin-active enzymes, in combination with WGS and functional coning to explore their functional potential.

4 Chapter 4: Exploration of Lignin-Degrading Microbial Communities Within Different Environments Using an Array of Nucleic Acid Sequencing Approaches

4.1 Introduction

This results chapter will address **Objectives 1 and 2** stated in **Introduction Section 1.6.2**, which encompass the exploration of environmental samples (detailed in **Methods and Materials Section 2.2**) using culture-independent methodologies, with a focus on identifying key microorganisms and their functional potential with regards to lignin degradation. Specifically, the culture-independent methodologies used to obtain the results presented in this chapter involved a combination of microbial SSU rRNA gene amplicon sequencing, metagenomic sequencing and nucleic acid stable isotope probing. These methodologies were used in combination to provide a detailed insight into microbial lignin degradation at community level and to potentially reveal how individual species contribute towards the metabolism of the lignin polymer at community scale.

As mentioned previously, lignocellulosic material is of high importance to industry however the presence of lignin within the biopolymer makes conversion into high-value products difficult (Kumar *et al*., 2008; Mathews *et al*., 2015). Current lignocellulose conversion processes involve energy-demanding machinery and expensive chemicals in order to remove lignin from the polymer, and results in significant underutilisation of the lignocellulosic biomass (Gabrielii *et al*., 2000; Vikman *et al*., 2002; Martínez *et al*., 2005; Bridgwater, 2006; Mathews *et al*., 2015; Kameshwar and Qin, 2016). Bioprocessing of lignocellulose, involving the use of lignin-degrading microorganism and their enzymes, has attracted attention as a potential alternative to conventional processing methods, with a focus on bacteria due to their suitability for industrial use (Kumar *et al*., 2009; Taylor *et al*., 2012; Ahmed *et al*., 2016; Tian *et al*., 2016).

So far, research has shown that specialized (primary) lignin-degrading bacteria employ a variety of oxidative enzymatic mechanisms (such as laccases and peroxidases) to break down the lignin polymer into its phenolic constituent components, making them available to secondary/tertiary degrader microorganisms (Guerriero *et al*., 2016; Kameshwar and Qin, 2016). Studies carried out by Lu *et al* (2014), Woo *et al* (2014) and Tian *et al* (2016) revealed that environments differ in terms of key lignin degrading

microorganisms and the mechanisms they use. Furthermore, many studies have identified bacterial species with the potential/ability to degrade lignin using these mechanisms, including the potent lignin-degrading enzyme DyP peroxidase, such as those carried out by Bandounas *et al*., 2011, Taylor *et al*., 2012 and Tian *et al*., 2016. However, despite recent research, little is known of how bacteria interact with and contribute towards the degradation of lignin at community scale (and at an individual scale) within the natural environment. Furthermore, the bacterial species which have been identified and characterised so far are not efficient enough to compete with conventional lignocellulose pre-treatment methods. Therefore, more information on community-level lignin degradation is required, involving the identification of previously unknown lignin-degrading species and their mechanisms, as well as that of the microorganisms they collaborate with within the environment. In addition, research focussed on environments which are abundant in plant material and are challenging for some microorganisms to inhabit (such as the anoxic regions and high temperatures in compost sanitisation) could provide candidate species with specialized lignin-degrading mechanisms. Further research will provide information vital to understanding the role these microorganisms play in the process of plant biomass degradation and carbon cycling within the natural environment, and to potentially improve or replace current lignocellulose conversion processes.

The more “traditional” method of exploring substrate-active microorganisms within complex environments involves enrichment and isolation of said microbes followed by functional assays, however many microorganisms cannot be cultured in a laboratory environment. Next generation sequencing (NGS) technologies, such as metagenomics, are powerful tools which can aid in searching for microorganisms with lignin-degrading potential. Several studies have already used NGS methodologies to explore the lignin-degradation potential of microbial communities within select environments, such as those carried out by DeAngelis *et al*., 2011 and Wang *et al*., 2016. These studies provided an insight into the functioning of previously unknown lignin-degrading microorganisms within their respective environments. An alternative method of screening environmental samples for microorganisms involved in the incorporation of a specific substrate is stable isotope probing (SIP), as detailed in the **Introduction Chapter**. With regards to microbial lignin degradation, this approach, combined with metagenomic methods would allow targeted analysis of complex environmental samples, screening for active lignin degraders within the samples and genes encoding the enzymes they use to degrade lignin. Such studies have only been carried out on a limited number of environments, but have shown the combination of SIP with NGS methodologies to be an

effective method of investigating lignin degradation (Darjany *et al*., 2014; Wilhelm *et al*., 2018).

In this study, high-throughput microbial SSU rRNA gene amplicon sequencing combined with metagenomic sequencing were used to explore microbial communities (and potential lignin-degrading mechanisms) present within various environmental samples, including those from extreme environments such as compost and plant-feeding animals/insects (**Objective 2**). This allowed reconstruction of microbial community profiles within each environmental sample, elucidating presence and relative abundance of microbial taxa and community diversity. Analysis of environmental metagenomes allowed identification in addition to the screening for genes encoding oxidative enzymes (such as laccases and peroxidases), and therefore the enzymatic potential of the entire microbial community within each environmental sample. Nucleic acid stable isotope probing (SIP) with ^{13}C-labelled lignin combined with high-throughput sequencing approaches were used to identify the key microorganisms involved in lignin degradation within these environments (**Objective 1**). Unfortunately, due to the lack of labelling observed within gradients, however, no SIP fractions were sent for microbial SSU rRNA gene amplicon sequencing/metagenomics. Samples derived from the gut and burrow of wood-boring bivalve molluscs (mentioned in **Methods and Materials Section 2.2**) also underwent similar analysis to that of the other 10 environmental samples, however the results of this analysis was incomplete and is therefore not presented in this thesis.

4.2 Results

The following results were derived from various analyses carried out on nucleic acids extracted from the environmental samples used in this study – Olduvai Gorge soils (SS, ZS and TM); elephant faeces (EF1 and EF2); compost soil (C54, C55 and C61) and Hickling sediment samples (HS(c) and HS(s)) (details of which can be found in **Methods and Materials Section 2.2**). Nucleic acids were extracted using two different methodologies. The first methodology involved the use of 6% hexadecyltrimethylammonium bromide (CTAB), and the second methodology involved the use of 3% sodium dodecyl sulphate (SDS). These methodologies were derived from Griffiths *et al* (2000) and Burgmann *et al* (2003), respectively (detailed in **Methods and Materials Section 2.5.1**). All environmental samples underwent nucleic acid extraction using both methodologies, followed by nucleic acid assessment (detailed below), in order to determine the best method of extraction from each sample.

4.2.1 Assessment of Nucleic Acid Extraction Methods

Agarose gel electrophoresis image analysis of the environmental sample nucleic acid extracts (shown below in **Figures 26, 27 and 28**) was carried out to determine DNA/RNA yield and quality. In addition, double-stranded DNA (dsDNA) concentration within the extracts was quantified using Qubit fluorometry to determine how much DNA was extracted from the samples (**Table 11**). CTAB extracts from the OG samples were not included in this table due to no nucleic acids being extracted (as discussed below).

From the gel images, optimal extracts (high abundance and quality of nucleic acids) were obtained from the Olduvai Gorge (OG) soil samples (SS, ZS and TM) (**Figure 26**) using the SDS protocol only. The OG CTAB extracts can be seen to contain no nucleic acids (indicated by an absence of bands in the gel image), however bands representing DNA (larger band at the top of the image) can be observed in the SDS extracts. Also, worth noting is the apparent lack of RNA within the OG SDS extracts, indicated by the lack of two bands below the DNA band representing small and large subunit RNA. Furthermore, a faint DNA band was observed within the TM SDS extract, indicating low DNA yield (which was reflected in the Qubit dsDNA quantification results in **Table 11**). Following extraction, the OG SDS extracts appeared dark in colour (proposedly due to high humic acid content) and were therefore further purified using a Zymo Research OneStep PCR Inhibitor Removal Kit before use (as detailed in **Methods and Materials Section 2.5.2**). Due to the reasons stated above, the OG SDS extracts were selected for further analysis.

The results obtained from elephant faeces (EF) samples (EF1 and EF2) (**Figure 26**) revealed that DNA and RNA was successfully extracted from these samples using both the SDS and CTAB extraction protocols. The bands representing RNA within the EF SDS extracts were observed to be fainter than that of the CTAB extracts (likely indicating lower RNA concentration). Furthermore, the dsDNA concentration of the EF SDS extracts (**Table 11**) was observed to be lower than that of the CTAB extracts. Therefore, the EF CTAB extracts were selected for further analysis.

Similar to the EF extracts, both the CTAB and SDS methodologies successfully extracted nucleic acids from the Hickling sediment (HS) samples (HS(s) and HS(c)) (**Figure 27**), indicated by strong DNA bands and lighter RNA bands. Little difference was observed between the CTAB and SDS extracts within the gel image, however the Qubit quantification result for these samples revealed higher dsDNA concentration within the HS(s) SDS extract. The HS SDS extracts were therefore selected for further analysis.

On the contrary to the OG extracts, analysis of the compost soil (CS) extracts (C54, C55 and C61) (**Figure 28**) revealed that the CTAB extraction methodology produced better quality DNA/RNA from these samples than the SDS methodology. Strong bands representing DNA and RNA were observed within the CTAB extracts, however the SDS method appeared to yield no nucleic acids at all. The Qubit quantification results, however, showed that dsDNA was present within both the CTAB and SDS extracts for all CS samples, with the CTAB extracts containing higher concentrations than the SDS extracts. According to the results of the agarose gel image analysis and Qubit quantification analysis, the CS CTAB extracts were selected for further analysis.

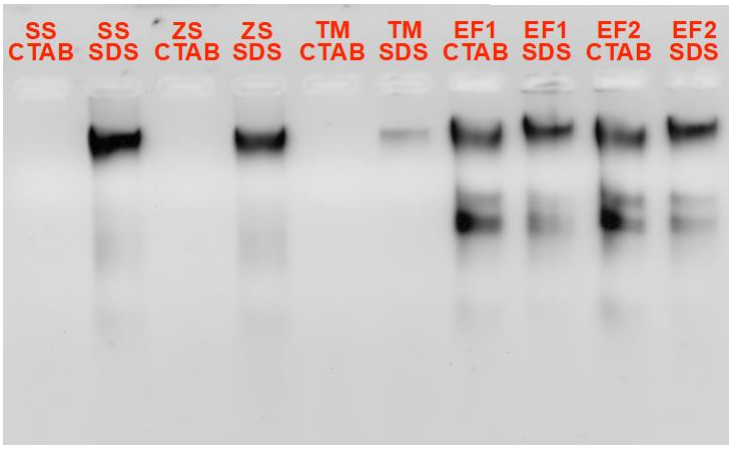

Figure 26 *- Agarose gel electrophoresis image of unaltered nucleic acid extracts obtained from Olduvai Gorge (OG) and elephant faeces (EF) samples. Image shows nucleic acid extracts obtained via CTAB and SDS extraction protocols. SS = Savanna Soil; ZS = Zebra-Grazed Soil; TM = Termite-Mound Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces. SDS = extract obtained via SDS extraction method; CTAB = extract obtained via CTAB extraction method.*

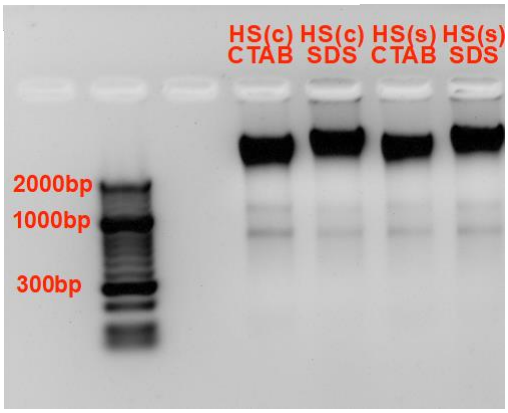

Figure 27 *- Agarose gel electrophoresis image of unaltered nucleic acid extracts obtained from Hickling sediment (HS) samples. Image shows nucleic acid extracts obtained via CTAB and SDS extraction protocols. DNA ladder = 50 bp Hyperladder. HS(c) = Hickling Sediment (Centre Broad); HS(s) = Hickling Sediment (Broad Shore). SDS = extract obtained via SDS extraction method; CTAB = extract obtained via CTAB extraction method.*

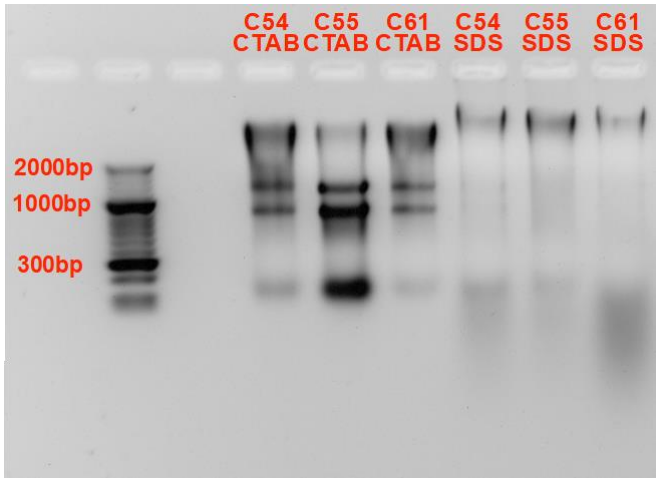

Figure 28 *- Agarose gel electrophoresis image of unaltered nucleic acid extracts obtained from compost soil (CS) samples. Image shows nucleic acid extracts obtained via CTAB and SDS extraction protocols. DNA ladder = 50 bp Hyperladder. C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1. SDS = extract obtained via SDS extraction method; CTAB = extract obtained via CTAB extraction method.*

Table 11 *– Environmental sample nucleic acid extract Qubit Fluorometer quantification. Measured using a Qubit dsDNA broad range kit which quantifies double-stranded DNA (dsDNA) labelled with fluorescent dye. SDS and CTAB extraction protocol). SS = Savanna Soil; ZS = Zebra-Grazed Soil; TM = Termite-Mound Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; HS(c) = Hickling Sediment (Centre Broad); HS(s) = Hickling Sediment (Broad Shore); C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1. SDS = extract obtained via SDS extraction method; CTAB = extract obtained via CTAB extraction method. CTAB extracts from the OG samples were not included in this table due to no nucleic acids being extracted.*

Sample	SS SDS	ZS SDS	TM SDS	EF1 CTAB	EF1 SDS	EF2 CTAB	EF2 SDS
dsDNA Concentration (ng/µl)	64.5	60.5	11.6	72	39.5	132	25.75

Sample	HS(c) CTAB	HS(c) SDS	HS(s) CTAB	HS(s) SDS	C54 CTAB	C54 SDS	C55 CTAB	C55 SDS	C61 CTAB	C61 SDS
dsDNA Concentration (ng/µl)	115	118	64.8	113	273	230	316	199	236	166

4.2.2 Environmental Sample Nucleic Acid Quality Assurance and PCR Analysis

PCR was carried out on these extracts with primers targeting the bacterial 16S rRNA gene (Amman *et al*., 1995), as well as functional genes encoding laccase-like multicopper oxidase (*lmco*) (Kellner *et al*., 2008) and dye-decolorising peroxidase (*dyp*) (Tian *et al*., 2016b). This was carried out in order to determine the ability to amplify these genes from these samples and, therefore, investigate their presence. Protocols for PCR

amplification of the bacterial 16S rRNA gene, as well as the *dyp* and *lmco* functional genes, can be found in the **Methods and Materials Sections 2.6.1.1, 2.6.1.3 and 2.6.1.4**, respectively. Degenerate primers targeting the *dyp* functional gene were obtained from a study carried out by Tian *et al* (2016), which used these primers to successfully detect and quantify a broad spectrum of genes coding for bacterial DyP-type peroxidases. Degenerate primers targeting the *lmco* functional gene were obtained from a study carried out by Kellner *et al* (2008), which used these primers to explore the diversity and distribution of bacterial *lmco* genes within forest and grassland soils. These primers were selected for use in this study due to laccases and peroxidases being important enzymes involved in lignin degradation, their target being bacterial genes and broad specificity (therefore allowing amplification of a larger diversity of genes).

Agarose gel electrophoresis analysis of bacterial 16S rRNA gene PCR products obtained from environmental samples (**Figure 29**) show the presence of this gene (expected product size: 1384 bp) within all samples, confirming the presence of bacterial communities within each. Similar results were obtained for the *lmco* (142 bp product size) and *dyp* (300-400 bp product size) functional gene PCR products (**Figures 30 and 31**, respectively), showing the presence of these genes within every sample. Non-specific amplification can also be observed within the results for both *lmco* and *dyp* primer sets, likely due to the degenerate nature of the primers. Degenerate primers, in brief, are designed to cover a wide range of diversity of a specific gene, therefore covering the same related gene from several different microorganisms. Degenerate primers are useful to amplify a desired gene from different microorganisms and a high degree of degeneracy can lead to the amplification of a greater diversity of genes, however high degeneracy can also lead to the amplification of non-desired genes (i.e. non-specific amplification) (Linhart and Shamir, 2007). Whilst non-specific amplification was observed in the results presented below, the desired genes were also amplified, therefore these results served their purpose. These results suggest that a portion of the microbial community within these samples possess the ability to produce laccase and DyP-type peroxidase enzymes.

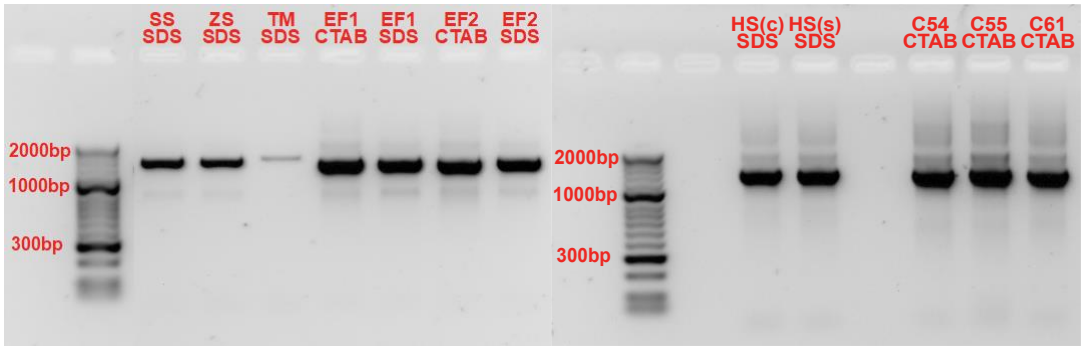

Figure 29 *– Agarose gel electrophoresis image of bacterial 16S rRNA gene PCR amplification products obtained from environmental sample nucleic acid extracts. DNA ladder = 50 bp Hyperladder. Expected product size: 1384 bp. SS = Savanna Soil; ZS = Zebra-Grazed Soil; TM = Termite-Mound Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore); C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1.*

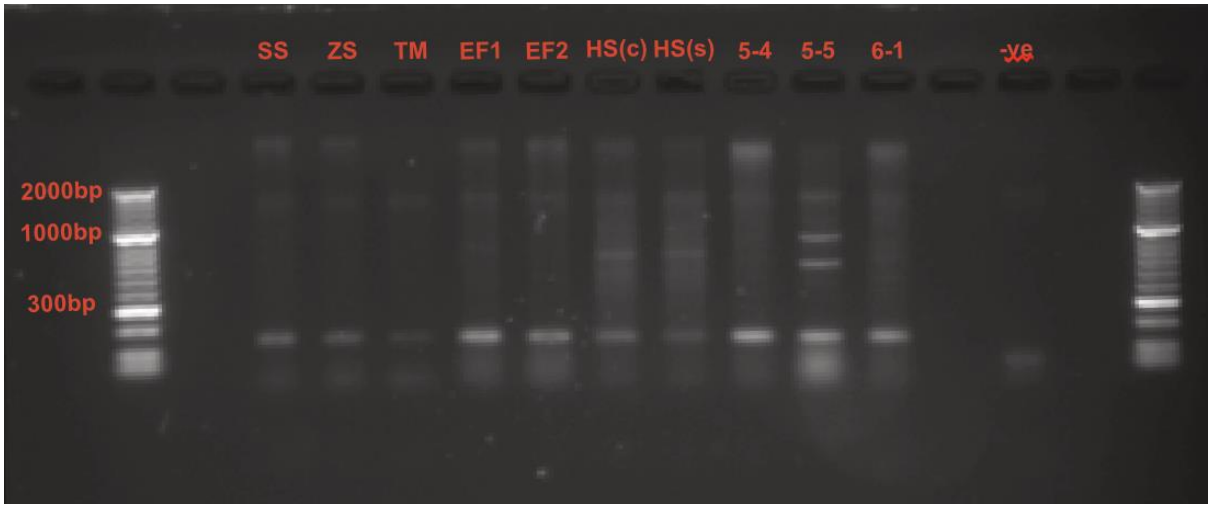

Figure 30 *– Agarose gel electrophoresis image of lmco functional gene PCR amplification products obtained from environmental sample nucleic acid extracts. Negative control (-ve) = water. DNA ladder = 50 bp Hyperladder. Expected product size: 142 bp. SS = Savanna Soil; ZS = Zebra-Grazed Soil; TM = Termite-Mound Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore); 5-4 = Compost Soil 5-4; 5-5 = Compost Soil 5-5; 6-1 = Compost Soil 6-1.*

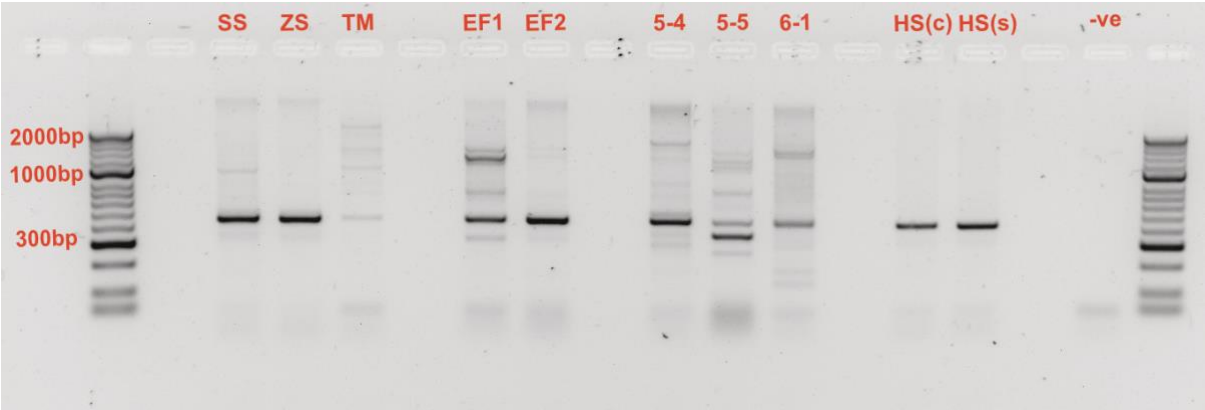

Figure 31 *- Agarose gel electrophoresis image of dyp functional gene PCR amplification products obtained from environmental sample nucleic acid extracts. Negative control (-ve) = water. DNA ladder = 50 bp Hyperladder. Expected product size: 300-400 bp. SS = Savanna Soil; ZS = Zebra-Grazed Soil; TM = Termite-Mound Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore).*

4.2.3 Microbial SSU rRNA Gene and Functional Gene qPCR Analysis

qPCR analysis was carried out in order to amplify and, ultimately, obtain abundances of the bacterial 16S rRNA gene and *dyp* functional genes within nucleic acid extracts from environmental samples (according to the protocols detailed in **Methods and Materials Sections 2.8.1 and 2.8.2**, respectively). In addition, a qPCR protocol involving primers targeting the *Imco* functional gene was also in development, however was not finished due to time limitations (as detailed in **Appendix Section 7.1**). A standard for bacterial 16S rRNA gene quantification was prepared from a PCR amplification product obtained from *E. coli*. A standard for *dyp* gene quantification was prepared (according to procedure detailed in **Methods and Materials Section 2.8.2**) from a PCR amplification product obtained from isolate T3A2 (*Pseudomonas* sp.) obtained from the enrichment (with lignin) of savanna soil carried out in this study (**Chapter 3**). The methodology used to obtain a *dyp* qPCR standard from isolate OG-T3A2 is detailed in **Methods and Materials Section 2.8.2** and **Appendix Section 7.2**.

Results from both the bacterial 16S rRNA gene and *dyp* functional gene qPCR runs, showing absolute gene abundances (copy number) per gram of each sample, are illustrated below in **Figure 32**. These results displayed higher bacterial 16S rRNA gene copy numbers in the Hickling Broad sediment, elephant faeces and compost nucleic acid extracts compared to the Olduvai Gorge samples (SS, ZS and TM). More specifically, the three compost samples (primarily compost sample C55) contain the highest bacterial

16S gene abundance, followed by the elephant faeces and Hickling sediment samples. From the absolute abundances shown in **Figure 32**, the *dyp* abundances can generally be seen to be similar between environments, with the exception of the TM sample, showing lower abundance than other samples.

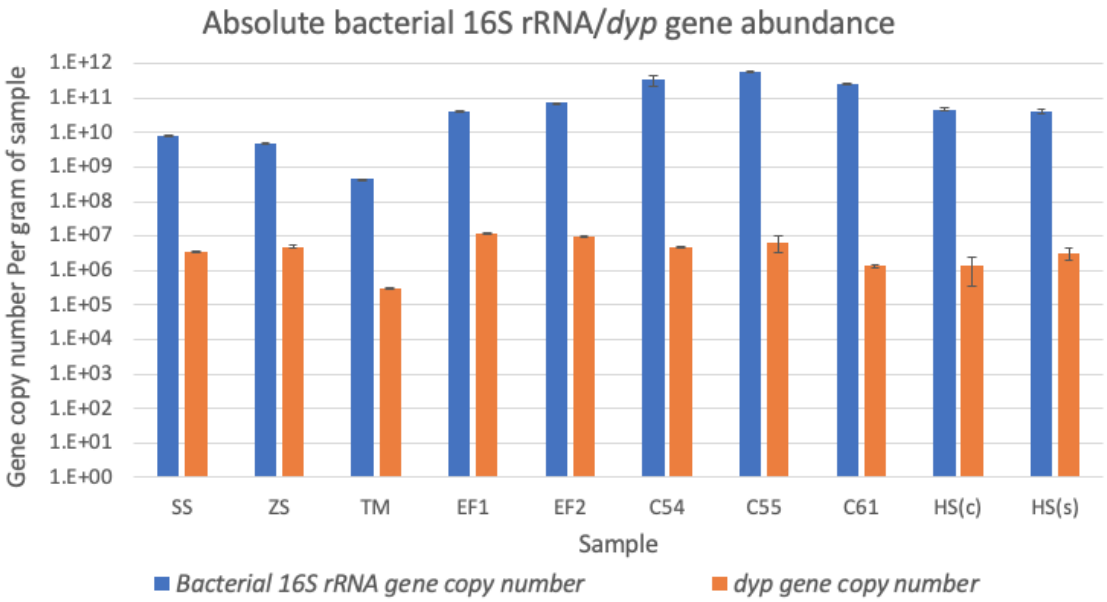

Figure 32 *- Absolute bacterial 16S rRNA gene and dyp gene abundance per gram of environmental sample (log scale). SS = Savanna Soil; ZS = Zebra-Grazed Soil; TM = Termite-Mound Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore).*

The samples showing lowest absolute 16S gene abundance were the Olduvai Gorge soils, especially within the TM sample (SS – 4.05E+09; ZS – 2.46E+09; TM – 2.29E+08). The lower gene abundances (16S and *dyp*) observed within the TM sample can likely be explained by the low concentration of nucleic acids extracted from this environment. These gene copy number values are lower than that observed by Ishii *et al* (2009), who carried out similar analysis on rice paddy soil samples and found bacterial 16S rRNA gene copy numbers to be around 5.33E+11 to 1.1E+12 in abundance. However, Ishii *et al* (2009) used different qPCR primers and different soil samples to those used in our study (similar analysis on Olduvai Gorge soils could not be found). The Olduvai Gorge soil samples were relatively older compared to the other samples tested in our study, which may explain the lower 16S gene abundances observed compared to similar studies.

The compost samples C54, C55 and C61 can be seen to have absolute bacterial 16S gene abundances of 1.64E+11, 2.93E+11 and 1.31E+11, respectively. A study carried out by Aigle *et al* (2021), involving bacterial 16S rRNA gene qPCR analysis of different

compost samples, found the quantity of this gene to be approximately within the range of E+10 to E+11. These results therefore support the findings of our study, despite Aigle *et al* using different compost samples to those used in our study.

The next highest samples with regards to absolute 16S rRNA gene abundance were the elephant faeces samples and Hickling sediment samples, all of which fell within the range of 1E+10 to 1E+11. Whilst little information could be found regarding bacterial 16S rRNA gene abundances within the elephant faeces environment, a study carried out by Givens *et al* (2014) which investigated the bacterial 16S rRNA gene abundance within sediment was found. Givens *et al* (2014) found that the bacterial 16S rRNA gene abundance within sediment fell between 1E+9 and 1E+10 which was slightly lower than the findings of our study. However, again, like the comparisons made with the OG samples, Givens *et al* used different primers and sediment samples to those used in our study, likely explaining these differences.

Analysing the relative abundance of *dyp* gene-carrying bacteria within the bacterial community (**Figure 33**), it can be seen that the three Olduvai Gorge samples, in particular the zebra-grazed soil sample, have the highest percentage abundance compared to other samples, followed by the elephant faeces samples. The Olduvai Gorge soil samples - SS, ZS and TM - contained relative percentage *dyp* gene abundances of 0.1%, 0.04% and 0.06%, respectively, which are higher than the other environmental samples (closest values were of EF1 with 0.03% and EF2 with 0.01%). These results suggest that the zebra-grazed soil sample, and the Olduvai Gorge soil samples in general, perhaps harbour more specialised bacterial communities for lignin degradation, as the DyP peroxidases are specialist enzymes used primarily by bacteria for lignin degradation (Colpa *et al*., 2014). As *dyp* gene abundances were observed to be similar within all samples in **Figure 32**, the lower *dyp* gene relative abundances observed in the compost samples in **Figure 33** are likely due to the large size and diversity of the microbiomes within these samples. Overall, however, these results illustrate that the *dyp* gene is present within all of the samples analysed, similar to the results obtained from PCR amplification of the *dyp* gene from environmental sample nucleic acids. These results, therefore, potentially indicate the presence of specialised lignin-degrading bacterial species within all samples.

The *dyp* primers used in this project were the same as those used by Tian *et al* (2016a), who created and used these primers to target a broad range of bacterial *dyp* genes within various environmental samples, whilst conserving specific targeting of the *dyp* gene. As these primers were designed to target a broad spectrum of *dyp* genes, it is likely that not

all *dyp* genes were amplified (and, therefore, quantified) in this study. Furthermore, as discussed above in **Section 4.2.2**, the degenerate nature of these primers led to non-specific amplification of other genes, therefore it is likely that this also had an impact of the gene quantifications obtained by qPCR shown in this section. Despite this, the PCR results using the *dyp* primers showed strong, clear bands for the *dyp* amplification products, therefore the quantifications obtained from qPCR in this study using the same primers were taken to be accurate.

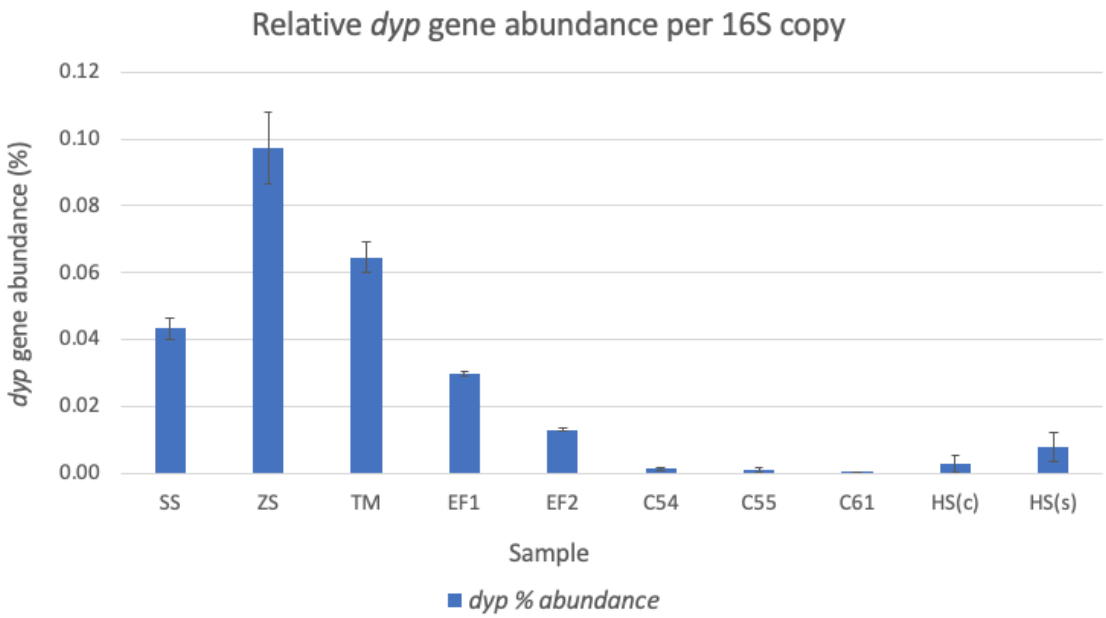

Figure 33 *- dyp gene percentage abundance relative to bacterial 16S rRNA gene copy number. SS = Savanna Soil; ZS = Zebra-Grazed Soil; TM = Termite-Mound Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore).*

4.2.4 Environmental Sample 16S/18S Amplicon Sequencing Data

4.2.4.1 Taxonomic Profiling of Amplicon Sequencing Data

In order to explore the microbial communities present within the environmental samples, their nucleic acid extracts were sent for high-throughput bacterial/archaeal 16S and eukaryotic 18S rRNA gene amplicon sequencing (Illumina MiSeq, NU-OMICS, Newcastle, UK) using 16S rRNA V4 primers (Kozich *et al*. 2013) and the Earth Microbiome 18S rRNA primers (Euk1391f-EukBr) (Amaral-Zettler *et al*., 2009; Stoeck *et al*., 2010), respectively. Both the 16S and 18S rRNA gene amplicon sequencing datasets were processed and analysed using QIIME 2 (Bolyen *et al*., 2019). The final

demultiplexed, trimmed and denoised amplicon sequencing variants (ASVs) were used as a final dataset for downstream analysis. The final ASV frequencies within each sample can be found below in **Table 12**. Whilst most samples can be seen to have ASV counts above 30,000 features, the HS samples had significantly lower counts than the rest (<10,000 features in both). Microbial community profiles for each environmental sample were constructed within Microsoft Excel from the taxonomic abundance information provided by QIIME 2, consisting of percentage relative abundance of microbial groups at various taxonomic levels (with the ASV counts in **Table 12** as '100%' abundance). The results of this analysis are presented for bacterial 16S classifications (phylum level – **Figure 34**; family level – **Figure 35**) and eukaryotic 18S classifications (order level – **Figure 36**; genus level – **Figure 37**). These taxonomic levels were chosen so as to efficiently present the most prominent groups of eukyarota and bacteria/archaea within the samples.

Table 12 *– Frequency (counts) of amplicon sequencing variants (ASV) within final (demultiplexed, trimmed and denoised) environmental sample datasets.*

Sample ID	SS	ZS	TM	EF1	EF2	C54	C55	C61	HS(c)	HS(s)
Feature Counts	78,448	39,124	37,753	63,310	38,193	95,695	115,711	63,472	7,654	85

4.2.4.1.1 Prokaryotic Abundances within the Environmental Samples

Across the Olduvai Gorge soil samples, the savanna soil (SS), zebra-grazed soil (ZS) and termite mound soil (TM) were similar with regards to microbial community composition at both phylum and genus levels of bacterial taxonomy (**Figures 34 and 35**), with similar presences of bacterial phyla/genera, however taxonomic abundances varied. In the Olduvai Gorge sample SS, the most abundant phyla of bacteria were the Acidobacteria (26.9%), Planctomycetes (18%), Actinobacteria (13.6%), Chloroflexi (14.1%) and Proteobacteria (11.6%). The most abundant genera within SS were of an unidentified group of the Acidobacteria phylum, an unclassified bacterial subgroup (13.7%) and the *Pyrinomonadaceae* family (5.2%). Like the SS sample, the predominant phyla within the ZS sample were Acidobacteria (26%), Planctomycetes (22.2%) and Chloroflexi (19.5%) and Actinobacteria (9.6%). Unlike the ZS sample, SS contained higher abundance of the Archaeal phylum Thaumarchaeaota (6% vs 1.6%, respectively). This archaeal phylum was also abundant within the TM sample (8.7%), and the most abundant bacterial phyla were similar to the other samples, with Actinobacteria (39.4% - much higher than other samples), Chloroflexi (14.25%), Planctomycetes (9.1%), Acidobacteria (7.9%) and Proteobacteria (7.38%).

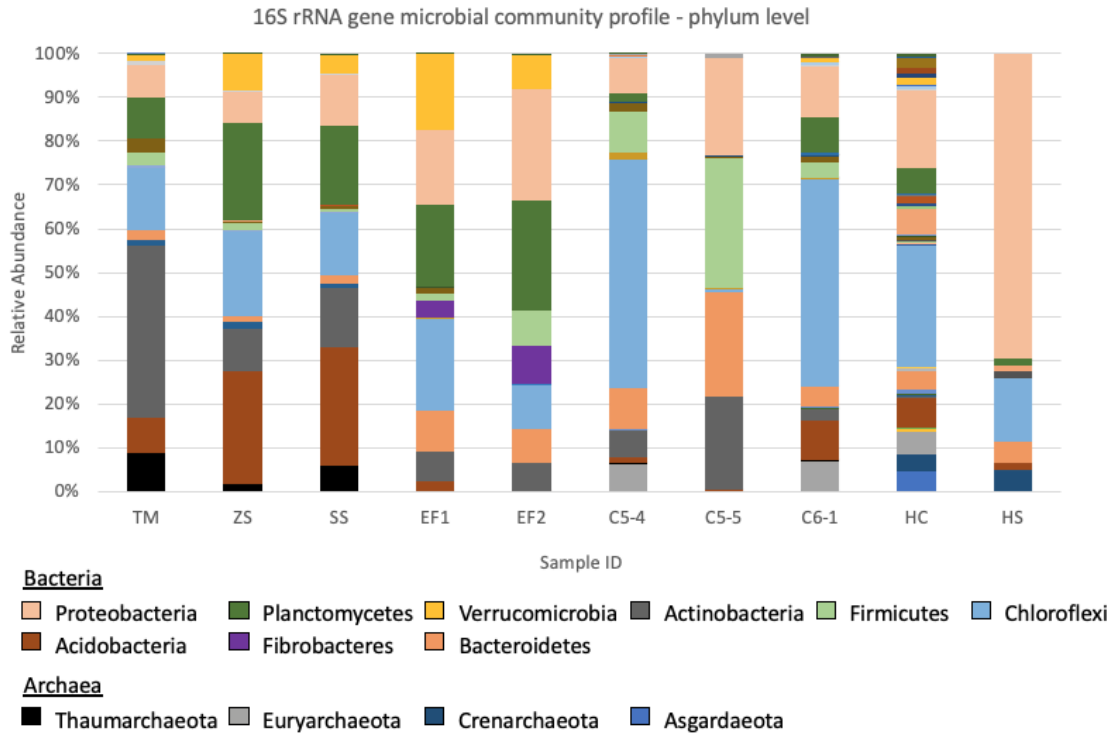

Figure 34 *- Bacterial/Archaeal 16S rRNA gene community profiles (phylum level) from Olduvai Gorge, elephant faeces, compost and Hickling sediment samples - 16S rRNA gene amplicon high-throughput sequencing (MiSeq) data. Community profiles shown on phylum level of taxonomy. TM = Termite-Mound Soil; ZS = Zebra-Grazed Soil; SS = Savanna Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore).*

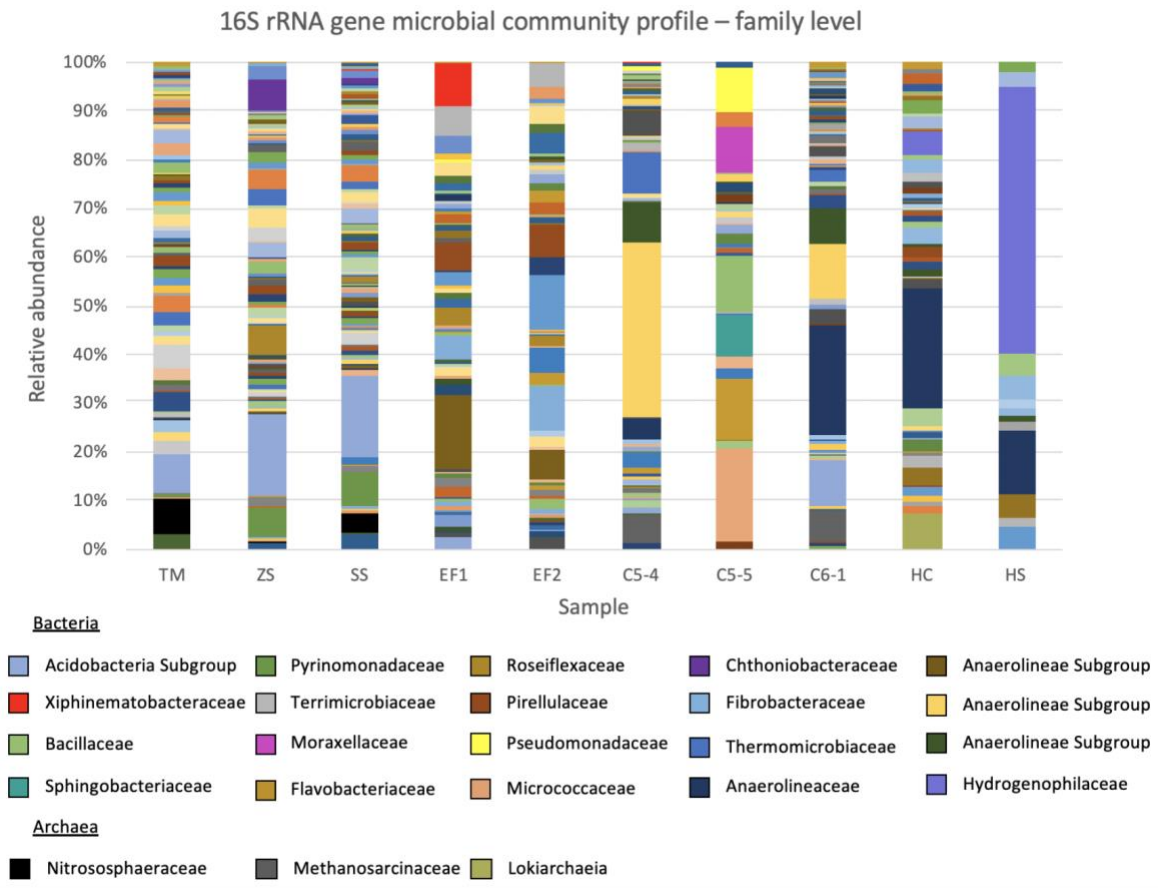

Figure 35 *- Bacterial/Archaeal 16S rRNA gene microbial community profiles (family level) from Olduvai Gorge, elephant faeces, compost and Hickling sediment samples - 16S rRNA gene amplicon high-throughput sequencing (MiSeq) data. Community profiles shown on family level of taxonomy. TM = Termite-Mound Soil; ZS = Zebra- Grazed Soil; SS = Savanna Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore).*

The similarity in community profiles between the SS and ZS samples, and their differences to sample TM, can likely be explained by their sampling source. The SS and ZS samples were similar in that they were both O-horizon soil samples taken from close proximity to vegetation, whereas the TM sample was taken from a modern termite mound. Actinobacteria, Chloroflexi, Firmicutes and Proteobacteria are common soil-inhabiting bacteria and encompass the majority of identified lignin-degrading bacteria discovered so far (Bugg *et al*., 2016). These bacterial families have also previously been shown to be prominent within termite mound soil. A review authored by Enagbonma and Babalola (2019) states that the soil surrounding termite mounds is nutrient-rich and therefore attracts a greater number and diversity of microorganisms to its vicinity. Furthermore, this review discusses the importance of microorganisms from families such as Actinobacteria, Acidobacteria and Proteobacteria in lignocellulosic decomposition, thus supporting the potential for termite mound soil as a source for industrially relevant candidates. Previous studies carried out on the gut microbiota of termite species

(*Bulbitermes* sp. and *Nasutitermes* spp.) have identified the Actinobacteria as one of the prominent phyla within this niche environment, amongst the Firmicutes and Spirochaetes (Kohler *et al*., 2012; Chew *et al*., 2018).

In both elephant faeces samples, the most prominent bacteria were of the phyla Proteobacteria (EF1 – 16%, EF2 – 25%), Chloroflexi (20%, 9%), Planctomycetes (18%, 24%), Bacteroidetes (9%, 7%) and Actinobacteria (8%, 9%). In the elephant 1 faeces sample (EF1), the most abundant bacterial groups were the Anaerolineae (Chloroflexi phylum) (17%), Spartobacteria (Chthoniobacteraceae order) (16%) and Planctomycetia (primarily Pirellulaceae family) (12%). In elephant 2 faeces (EF2), aside from higher Planctomycetia (Isosphaeraceae and Pirellulaceae families) abundance (23%), the Gammaproteobacteria (Alteromonadales order), Alphaproteobacteria (Hyphomicrobiaceae family), Bacilli (Paenibacillaceae family), Spartobacteria (Chthoniobacteraceae order) and Fibrobacteria were the most abundant classes (9%, 14%, 8%, 7% and 8%, respectively). Chthoniobacteraceae prevalence was much higher in EF1 than in EF2 (16% vs 7%), and similar level of abundance can be seen between EF2 and the Olduvai Gorge ZS sample (7% vs 6%).

A metagenomic study of elephant faecal microbiota, carried out by Ilmberger *et al* (2014) found that the most prominent microorganisms were of the phyla Bacteroidetes and Firmicutes, with Proteobacteria being more prevalent in younger elephants (approximately three weeks old). The results obtained by Ilmberger *et al* are dissimilar to the findings of this study in that our results show the Planctomycetes, Verrucomicrobia and Chloroflexi (average relative abundances of 22%, 12% and 15%; **Figure 34**) to be more prevalent than the Bacteroidetes and Firmicutes (average relative abundances of 8% and 4%, **Figure 34**). However, it is likely that the elephant faeces samples used by Ilmberger *et al* were different to those used in this study and there are not enough studies to investigate the variability of the elephant faecal microbiome. Like the Actinobacteria, many genera of the Proteobacteria and Firmicutes phyla have gained interest for their lignin degrading ability (Tian *et al*., 2014).

In the compost samples C5-4 and C6-1 the predominant bacterial classes were *Anaerolineae* (Chloroflexi phylum) (C5-4 – 41%, C6-1 – 37%), *Clostridia* (Firmicutes phylum) (7%, 2%), several classes of Proteobacteria (7%, 10%) and the archaeal class of methanogenic *Methanomicrobia* (6%, 6%). Both samples were dominated by groups of *Anaerolineae* and *Anaerolineaceae*, being the most prevalent in both samples. *Thermomicrobia* (Sphaerobacterales order) (9%), *Anaerolineae* (41%), *Clostridia* (7%) and *Flavobacteria* (*Flavobacteriaceae* family) (6%) were more abundant in C5-4 than in

C6-1, whereas the Planctomycetia (6%) and Acidobacteria (9%) were more prevalent within C6-1. Sequences from unassigned taxonomy were higher in the C6-1 sample compared to the C5-4 sample (4% vs 1%). Compost sample C5-5, on the other hand, was much less diverse and more prevalent in (in order of most abundant) Firmicutes (primarily *Bacillaceae*) (29.8%), Bacteroidetes (primarily *Flavobacterium*) (23.8%), Proteobacteria (*Pseudomonas* and *Acinetobacter* genera) (22.3%) and Actinobacteria (primarily *Glutamicibacter*) (21.6%).

Compost samples C54 and C55 were both similar in nature (being both composed of green and food waste and both being maintained in-vessel), the only difference being their stage in the composting process, with C54 being mature and C55 being sanitized. C61, on the other hand was composed of green waste only, was maintained ex-situ and was mature. Whilst compost samples C54 and C55 were similar in nature, samples C54 and C61 shared the closest similarity in regards to microbial presence and abundance, suggesting the unimportance of maintenance and composition of compost material on the resident microbial community structure. Compost can be a challenging environment for microorganisms to inhabit, with areas of anaerobic conditions and temperatures reaching up to 60 °C, which can limit microbial activity (Tuomela *et al*., 2000). This environment can therefore provide optimum growth conditions for thermophilic anaerobic organisms, such as the *Anaerolineae* prominent in all compost samples, or the thermophilic *Thermomicrobiaceae* observed in C54 (Yan *et al*., 2018). Previous studies have identified the thermophilic *Actinomycetes* (Actinobacteria) and *Bacillus* (Firmicutes) which are well known to solubilize/degrade the lignin polymer (Tuomela *et al*., 2000; Tian *et al*., 2014; Wang *et al*., 2016). The *Thermoactinomycetaceae* were present within the compost samples results of our study (**Figure 35**, not identified in legend) but at relative abundances <1%. Other examples of phyla commonly enriched in the composting process are the Proteobacteria, Chloroflexi and Bacteroidetes, shown through metagenomic analysis of the compost microbiome carried out by Wang *et al* (2016) (average relative abundances of 14%, 33% and 12%; **Figure 34**; Chloroflexi not present within C55).

The Hickling sediment centre and shore samples (HS(c) and HS(s), respectively) were very similar with regards to their community profiles, however HS(c) appeared to have higher diversity of present microorganisms than HS(s). Both samples were abundant in the bacterial class *Anaerolineae* (Chloroflexi phylum) (HS(c) - 18%, HS(s) - 17%), the class Betaproteobacteria (Proteobacteria phylum) (7%, 14%) and the archaeal phylum Crenarchaeota (7%, 7%). Within the HS(c) sample, the Planctomycetes (classes Phycisphaerae and Planctomycetia (*Pirellulales* family)) (6%) and Acidobacteria (4%)

were more predominant, whereas the Firmicutes (*Clostridiales* families *Peptococcaceae* and *Lachnospiraceae*) (5%), Bacteroidetes (Bacteroidales order) (5%) and Proteobacteria (specifically Betaproteobacteria) (22%) were more prevalent within HS(s). The Proteobacteria families varied between each sample, with HS(c) having more Deltaproteobacteria (primarily *Desulfobacteraceae* and *Desulfarculaceae* families) (4% vs 1%) and HS(s) having more Betaproteobacteria (*Alcaligenaceae*, *Comamoadaceae* and *Hydrogenophilaceae* families) (14% vs 7%) and Gammaproteobacteria (primarily *Chromatiaceae* family) (5% vs 4%).

4.2.4.1.2 Eukaryotic Abundances within the Environmental Samples

As for eukaryotic abundances within the OG samples (**Figures 36 and 37**), the SS and ZS soil samples were similar at order level, with the most prominent groups being Fungi (48.7% and 33.5%, respectively), Metazoa (primarily *Dorylaimida* genus) (12.8% and 13.1%, respectively) and Ciliophora (11.1% and 13.1%, respectively). Most of the fungal content of these samples were only identified to the taxonomic level of phylum, however the fungal presence within SS was partly composed of the *Aspergillaceae* family (8.8%) These samples were also prominent in Cercozoa (6.6% and 6.2%, respectively) and unidentified Eukaryota (7.4% and 9.1%, respectively). The ZS sample, however, contained relatively higher abundance of Charophyta than SS (7.8% vs 2.3%). Within the TM sample, compared to the SS and ZS samples, fungi were less abundant (19.6%), however, groups such as the unidentified Eukaryota (16.5%) and Opisthokonta (8.6%) were more prevalent, in addition to the higher abundance of unidentified content (20.6%).

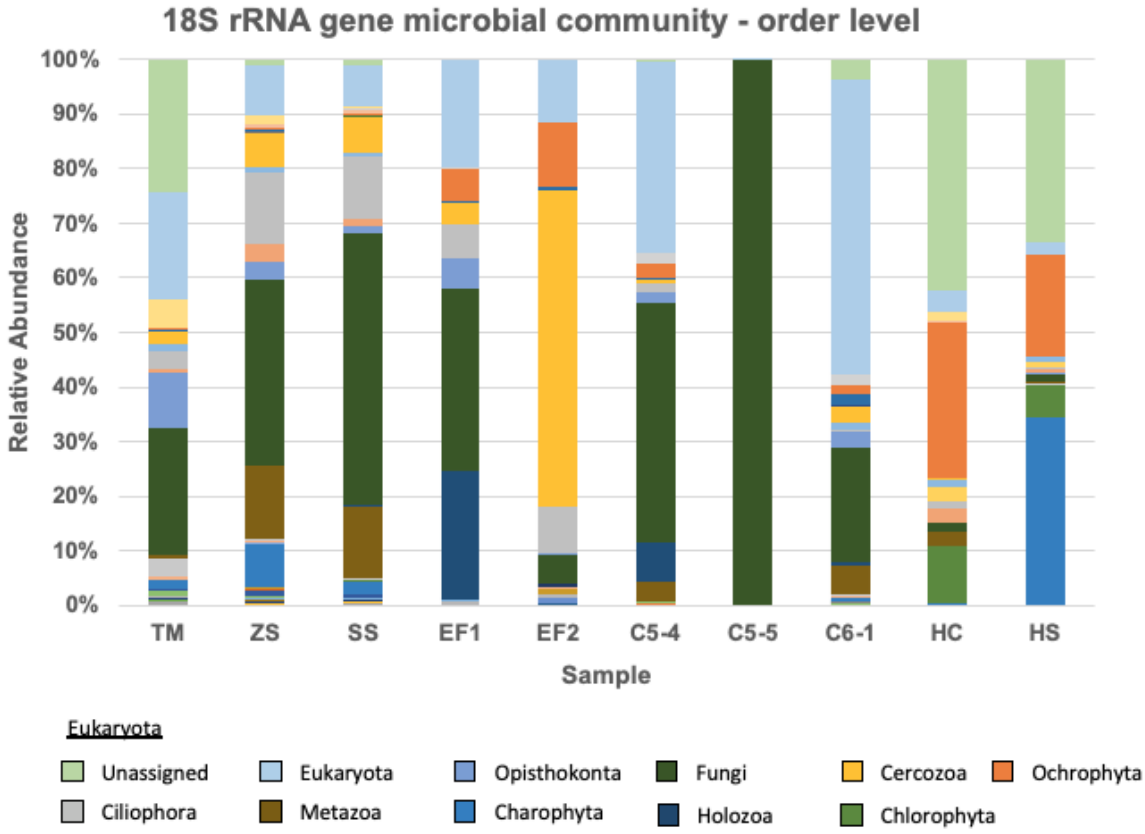

Figure 36 *- Eukaryotic 18S rRNA gene microbial community profiles (order level) from Olduvai Gorge, elephant faeces, compost and Hickling sediment samples - 18S rRNA gene amplicon high-throughput sequencing (MiSeq) data. Community profiles shown on order level of taxonomy. TM = Termite-Mound Soil; ZS = Zebra-Grazed Soil; SS = Savanna Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore).*

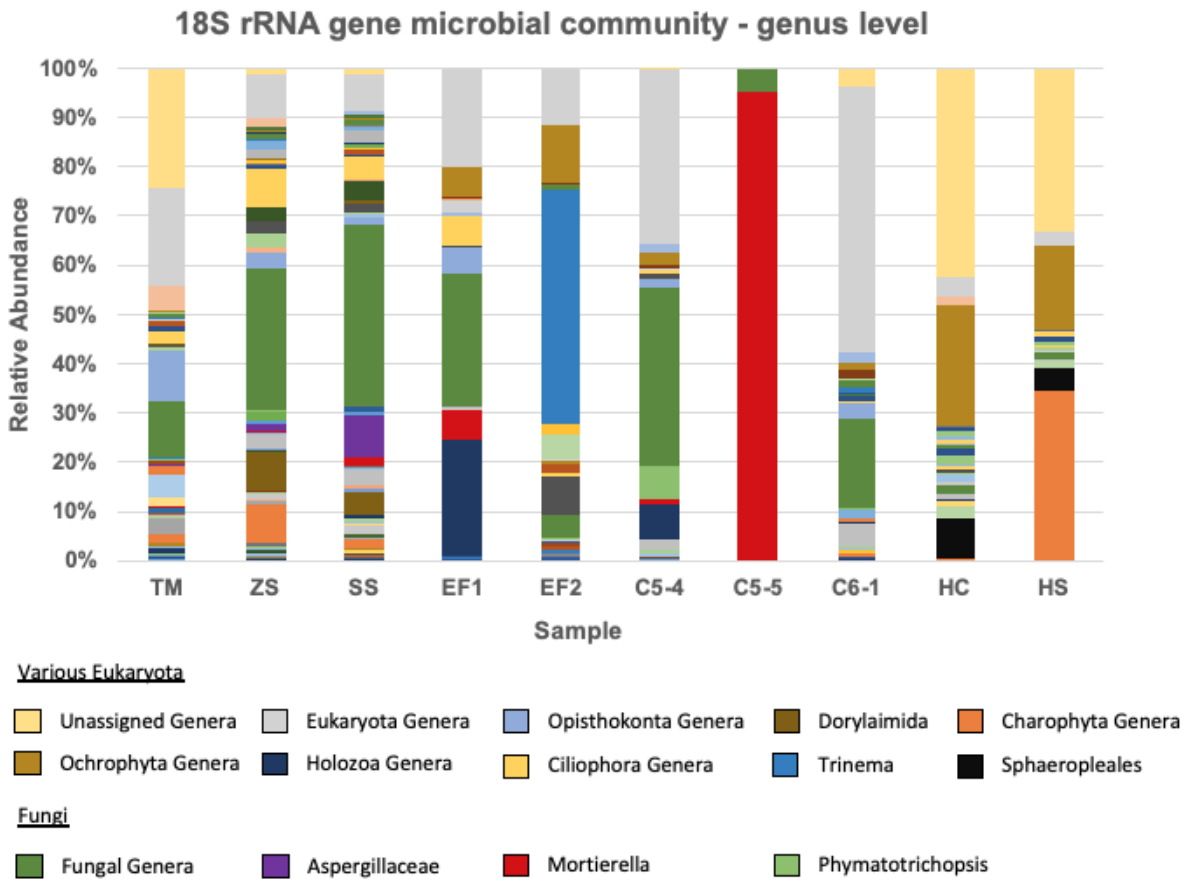

Figure 37 *- Eukaryotic 18S rRNA gene microbial community profiles (genus level) from Olduvai Gorge, elephant faeces, compost and Hickling sediment samples - 18S rRNA gene amplicon high-throughput sequencing (MiSeq) data. Community profiles shown on order level of taxonomy. TM = Termite-Mound Soil; ZS = Zebra-Grazed Soil; SS = Savanna Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore).*

Within the EF samples, both samples varied drastically with regards to Eukaryotic abundances, with EF1 being more prominent in Fungi (33.3%), Holozoa (23.3%) and unidentified Eukaryota (19.8%), whereas EF2 was more prominent in Cercozoa (primarily *Trinema* genus) (57.9%), Ochrophyta (11.6%) and Ciliophora (primarily *Hypotrichia* family) (8.5%).

Upon analysis of the eukaryotic community profiles of the compost samples, both C54 and C61 were found to be dominated by an unidentified order and genus of Eukaryota (30.8% and 47.4%, respectively) and Fungi (38.7% and 18.5%, respectively), whereas C55 was completely dominated by Fungi (97.9%) of the *Mortierella* genus. Whilst the fungal genus *Mortierella* have previously been shown to be present within compost environments (at low abundances), the almost complete domination of compost sample C55 by this genus was unexpected (**Figure 37**). The fact that C55 had undergone the sanitization step – involving high temperatures (>50 ºC) caused by microbial metabolism,

which likely limited many fungal competitors from proliferation, due to the stringent proliferation conditions of fungi - could have limited competition from other microorganisms (Tuomela *et al*., 2000).

4.2.4.2 Beta-Diversity Analysis of Bacterial 16S rRNA Gene Amplicon Sequencing Data

The bacterial community structures of the environmental samples were compared by principal coordinate analysis (PCoA) using the beta-diversity statistical analysis method – Unifrac distances. When carrying out diversity analyses with QIIME 2, a sampling depth value is used which is a parameter based on the desired rarefaction depth. A sampling depth values of 1103 was used in the analysis shown. This value must be, as stated by QIIME 2, as high as possible whilst retaining as many samples as possible. For this reason, environmental sample HS(c) (HC) was not included in this analysis as the sampling depth value used was too high to include it. Several sampling depth values were tested, however the sampling depth value required to include HS(s) (HS) (sampling depth value = ~50) was too low and therefore compromised the results of other samples. This can likely be explained by the low ASV counts for this sample seen in **Table 12** above.

Both unweighted (**Figure 38**) and weighted (**Figure 39**) Unifrac distances were explored at amplicon sequencing variant (ASV) level, in order to determine the primary driving factors within the microbial communities present, more specifically whether species abundance impacted diversity or not. The unweighted Unifrac (ignoring abundance, taking only species presence/absence into account) PCoA results, illustrated in **Figure 38**, showed clear clustering of samples within their respective sample sets, indicating sample similarity (i.e. distance between data points reflects sample relatedness). The EF samples (EF1 and EF2) were the best example of this, clustering closely together. Some minor differences could be observed between the EF samples in the community profiles presented above (**Figures 34-37**), however both samples appeared similar in taxa presence and abundance, reflecting the results of beta diversity analysis. The OG (SS, ZS and TM) and CS (C54, C55 and C61) sample sets were, on the other hand, more dissimilar between samples. Within the OG sample set, samples SS and ZS both clustered close together, however the TM sample clustered further away from the other two samples, indicating difference in community diversity with regards to species presence (reflected within the community profiles above – **Figures 34-37**). Within the CS sample set, the samples are clustered, however the samples seem to vary from one another more significantly than the other sample sets, with samples C54 and C55 being

closer together (correlating with their similarity) but sample C61 being further away. This similarity in C54 and C55 was not observed within the community profiles shown above (**Figures 34-37**), as C55 was observed to be different to both C54 and C61 in taxa presence and abundance. Sample HC can be seen as an outlier aside from the other data points, indicating its dissimilarity from the other samples. The unweighted Unifrac PCoA analysis of the environmental samples showed that microbial communities were similar within the sample sets, but that the communities between sample type (soil, faeces, compost and sediment) were distinct. Furthermore, the TM sample microbial community was much more distinct from other samples of the OG sample, and more so was observed for samples within the compost samples.

The results of weighted Unifrac PCoA analysis (**Figure 38**) (accounting for the relative abundance of species) showed similar results to that of the unweighted Unifrac PCoA analysis, with clustering of samples within their sample sets. The first three axes of the weighted Unifrac PCoA explained 75% of the variance, whilst for the unweighted Unifrac only 56% was explained. The weighted Unifrac PCoA results showed clustering of data-sets, with the EF samples clustering closely together and similar distances between the OG and CS sample sets observed in the unweighted Unifrac distances. Generally, however, distances were larger between data points within the data-sets compared to the unweighted results, likely due to the species abundances creating greater difference between samples. The HC sample, however, appeared to cluster closely with CS sample C61, which was not observed to such an extent within the unweighted Unifrac results.

From the data presented in these figures, the unweighted Unifrac metric (ignoring abundance) illustrates closer clustering of samples to that of the weighted Unifrac PCoA results (accounting for abundance), the clusters of which were much less defined. As the clusters of data-sets (OG/EF/CS) remain the same between the two plots, this indicates that abundance was not an important driver in community diversity between the sample-sets and that diversity was influenced more by species presence/absence. However, the larger distances observed between individual samples within data-set clusters in the weighted Unifrac test suggests that abundance was a driving factor within these data-sets. In both figures, the similarity in SS and ZS community diversity (in addition to the similarity between EF samples) can likely be explained by similarity in their origin – both SS and ZS were O-horizon savanna soils. The distance of TM from the other OG samples can likely be explained by TM coming from a termite mound, and likely containing microorganisms specific to that microbiome. Sample origin can also likely explain the differences observed between compost samples, with each being composed of varying source material, differing in maintenance method and maturity. Compost

sample C54 and C55 can be seen to cluster closer together than C61 likely due to C54 and C55 being composed of similar material (green and food waste) and both being maintained in-vessel, the only difference being that C54 was mature and C55 was sanitized. Sample C61, on the other hand, was composed of green waste and maintained ex-situ, both factors likely having an impact on microbial diversity within the sample and making it deviate from the others. However, as mentioned above, the taxonomic profiles of these samples showed different results, with C54 being more similar to C61 and C55 being dramatically different to both of these samples. This could perhaps be due to the presence of specific taxa being similar between C54 and C55 (therefore clustering closer together in the unweighted Unifrac results) however the abundances of these taxa being very different (therefore clustering further apart in the weighted Unifrac results).

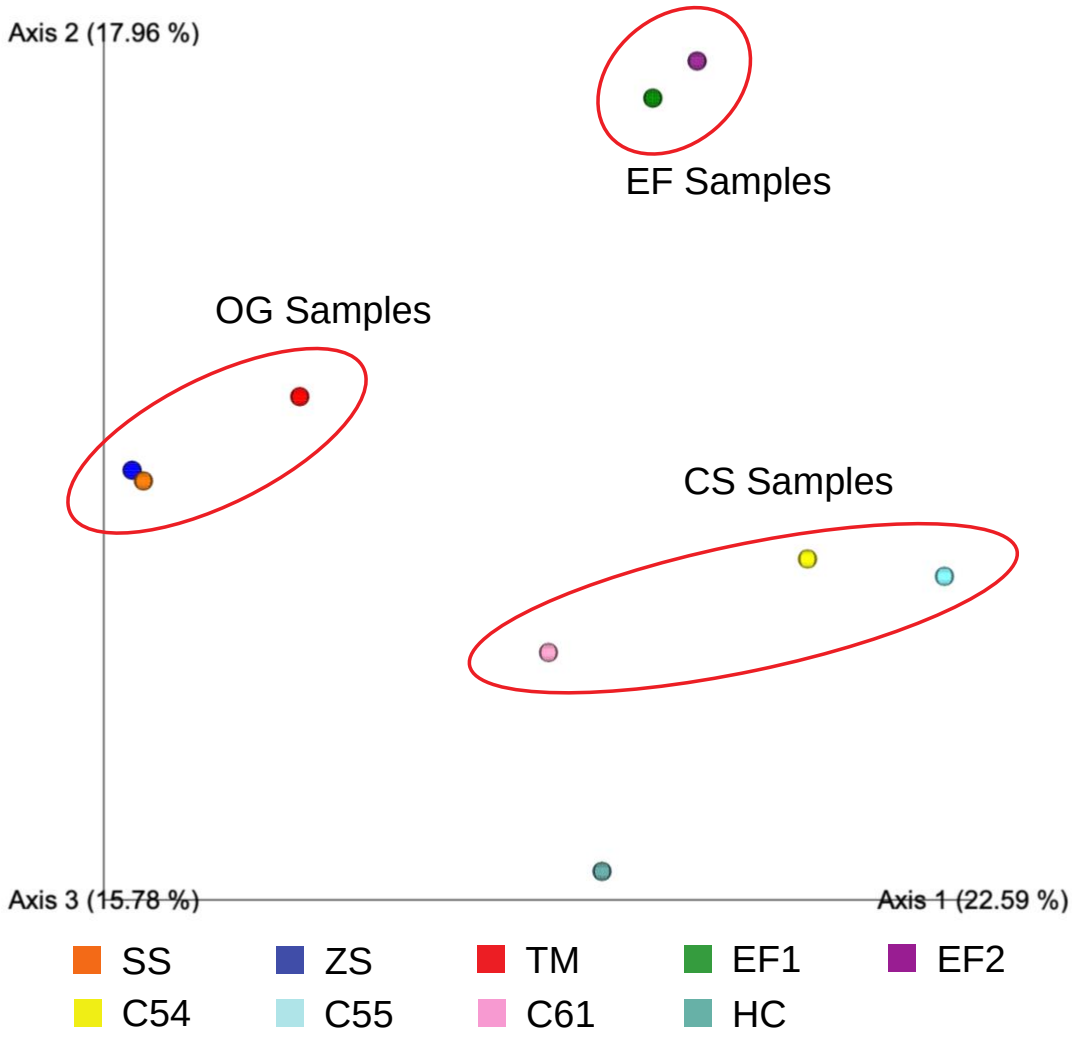

Figure 38 *- Principal coordinate analysis (PCoA) of the microbial community structures from environmental samples using* ***unweighted*** *Unifrac distances. Axes 1, 2 and 3 describe approximately 56% of variance between samples. Sample sets and clustering indicated by red circles. Sample type indicated by colour of data point: OG samples - SS in orange, ZS in dark blue, TM in red; EF samples – EF1 in green, EF2 in purple; CS samples – C54 in yellow, C55 in light blue, C61 in pink; HS samples – HC in teal blue. SS = Savanna Soil; ZS = Zebra-Grazed Soil; TM = Termite-Mound Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1; HC = Hickling Sediment (Centre Broad).*

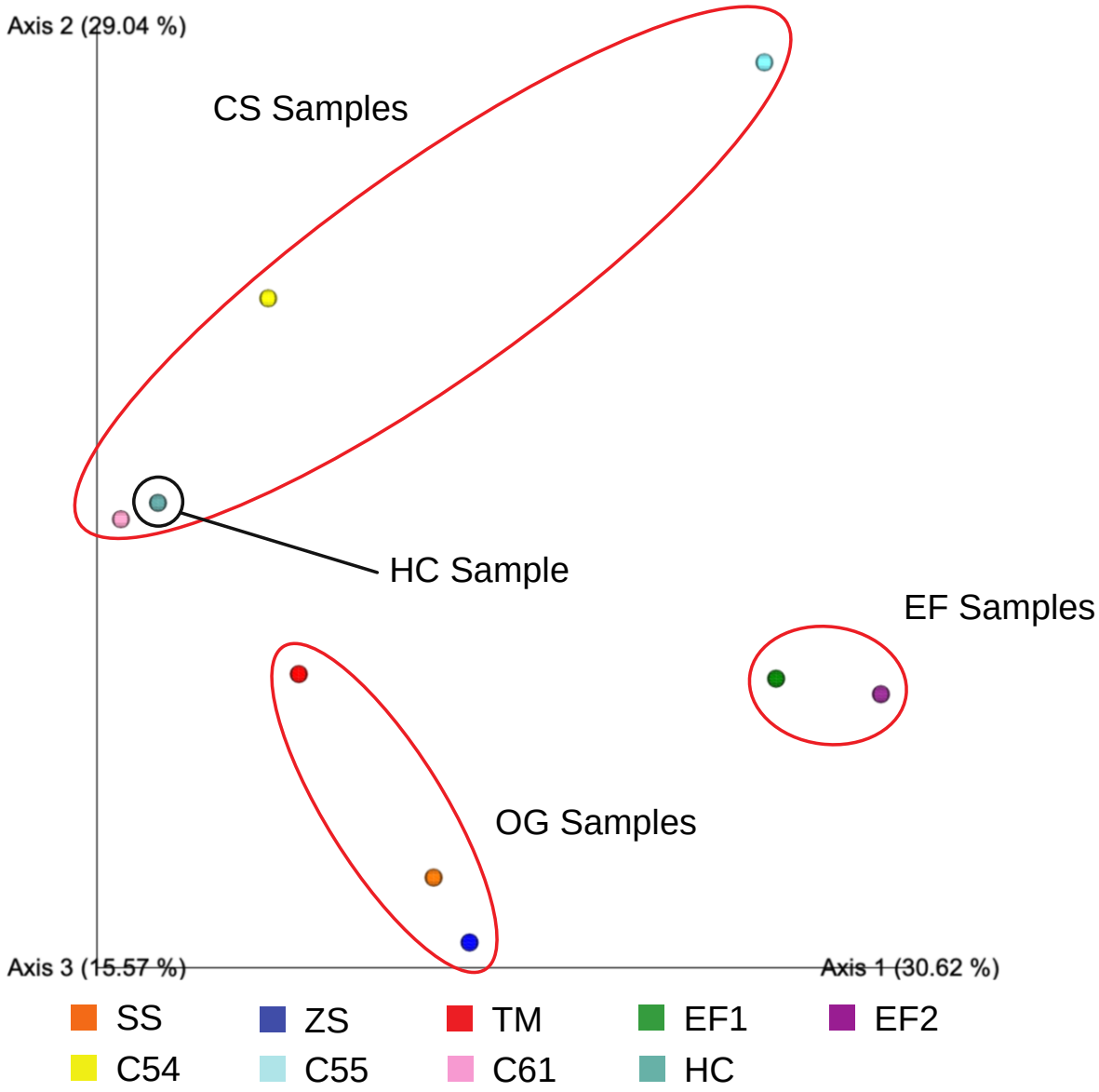

Figure 39 *- Principal coordinate analysis (PCoA) of the microbial community structures from environmental samples using* ***weighted*** *Unifrac distances. Axes 1, 2 and 3 describe approximately 75% of variance between samples. Sample sets and clustering indicated by red circles (HC sample indicated by black circle). Sample type indicated by colour of data point: OG samples - SS in orange, ZS in dark blue, TM in red; EF samples – EF1 in green, EF2 in purple; CS samples – C54 in yellow, C55 in light blue, C61 in pink; HS samples – HC in teal blue. SS = Savanna Soil; ZS = Zebra-Grazed Soil; TM = Termite-Mound Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1; HC = Hickling Sediment (Centre Broad).*

4.2.5 Environmental Sample Metagenomic Sequencing Data

Environmental sample nucleic acid extracts, following RNase digestion and gDNA purification (Zymo DNA clean-up), were sent for metagenomic sequencing (Edinburgh Genomics, Edinburgh, UK) in duplicate (on two 96-well plates, both sequenced on Illumina NovaSeq 6000, paired-end reads, 150 bp length) from which sequencing data

was obtained in two batches (batch 1 and 2). Furthermore, nucleic acid extracts from anoxic elephant faeces (E1-12-A1 and E2-12-A1) ^{12}C SIP incubation samples (see **Section 4.2.6** below) underwent RNA purification, reverse transcription into double stranded cDNA and were sent for metatranscriptomic sequencing (both cDNA samples pooled as individual samples were below nucleic acid concentration requirements). These samples were included due to space being available on a colleagues metatranscriptomic sequencing plate. Due to time limitation, however, this metatranscriptome was not analysed in this study.

Whilst most samples passed library preparation for batch 2, the Hickling sediment (shore) sample did not. Whilst the reason for this is unknown to the author, this sample also displayed low final ASV counts in the amplicon sequencing results shown above (**Table 12**), perhaps due to the same reason. Furthermore, whilst data was obtained for compost sample C55, Edinburgh Genomics reported that compost sample C61 failed (due to the plate well being empty), therefore it is thought that C61 might have been mistakenly added to the C55 well during plate preparation. Information on read and base counts from batch 1 and 2 of the metagenomics sequencing data can be found in **Table 13**.

Table 13 *- Read count and base count information for batches 1 and 2 of metagenomic sequencing data (Illumina NovaSeq 6000, paired-end reads, 150 bp length) obtained from Edinburgh Genomics (Edinburgh, UK). EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore); SS = Savanna Soil; ZS = Zebra-Grazed Soil; TM = Termite-Mound Soil; C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1; EF metatranscriptome = pooled EF1 and EF2 anoxic ^{12}C SIP incubation samples.*

	Batch 1		Batch 2		
Sample	**Total Reads**	**Total Bases**	**Total Reads**	**Total Bases**	**Notes**
SS	54,128,091	8,119,213,650	94,846,084	14,226,912,600	
ZS	52,063,931	7,809,589,650	120,693,915	18,104,087,250	
TM	46,445,650	6,966,847,500			Not included in batch 2 sequencing plate
EF1	53,614,023	8,042,103,450	93,793,818	29,069,072,700	
EF2	59,260,562	8,889,084,300	123,183,268	18,477,490,200	
C54	50,116,761	7,517,514,150	133,612,153	20,041,822,950	
C55	56,475,031	8,471,254,650	285,105,855	42,765,878,250	C55 and C61 may have been combined in batch 2 sequencing
C61	62,434,238	9,365,135,700			
HS(c)	51,619,988	7,742,998,200			Failed library prep. In batch 2 sequencing
HS(s)	50,444,231	7,566,634,650	93,600,327	14,040,049,050	Sequenced twice in batch 2 data, read and base counts combined from both runs.
EF metatranscriptome			113,295,669	16,994,350,350	Sequenced twice in batch 2 data, read and base counts combined from both runs.

Metagenomic data (batch 1 and 2) underwent processing and analysis using various bioinformatic tools on an instance within the Cloud Infrastructure for Microbial Bioinformatics (CLIMB) (www.climb.ac.uk) (according to methodologies detailed in the **Methods and Materials Section 2.7.2**). The two batches of metagenomic sequencing data were duplicates and were, therefore, co-assembled in order to obtain more comprehensive assemblies allowing better annotation and analysis. The trimmed reads and assemblies obtained from batch 1 and 2 data were used as final datasets for downstream analysis (e.g. DAS_tool binning and CAZy gene annotation pipeline). Information on Megahit assemblies obtained from environmental sample metagenomes can be found below in **Table 14**, including number of contigs within assembly, total length of assembly in base pairs (bp) and N50 values. The N50 value is a length-weighted average statistic which, in short, refers to the smallest contig size (in bp) required to cover 50% of the metagenome, i.e. 50% of the assembly is composed of contigs equal to or higher than the N50 value, therefore providing an insight into quality value of assembled contigs (Goicoechea *et al*., 2010). For the samples that did not have batch 2 sequencing data, or those that could not be done due to time limitation, metagenomic assemblies were generated using batch 1 data only (highlighted as red data series in **Table 14** below).

Table 14 *– Environmental sample Megahit assembly information including: assembly size (in number of contigs and total base pairs (bp)); minimum, maximum and average contig size (in bp) and N50 values. Data series highlighted in green represent assemblies composed of batch 1 and 2 sequencing data, whereas data series highlighted in red represent assemblies composed of batch 1 data only (due to lack of batch 2 data). SS = Savanna Soil; ZS = Zebra-Grazed Soil; TM = Termite-Mound Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore)*

	Megahit					
Sample	**Contigs**	**Total bp**	**Min. bp**	**Max. bp**	**Avg. bp**	**N50**
SS	5.97E+06	3.46E+09	2.00E+02	1.52E+05	5.79E+02	6.10E+02
ZS	7.35E+06	4.52E+09	2.00E+02	1.33E+05	6.14E+02	6.52E+02
TM	2.87E+06	1.57E+09	2.00E+02	6.64E+04	5.46E+02	5.67E+02
EF1	3.92E+06	3.81E+09	2.00E+02	1.68E+06	9.71E+02	1.54E+03
EF2	2.13E+06	2.28E+09	2.00E+02	1.45E+06	1.07E+03	1.93E+03
C54	2.54E+06	2.64E+09	2.00E+02	1.31E+06	1.04E+03	1.83E+03
C55	9.88E+05	8.07E+08	2.00E+02	3.20E+05	8.16E+02	1.26E+03
C61	2.36E+06	1.98E+09	2.00E+02	4.15E+05	8.39E+02	1.06E+03
HS(c)	2.77E+06	1.66E+09	2.00E+02	1.89E+05	5.99E+02	6.12E+02
HS(s)	2.62E+06	1.55E+09	2.00E+02	1.43E+05	5.93E+02	6.04E+02

4.2.5.1 Taxonomic Profiling of Environment Metagenomes

The Kraken2 taxonomic sequence classification system (Wood *et al*., 2019) was used to generate taxonomic profiles of the environmental sample metagenomes (trimmed reads) (as described in the **Methods and Materials Section 2.7.2.2**). As the full Kraken2 database for sequence classification was very large, a smaller pre-formatted database was used for taxonomic assignment (16 GB PlusPFP database – standard database including Protazoa, Fungi and Plant references – details of which can be found at https://benlangmead.github.io/aws-indexes/k2). Relative abundances of taxonomic groups were calculated relative to total classified reads at both phylum (**Figure 40**) and family (**Figure 41**) levels of taxonomy.

For all datasets, only 4.02%-35.6% of the reads per metagenome were classified by kraken2 (shown in **Table 15**). This low number is likely due to the fact that most reads would not have been reliably assigned to a specific genome because a lot of diversity was not represented in the reference database for this tool in contrast to the SSU databases (Lu and Salzberg, 2020). Of the low number of classified reads, at phylum level of taxonomy (**Figure 40**), all samples were shown to be mostly classified as Proteobacteria (20-80%) and Actinobacteria (15-70%) phyla, with low abundances of Firmicutes (2-10%). In addition, the compost samples C54 and C61 contained Chloroflexi (12% and 2%, respectively) and Euryarchaeota (5%), whereas C55 contained Bacteroidetes (6%). The Hickling sediment samples (HS and HC) both contained low abundances of Cyanobacteria (2%) and elephant faeces sample EF2 contained low abundances of Planctomycetes (2%). Comparing these results to that of microbial SSU rRNA gene amplicon sequencing of these samples (**Figures 34-37** above), in general, similar patterns of phylum presence can be observed, however relative abundances differ. For example, amplicon sequencing results obtained for the compost samples showed significantly higher Chloroflexi and Bacteroidetes abundances, and significantly lower Proteobacteria and Actinobacteria abundances, however the absence of Chloroflexi within C55 is mirrored within the metagenome Kraken2 taxonomic profile. These differences arose from the reasons mentioned above, with SSU amplicon sequencing being the more reliable taxonomic classification method with better resolution. Although PCR bias can also have an impact on amplicon sequencing data.

Table 15 *– Kraken2 percentage classified reads within environmental samples. SS = Savanna Soil; ZS = Zebra-Grazed Soil; TM = Termite-Mound Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore).*

Sample ID	SS	ZS	TM	EF1	EF2	C54	C55	C61	HS(c)	HS(s)
% Classified Reads	9.13	9.14	8.27	9.74	12.38	14.55	35.60	8.82	4.05	4.02

At the family level (**Figure 41**), the metagenome Kraken2 taxonomic profiles between samples were much more varied. The compost samples C54 and C61 contained a more mixed community of microorganisms (such as *Streptomycetaceae* and *Microbacteriaceae* in C54, and *Burkholderiaceae* and *Comamonadaceae* in C61) compared to C55, which is dominated by the *Micrococcaceae* followed by *Pseudomonadaceae*. Both elephant faeces samples contain similar patterns of taxonomic abundances, with high relative abundances of *Microbacteriaceae*, *Mycobacteriaceae* and *Bradyrhizobiaceae*. Like the EF samples, the Hickling sediment samples also have similar taxonomic profiles, containing primarily *Comamonadaceae* and *Burkholderiaceae* families of bacteria. The SS and ZS samples from the Olduvai Gorge soils are more abundant in *Bradyrhizobiaceae* and *Micromonosporaceae* compared to the TM sample, which contains higher abundances of *Nocardioidaceae* and *Streptomycetaceae*. Like the taxonomic profiles at phylum level of taxonomy, similar discrepancies at family level can be observed to that of the amplicon sequencing results (**Figure 34-37** above). Using the compost samples as an example again, the Kraken2 profiles show C55 to contain a high abundance of *Micrococcaceae* (~65%) and *Pseudomonadaceae* (~12%). Whilst these microbial groups are also abundant within the amplicon sequencing profile for this sample, the relative abundance of *Micrococcaceae* is significantly lower, in addition to the previous profile containing other abundant families not observed in the Kraken2 profiles (such as the *Flavobacteriaceae* and *Bacillaceae*). This is likely due to the lack of reference information contained within the Kraken2 database compared to that used in analysis of the amplicon sequencing data, resulting in a lack of diversity and amplifying abundance values of identified taxonomic groups in the Kraken2 profiles. As mentioned above, only 4.02%-35.6% of the reads per metagenome were classified, therefore indicating a significant lack of identified lineages for which no kmer-based identification was possible. Despite this lack in kmer-based identification, the Kraken2 profiles aided in confirming the microbial lineages classified in the amplicon sequencing taxonomic profiles obtained via the 16S rRNA framework.

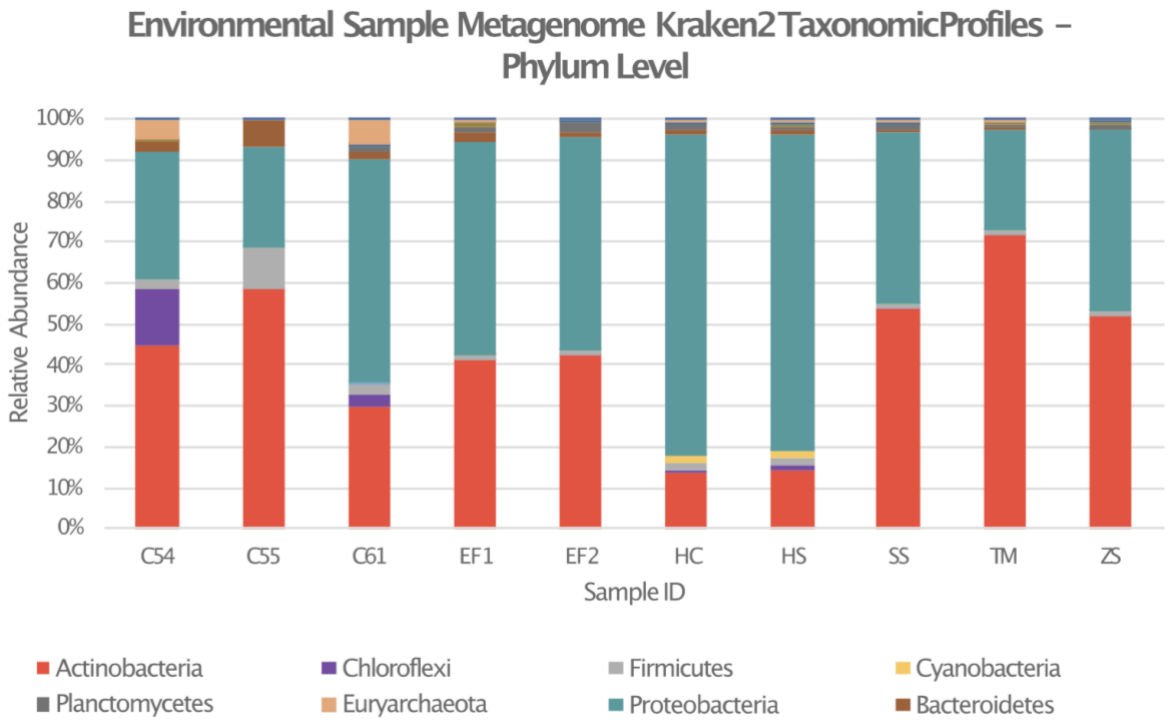

Figure 40 *- Metagenome Kraken2 taxonomic profiles (phylum level) of classified reads from Olduvai Gorge, elephant faeces, compost and Hickling sediment samples. Abundance of taxonomic groups calculated relative to total classified reads. C54 = Compost Soil 5-4; C55 = Compost Soil 5-5; C61 = Compost Soil 6-1; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore); SS = Savanna Soil; TM = Termite-Mound Soil; ZS = Zebra-Grazed Soil.*

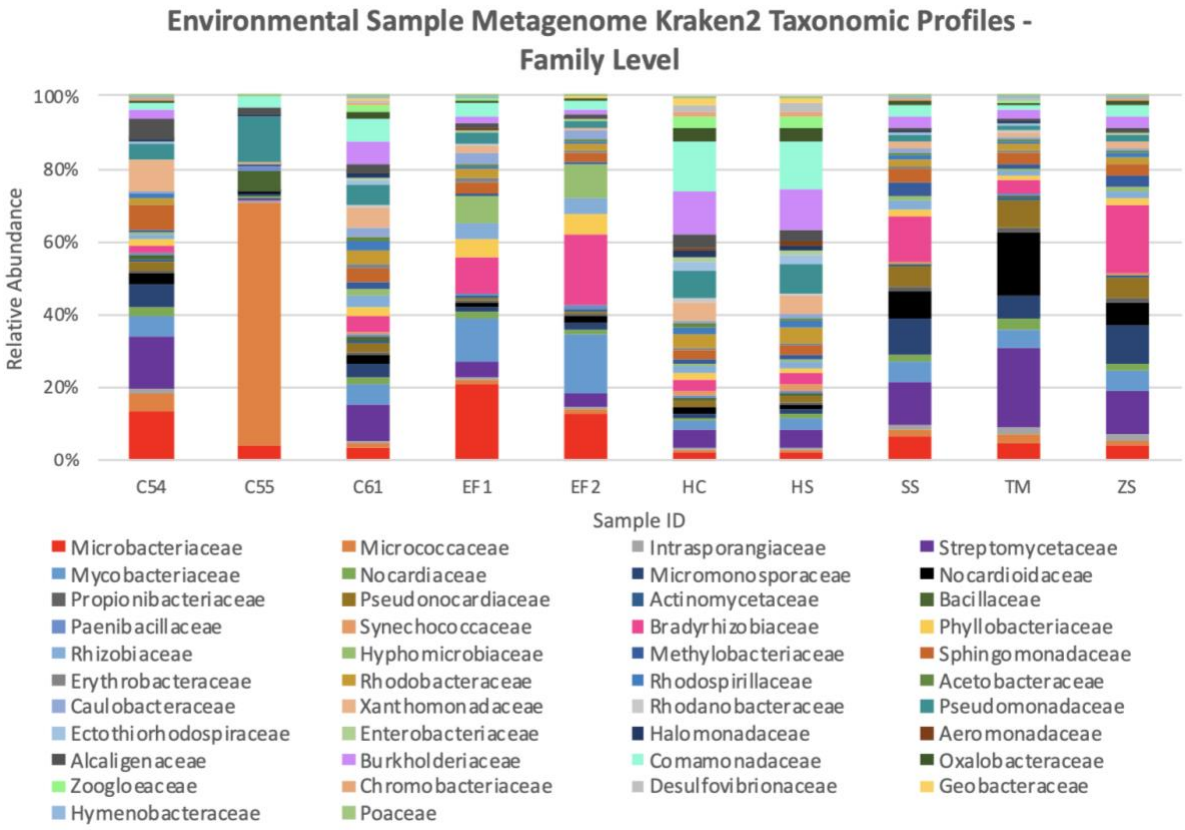

Figure 41 *- Metagenome Kraken2 taxonomic profiles (Family level) of classified reads) from Olduvai Gorge, elephant faeces, compost and Hickling sediment samples. Abundance of taxonomic groups calculated relative to total classified reads (>0.001%). C54 = Compost Soil 5-4; C55 = Compost Soil 5-5; C61 = Compost Soil 6-1; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore); SS = Savanna Soil; TM = Termite-Mound Soil; ZS = Zebra-Grazed Soil.*

4.2.5.2 CAZy Functional Profiling of Environment Metagenomes

A bioinformatic pipeline for identification and quantification of genes encoding Carbohydrate-Active enZYmes (CAZymes) was set up and used according to the methodologies used by Wilhelm *et al* (2018) (detailed in the **Methods and Materials Section 2.7.2.3**). CAZymes encompass microbial enzymes active primarily against lignocellulosic polysaccharides (such as cellulose and xylan), as well as lignin active enzymes (such as laccases and peroxidases). A database composed of these CAZymes was therefore prepared and the environmental sample metagenomes were blasted (DIAMOND BLASTx - minimum e-value of 0.0001) against it in order to identify genes present for lignin degradation and microorganisms containing them. The resulting DIAMOND blastx results were fed into MEGAN6 (community edition) (Huson *et al*., 2016) for taxonomic and functional analysis. The EC and InterPro2Go databases (included in MEGAN6) were used for extracting the lignocellulose degrading enzyme hits. This

analysis was carried out with metagenomic assemblies, as well as metagenome assembled genomes (MAGs) (obtained from the DAS_Tool binning pipeline, detailed below), to provide important information on presence/absence and diversity of lignin degrading mechanisms within these environments and obtain detailed information on potential key lignin degrading microorganisms within them, respectively. The CAZy functional profiles of the metagenome assemblies can be found below in **Figure 42** and **Figure 43**.

Analysis of the lignocellulose polysaccharide-active CAZy genes identified within the metagenome assemblies (**Figure 42**) revealed that the enzymes required for degrading cellulose (cellulase, cellulose 1,4-beta-cellobiosidase) and xylan (a group of hemicelluloses) (xylan 1,4-beta-xylosidase, endo-1,4-beta-xylanase, oligosaccharide reducing-end xylanase) were present within all environmental samples. As the environmental samples were derived from environments which were likely abundant in lignocellulose material, the presence of lignocellulose-degrading microbial communities and therefore enzymes were expected. Cellulases and beta-xylosidases are common enzymes expressed by microorganisms, which is especially the case in lignocellulose-rich environments (Jayasekera and Ratnayake, 2019; Bajpai, 2014). Worth noting is the absence of oligosaccharide reducing-end xylanase within OG samples TM and SS, as well as EF2. The microbial xylan degradation pathway involves the breakdown of the xylose backbone into xylan by several different enzymes, which endo-1,4-beta-xylosidases then attack to create xylo-oligomers. The xylo-oligomers are then transported into the cell to begin breakdown by additional beta-xylosidases (Erlandson *et al*., 2001). Therefore, the absence of oligosaccharide reducing-end xylanase within these samples may infer that microorganisms have the ability to degrade polymeric xylan, but are unable to metabolise the resulting oligomers.

In addition to the presence of polysaccharide-degrading enzyme genes, all environmental samples contained genes encoding lignin-active enzymes (such as peroxidases and laccase) (**Figure 43**). CAZy genes responsible for heme-dependent peroxidase, peroxidase, DyP-type peroxidase and laccase were present in every environment except EF1 and HS, which lacked the genes for heme-dependent peroxidase. CAZyme gene content of elephant faeces and composts have been investigated previously (Zhang *et al*., 2019; Wang *et al*., 2016). Zhang *et al* (2019) showed that most CAZymes in the elephant faeces microbiome were for hemicellulose degradation, compared to cellulose degradation, in addition to vanillyl alcohol oxidases required as part of lignin degradation pathways. Whilst our study did not include an analysis of genes encoding vanillyl alcohol oxidases, the results in **Figure 42** show the

presence of genes required for both cellulose (cellulase) and hemicellulose (e.g. xylan 1,4-beta-xylosidase) degradation. Furthermore, Jakeer *et al* (2020), exploring the CAZy gene content of an adult elephant faecal metagenome, also uncovered the presence of the AA2 family of enzymes, encompassing microbial peroxidases, which were present within the results shown in **Figure 43**.

With regards to compost environments, Wang *et al* (2016) showed that, whilst cellulase and hemicellulase CAZy gene content was high within the compost microbiome analysed, so too was the AA2 CAZy gene family (encompassing peroxidases), similar to the results of this study (**Figures 42 and 43**). Furthermore, Tian *et al* (2016), using primers targeting the genes encoding bacterial type-P DyP peroxidases, discovered the presence of these genes within several environments, including compost (also shown to be present in our study). Composts are, however, extremely complex environments which are in constant flux, therefore the presence/abundance of microorganisms and their ligninolytic mechanisms likely change throughout the composting process. For example, whilst fungi are important in the degradation of lignocellulose throughout the composting process, Antunes *et al* (2016) were unable to find genes coding for fungal peroxidases within thermophilic compost, suggesting that the primary degraders during this stage were bacteria. However, as our results show the presence of peroxidase genes from all microorganisms within the samples (bacteria, fungi, etc.), it could not discerned which kingdom or species of organism these peroxidase genes came from (i.e. fungal peroxidases vs. bacterial peroxidases).

The results obtained by Zhang *et al*, Jakeer *et al*, Wang *et al* and Tian *et al* mirror those illustrated in our study, however, to the authors knowledge, genes coding for DyP-type peroxidases have not previously been shown to be present within the elephant faeces microbiome. As mentioned in the **Introduction Chapter**, the DyP peroxidases are important enzymes in bacterial lignin degradation and are almost exclusively produced by bacteria. The presence of *dyp* genes shown within the elephant faeces samples in this study therefore potentially indicates (1) the presence of bacterial lignin degraders within the elephant faeces (and possibly the elephant gut) environment(s) and (2) suggests the possible importance of bacteria in the lignin-degradation process within this environment.

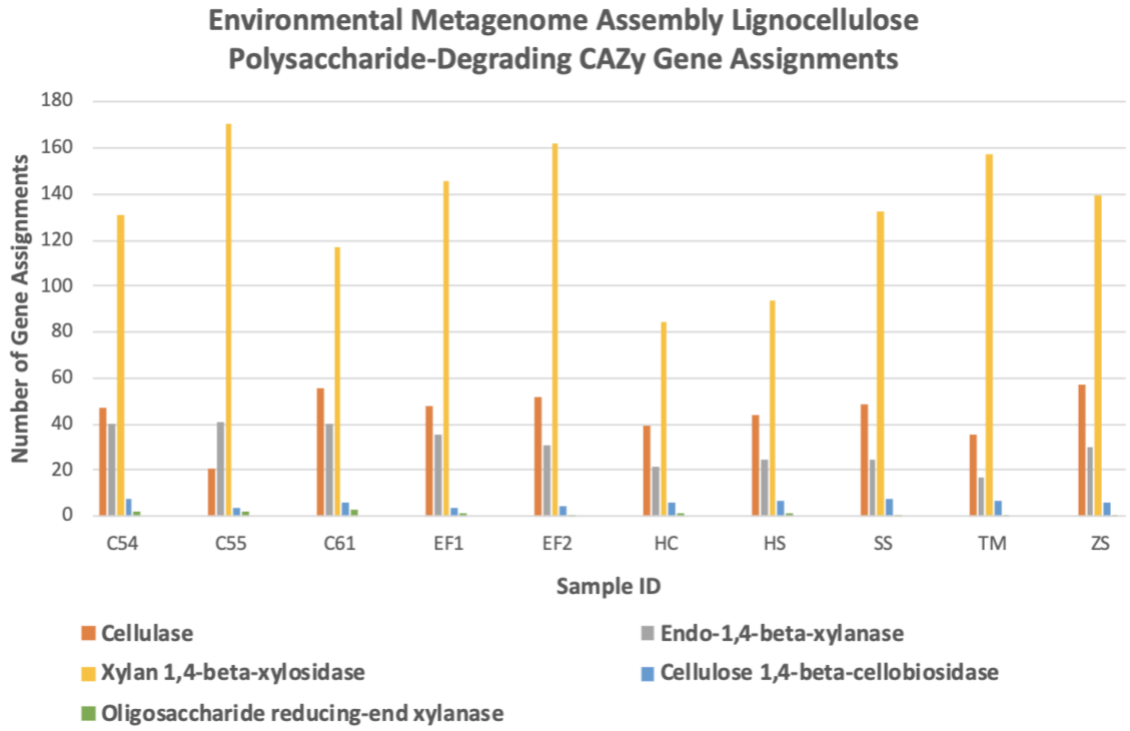

Figure 42 *- Environmental sample metagenome assemblies lignocellulose polysaccharide-active CAZyme number of gene assignments. CAZy gene assignments derived from the EC and InterPro2Go enzyme/protein databases. C54 = Compost Soil 5-4; C55 = Compost Soil 5-5; C61 = Compost Soil 6-1; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore); SS = Savanna Soil; TM = Termite-Mound Soil; ZS = Zebra-Grazed Soil.*

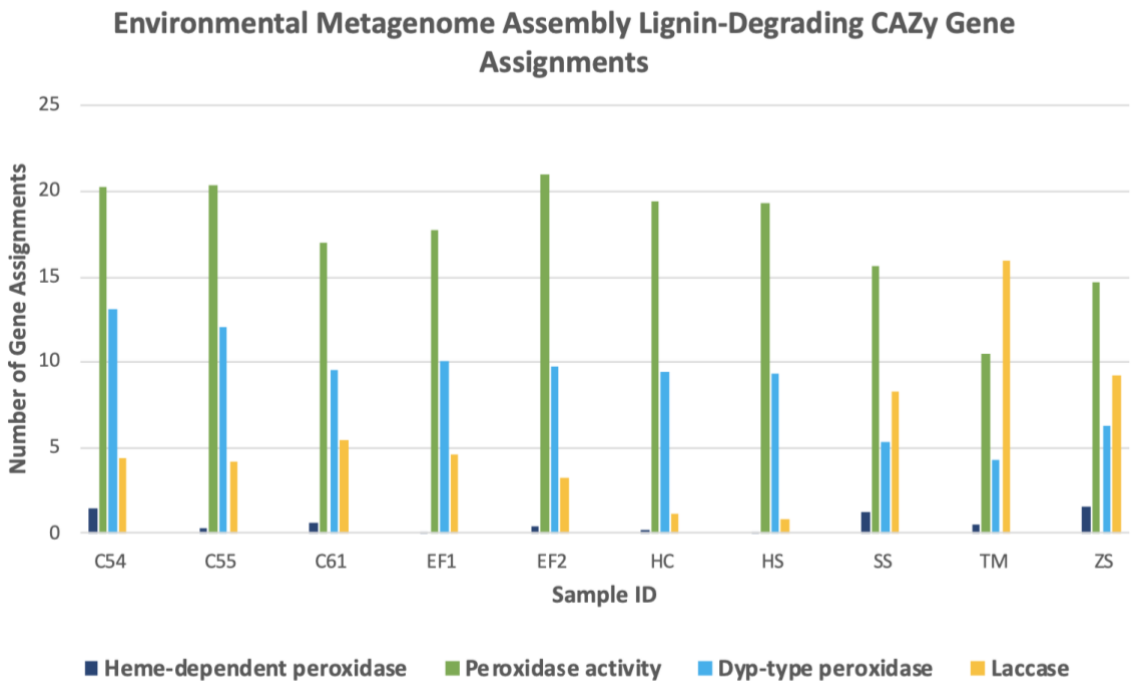

Figure 43 *- Environmental sample metagenome assemblies lignocellulose lignin-active CAZyme number of gene assignments. CAZy gene assignments derived from the EC and InterPro2Go enzyme/protein databases. C54 = Compost Soil 5-4; C55 = Compost Soil 5-5; C61 = Compost Soil 6-1; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; HC = Hickling Sediment (Centre Broad); HS = Hickling Sediment (Broad Shore); SS = Savanna Soil; TM = Termite-Mound Soil; ZS = Zebra-Grazed Soil.*

4.2.5.3 DAS_Tool Binning with Assembled or Co-Assembled Metagenomes

In addition to metagenome data processing and taxonomic/functional analysis, the co-assembled EF metagenomes were binned into metagenome assembled genomes (MAGs) using a binning pipeline on the CLIMB instance. The EF samples were selected for binning due to their similarity within results obtained from beta-diversity analysis of the amplicon sequencing data. The EF1 and EF2 metagenomes were co-assembled to provide a larger, more complete dataset, therefore providing more reliable downstream results. The binning pipeline encompassed several binning algorithms - Metabat2 (Kang *et al*., 2015), Binsanity (Graham *et al*., 2017), Maxbin2 (Wu *et al*., 2015) and CONCOCT (Alneberg *et al*., 2014) - to create metagenome bins which were subsequently fed into DAS_Tool (Sieber *et al*., 2018), which created high-quality consensus bins (as detailed in **Methods and Materials Section 2.7.2.1**). All binning tools were run with minimum contig length >2000 bp. The resulting bins from DAS_Tool consensus binning then underwent analysis with CheckM (v1.1.2) (Parks *et al*., 2016) to determine bin completeness, contamination and taxonomic assignment, amongst other characteristics.

Of the 128 resulting bins obtained from the DAS_Tool binning of the EF metagenomic co-assembly, 25 genome bins were selected based on high completeness (>90%) and low contamination (<10%). The resulting bins and their details can be found below in **Table 16**. Bins were also selected for further study based on their taxonomic assignment (assigned taxonomy used to determine relevance as potential lignin degrader). Preferably, all 128 bins would have underwent analysis, however, bins had to be selected using taxonomic assignment due to time limitations, as discussed below in the **Future Work Section** of this chapter. CheckM's SSU finder was also run on the consensus bins in order to identify microbial SSU rRNA genes (16S or 18S) (SSU finder results not shown). This was done to confirm the taxonomy of bins assigned by CheckM based on other genes.

The MAGs selected from the EF co-assembled metagenomes were analysed with the CAZy Diamond MEGAN6 pipeline (like the environmental sample assemblies) in order to identify potential lignocellulose-active microorganisms. From the absolute lignocellulose-active CAZy gene counts displayed in **Table 17** below, overall, counts corresponding to the gene encoding xylan 1,4-beta-xylosidase were consistently present within the consensus bins, with higher gene counts compared to other enzymes and some bins only containing the gene encoding this enzyme. Another common enzyme-encoding gene throughout the bins were for cellulase enzymes, used by microorganisms to degrade cellulose within lignocellulose, observed within almost every bin. As mentioned above, the production of cellulase and beta-xylosidase enzymes is common amongst microorganisms (Jayasekera and Ratnayake, 2019; Bajpai, 2014). Particular bins of interest, with regards to lignin degradation are bins 5 (*Verrucomicrobiaceae* family), 10 (*Phyllobacteriaceae* family, *Rhizobiales* order) and 12 (*Gammaproteobacterium* class), which possess genes encoding cellulase, xylan 1,4-beta-xylosidase and peroxidase/DyP peroxidase enzymes, meaning these microorganisms have the genetic potential to degrade all major components of lignocellulose. Bins 6 (*Verrucomicrobiaceae* family), 16 (*Flavobacterium* genus) and 19 (*Planctomycetaceae* family) were found to contain none of the lignocellulose-active enzymes searched for in this study.

This information indicates that these microorganisms could be key degraders of lignocellulosic material within the elephant gut. The Gammaproteobacteria, Verrucomicrobia and Rhizobiales were present within the 16S rRNA gene amplicon sequencing data (**Figure 34**, Gammaproteobacteria and Rhizobiales groups not shown) of both EF1 and EF2 samples (9% and 12%, 17% and 8%, 5% and 11%, respectively). A study carried out by Budd *et al.* (2020) on the elephant gut microbiome showed that

the Verrucomicrobia were highly abundant within these environments, following the Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria. Furthermore, Jakeer *et al*. (2020), through metagenomic sequencing of elephant feaces, found that *Proteobacteria* were the dominant bacterial group (91% relative abundance) which were responsible for the majority of glycosyl transferase and glycosylase CAZy genes found, as well as peroxidase genes. The study also found that Gammaproteobacteria and Rhizobiales were prominent taxonomic groups within elephant faeces, similar to the results of our study (**Figure 12**, Gammaproteobacteria and Rhizobiales groups not shown). As with the results shown in the previous section, the data shown here was produced from a metagenomic (co-)assembly, therefore no abundance information could be obtained. This data does, however, provide valuable information on potentially key lignin-degrading species within these samples, such as the MAGs obtained from the Gammaproteobacteria, *Verrucomicrobiaceae* and *Phyllobacteriaceae* groups of bacteria. To the knowledge of the author of this thesis, there is no previous evidence of the *Verrucomicrobiaceae* and *Phyllobacteriaceae* families of bacteria being involved in the lignin-degradation process within the elephant gut. Members of the Rhizobiales order (of which the *Phyllobacteriaceae* are affiliated) obtained from composted wheat straw (Taylor *et al*., 2012) have, however, been shown previously to metabolise lignin and lignin model compounds. Furthermore, members of the Verrucomicrobia have also been previously shown to potentially have roles in lignin metabolism in studies such as those carried out by DeAngelis *et al* (2011) and Puentez-tellez *et al* (2019).

Table 16 *- Selected DAS_Tool consensus bins of the co-assembled metagenomes from the elephant faeces samples (EF1 and EF2). Bins were obtained via binning with Metabat2, CONCOCT, Maxbin2 and Binsanity, the results of which fed into DAS_Tool to obtain consensus bins. Binning algorithms ran with contig length > 2000bp.*

Bin ID	Binning Algorithm	Taxonomy	Completeness	Contamination	GC (%)	Genome size	# contigs	N50 contigs	Predicted genes	SSU_finder hits
Bin 1	MetaBat2	Verrucomicrobiaceae	99.3243	1.3514	0.5926	4858953	73	109081	4290	1
Bin 2	MetaBat2	Rhizobiales	98.8965	1.2658	0.6169	4942283	63	152214	4604	0
Bin 3	MetaBat2	Chloroflexi	97.2727	1.0909	0.5951	5359150	91	94759	4654	2
Bin 4	MetaBat2	Sphingomonadales	97.2181	1.6431	0.6427	2618859	80	54758	2583	1
Bin 5	MetaBat2	Actinomycetales	97.5626	1.9268	0.7286	3446418	147	37160	3230	0
Bin 6	MetaBat2	Verrucomicrobiaceae	97.0721	0.6757	0.5691	4247584	266	21668	4105	1
Bin 7	MetaBat2	Verrucomicrobiaceae	97.2973	2.7027	0.6268	4784557	91	109821	4509	0
Bin 8	MetaBat2	Acetobacteraceae	96.7662	0.1156	0.6962	5856318	372	23672	5582	0
Bin 9	MetaBat2	Actinobacteria	97.9297	4.6249	0.6835	3955836	339	18469	4002	0
Bin 10	MetaBat2	Flammeovirgaceae	98.5119	0.6696	0.4884	4516652	95	97124	3829	0
Bin 11	MetaBat2	Phyllobacteriaceae	96.1239	1.8443	0.6578	5477995	351	22659	5255	0
Bin 12	Binsanity-lc	Planctomycetaceae	97.6335	4.2146	0.5646	5918352	223	41289	5047	2
Bin 13	Binsanity-lc	Gammaproteobacteria	96.5517	2.0847	0.5490	3825062	125	58765	3372	5
Bin 14	CONCOCT	Burkholderiales	96.9136	0.2058	0.6684	4441903	319	22439	4306	1
Bin 15	CONCOCT	Proteobacteria	98.3926	4.8240	0.4987	3814590	411	13911	3569	1
Bin 16	CONCOCT	Rhizobiales	95.4348	0.2174	0.5636	3173315	210	24755	3154	0
Bin 17	CONCOCT	Comamonadaceae	97.7559	0.4673	0.6602	4161507	165	38330	4094	0
Bin 18	CONCOCT	Flavobacterium	99.6466	0.0000	0.3375	3497454	33	301600	3100	0
Bin 19	CONCOCT	Cohnella	98.1183	1.1241	0.4554	5476055	63	227399	4945	2
Bin 20	CONCOCT	Rhizobiales	99.3151	1.0475	0.6200	5414173	47	215312	4302	0
Bin 21	MaxBin2	Bacteria	98.9011	1.6484	0.6080	4376894	201	41173	3856	0
Bin 22	MaxBin2	Planctomycetaceae	96.5909	1.1364	0.5444	3104577	4	974755	2694	1
Bin 23	MaxBin2	Gammaproteobacteria	96.5522	3.4406	0.5006	3906289	347	19231	3614	0
Bin 24	MaxBin2	Planctomycetaceae	97.7273	0.0000	0.6707	4606097	65	147329	3687	2
Bin 25	MaxBin2	Paenibacillaceae	98.3871	1.1557	0.4264	5329022	86	189938	4959	2

Table 17 *- Lignocellulose-active CAZyme gene presence and absolute counts in selected DAS_Tool consensus bins of the co-assembled metagenomes from the elephant faeces samples (EF1 and EF2). CAZy gene assignments derived from the EC and InterPro2Go enzyme/protein databases. Lignocellulose-active CAZy genes identified in bins highlighted in green.*

		CAZy Gene Lignocellulose-Active Gene Absolute Counts					
Bin ID	**Highest Taxonomic Rank**	**Laccase**	**Cellulase**	**Endo-1,4-beta-xylanase**	**Xylan 1,4-beta-xylosidase**	**Peroxidase activity**	**DyP-type peroxidase**
Bin 1	f__Verrucomicrobiaceae	0	1	0	2	0	0
Bin 2	o__Rhizobiales	0	0	0	5	0	0
Bin 4	o__Sphingomonadales	0	0	0	1	0	0
Bin 5	o__Actinomycetales	0	4	7	10	0	0
Bin 6	f__Verrucomicrobiaceae	1	1	0	8	1	0
Bin 7	f__Verrucomicrobiaceae	0	0	0	0	0	0
Bin 8	f__Acetobacteraceae	0	0	0	17	0	0
Bin 9	c__Actinobacteria	0	0	2	14	0	0
Bin 10	f__Flammeovirgaceae	0	3	0	1	0	0
Bin 11	f__Phyllobacteriaceae	0	2	0	22	2	1
Bin 12	f__Planctomycetaceae	1	0	0	3	0	0
Bin 13	c__Gammaproteobacteria	0	6	2	3	1	1
Bin 14	o__Burkholderiales	0	0	0	15	0	0
Bin 16	o__Rhizobiales	0	0	0	5	0	0
Bin 17	f__Comamonadaceae	0	0	0	12	0	0
Bin 18	g__Flavobacterium	0	0	0	0	0	0
Bin 19	g__Cohnella	0	1	1	1	0	0
Bin 20	o__Rhizobiales	0	1	0	7	0	0
Bin 22	f__Planctomycetaceae	0	0	0	0	0	0
Bin 23	c__Gammaproteobacteria	0	6	3	8	0	0
Bin 24	f__Planctomycetaceae	0	0	0	1	0	0
Bin 25	f__Paenibacillaceae	0	1	2	0	0	0

4.2.6 DNA/RNA-SIP with Environmental Samples and ^{13}C-Labelled Lignin

Incubation with ^{13}C-labelled lignin was carried out with the Olduvai Gorge soil, elephant faeces and compost soil environmental samples. The environmental samples underwent SIP incubation in conditions similar to natural environmental conditions (in dark) with ^{13}C-labelled lignin, ^{12}C-lignin (natural abundance control), $^{13}CO_2$ (secondary labelling control) and no substrate control replicates. Microcosm preparation methodology and incubation details can be found in the **Methods and Materials Section 2.9.1**. Samples taken from microcosms half-way through incubation ($T_{0.5}$) and after (post-) incubation (T_1) were flash-frozen in liquid nitrogen and stored at -80 °C. Information on the sampling of SIP microcosms can also be found in **Methods and Materials Section 2.9.1**. Like in the enrichment experiments shown in **Chapter 3**, lignin degradation/removal was not tracked in the SIP experiments shown in this section. However, the SIP microcosms were incubated in the same conditions as the enrichment cultures in **Chapter 3**, from which lignin degrading microorganisms were obtained, therefore lignin was likely being degraded within the SIP microcosms also.

Post-incubation (T_1) samples from the Olduvai Gorge soil, elephant faeces and compost soil SIP incubations underwent DNA/RNA extraction using both CTAB and SDS extraction protocols (detailed in **Methods and Materials Section 2.5.1**). As mentioned above, CTAB extraction was used for nucleic acid extraction from compost soil and elephant faeces samples, whereas SDS extraction was used for the Olduvai Gorge samples. The SDS DNA/RNA extracts from Olduvai Gorge samples were dark in colour (likely due to high humic acid content) and therefore underwent further purification using Zymo Research OneStep PCR Inhibitor Removal Kit. Gel electrophoresis analysis of all SIP DNA/RNA extracts confirmed that both genomic DNA and RNA are present within each extract before gradient ultracentrifugation and fractionation.

4.2.6.1 DNA-SIP

^{13}C-labelled (heavy) and non-labelled (light) nucleic acids were separated using cesium chloride (CsCl) density ultracentrifugation and gradient fractionation (methodology detailed in **Methods and Materials Section 2.9.2**) and bacterial 16S rRNA gene qPCR analysis was carried out on the nucleic acid fractions (in order to identify labelled fractions) (methodology can be found in **Methods and Materials Section 2.8.1**). Bacterial 16S rRNA gene relative abundance was plotted against CsCl concentration (obtained from refractometer measurements) in order to identify the SIP fractions

containing labelled DNA derived from lignin-degrading communities. The results of DNA SIP fraction qPCR analysis can be found in **Figure 44** (Olduvai Gorge samples), **Figure 45** (elephant faeces samples) and **Figure 46** (compost samples) below. Results were not obtained from Olduvai Gorge sample TM due to the quantities of DNA obtained from this sample being too low.

Upon reviewing data obtained from the DNA SIP fraction qPCR analysis (**Figures 44, 45 and 46**), it was found that, whilst gradients had formed correctly (evidenced by clear and defined peaks), no labelling could be observed in any of the ^{13}C-labelled fractions obtained (from all environmental sample incubations). Within each graph in each figure, peaks can clearly be observed in both the ^{12}C (blue) and ^{13}C (orange) gradients, representing the 'light' DNA fractions of the non-labelled microbial community, however, no secondary 'heavy' labelled peaks were observed. These results suggested that, perhaps, the peak representing nucleic acids from the labelled lignin-degrading microbial community had been overshadowed by the non-labelled peak, or that the labelled and non-labelled had not sufficiently separated within the gradients.

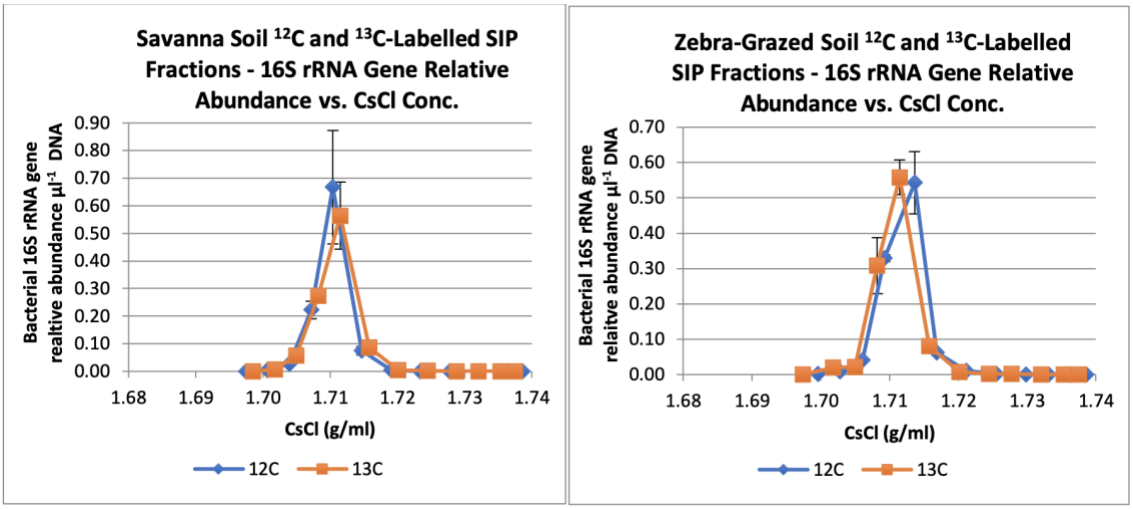

Figure 44 *- Distribution of bacterial 16S rRNA genes within savanna soil (left) and zebra-grazed soil (right) in DNA-SIP after 4 weeks of incubation at 25 ºC with 3 mg lignin (^{12}C or ^{13}C). Lignin was added to 3 g soil suspended in 2.5 ml sterile water. Distribution of Bacterial 16S rRNA genes was measured by qPCR of DNA from gradient fractions. The copy numbers represent the mean from repeated qPCR analyses.*

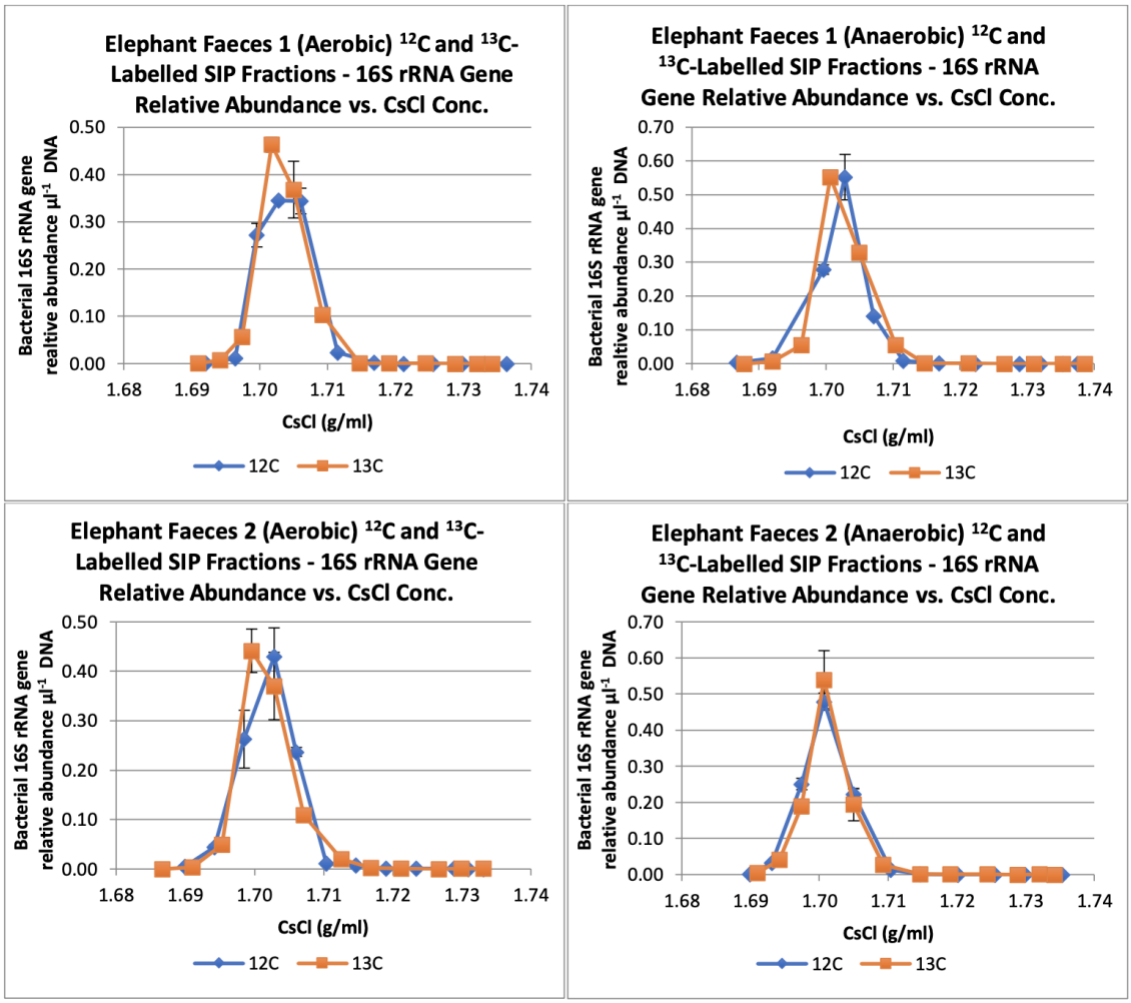

Figure 45 *- Distribution of Bacterial 16S rRNA genes within elephant faeces samples 1 (aerobic – top left; anaerobic – top right) and 2 (aerobic – bottom left; anaerobic – bottom right) in DNA-SIP after 2 weeks of incubation at 37 ºC with 3 mg lignin (^{12}C or ^{13}C). Lignin was added to 3 g faeces sample suspended in 3 ml sterile water. Anoxic microcosms were flushed with N_2/CO_2 air mixture. Distribution of Bacterial 16S rRNA genes was measured by qPCR of DNA from gradient fractions. The copy numbers represent the mean from repeated qPCR analyses.*

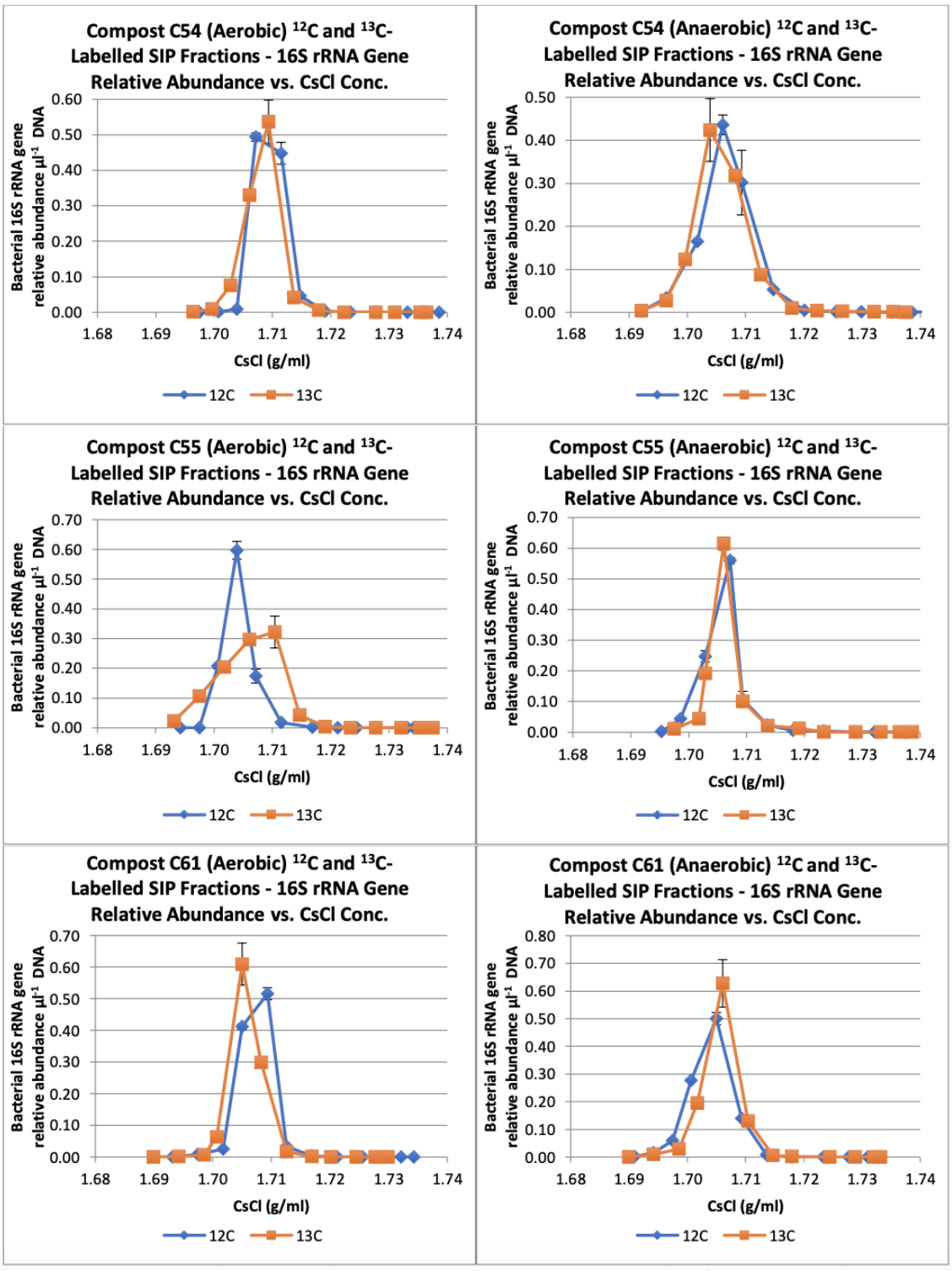

Figure 46 *- Distribution of Bacterial 16S rRNA genes within compost samples C54 (aerobic – top left; anaerobic – top right), C55 (aerobic – middle left; anaerobic – middle right) and C61 (aerobic – bottom left; anaerobic – bottom right) in DNA-SIP after 2 weeks of incubation at 50 ºC with 3 mg lignin (^{12}C or ^{13}C). Lignin was added to 3 g compost sample suspended in 2.5 ml sterile water. Anoxic microcosms were flushed with N_2/CO_2 air mixture. Distribution of Bacterial 16S rRNA genes was measured by qPCR of DNA from gradient fractions. The copy numbers represent the mean from repeated qPCR analyses.*

To investigate whether or not any labelled microbial community might have been missed by the bacterial 16S rRNA gene qPCR, PCR was additionally carried out using primers targeting the functional *dyp* gene in order to more specifically target nucleic acids from the lignin-degrading communities (protocol can be found in **Methods and Materials Section 2.6.1.3**). *dyp* PCR was carried out on the ^{12}C and ^{13}C-labelled fractions of anaerobic elephant faeces sample 1 (**Figure 47**) and savanna soil (**Figure 48**) SIP incubations; gel images from which can be found below. The results from this analysis, like the bacterial 16S qPCR results, showed no evidence of labelling in any of the heavy fractions from each incubation analysed (expected *dyp* gene product size of 300-400 bp in length). Like in the bacterial 16S qPCR results, the non-labelled DNA 'peaks' can be observed as nucleic acid bands in the gel images below in fraction 8-9 for the anaerobic EF1 fractions and fractions 7-9 for the savanna soil fractions, whereas, if ^{13}C labelling was present, bands would also be present within the heavier fractions (e.g. F1-F6), which was not the case.

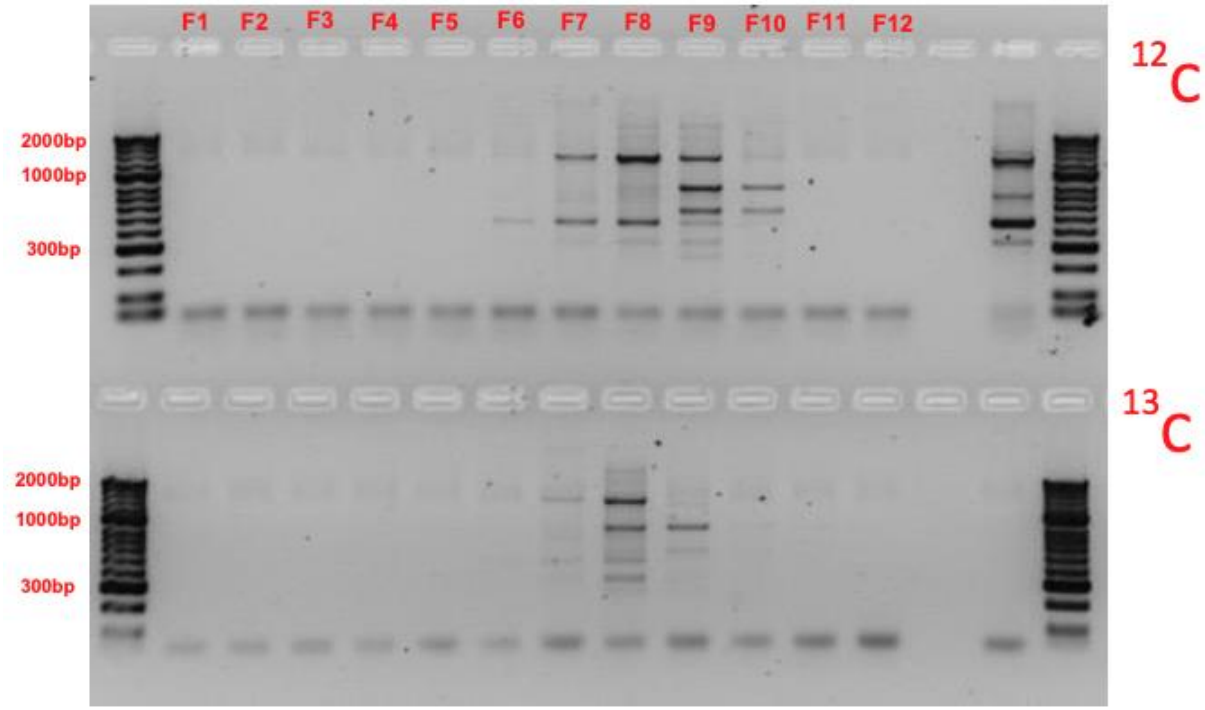

Figure 47 *- Distribution of dyp functional genes within anaerobic elephant faeces sample 1 in DNA-SIP after 2 weeks of incubation at 37 ºC with 3 mg lignin (^{12}C (top lanes) or ^{13}C (bottom lanes)). Lignin was added to 3 g faeces sample suspended in 3 ml sterile water. Anoxic microcosms were flushed with N_2/CO_2 air mixture. Distribution dyp genes was determined by dyp functional gene PCR amplification of DNA from gradient fractions (F1-F12) followed by gel electrophoresis image analysis. Expected product size: 300-400 bp.*

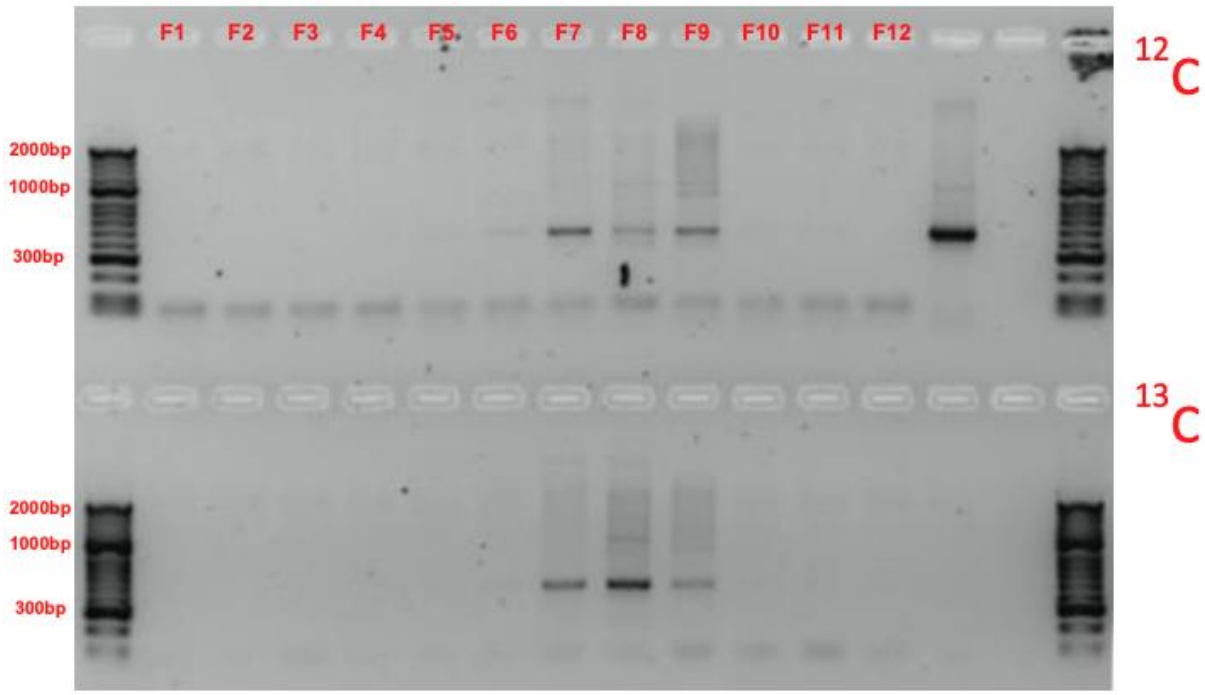

Figure 48 *- Distribution of dyp functional genes within savanna soil in DNA-SIP after 4 weeks of incubation at 25 ºC with 3 mg lignin (^{12}C (top lanes) or ^{13}C (bottom lanes)). Lignin was added to 3 g soil sample suspended in 2.5 ml sterile water. Distribution dyp genes was determined by dyp functional gene PCR amplification of DNA from gradient fractions (F1-F12) followed by gel electrophoresis image analysis. Expected product size: 300-400 bp.*

4.2.6.2 RNA-SIP

Following the lack of success in identifying ^{13}C labelling within the DNA-SIP fractions, it was therefore decided to instead use RNA-SIP which would increase sensitivity (due to higher RNA synthesis rates compared to DNA) and the chance to detect labelled fractions (Manefield *et al*., 2002). The total nucleic acid extracts from SIP incubation samples underwent DNase digestion followed by RNA purification to isolate RNA (according to the procedures detailed in **Methods and Materials Section 2.9.3**). RNA extracts underwent CsTFA gradient ultracentrifugation and fractionation according to methodology detailed in **Methods and Materials Section 2.9.4**. In order to avoid gradient overloading (observed in previous results – not shown), 400 ng RNA was added instead of 500 ng (the routine concentration added to RNA gradients). Following gradient ultracentrifugation and fractionation, RNA fractions were reverse transcribed into cDNA according to methodology detailed in **Methods and Materials Section 2.9.3**. Like the DNA fractions, cDNA fractions were analysed for labelling by carrying out bacterial 16S rRNA gene qPCR analysis on each fraction (bacterial 16S rRNA gene relative abundance plotted against CsTFA concentration).

The bacterial 16S rRNA gene qPCR results of successful gradients can be found below in **Figure 49** (aerobic/anaerobic elephant faeces 1 and 2 ^{12}C and ^{13}C gradients) and **Figure 50** (aerobic C54 and aerobic/anaerobic C61 compost sample ^{12}C and ^{13}C gradients). The results obtained from bacterial 16S rRNA gene qPCR analysis of the RNA SIP cDNA fractions showed that the RNA gradients had formed correctly, with clear defined 16S gene peaks within the ^{12}C and ^{13}C gradients. Despite this, like in the DNA-SIP results, no ^{13}C labelling was observed within any of the analysed ^{13}C gradients. Unfortunately, as no ^{13}C labelling could be identified within any of the results obtained, no fractions were sent for NGS analysis, therefore no conclusions towards the aims of this study could be made. As to the reasoning behind the lack of ^{13}C labelling, it could be possible that the samples were not incubated with the ^{13}C substrate for a long enough period, resulting in a lack of incorporation of the ^{13}C-lignin substrate by lignin-active microorganisms. It could also be possible that the lignin-degrading community within the natural environment is so small that the non-labelled fraction 'peaks' overshadowed the ^{13}C labelled 'peaks', leading to the peak going unidentified. In addition, it could be possible that the labelled and non-labelled nucleic acids were not sufficiently separated during ultracentrifugation (leading to both peaks appearing as one). Due to the lack of labelling observed in the primary substrate (^{13}C-lignin) incubations, the $^{13}CO_2$ control samples were not analysed.

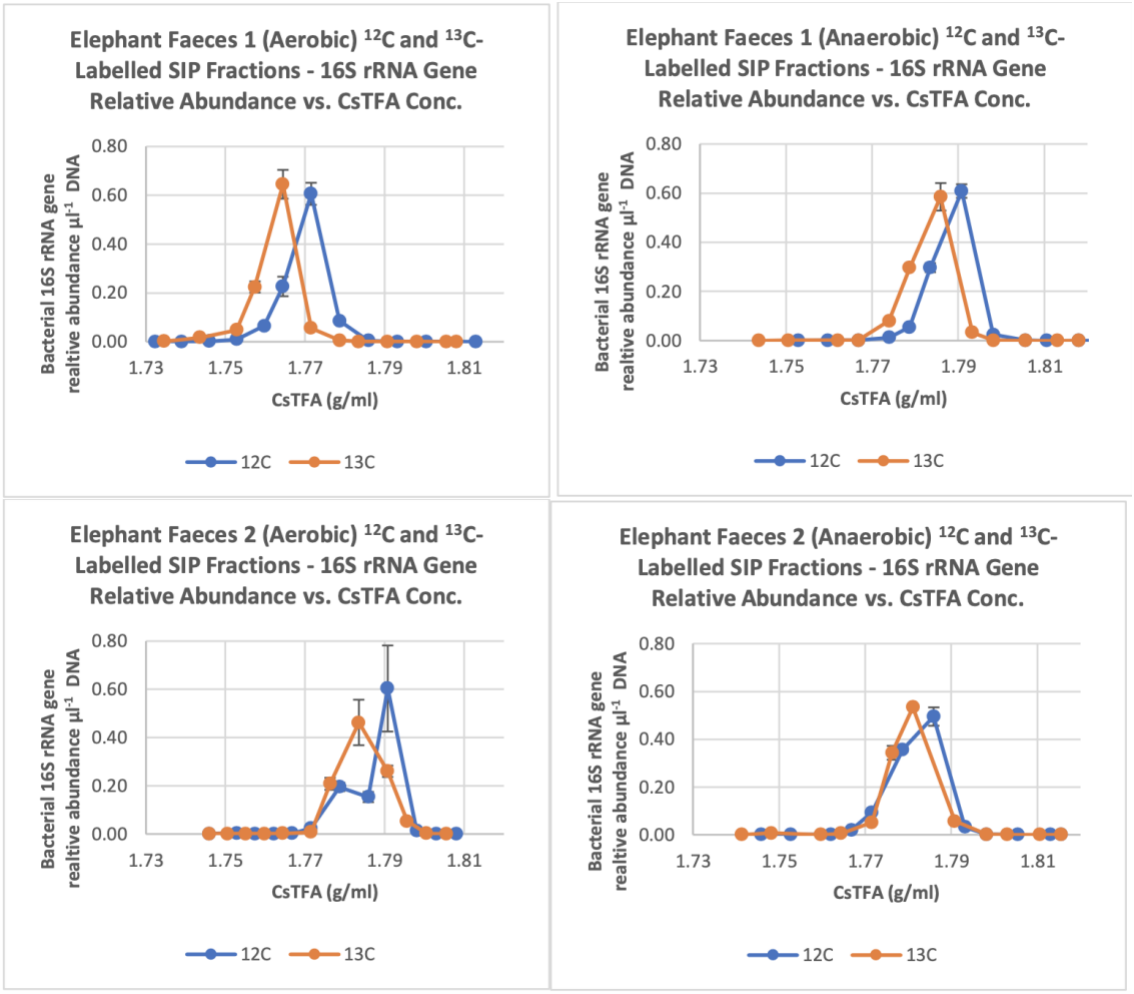

Figure 49 *- Distribution of Bacterial 16S rRNA genes within elephant faeces samples 1 (aerobic – top left; anaerobic – top right) and 2 (aerobic – bottom left; anaerobic – bottom right) cDNA in RNA-SIP after 2 weeks of incubation at 37 ºC with 3 mg lignin (^{12}C or ^{13}C). Lignin was added to 3 g faeces sample suspended in 3 ml sterile water. Anoxic microcosms were flushed with N_2/CO_2 air mixture. Distribution of Bacterial 16S rRNA genes was measured by qPCR of cDNA (reverse transcribed) from RNA gradient fractions. The copy numbers represent the mean from repeated qPCR analyses.*

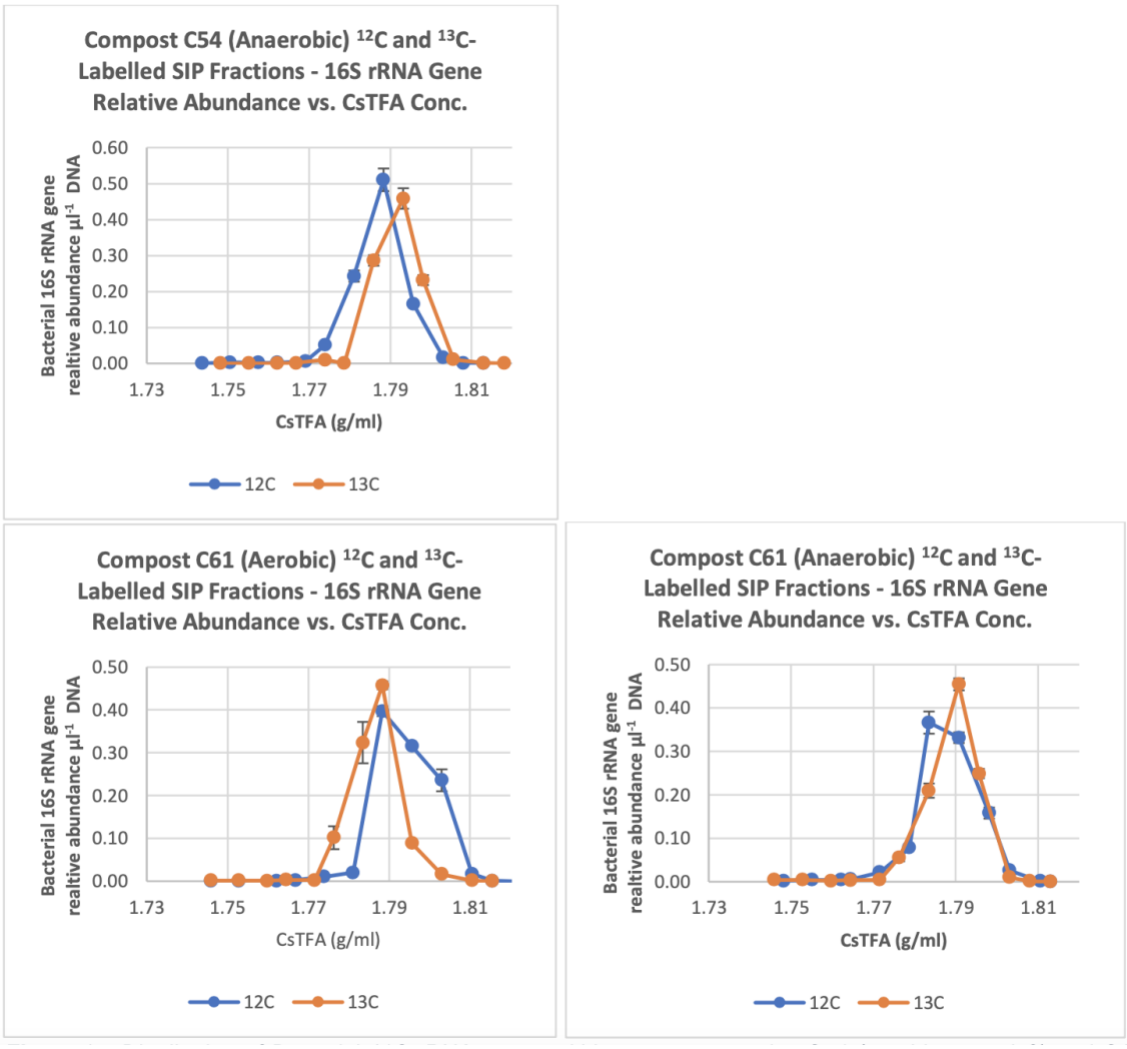

Figure 50 *- Distribution of Bacterial 16S rRNA genes within compost samples C54 (aerobic – top left) and C61 (aerobic – bottom left; anaerobic – bottom right) cDNA in RNA-SIP after 2 weeks of incubation at 37 ºC with 3 mg lignin (^{12}C or ^{13}C). Lignin was added to 3 g faeces sample suspended in 3 ml sterile water. Anoxic microcosms were flushed with N_2/CO_2 air mixture. Distribution of Bacterial 16S rRNA genes was measured by qPCR of cDNA (reverse transcribed) from RNA gradient fractions. The copy numbers represent the mean from repeated qPCR analyses.*

4.3 Conclusions

In summary, through qPCR analysis using primers targeting the bacterial 16S rRNA gene and *dyp* functional gene, the *dyp* gene was found to be present within all 10 samples. This indicated that all samples contained a bacterial community with the potential ability to utilise DyP-type peroxidase enzymes. Furthermore, it was found that the OG and EF samples had greater abundances of these genes compared to the CS and HS samples, suggesting that the OG and EF samples potentially had a larger abundance of lignin-active bacterial communities.

Taxonomic profiles constructed from the microbial SSU rRNA gene amplicon sequencing data revealed that each environmental sample had a distinct microbial community, however similarities were observed within sample subgroups. To support this, beta-diversity analysis using weighted and unweighted Unifrac distances showed that the sample sets (OG, EF and CS) were taxonomically unique from one another and similarities were further illustrated. The EF samples were closely similar in taxonomy, with the same being observed for the OG samples SS and ZS. Within the OG sample set, TM was dissimilar to the other samples, clustering separately from the former, likely due to difference in sample type. The individual CS samples were also seen to cluster separately and therefore be taxonomically diverse from one another, suggesting the impacts of compost composition and maintenance methods on microbial community structure.

Kraken2 taxonomic profiles, constructed from environmental sample metagenomic trimmed reads, yielded similar results to that of the microbial SSU rRNA gene amplicon sequencing profiles. However, whilst these results mirrored one another in terms of taxa presence, the Kraken2 profiles showed dissimilar taxa abundance values, in addition to lower depth of diversity. The reason for this likely resulted from the smaller reference database used by Kraken2, with lower amount of reference data leading to lower number of assigned reads. Therefore, this data was only used to reinforce the taxonomic profiles obtained from the microbial SSU rRNA gene amplicon sequencing data.

Functional profiles generated using the environment metagenome assemblies involved exploring the presence and diversity of CAZy genes. These results revealed the presence of various genes encoding enzymes active against the major components of lignocellulose in all samples. The presence of cellulose-degrading cellulase, several enzymes involved in the degradation of xylan (hemicellulose) and lignin-degrading peroxidases and laccase indicated the presence of lignocellulose-degrading and, more specifically, lignin-degrading communities within these samples. Furthermore, the identification of genes encoding DyP peroxidases within the elephant faeces samples was, to the best of the authors knowledge, novel information. The presence of these genes indicated the potential importance of bacteria in the microbial lignin degradation process within the elephant faeces/gut environment.

A deeper exploration of the EF samples using a DAS_Tool binning pipeline followed by CAZy gene analysis revealed the partial identity of potential key lignin-degrading microorganisms. CAZy gene analysis of selected EF MAGs showed that microorganisms of the *Phyllobacteriaceae* family and Gammaproteobacteria class harboured genes

encoding peroxidase and DyP-type peroxidase. Furthermore, bins derived from microorganisms of the *Planctomycetaceae* and *Verrucomicrobiaceae* families harboured genes encoding laccase with the former containing genes also encoding peroxidase.

4.4 Limitations and Future Work

Nucleic acid SIP techniques were employed to isolate nucleic acids from lignin-active microbial communities in hopes of identifying these communities via NGS techniques. Despite attempting both DNA- and RNA-SIP, as well as screening SIP fractions using qPCR with primers targeting the bacterial 16S rRNA gene and *dyp* functional gene, no labelled fractions were obtained. Due to the impacts of the COVID-19 pandemic limiting laboratory work and restricted time towards the end of this study, this work could not be completed. As mentioned above, it is thought that the environmental samples had not been enriched for a long enough period of time to permit sufficient labelling, or the ratio of labelled community to the non-labelled community was so low that labelling could not be detected. Both issues could perhaps have been sufficiently addressed by repeating the SIP enrichment experiment with ^{13}C-labelled lignin, however this was impossible due to the restricted time and the significant monetary cost of ^{13}C-labelled lignin. Plans were also made to analyse the phospholipid fatty acid (PLFA) content (PLFA-SIP) using GC-MS analysis, in coordination with Dr Clay Magill (Lyell Centre, HWU). SIP incubation samples were to undergo PLFA extraction followed by GC-MS analysis in order to detect and identify ^{13}C-labelled PLFAs within the SIP samples, however these plans did not come to fruition. In future, the SIP incubations would be repeated with longer incubation periods and SIP fractions would be analysed using qPCR primers targeting a wider variety of lignin-active enzyme genes (e.g. laccases via the *lmco* gene, **Appendix 7.1**). Furthermore, methodologies regarding the monitoring of lignin within the SIP microcosms or microbial incorporation of the ^{13}C-lignin substrate would be researched and included in the experiment to confirm lignin degradation/incorporation. In addition, the SIP samples would undergo PLFA analysis, as stated above, to detect labelling.

The COVID-19 pandemic and restricted laboratory time following the pandemic not only impacted the progress of SIP, but other areas of this study. For example, before the pandemic, a qPCR methodology was being developed to target the *lmco* functional gene, much like the methodology used to target the *dyp* gene. This method was to be used to quantify this gene within the unenriched (natural) environmental samples and also to be used to screen SIP fractions to detect labelling. This assay may have permitted the quantification of laccase-coding genes within the environmental samples, providing more information on the lignin-degrading potential of the microbial communities within. In

addition, the *lmco* qPCR assay may have permitted the identification of labelled SIP fractions as laccases are expressed by both bacteria and fungi.

The primers used to amplify the 16S and functional genes (*dyp* and *lmco*) within this project were chosen to specifically target these genes. However, as discussed above, the degeneracy of these wide-spectrum primers can lead to potential biases being introduced into the data obtained. The PCR methodologies used to amplify these genes were taken from previous studies and involved some optimisation work to reduce non-specific amplification (e.g. changing the length of elongation time used in the method obtained from Tian *et al* (2016)). Due to the work involving these primers being a small part of the project as a whole, time was not taken to fully investigate the potential biases these primers had on the data obtained.

Two different extraction methodologies were used to obtain the nucleic acids from the environmental samples used in this project (methods involving the use of CTAB and SDS reagents). Nucleic acids were extracted from the Olduvai Gorge and Hickling sediment samples using the SDS method, whereas the elephant faeces and compost samples underwent extraction using the CTAB method. It is possible that the difference in extraction methods used between samples could have introduced bias in the nucleic acids obtained and therefore affected the data obtained from these samples. A study performed by Teng *et al* (2018) investigated the impact of DNA extraction methodologies on the 16S amplicon sequencing results obtained from oral microbiota samples. This study found that the selection of DNA extraction methodology had a significant impact on the bacterial diversity of amplicon sequencing results. Therefore, it is likely that the different methods used to extract nucleic acids in this project had an impact on the results obtained from samples (for example the taxonomic profiles obtained via amplicon sequencing). Despite this, whilst testing the different extraction methods on samples, it was found that certain extraction methods only extracted usable nucleic acids from certain samples (e.g. the CTAB method did not extract usable nucleic acids from the Olduvai Gorge samples, therefore the SDS method had to be used). Therefore, the impact of differences in extraction method was difficult to mitigate. Initially, it was planned to send all extracts (CTAB and SDS) from all samples for analysis (amplicon sequencing and metagenomic sequencing), however, this would have been expensive and it was later decided that this would not be possible.

Full analysis of the metagenomic sequencing data was not possible to complete due to the complexity of metagenomic data, data processing time and numerous potential channels for investigation. For example, the metatranscriptome obtained for the anoxic

EF SIP incubation samples was not analysed. In addition, whilst binning results were obtained for the elephant faeces samples, initial plans were made to carry out metagenome binning for all environmental samples, which was not possible. Binning all metagenomes could have provided valuable information on potential key lignin-degrading microorganisms and allowed a comparison to be made between environments. Furthermore, CAZy gene analysis of all MAGs (including those obtained from the EF metagenomes) could have potentially allowed the discovery of novel lignin-degrading species, instead of the selection method used in this study, which involved selection based on known lignin-degradation capability.

Furthermore, as the CAZy gene analysis (shown in **Section 4.2.5.2**) was carried out on metagenomic assemblies, no reliable abundance information could be obtained, which would have been possible through carrying out this analysis with the metagenomic raw reads. Whilst the analysis carried out in this study provided information on the presence and diversity of microbial CAZy genes within the environmental samples, analysis of the metagenomic raw reads would have allowed quantification of these genes and therefore provided an insight into the potential enzymes more commonly used in these environments. Furthermore, the genes explored within this CAZy analysis encompass only a fraction of enzyme-encoding genes involved in the microbial degradation of lignocellulosic material. The genes explored in our study were selected for analysis due to their relevance in degradation of the 3 primary components of lignocellulose and their direct degenerative ability (for example, enzymes such as the glycosyltransferases possess no direct degradation properties) (Jakeer *et al*., 2020).

If the research encompassed within this chapter were to be continued, binning would therefore be carried out with all sample metagenomes, followed by subsequent CAZy analysis of raw reads and MAGs. Future work would also involve the exploration/quantification of more CAZy genes, alongside the genes investigated in this study. Investigating a larger catalogue of CAZy genes would potentially provide a more thorough insight into the presence/diversity of microbial degradation mechanisms.

Finally, as mentioned in **Introduction Section 4.1** of this chapter, gut and burrow samples derived from wood-boring bivalve molluscs also underwent culture-independent analysis, in addition to the 10 environmental samples discussed in this chapter. This analysis involved microbial SSU rRNA gene amplicon sequencing (bacterial 16S and eukaryotic 18S) as well as metagenomic sequencing, followed by Kraken2 taxonomic profiling and CAZy gene analysis. Despite the significant period of time committed to the analysis of these samples, the resulting data was not to a high enough standard to

present in this thesis. However, these results may be used in a future publication, following refinement.

4.5 Final Words

Despite the limitations detailed above, the results of this study successfully display the utility and benefits of using a combination of nucleic acid sequencing methodologies to explore microbial communities within complex environments. Using these methodologies, this study illustrated the diversity and abundance of microbial species within a range of unique environments, as well as the presence of genes encoding lignin-active enzymes within. Furthermore, this study showed the presence and diversity of potential lignin-degrading mechanisms being used by microorganisms within these environments, in addition to identifying several key lignin-degrading microbial groups within the elephant faeces microbiome. As mentioned above, however, the metagenomic sequencing data used in this study has only partially been explored. Future studies would involve the identification of more lignin-degrading groups within these environments and a deeper exploration of their individual and community-level mechanisms for lignin degradation.

5 Chapter 5: Discussion

Whilst the results of this project have already been discussed in previous sections, this chapter will cover the progress and limitations of the work carried out, as well as providing a deeper discussion into the novelty of the research findings. In this study, a wide variety of culture-dependent and culture-independent methodologies were employed in order to explore the microbial ecology of several different environmental samples, specifically with a focus on communities active against the lignin polymer. With reference to the objectives of this project stated in the **Introduction Chapter**: Several lignin-degrading bacterial microorganisms were isolated from Olduvai Gorge soil, elephant faeces and compost samples (**Objective 3**), all of which showed oxidative enzyme and lignin degrading activities (via enzyme assay and phenolic compound GC-MS analysis) (**Objectives 4 and 5**). However, time limitations did not permit whole genome sequencing to be carried out on the isolates. The microbial communities within these environments were investigated (via microbial SSU rRNA gene amplicon and metagenomic sequencing), as well as several genes relevant to lignin degradation and potential lignin-degrading microorganisms (**Objective 2**). Despite this, work carried out involving the identification of lignin-active microorganisms via DNA- and RNA-SIP was unsuccessful (**Objective 1**).

5.1 Objective Progress

Objective 2: Investigating the identity of and mechanisms being employed by lignin degraders using microbial SSU rRNA and functional gene PCR and qPCR, amplicon sequencing and metagenomics/transcriptomics:

The work encompassed in this project provides a good foundational study for further work, providing information key to building additional research objectives. For example, the microbial SSU rRNA gene amplicon sequencing results, in addition to the metagenome Kraken2 taxonomic profiles, provide an overview of the microbial communities present. Furthermore, the CAZy gene functional analysis of the environment metagenomes provide an initial insight into potential mechanisms being used by microorganisms to break down lignin. In addition, the analysis of CAZy genes present within the EF MAGs (obtained via binning) allowed the identification of potential key lignin-degrading species. With more work, the abundance of lignin-active CAZy genes within each environment, as well as the responsible taxa, could be revealed, providing an abundance of information on lignin-degrading communities. **Objective 2** of

this study was, therefore, partially fulfilled, however the lignin-degrading mechanisms of specific microbial species was not investigated, except the limited analysis carried out on the EF metagenome bins.

Objective 3: Identification and characterisation of key lignin degrading species, isolated from environmental samples through enrichment methods, and their lignin-degrading mechanisms; Objective 4: Determination of isolate and enzyme functionality using functional cloning and lignin-degrading enzyme activity assays; Objective 5: Determination of isolate lignin breakdown mechanisms and resulting aromatic compounds (with industrial relevance) through advanced biochemical analysis:

Project **Objectives 3, 4 and 5**, encompassing the enrichment, isolation and functional analysis of lignin-degrading microorganisms from the environmental samples were successful, excluding the originally planned WGS and functional cloning work. Whilst the isolates obtained in this study were already known lignin-degraders, having these isolates in pure culture will allow further analysis to be carried out. Future work, as stated above in **Chapter 3**, would involve further investigation of the functionality and industrial potential of these isolates. Additional growth experiments (more temperature and pH replicates), carried out in combination with enzyme assays (perhaps targeting more lignin-active enzymes) would provide a deeper understanding of optimal lignin degradation capacity, further information on isolate functionality and reveal their potential for industrial use. For example, the β-etherase enzyme, which cleaves the β-O-4 aryl ether bonds within lignin and has been shown to be important in microbial lignin degradation pathways, would be interesting to analyse with the isolates (Bugg *et al*., 2016). In addition to testing the growth optima of individual isolates, the compost samples could also have been enriched in different conditions to replicate the stages of the composting process, and to obtain lignin-degrading isolates from each stage. Nucleic acid extraction with isolate pure cultures, followed by WGS would provide a deeper insight into the lignin-degrading functional capabilities of these isolates, as well as their true identities to species level. Furthermore, in addition to the catalogue of lignin-derived products explored in the GC-MS analysis, investigation of more products (such as direct derivatives of monolignol components) would have provided more information on lignin catabolic pathways of the isolates. Following on from the aforementioned work, a study involving the preparation and functional analysis of consortia encompassing the isolates would be a worthy endeavour, in order to determine the potential for increased lignin-degrading capacity using several key degraders.

Objective 1: Screening of environmental samples and identification of microorganisms involved in lignin degradation using stable isotope probing:

The lack of success in obtaining labelled DNA/RNA fractions from the environmental samples using SIP (**Objective 1**) was unfortunate, however carrying out this work alongside the other research objectives of this project was ambitious. As mentioned previously, it is likely that the environmental samples were not incubated with the ^{13}C-lignin substrate for a long enough period of time. As lignin is an extremely large and complex polymer, microorganisms present within the environmental samples would have required more time to fully incorporate this substrate. The SIP incubations time periods were selected based on the activity observed within the enrichment cultures, in that significant colour changes were observed within 1-2 weeks. The Olduvai Gorge samples were incubated for 28 days, whereas the elephant faeces and compost samples were incubated for 14 days. A longer incubation time would likely have been required in order to achieve suitable labelling, such as the incubation time of 60 days used by Wilhelm *et al* (2018). Furthermore, Wilhelm *et al* (2018) experienced additional difficulties with ^{13}C-lignin labelling in that the organic matter already present within the environmental samples prevented uptake of the ^{13}C substrate. This study therefore diluted the environmental samples with autoclaved mineral soil to reduce the presence of organic matter which, upon reflection, would have aided the work carried out in this project also.

Despite some objectives being incomplete, the results provided by this study outline the participation of specific groups/species which contribute towards lignin degradation within different, and in some cases niche, environments. This information is fundamental to understanding the process of microbial degradation of plant material within natural environments and, ultimately, the fate of lignocellulosic carbon within these environments. Identification of specialised species involved in lignin degradation will inform targeted studies aimed at isolating these species for potential utilization in industrial processes. These species could have a valuable impact on current and future-generation lignocellulose conversion processes, providing an effective means of removing lignin from plant material for production of biofuels or high-value chemicals.

5.2 Isolation of Microbial Lignin-Degrading Microorganisms and Comparisons Between Environments

Aside from identifying key lignin-degrading species and exploring the mechanisms they use within individual environments; an additional objective of this study was to explore

the differences in these microorganisms and mechanisms between the environmental samples used. From the enzyme assay results in **Chapter 3**, **Section 3.2.5** (**Figure 21**), it was shown that most isolates produced peroxidase enzyme, except the anoxic isolates which showed low activity due to reasons already discussed above. The results obtained from CAZy gene analysis of the metagenome assemblies (**Chapter 4**, **Section 4.2.5.2**, **Figure 43**) displayed the presence of genes coding for several lignin-active enzymes, encompassing laccases and peroxidases, which were present within all environments analysed. These results were supported by the PCR and qPCR analysis using primers targeting the *lmco* (coding for laccase-like multicopper oxidase) and *dyp* (DyP peroxidase) functional genes, showing the presence of these genes within all environments (**Chapter 4**, **Section 4.2.2**, **Figures 30 and 31**) and higher *dyp* gene abundances within the OG soils (**Chapter 4**, **Section 4.2.3**, **Figure 33**). Whilst the presence of these enzymes and enzyme-encoding genes within the isolates and environments is interesting, these results did not enable a comparison to be made between environments. In order to make said comparison, as previously stated, CAZy gene analysis with metagenomic raw reads would have been required, to reveal the abundances of these genes and their differences between environments. Furthermore, metagenome binning with all environmental samples would have been required in order to determine the differences in potential key lignin degraders as well as potential mechanisms between environments. However, time only permitted CAZy gene analysis with the metagenomic assemblies and binning to be carried out with the EF co-assembled metagenomes.

Despite the impacts of these limitations, the varied catalogue of microbial species obtained from the lignin enrichment cultures provides evidence for differences in lignin degraders between the environments explored. The majority of these isolates were enriched with lignin as sole carbon source in minimal medium, leaving little alternative substrates for these isolates to grow on. Through GC-MS analysis of the alkali lignin substrate, it was shown that this substrate contained several lignin-derived molecules (such as vanillin, guaiacol and acetovanillone) in addition to polymeric lignin, likely present due to the alkali processes used to produce the substrate. Whilst it is possible that the isolates used these phenolic compounds to proliferate as opposed to the polymeric lignin, the GC-MS results also showed increases in larger compounds (such as vanillin and syringaldehyde) in the majority of isolate cultures, suggesting active catabolism of polymeric lignin.

From the OG samples, differing species of the *Pseudomonas*, *Enterobacter* and *Ochrobactrum* genera were isolated (**Chapter 3**, **Section 3.2.2**, **Figure 12**). Whilst two

additional and, again, different *Pseudomonas* spp. were obtained from the EF samples, species of *Paenibacillus*, *Bacillus* and *Brevundiomonas* were also obtained, the latter two possibly being facultative anaerobes. Another *Paenibacillus* sp. was obtained from the compost samples, in addition to a species of *Lysinibacillus*. All of the microorganisms identified above were known for lignin-degradation within their respective environments (except *Micrococcus* sp. from the EF samples) and, whilst some genera overlapped between environments, several were unique to their respective environment. A prime example were the anoxic microorganisms obtained from the EF samples, which were likely adapted to lignin degradation within this niche environment. These findings support the results of previous studies, as stated before, in that lignin-degrading microorganisms vary in identity between environments (Bugg *et al*., 2016). However, this is to be expected from environmental samples which are so different in terms of organic composition and natural conditions such as temperature, pH and oxygen availability. Due to the significant differences between the environmental samples it is, therefore, difficult to compare the identities of key lignin degrading species.

5.3 Novelty of Identified Lignin-Degrading Species

5.3.1 *dyp* Functional Genes Identified within *Ochrobactrum* spp.

Exploring relevant literature, many of the potential lignin-degrading microorganisms that were isolated via enrichment or identified within the metagenome data were found to be known lignin-degraders. For example, the *Pseudomonas* and *Enterobacter* isolates, species from the same genera as the well-known lignin-degraders *Pseudomonas putida* and *Enterobacter lignolyticus*. Whilst these isolates, as well as the other species isolated, could not be identified to species level, the presence of genes coding for laccase (*lmco*) and DyP-type peroxidase (*dyp*), as well as their performance within the enzyme assays and GC-MS analysis, provide evidence for their participation in lignin degradation. From the *lmco* and *dyp* functional gene screening results, it was found that an *Ochrobactrum* isolate had the *dyp* gene, which had not been shown by previous studies. These results are important as they provide evidence to suggest that this species may play a more important role in lignin degradation within the soil environment than previously thought. The two different *Ochrobactrum* spp. investigated in this study were derived from savanna soil and termite-mound soil. Strains of *Ochrobactrum* have been isolated from soil (Taylor *et al*., 2012) and the guts of termite species (Tsegaye *et al*., 2019) previously, and have been shown to synergistically contribute to lignocellulose digestion within the termite gut (Marina *et al*., 2013). Furthermore, *Ochrobactrum* have been investigated for

their potential use in delignification of lignocellulosic material and use in the biofuel industry. A study carried out by Tsegaye *et al* (2018), investigating the delignification of wheat straw by *Ochrobactrum oryzae* BMP03, provided evidence to suggest that this strain actively removed lignin from the lignocellulose complex whilst preserving the cellulose and hemicellulose components. This property is, therefore, of great interest to industrial sectors which utilise the polysaccharide component of lignocellulosic biomass (such as the biofuel sectors). These studies provide significant evidence for the *Ochrobactrum* genus being a key lignin degrading group of bacteria, and highlight the importance of investigating niche environments for industrial candidates. The results of this study showing *Ochrobactrum* to harbour a gene responsible for production of DyP-type peroxidase contributes to this evidence, as this enzyme is heavily associated with key lignin-degrading microorganisms. However, these results imply only functional potential, and further functional analysis would be required to determine DyP peroxidase expression.

5.3.2 Industrial Relevance of EF *Pseudomonas* spp.

In addition to the *Ochrobactrum* isolates, the most notable of the isolates obtained with regards to lignin-degrading potential in this study, were the EF *Pseudomonas* spp.. These isolates harboured both the *dyp* and *lmco* functional genes, demonstrated high peroxidase activity in the enzyme assays and significant catabolic potential in the GC-MS analysis. These properties make the EF *Pseudomonas* spp. interesting candidates for further study, as their high enzymatic and lignin-degrading activity suggest their potential for use in an industrial setting. Whilst it was difficult to compare enzymatic activity to similar studies, due to differences in growth conditions and species used, these isolates significantly out-performed the *Pseudomonas putida* and *Pseudomonas fluorescens* strains used by Salvachua *et al* (2018), the former being known for its lignin-degrading abilities. In addition to their enzymatic properties, analysis of the lignin-derived compounds present after incubation of these isolates with lignin (using GC-MS) revealed high levels of vanillin and syringaldehyde. Vanillin is an important additive used in the production of a range of industrial products (such as foods, perfumes and pharmaceuticals) and is primarily synthesised chemically. Like many commercial markets today, there has recently been a demand for more “natural” and “environmentally friendly” products which has created an interest for researchers to develop alternative methods of producing vanillin. A review conducted by Priefert *et al* (2001) explored various biotechnological methods for vanillin production from lignin or lignin-derived compounds, and stressed the importance of microbial transformation for vanillin production. Several bacterial strains have already been shown to be effective in

vanillin production, such as *Amycolatopsis* sp. used by Rabenhorst and Hopp (1997) which produced more than 10 g/L of vanillin from ferulic acid, however ferulic acid is expensive and chemical synthesis remains a more profitable conversion method (Priefert *et al*., 2001). Whilst the yields of vanillin produced by the EF *Pseudomonas* spp. in this study are unknown, further investigation into these properties could reveal their potential for vanillin production. For industrial use, however, the optimum growth conditions of the EF *Pseudomonas* spp. would require further analysis. From the growth experiments carried out in this study, the optimum growth conditions of the EF *Pseudomonas* spp. were 25 °C temperature and pH 7, and little enzyme activity was observed at 37 °C. Ideally, for industrial use, these isolates would function in harsher conditions (e.g. higher temperature, lower pH). For this reason, further tests would be required to investigate alternative growth conditions, or the potential for genetic engineering.

5.3.3 Potential Lignin Degraders within the Elephant Faeces Microbiome

Whilst all isolates obtained within this study were from genera containing lignin-degrading candidates, some of the microorganisms identified were not known for degradation within their respective environment. To the best of the authors knowledge, the MAGs obtained for species of *Verrucomicrobiaceae* from binning of the EF metagenomes, as well as the *Micrococcus* sp. isolated from the EF enrichment cultures, were not previously known to degrade lignin in the EF environment. As mentioned previously, studies involving CAZy analysis of elephant faeces metagenomes have previously been carried out, however these studies have primarily focussed on cellulose and hemicellulose-degrading communities (Ilmberger *et al*., 2014; Zhang *et* al., 2019; Jakeer *et al*., 2020). These studies, however, identified the *Pseudomonas*, *Brevundimonas*, *Bacillus* and *Paenibacillus* genera to be prevalent within the elephant faeces microbiome, and the genera responsible for lignin-degrading CAZy genes were identified primarily as Proteobacteria (AA2 family of CAZymes, encompassing peroxidases). In this project, one of the EF MAGs which had CAZy genes coding for laccase and peroxidase was of the *Phyllobacteriaceae* family which fall under the Proteobacteria phylum mentioned above. The *Phyllobacteriaceae* family (specifically *Aminobacter*) have previously been enriched with lignin from several environments, and members have been shown to be potential lignin degraders through the presence of genes coding for laccase within their genome (Mendes *et al*., 2021). Whilst the Verrucomicrobia phylum and *Micrococcus* genus have both been implicated as encompassing lignin-degrading species, these groups have not yet been identified as

lignin-degraders within elephant faeces (DeAngelis *et al*., 2010; Taylor *et al*., 2012; Puentez-Tellez *et* al., 2019).

5.4 Abundance of Lignin Degrading Isolates within Environmental Samples

The enrichment methods carried out in this study resulted in the isolation of a varied catalogue of lignin-degrading bacteria. The combination of species identification (via 16S rRNA gene sequencing and analysis) and environmental sample microbial SSU rRNA gene amplicon sequencing, allowed the tentative relative abundances of these species to be explored. In order to determine the relative abundance of isolates within the environmental samples they were isolated from, isolate genus/family identities were searched for within the results obtained from microbial 16S rRNA gene amplicon sequencing of these samples. The results below (in **Table 18**) show the relative abundances of the identified isolates within their respective environments (highlighted in yellow). A variety of *Pseudomonas*, *Ochrobactrum*, *Enterobacter* and *Stenotrophomonas* spp. were isolated from the OG samples. Of these isolates, the *Pseudomonas* genus was low in abundance (0%-0.02%), however the taxonomic groups encompassing *Ochrobactrum* (Rhizobiales order), *Stenotrophomonas* (*Xanthomonadaceae* family) and *Enterobacter* (Gammaproteobacteria class) were higher in abundance (1.17%-4.06%, 0.11%-0.28% and 1.84%-3.01%, respectively). Whilst the specific genera were not identified, it is likely that these genera were included in higher taxonomic ranks but went unidentified within the amplicon sequencing data. The isolates obtained from the EF samples encompassed microorganisms from the *Micrococcus*, *Pseudomonas*, *Bacillus* and *Paenibacillus* genera. Within the EF samples, all microorganisms isolated were identified within the amplicon sequencing data (*Micrococcaceae*, *Pseudomonas*, *Bacillus* and *Paenibacillus*), with the same being observed for the compost isolates. Within the EF samples, the *Pseudomonas* and *Micrococcus* genera were present at 0.04-0.08% and 0.00-0.07% relative abundance, respectively, however the *Bacillus* and *Paenibacillus* genera were present at higher relative abundances, 0.55-0.94% and 0.48-5.36%, respectively. Within the compost samples, the *Paenibacillus* and *Lysinibacillus* genera were present at 0.01-0.39% and 0.07-1.61%, respectively.

Generally, from the table below, it can be seen that the microorganisms isolated in this project are present in low abundances within their respective environments. Despite selecting conditions which were as close as possible to the natural conditions of these environments, these conditions likely created a bias for the enrichment of these

microorganisms, therefore explaining the isolation of only these microorganisms. However, this is to be expected in an enrichment study. Specifically, in the case of the compost samples, the selection of enrichment conditions was difficult due to the variability of conditions during the composting process. The conditions used in this study were selected so as to enrich thermotolerant species. Whilst this analysis provided tentative information on the isolates taxonomic groups, a method to obtain more accurate abundance information would have involved the isolate 16S rRNA gene sequences being searched for within the 16S rRNA gene amplicon sequencing data. However, time did not permit this analysis to be carried out.

Oxidative enzyme assays carried out with these isolates showed that all were capable of producing lignin-degrading peroxidase enzymes, and were potentially involved in the degradation of lignin to some degree. The presence of these isolates in low abundance indicated they were not prominent within their respective environment. It is, therefore, possible that lignin-degrading microorganisms are present in low abundances within the environment as their function in lignin degradation is niche, making it difficult for these microbes to compete with more prominent species.

Table 18 *- Relative abundance of isolate family/genus identified within microbial SSU rRNA gene amplicon sequencing results. Microorganisms isolated from samples highlighted in yellow. TM = Termite-Mound Soil; ZS = Zebra-Grazed Soil; SS = Savanna Soil; EF1 = Elephant 1 Faeces; EF2 = Elephant 2 Faeces; C5-4 = Compost Soil 5-4; C5-5 = Compost Soil 5-5; C6-1 = Compost Soil 6-1.*

	Sample Relative Abundance (%)							
family/genus	**TM**	**ZS**	**SS**	**EF1**	**EF2**	**C5-4**	**C5-5**	**C6-1**
Micrococcaceae	0.84	0.03	0.35	0.07	0	0	0.01	0.03
Pseudomonas	0	0.01	0.02	0.08	0.04	0.64	8.75	0.09
Bacillus	2.24	1.54	0.45	0.94	0.55	0.40	10.40	0.10
Paenibacillus	0.10	0	0.00	0.48	5.36	0.01	0.39	0.02
Lysinibacillus	0	0	0	0	0	0.08	1.61	0.07
Rhizobiales order (*Ochrobactrum*)	1.17	2.49	4.06	4.90	10.99	1.31	0.07	1.23
Xanthomonadaceae family (*Stenotrophomonas*)	0.18	0.11	0.28	0.08	0.05	0.56	0.05	0.28
Gammaproteobacteria class (*Enterobacter*)	1.84	2.16	3.01	8.68	11.98	5.14	22.08	6.23

5.5 Summary

To conclude, this study both confirmed and built upon the current knowledge surrounding microbial (and more specifically, bacterial) lignin degradation within various environments. This project displayed the effectiveness of using a combination of culture-

dependent and culture-independent methodologies, to explore substrate-specific communities at the individual and community levels. The results provided by this study outline the participation of specific bacterial groups/species which contribute towards lignin degradation within different, and in some cases niche, environments. Bacterial groups and their lignin-degrading mechanisms were explored, some of which were previously unknown. Most notable being the *Pseudomonas* spp., which performed well in functional assays and have significant potential for use in industrial lignocellulose bioconversion. Additionally, new information on potential lignin degrading species was revealed, such as the functional potential of OG *Ochrobactrum* spp. to produce DyP peroxidase enzymes. Through microbial SSU rRNA gene amplicon and metagenomic sequencing, the diversity and functionality of lignin degrading communities was investigated, revealing the prominence of genes coding for lignin-active enzymes within the environment. Furthermore, through metagenomic binning, a new potential lignin degrading family of bacteria was revealed.

The in-depth taxonomic profiles and CAZy gene functional profiles generated for each environment, as well as the characterisation of lignin-degrading species from these environments, provide many lucrative channels for further research. With more work, the specific functioning of the lignin-degrading species identified within this study could be revealed, as well as their contributions to lignocellulose catabolism within the environment and their industrial potential. Furthermore, the contributions of specific species and the dynamics involved in lignin catabolism could be further explored through metagenomic binning and deeper analysis of community CAZy gene content.

The research carried out in this project is fundamental to understanding the process of microbial degradation of plant material within natural environments and, ultimately, the fate of lignocellulosic carbon within these environments. Identification of specialized species involved in lignin degradation will inform targeted studies aimed at isolating these species for potential utilization in industrial processes. These species could have a valuable impact on current and future-generation lignocellulose conversion processes, providing an effective means of removing lignin from plant material for production of biofuels or high-value chemicals.

6 Bibliography

Agrawal, A., Kaushik, N., Biswas, S., 2014. Derivatives and applications of lignin – an insight. Scitech J. 1, 30–36.

Ahmad, M., Roberts, J.N., Hardiman, E.M., Singh, R., Eltis, L.D., Bugg, T.D.H., 2011. Identification of DypB from *Rhodococcus jostii* RHA1 as a Lignin Peroxidase. Biochemistry 50, 5096–5107. https://doi.org/10.1021/bi101892z

Ahmad, M., Taylor, C.R., Pink, D., Burton, K., Eastwood, D., Bending, G.D., Bugg, T.D.H., 2010. Development of novel assays for lignin degradation: comparative analysis of bacterial and fungal lignin degraders. Mol. Biosyst. 6, 815. https://doi.org/10.1039/b908966g

Ahmed, A., Babalola, O., McKay, T., 2016. Lignin degradation by two isolated Bacillus sp. strains and their co-culture potential in the production of platform chemicals, in: 3rd International Conference on Sustainable Agriculture and Environment. pp. 10–19.

Aigle, A., Bourgeois, E., Marjolet, L., Houot, S., Patureau, D., Doelsch, E., Cournoyer, B., Galia, W., 2021. Relative weight of organic waste origin on compost and digestate 16S rRNA gene bacterial profilings and related functional inferences. Front. Microbiol. 12, 1–18. https://doi.org/10.3389/fmicb.2021.667043

Akin, D.E., Rigsby, L.L., Sethuraman, a, Morrison, W.H., Gamble, G.R., Eriksson, K.E.L., 1995. Alterations in structure, chemistry and biodegradability of grass lignocellulose treated with the white-rot fungi *Ceriporiopsis subvermispora* and *Cyathus stercoreus*. Appl. Environ. Microbiol. 61, 1591–1598.

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Loman, N.J., Andersson, A.F., Quince, C., 2013. CONCOCT: Clustering cONtigs on COverage and ComposiTion 1–28.

Amann, R.I., Ludwig, W., Schleifer, K., 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. 59, 143–169.

Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W., Huse, S.M., 2009. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable

regions of small-subunit ribosomal RNA Genes. PLoS One 4, 1–9. https://doi.org/10.1371/journal.pone.0006372

Antonopoulos, V.T., Rob, A., Ball, A.S., Wilson, M.T., 2001. Dechlorination of chlorophenols using extracellular peroxidases produced by *Streptomyces albus* ATCC 3005. Enzyme Microb. Technol. 29, 62–69. https://doi.org/10.1016/S0141-0229(01)00357-X

Antunes, L.P., Martins, L.F., Pereira, R.V., Thomas, A.M., Barbosa, D., Lemos, L.N., Silva, G.M.M., Moura, L.M.S., Epamino, G.W.C., Digiampietri, L.A., Lombardi, K.C., Ramos, P.L., Quaggio, R.B., De Oliveira, J.C.F., Pascon, R.C., Da Cruz, J.B., Da Silva, A.M., Setubal, J.C., 2016. Microbial community structure and dynamics in thermophilic composting viewed through metagenomics and metatranscriptomics. Sci. Rep. 6, 1–13. https://doi.org/10.1038/srep38915

Asina, F.N.U., Brzonova, I., Kozliak, E., Kubátová, A., Ji, Y., 2017. Microbial treatment of industrial lignin: Successes, problems and challenges. Renew. Sustain. Energy Rev. 77, 1179–1205. https://doi.org/10.1016/j.rser.2017.03.098

Bandounas, L., Wierckx, N.J.P., Winde, J.H. De, Ruijssenaars, H.J., 2011. Isolation and characterization of novel bacterial strains exhibiting ligninolytic potential. BMC Biotechnology. 11:94.

Bertani, G., 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli.* J. Bacteriol. 62, 293–300. https://doi.org/10.1128/JB.62.3.293-300.1951

Bianchetti, C.M., Harmann, C.H., Takasuka, T.E., Hura, G.L., Dyer, K., Fox, B.G., 2013. Fusion of dioxygenase and lignin-binding domains in a novel secreted enzyme from cellulolytic *Streptomyces* sp. SIRexaa-e. J. Biol. Chem. 288, 18574–18587. https://doi.org/10.1074/jbc.M113.475848

Bourbonnais, R., Paice, M. G., 1990. Oxidation of non-phenolic substrates: An expanded role for laccase in lignin biodegradation. FEBS Letters. Volume 267, Issue 1. Pages 99-102. ISSN 0014-5793. https://doi.org/10.1016/0014-5793(90)80298-W

Blánquez, A., Ball, A., González-Pérez, J., Jiménez-Morillo, N., González-Vila, F., Arias, M.E., Hernández, M., 2017. Laccase SilA from *Streptomyces ipomoeae* CECT 3341 , a

key enzyme for the degradation of lignin from agricultural residues? PLoS One 12, e0187649. https://doi.org/10.6084/m9.figshare.5540119.

Bohacz J., Korniłłowicz-Kowalska T. Modification of post-industrial lignin by fungal strains of the genus *Trichoderma* isolated from different composting stages. J Environ Manage. 2020 Jul 15;266:110573. doi: 10.1016/j.jenvman.2020.110573. Epub 2020 Apr 17. PMID: 32314744.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu YX, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, and Caporaso JG. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 37: 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bridgwater, T., 2006. Review: Biomass for energy. J. Sci. Food Agric. 86, 1755–1768. https://doi.org/10.1002/jsfa

Brown, M.E., Chang, M.C.Y., 2014. Exploring bacterial lignin degradation. Curr. Opin. Chem. Biol. 19, 1–7. https://doi.org/10.1016/j.cbpa.2013.11.015

Buchfink, B., Reuter, K., Drost, H.G., 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods 18, 366–368. https://doi.org/10.1038/s41592-021-01101-x

Budd, K., Gunn, J.C., Finch, T., Klymus, K., Sitati, N., Eggert, L.S., 2020. Effects of diet, habitat, and phylogeny on the fecal microbiome of wild African savanna (Loxodonta africana) and forest elephants (L. cyclotis). Ecol. Evol. 10, 5637–5650. https://doi.org/10.1002/ece3.6305

Bugg, T.D.H., Ahmad, M., Hardiman, E.M., Rahman, R., 2011a. Pathways for degradation of lignin in bacteria and fungi. Nat. Prod. Reports Curr. Dev. Nat. Prod. Chem. 28, 1871–1960. https://doi.org/10.1039/c1np00042j

Bugg, T.D.H., Ahmad, M., Hardiman, E.M., Singh, R., 2011b. The emerging role for bacteria in lignin degradation and bio-product formation. Curr. Opin. Biotechnol. 22, 394–400. https://doi.org/10.1016/j.copbio.2010.10.009

Bugg, T.D.H., Rahmanpour, R., Rashid, G.M.M., 2016. Bacterial enzymes for lignin oxidation and conversion to renewable chemicals. In: Fang, Z., Smith, Jr., R. (eds) Production of Biofuels and Chemicals from Lignin. Biofuels and Biorefineries. Springer, Singapore. 131–146. https://doi.org/10.1007/978-981-10-1965-4

Burgmann, H., Widmer, F., Sigler, W. V, Zeyer, J., 2003. mRNA extraction and reverse transcription - PCR protocol for detection of nifH gene expression by *Azotobacter vinelandii* in soil. Applied and Environmental Microbiology. 2003 Apr;69(4):1928-1935. https://doi.org/10.1128/AEM.69.4.1928

Chandra, R., Raj, A., Purohit, H.J., Kapley, A., 2007. Characterisation and optimisation of three potential aerobic bacterial strains for kraft lignin degradation from pulp paper waste. Chemosphere 67, 839–846. https://doi.org/10.1016/j.chemosphere.2006.10.011

Chandra, R., Raj, A., Purohit, H.J., Kapley, A., 2007. Investigating the degradation process of kraft lignin by β-proteobacterium, *Pandoraea* sp. ISTKB. Chemosphere 67, 839–846. https://doi.org/10.1016/j.chemosphere.2006.10.011

Chantarasiri, A., Boontanom, P., 2017. Decolorization of synthetic dyes by ligninolytic *Lysinibacillus sphaericus* JD1103 isolated from Thai wetland ecosystems. AACL Bioflux 10, 814–819.

Chen, H., Fu, X., 2016. Industrial technologies for bioethanol production from lignocellulosic biomass. Renew. Sustain. Energy Rev. 57, 468–478. https://doi.org/10.1016/j.rser.2015.12.069

Chew, Y.M., Lye, S.F., Md. Salleh, M., Yahya, A., 2018. 16S rRNA metagenomic analysis of the symbiotic community structures of bacteria in foregut, midgut, and hindgut of the wood-feeding termite *Bulbitermes* sp. Symbiosis 1–11. https://doi.org/10.1007/s13199-018-0544-5

Colpa, D.I., Fraaije, M.W., Van Bloois, E., 2014. DyP-type peroxidases: A promising and versatile class of enzymes. J. Ind. Microbiol. Biotechnol. 41, 1–7. https://doi.org/10.1007/s10295-013-1371-6

Coyotzi, S., Pratscher, J., Murrell, J.C., Neufeld, J.D., 2016. Targeted metagenomics of active microbial populations with stable-isotope probing. Curr. Opin. Biotechnol. 41, 1–8. https://doi.org/10.1016/j.copbio.2016.02.017

Dao, T.T., Gentsch, N., Mikutta, R., Sauheitl, L., Shibistova, O., Wild, B., Schnecker, J., Petr, Č., Gittel, A., Lashchinskiy, N., Urich, T., Hana, Š., Richter, A., Guggenberger, G., 2017. Fate of carbohydrates and lignin in north-east Siberian permafrost soils. https://doi.org/10.1016/j.soilbio.2017.10.032

Darjany, L.E., Whitcraft, C.R., Dillon, J.G., 2014. Lignocellulose-responsive bacteria in a southern California salt marsh identified by stable isotope probing. Front. Microbiol. 5, 1–1. https://doi.org/10.3389/fmicb.2014.00263

Datta, R., Kelkar, A., Baraniya, D., Molaei, A., Moulick, A., Meena, R.S., Formanek, P., 2017. Enzymatic degradation of lignin in soil: A review. Sustain. 9. https://doi.org/10.3390/su9071163

de Vries, R.P., Visser, J., Ronald, P., de Vries, R., 2001. *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. Microbiol. Mol. Biol. Rev. 65, 497–522. https://doi.org/10.1128/MMBR.65.4.497

DeAngelis, K.M., Allgaier, M., Chavarria, Y., Fortney, J.L., Hugenholtz, P., Simmons, B., Sublette, K., Silver, W.L., Hazen, T.C., 2011. Characterization of trapped lignin-degrading microbes in tropical forest soil. PLoS One 6. https://doi.org/10.1371/journal.pone.0019306

DeAngelis, K.M., D'Haeseleer, P., Chivian, D., Fortney, J.L., Khudyakov, J., Simmons, B., Woo, H., Arkin, A.P., Davenport, K.W., Goodwin, L., Chen, A., Ivanova, N., Kyrpides,

N.C., Mavromatis, K., Woyke, T., Hazen, T.C., 2011. Complete genome sequence of "*Enterobacter lignolyticus*" SCF1. Stand. Genomic Sci. 5, 69–85. https://doi.org/10.4056/sigs.2104875

DeAngelis, K.M., Gladden, J.M., Allgaier, M., D'haeseleer, P., Fortney, J.L., Reddy, A., Hugenholtz, P., Singer, S.W., Vander Gheynst, J.S., Silver, W.L., Simmons, B.A., Hazen, T.C., 2010. Strategies for enhancing the effectiveness of metagenomic-based enzyme discovery in lignocellulolytic microbial communities. Bioenergy Res. 3, 146–158. https://doi.org/10.1007/s12155-010-9089-z

DeAngelis, K.M., Sharma, D., Varney, R., Simmons, B., Isern, N.G., Markilllie, L.M., Nicora, C., Norbeck, A.D., Taylor, R.C., Aldrich, J.T., Robinson, E.W., 2013. Evidence supporting dissimilatory and assimilatory lignin degradation in *Enterobacter lignolyticus* SCF1. Front. Microbiol. 4, 1–14. https://doi.org/10.3389/fmicb.2013.00280

Drula, E., Garron, M., Dogan, S., Lombard, V., Henrissat, B., Terrapon, N. The carbohydrate-active enzyme database: functions and literature, Nucleic Acids Research, Volume 50, Issue D1, 7 January 2022, Pages D571–D577, https://doi.org/10.1093/nar/gkab1045

Eastman, A.W., Heinrichs, D.E., Yuan, Z.C., 2014. Comparative and genetic analysis of the four sequenced *Paenibacillus polymyxa* genomes reveals a diverse metabolism and conservation of genes relevant to plant-growth promotion and competitiveness. BMC Genomics 15, 1–22. https://doi.org/10.1186/1471-2164-15-851

Enagbonma, B.J., Babalola, O.O., 2019. Environmental sustainability: A review of termite mound soil material and its bacteria. Sustainability, 11, 3847. https://doi.org/10.3390/su11143847

Erlandson, K.A., Delamarre, S.C., Batt, C.A., 2001. Genetic evidence for a defective xylan degradation pathway in *Lactococcus lactis*. Appl. Environ. Microbiol. 67, 1445–1452. https://doi.org/10.1128/AEM.67.4.1445-1452.2001

Fuchs, G., 2008. Anaerobic metabolism of aromatic compounds. Ann. N. Y. Acad. Sci. 1125, 82–99. https://doi.org/10.1196/annals.1419.010

Gabrielii, I., Gatenholm, P., Glasser, W.G., Jain, R.K., Kenne, L., 2000. Separation, characterization and hydrogel-formation of hemicellulose from aspen wood. Carbohydr. Polym. 43, 367–374.

Galbe, M., Zacchi, G., 2007. Pretreatment of lignocellulosic materials for efficient bioethanol production. Adv. Biochem. Eng. Biotechnol. 108, 41–65. https://doi.org/10.1007/10_2007_070

Givens, C.E., Bowers, J.C., Depaola, A., Hollibaugh, J.T., Jones, J.L., 2014. Occurrence and distribution of *Vibrio vulnificus* and *Vibrio parahaemolyticus* - potential roles for fish, oyster, sediment and water. Lett. Appl. Microbiol. 58, 503–510. https://doi.org/10.1111/lam.12226

Goicoechea, J.L., Ammiraju, J.S.S., Marri, P.R., Chen, M., Jackson, S., Yu, Y., Rounsley, S., Wing, R.A., 2010. The future of rice genomics: Sequencing the collective *Oryza* genome. Rice 3, 89–97. https://doi.org/10.1007/s12284-010-9052-9

Gong, G., Zhou, S., Luo, R., Gesang, Z., Suolang, S., 2020. Metagenomic insights into the diversity of carbohydrate-degrading enzymes in the yak fecal microbial community. BMC Microbiol. 20, 1–15. https://doi.org/10.1186/s12866-020-01993-3

Graham ED, Heidelberg JF, Tully BJ. (2017) BinSanity: unsupervised clustering of environmental microbial assemblies using coverage and affinity propagation. PeerJ 5:e3035 https://doi.org/10.7717/peerj.3035

Graham, J.E., Clark, M.E., Nadler, D.C., Huffer, S., Chokhawala, H.A., Rowland, S.E., Blanch, H.W., Clark, D.S., Robb, F.T., 2011. Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. Nat. Commun. 2. https://doi.org/10.1038/ncomms1373

Granja-Travez, R.S., Persinoti, G.F., Squina, F.M., Bugg, T.D.H., 2020. Functional genomic analysis of bacterial lignin degraders: diversity in mechanisms of lignin oxidation and metabolism. Appl. Microbiol. Biotechnol. 104, 3305–3320. https://doi.org/10.1007/s00253-019-10318-y

Griffiths, R.I., Whiteley, A.S., Donnell, A.G.O., Bailey, M.J., 2000. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA-

and rRNA-based microbial community composition. Applied and Environmental Microbiology. 66, 5488–5491.

Guerriero, G., Hausman, J.F., Strauss, J., Ertan, H., Siddiqui, K.S., 2016. Lignocellulosic biomass: Biosynthesis, degradation, and industrial utilization. Eng. Life Sci. 16, 1–16. https://doi.org/10.1002/elsc.201400196

Hayat, R., Ahmed, I., Paek, J., Sin, Y., Ehsan, M., Iqbal, M., Yokota, A., Chang, Y.H., 2014. *Lysinibacillus composti* sp. nov., isolated from compost. Ann. Microbiol. 64, 1081–1088. https://doi.org/10.1007/s13213-013-0747-1

Hermosilla, E., Schalchli, H., Mutis, A., Diez, M.C., 2017. Combined effect of enzyme inducers and nitrate on selective lignin degradation in wheat straw by *Ganoderma lobatum.* Environ. Sci. Pollut. Res. 1–13. https://doi.org/10.1007/s11356-017-9841-4

Hrdlicka, L., Skulcova, A., Haz, A., 2014. Degradation of lignin via Fenton reaction.

Huang, X.F., Santhanam, N., Badri, D. V., Hunter, W.J., Manter, D.K., Decker, S.R., Vivanco, J.M., Reardon, K.F., 2013. Isolation and characterization of lignin-degrading bacteria from rainforest soils. Biotechnol. Bioeng. 110, 1616–1626. https://doi.org/10.1002/bit.24833

Huson, D, Albrecht, B, Bagci, C, Bessarab, I, Gorska, A, Jolic, D, and Williams, RB (2018). MEGAN- LR: New algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. Biology Direct, 13(6).

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119. https://doi.org/10.1186/1471-2105-11-119

Ilmberger, N., Güllert, S., Dannenberg, J., Rabausch, U., Torres, J., Wemheuer, B., Alawi, M., Poehlein, A., Chow, J., Turaev, D., Rattei, T., Schmeisser, C., Salomon, J., Olsen, P.B., Daniel, R., Grundhoff, A., Borchert, M.S., Streit, W.R., 2014. A comparative metagenome survey of the fecal microbiota of a breast-and a plant-fed asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes. PLoS One 9, 1–12. https://doi.org/10.1371/journal.pone.0106707

Ishii, S., Yamamoto, M., Kikuchi, M., Oshima, K., Hattori, M., Otsuka, S., Senoo, K., 2009. Microbial populations responsive to denitrification-inducing conditions in rice

paddy soil, as revealed by comparative 16S rRNA gene analysis. Appl. Environ. Microbiol. 75, 7070–7078. https://doi.org/10.1128/AEM.01481-09

Jackson, C.A., Couger, M.B., Prabhakaran, M., Ramachandriya, K.D., Canaan, P., Fathepure, B.Z., 2017. Isolation and characterization of *Rhizobium* sp. strain YS-1r that degrades lignin in plant biomass. J. Appl. Microbiol. 122, 940–952. https://doi.org/10.1111/jam.13401

Jakeer, S., Varma, M., Sharma, J., Mattoo, F., Gupta, D., Singh, J., Kumar, M., Gaur, N.A., 2020. Metagenomic analysis of the fecal microbiome of an adult elephant reveals the diversity of CAZymes related to lignocellulosic biomass degradation. Symbiosis 81, 209–222. https://doi.org/10.1007/s13199-020-00695-8

Janusz, G., Pawlik, A., Sulej, J., Świderska-Burek, U., Jarosz-Wilkołazka, A., Paszczyński, A., 2017. Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. FEMS Microbiol. Rev. 1–22. https://doi.org/10.1093/femsre/fux049

Jiménez, D.J., Korenblum, E., Van Elsas, J.D., 2014. Novel multispecies microbial consortia involved in lignocellulose and 5-hydroxymethylfurfural bioconversion. Appl. Microbiol. Biotechnol. 98, 2789–2803. https://doi.org/10.1007/s00253-013-5253-7

Joseph, B., Kuddus, M., Ramteke, P. W., 2014. Extracellular alkaline thermostable laccase from *Micrococcus* species: partial purification and characterization. Current Biotechnology, Volume 3, Number 2, pp. 145-151(7). Bentham Science Publishers. https://doi.org/10.2174/22115501113026660035

Kai, D., Tan, M.J., Chee, P.L., Chua, Y.K., Yap, Y.L., Loh, X.J., 2016. Towards lignin-based functional materials in a sustainable world. Green Chem. 18, 1175–1200. https://doi.org/10.1039/C5GC02616D

Kaiser, K., Benner, R., 2012. Characterization of lignin by gas chromatography and mass spectrometry using a simplified CuO oxidation Method. Anal. Chem. 84, 1, 459–464.

Kameshwar, A.K.S., Qin, W., 2016. Recent developments in using advanced sequencing technologies for the genomic studies of lignin and cellulose degrading microorganisms. Int. J. Biol. Sci. 12, 156–171. https://doi.org/10.7150/ijbs.13537

Kamimura, N., Takahashi, K., Mori, K., Araki, T., Fujita, M., Higuchi, Y., Masai, E., 2017. Bacterial catabolism of lignin-derived aromatics: New findings in a recent decade: Update on bacterial lignin catabolism. Environ. Microbiol. Rep. 9, 679–705. https://doi.org/10.1111/1758-2229.12597

Kang, D.D., Froula, J., Egan, R., Wang, Z., 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities 1–15. https://doi.org/10.7717/peerj.1165

Kellner, H., Luis, P., Zimdars, B., Kiesel, B., Buscot, F., 2008. Diversity of bacterial laccase-like multicopper oxidase genes in forest and grassland Cambisol soil samples. Soil Biol. Biochem. 40, 638–648. https://doi.org/10.1016/j.soilbio.2007.09.013

Kerem, Z., Jensen, K.A., Hammel, K.E., 1999. Biodegradative mechanism of the brown rot basidiomycete Gloeophyllum trabeum: Evidence for an extracellular hydroquinone-driven fenton reaction. FEBS Lett. 446, 49–54. https://doi.org/10.1016/S0014-5793(99)00180-5

Köhler, T., Dietrich, C., Scheffrahn, R.H., Brune, A., 2012. High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). Appl. Environ. Microbiol. 78, 4691–4701. https://doi.org/10.1128/AEM.00683-12

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. 79, 5112–5120. https://doi.org/10.1128/aem.01043-13

Kumar, M., Singh, J., Singh, M.K., Singhal, A., Thakur, I.S., 2015. Investigating the degradation process of kraft lignin by β-proteobacterium, *Pandoraea* sp. ISTKB 15690–15702. https://doi.org/10.1007/s11356-015-4771-5

Kumar, P., Barrett, D.M., Delwiche, M.J., Stroeve, P., 2009. Methods for pretreatment of lignocellulosic biomass for efficient hydrolysis and biofuel production. Ind. Eng. Chem. (Analytical Ed. 48, 3713–3729. https://doi.org/10.1021/ie801542g

Kumar, R., Singh, S., Singh, O. V., 2008. Bioconversion of lignocellulosic biomass: Biochemical and molecular perspectives. J. Ind. Microbiol. Biotechnol. 35, 377–391. https://doi.org/10.1007/s10295-008-0327-8

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096

Lambertz, C., Ece, S., Fischer, R., Commandeur, U., 2016. Progress and obstacles in the production and application of recombinant lignin-degrading peroxidases. Bioengineered 7, 145–154. https://doi.org/10.1080/21655979.2016.1191705

Lane, D.J., 1991. 16S/23S rRNA sequencing. E. Stackebrandt, M. Goodfellow (Eds.), Nucleic acid techniques in bacterial systematics, Wiley, Chichester, England (1991), pp. 205-248

Laureano-pérez, L., Teymouri, F., Alizadeh, H., Dale, B.E., 2005. Understanding factors that limit enzymatic hydrolysis of biomass. Appl. Biochem. Biotechnol. 121–124, 1081–1099. https://doi.org/10.1007/978-1-59259-991-2

Lee, K.H., Wi, S.G., Singh, A.P., Kim, Y.S., 2004. Micromorphological characteristics of decayed wood and laccase produced by the brown-rot fungus *Coniophora puteana*. J. Wood Sci. 50, 281–284. https://doi.org/10.1007/s10086-003-0558-2

Li, D., Liu, C., Luo, R., Sadakane, K., Lam, T., 2015. Sequence analysis MEGAHIT : an ultra-fast single- node solution for large and complex metagenomics assembly via succinct de Bruijn graph 31, 1674– 1676. https://doi.org/10.1093/bioinformatics/btv033

Li, H.-F., Fujisaki, I., Su, N.-Y., 2013. Predicting habitat suitability of *Coptotermes gestroi* (*Isoptera: Rhinotermitidae*) with species distribution models. J. Econ. Entomol. 106, 311–21. https://doi.org/10.1603/ec12309

Li, J., Yuan, H., Yang, J., 2009. Bacteria and lignin degradation. Front. Biol. China 4, 29–38. https://doi.org/10.1007/s11515-008-0097-8

Linhart C., Shamir R., 2007. Degenerate Primer Design. In: Yuryev A. (eds) PCR Primer Design. Methods in Molecular Biology™, vol 402. Humana Press. https://doi.org/10.1007/978-1-59745-528-2_11

Liu, Y., Hu, T., Wu, Z., Zeng, G., Huang, D., Shen, Y., He, X., Lai, M., He, Y., 2014. Study on biodegradation process of lignin by FTIR and DSC. Environ. Sci. Pollut. Res. 21, 14004–14013. https://doi.org/10.1007/s11356-014-3342-5

Lu, L., Zeng, G., Fan, C., Zhang, J., Chen, A., Chen, M., Jiang, M., Yuan, Y., Wu, H., Lai, M., He, Y., 2014. Diversity of two-domain laccase-like multicopper oxidase genes in *Streptomyces* spp.: Identification of genes potentially involved in extracellular activities and lignocellulose degradation during composting of agricultural waste. Appl. Environ. Microbiol. 80, 3305–3314. https://doi.org/10.1128/AEM.00223-14

Lv, Y., Chen, Y., Sun, S., Hu, Y., 2014. Interaction among multiple microorganisms and effects of nitrogen and carbon supplementations on lignin degradation. Bioresour. Technol. 155, 144–151. https://doi.org/10.1016/j.biortech.2013.12.012

Malherbe, C., Cloete, T.E., 2002. Lignocellulose biodegradation: Fundamentals and applications. Rev. Environ. Sci. Bio/Technology 1, 105–114. https://doi.org/10.1023/A

Manefield, M., Whiteley, A.S., Griffiths, R.I., Bailey, M.J., 2002. RNA stable isotope probing , a novel means of linking microbial community function to phylogeny. Appl. Environ. Microbiol. 68, 5367–5373. https://doi.org/10.1128/AEM.68.11.5367

Manter, D.K., Hunter, W.J., Vivanco, J.M., 2011. *Enterobacter soli* sp. nov.: A lignin-degrading γ-Proteobacteria isolated from soil. Curr. Microbiol. 62, 1044–1049. https://doi.org/10.1007/s00284-010-9809-9

Marina, G., Chacko, D., Lo S., Mathew, G., Yang, J., Mahalingam, J., Mahamadali, T., Huang, C., 2013. Synergistic collaboration of gut symbionts in *Odontotermes formosanus* for lignocellulosic degradation and bio-hydrogen production. Bioresour Technol 145:337–344. https://doi.org/10.1016/j.biortech.2012.12.055

Martínez, Á.T., Speranza, M., Ruiz-Dueñas, F.J., Ferreira, P., Camarero, S., Guillén, F., Martínez, M.J., Gutiérrez, A., Del Río, J.C., 2005. Biodegradation of lignocellulosics: Microbial, chemical, and enzymatic aspects of the fungal attack of lignin. Int. Microbiol. 8, 195–204. https://doi.org/im2305029

Mathews, S.L., Pawlak, J., Grunden, A.M., 2015. Bacterial biodegradation and bioconversion of industrial lignocellulosic streams. Appl. Microbiol. Biotechnol. 99, 2939–2954. https://doi.org/10.1007/s00253-015-6471-y

Mendes, I.V., Garcia, M.B., Bitencourt, A.C.A., Santana, R.H., de Castro Lins, P., Silveira, R., Simmons, B.A., Gladden, J.M., Kruger, R.H., Quirino, B.F., 2021. Bacterial diversity dynamics in microbial consortia selected for lignin utilization. PLoS One 16, 1–20. https://doi.org/10.1371/journal.pone.0255083

Mengistu, T., Gebrekidan, H., Kibret, K., Woldetsadik, K., Shimelis, B., Yadav, H., 2018. Comparative effectiveness of different composting methods on the stabilization, maturation and sanitization of municipal organic solid wastes and dried faecal sludge mixtures. Environ. Syst. Res. 6. https://doi.org/10.1186/s40068-017-0079-4

Mongkolthanaruk, W., Tongbopit, S., Bhoonobtong, A., 2012. Independent behaviour of bacterial laccases to inducers and metal ions during production and activity. African J. Biotechnol. 11, 9391–9398. https://doi.org/10.5897/ajb11.3042

Murugesan, K., Nam, I.H., Kim, Y.M., Chang, Y.S., 2007. Decolorization of reactive dyes by a thermostable laccase produced by *Ganoderma lucidum* in solid state culture. Enzyme Microb. Technol. 40, 1662–1672. https://doi.org/10.1016/j.enzmictec.2006.08.028

Neufeld, J.D., Vohra, J., Dumont, M.G., Lueders, T., Manefield, M., Friedrich, M.W., Murrell, J.C., 2007. DNA stable-isotope probing. Nat. Protoc. 2, 860–866. https://doi.org/10.1038/nprot.2007.109

Palonen, H., Thomsen, A.B., Tenkanen, M., Schmidt, A.S., Viikari, L., 2004. Evaluation of wet oxidation pretreatment for enzymatic hydrolysis of softwood. Appl. Biochem. Biotechnol. 117, 1–17. https://doi.org/10.1385/ABAB:117:1:01

Papinutti, V.L., Diorio, L.A., Forchiassin, F., 2003. Production of laccase and manganese peroxidase by *Fomes sclerodermeus* grown on wheat bran. Soc. Ind. Biotechnol. 30, 157–160. https://doi.org/10.1007/s10295-003-0025-5

Parks, DH., Imelfort, M., Skennerton, CT., Hugenholtz, P., Tyson, GW., 2014. Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research, 25: 1043-1055.

Passardi, F., Cosio, C., Penel, C., Dunand, C., 2005. Peroxidases have more functions than a Swiss army knife. Plant Cell Rep. 24, 255–265. https://doi.org/10.1007/s00299-005-0972-6

Pepe-Ranney, C., Campbell, A.N., Koechli, C., Berthrong, S., Buckley, D.H., 2015. Unearthing the microbial ecology of soil carbon cycling with DNA-SIP Introductory Paragraph 107, 33–1. https://doi.org/10.1101/022483

Pometto, A.L., Crawford, D.L., 1986. Effects of pH on lignin and cellulose degradation by *Streptomyces viridosporus*. Appl. Environmental Microbiol. 52, 246–250.

Poomima, P., Velan, M. 2018. A novel laccase producing *Brevundimonas* sp. MVSP from paper and pulp industry waste water. Journal of Environmental Biology. vol. 39, no. 4, pp. 447-453.

Prabhakaran, M., Couger, M.B., Jackson, C.A., Weirick, T., Fathepure, B.Z., 2015. Genome sequences of the lignin-degrading *Pseudomonas* sp. strain YS-1p and *Rhizobium* sp. strain YS-1r isolated from decaying wood. Genome Announc. 3, e00019-15. https://doi.org/10.1128/genomeA.00019-15

Priefert, H., Rabenhorst, J., Steinbüchel, A., 2001. Biotechnological production of vanillin. Appl. Microbiol. Biotechnol. 56, 296–314. https://doi.org/10.1007/s002530100687

Puentes-téllez, P.E., Salles, J.F., 2019. Dynamics of abundant and rare bacteria during degradation of lignocellulose from sugarcane biomass. Microb. Ecol. Feb;79(2):312-325.

Ramachandra, M., Crawford, D.L., Hertel, G., 1988. Characterization of an extracellular lignin peroxidase of the lignocellulytic Actinomycete *Streptomyces viridosporus*. Appl. Environ. Microbiol. 54, 3057–3063.

Reimer, L.C., Carbasse, J.S., Koblitz, J., Ebeling, C., Podstawka, A., Overmann, J. BacDive in 2022: the knowledge base for standardized bacterial and archaeal data. Nucleic Acids Research; database issue 2022.

Ruiz-Dueñas, F.J., Martínez, Á.T., 2009. Microbial degradation of lignin: How a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. Microb. Biotechnol. 2, 164–177. https://doi.org/10.1111/j.1751-7915.2008.00078.x

Salvachúa, D., Karp, E.M., Nimlos, C.T., Vardon, D.R., Beckham, G.T., 2015. Towards lignin consolidated bioprocessing: simultaneous lignin depolymerization and product generation by bacteria. Green Chem. 17, 4951–4967. https://doi.org/10.1039/C5GC01165E

Santhanam, N., Vivanco, J.M., Decker, S.R., Reardon, K.F., 2011. Expression of industrially relevant laccases: Prokaryotic style. Trends Biotechnol. 29, 480–489. https://doi.org/10.1016/j.tibtech.2011.04.005

Seemann, T., 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Shi, Y., Chai, L., Tang, C., Yang, Z., Zhang, H., Chen, R., Chen, Y., Zheng, Y., 2013. Characterization and genomic analysis of kraft lignin biodegradation by the Beta-proteobacterium *Cupriavidus basilensis* B-8. Biotechnol. Biofuels 6, 1. https://doi.org/10.1186/1754-6834-6-1

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 3, 836–843 (2018). https://doi.org/10.1038/s41564-018-0171-1

Souza, W.R. De, 2013. Microbial degradation of lignocellulosic biomass, InTech. https://doi.org/10.5772/54325

Spiker, J.K., Crawford, D.L., Thiel, E.C., 1992. Oxidation of phenolic and non-phenolic substrates by the lignin peroxidase of *Streptomyces viridosporus* T7A. Appl. Microbiol. Biotechnol. 37, 518–523. https://doi.org/10.1007/BF00180980

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H.-W., & Richards, T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Molecular Ecology, 19 Suppl 1, 21–31. http://doi.org/10.1111/j.1365-294X.2009.04480.x

Sun, S., Sun, S., Cao, X., Sun, R., 2016. The role of pretreatment in improving the enzymatic hydrolysis of lignocellulosic materials. Bioresour. Technol. 199, 49–58. https://doi.org/10.1016/j.biortech.2015.08.061

Sun, Y., Cheng, J., 2002. Hydrolysis of lignocellulosic materials for ethanol production: a review. Bioresour. Technol. 83, 1–11. https://doi.org/10.1016/S0960-8524(01)00212-7

Taniguchi, M., Takahashi, D., Watanabe, D., Sakai, K., Hoshino, K., Kouya, T., Tanaka, T., 2010. Effect of steam explosion pretreatment on treatment with *Pleurotus ostreatus* for the enzymatic hydrolysis of rice straw. J. Biosci. Bioeng. 110, 449–452. https://doi.org/10.1016/j.jbiosc.2010.04.014

Taylor, C.R., Hardiman, E.M., Ahmad, M., Sainsbury, P.D., Norris, P.R., Bugg, T.D.H., 2012. Isolation of bacterial strains able to metabolize lignin from screening of environmental samples. J. Appl. Microbiol. 113, 521–530. https://doi.org/10.1111/j.1365-2672.2012.05352.x

Teng, F., Darveekaran Nair, S.S., Zhu, P. et al. Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. Sci Rep 8, 16321 (2018). https://doi.org/10.1038/s41598-018-34294-x

Thakur, V.K., Thakur, M.K., 2015. Recent advances in green hydrogels from lignin: A review. Int. J. Biol. Macromol. 72, 834–847. https://doi.org/10.1016/j.ijbiomac.2014.09.044

Thomas, T., Gilbert, J., Meyer, F., 2012. Metagenomics - a guide from sampling to data analysis. Microb. Inform. Exp. 2, 3. https://doi.org/10.1186/2042-5783-2-3

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., …, Jansson, J. K., Gilbert, J. A., Knight, R., & The Earth Microbiome Project Consortium. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. Nature, 551:457-463. doi:10.1038/nature24621

Tian, J.H., Pourcher, A.M., Bouchez, T., Gelhaye, E., Peu, P., 2014. Occurrence of lignin degradation genotypes and phenotypes among prokaryotes. Appl. Microbiol. Biotechnol. 98, 9527–9544. https://doi.org/10.1007/s00253-014-6142-4

Tian, J., Pourcher, A., Klingelschmitt, F., Roux, S. Le, Peu, P., 2016. Class P dye-decolorizing peroxidase gene: Degenerated primers design and phylogenetic analysis. J. Microbiol. Methods 130, 148–153. https://doi.org/10.1016/j.mimet.2016.09.016

Tian, J.H., Pourcher, A.M., Peu, P., 2016. Isolation of bacterial strains able to metabolize lignin and lignin-related compounds. Lett. Appl. Microbiol. 63, 30–37. https://doi.org/10.1111/lam.12581

Tsegaye, B., Balomajumder, C., Roy, P., 2019. Isolation and characterization of novel lignolytic, cellulolytic, and hemicellulolytic bacteria from wood-feeding termite *Cryptotermes brevis*. Int. Microbiol. 22, 29–39. https://doi.org/10.1007/s10123-018-0024-z

Tuncer, M., Kuru, A., Isikli, M., Sahin, N., Çelenk, F.G., 2004. Optimization of extracellular endoxylanase, endoglucanase and peroxidase production by *Streptomyces* sp. F2621 isolated in Turkey. J. Appl. Microbiol. 97, 783–791. https://doi.org/10.1111/j.1365-2672.2004.02361.x

Tuomela, M., Vikman, M., Hatakka, A., It, M., 2000. Biodegradation of lignin in a compost environment : a review. Bioresource Technology. 72, 2, 169-183.

Ueda, J., Kurosawa, N., 2015. Characterization of an extracellular thermophilic chitinase from *Paenibacillus thermoaerophilus* strain TC22-2b isolated from compost. World J. Microbiol. Biotechnol. 31, 135–143. https://doi.org/10.1007/s11274-014-1754-5

Vaz-Moreira, I., Faria, C., Nobre, M.F., Schumann, P., Nunes, O.C., Manaia, C.M., 2007. *Paenibacillus humicus* sp. nov., isolated from poultry litter compost. Int. J. Syst. Evol. Microbiol. 57, 2267–2271. https://doi.org/10.1099/ijs.0.65124-0

Vikman, M., Karjomaa, S., Kapanen, A., Wallenius, K., Itävaara, M., 2002. The influence of lignin content and temperature on the biodegradation of lignocellulose in composting conditions. Appl. Microbiol. Biotechnol. 59, 591–598. https://doi.org/10.1007/s00253-002-1029-1

Wan, C., Li, Y., 2010. Microbial pretreatment of corn stover with *Ceriporiopsis subvermispora* for enzymatic hydrolysis and ethanol production. Bioresour. Technol. 101, 6398–6403. https://doi.org/10.1016/j.biortech.2010.03.070

Wang, C., Dong, D., Wang, H., Müller, K., Qin, Y., Wang, H., Wu, W., 2016. Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition. Biotechnol. Biofuels 9, 22. https://doi.org/10.1186/s13068-016-0440-2

Wang, Y., Liu, Q., Yan, L., Gao, Y., Wang, Y., Wang, W., 2013. A novel lignin degradation bacterial consortium for efficient pulping. Bioresour. Technol. 139, 113–119. https://doi.org/10.1016/j.biortech.2013.04.033

Weissenböck, N. M., Schober, F., Fluch, G., Weiss, C., Paumann, T., Schwarz, C., Arnold, W., 2010. Reusable biotelemetric capsules: A convenient and reliable method for measuring core body temperature in large mammals during gut passage. Journal of Thermal Biology. Volume 35, Issue 3. Pages 147-153. https://doi.org/10.1016/j.jtherbio.2010.02.001

Wenzel, M., Schonig, I., Berchtold, M., Kampfer, P., Konig, H., 2002. Aerobic and facultatively anaerobic cellulolytic bacteria from the gut of the termite *Zootermopsis angusticollis*. J. Appl. Microbiol. 92, 32–40.

Wilhelm, R.C., Singh, R., Eltis, L.D., Mohn, W.W., 2018. Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing. ISME J. 13, 413–429. https://doi.org/10.1038/s41396-018-0279-6

Woo, H.L., Hazen, T.C., Simmons, B.A., DeAngelis, K.M., 2014. Enzyme activities of aerobic lignocellulolytic bacteria isolated from wet tropical forest soils. Syst. Appl. Microbiol. 37, 60–67. https://doi.org/10.1016/j.syapm.2013.10.001

Woo, P.C.Y., Lau, S.K.P., Teng, J.L.L., Tse, H., Yuen, K.Y., 2008. Then and now: Use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. Clin. Microbiol. Infect. 14, 908–934. https://doi.org/10.1111/j.1469-0691.2008.02070.x

Wood, D.E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. Genome Biol 20, 257 (2019). https://doi.org/10.1186/s13059-019-1891-0

Wu, YW., Tang, YH., Tringe, S.G. et al. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome 2, 26 (2014). https://doi.org/10.1186/2049-2618-2-26

Xu, Z., Qin, L., Cai, M., Hua, W., Jin, M., 2018. Biodegradation of kraft lignin by newly isolated *Klebsiella pneumoniae*, *Pseudomonas putida*, and *Ochrobactrum tritici* strains. Environ. Sci. Pollut. Res. 1–11. https://doi.org/10.1007/s11356-018-1633-y

Yan, J., Wang, L., Hu, Y., Tsang, Y.F., Zhang, Y., Wu, J., Fu, X., Sun, Y., 2018. Plant litter composition selects different soil microbial structures and in turn drives different litter decomposition pattern and soil carbon sequestration capability. Geoderma 319, 194–203. https://doi.org/10.1016/j.geoderma.2018.01.009

Yu HY, Zeng GM, Huang GH, Huang DL, Chen YN. Lignin degradation by *Penicillium simplicissimum*. Huan Jing ke Xue, Huanjing Kexue. 2005 Mar;26(2):167-171. PMID: 16004322.

Zakzeski, J., Bruijnincx, P.C.A., Jongerius, A.L., Weckhuysen, B.M., 2010. The catalytic valorization of lignin for the production of renewable chemicals. Chem. Rev. 110, 3552–3599. https://doi.org/10.1021/cr900354u

Zhan, M., Wang, L., Xie, C., Fu, X., Zhang, S., Wang, A., Zhou, Y., Xu, C., Zhang, H., 2020. Succession of gut microbial structure in twin giant pandas during the dietary change stage and its role in polysaccharide metabolism. Front. Microbiol. 11. https://doi.org/10.3389/fmicb.2020.551038

Zhang, C., Xu, B., Huang, Z., Lu, T., 2019. Metagenomic analysis of the fecal microbiomes of wild asian elephants reveals microflora and enzymes that mainly digest hemicellulose. J. Microbiol. Biotechnol. 29, 1255–1265. https://doi.org/10.4014/jmb.1904.04033

Zhou, H., Guo, W., Xu, B., Teng, Z., Tao, D., Lou, Y., Gao, Y., 2017. Screening and identification of lignin-degrading bacteria in termite gut and the construction of LiP-expressing recombinant *Lactococcus lactis*. Microb. Pathog. 112, 63–69. https://doi.org/10.1016/j.micpath.2017.09.047

7 Appendices

7.1 *Unfinished lmco Functional Gene* PCR and qPCR Work

In addition to the *dyp* functional gene, a *lmco* functional gene qPCR method was in development before the COVID-19 pandemic. A *lmco* PCR amplification product obtained from compost isolate C2 (*Paenibacillus* sp.) (according to protocol detailed in **Methods and Materials Section 2.6.1.4**) was sent for Sanger sequencing, the data from which underwent phylogenetic analysis via MegaX sequence editing and alignment with closest neighbouring sequences obtained from the GenBank database in NCBI using BLASTn searches. From the *lmco* gene phylogenetic tree (**Figure 51**), the C2 *lmco* amplification product can be seen to align closest to the laccase genes of uncultured bacterial isolates and to a gene within the genome of *Paenibacillus mucilaginosus* (same genus as isolate C2).

This amplification product was diluted to 10^8 gene copy numbers, from which a qPCR standard dilution series was to be created (10^7-10^1), however time limitations prevented the continuation of this work. This standard dilution series would subsequently have undergone qPCR analysis to investigate possible amplification bias of the *lmco* gene (involving analysis of the amplification and melt curves). Furthermore, the *lmco* qPCR method would have been tested with environmental sample nucleic acid extracts in order to determine *lmco* gene abundance (relative to bacterial 16S rRNA gene abundance) within environmental samples. Furthermore, this method would have then been used to identify ^{13}C labelling within RNA SIP cDNA fractions.

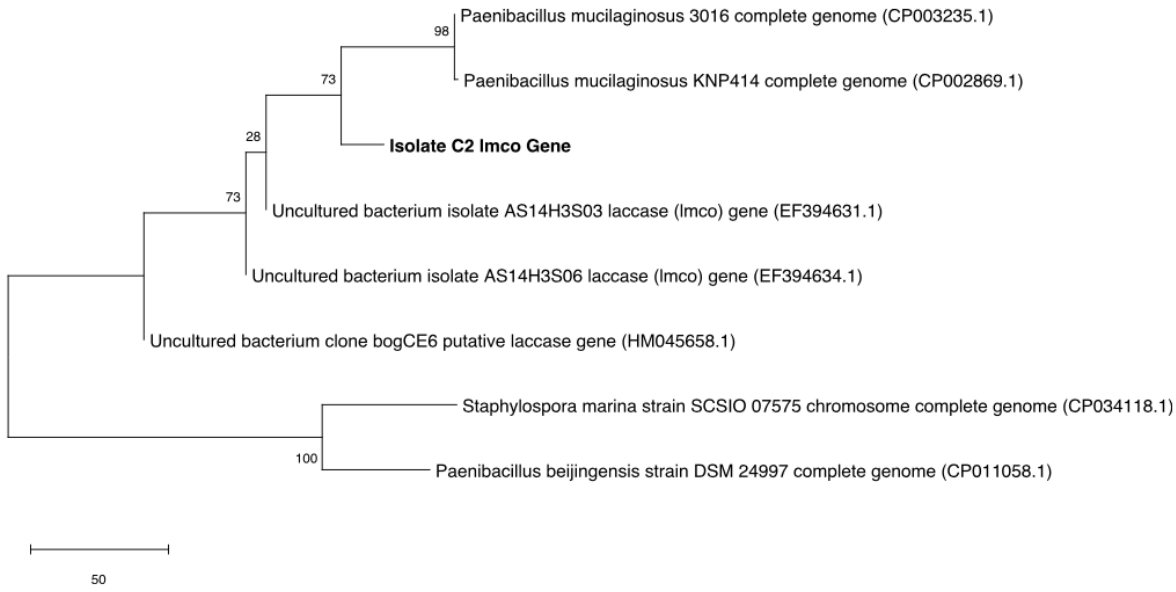

Figure 51 *- Compost isolate C2 lmco gene phylogenetic neighbour-joining tree with neighbouring sequences obtained from the GenBank database in NCBI using BLASTn searches. Isolate sequences highlighted in bold font. Tree generated using MEGAX with bootstrap analysis of 500 replicates.*

7.2 *dyp* qPCR Standard Preparation

In order to obtain a standard for qPCR analysis using the *dyp* primer set, PCR with primers targeting the *dyp* functional gene was carried out with all Olduvai Gorge and elephant faeces isolates (PCR methodology detailed in **Section 2.6.1.3**). Products were successfully obtained and purified from OG isolates T3A2, T5C2, and T7A2 and EF isolate T4C1 (gel image of *dyp* PCR products in **Figure 52** below). *dyp* PCR products obtained from isolates were sent for Sanger sequencing by Source BioScience. Good quality sequences (long sequence length without background noise) were searched against the NCBI GenBank nucleotide database using blastn searches. The isolate *dyp* sequences were aligned with similar/neighbouring sequences obtained from blastn searches and a neighbour-joining phylogenetic tree was constructed from this alignment using MEGA7 software (**Figure 53**). The phylogenetic tree shows that the *dyp* gene products obtained from these four isolates vary phylogenetically from one another, however OG-T3A2 and T5C2 (both *Pseudomonas* spp.) are closer related compared to EF-T4C1 (unknown identity) and OG-T7A2 (*Enterobacter* sp.).

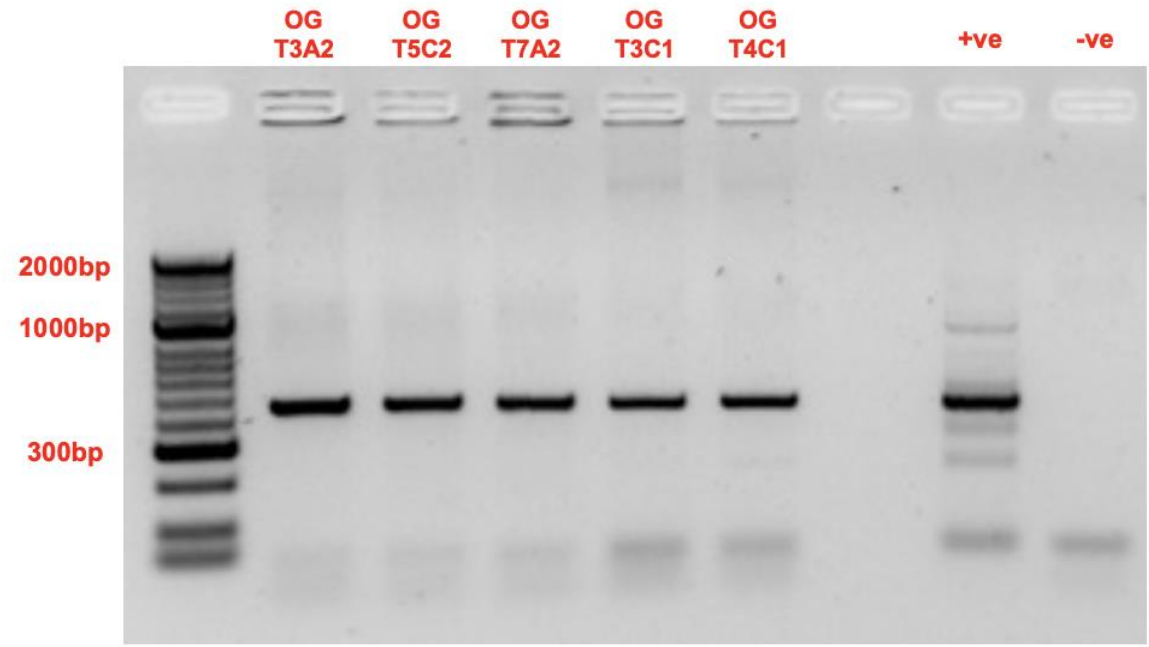

Figure 52 *- Agarose gel electrophoresis image of dyp functional gene PCR products obtained from Olduvai Gorge and Elephant Faeces isolates OG T3A2, OG T5C2, OG T7A2, EF T3C1 and EF T4C1. Positive control = previous dyp PCR product from EF T4C1; negative control = water. DNA ladder = 50 bp Hyperladder.*

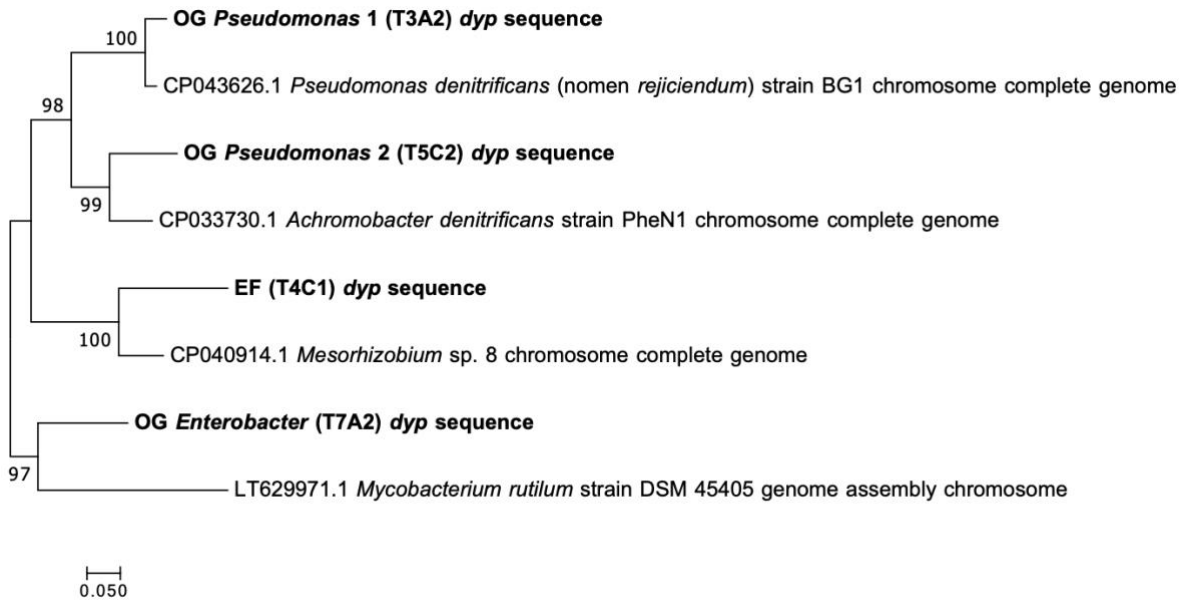

Figure 53 *- Isolate dyp phylogenetic neighbour-joining tree with neighbouring sequences obtained from the GenBank database in NCBI using BLASTn searches. Isolate sequences highlighted in bold font. Tree generated using MEGA7 with bootstrap analysis of 500 replicates.*

dyp PCR products from OG-T3A2, OG-T5C2, OG-T7A2 and EF-T4C1 were tested as standards for qPCR testing (qPCR conditions detailed below) in a dilution series of 10^7-

10^1 copy (**Figure 54**). Melt curve data (**Figure 54, B.**) obtained from the standard dilution series highlighted that OG-T3A2 and T7A2 share a common template melting temperature, however OG-T5C2 results suggest the possible formation of primer dimers (early melt curve peaks) and EF T4C1 results show different template melting temperature (possible lower G-C content). Amplification curve results (**Figure 54, A.**) show that highest amplification levels were obtained for OG T3A2 and T7A2, whereas OG T5C2 and EF T4C1 showed lower amplification levels (perhaps due to higher melting point or formation of primer dimers). Based on these results, the OG T3A2 standard was selected for future dyp qPCR runs due to high amplification (through high ΔRn values) and consistent melt curve results. The coefficient of determination (R^2) value (**Figure 54, C.**) of 0.999 (maximum $R^2 = 1$) suggests that data fit the standard curve of OG-T3A2 well. The efficiency percentage (**Figure 54, C.**) of 72.812% is lower than desired optimum (90-110%), possibly due to primer dimer formation in OG- T5C2 and high G-C content in EF-T4C1.

Original dyp qPCR conditions used by Tian *et al* (2016): Initial denaturation at 95 ºC for 2 mins followed by 40 cycles of 95 ºC for 30 secs, 60 ºC for 30 secs and 72 ºC for 90 secs followed by final elongation of 72 ºC for 7 mins.

dyp qPCR conditions used in this experiment: Initial denaturation at 94 ºC for 3 min followed by 30 cycles of 94 ºC for 30 secs, 60 ºC for 40 secs and 72 ºC for 30 secs followed by a brief denature step of 94 ºC for 15 secs and then 100 cycles of 75-95 ºC (+0.2 ºC per cycle) (conditions adapted from Tian et al (2016a)).

The qPCR conditions used in this experiment were based on the *dyp* PCR conditions (detailed in **Section 2.6.1.3**) as these conditions resulted in reduced non-specific amplification in previous tests (Tian *et al*., 2016). The qPCR conditions used by Tian *et al* (2016) are similar to that used in this experiment, however, denature temperature was reduced (95 ºC to 94 ºC), primer binding cycle time increased (30s to 40s) and elongation cycle time decreased (90s to 30s). An elongation time of 90s seemed much too long for an amplification product of 300-400 bp length, therefore it was reduced to minimize non-specific amplification.

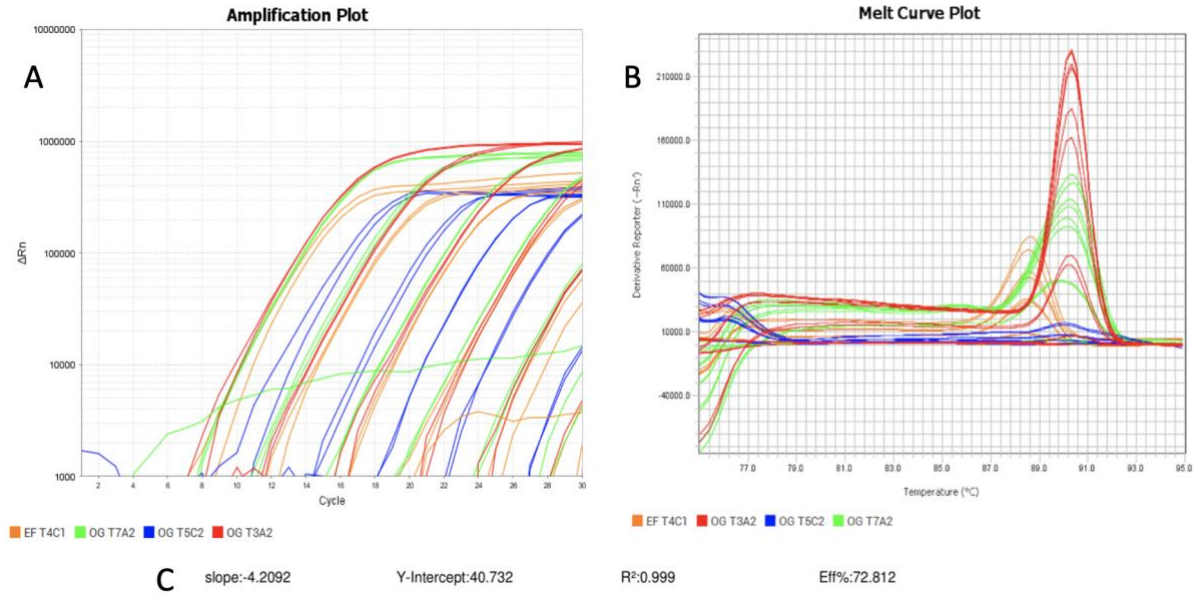

Figure 54 *- Results of qPCR analysis of isolate dyp products prepared as standard dilutions (10 – 10 copy number). Isolates OG T3A2 (Pseudomonas sp.), OG T5C2 (Pseudomonas sp.), OG T7A2 (Pseudomonas sp.) and EF T4C1 (unidentified). A – Amplification curve plot ; B – Melt curve plot of isolate standard dilutions; C – Standard curve details of OG T3A2 standard dilution series.*