

**Mining microbial compost
communities for lignocellulose
degrading proteins**

Nicola Oates

PhD

University of York

Biology

November 2016

Abstract

The production of second generation biofuels from agricultural residues is an attractive alternative to the use of conventional first generation feedstocks, which are also important food resources. However, these alternative feedstocks predominately consist of lignocellulose, the main structural component of the plant cell wall, and expensive physicochemical and enzymatic pre-treatments are required before fermentation into biofuel. Therefore, the discovery of novel enzymes capable of deconstructing lignocellulose, in conditions that would be amenable to industry, is an important goal.

The work, presented in this thesis, has explored the degradation of lignocellulose by a community of composting microbes, enriched for growth on wheat straw. Culturable members of the community, were isolated and assessed for their enzymatic activities towards lignin, cellulose and xylan. From these studies, a promising Ascomycota was identified, *Graphium* sp., which was capable of utilising both crystalline cellulose and xylan as carbon sources for growth.

Transcriptomic studies were performed on *Graphium* sp. with and without wheat straw present, representing the first molecular information generated from an organism of this genus. From this 680 putative proteins were annotated as containing carbohydrate active domains. Proteomics added further depth to the analysis, with investigations focused on secreted proteins both located in the culture supernatant and bound to the insoluble lignocellulose substrate.

Six secreted proteins were identified as targets for further analysis, and three of these were successfully isolated either from the native host, or a heterologous system. This included a lytic polysaccharide monooxygenase that appeared active on both chitin and cellulose, and a GH7 whose activity on cellulose was demonstrated. An intriguing protein, which showed low homology to a dioxygenase, was also expressed, though its role in the lignocellulose degrading environment has yet to be established.

Table of Contents

Abstract.....	2
Table of Contents.....	3
List of figures	10
List of Tables.....	14
Acknowledgements.....	15
Author’s Declaration	16
1 General introduction	17
1.1 Importance of transportation fuel.....	17
1.2 Biofuel as an alternative to petroleum.....	18
1.3 Second generation biofuel.....	20
1.4 Components of the plant cell wall	22
1.4.1 Cellulose.....	23
1.4.2 Hemicellulose.....	24
1.4.3 Lignin	25
1.5 Lignocellulose pre-treatments.....	27
1.6 Prospects for lignocellulose degradation.....	28
1.7 Enzymes involved in the degradation of lignocellulose.....	30
1.7.1 Enzyme classifications.....	30
1.7.2 Cellulose degradation	31
1.7.3 Hemicellulose degradation	35
1.7.4 Oxidative degradation of polysaccharides.....	36
1.7.5 Carbohydrate binding module	38
1.8 Microbial lignocellulose degraders.....	39
1.9 Aims of this thesis	43

2	Methods and Materials.....	44
2.1	Chemical reagents and substrates.....	44
2.1.1	Biomass	44
2.1.2	Organisms	44
2.1.3	Buffers.....	44
2.2	Microbiology methods.....	45
2.2.1	Preparation of Hutner’s trace elements	45
2.2.2	Preparation of media	45
2.2.3	Optimised Media.....	45
2.2.4	Agar plate and slant preparation	45
2.2.5	Spore preparation and harvesting	45
2.2.6	Spore and protoplast counting	46
2.3	Molecular biology techniques.....	46
2.3.1	DNA extraction.....	46
2.3.2	cDNA synthesis.....	48
2.3.3	Polymerase Chain Reaction (PCR).....	48
2.3.4	Agarose gel electrophoresis.....	49
2.3.5	Plasmid Extraction.....	49
2.3.6	PCR clean-up	49
2.3.7	Nucleotide Quantification.....	49
2.3.8	Sanger DNA Sequencing.....	50
2.3.9	In-fusion cloning.....	50
2.3.10	Transformation of competent cells	50
2.3.11	Periplasmic extraction.....	50
2.3.12	Mini-prep	51
2.4	Proteomic methods	51
2.4.1	Bradford Assay for protein quantification	51

2.4.2	Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)	51
2.4.3	Protein concentration by centrifugation	52
2.4.4	Buffer exchange by centrifugation.....	52
2.4.5	Anion exchange chromatography.....	52
2.4.6	Size exclusion chromatography	52
2.5	Data analysis	52
2.6	Enzyme assays methods	53
2.6.1	Qualitative Plate Assays	53
2.6.2	Reducing Sugar Assay.....	53
2.6.3	Glycospot plates.....	53
2.6.4	Para-nitrophenol (<i>pNp</i>) assays.....	54
3	Characterisation of a fungal compost isolate	55
3.1	Introduction	55
3.2	Methods and materials.....	57
3.2.1	Isolation of bacteria and fungi from an enriched wheat straw degrading community	57
3.2.2	Qualitative plate assays to identify potential lignocellulolytic activities.....	57
3.2.3	Identification of bacteria and fungi	58
3.2.4	Phylogenetic analysis of compost isolates.....	58
3.2.5	Targeted Amplicon Sequencing of 16S and ITS DNA	58
3.2.6	Bioinformatics analysis of the compost community.....	59
3.2.7	Conditions for fungal growth	59
3.2.8	Media optimisation of <i>Graphium</i> sp.....	60
3.2.9	Estimation of growth through dried biomass	60
3.2.10	Estimation of growth by total protein	61
3.2.11	Folin-Ciocalteu method for measurement of phenolics.....	61
3.3	Results.....	62

3.3.1	Lignocellulolytic fungal and bacterial isolates identified from the compost community	62
3.3.2	Community profiling through targeted amplicon sequencing	65
3.3.3	Composition of the prokaryotic composting community	65
3.3.4	Composition of the eukaryotic composting community	66
3.3.5	<i>Graphium</i> sp. growth in monoculture on lignocellulose derived substrates.....	68
3.3.6	<i>Graphium</i> sp. has increased tolerance to kraft lignin	69
3.3.7	Media optimisation for growth of <i>Graphium</i> sp. to increase production of both cellulases and xylanases.....	70
3.3.8	Comparison of <i>Graphium</i> sp. and <i>Trichoderma hazianum</i> growth on lignocellulose derived substrates.....	71
3.3.9	Extended time-course of <i>Graphium</i> sp. growth.....	74
3.3.10	Enzymatic activity	76
3.4	Discussion.....	77
3.4.1	The composting community	77
3.4.2	<i>Graphium</i> sp. adaptation to growth on lignocellulose	79
3.4.3	Determining critical stages of <i>Graphium</i> sp. growth in the degradation of wheat straw	80
4	Transcriptomic analysis of <i>Graphium</i> sp. During growth on wheat straw	81
4.1	Introduction	81
4.2	Materials and methods.....	83
4.2.1	Experimental Setup.....	83
4.2.2	RNA extraction	83
4.2.3	Clean up and enrichment of mRNA	83
4.2.4	RNA sequencing	84
4.2.5	Assembly of the transcriptome.....	84
4.2.6	Filtering	84
4.2.7	Bioinformatics	84
4.3	Results.....	86

4.3.1	RNA extraction from <i>Graphium</i> sp.....	86
4.3.2	mRNA enrichment from total RNA	86
4.3.3	RNA Sequencing of <i>Graphium</i> sp.....	87
4.3.4	Carbohydrate Active Enzymes	89
4.3.5	Transcriptomic response to wheat Straw	91
4.3.6	Carbohydrate binding modules	100
4.3.7	Cluster Analysis	102
4.4	Discussion.....	105
4.4.1	Optimisation of RNA extraction	105
4.4.2	Annotation of the <i>Graphium</i> sp. transcriptome	106
4.4.3	Differential expression patterns across wheat straw time course	107
4.4.4	Identification of novel enzymes from the <i>Graphium</i> sp. transcriptomics	107
5	Proteomics analysis of <i>Graphium</i> sp. during growth on wheat straw.....	109
5.1	Introduction	109
5.2	Methods.....	112
5.2.1	Protein Collection.....	112
5.2.2	Harvesting of supernatant protein	112
5.2.3	Harvesting of proteins attached to the insoluble components of the culture	112
5.2.4	Storage of Proteins.....	112
5.2.5	Liquid Chromatography-tandam mass spectrometry (LC-MS/MS) Analysis	113
5.2.6	Spectra Analysis	114
5.3	Results.....	115
5.3.1	Proteomics of <i>Graphium</i> sp. supernatant.....	115
5.3.2	Carbohydrate Active Enzymes	116
5.3.3	Comparison of the Biotin-tagged and Supernatant Proteins	123
5.3.4	Carbohydrate active enzymes in the biotin-tagged and supernatant fractions.....	126
5.4	Discussion.....	132

5.4.1	A comparison of the biotin labelled and supernatant fractions.....	135
6	Investigation of single lignocellulose active proteins.....	138
6.1	Introduction	138
6.2	Methods.....	141
6.2.1	Plasmid preparation.....	141
6.2.2	Protoplast formation.....	143
6.2.3	Fungal transformation	144
6.2.4	Bacterial transformation.....	144
6.2.5	Expression trials	145
6.2.6	Heterologous protein expression in <i>E. coli</i>	145
6.2.7	Purification of heterologously expressed LPMO.....	145
6.2.8	Purification of the GH7 cellulase.....	146
6.2.9	Activity assays	147
6.3	Results.....	148
6.3.1	Fungal cloning	148
6.3.2	Cloning, Expression and Purification of an LPMO.....	148
6.3.3	Enzymatic activity of the lytic polysaccharide monooxygenase	150
6.3.4	Cloning and expression of c6332_g2_i1	152
6.3.5	Activity assays for c6332_g2_i1	153
6.3.6	Purification of a GH7	155
6.3.7	Enzymatic assays for the GH7	156
6.4	Discussion.....	159
6.4.1	Fungal transformants	159
6.4.2	The lytic polysaccharide monooxygenase	159
6.4.3	c6332_g2_i1	161
6.4.4	The GH7.....	162
7	Final Discussion	164

7.1	Future directions.....	166
	Abbreviations	168
	References.....	170

List of figures

Figure 1-1 – Global rise in temperature correlates with CO ₂ atmospheric concentration	17
Figure 1-2 – The manufacture of photosynthetically fixed carbon present in lignocellulose into liquid transportation fuels.	20
Figure 1-3 - Plant cell wall structure.	23
Figure 1-4 – Model of xylan structure.....	25
Figure 1-5 – Structure of the three major monolignol monomers that are incorporated into the lignin polymer.....	25
Figure 1-6 – The cleavage of the cellulose chain through the action of glycoside hydrolases.....	31
Figure 1-7 – The inverting mechanism employed by glycoside hydrolases in the cleavage of cellulose.	32
Figure 1-8 - The retaining mechanism employed by glycoside hydrolases in the cleavage of cellulose.	33
Figure 1-9 – The sites of enzymatic attack on the cellulose chain.....	34
Figure 1-10 – Endo- and exo-glucanase structures.....	35
Figure 1-11 – Schematic of the CAZymes involved in the degradation of xylose chains in ascomycetes fungi.	36
Figure 1-12 – Reaction pathway of the LPMOs.....	37
Figure 1-13 – Representation of a bacterial cellulosome	39
Figure 2-1 Layout of the haemocytometer for cell counting.....	46
Figure 3-1 - Libraries of morphologically distinct isolates.	57
Figure 3-2 Central composite design for the optimisation of enzyme production..	60
Figure 3-3 Molecular phylogenetic analysis of the bacterial isolates.....	63
Figure 3-4 - Molecular phylogenetic analysis of the fungal isolates.....	64

Figure 3-5 – The composition of the bacterial community as accessed by 16S ribosomal DNA.....	66
Figure 3-6 – The composition of the fungal community as accessed by ribosomal DNA targeted amplicon sequencing.	68
Figure 3-7 – Growth of <i>Graphium</i> sp. on lignocellulose derived material.....	69
Figure 3-8 – Effect of kraft lignin on <i>Graphium</i> sp. and <i>Trichoderma harzianum</i> growth.....	70
Figure 3-9 – Optimisation of <i>Graphium</i> sp. growth media	71
Figure 3-10 – Comparison of the growth of <i>Graphium</i> sp. and <i>Trichoderma harzianum</i> on lignocellulose derived substrates	72
Figure 3-11 – Enzyme activities for <i>Graphium</i> sp. and <i>Trichoderma harzianum</i> cultures grown on different substrates	73
Figure 3-12 – Agar plate assays for endocellulase activity.	74
Figure 3-13 – Growth of <i>Graphium</i> sp. on wheat straw over a period of one month.....	75
Figure 3-14 – Cellulase and xylanase activity in <i>Graphium</i> sp. wheat straw cultures	76
Figure 4-1 - mRNA enrichment from total RNA harvested from wheat straw grown <i>Graphium</i> sp. cultures.	87
Figure 4-2 – Hierarchal clustering of RNA sequencing results and the distribution of contigs between samples	89
Figure 4-3 – Abundance of carbohydrate active enzyme classes	92
Figure 4-4 – Abundance and count of CAZyme families across the time points	93
Figure 4-5- The abundance of the top five glycoside hydrolase families.	94
Figure 4-6 - Expression profiles of contigs contain the glycoside hydrolase 45 family	96
Figure 4-7 Transcriptomic expression profile of the auxiliary activities families.....	97
Figure 4-8 Variation in the expression pattern of the auxiliary activity family nine (AA9).....	98
Figure 4-9 Transcriptomic expression profile of the top eight carbohydrate esterase families	99

Figure 4-10 – The abundance of the top five contigs that contain carbohydrate esterase domains 1	100
Figure 4-11 – CBM1s with no identified catalytic domain.....	102
Figure 4-12 – Cluster analysis of contigs across samples for novel lignocellulolytic protein discovery	103
Figure 4-13 - Expression of contigs of interest identified from their expression profile.....	104
Figure 5-1 Euclidean distances between proteomic samples.....	116
Figure 5-2 - Correlation between the transcriptomic and proteomic data	116
Figure 5-3 - Relative molar percentages of the CAZyme classes detected within the supernatant...	118
Figure 5-4 - Relative molar percentages of the top 15 families of the glycoside hydrolase (GH) class detected within the supernatant.....	119
Figure 5-5 - Relative molar percentages of families contained within the auxiliary activities (AA) class.	120
Figure 5-6 - Relative molar percentages of proteins of interest with the supernatant.....	121
Figure 5-7 - Relative molar percentages of the families contained within the carbohydrate esterase (CE) class	122
Figure 5-8 - Relative molar percentage of proteins which were identified as containing a carbohydrate binding domain family one (CBM1) but no carbohydrate active domain.....	123
Figure 5-9 - Subset of proteins contain in the supernatant and biotin-tagged fractions.....	125
Figure 5-10 Predicted features of proteins detected in the supernatant and the biotin labelled fractions.	126
Figure 5-11 - Distribution of carbohydrate active families across the biotin labelled (BF) and supernatant (SNT) fractions.....	127
Figure 5-12 – Distribution of the CAZymes in the biotin-tagged and supernatant samples.	129
Figure 5-13 – Heatmap showing of contigs containing carbohydrate binding domains.....	130
Figure 6-1 Contigs of interest for heterologous expression	138

Figure 6-2 Set up of equipment for the fungal transformation.....	144
Figure 6-3- Colony PCR of Stellar cells transformed with the gene of interest in the PrgG vector	148
Figure 6-4 – SDS-PAGE of periplasmic proteins harvested from LPMO expressing Rosetta cells	149
Figure 6-5 - Purification of a periplasmically expressed LPMO..	150
Figure 6-6 – LPMO activity on chitin, PASC and wheat straw.....	151
Figure 6-7 LPMO activity as assayed by Glycospot plates	152
Figure 6-8 Expression trials of the putative dioxygenase	153
Figure 6-9 Sugar release after the periplasmic fraction of cells transformed with the unknown protein incubation with PASC, wheat straw and xylan.....	154
Figure 6-10 Tannic acid and Congo Red visual changes in the presence of the unknown protein. ...	155
Figure 6-11 - Purification of a GH7 cellulase from the native system.	156
Figure 6-12 - Substrate specificity of a GH7 purified from <i>Graphium</i> sp.....	157
Figure 6-13 pH optima of the GH7 from <i>Graphium</i> sp	158

List of Tables

Table 2-1 – Polymerase chain reaction components.....	48
Table 2-2 – Polymerase chain reaction thermocycling conditions.....	49
Table 3-1 Primers for the amplification of ITS and 16S ribosomal DNA.....	58
Table 3-2 - Number of reads generated through targeted amplicon sequencing.....	65
Table 3-3 – Most abundant eukaryotic operational taxonomic units	67
Table 4-1 – RNA extraction conditions	86
Table 4-2 - Number of reads generated by RNA sequencing.....	88
Table 4-3 – The number of filtered contigs present within each sample	88
Table 4-4 - Total number of CAZymes detected in <i>Graphium</i> sp. transcriptome.....	90
Table 4-5 - Number of CAZyme domains in each condition	91
Table 4-6 – CBM containing contigs, and their CAZyme catalytic domains.....	101
Table 5-1 - Spectra obtained from the Liquid Chromatography-tandam mass spectrometry (LC-MS/MS) analysis.....	115
Table 5-2 - Carbohydrate active enzymes detected within the supernatant	117
Table 5-3 - Spectra obtained from LC-MS/MS analysis of the supernatant and biotin-tagged samples	124
Table 6-1 - Primers for the amplification of target genes from cDNA.....	141
Table 6-2 - Primers used for amplification of targets for fungal expression with overhangs for Infusion cloning into pIGF-pyrG	142
Table 6-3 -Primers for the linearization of selected vectors.....	145
Table 6-4 – Solutions for the disruption of affinities between the protein and insoluble fraction of the wheat straw grown <i>Graphium</i> sp. cultures.....	146

Acknowledgements

I owe a great deal of thanks to everybody who has helped me during the course of my PhD and my journey to get here.

Firstly my two supervisors Neil and Simon who have expertly guided my project, along with my TAP panel Gavin Thomas and James Chong.

Anna Alessi and Susannah Bird whose support and encouragement was been critical during the development of my project, and Shikin for here company as we learnt together.

Chong, Masa and Andre, whose visits to CNAP taught me a great deal about cloning, expression and protein purification, and who without their expertise I would not have achieved half as much, and everybody in CNAP who have made my time here so enjoyable and fun, and who were always happy to give ideas and advice.

My housemates Liz, Sarah, Milena and Lucie who throughout this process have given me support and love. Along with Becky, Zoe, Ellie and Jo whose support in the last few months has been incredible.

And finally a huge thank you to my family, who have helped my every step of the way and made the completion of this thesis possible.

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. All work is my own. No part of this work has previously been presented for an award at this, or any other University or Institution.

1 General introduction

1.1 Importance of transportation fuel

The transportation sector relies on oil as its principal fuel. In the United Kingdom, 96 % of the primary fuel consumption of this industry was petroleum based. Global consumption of oil rose to 4,211.1 million tonnes in 2014, an increase of 0.8 % on the previous year, and 8.8 % in the last decade (BP, 2015). This continued use is problematic, as oil, formed over many millions of years, is considered a non-renewable resource. Furthermore, the combustion of oil and its derivatives contribute to carbon dioxide emissions, which are associated with the rise in global temperature that now stands at 1 °C above pre-industrial levels (Met Office, 2015). This is a major environmental, social and political issue. There is, therefore, growing pressure to find clean renewable alternatives to petroleum in the transportation industry.

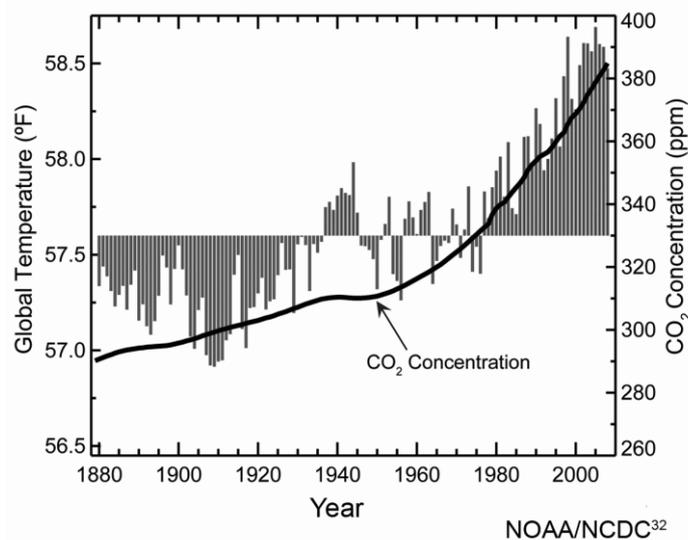


Figure 1-1 – Global rise in temperature correlates with CO₂ atmospheric concentration.

Sourced from <http://www.ncdc.noaa.gov/indicators/>.

1.2 Biofuel as an alternative to petroleum

Energy produced directly from the photosynthetically fixed carbon in plant biomass is an attractive alternative to conventional fossil fuels. Biofuels currently consist of two main categories: biodiesel derived from natural oils, and bioethanol fermented from sugars stored in biomass - either in simple or complex carbohydrate forms (Jordan et al., 2012).

Alongside the advantage of being a renewable resource, there are also considerable carbon savings gained with the use of biofuel, since carbon is cycled from the atmosphere to fuel and back again in a much shorter period. Additionally, as the feedstock for biofuel is commonly plant matter, there are less geological constraints on its production. This reduces dependence on fuel producing countries, which are often politically unstable, and brings new industry and jobs to the producing country. Biofuel may also prove an attractive resource for smallholders, able to manufacture their fuel on site from surplus produce - thus reducing their environmental impact whilst enhancing profits and quality of life (Branca et al., 2016, Ewing and Msangi, 2009, Demirbas and Demirbas, 2007).

Bioethanol can be blended into petrol at quantities of up to 15 % and used in conventional, unmodified engines. The use of these blends carries the dual advantage of both limiting the amount of infrastructure change required, and lessening the amount of air pollution produced from the combustion of petroleum, as blends contain a relatively high percentage of oxygen thereby reducing the amount of carbon monoxide produced by incomplete combustion (Malca and Freire, 2006). The addition of ethanol also acts as an anti-knocking agent, raising the temperature and pressure at which auto-ignition occurs and decreasing the need for other, more toxic additives, such as lead (Malca and Freire, 2006, Balat and Balat, 2009).

When higher blends of ethanol are desired, for example in countries such as Brazil where the bioethanol economy is more mature, bioethanol can also be used in flexi-fuel engines. These are combustion engines that allow the use of both petroleum and ethanol in any ratio (Balat and Balat, 2009). Alongside the modified engines, infrastructure must be provided to facilitate this technology with the provision of ethanol pumps at gas stations. These have become commonplace in Brazil, and have been growing in popularity in countries such as the United States and European countries, but are still lacking in the majority of areas (Pouliot and Babcock, 2014).

The environmental benefits conferred by biofuel have resulted in governments issuing mandates designed to encourage the use and development of these fuels. In the U.S.A, the renewable fuel standard was implemented whereby a certain percentage of transportation fuel (in 2016 9.63%) must originate from a renewable source each year (United States Congress, 2005), and the EU pledged that 10 % of transportation fuels produced in its member countries will be renewable by 2020 (Council of the European Union, 2009).

Despite biofuels presenting some clear environmental advantages over fossil fuels, current feedstocks for the majority of commercial bioethanol are the simple, readily available sugars from plant biomass, due to their relative ease in production. This has led to some controversy, as these fuels, termed first generation biofuels, have faced increasing scrutiny as part of the fuel vs. food debate. First generation biofuels have been heavily criticised for using resources that could otherwise be directed into the food and feed industries, and have been accused of contributing to a rise in global food prices (Algieri, 2014).

First generation biofuels have also been seen to impact biodiversity negatively. With the world's population estimated to rise from 7.5 billion in 2016 to 9.7 billion by 2050 (Nations, 2015), the need to produce more food is apparent. This, combined with economic pressure for biofuel production has led to major land use changes and agricultural expansion (Laurance et al., 2014, Wright et al., 2017). This has led land to be intensively farmed and large quantities of fertilisers applied, leading to soil erosion, eutrophication and negative effects on human health (Morales et al., 2015, Sheppard et al., 2011). Due to the range of feedstocks that biofuel can be produced from, land use changes have occurred on a global scale, including biodiversity hotspots and pristine forest (Fitzherbert et al., 2008). For example oil palm plantations support fewer species than the native forest in South-East Asia, introduce pest species to the area, and negatively affect soil quality (Fitzherbert et al., 2008, Mukherjee and Sovacool, 2014). Land use changes driven by the production of first generation feedstocks could even offset the carbon savings generated by biofuel production (Lapola et al., 2010), and the increasingly negative view of first generation biofuel has led to changes to EU legislation, which now limits the use of percentage of transportation fuel derived from sources competing with the food industry to 7 % (Chatain, 2015).

1.3 Second generation biofuel

Second generation biofuel, made from inedible plant biomass, address the criticisms levelled at the first generation fuels by using feedstocks derived from either agricultural residues (Gupta and Verma, 2015), or dedicated low input crops that have been grown on marginal land (Gelfand et al., 2013). Compared to their first generation predecessors, second generation biofuels, as well as not impacting on food security, have the advantage of lower carbon emissions over their life cycle (Fazio and Monti, 2011, Havlik et al., 2011). However, due to a number of uncertainties, such as the complexity of the production and variation in feedstocks, there is considerable variation in the predicted savings they confer (Whitaker et al., 2010).

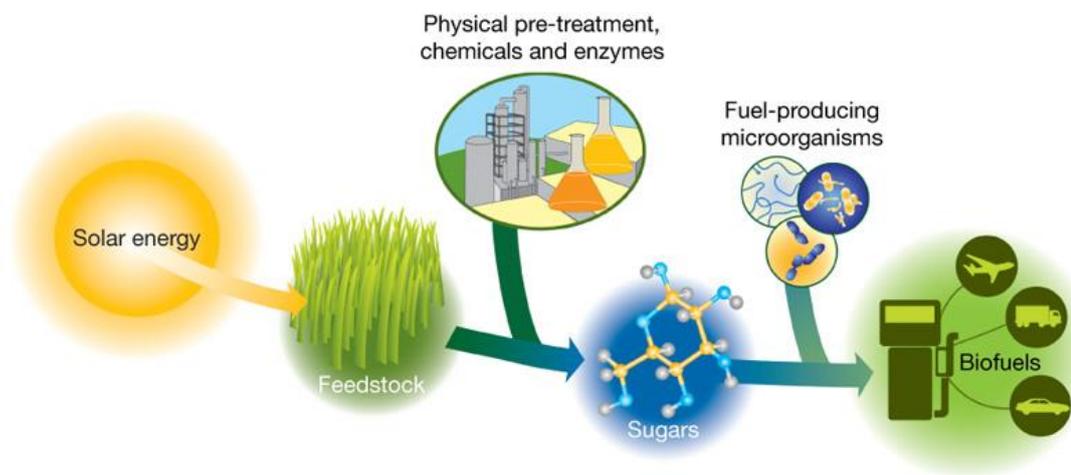


Figure 1-2 – The manufacture of photosynthetically fixed carbon present in lignocellulose into liquid transportation fuels. Sourced from Rubin et al (Rubin, 2008).

Several different feedstocks are used for second-generation biofuel production. In Brazil, sugarcane bagasse, the remaining fibre left after sugar extraction, may make a promising feedstock. It amounts to 0.6 kg per 1kg of total sugar cane used in food manufacturing, and is an abundant resource with 180 Tg of dry sugarcane bagasse produced each year (Kim and Dale, 2004).

In the USA, low input grass crops, such as miscanthus or switchgrass, have been considered as appropriate feedstocks. Although these are dedicated biofuel crops, they are capable of growth on marginal land, and have been demonstrated to increase biodiversity in these environments when properly managed (Werling et al., 2014, Hartman et al., 2011).

Europe, like Brazil, has focussed on the potential to produce biofuel from agricultural residues. Wheat straw, one of Europe's primary agricultural residues, is abundant with 194 Tg of straw produced annually and has been modelled to have beneficial greenhouse gas emissions compared to petrol making it an ideal feedstock in this region (Kim and Dale, 2004, Wang et al., 2013).

In spite of these promising feedstocks and the environmental benefits they confer, second generation biofuels are still not currently produced at a widespread commercial level. The major barrier to be overcome is the cost of required processing, which although varies widely in estimations (14–26 \$/GJ), is not yet sufficiently favourable to drive the replacement of petroleum based fuels (Figure 1-2) (Kumar et al., 2016, Cazzola et al., 2013, van Eijck et al., 2014). The high cost of bioethanol is driven by the need for intensive pre-treatments. As dedicated bioethanol crops and agricultural residues are predominantly comprised of lignocellulose, the structural component of the plant cell wall, and this material must be broken down into its component sugars (Chaturvedi and Verma, 2013). Typically, various physical and chemical pre-treatments are used to expose the component polysaccharides and these are energy intensive, require specialised equipment and often require large quantities of water (Alvira et al., 2010). The subsequent enzymatic saccharification step, to hydrolyse the exposed polysaccharides into fermentable sugars also contributes to the cost of bioethanol production (Jordan et al., 2012). These enzymes are expensive to produce, currently suffer from low efficiencies and can be inhibited by molecules released during chemical pre-treatments (Jonsson et al., 2013, Jönsson and Martín, 2016).

1.4 Components of the plant cell wall

Lignocellulose is a complex composite material that exists within the secondary cell wall of plant cells. It is formed after the plant cell's growing has ceased, and is abundant in cell types that require greater mechanical strength, having evolved to give plants durability and structure, as well as to protect them from microbial attack (Harris and DeBolt, 2010, Himmel et al., 2007). The secondary cell wall differs from the primary cell wall due to the addition of lignin and exclusion of most pectins, which gives it a more rigid structure and a greater hydrophobicity (Zhong and Ye, 2014, Vanholme et al., 2012, Vanholme et al., 2010).

Lignocellulose itself consists of three main polymers: cellulose, hemicellulose and lignin, which are closely connected through complex interactions and extensive cross-linking (Figure 1-3). These polymers, described in the following sections, are present in plant cell walls in varying ratios and with different amendments, depending on species and cell type.

In wheat straw, the agricultural residue of most relevance to this thesis, the largest fraction of polymers is cellulose, representing 33-47 % of the biomass. This is followed by hemicellulose that represents an estimated 21 to 26 %, and finally lignin, at 11 to 22.9 %. It should be noted, however, that the composition of lignocellulose is variable and dependent on environmental factors, since plants that have been exposed to more stresses have been found to have more lignin (Cano-Delgado et al., 2003).

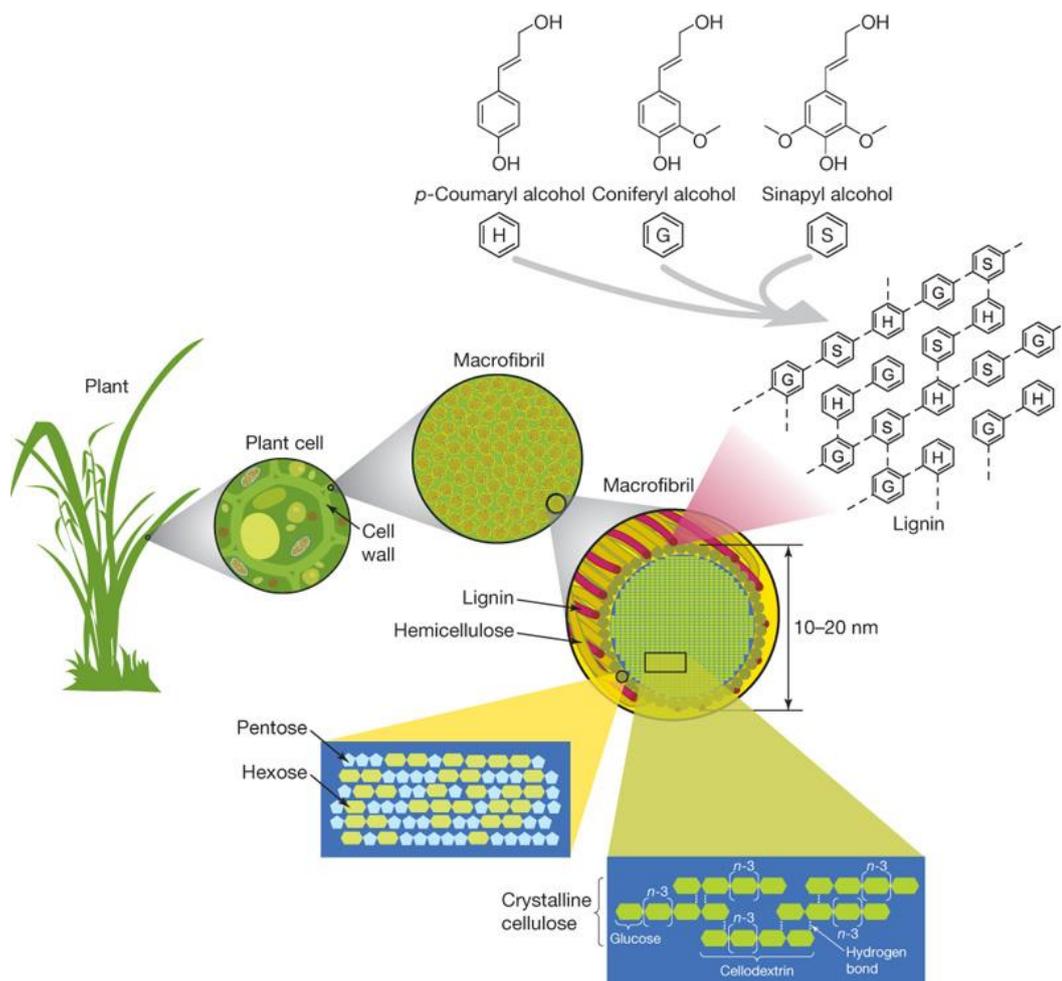


Figure 1-3 - Plant cell wall structure. Cellulose, hemicellulose and lignin cross-link to form an insoluble, macromolecular composite material. Sourced from Rubin et al. (Rubin, 2008).

1.4.1 Cellulose

Cellulose is the most abundant organic polymer on earth and the most stable polysaccharide (Cosgrove, 2005). It consists of unbranched chains of β -1,4 linked glucose units with high degrees of polymerisation. Within the cellulose chain, glucose molecules are flipped 180° relative to each other, such that cellobiose acts as the repeating unit (Marriott et al., 2016). This allows the polysaccharides to lie flat against each other and form tightly packed parallel sheets, with hydrogen bonding between adjacent chains (Himmel et al., 2007).

These parallel layers then interact through van der Waals bonds to form ordered microfibrils, and although some amorphous regions do exist, they tend to have a crystalline nature that renders them recalcitrant to enzymatic degradation (Nishiyama et al., 2002). Cross-linking of

other plant cell wall components, including hemicellulose and lignin, further increase this stability (Valent and Albersheim, 1974).

Despite cellulose being of great importance in both the biofuel and paper making industries, there are still unanswered questions about its structure in the secondary plant cell wall. This includes the dimensions of the microfibrils that cellulose forms, the intricacies in how and where this structure interacts with the other plant cell wall polymers, and the effect that this has on the enzymes that deconstruct the lignocellulose structure (Cosgrove, 2014).

1.4.2 Hemicellulose

The term hemicellulose denotes a family of diverse polysaccharides that can consist of 5- and 6- carbon sugars, decorated with branches of various lengths. These polysaccharides interact with cellulose through hydrogen bonding, thereby forming a coating around the microfibrils which helps cement them to one another, whilst conferring a degree of flexibility. This plasticity can be maintained due to the presence of the side chain substitutions that prevent chains from forming crystalline structures (Scheller and Ulvskov, 2010).

The composition of hemicellulose found within the plant cell wall varies with species and location. In gymnosperms, which include lumber producing trees like conifers and pine, the main hemicelluloses present are galactoglucomannans. These are linear chains of β -1,4 linked mannose and glucose residues substituted with randomly distributed glucose, galactose and mannose units.

In angiosperms, however, the grouping encompassing all flowering plants, the predominant secondary cell wall hemicellulose is xylan. Polysaccharides designated as xylan are characterised by their backbone of 1,4-linked β -D-xylopyranose residues, but decorations can vary - again dependent on species (Marriott et al., 2016). In the dicot secondary cell wall the typical xylans are glucuronoxylans (GX), where the xylose backbone is decorated with α -1,2 glucuronic acid and 4-O-methyl glucuronic acid residues ([Me]GlcA). Typical xylans of the secondary cell wall of grasses, such as those in wheat straw, are highly substituted with single or short chains of arabinosyl, glucuronosyl and xylose residues, and known as glucuronoarabinoxylans (GAX) (Figure 1-4) (Vogel, 2008).

As well as carbohydrate substitutions, acids can also be used to decorate these xylan chains. In GAX, arabinose residues are substituted with esters of ferulic acid, which can be oxidised to form di-ferulates that link xylans together and form covalent attachments to the third

component of the secondary cell wall - lignin (Lam et al., 2001, Iiyama et al., 1990). The xylan backbone is also extensively decorated with acetyl groups (Gille and Pauly, 2012), conferring a hydrophobicity to these chains, and making them less susceptible to attack by xylanases through steric hindrance (Ebringerova and Heinze, 2000, Biely et al., 1986).

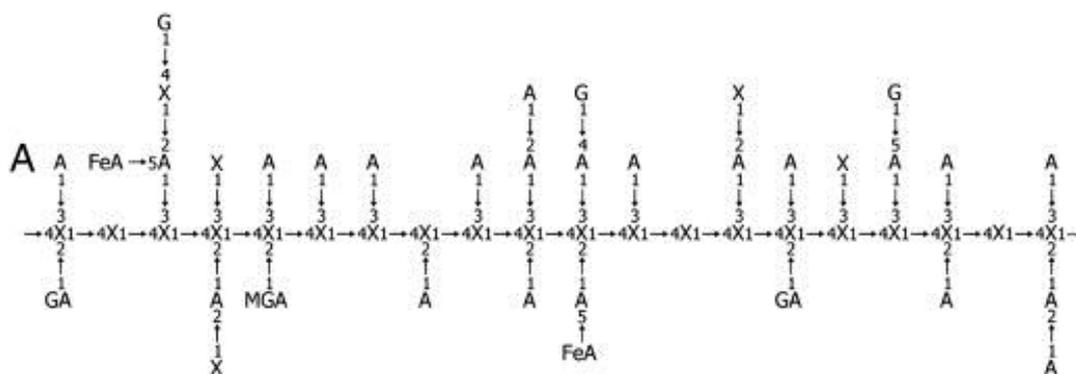


Figure 1-4 – Model of xylan structure. The xylose (X) backbone is shown with arabinosyl (A) and glucuronosyl (G) sidechain with additional ferulic acid (FeA) and methyl glucuronic acid residues (MGA). Adapted from Kumar et al. (Kumar et al., 2008).

1.4.3 Lignin

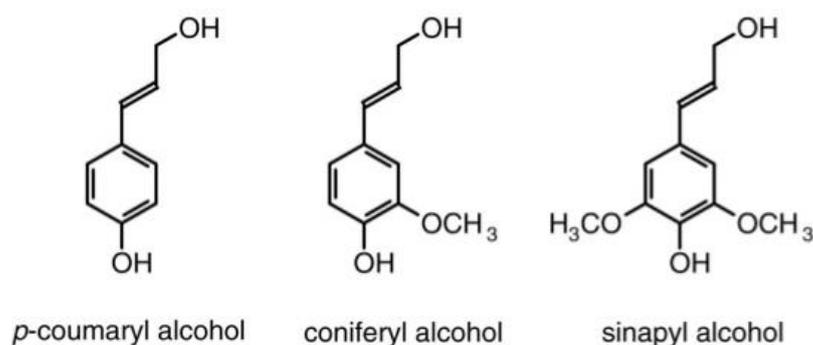


Figure 1-5 – Structure of the three major monolignol monomers that are incorporated into the lignin polymer.

Lignin, the third major component of the secondary plant cell wall, is a more complex polymer, with no defined repeat structure. It is deposited in the plant cell wall to cover and penetrate the cellulose and hemicellulose structures, preventing cell wall degrading enzymes from accessing the sugar-rich polymers, and adding to the structure's hydrophobicity.

Lignin is constructed from three monolignol monomers: p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol (Figure 1-5). Once assimilated into the lignin structure as phenylpropanoids, these are known respectively as guaiacyl propanol (G), synapyl (S) and p-hydroxyphenyl (H). These monolignols are thought to be polymerised through plant peroxidase mediated dehydrogenation of the phenolic molecule, initiating the generation of free radicals, which then couple to the growing lignin polymer (Boerjan et al., 2003). This coupling reaction acts to quench the free radicals. The resultant bond formed during the quenching is dependent on the location of the radical in the phenolic ring at the time of incorporation. This results in a non-regular phenylpropanoid matrix with high oxidation potentials and random linkages, where the lack of defined repeat structure prevents the evolution of enzymes with specific active sites (Davin and Lewis, 2005).

Like the hemicelluloses, the composition of lignin varies with species. In gymnosperms, the dominant phenylpropanoid incorporated is G, with small quantities of H, while angiosperms are a mixture of all three. Within angiosperms, monocots and dicots also have different ratios of the phenylpropanoids. The differing rates at which these lignin monomers are incorporated is thought to affect the ease at which lignin is degraded. Monomers with additional methoxy groups have additional potential reactive sites, and can, therefore, create structures with a more diverse range of linkages (Marriott et al., 2016).

1.5 Lignocellulose pre-treatments

Currently, due to lignocellulose being extremely recalcitrant to degradation, chemical and mechanical methods of pre-treatment are required to access the sugar-rich polymers stored within the plant cell structure. Once exposed, these polymers are treated with high loadings of cellulolytic enzymes to release the fermentable sugars required for bioethanol production (Alvira et al., 2010).

Mechanical methods of pre-treatment include the milling or grinding of the biomass. This enhances the accessibility of cellulolytic cocktails to the structure, and reduces the degree of polymerisation. Chemical methods include treatments in which the biomass is soaked in alkali solutions at high temperatures and at high pressures. This swells the cellulose fibres, disrupts crystallinity, and partially solubilises the lignin. Physical and chemical pre-treatment methods can also be combined in ammonia fibre explosion, where biomass, soaked in ammonia or water, is heated under pressure before being rapidly depressurised. Depressurisation allows the liquid contained within the biomass to expand, ripping open the structure (Alvira et al., 2010). Although these pre-treatments do dramatically increase the digestibility of lignocellulose, they require high-energy inputs and costs, thus diminishing the environmental benefits of using second-generation biofuels (Wang et al., 2013).

The cost of producing second-generation biofuel is further increased by the high loadings of cellulolytic cocktails required in the saccharification step, in which the exposed polymers are broken down into their component sugars. This can also be complicated by the use of harsh chemicals in the pre-treatment stages as these can affect downstream processing through the production of molecules inhibitory to the cellulolytic enzymes (Kim et al., 2011, Zhai et al., 2016). This requires the addition of costly detoxification steps, often using up large volumes of water (Frederick et al., 2014).

Once the polysaccharides have been broken down into their component sugars, yeast-based fermentation is used for ethanol production. As natural yeasts cannot ferment 5-carbon sugars from the hemicellulose, genetically modified strains have been developed to effectively use the sugars released in the previous stages (Nielsen et al., 2013, Zhou et al., 2012).

1.6 Prospects for lignocellulose degradation

As discussed in the previous sections, the production of second-generation biofuel is an energy intensive process with varying environmental caveats. These caveats have prevented its widespread deployment as an alternative to petroleum-based fuels. The commercial viability of biofuel production may be achieved by developing more effective pre-treatments, improving the digestibility of biomass through genetic engineering or selective breeding (Vanholme et al., 2012, Badhan and McAllister, 2016, Marriott et al., 2014), and optimising saccharification cocktails (Bussamra et al., 2015, Chandel et al., 2012, Gao et al., 2010, Sun and Cheng, 2002).

One alternate method of pre-treatment currently being explored involves ionic liquids, where the structure of lignocellulose is effectively dissolved by the disruption of the hydrogen bonds that hold the cellulose chains together (Swatloski et al., 2002). However, before these can become a viable commercial alternative, methods of recycling the ionic liquid to lower the cost must be developed (Klein-Marcuschamer et al., 2011), and inhibition of these liquids on cellulases must be overcome (Turner et al., 2003, Xu et al., 2016, Wolski et al., 2016, Park et al., 2012).

Another alternative pre-treatment strategy is the use of biological treatments from white-rot fungi, the only organisms that have been demonstrated to significantly degrade lignin. These are discussed in Section 1.8. Though these biological pre-treatments have been investigated, ligninases are not currently included in commercial cocktails. It has, however, been hypothesised that these fungi could be grown directly on biomass, to selectively degrade the lignin and hemicellulose components of the plant cell wall. These treatments would be both environmentally friendly and cost-effective, but are currently considered too slow to be incorporated into industrial pipelines (Talebnia et al., 2010, Garcia-Torreiro et al., 2016).

The discovery of more efficient saccharification enzymes is also being targeted, as currently the production of these cocktails is a major factor in the commercial infeasibility of bioethanol production (Klein-Marcuschamer et al., 2012). If enzymatic cocktails could be applied at lower concentrations, this would go some way towards improving the commercial prospects of second generation biofuels. Further reductions in the cost of second generation biofuels could be achieved by the discovery of enzymes with high stabilities that could be reused in multiple saccharification reactions (Gregg and Saddler, 1996, Eriksson et al., 2002). Additionally, the discovery of cellulases capable of working in the inhibitor-rich post pre-treatment hydrolysates

could reduce the need for extensive clean up stages after biomass pre-treatment (Oriente et al., 2015).

1.7 Enzymes involved in the degradation of lignocellulose

Lignocellulolytic enzymes are found throughout all domains of life, and as such it is unlikely that we have explored the total diversity that they have to offer (Yang et al., 2011). Organisms with the ability to metabolise lignocellulose have been described in radically different environments, including bacterial symbiotes in ruminant animals (Hess et al., 2011), microbial leaf litter communities (Dougherty et al., 2012, Allgaier et al., 2010), and marine animals (King et al., 2010).

The development of next-generation sequencing technology as a staple of biological investigation has been a useful tool in the discovery of new enzymes (Kameshwar and Qin, 2016, Rubin, 2008), and as the technology further advances and becomes more cost effective this is only set to increase. Next generation sequencing is of particular use for the discovery of lignocellulolytic enzymes from complex communities of microorganisms – for example, those present in leaf litter and compost, where the majority are unable to be studied in isolation. Metagenomic and metatranscriptomic studies have been carried out in a wide range of environments for enzymatic discovery. These have provided enormous amounts of information about the communities that they contain, as well as their member's interactions (Hongoh, 2011, King et al., 2010). In particular, meta-type studies's detailed analysis on individual organisms can provide a comprehensive knowledge of the mechanisms and physiology employed in metabolising lignocellulose, as techniques such as transcriptomics can create a detailed understanding of how organisms respond to lignocellulose (Sun et al., 2012).

1.7.1 Enzyme classifications

Since carbohydrate active enzymes are so ubiquitous and diverse, a library, known as the CAZyme (Carbohydrate Active enZyme) database was developed to aid their identification (Lombard et al., 2014). Here, enzymes are grouped based on similarities in their amino acid sequences using the methods established by Bernard Henrissat in 1999 (Henrissat, 1991, Cantarel et al., 2009).

Within the CAZyme database enzymes are broken down into classes. These classes include: carbohydrate esterases (CEs), glycoside hydrolases (GHs), glycosyl transferases (GTs), and polysaccharide lyases (PLs), along with the auxiliary activities (AAs) which describe enzymes that do not act on carbohydrates themselves but are closely associated with carbohydrate breakdown (Levasseur et al., 2013b, Campbell et al., 1997, Coutinho et al., 2003, Lombard et

al., 2010). Carbohydrate binding modules (CBMs) are also designated within the CAZyme database (Boraston et al., 2004).

Classes are further subdivided into families with each class containing at least one protein that has been biochemically characterised. Currently, the 332,698 classified sequences within the GH class are divided among 135 families, the 13,682 sequences in 13 AA families and the 37,705 in 16 CE families (Lombard et al., 2014). As these families have been defined by their primary amino acid sequence, and amino acid structure is intrinsically linked to protein structure, enzymes within groupings tend to share similar 3D-structures and mechanisms. For example the GH6 family are cellulases that follow the inverting mechanism of hydrolysis. However it should be noted that enzymes with differing enzyme activities and substrate preferences can also be contained within the same family. This is epitomised by the GH1 family which despite consisting of proteins all possessing a TIM barrel fold, contains enzymes with differing substrate specificities, including mannosidases, fucosidases, xylanases, glucosidases and galactosidases (Pujadas and Palau, 1999).

This classification system, since its introduction, has undergone multiple modifications, with the constant addition of newly discovered families and refinement of sequence comparison methods (Henrissat and Bairoch, 1996). These modifications have included the introduction of clans - a method to group families based on structure similarities, and sub-families – a more refined classification within select families. Perhaps most significantly, the addition of a new class, the auxiliary activities was made in 2013 (Levasseur et al., 2013b) to address the previous miss-assignment of the GH61 and CBM33 families (Harris et al., 2010, Forsberg et al., 2011).

1.7.2 Cellulose degradation

Enzymes responsible for the degradation of cellulose were first identified as being hydrolytic, which through the incorporation of a water molecule cause the cleavage of a glycosidic bond within the polysaccharide chain (Figure 1-6).

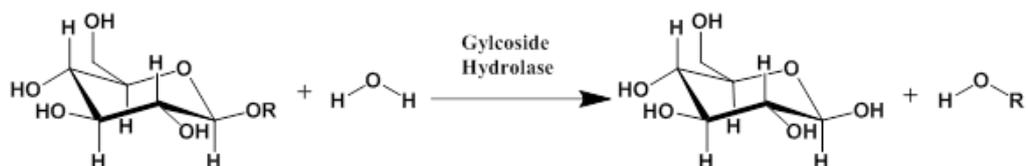


Figure 1-6 – The cleavage of the cellulose chain through the action of glycoside hydrolases.

These enzymes can work through either a retaining or inverting mechanism, depending on the maintenance of the anomeric carbon's stereochemistry (Koshland, 1953). Enzymes that employ the inverting mechanism of glycoside hydrolysis do so through the action of two carboxylic acid residues (Asp and Glu), one of which, deprotonated in the active state, acts as the general base, and the other as the general acid (Figure 1-7). This mechanism proceeds through the protonation of the glycosidic oxygen by the general acid and the concurrent deprotonation, and activation of a molecule of H₂O. This activated H₂O then, acting as a nucleophile, attacks the anomeric carbon, resulting in its hydroxylation and the inversion of the substrate stereochemistry (Davies and Henrissat, 1995, Vocadlo and Davies, 2008).

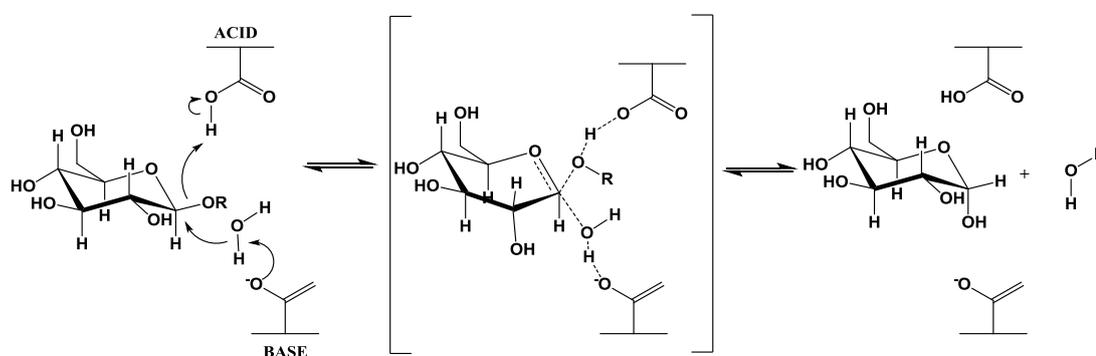


Figure 1-7 – The inverting mechanism employed by glycoside hydrolases in the cleavage of cellulose.

The retaining mechanism, unlike like the inverting mechanism which proceeds via a single displacement reaction, uses a two-step, double-displacement (Figure 1-8). The enzymes Asp and Glu residues also coordinate this reaction, although the distances between these in retaining enzymes is slightly shorter compared to those in the inverting enzymes. One of these residues, acting as an acid, donates a proton to a glycosidic oxygen, whilst the other, acting as a nucleophile, attacks the anomeric carbon, forming a glycosyl enzyme intermediate. The second step then involves the residue that was previously acting as an acid acting as a base. This base serves to deprotonate a water molecule as it attacks the anomeric carbon, hydrolysing the bond between the glycosyl and enzyme to create a product, which maintains stereochemistry of the initial substrate.

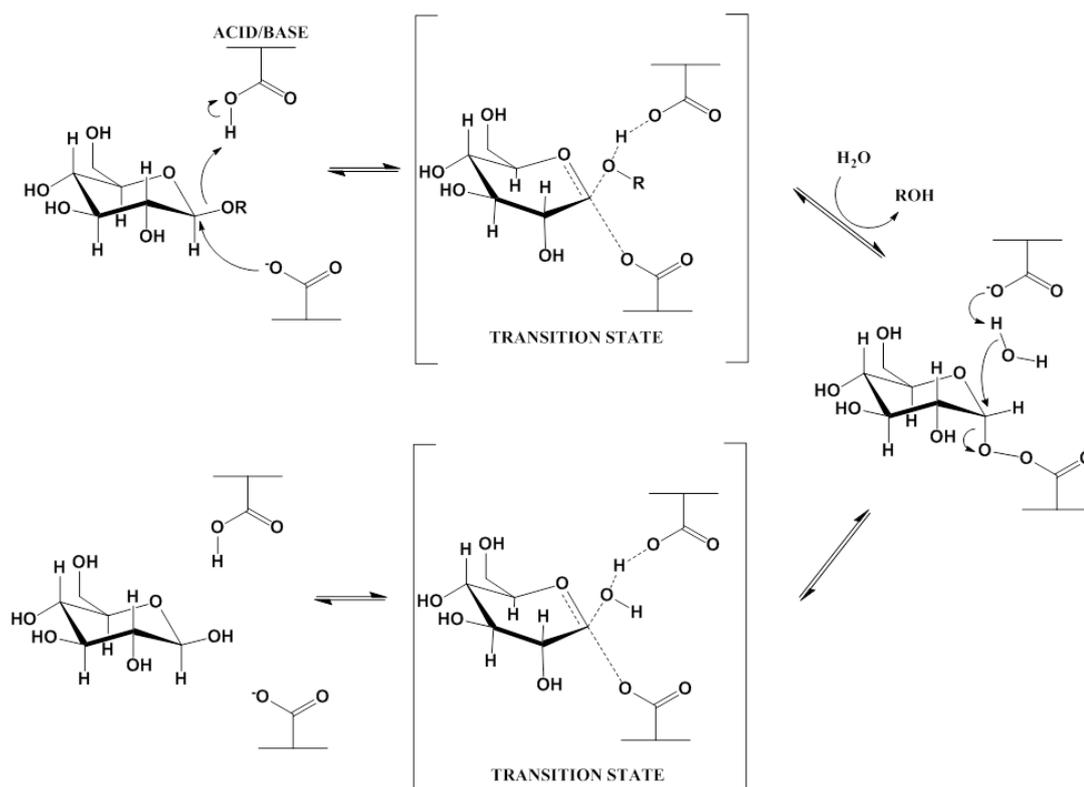


Figure 1-8 - The retaining mechanism employed by glycoside hydrolases in the cleavage of cellulose.

As well as being either inverting or retaining, cellulases are commonly defined as having one of three actions. Endo-acting endoglucanases randomly cleave at internal glycosidic bonds at amorphous regions. Exo-acting cellobiohydrolases cleave progressive oligosaccharides or cellobiose from the ends of cellulose. Finally, β -glucosidases hydrolyse cellobiose into its glucose components (Figure 1-9) (Lynd et al., 2002). These enzymes work in a synergistic manner, with endo-acting enzymes creating more initiation points for the cellobiohydrolases to act on, and β -glucosidases helping to relieve product inhibition. Cellobiohydrolases also can be defined as operating from the reducing end or the non-reducing end of the cellulose chain; these differing modes of action are contained within the GH7 and GH6 family respectively.

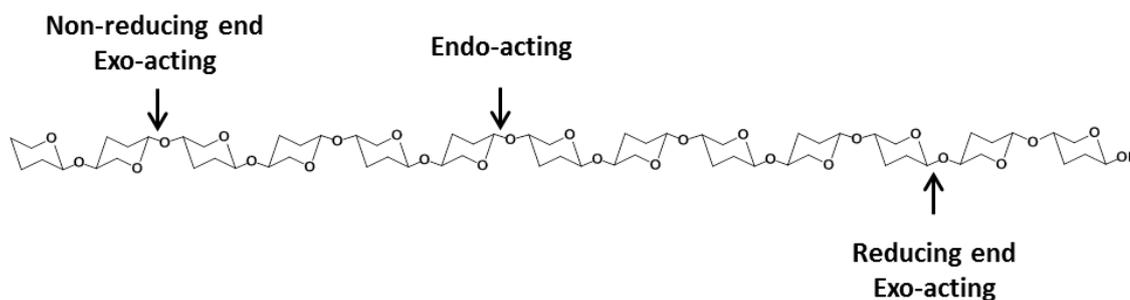


Figure 1-9 – The sites of enzymatic attack on the cellulose chain.

The structure of the enzyme plays a major role in determining the action that it displays (Figure 1-10). Endoglucanase have a groove in their structure, which allows the protein to interact with cellulose chains at amorphous sites and subsequently cause chain breaks. These chain breaks can be used by exoglucanases as initiation points. Exoglucanases require initiation points, as rather than possessing a groove structure, they have a tunnel-like shape that cannot incorporate the middle of the chain into itself. Therefore, in order to position the cellulose chain within the tunnel, the enzyme must first dislocate an end from the crystalline microfibril before threading it through the tunnel, where it can be bound in as many as nine positions. Once the polysaccharide is in position, multiple catalytic events can occur to progressively move along the chain, catalysing multiple cleavages and expelling the product, often cellobiose, before the enzyme is finally deabsorbed from the crystalline cellulose (Payne et al., 2015).

A more detailed understanding of the enzymes has challenged the traditional viewpoint that there are just three modes of action in cellulose degradation. Evidence that some endoglucanases do not produce degradation patterns compatible with their proposed random attack, has led to the discovery of processive endoglucanases (Tomme et al., 1996). These are capable of cleaving at amorphous sites in the same manner as endoglucanase, and can then progressively degrade soluble oligosaccharides like cellobiohydrolases. This demonstrates that hydrolytic classifications are not as distinct as previously thought.

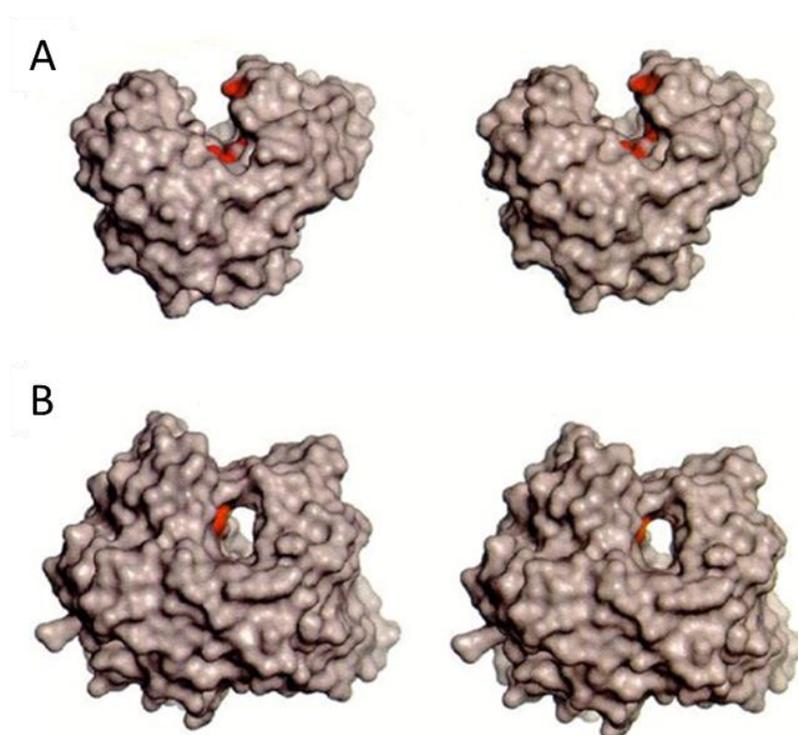


Figure 1-10 – Endo- and exo-glucanase structures. (A) An endocellulase from *T. fusca* showing an exposed cleft like structure. (B) An exoglucanase from *T. reesei* showing an enclosed tunnel active site. Figure adapted from Davies et al (Davies and Henrissat, 1995).

1.7.3 Hemicellulose degradation

Hemicellulases are more varied than cellulases, as hemicellulose structures are more diverse (Figure 1-11). Xylans, the main hemicellulose in grasses, are depolymerised through the action of endo- β -1,4-xylanase enzymes, which break the xylose backbone into smaller oligosaccharides. These smaller oligos can then be acted on by β -1,4-xylosidase to release xylose monomers, which are contained within multiple CAZyme families, including GH 3, 39, 43, 52, and 54.

Xylans, particularly those derived from the secondary cell wall of cereal crops, are extensively decorated and additional enzymes are required for the removal of side chains (Glass et al., 2013). These enzymes are critical for the effective degradation of xylan since if side chains are not removed, they may sterically inhibit the action of the backbone cleaving xylanases. Side-chain modifying enzymes include those capable of cleaving the glycosidic bonds formed between the arabinofuranosyl, contained within multiple families, including GH3, GH10, GH43, GH51, GH54 and GH62, and [Me]GlcA residues within the GH67 family.

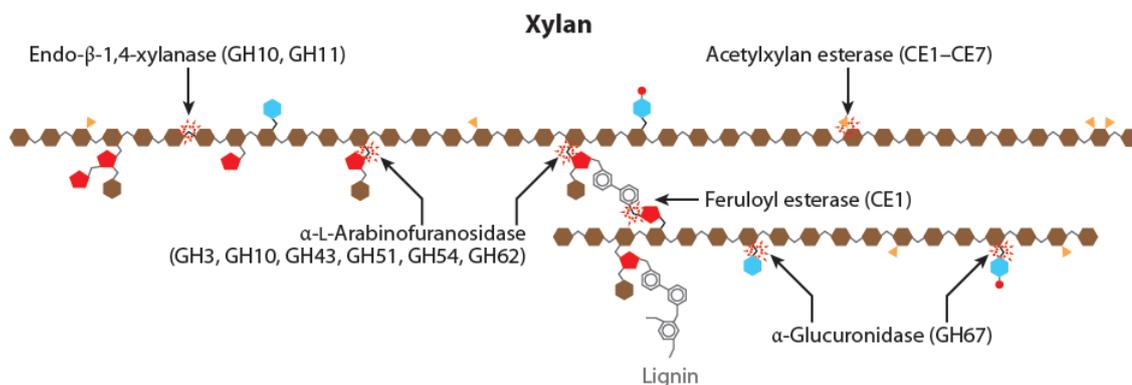


Figure 1-11 – Schematic of the CAZymes involved in the degradation of xylose chains in ascomycetes fungi. Reproduced from Glass et al. (Glass et al., 2013).

Carbohydrate esterases are similarly required to allow the efficient breakdown of the xylans contained within the wheat straw cell wall. This is because the acetylation of the xylose backbone affects the capacity of the other xylanase enzymes to degrade the xylan effectively, and because the ferulic modifications of the arabinofuranosyl groups can be cross-linked to lignin. Xylan esterases contained within the families CE1-CE7, CE12 and CE16 catalyse the deacetylation of the extensively decorated xylan backbone, whilst some contained within the CE1 family catalyse the deacetylation of the ferulic acid groups (Dodd and Cann, 2009). The addition of these carbohydrate esterase enzymes to industrial preparations of xylanolytic glycoside hydrolases has been shown to increase sugar release significantly (Goldbeck et al., 2016). Likewise, the solubilisation of xylan in lignocellulose has been demonstrated to increase the ability of the cellulose fraction to be depolymerised by cellulases (Zhang et al., 2012).

1.7.4 Oxidative degradation of polysaccharides

The hypothesis that the enzymatic degradation of cellulose and xylan is purely hydrolytic has been disproven in recent years, with the discovery of the lytic polysaccharide monooxygenases (LPMOs). These, characterised as having an oxidative action after being miss-identified as hydrolytic enzymes in the CAZy GH61 and CBM33 families (Quinlan et al., 2011, Zifcakova and Baldrian, 2012, Forsberg et al., 2011), were first discovered to induce chain breaks in chitin (Vaaje-Kolstad et al., 2010). Since their initial discovery, they have been found to work on different polysaccharides, including cellulose (Quinlan et al., 2011), xylan (Agger et al., 2014) and starch (Vu et al., 2014). Furthermore, they have undergone a rapid expansion as numerous enzymes have been found in cellulolytic organisms, with some, such as *Chateomium globsum*, containing as many as 39 distinct LPMOs.

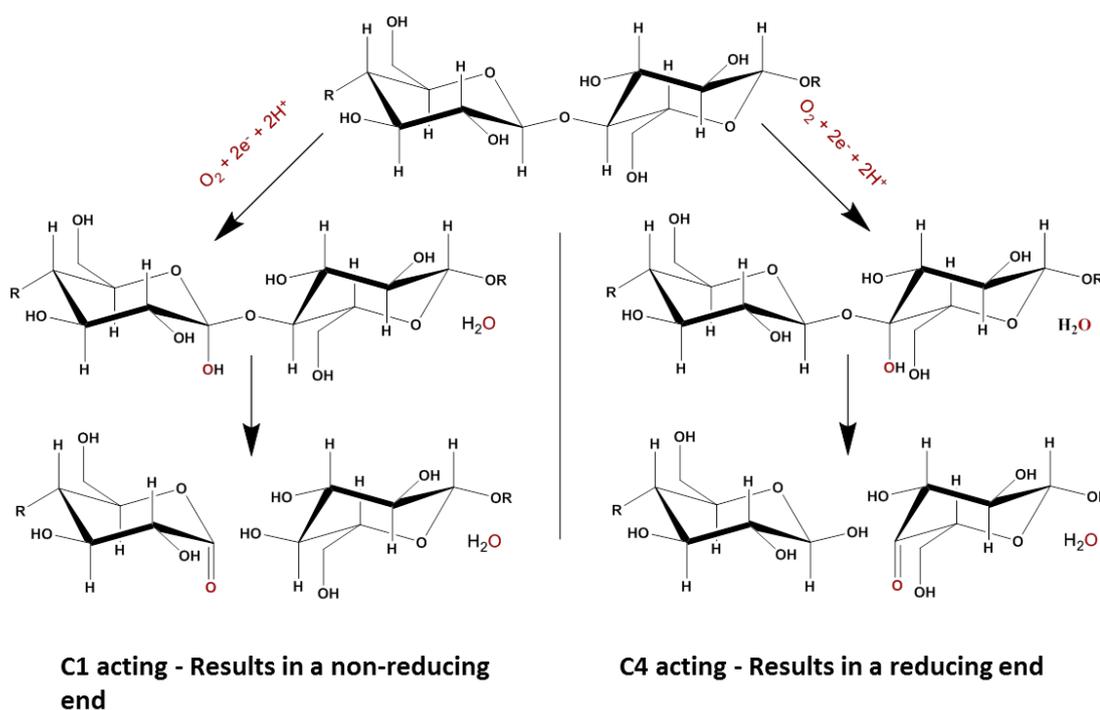


Figure 1-12 – Reaction pathway of the LPMOs. LPMOs can cleave at either the C1 or C4 position of the sugar residue, resulting in reducing or non-reducing end product.

LPMOs are copper containing oxidative enzymes that possess a common feature, termed the histidine brace (Hemsworth et al., 2013). This brace, consisting of two histidines, one always present at the N-terminus of the protein, and a tyrosine, coordinate the copper-containing active site. It is this unique active site that allows LPMOs to depolymerise polysaccharides, by making chain breaks at either the C1 or C4 position of a substrate's repeating unit (Figure 1-12).

Chain breaks are induced through the insertion of molecular oxygen, activated by the copper cofactor. To function, this copper active site requires an electron donor. Much of the in vitro characterisation work performed on these proteins used small molecular reductants, such as gallic acid or ascorbate, though there has been evidence to suggest that this role is played, in the natural system, by cellobiose dehydrogenases (Phillips et al., 2011, Langston et al., 2011). These enzymes have been shown to activate LPMOs in several species, and their deletion in *Neurospora crassa* resulted in a considerable loss of cellulose degradation activity (Phillips et al., 2011). However, as LPMOs are produced by fungi that do not have any cellobiose dehydrogenases, it is clear there must be other electron donors in other fungal systems. Lignin degradation products have been proposed as alternative electron donors, due to evidence that

LPMOs do not require an additional electron donor if the substrate is in contact with lignin (Kracher et al., 2016).

Alternative roles for cellobiose dehydrogenase have been proposed to degrade polysaccharides as well; they may be involved in relieving product inhibition or producing hydrogen peroxide for lignin degrading reactions like the Fenton reaction or for the action of peroxidases (Phillips et al., 2011, Henriksson et al., 2000).

1.7.5 Carbohydrate binding module

Many CAZymes are multi-modular and contain catalytic domains that are linked, via a threonine and serine-rich peptide, to carbohydrate binding modules (CBMs). These modules act to bring the catalytic domain into close contact with the appropriate substrate, allowing a high concentration of the enzyme to be maintained on the carbohydrate's surface. It has also been suggested that they aid the action of the enzyme by disrupting the structure of the substrate, although this is an area of some contention (Boraston et al., 2004).

To date, 80 distinctive carbohydrate domains have been described in the CAZyme database, encompassing 78,874 sequences. These families have specificity to a wide range of carbohydrates, and are found in association with all classes of CAZyme catalytic domains. They can be broadly described as belonging to three sub-types, based on their structure. Type A CBMs bind to crystalline chitin or cellulose and have flat faces. Type B have grooves to bind internally on single chains in amorphous regions of glycans, whereas Type C binds to ends of the glycan chains (Gilbert et al., 2013). There has also been evidence suggesting that the CBMs can direct the protein to specific regions within the plant cell wall, including those that are being actively degraded (Montanier et al., 2009).

Investigations to define the importance of CBM have demonstrated that the truncation or removal of the CBM can have a major effect on the activity of the remaining catalytic domain. The removal of the CBM was shown to decrease the rate of degradation of insoluble substrates (Arai et al., 2003, Wu et al., 2007). However, there are also many examples in which the catalytic domain is not appended to a CBM. This may present an evolutionary advantage in environments that have large amounts of dry matter, or in systems with high solid loadings, because the enzymes here are in closer contact with their substrates.

1.8 Microbial lignocellulose degraders

Bacteria and fungi have achieved the ability to degrade lignocellulose in environments ranging from soil and compost, to the guts of ruminant animals. Different groups of organisms have developed different strategies of lignocellulose depolymerisation. Within bacteria, there appear to be two main strategies employed by either anaerobic or aerobic organisms. Aerobic bacteria secrete hydrolytic enzymes into the extracellular environment. Anaerobic microbes can have a structure called a cellulosome attached to their own cell wall, where hydrolytic enzymes are held in place by interactions between cohesins and dockerins (Lynd et al., 2002). These structures can be formed out of hundreds of hydrolytic enzymes, and a simplified representation is shown in Figure 1-13. They are thought to benefit the organism by bringing synergistic enzymes in close contact, and by coordinating the efficient targeting of the substrate (Fontes and Gilbert, 2010).

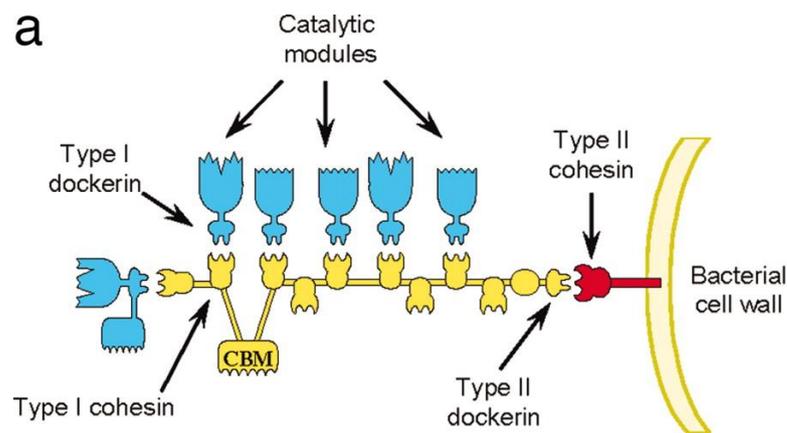


Figure 1-13 – Representation of a bacterial cellulosome sourced from Carvalho et al. (Carvalho et al., 2007).

There is some uncertainty over the extent to which bacteria can degrade lignin. Several papers in recent years have suggested that bacteria are capable of modifying the lignin structure (Bugg et al., 2011a, Taylor et al., 2012, Bugg et al., 2011b, Chen et al., 2012). Fungi capable of wood biodegradation can be separated into three broad physiological groups by their method of decay: white rot, brown rot and soft rot.

White rot fungi, named for their ability to degrade the dark lignin, often selectively leave behind pale cellulose, and are the only grouping that has been shown to degrade and mineralise lignin fully. Lignin, as discussed in Section 1.4.3, is incredibly recalcitrant to degradation and is a very stable structure with a high oxidative potential. It contains multiple

types of random linkages that prevent the evolution of a specific enzyme active site. The depolymerisation of lignin that white rots mediate is, therefore, catalysed by radical chemistry, and four enzymes involved in this process have been identified. Lignin peroxidases are heme-containing glycoproteins which rely on activation by hydrogen peroxide for the oxidative depolymerisation of lignin. Manganese peroxidases are also heme-containing glycoproteins, similar to lignin peroxidases but requiring the oxidation of Mn^{2+} to Mn^{3+} to complete the catalytic cycle. Finally, versatile peroxidases have the oxidation ability of both lignin peroxidases and manganese peroxidases. In addition to these peroxidases, laccase - a multicopper enzyme associated with detoxification, has also been identified as having an involvement in lignin degradation (Martinez, 2002, Cullen and Kersten, 1996).

Brown rot fungi are thought to have evolved from white rot fungi, losing key enzymes along the way to cast off the energetically expensive lignin removal strategies. Instead, brown rot fungi rely on non-enzymatic chemical processes such as the Fenton reaction to disrupt the structure of ligninocellulose (Floudas et al., 2012, Hori et al., 2013, Eastwood et al., 2011). Evidence for this mode of attack was described in the brown rot fungi *Gloeophyllum trabeum*, where it was observed to depolymerise lignocellulose despite its enzymes being physically separated from the substrate (Green and Highley, 1997).

The Fenton reaction occurs when Fe^{2+} is oxidised to Fe^{3+} by a molecule of hydrogen peroxide, resulting in the generation of a hydroxyl radical (**Error! Reference source not found.**). This hydroxyl radical is a powerful oxidising agent, and is capable of depolymerising the holocellulose fraction of the plant cell wall whilst modifying the lignin structure (Arantes et al., 2012).

Equation 1 - The Fenton Reaction



The Fenton reaction is chemical and can depolymerise cellulose and hemicellulose in the absence of hydrolytic enzymes. This is an advantage when degrading a composite material like lignocellulose, where the lignin fraction physically blocks the access of enzymes to the polysaccharides. Brown rot fungi, unlike soft rot or white rot, can therefore extensively depolymerise cellulose and hemicellulose, whilst keeping the morphology of the plant cell wall largely unchanged and without causing substantial lignin removal. This is because the Fenton reaction can be initiated at sites proximal to the substrate (Green and Highley, 1997). However, although enzymes are not directly involved in this mode of attack, the reaction requires very precise conditions and the fungi that utilise it must have mechanisms to create

these. These include the ability to reduce and solubilise iron, to create an acidic environment, and to produce H₂O₂ (Arantes et al., 2012).

Another mechanism of attack is shown by soft rot fungi. Unlike white rots and brown rots, which are typically basidiomycetes, soft rots are from the ascomycetes and deuteromycete phyla. They are often reported as having a higher tolerance to pH fluctuations and high moisture contents. Consequently, they have been isolated from a wide variety of environments (Kim and Singh, 2000). Commonly studied soft rot fungi include the industrially relevant species of the *Trichoderma*, *Penicillium*, *Fusarium* and *Aspergillus* genus (Hamed, 2013, Ravalason et al., 2012, Huang et al., 2015, Adav et al., 2010, Deshmukh et al., 2016). These fungi are capable of extensively degrading lignocellulose through the secretion of large amounts of enzymes, close to the site of attack due to the penetrating nature of their filamentous hyphae. This causes the characteristic softening of the lignocellulose as the plant cell walls lose their structure. They are not, however, known for their ability to deconstruct lignin, and though there have been reports that they possess capacity to modify linkages found in lignin, it is not clear to which extent this occurs.

Due to the ability of soft rot fungi to secrete large volumes of enzymes, they have been used for the preparation of industrial saccharification cocktails. Historically, one of the most studied ascomycete fungi for biofuel production is the sordariomycetes *Trichoderma reesei*. Originally isolated during World War II in the Solomon Islands, it was first studied due to its ability to degrade tents and clothes belonging to the U.S Army. Since then, it has undergone numerous genetic modifications; first through random radiation mutagenesis, and more recently through molecular cloning and targeted deletion (Bischof et al., 2016, Juturu and Wu, 2014). This strain modification has increased the saccharification capacity of this model organism, and has been enabled, in part, by a detailed understanding of the organism's arsenal of cellulolytic enzymes, as well as their interactions (Peterson and Nevalainen, 2012).

Neurospora crassa, *Aspergillus niger* and *Chaetomium globosum* as well as others, have also undergone detailed molecular studies. These have provided information on the physiology and enzymatic systems of these organisms (Sun et al., 2012, Phillips et al., 2011, Adav et al., 2010, Glass et al., 2013). Cellulases are also commonly heterologously expressed for detailed biochemical analysis, and information gathered from these investigations feed through into the preparation of more effective cellulase cocktails (Laothanachareon et al., 2015).

However, only small subsets of cellulolytic organisms have been studied in detail. The vast

majority of enzymes used in the saccharification reaction during bioethanol production originate from *T. reesei* and *A. niger*. It is, therefore, likely that species cultured from different environments will contain enzymes with differing characteristics. These may include pH and temperature optima, and mechanisms that may prove vital in the development of second generation biofuels. Hence information gained through discovery and characterisation of novel lignocellulolytic organisms from diverse environments is an important resource for the second generation biofuel industry.

1.9 Aims of this thesis

Novel lignocellulolytic enzymes and proteins are of interest for the production of second generation bioethanol. This thesis aims to investigate the microbial members of composting communities that have been enriched for the breakdown of wheat straw with a particular focus on discovery of novel lignocellulolytic enzymes.

The thesis describes the identification and isolation of both bacterial and fungal members of the composting community. High-throughput sequencing techniques are used to describe the dynamics of these cultures over an eight-week time-course. Subsequently, a single member of the compost community that showed significant lignocellulose degrading abilities was studied in isolation. A comprehensive review of its cellulolytic system was performed, using a combination of transcriptomics and novel proteomic techniques. This allowed the identification of a diverse arsenal of cell wall attacking enzymes, as well as novel proteins that may prove relevant to the production of new, cost-effective saccharification cocktails.

2 Methods and Materials

2.1 Chemical reagents and substrates

Reagents and chemicals described in this thesis were purchased from Sigma-Aldrich (Poole, UK), Fisher Scientific (Loughborough, UK), Melford (Ipswich, UK), New England Biolabs (USA), GE Healthcare (London, UK), Promega (Southampton, UK), Qiagen (West Sussex, UK), Expedeon (Swavesey, UK), Clontech Laboratories (USA), Invitrogen (Paisley, UK), Cambio (Cambridge, UK), Cambridge Biosciences (Cambridge, UK) and Agilent (USA).

All water used in experiments, unless otherwise stated, was obtained with an Elga PureLab Ultra water polisher (High Wycombe, UK).

2.1.1 Biomass

Wheat straw was milled to 2 mm particles at the Biorenewable Development Centre (York, UK). Xylan mentioned throughout this thesis was beechwood xylan purchased from Sigma-Aldrich.

Phosphoric acid swollen cellulose (PASC) was prepared from avicel (Sigma-Aldrich). Briefly, for every gram of Avicel used, 30 mL of 85 % phosphoric acid was added. This was stirred for one hour, on ice, before 100 mL of acetone per gram of avicel was added, filtered on a glass-filter funnel and washed three times with 100 ml ice cold acetone. This was washed, again, with 500 mL of water and homogenized with a blender.

2.1.2 Organisms

Escherichia coli strains Rosetta™ 2(DE3)(Novazymes) and Stellar™ Competant Cells (Clontech) were purchased for heterologous expression of proteins in bacteria. *Aspergillus niger* D15 was kindly received from Peter Punt (TNO, Netherlands) for heterolous expression of proteins in eukaryotes, and *Trichoderma harzianum* IOC 3844 for growth comparisons was gifted by Vanessa Rocha (Federal University of Rio De Janeiro, Brazil).

2.1.3 Buffers

Phosphate buffered saline was prepared to x10 concentration (NaCl 80 g/L, KCl 2g/L, Na₂HPO₄ 14.4 g/L, KH₂PO₄ 2.4 g/L) and diluted as necessary with H₂O. After preparation the solution was pH adjusted to 7.4, autoclaved and stored at room temperature.

Britton-robson buffer was prepared as a universal pH buffer; it was prepared with 0.04 M H_3BO_3 , 0.04 M H_3PO_4 and 0.04 M CH_3COOH . The desired pH was then achieved by the addition of 0.2 NaOH.

2.2 Microbiology methods

2.2.1 Preparation of Hutner's trace elements

Hunter's trace elements were prepared to the following specifications. - $\text{Na}_2\text{EDTA} \cdot 2\text{H}_2\text{O}$ 50 g/L, $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ 22 g/L, H_3BO_3 11.4 g/L, $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ 0.506 g/L, $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 0.4499 g/L, $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ 0.161 g/L, $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ 0.157 g/L, $(\text{NH}_4)_6\text{Mo}_7\text{O}_{24} \cdot 4\text{H}_2\text{O}$ 0.110 g/L. Trace elements were stored away from light at 4 °C.

2.2.2 Preparation of media

The minimal media referenced throughout this thesis was based on *A. niger* minimal media and contained KCl 0.52 g/L, KH_2PO_4 0.815 g/L, K_2HPO_4 1.045 g/L, MgSO_4 1.35 g/L, NaNO_3 1.75 g/L, Hutner's trace elements.

2.2.3 Optimised Media

The optimised media for *Graphium* sp. growth consisted of yeast extract 8.55 g/L, KCl 0.52 g/L, KH_2PO_4 0.815 g/L, K_2HPO_4 1.045 g/L, MgSO_4 1.35 g/L, NaNO_3 1.75 g/L and Hutner's trace elements (Section 2.2.1).

2.2.4 Agar plate and slant preparation

Agar plates were prepared by the addition of 1 % (w/v) agar to the media of choice before autoclaving. The agar was then allowed to cool, before the addition of antibiotics, if appropriate and poured into 9 cm petri dishes.

2.2.5 Spore preparation and harvesting

Slants for spore growth were prepared using 10 mL of potato dextrose agar (PDA) in . Once solidified this were inoculated with a single colony of the fungus of interest and incubated at 30 °C, until sporulation had occurred. Then under sterile conditions, 500 μL of sterile saline solution (0.5 % (w/v) NaCl, 0.2 % (v/v) Tween 20) was pipetted on to the surface on the slant, and a sterile loop using to gently release the spores from the mycelium into the solution. This was then pipetted into a sterile Eppendorf tube and vortexed vigorously to break up any

remaining fungal structures. A 1/100 solution was then created with more sterile saline solution and 10 µL was pipetted onto a clean haemocytometer for counting (Section 2.2.6).

2.2.6 Spore and protoplast counting

A haemocytometer was routinely used for counting both spores and fungal protoplast. The sample for counting was well mixed and 10 µL was pipetted onto the clean haemocytometer as per the manufactures instructions. Under x10 magnification the number of cells were counted in the four squares located in the corners of the counting square (Figure 2-1), which have a surface area of 1 mm² and a volume of 0.1 mm³. Therefore, to determine the number of cells per mL of suspension **Equation 2** was used.

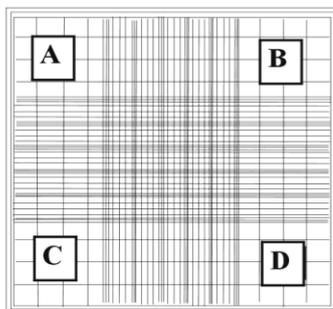


Figure 2-1 Layout of the haemocytometer for cell counting. Squares labelled A, B, C and D, which each have a volume of 0.1 mm³, were used for counting.

Equation 2 - Calculation for the number of cells per millilitre of suspension. N is the number of cells counted in four large squares on a haemocytometer.

$$\text{Number of cells per mL of suspension} = \frac{N}{4} \times 10^4$$

2.3 Molecular biology techniques

2.3.1 DNA extraction

2.3.1.1 Disruption of cells in composting cultures for DNA extraction

DNA was extracted from the composting cultures using a method adapted from Griffiths *et al* (Griffiths et al., 2000). From the composting shake flask 20 mL aliquots were harvested weekly.

The biomass was separated from the liquid fraction by centrifugation performed at 4000 xg at 4 °C, and 0.5 grams of biomass removed to a 2 mL screw cap tube. To this 500 µL of cetyltrimethylammonium bromide (CTAB) buffer (2% (w/v) CTAB 100mM Tris-HCl (pH 8.0), 20mM Ethylenediaminetetraacetic acid (EDTA) (pH 8.0), 2M NaCl, 2 % (w/v) polyvinylpyrrolidone (Mr 40.000), 5% 2-mercaptoethanol (v/v), 10mM ammonium acetate) was added along with 0.5 grams of zirconia beads and 0.5 mL of phenol: chloroform: isoamyl alcohol (25: 24: 1, pH 8.0), before briefly vortexing. The material was then bead beaten using a TissueLyser II (Qiagen) for 5 minutes at speed 28/sec. Samples were then processed as in Section 2.3.1.3.

2.3.1.2 Disruption of fungal cells for DNA extraction

Approximately 0.5 grams of fungal biomass was harvested through mirapore cloth after growth in shake flasks with an appropriate media (NA or PDA). The fungal biomass was ground under liquid nitrogen using a pestle and mortar, before 500 µL CTAB extraction buffer was added, and the sample transferred to a 2 mL screw cap tube. To this, 800 µL of a Phenol/Chloroform/Isoamyl alcohol (25:24:1) mix was added and vortexed briefly and samples processed as in Section 2.3.1.3.

2.3.1.3 Phenol-Chloroform DNA extraction method

A modified Phenol-Chloroform method was used to extract DNA after cell lysis. The sample was spun for five minutes at max speed to achieve separation of the phases, before the aqueous layer was removed to a fresh 2 mL Eppendorf tube. To the aqueous phase Chloroform:Isoamyl Alcohol (21:1) was added and this was spun and the aqueous phase transferred to a fresh tube, to remove any remaining phenolics. To precipitate the DNA within the sample, an equal volume of ice-cold 100 % isopropanol was added and incubated for 1 hour. DNA was pelleted by centrifugation at 13,000 rpm for 10 minutes and supernatant was removed without disturbing the pellet. The pellet was then washed with 80 % ethanol, before being re-suspended in DNase free water.

2.3.1.4 SV DNA purification columns for DNA extraction from prokaryotes

Wizard® SV DNA purification columns (Promega) were used to extract DNA from bacteria isolates after growth in an appropriate media (NA or PDA). An adapted protocol was performed. Briefly, 5 mL cultures of cells that had been grown overnight were centrifuged at 4,000 xg. Supernatant was removed and the cells were resuspended in 100 µL of lysis buffer (100 µg/mL lysozyme) and incubated at 37 °C for 30 minutes. To this 150 µL of Wizard® SV

Lysis Buffer was added and vortexed before being applied to a minicolumn. This was centrifuged for 3 minutes at 13,000 xg , and washed with 650 μL column wash solution, before being centrifuged again. The binding matrix was then dried with an additional centrifugation and transferred to a fresh 1.5 mL eppendorf, before 150 μL of nuclease free water was applied, incubated for 2 minutes, and centrifuged again for 1 minute to elute the DNA.

2.3.2 cDNA synthesis

cDNA was synthesised from RNA, that had been DNase treated with RTS Dnase kits (Mobio) using the standard protocol described by the manufacturer. cDNA synthesis was performed using SuperScript II Reverse Transcriptase (Invitrogen) kits and the standard protocol with 100 ng of random hexamers (Thermo Scientific) per 20 μL of reaction prepared.

2.3.3 Polymerase Chain Reaction (PCR)

PCR reactions were performed on a PTC-200 Peltier Thermal Cycler (MJ Research). Reactions were performed with Phusion® High-Fidelity DNA Polymerase as per manufactures instructions. The PCR reaction mix and temperature cycling used are described in Table 2-1 and Table 2-2 for 20 μL reactions.

Table 2-1 – Polymerase chain reaction components.

Component	Concentration	Final Concentration
Nucleases free H ₂ O	To 20 μL	N/A
5X Phusion HF buffer	4 μL	1X
10 mM dNTPs	1 μL	200 μM
10 μM forward primer	1 μL	0.5 μM
10 μM reverse primer	1 μL	0.5 μM
Template	-	-
Phusion polymerase	0.2 μL	0.4 units

Table 2-2 – Polymerase chain reaction thermocycling conditions.

Step	Temperature	Time
Initial denaturation	98 °C	30 seconds
Denaturation	98 °C	10 seconds
Annealing	As determined for primer	30 seconds
Extension	72 °C	30 seconds per kb
Cycling	Cycle between denaturation and extension	29 cycles
Final extension	72 °C	10 minutes
Hold	4 °C	N/A

Primers were purchased from Intergrated DNA technologies. Primer annealing temperature was calculated using ThermoFisher's T_m calculator (<https://www.thermofisher.com>).

2.3.4 Agarose gel electrophoresis

DNA fragments were separated by agarose gel electrophoresis. To prepare agarose gels 1 % (w/v) agarose was dissolved, by microwave heating, in 0.5 % TBE buffer. After cooling 0.00005 % (v/v) ethidium bromide was added and the solution was poured into a cast and the well comb added. This was then left to set for 30 minutes, before being placed in an electrophoresis tank containing 0.5 % TBE buffer and comb removed. An appropriate volume of sample buffer was added to DNA samples and mixed with pipetting before being loading into each well, alongside a commercial DNA ladder. An electric current was then generated at 120 Vs for 30 minutes by a BioRad PowerPac 3000 to migrate the negatively charged nucleotides towards the cathode. After completion UV illumination was used to visualise DNA bands using a UVitec gel documentation system.

2.3.5 Plasmid Extraction

Plasmids were extracted from transformed cells using QIAprep Miniprep Kits (QIAGEN, USA) following the manufacturers instructions.

2.3.6 PCR clean-up

DNA from PCR reactions were purified using according to manufacturer's instructions using Wizard PCR clean up kits and typically eluted into 30 μ L nuclease free H₂O.

2.3.7 Nucleotide Quantification

DNA and RNA were quantified using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific).

2.3.8 Sanger DNA Sequencing

DNA was sequenced by GATC Biotech, Germany, using the **LIGHTRUN** sanger sequencing service. Samples sent for this service contained 5 μ L of 80 - 100 ng of purified DNA in DNase-free H₂O along with 5 μ L of 5 μ M the appropriate primer. DNA fragments were sequenced using both forward and reverse primers to ensure complete coverage of the amplicon, reads were consolidated using BioEdit software.

2.3.9 In-fusion cloning

All cloning was performed using Clontech's InFusion system. This system fuses the gene of interest and linearized vector by recognising 15 bp complimentary regions at their ends.

15 bp overlaps on the gene of interest were added to the target gene, and destination vector was linearised via PCR. Agarose gel electrophoresis was used to confirm that a single band of the correct size had been obtained for both the GOI and vector.

The cloning reaction was set up as per the maufactors instructions with 2 μ L In-Fusion HD Enzyme Premix, 2 μ L PCR linearised vector, 1 μ L PCR fragment and 1 μ L Cloning Enhancer, before being brought up to 10 μ L with 10 μ L DNase free water. The reaction was then mixed, and incubated for 15 minues at 37 °C, followed by a 15 minute incubation at 50 °C and placed on ice. The 2 μ L of this reaction was then immediately use for the transformation of competent cells, whilst the remaining was stored at -20 °C.

2.3.10 Transformation of competent cells

Competent cells were thawed on ice, whilst 2 μ L of In-fusion cloning reaction was pipetted into round-bottom 10 mL tubes. Once thawed, 50 μ L of cells were gently mixed with the cloning reaction and the tubes left on ice for a further 20 mins. The cells were then heat-shocked for 45 secs by placing them in a water bath, pre-heated to 42 °C before placing them back in ice for 2 mins. SOC, warmed to 37 °C, was then added to the transformation to a final volume of 500 μ L. The cells were left to shake (180 rpm) at 37 °C for an additional hour before 100 μ L was spread onto an agar plate with an appropriate antibiotic.

2.3.11 Periplasmic extraction

Periplasmic extractions were performed on transformed bacterial cells (Hemsworth et al., 2014, Neu and Heppel, 1965). Cells were harvested at 4,500 xg before 3 mL of 50 mM Tris-HCL, 20 % (w/v) sucrose at pH 8.5 was added for each gram of cell paste along with 25 µg/mL of lysozyme. The cells were gently resuspend and left to incubate for an hour on ice, shaking at 180 rpm at 4 °C. After which 1M MgSO₄ was added to a final concentration of 5 mM and again this was left for a further 30 minute incubation. Supernatant was then collected after the cells were pelleted by centrifugation at 4,500 g for 30 minutes. The remaining cell pellets were then washed with 3 mL of ice cold water per gram of cells that were originally harvested. This was left for 30 minutes on ice, before the cells were re-pelleted and the supernatant could be added to the fraction collected in the previous step.

2.3.12 Mini-prep

Plasmids were purified from transformed cells using QIAgen Miniprep kits using the standard protocol as described by the manufacturers. Plasmids were purified from cells that had been grown overnight in 5 mL of LB, shaking at 180 rpm in 37 °C, these were

2.4 Proteomic methods

2.4.1 Bradford Assay for protein quantification

The protein concentration in solutions was measured by the Bradford assay, performed in 96-well plate format. Briefly, 300 µL samples, diluted to be within the sensitivity range of the assay, were mixed with 10 µL of Quick-Start Bradford Protein Assay solution (Bio-Rad), and there absorbance measured after 10 minutes incubation at room temperature. Concentrations where then determined through the comparison of known protein concentrations of bovine serum albumin (BSA).

2.4.2 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated by size using sodium dodecyl sulfate polyacrylamide gel electrophoresis and ran on Mini-Protean Tetra cell apparatus (Bio-Rad, USA). The separating layer was prepared by mixing 2.5 ml 1.5 M Tris-HCl, pH 8.8; 100 µl 10% SDS; 3.3 ml 30 % acrylamide; 4.00 ml H₂O; 100 µl 10 % ammonium persulphate and 15 µl Tetramethylethylenediamine (TEMED). This was immediately pipetted into a 0.75 mm thick gel cast, and a 50 % isopropanol covering added, before being left to set. The isopropanol layer was then poured off and gel surface rinsed with H₂O, before the separating layer - prepared by mixing 1.25 ml 1 M Tris-HCl, pH 6.8; 100 µl 10 % SDS; 1.3 ml 30 % acrylamide; 7.35 ml H₂O; 100

10 µl 10 % ammonium persulphate and 10 µl TEMED, was pipetted over top and well-comb added. Once solidified, gels were assembled with the electrophoresis apparatus and tank filled with running buffer (25 mM Tris, 192 mM glycine, 0.1 % SDS, pH 8.3). Samples, mixed with an appropriate amount of sample buffer (250 mM Tris-HCl pH 6.8, 50 % glycerol, 0.5 % bromophenol blue, 10 % SDS, 500 mM 2-mercaptoethanol) and heated to 100 °C for 5 minutes, were loaded into the wells alongside a commercial protein marker and electrophoresis was performed at 100 V until the samples had migrated through the stacking layer, and subsequently 200 V until the dye front had ran off the gel. Gels were stained with InstantBlue (Sigma-Aldrich), as per manufacturer's instructions.

2.4.3 Protein concentration by centrifugation

Proteins were concentrated via centrifugation using 20 mL Vivaspin 20 10,000 MWCO (Sartorius) concentrators, with PES membranes. Samples were routinely spun in spin out bucket at 4,000 g until the desired protein concentration had been reached.

2.4.4 Buffer exchange by centrifugation

Buffer exchanges were performed using the vivaspin 20 10,000 MWCO as per the manufactures instructions using centrifugation.

2.4.5 Anion exchange chromatography

Anion exchange chromatography was performed using the ÄKTA fast protein liquid chromatography (FPLC) system (GE Healthcare, UK) on a DEAE-Sepharose FF column (GE Healthcare, UK). The column was equilibrated with 50 mM Tris buffer at a pH at least one value over the desired proteins pI. The sample containing the target protein was then loaded onto the column and eluted with a gradient of 50 mM Tris-HCl and 500 mM NaCl in 2 mL fractions.

2.4.6 Size exclusion chromatography

Size exclusion chromatography was performed on the ÄKTA using a HiLoad Superdex 75 16/60 column (GE Healthcare, UK) equilibrated with 50 mM Tris-HCl pH 7.5 and 200 mM NaCl. Fractions which contained protein peaks, according to the UV trace were evaluated by SDS-PAGE.

2.5 Data analysis

Data manipulation was performed in R (R Core Development Team, 2010), with the use of the following packages tidyR (Wickham, 2016), plyr (Wickham, 2011), dply (Wickham and Francois, 2016) and Biostrings (Pages et al.). All plots were produced using R and the packages gplot (Warnes et al., 2016) and ggplot2 (Wickham, 2009). Eulidecean clustering was performed with the package ape and vegan (Oksanen et al., 2016, Paradis et al., 2004).

2.6 Enzyme assays methods

2.6.1 Qualitative Plate Assays

Qualitative agar plate assays were used to screen for lignocellulose degrading abilities. All plates were prepared with 1.5 % agar and minimal media (Section 2.2.2) and autoclaved. Assays for the indication of lignin modifying activity including remazol brilliant blue, a azo-dye whose decolourisation has been linked to lignin-degrading enzymes (Archibald, 1992), and 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid (ABTS) an colorimetric assay for the detection of laccases (Oriente et al., 2015). Endocellulase activity was detected with the use of carboxymethylcellulose (CMC), as its degradation can be detected by the production of clearance zones after being dyed with congo red (Pointing, 1999).

Remazol brilliant blue agar was supplemented with 0.1 % (w/v) remazol brilliant blue with 0.2 % (w/v) glucose. Cellulase activity was detected with 2 % (w/v) CMC. CMC plates were dyed for 15 minutes with 2 % (w/v) congo red, before 5 % (v/v) acetic acid was applied to darken the area. For the detection of xylanase activity 0.1 % (w/v) Azcl xylan was used and release of blue dye visually assessed over time. Xylanase activity was alternatively detected using 2 % beechwood xylan and processing as the CMC plates.

2.6.2 Reducing Sugar Assay

The ability of an enzyme to cleave polysaccharides and produce products with reducing ends was assessed after incubating the enzyme with the 2 % (w/v) of appropriate polysaccharide substrate in 200 μ L of 50 mM sodium phosphate at the desired pH and temperature. Unless otherwise stated the pH used for assays was 6.8 and temperature 30 °C. Before and after incubation 10 μ L aliquots mixed with p-hydrobenzoic acid (PAHBAH), heated to 70 °C for 10 minutes, and colour change detected at 415 nm using a microtitre Tecan Safire2 plate reader (Lever, 1972). A stock solution of the appropriate monosaccharide was diluted from 1 mg/ml⁻¹ to and assayed to obtain a standard curve.

2.6.3 Glycospot plates

Glycospot plates were used to assay the degradation of select polysaccharides. Plates contain a chromogenic polysaccharide hydrogel (CPH) substrate, created from the dyeing of the polysaccharide of question with a chlorotriazine dyes via nucleophilic aromatic substitution and then crosslinking with 1,4-butanediol diglycidyl ether. The resulting hydrogels were then used to detect the breakdown of the polysaccharide in question through the release of the bound dye in a 96-well plate format (Kracun et al., 2015).

The reaction was performed according to manufacturer's instructions. Briefly, appropriate wells of the glycoplates were prepared by the addition of 200 μL of the supplied activation solution, and centrifuged at 2,700 $\times g$ for ten minutes. The activated wells were then washed twice with 100 μL of water to remove remaining stabiliser. A 145 μL reaction mixture, consisting of 25 mM ammonium acetate buffer, with and without 0.1 mg of protein, was applied to each activated well and incubated for 16 hours at 30 $^{\circ}\text{C}$, shaking at 150 rpm. After incubation, the reaction mixture in the glycospot spot plate was transferred into a clean 96-well optical plate via centrifugation (10 minutes at 2700 $\times g$), and dye release visualised with microtitre Tecan Safire2 plate reader at the appropriate wavelength.

2.6.4 Para-nitrophenol (*pNp*) assays

Cellulase activity was assayed with para-nitrophenol (*pNp*) linked substrates. Samples for assaying were incubated with the suitable *pNp*-substrate (0.5 mM) for 16 hours in 100 mM 6.8 sodium phosphate in reaction mixes totalling 50 μL . A 20 μL aliquot of this reaction mixture was then incubated with 50 μL 1M sodium carbonate. The absorbance of free *pNp* was then measured on the microtitre Tecan Safire2 plate reader at 405 nm and quantified through comparison with a standard curve of *pNp* dilutions.

3 Characterisation of a fungal compost isolate

3.1 Introduction

Composting is the process in which organic matter is decayed to a stable humus-like substance that can be used as a soil enhancer. The bioconversion of this organic matter is driven by complex microbial communities, which secrete an enormous diversity of enzymes to catalyse the breakdown of the biomass present. As lignocellulose is one of the main components of this plant biomass, these communities have evolved the ability to efficiently access the carbon locked into this complex material and are promising sources of cellulases, hemicellulases and lignin modifying enzymes (Martins et al., 2013, Ventorino et al., 2015, Tuomela et al., 2000, Allgaier et al., 2010).

Typically, the first stage of the composting process is a thermophilic, self-heating, stage in which the temperature can reach up to 80 °C as the easily accessible carbon and proteins stored within the plant biomass are metabolised. Following this, there is a cooling and maturation period where the temperature drops between 20 - 50 °C. This latter stage is when much of the recalcitrant lignocellulose matter is decayed (Yu et al., 2007, Lopez-Gonzalez et al., 2014), and consequently is the stage of most interest for the bioprospecting of novel lignocellulolytic enzymes.

Though there have been numerous studies on the microorganisms present in compost, characterising compost communities and identifying novel proteins from them is a challenging endeavour. The communities within compost are variable and are dependent on multiple factors, including starting material (Simmons et al., 2014), moisture content (Wang et al., 2015, Zhang et al., 2011), and carbon to nitrogen ratio (Eiland et al., 2001). The organisms present are also form complex biological interactions with each other, further complicating their study.

Not all members of the composting community are expected to have the ability to directly metabolise lignocellulose as some may act as secondary consumers, aiding the recycling of bacterial and fungal biomass. Those that can metabolise lignocellulose may also show a preference for specific polysaccharides, creating a system in which enzymes from different organisms have a synergistic effect when combined (Deng and Wang, 2016, Taha et al., 2015). Other organisms form antagonistic relationships through the production of antimicrobial compounds as they compete for limited resources (Sotoyama et al., 2016).

The study of these complex communities is further hampered, as many microorganisms from natural environments are likely not to be culturable (Rappe and Giovannoni, 2003, Fracchia et al., 2006). However, recent advances in next-generation sequencing techniques have resulted in an increased understanding of these systems (Carvalhais et al., 2012).

The composition of complex communities can now be elucidated by the targeted amplicon sequencing of ribosomal DNA, and meta -genomic, -transcriptomic and -proteomic studies have allowed lignocellulolytic enzymes to be identified and characterised (Allgaier et al., 2010, Dougherty et al., 2012), as well as revealing the presence of numerous proteins whose roles have not been defined (Rigden et al., 2014).

Isolation of key members of the compost communities, however, is not without merit. Studying the organism in isolation allows for a more complete characterisation of the species, and its physiology. Growth conditions can be manipulated and optimised in monocultures in ways that are not feasible in community-based experiments, where the inoculum may fluctuate across repeats, and genetic techniques can be used for strain optimisation. In the case of *T. reesei*, cellulase activities were increased after the fusion of a bacterial endocellulase catalytic domain to its native cellobiohydrolase by 39 % (Liu et al., 2008), and it has been suggested that the most appropriate organisms for the production of enzymes in biotechnology will be those that have undergone significant strain modification.

In this chapter, a compost community is studied for its lignocellulolytic abilities, and wheat straw enriched compost communities analysed to identify organisms, which then can be studied in monoculture for the discovery of novel enzymes.

3.2 Methods and materials

3.2.1 Isolation of bacteria and fungi from an enriched wheat straw degrading community

Two litre shake flasks, containing 1 L minimal media (Section 2.2.3) and 5 % (w/v) milled wheat straw, were inoculated with 1 % (w/v) compost. The inoculum was collected from composting wheat straw that had been developed over the period of a year and watered at regular intervals. The inoculum was prepared by blending until homogenised and used on the day of preparation.

These shake flasks were incubated at 30 °C and shaken at 150 rpm. Weekly 10 mL samples containing both the solid and liquid fractions were aseptically collected over the course of eight weeks. The samples were then serially diluted with x1 PBS (Section 2.1.3) to concentrations ranging between 10^{-1} and 10^{-7} . From these dilutions 100 μ L samples were used to create spread plates on both nutrient agar (NA) and potato dextrose agar (PDA). After five days of incubation, isolates that appeared morphologically distinct were selected and transferred to agar plate libraries (Figure 3-1).

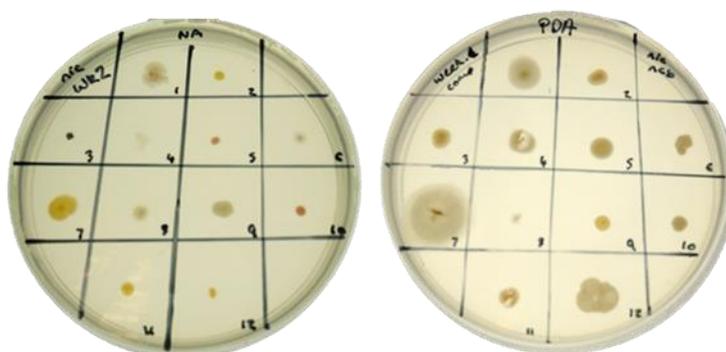


Figure 3-1 - Libraries of morphologically distinct isolates. Isolates were obtained via spread plates of a compost-inoculated wheat straw degrading shake flask culture on a weekly basis on either Nutrient Agar (NA) or Potato Dextrose Agar (PDA), before being stored in libraries.

3.2.2 Qualitative plate assays to identify potential lignocellulolytic activities

Isolates were transferred using sterile pipette tips from the agar plate libraries to CMC, azu-xylan, ABTS, and remazol brilliant blue agar plates to assess their potential to degrade plant cell wall components. The preparation and use of these assays are described in Section 2.6.1.

3.2.3 Identification of bacteria and fungi

Isolates that appeared to be active against at least one component of the plant cell wall were identified. Bacteria were identified from their 16S ribosomal RNA (rRNA) gene and fungi from their internal transcribed spacer 1 (ITS1), a region of DNA situated between the 18S and 5.8S. These regions were chosen as they are commonly used in microbial ecology as barcodes for microbial identification, and as such sequence databases curated from a diverse range of species have been established and are publically available (Schoch et al., 2012, Klindworth et al., 2013).

Ribosomal DNA regions were amplified by PCR performed on genomic DNA, harvested from bacterial isolates using SV DNA purification columns (Section 2.3.1.4), and fungal isolates using a modified phenol-chloroform method (Section 2.3.1.3). The PCR reaction was performed as described in Section 2.3.3 using the respective primers for fungal and bacterial ribosomal regions **Table 3-1**. Once a single product had been confirmed after visualisation on an agarose gel, the amplicon was purified from the PCR reaction mixture (Section 2.3.6) and submitted for Sanger sequencing at the University of York (Section 2.3.8).

Table 3-1 Primers for the amplification of ITS and 16S ribosomal DNA.

Target	Primers
ITS (ITS1-5.8S)	Fw - TCCGTAGGTGAACCTGCGG
	Rv - CGCTGCGTTCTTCATCG
16S (7f-1510r)	Fw - AGAGTTTGATYMTGGCTCAG
	Rv - ACGGYTACCTTGTTACGACTT

3.2.4 Phylogenetic analysis of compost isolates

Sequences obtained from bacterial isolates were BLASTn searched against the NCBI collection of 16S Ribosomal RNA sequences, and the fungal isolates against the complete non-redundant nucleotide database. Phylogenetic trees were created using the MEGA software (version 6) (Hall, 2013), and sequence alignments were performed using the MUSCLE method before a maximum-likelihood tree created (Edgar, 2004). Bootstrapping was conducted with 1,000 repeats to estimate the reliability of the clusters generated.

3.2.5 Targeted Amplicon Sequencing of 16S and ITS DNA

Genomic DNA was harvested from the compost cultures using a modified CTAB protocol adapted for use on materials with high phenolic contents (Section 2.3.1.3). Regions for amplicon sequencing were then amplified using PCR (Section 2.3.3) before being sequenced at the Biorenewables Development Centre (BDC), York, U.K. using an ion torrent platform.

3.2.6 Bioinformatics analysis of the compost community

Ribosomal DNA sequence data generated via targeted amplicon sequencing was analysed using the open access software Qiime on the University of York's Technology Facilities linux server (Caporaso et al., 2010). Each fastq file generated from the IonTorrent platform was first demultiplexed, and then converted into both fasta and qual file types using Qiime's python script `convert_fastaqual_fastq.py`. Then, to remove the primer sequences from the reads, the script `split_libraries.py` was used along with a mapping file generated as per Qiime's instructions. Low quality reads were removed by filtering out reads under 180 bp long and those without recognisable primers. As the fasta files contain a mixture of reads, starting from both the forward and reverse primer the script was run again on each sample, with the reverse primer labelled as the forward and vice versa. Once this was performed for every sample, the samples in which the reverse primer came first were reverse complemented to correct their orientation using the `adjust_seq_orientation.py` script. These orientation-corrected fasta files, along with the fasta files already in the correct orientation, were then combined into two files, one containing fungal sequences and the other bacterial, using BASH. Operational taxonomic units (OTUs) were then created from the fasta files. These files were picked using the open-reference OTU picking process, where reads are clustered against a reference sequence collection and any reads which do not show significant similarity to the reference sequences are subsequently clustered de novo. To perform this, the script `pick_open_reference_otus.py` was used. This step also includes taxonomy assignment, sequence alignment, and tree building steps. For the taxonomy assignments of bacterial sequences the default reference database was used, (greengenes gg_13_8 97_otus database) (DeSantis et al., 2006, McDonald et al., 2012), and for the fungal ITS sequences the UNITE (alpha release 12_11) database was used (Abarenkov et al., 2010).

3.2.7 Conditions for fungal growth

Fungi were grown in monoculture on multiple substrates. Liquid shake flask cultures were inoculated with spores at a concentration of 10^5 spores per mL (Section 2.2.65-6), then incubated at 30 °C, and shaken at 180 rpm.

Biomass and supernatant samples were collected by centrifugation at 4,500 rpm for 20 minutes at 4 °C. The supernatant was then gently poured off and, filter sterilised through 0.22 µm polyethersulfone (PES) filter units. Five 1 mL aliquots of the liquid fraction was then taken and stored in 1.5 mL Eppendorf tubes at -80 °C after flash freezing. The remaining

culture supernatant was stored in 50 mL falcon tubes at $-20\text{ }^{\circ}\text{C}$, after being used for agar plate assays (Section 2.6.1), reducing sugar assays (Section 2.6.2) and pH measurements.

3.2.8 Media optimisation of *Graphium sp.*

Media was optimised using a central composite design with rotation (Bezerra et al., 2008). It was optimised for the production of both cellulase and xylanase enzymes as assessed by measuring reducing sugar release after incubation on CMC and xylan. The concentrations of both sodium nitrate and yeast extract were varied as part of the optimisation and the matrix used is presented in **Figure 3-2**. This experimental design was chosen as a linear regression can be used to build a quadratic model of the response variable, in this case, cellulase and xylanase activity, without using a full three-level factorial experiment, thus reducing the number of replicates required. Statistica 6.0 software was used to create the experimental design and analyse the results.

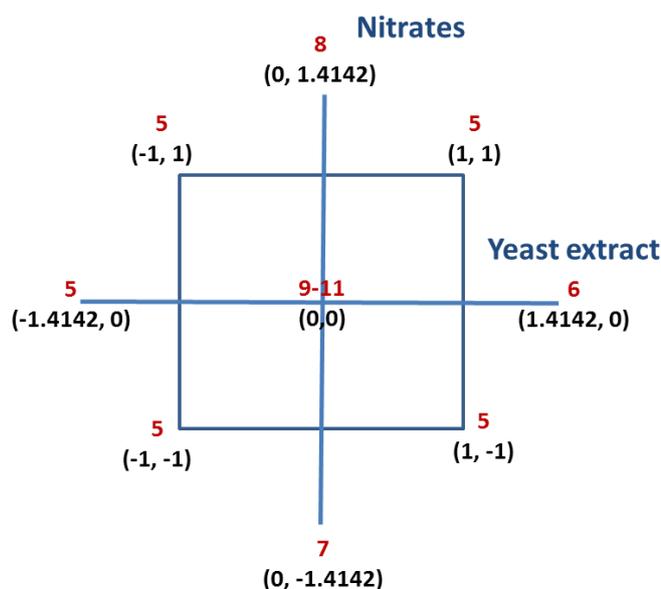


Figure 3-2 Central composite design for the optimisation of enzyme production from *Graphium sp.* Eleven cultures, indicated in red, were grown as part of this optimisation. These varied with yeast extract and sodium nitrate concentration according to the design matrix presented. The sodium nitrate concentration was varied between 0 g/L and 3.5 g/L and yeast extract was varied between 0 % and 1 % (w/v).

3.2.9 Estimation of growth through dried biomass

Growth of fungi was assessed using the dried weight of the biomass present within the culture. Cultures were transferred to pre-weighed and freeze-dried falcon tubes and chilled for 5 minutes. They were then centrifuged at 4,500 rpm, and the supernatant removed. The

biomass was gently rinsed with x1 PBS and tubes were flash frozen in liquid nitrogen and lyophilized. Each tube was then re-weighed to calculate the dry weight of the biomass present.

3.2.10 Estimation of growth by total protein

The total protein content of the cultures was used as an indicator of growth on insoluble materials such as wheat straw. Total protein was extracted by boiling 100 µg of freeze dried biomass in 1 mL of 0.2 % (w/v) SDS for 5 minutes to lyse all cells present. Protein was then collected by centrifugation at 14,000 rpm and the supernatant collected into a fresh 50 mL falcon tube. This was repeated three times, without heating, and with vigorous vortexing between each centrifuge step, to wash the biomass of any remaining protein.

Extracted protein was precipitated with five volumes of ice-cold acetone overnight at -20 °C, before being centrifuged at 4500 rpm and the resulting pellet washed with 80 % (v/v) ice-cold ethanol. The ethanol-protein mix was then centrifuged again, and the supernatant removed and the pellet air-dried. The protein was then solubilised in 3 mL of H₂O and quantified using the Bradford assay (Section 2.4.1).

3.2.11 Folin-Ciocalteu method for measurement of phenolics

The phenolic content of the culture supernatant was measured by the Folin-Ciocalteu method (Blainski et al., 2013). Commercially available Folin-Ciocalteu reagent was, on the day of use, diluted 1:1 with H₂O. 50 µL of this reagent, along with 250 µL 20 % (w/v) sodium carbonate (Na₂CO₃), was then added to 20 µL samples of the culture supernatant that had been diluted with 80 µL H₂O. These were then incubated at room temperature, for 30 minutes in Corning 96 Well Plates. Colour change was then detected at 650 nm using a microtitre Tecan Safire2 plate reader. Known quantities of tannic acid, diluted with water were used to obtain a standard curve, from which tannic acid equivalent (µg/mL) could be calculated.

3.3 Results

3.3.1 Lignocellulolytic fungal and bacterial isolates identified from the compost community

In total 15 bacterial isolates were identified that showed indications of activity against at least one of the three principal lignocellulose components (Figure 3-3). However, despite the time course being run over a period of eight weeks, bacteria with the desired activities could only be identified from samples harvested during the first two weeks of incubation.

In total 14 out of the 15 bacterial isolates identified showed cellulase activity, five had xylanase activity, and two showed indications of lignin degrading potential. Many of the identified bacterial isolates showed genetic similarities to genera that have previously been confirmed as having lignocellulolytic activities. One of the most interesting of which was identified as a *Microbacterium* sp. a genus that has recently been described as lignolytic through the action of superoxide dismutases (Taylor et al., 2012). Other identified bacterial species included *Candidimonas bauzanensis*, a bacterium that has previously been isolated from activated sludge and has not as yet been recognised as being lignocellulolytic (Felfoldi et al., 2014), and *Pseudomonas composti*, a species described from composting communities, and the only prokaryotic isolate that showed a preferential ability to degrade xylan, rather than CMC (Gibello et al., 2011).

From the same composting shake flask fungi were also isolated, and these are described in Figure 3-4. These, unlike the bacterial isolates, remained active throughout the eight-week time course and notably one isolate that showed similarity to *Graphium* sp. was present throughout the time course and showed activity in all assays tested. Other notable fungal isolates included those that were identified as *C. globosum* and *A. niger*; both established cellulose degraders, and a cellulolytic yeast species with a 96 % identity to its closest relative, *Sakaguchia dacryoidea* (Sequence ID: KY105305.1), in the NCBI nucleotide collection database.

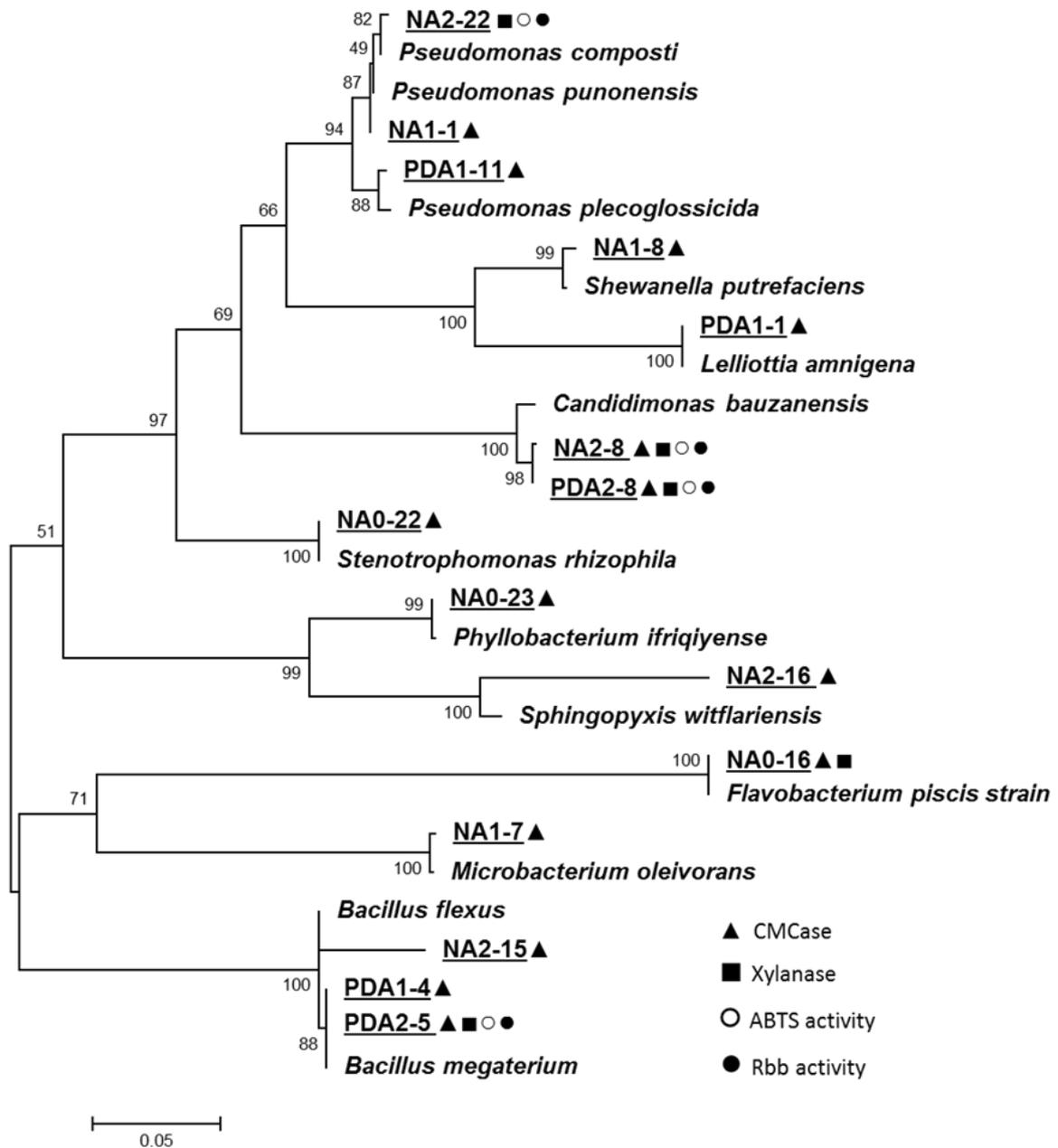


Figure 3-3 Molecular phylogenetic analysis of the bacterial isolates. The evolutionary history of the bacterial isolates and their nearest BLAST hit was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model, and the tree with the highest log likelihood (-4304.9790) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches.

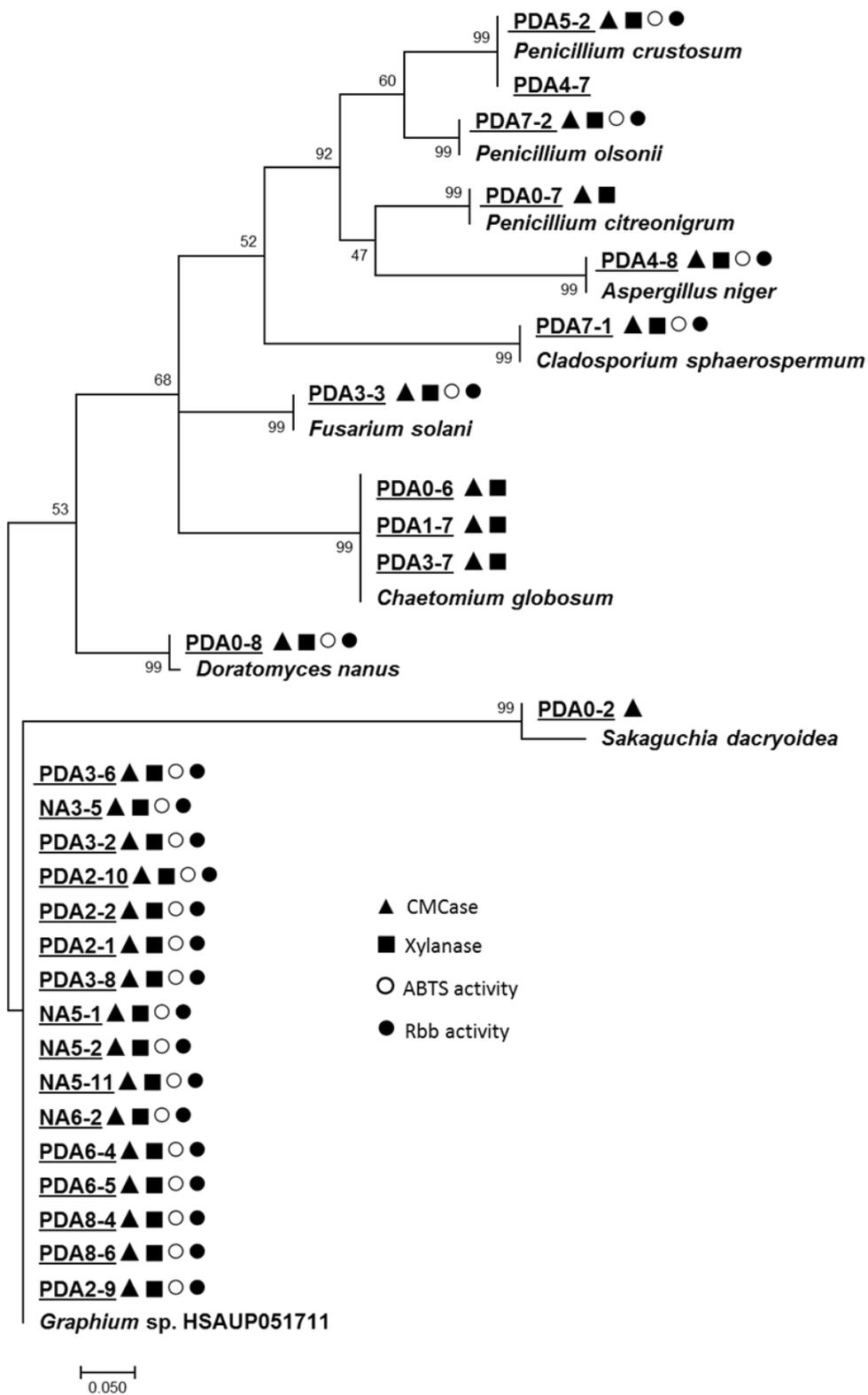


Figure 3-4 - Molecular phylogenetic analysis of the fungal isolates. The evolutionary history of the isolated identified as fungi and their nearest BLAST hit, was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model. The tree with the highest log likelihood (-4304.9790) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches.

3.3.2 Community profiling through targeted amplicon sequencing

Alongside the isolation work, the eukaryote and prokaryote communities within the compost were probed using targeted amplicon sequencing. The ITS region and the 16S region of the fungal and bacterial rRNA genes were amplified and sequenced using an ion torrent platform. Amplification of the ribosomal regions was performed by Anna Alessi (University of York, UK), and the sequencing was performed by the Biorenewables Development Centre (York, UK). Table 3 -2 contains the number of useable reads and the subsequent number of operational taxonomic units that were generated through the open source bioinformatics software Qiime, the process of which is described in Section 3.2.6.

Table 3-2 - Number of reads generated through targeted amplicon sequencing of ribosomal DNA. Targeted amplicon sequencing was performed for the identification of the prokaryotic and eukaryotic composting communities. Bacterial members were identified from the 16S ribosomal gene and the fungal members from the ITS1 region. The total number of reads generated from the ion torrent sequencing platform are listed for each sample, as well as the number of reads remaining after filtering for those in which the amplification primer could be identified. The filtered reads were then clustered into OTUs with 97 % similarity.

	Week	Reads	Filtered reads	OTUs
16S	1	570,143	387,324	10,190
	2	406,750	259,804	9,394
	3	544,289	362,234	8,761
	4	592,422	386,377	12,679
	5	527,888	154,962	4,431
	8	558,176	356,461	10,396
ITS	1	655,872	519,419	2,990
	2	563,242	439,733	3,156
	3	560,866	407,012	3,115
	4	576,44	452,384	3,466
	5	436,598	346,767	3,839
	8	419,384	319,775	4,872

3.3.3 Composition of the prokaryotic composting community

16S sequencing generated over three million reads, which were clustered together to form 25,304 OTUs, from which 182 genera could be identified (Figure 3-5).

The genera *Luteolibacter*, whose relative abundance remained approximately constant throughout the time course, and *Truepera*, a grouping that increased in relative abundance as the time course progressed, made up the majority of the reads identified to genus level. Strikingly neither have been

previously associated with lignocellulose degradation. Interestingly all the culturable bacterial isolates characterised as having lignocellulose degrading activities, whilst identified within the OTU libraries, were only present at low relative read abundances.

Although not all OTUs could be assigned to a genus, a large proportion could be identified to the phyla level. The most abundant phyla within the composting bacterial community were the gram-negative bacteroidetes, verrucomicrobia and proteobacteria; these represented an average of 31, 19.8 and 15.5 % of the total reads across the time course respectively. However due to the immense diversity contained within these prokaryote phyla generalisations about the role that these play within the community cannot be made.

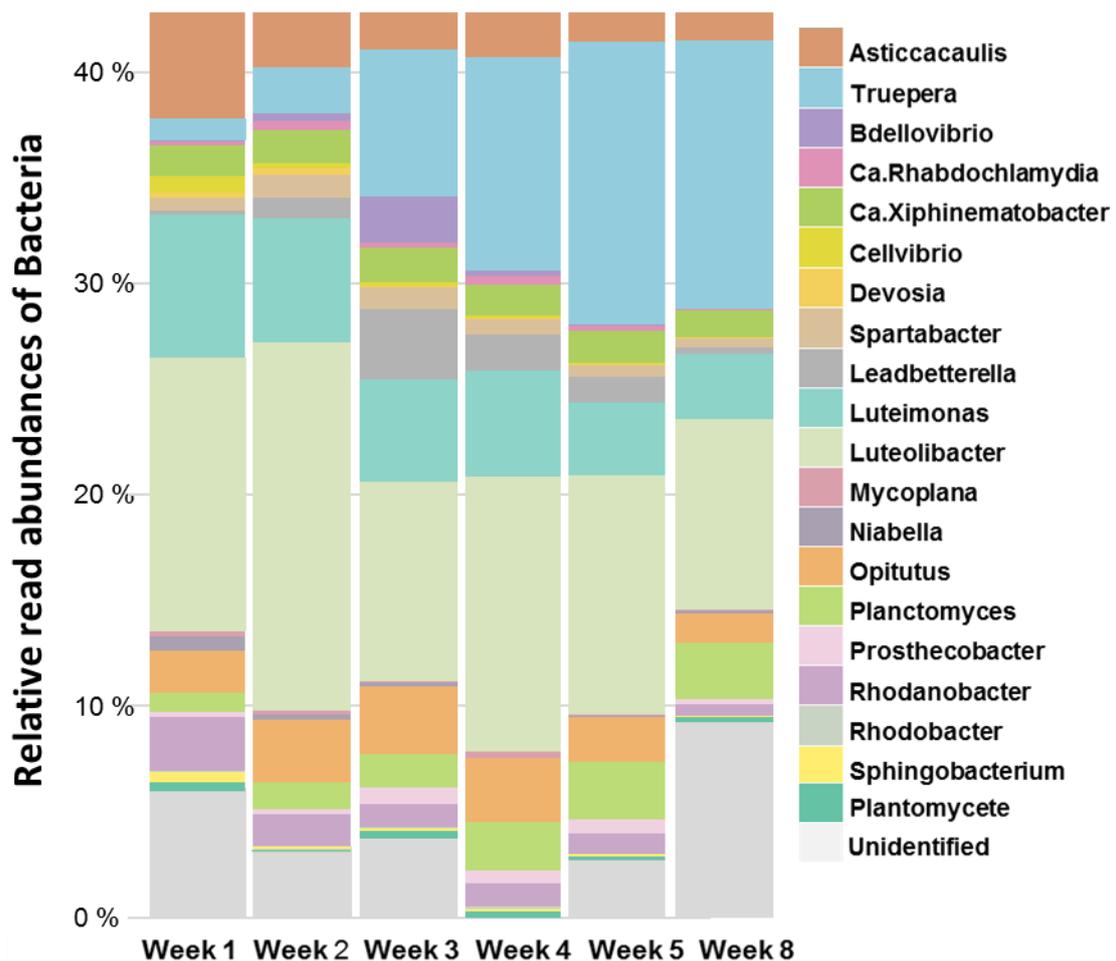


Figure 3-5 – The composition of the bacterial community as accessed by 16S ribosomal DNA. Sequences generated on the ion torrent platform after the amplification of a region of the 16S ribosomal DNA were processed with Qiime and identified to the nearest genus in the GreenGenes database.

3.3.4 Composition of the eukaryotic composting community

The eukaryotic community, as described by the sequencing of ITS region, predominantly consisted of reads that had no match within the UNITE fungal database, and in total 96.5 % of generated OTUs were not recognised as fungal. When the most abundant five OTUs, representing an average relative read abundance of 58.7 % of the total unidentified reads across all the timepoints, were BLASTn searched against the NCBI non-redundant database they showed homologies to protozoa ITS regions (Table 3-3).

Table 3-3 – Most abundant eukaryotic operational taxonomic units. Sequences generated by targeted amplicon sequencing of the ITS region where processed with Qiime to operational taxonomic units and BLAST searched to identify the species with the highest identity in the non-redundant NCBI library.

Sequence	Closest Match	Species	Identity	Query Cover
denovo5275	gb GU586187.1	<i>Epistylis chrysemydis</i>	87%	80%
denovo5614	gb KT358504.1	<i>Epistylis portoalegrensis</i>	91%	70%
denovo8367	gb KM222089.1	<i>Zoothamnium hentscheli</i>	87%	100%
denovo1426	gb AF429890.1	<i>Epistylis plicatilis</i>	89%	97%
denovo6006	gb AF429891.1	<i>Epistylis urceolata</i>	83%	100%

The remaining 252 OTUs which could be identified as fungal were able to be assigned to 30 genera (Figure 3-6), and the culturable isolates described in Section 3.3.1 were all identified in the sequencing.

Although fungi were the minority component of the community, a clear change in composition of the fungal community could be observed over the time course. The basidiomycota genus *Wallemia*, a grouping that contains species capable of growing in environments with high salt levels and low water availability dominated until week four. After which the Ascomycota genus *Graphium* rose to represent 25 % and 76 % of the identified fungal community in weeks five and eight.

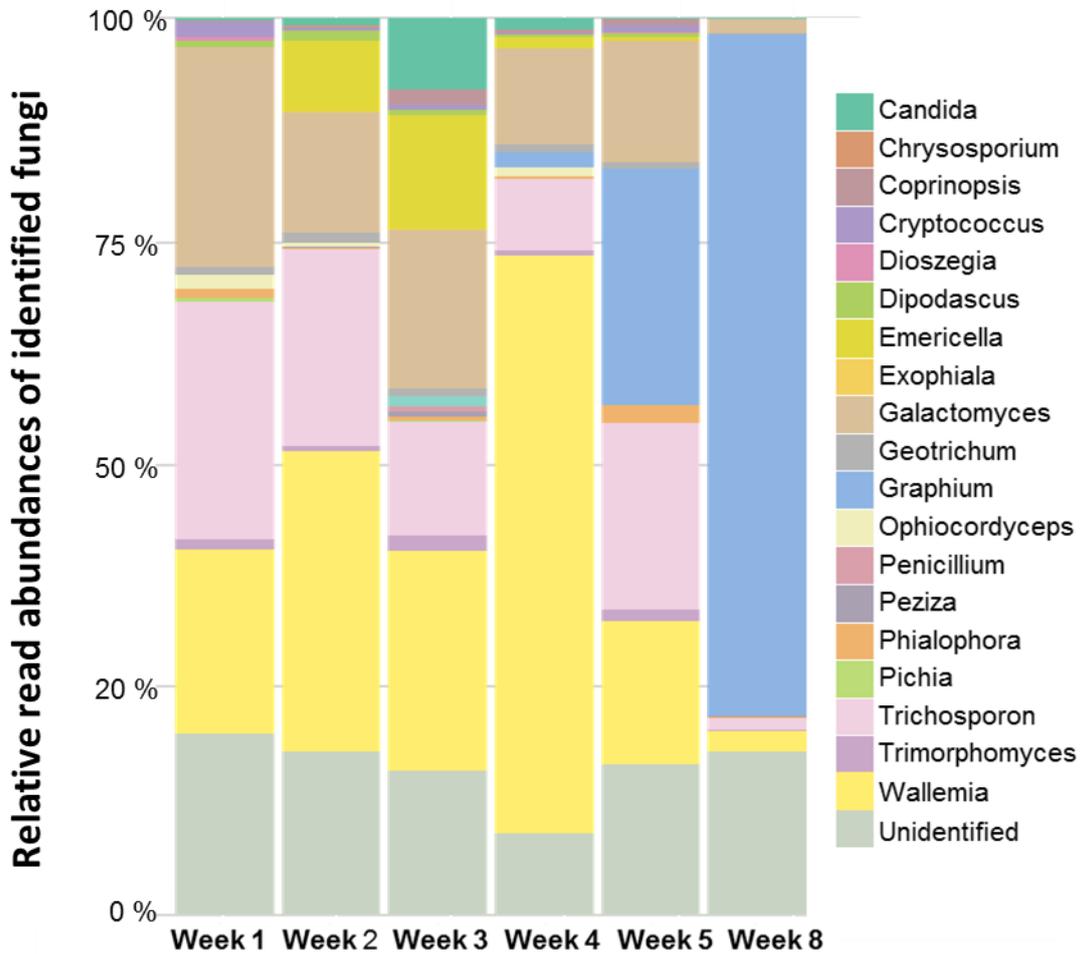


Figure 3-6 – The composition of the fungal community as accessed by ribosomal DNA targeted amplicon sequencing. Sequences generated on the ion torrent platform after the amplification of the ITS1 region were processed with Qiime and filtered to include only recognisable fungal sequences. These were then identified to the nearest genus in the UNITE database.

3.3.5 *Graphium* sp. growth in monoculture on lignocellulose derived substrates

As *Graphium* sp. was isolated from weeks two to eight of the compost time course and appeared as one of the most abundant members of the fungal community according to the ITS sequencing data, it was chosen to be studied in further depth. Its identity was confirmed through the sequencing of its 18S ribosomal region. This showed a 98 % match to *Graphium fructicola* (AB007659.1) and *Graphium tectonae* (U43907.1), in agreement with assumption that this fungus belongs to the *Graphium* genus.

Graphium sp.'s growth on various lignocellulolytic materials was assessed, as although it was isolated from a community growing on lignocellulose, it remained to be seen if it was able to utilise wheat straw as its sole carbon source in monoculture.

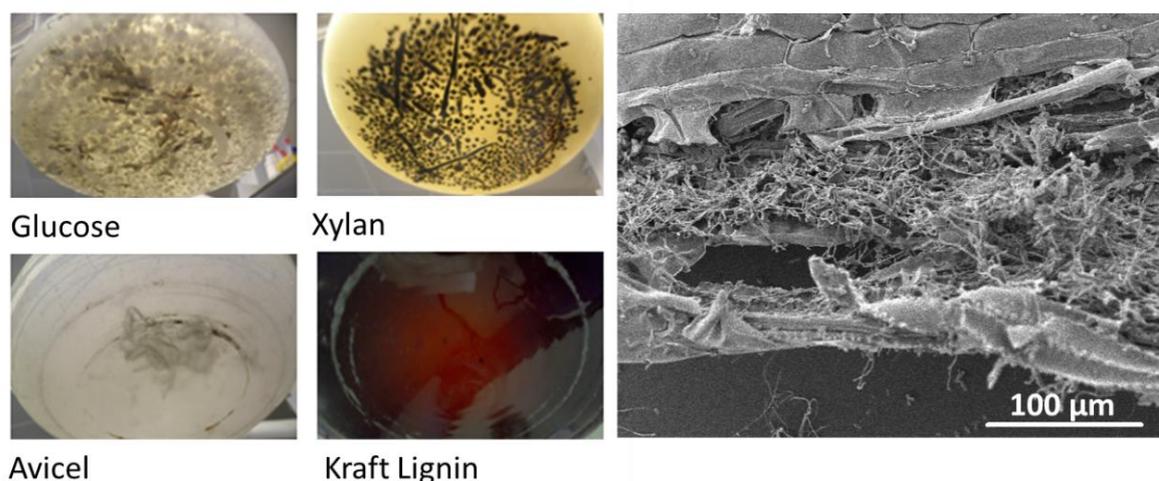


Figure 3-7 – Growth of *Graphium* sp. on lignocellulose derived material. *Graphium* sp. was grown in minimal media supplemented with 0.5 % glucose (A), 0.5 % beechwood xylan (B), 1.5 % avicel (C), 0.5 % kraft lignin (D) and photographed after seven days of growth. Samples taken from 1.5 % wheat straw cultures after seven days of growth were imaged using scanning electron microscopy.

To assess the growth of *Graphium* sp. on various carbon sources, spores were grown in 2 L shake flasks with glucose, avicell, xylan, kraft lignin or wheat straw as the sole carbon sources in minimal media, the preparation of which is described in Section 2.2.2. *Graphium* sp. appeared able to utilise all of these substrates as sole carbon sources for growth (Figure 3-7) and scanning electron microscopy, performed by the Technology Facility at the University of York (York, UK), confirmed extensive growth on wheat straw, with apparent penetration of the filamentous hyphae into the remaining structures of the plant cell walls.

3.3.6 *Graphium* sp. has increased tolerance to kraft lignin

Visual inspection of lignin cultures suggested that fungal growth had occurred, but enzymatic activity could not be estimated due to the dark colour of the supernatant. However it is unusual for a fungus to be able to grow using kraft lignin as a sole carbon source, both due to the difficulty in obtaining carbon for metabolism and inhibitory substances contained within this material. To test the tolerance of *Graphium* to lignin derivatives, cultures were grown on glucose media supplemented with varying percentages of kraft lignin. This experiment was also performed in parallel using the fungus *T. harzianum*, a well-studied industrially relevant fungal species (Section 2.1.2). The growth of *Graphium* sp. in high percentages of kraft lignin on glucose was not impaired and there was even a significant increase in biomass, in contrast to *T. harzianum* which showed a significant reduction in growth at higher lignin concentrations (Figure 3-8).

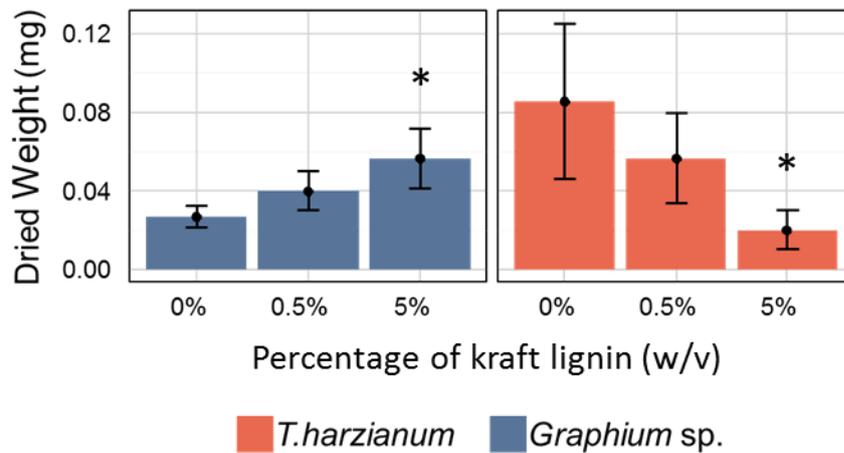


Figure 3-8 – Effect of kraft lignin on *Graphium* sp. and *Trichoderma harzianum* growth. Growth was assessed by dried weight measurements of *Graphium* sp. and *Trichoderma harzianum* after supplementation with varying percentages of kraft lignin. The cultures were grown for seven days in 200 mL flasks containing 25 mL minimal media with 0.5 % glucose (w/v) and supplemented with either 0 %, 0.5 % or 5 % kraft lignin (w/v). Cultures were grown in biological replicates, and error bars represent the standard deviations of these. Asterisks mark significant difference from the 0 % kraft lignin cultures assessed by one-way ANOVA to $P < 0.05$.

3.3.7 Media optimisation for growth of *Graphium* sp. to increase production of both cellulases and xylanases

Graphium sp. was successfully grown on wheat straw as a sole carbon source, using minimal media first described for the growth of *A. niger*. As this media has not been optimised for the growth of *Graphium* sp. on lignocellulose, there was scope to improve the number of enzymes that could be produced by this uncharacterised fungal species.

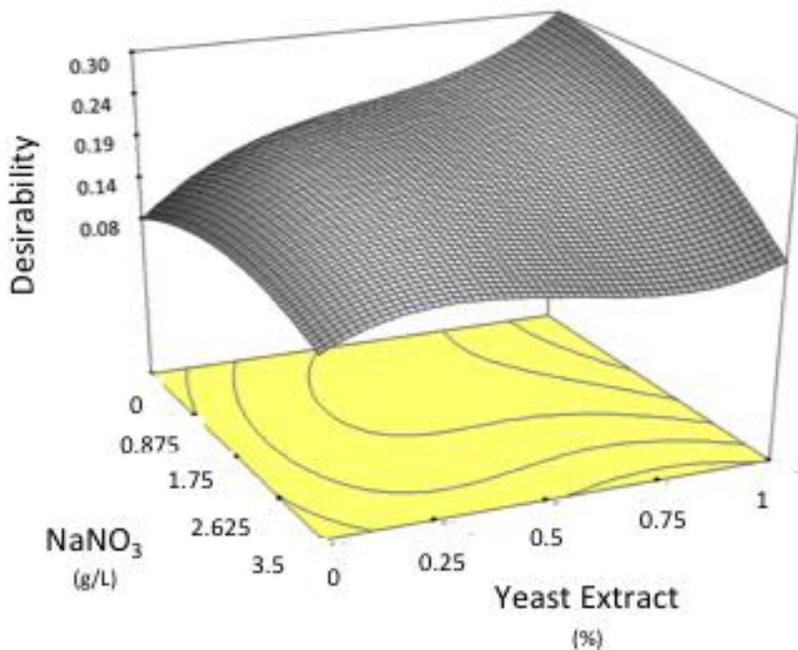


Figure 3-9 – Optimisation of *Graphium sp.* growth media. A central composite design was used to create a response surface morphology to varying concentrations of yeast extract and sodium nitrate. Both cellulase and xylanase production was improved with a high yeast extract and low nitrate concentrations.

Enzyme production was optimised for the growth of *Graphium sp.* on wheat straw, by varying nitrate and yeast extract concentrations, with the use of a central composite design with rotation. Low concentrations of sodium nitrate and high concentrations of yeast extract gave the highest enzyme activities; a two-fold increase in xylanase activity and threefold increase in cellulase activity could be achieved (Figure 3-9). The components of the optimised media are listed in Section 2.2.3.

3.3.8 Comparison of *Graphium sp.* and *Trichoderma harzianum* growth on lignocellulose derived substrates

The growth and enzyme production of *Graphium sp.* and *T. harzianum* on avicel, wheat straw and xylan were compared. Growth was performed in triplicate in 100 mL shake flasks containing 25 mL of the optimised media described in the previous section. Enzyme activities were calculated for all cultures on the first, second, fourth and seventh day of growth using reducing sugar assays. Growth was estimated, either by the dry weight of the fungal biomass or by calculating the total protein content.

Graphium sp. grown with either glucose or xylan produced a lower dried weight of biomass after seven days of incubation than *T. harzianum* grown under the same conditions. However, when

grown on wheat straw, more protein could be extracted from *Graphium* sp. cultures than from the *T. harzianum*, and the dry weight of the total biomass was higher – suggesting that *Graphium* sp. has a greater ability to metabolise wheat straw.

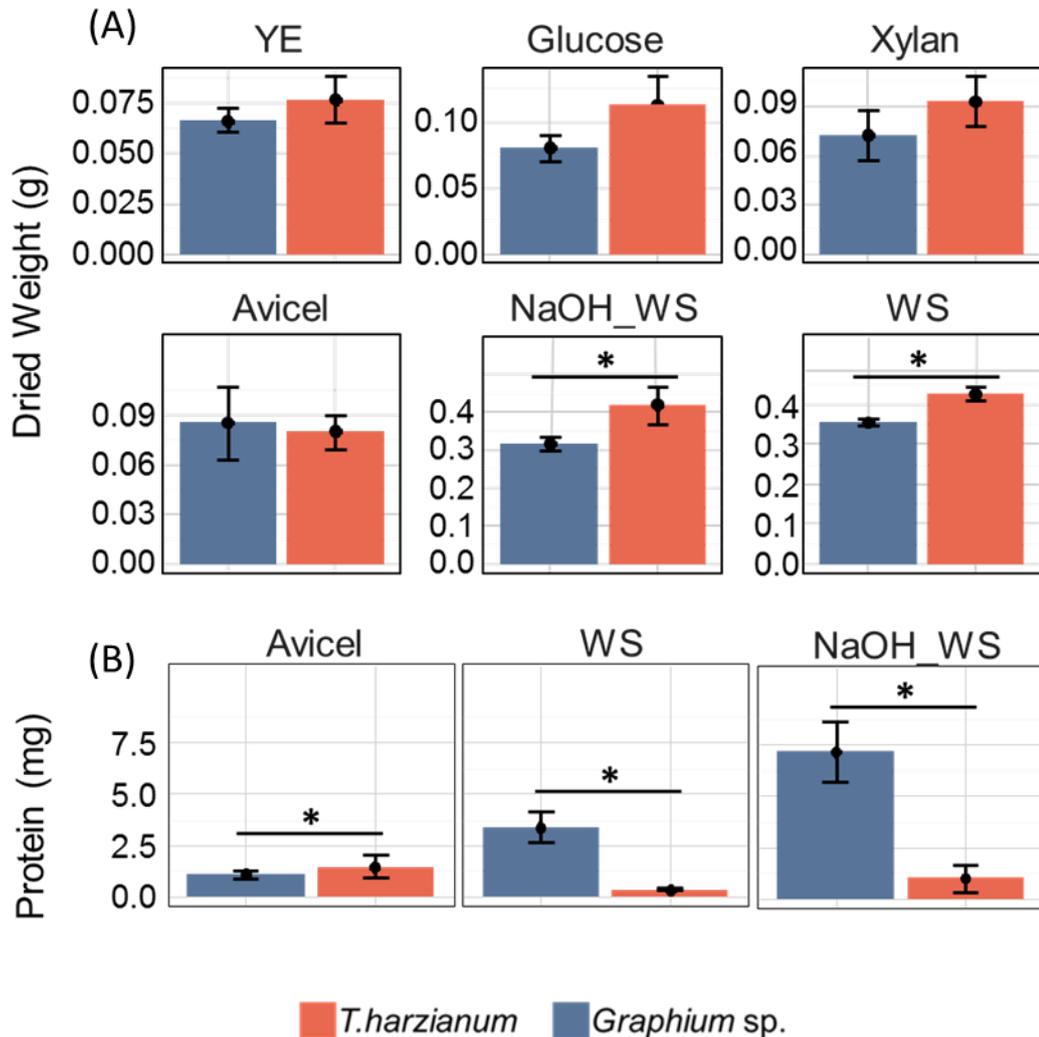


Figure 3-10 – Comparison of the growth of *Graphium* sp. and *Trichoderma harzianum* on lignocellulose derived substrates. (A) Dry weight of *Graphium* sp. and *T. harzianum* grown on the optimised medium with no carbon supplement (YE) or optimised medium with additions of either 0.5 % (w/v) glucose, 0.5 % (w/v) xylan, 1.5 % (w/v) NaOH treated milled wheat straw (NaOH_WS), or 1.5 % (w/v) milled wheat straw (WS), after seven days of growth. (B) Total protein content of the NaOH treated wheat straw and milled wheat straw grown cultures. Error bars represent standard deviations of three biological replicates. Asterisks mark significant difference as assessed by one-way ANOVAs to P<0.05.

Enzyme assays also suggested that *Graphium* sp. had a greater affinity for growth on wheat straw than *T. harzianum*. Supernatants harvested from *Graphium* sp. grown on wheat straw produced higher quantities of reducing sugars after incubation on either cellulose or xylan compared to the

supernatant from *T. harzianum*. Whereas the supernatant from *T. harzianum* grown in xylan contained the greatest xylanase activity compared to *Graphium* sp. (Figure 3-11).

To examine whether the lignin fraction of the wheat straw biomass was inhibiting the growth of *T. harzianum*, both fungi were grown on NaOH-treated wheat straw, a pretreatment that partially removes lignin from wheat straw. Although *T. harzianum* showed a slight increase in growth on the treated wheat straw compared to untreated, the total protein was significantly less than for *Graphium* sp., which had a two-fold increase in total protein on this substrate in comparison to the untreated wheat straw (Figure 3-10).

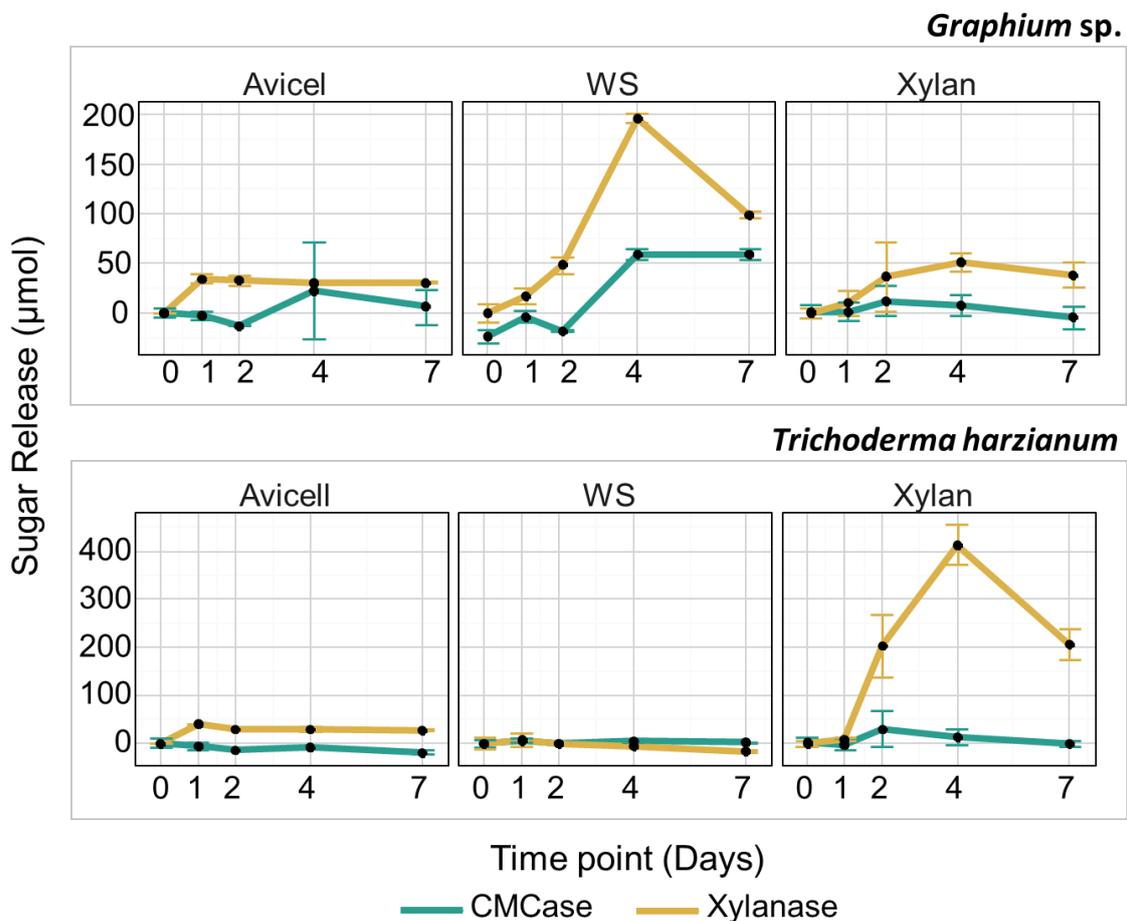


Figure 3-11 – Enzyme activities for *Graphium* sp. and *Trichoderma harzianum* cultures grown on different substrates. Enzyme activity from cultures grown in optimised media and 1.5 % avicel, 1.5 % wheat straw (WS) and 0.5 % xylan were estimated by calculating the release of reducing sugar after an one hour incubation on 2 % xylan or CMC. The supernatant to substrate ratio was 1:9 and the assays with performed in triplicate for each culture. Averages of each biological replicate were then used to calculate the standard deviation for error bars.

Graphium sp. may be able to grow better on wheat straw compared to *T. harzianum* due to an increased ability to deconstruct cellulose. Avicel grown cultures of both *Graphium* sp. and

T. harzianum had undetectable enzymatic activities against CMC and xylan in reducing sugar assays and no significant difference in the amount of protein secreted (Figure 3-10). However, agar plate assays revealed distinct clearance zones on both CMC and xylan after 16 hours of incubation with *Graphium* sp. supernatant which were not observed with *T. harzianum* supernatant (Figure 3-12).

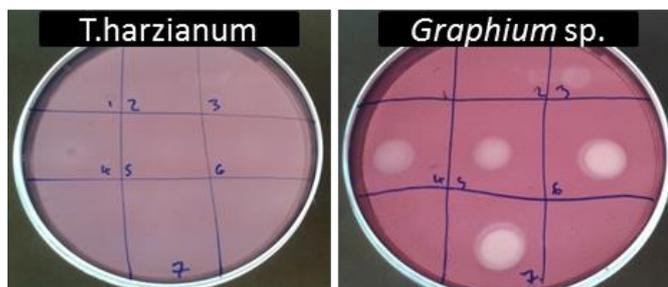


Figure 3-12 – Agar plate assays for endocellulase activity. Filtered supernatant from cultures of *Trichoderma reesei* and *Graphium* sp. grown on 1.5 % Avicel for seven day were assayed for the ability to degrade carboxymethylcellulose. From each culture 10 μ L of supernatant was plated on to agar and incubated for 16 hours at 30 °C before zones of degradation were visualised with Congo Red. Blue text indicates the timepoint (day) at which the samples were taken.

3.3.9 Extended time-course of *Graphium* sp. growth

Since *Graphium* sp. appeared particularly suited to growing on wheat straw, a time course was performed over a period of one month during which growth, enzyme activities, and biomass changes were observed.

Graphium sp. and *T. harzianum* were grown, as before, in 25 mL of optimised media with 1.5 % (w/v) wheat straw. Flasks were inoculated in triplicate with 10^{15} fungal spores per ml and incubated at 30 °C, shaking at 180 rpm. Biomass and supernatant samples were taken at ten different time points: days zero, one, two, four, seven, ten, 14, 18, 22 and 28.

The total protein content of these cultures peaked at the fourth day of growth (Figure 3-13A), after which the fungi appeared to decline. The dried weight of the insoluble fraction of the fungal culture is shown in (Figure 3-13B) and clearly demonstrated that *Graphium* sp. is capable of deconstructing the wheat straw structure, as the weight reduces below the control level. However the rate at which wheat straw biomass is lost and the time points during which this loss occurs, are unclear due to the masking effect of the fungal biomass.

The supernatant of the growth culture also changed over time. It increased rapidly in pH, rising from its starting point of 6.5 to nearly 9 (Figure 3-13C), and also darkened considerably. Accurate quantification of the amount of protein present within the supernatant was not possible, as a

pigment from the breakdown of lignocellulose (which caused the darkening) reacted with the Bradford assays (Halvorson and Gonzalez, 2008). The total phenolic content of the supernatant, as measured by the Folin-reagent assay, also increased over time, something that was not observed with *T. harzianum* when the same investigation was performed (Figure 3-13D).

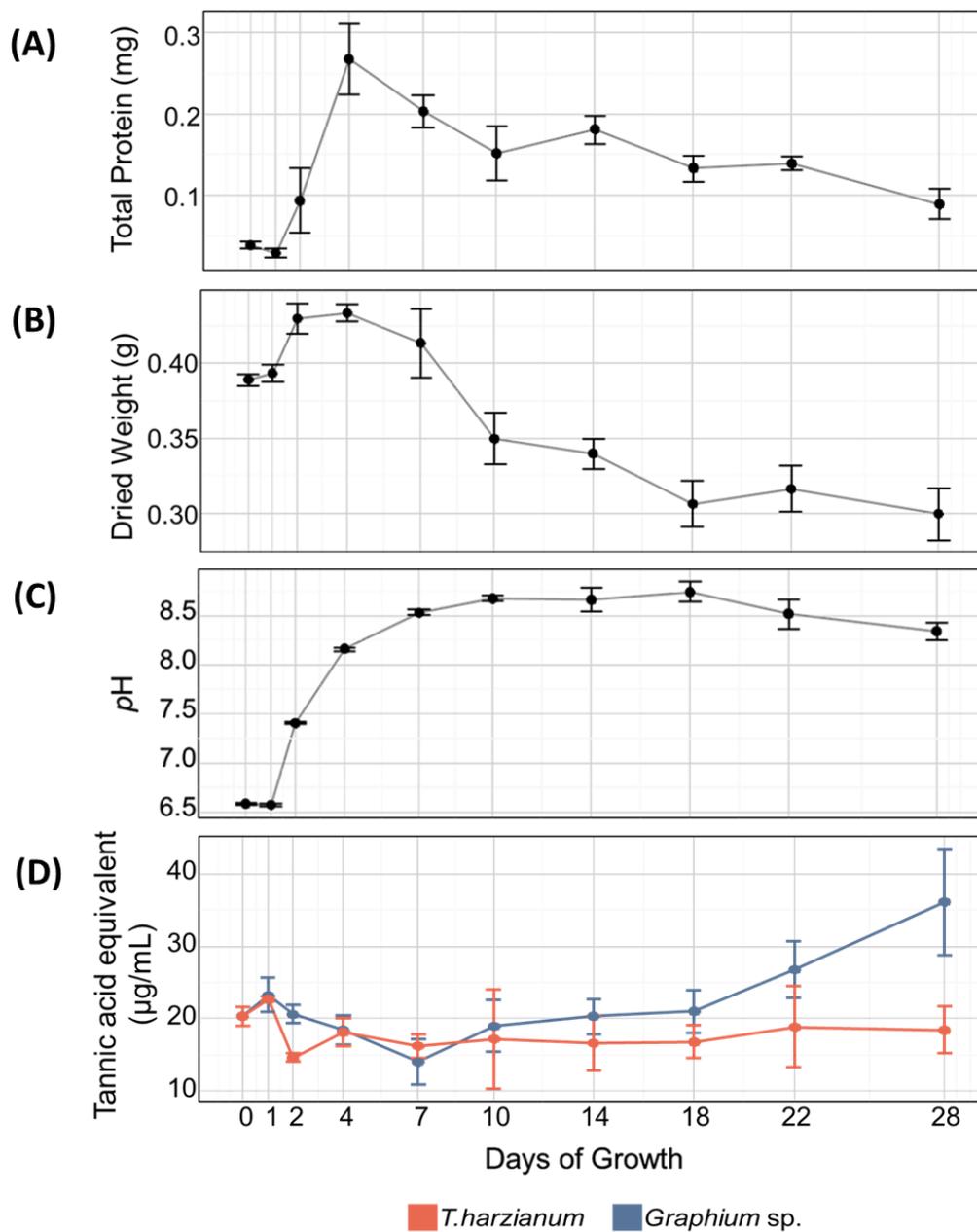


Figure 3-13 – Growth of *Graphium sp.* on wheat straw over a period of one month. The growth of *Graphium sp.* on wheat straw estimated by the (A) total protein present in the culture and (B) the dried weight of the total biomass within the culture. (C) The pH of the culture was also monitored

throughout the month long timecourse. (D) The total phenolic content of the supernatant, as estimated by the Folin-Ciocalteu method and compared to *Trichoderma harzianum*.

3.3.10 Enzymatic activity

Reducing sugar assays clearly show that the peak of xylanase activity in the *Graphium* sp. culture occurs at day four, and that by day seven the activity had decreased by two-thirds. After this sharp decline xylanase activity was stable for the next seven days followed another significant reduction, which occurred between days 14 – 18.

CMC and xylan agar assays were also used to assess cellulase and xylanase activity. The agar plate assays on CMC and beechwood xylan largely agreed with reducing sugar assays, although enzyme activity was detected earlier, with xylanase activity detected 24 hours after the initial inoculation (Figure 3-14)

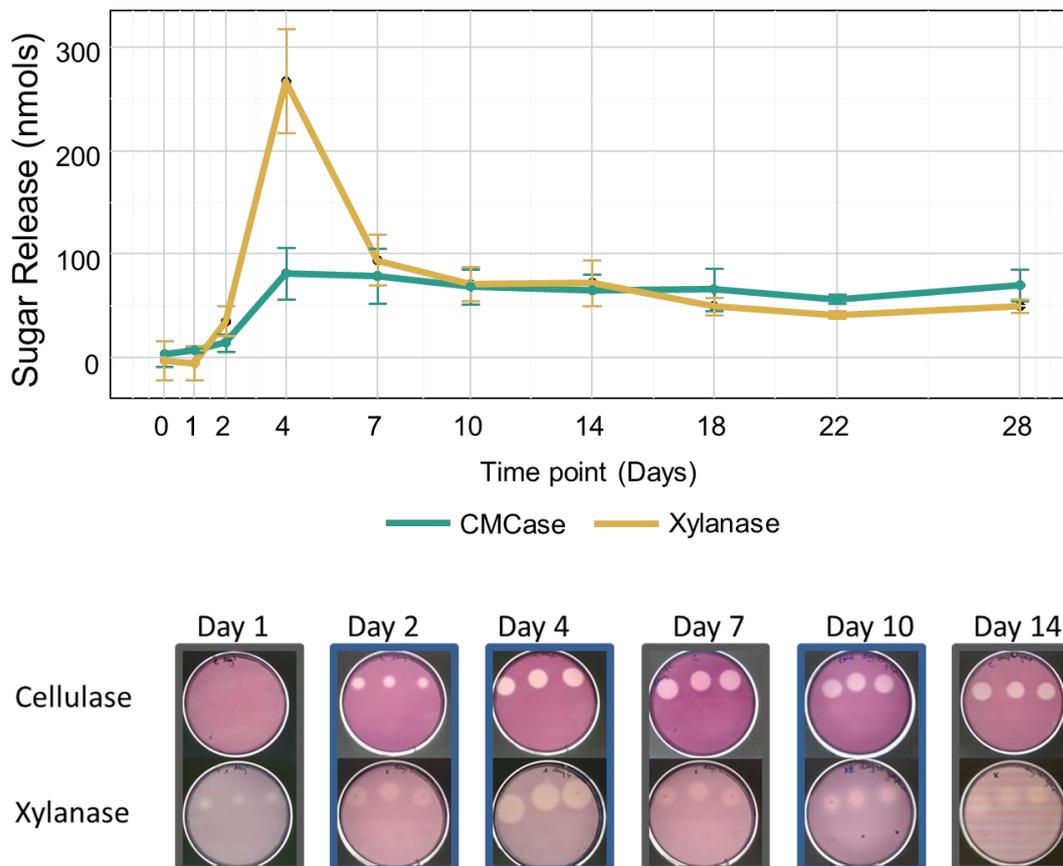


Figure 3-14 – Cellulase and xylanase activity in *Graphium* sp. wheat straw cultures. (A) Reducing sugar release after one-hour incubation on 2 % CMC or xylan. (B) Agar plate assays, performed in triplicate on the same substrates.

3.4 Discussion

3.4.1 The composting community

In total, 32 isolates that showed activity against lignocellulose were identified from the wheat straw enriched compost culture. Of these, 15 were bacterial and were only identified within the first two weeks of sampling. The appearance of bacteria capable of producing secreted oxidases and laccases in the second week of the time course, was likely due to the easily assessable polysaccharides having been degraded by this stage, creating a selective pressure for organisms able to mobilise lignin to access the remaining polysaccharides. Alternatively, these activities could be advantageous due to their detoxification abilities, allowing the bacteria to cope with the toxic phenolic by-products of wheat straw degradation (Bilal et al., 2016, Deshmukh et al., 2016).

It was also noteworthy that bacteria with lignocellulolytic ability were not observed on plates after two weeks of incubation, despite bacteria being isolated from the each timepoint, and the community, as assessed by targeted amplicon sequencing, remaining diverse. This may be a result of the bacteria present becoming secondary consumers, involved in the recycling of the microbial biomass. In addition it may represent the fungal community becoming responsible for the bulk of lignocellulose degradation occurring within the shake flask composting culture. Fungi are well recognised as being adept enzyme secretors, and in a liquid environment with constant agitation the fungal members of the community may be secreting enough lignocellulolytic enzymes to the cultural supernatant to sustain the bacteria present, removing the selective pressure within the bacterial community to be able to degrade lignocellulose. This hypothesis is supported by evidence that within co-cultures of bacteria and fungi, when incubated together on lignocellulose, a reliance on the fungal-produced cellulolytic enzymes, with twice as many present in the secretome compared to the bacterial enzymes is observable (Taha et al., 2015). However this is expected to be species specific and highly variable. It should also however be noted that these experiments have limitations. Many microbes from complex environments are not culturable (Rappe and Giovannoni, 2003), and the agar plate assays used are not able to detect the full range of lignocellulose degrading activities (Pointing, 1999). Therefore it is possible that bacterial isolates may have activity against lignocellulose that has, as yet, not been detected.

The identified lignocellulolytic culturable bacterial isolates were not found to be prominent members of the prokaryotic community according to targeted amplicon sequencing. In fact,

the species that represented the highest relative read abundances were those that belonged to the *Luteolibacter* and *Trupera* genera, neither of which have been reported to be capable of lignocellulolytic activities. These species belong to the Verrucomicrobia and Dienococcus-Thermus phyla, which have been seen in other compost enrichments to dominate the community (Knerr and Tripepi, 2014, Partanen et al., 2010), although species composition is variable across composting environments, and due to the immense diversity contained within these prokaryote phyla, generalisation about the role that they play within the community are often not satisfactory.

Interestingly, the number of protist reads identified in the examination of the eukaryotic community outnumbered the identified fungal sequences by a considerable amount; however, it should be noted that an absolute quantification of the number of individual organisms is not clear, as copy numbers of rRNA genes can vary between species (Weber and Pawlowski, 2013, Gong et al., 2013). Compared to the bacterial and fungal members of lignocellulolytic systems, far less is known about protists. Flagellates have, however, been found to be active in lignocellulose degradation in the termite gut alongside prokaryotes and a symbiotic role has been proposed (Ohkuma, 2008), and a cellulase from this environment has been cloned and characterised (Inoue et al., 2005), suggesting that they may be active members of the lignocellulolytic degrading community.

In contrast to the bacterial isolates, which were only seen in the first few weeks of the culture, the fungal isolate *Graphium* sp., was repeatedly identified across the last six weeks of the time course and targeted amplicon sequencing showed it accounted for a high proportion of the fungal amplicons in weeks five and eight. This isolate showed activity in all of the agar plate assays performed and belongs to a genus that has been identified as containing both cellulase and xylanase activities.

Graphium sp. was first described as having endocellulase activity after being isolated from emulsion paint that contained cellulose as a thickening agent (Tothill et al., 1988). It was later also identified as having xylanase activity after being discovered as part of a leaf litter community (Savitha et al., 2009). Species of this genus have also been characterised as having the ability to degrade xenobiotics and organic pollutants and bioremediate heavy metal containing soils (Vargas-Garcia et al., 2012, Skinner et al., 2008, Santos and Linardi, 2004). This ability to withstand environments that have toxic effects on other organisms, may in part explain why *Graphium* sp. becomes such a prominent member of the fungal community, rising

from being undetectable in the first three weeks of isolation to representing more than 75 % of the relative read abundances in the eighth week.

3.4.2 *Graphium* sp. adaptation to growth on lignocellulose

Graphium sp. appeared to be an organism well adapted to growth on wheat straw. It was seen to utilise wheat straw, avicel and xylan as sole carbon sources, and grew in cultures containing high percentages of kraft lignin, a substrate that was found to inhibit the growth of the well characterised lignocellulolytic fungus *T. harzianum*.

Graphium sp. also appeared to have a better capacity to grow on wheat straw than *T. harzianum*, with more enzymatic activity being detected in the culture supernatant from the wheat straw culture, and a greater amount of the initial biomass either being solubilised or lost through respiration. However, when beechwood xylan is the primary carbon source the opposite appeared to be true, and *T. harzianum* had the highest enzymatic abilities and produced the greatest amount of biomass after seven days of growth.

Graphium sp. ability to display better growth on wheat straw compared to *T. harzianum* may be due to its capacity to deal with inhibitory substances produced from the degradation of lignocellulose. When grown in glucose media, *Graphium* sp. cultures supplemented with kraft lignin produced more dried biomass than *T. harzianum*. This enhancement could be because *Graphium* sp. requires the presence of lignin to be able to deconstruct lignocellulose, and it has been suggested that small molecular weight compounds produce during the degradation of lignin can act as electron donors to complete the catalytic cycle of the LPMOs (Westereng et al., 2015). Though it should be noted that Kraft lignin is an industrial by-product, and so is likely to contain significant contaminants.

Another reason for the enhanced growth of *Graphium* sp. on wheat straw could be due to a better ability to deconstruct crystalline cellulose. It appeared that although the growth of *Graphium* sp. on Avicel did not yield any significant results from the reducing sugar assays performed; however agar plate assays showed that *Graphium* was able to produce clearance zones that were not seen with *T. harzianum* culture supernatant samples. The appearance of activity on agar plates but not in reducing sugar assays could be a result of the longer incubation or an increased sensitivity. It could, however, be a result of the production of non-reducing ends from the polysaccharides due to the action of LPMOs, or the oxidation of free sugars and short polysaccharides into aldonic acids through glucose oxidases and cellobiose dehydrogenases.

3.4.3 Determining critical stages of *Graphium* sp. growth in the degradation of wheat straw

As *Graphium* sp. appeared to be a capable degrader of lignocellulose, a longer time course over a month was performed to identify key stages from which novel lignocellulolytic enzymes might be discovered. Growth on wheat straw was chosen as a focus for protein discovery, as model substrates often cannot fully replicate the native lignocellulose structure in terms of its degree of polymerisation, crystallinity and fibre shape. Complex lignocellulose material can also induce different enzymes than the model substrates, and are more representative of the environments that industrial enzymatic cocktails will be used.

Enzyme activity on lignocellulose is often described as sequential, whereby the easily accessible sugars in hemicellulose are attacked first, followed by the underlying sugars in cellulose. During growth of *Graphium* sp., cellulolytic activity appeared within the first few days but lagged behind xylanase activity, which, appeared as early as 24 hours after inoculation on agar plate assays. Xylanase activity peaked at day four in both assays and subsequently decreased thereafter, whereas cellulose activity peaked at day seven and stayed constant from this point onwards. The fungal growth, like the xylanase activity, appeared to peak on the fourth day with the highest levels of protein appearing at that time point. The pH of the culture also rose rapidly in the first few days of growth, from 6.8 in the optimised media to over 8.5 after ten days of growth. After the tenth day of growth, the total phenolic content of the culture increased. However, this assay cannot distinguish between phenolic content that is produced during the breakdown of lignin or phenolic pigments such as melanin, a pigment fungi are known to secrete to protect against UV radiation and in some cases oxidative stress (Eisenman and Casadevall, 2012).

4 Transcriptomic analysis of *Graphium* sp. During growth on wheat straw

4.1 Introduction

DNA sequencing has become an increasingly attractive technique for the identification of novel genes (Clark et al., 2016). The popularity of these sequencing technologies has been largely driven by the rapid advancements that have occurred in the last decade, as the practice has become both more cost-effective and the depth of sequencing attainable has increased.

The Illumina platform is one of the most widely used methods of next-generation sequencing and employs the sequencing by synthesis approach, whereby fluorescently marked nucleotides are simultaneously washed over immobilised target sequences, and their incorporation imaged (Mardis, 2008). In the last fifteen years, this technique has achieved an exponential growth in the number of gigabases a single machine can output, with current devices capable of producing millions of reads per run.

To cope with the huge numbers of reads generated in these experiments, advancements have also taken place in the development of tools able to assemble them using overlapping regions into consensus sequences (contigs). Aided by the increased depth of the sequencing achievable, *de-novo* assembly methods are now available to allow the transcriptomic study of non-model organisms, as, if sequencing of a sufficient depth is acquired a reference genome is no longer a prerequisite (Grabherr et al., 2011). Curated databases of annotated enzyme families as well as Hidden Markov Models (HMM), developed to recognise characteristic regions of sequence in protein families, are also publically available online, alongside tools designed to aid the identification of open reading frames, signal peptides and transmembrane domains (Finn et al., 2011, Finn et al., 2006, Sonnhammer et al., 1998, Petersen et al., 2011, Emanuelsson et al., 2007).

However, despite these tools being of use in the identification of known protein functions, if the information contained within the models and databases is incomplete or incorrect they will be unable to assign the correct function to proteins of interest. The fungal LPMO family epitomises this, as after initially been miss-assigned as belonging to a glycoside hydrolase family it was later shown to employ an oxidative mechanism of attack, and consequently reassigned to the newly designated AA class (Hemsworth et al., 2015).

As it is hypothesised that there are still many enzyme families involved in lignocellulose degradation whose roles have not yet been described, transcriptomic studies in which differential expression of genes can be investigated between conditions, with and without lignocellulose, may be able to identify genes of interest to the biofuel industry without prior knowledge of sequence.

However, this type of investigation on insoluble complex substrates, such as wheat straw, presents multiple challenges. The extraction of mRNA of a suitable quality for sequencing is hampered by the large amount of polyphenolics located within these environments, and the harsh treatments required for cell lysis and quantity of endogenous fungal ribonucleases present can often result in partially degraded mRNA (Patyshakuliyeva et al., 2014, Wang et al., 2012, Yang et al., 2007, Leite et al., 2012, Latge, 2007). As a result, relatively few studies have examined the transcriptomics of a single fungus during the degradation of lignocellulose on a prolonged timescale.

The results of a transcriptomic experiment performed on *Graphium* sp. during a ten-day time course of growth on wheat straw are presented in this chapter. This investigation was conducted on *Graphium* sp. as characterisations reported in Chapter 3, appeared to have the potential for containing novel lignocellulolytic enzymes.

Transcriptomics was deemed the most appropriate method to identify the enzymes responsible for the lignocellulolytic activities reported in the previous chapter since as the genome of this organism is not available, and no transcriptomic study has previously been performed any proteomic study would be hampered by the lack of genetic information.

4.2 Materials and methods

4.2.1 Experimental Setup

The fungal biomass was collected from the cultures described in Chapter 4 after two, four and ten days of growth on wheat straw using the method outlined in Section 3.2.7. Cultures grown on glucose were also harvested after four days of growth to represent the peak of growth on this substrate (see Section 3.3.9 for detailed growth characterisations). Care was taken to process the biomass as fast as possible, and all procedures were performed on ice, unless otherwise stated, with DEPC treated solutions and sterile equipment to prevent RNA degradation.

4.2.2 RNA extraction

To control for varying amounts of cell growth, aliquots of either 0.5 g, 0.3 g and 0.1 g of biomass from the wheat straw cultures were weighed into 2 ml screw cap tubes that contained 3x 3 mm tungsten carbide beads and 1 mL Trizol (Life Technologies). The cells were then disrupted in a TissueLyser II (Qiagen) for either 2x2 minutes or 2x5 minutes at 28/seconds, dependent on the stage of growth. Total RNA was then extracted with the standard Trizol method as per manufactures instructions and extracted RNA was resuspended in 50 µL of nuclease free water. The quality of RNA was assessed by visualisation on agarose gels. To obtain enough RNA for processing six technical replicates were performed for each biological replicate. These were stored at -80 °C after being flash frozen in liquid nitrogen before further processing could occur.

4.2.3 Clean up and enrichment of mRNA

The RNA samples were treated for DNA contamination with RTS DNase kits (Mobio) using standard methods described by the manufacturers. The samples were then cleaned with ZymoResearch RNA Clean & ConcentratorTM-5 kits, using the manufacturer's protocol to separate small and large RNA fragments into different fractions. RNA fragments greater than 200 nt were elution into 50 µl of RNase-free water before RNA concentration and quality was evaluated with the 2200 TapeStation (Aligent). Once total RNA of a suitable quantity and quality was obtained, samples could be enriched for messenger RNA (mRNA). Preliminary experiments to test mRNA enrichment were performed using either Illumina's Ribo-Zero Gold rRNA Removal (Epidemiology) Kits, ThermoFisher's Dynabeads[®] or Qiagen's Oligotex mRNA Mini Kit using instructions provided by the manufacturer. After processing, the samples were

again evaluated using the TapeStation to access rRNA depletion. Samples sent for sequencing were processed with the Ribo-Zero kits.

4.2.4 RNA sequencing

The Genome Analysis Centre (TGAC), Norwich, U.K, performed the RNA sequencing on an Illumina HiSeq platform. As per the requirements of the sequencing service, 100 ng of enriched mRNA was provided for each sample.

From the proved mRNA, cDNA libraries were constructed using the adapted TruSeq RNA v2 protocol (Illumina 15026495 Rev.B). Libraries were then normalised using elution buffer (Qiagen) and pooled in equimolar amounts into one final 12 nM pool. These were then diluted to a final concentration of 10 pM, spiked with 1 % PhiX and loaded onto the Illumina cBotTemplate, for hybridisation and first extension, using the TruSeq Rapid PE Cluster Kit v1 before the flow cell was transferred onto the Illumina HiSeq2500. Here, the remainder of the clustering process was conducted, and the library pool was run in a single lane for 100 cycles of each paired-end read before samples were demultiplexed. One base-pair mismatch per library was allowed, and reads were converted to FASTq.

4.2.5 Assembly of the transcriptome

The raw data was subject to rRNA removal by catching the remaining paired reads after mapping to a modified rRNA_115_tax_silva_v1.0 ribosomal set, using BOWTIE2. The reads were further trimmed to remove adaptor sequences with the ngsShoRT_2.1 method, and libraries were pooled before being assembled by Trinity Software to obtain 37,720 contigs. Then, using this assembly as a reference, the original (unprocessed) individual libraries were mapped and the number of reads counted for each contig. Counts per million (CPM) were converted to reads per kilobase of exon per Million reads mapped (RPKM) to normalise for both the depth of sequencing achieved in each sample and length of the contig.

4.2.6 Filtering

Contigs were filtered to limit the number of false sequences. Contigs that had an RPKM of less than one, and were present in less than two out of the three biological replicates were filtered out of the analysis.

4.2.7 Bioinformatics

Emboss GETORF (<http://www.bioinformatics.nl/cgi-bin/emboss/getorf>) was used to generate putative protein coding sequences in all six reading frames from the transcriptomic libraries by translating regions over 300 bp long between potential start and stop codons. These putative open reading frames (ORFs) were searched against the NCBI non-redundant protein database and KOG database using BLASTp, the Pfam and dbCAN databases using HMMER3 (Finn et al., 2006, Tatusov et al., 2003). Annotations were subsequently mapped back to the contig from which the ORF originated from. Local BLAST searches using unique were performed using BLAST+ 2.3.0 (Camacho et al., 2009, Altschul et al., 1990). Signal peptides were predicted from ORFs using SignalP 4.0 (Emanuelsson et al., 2007, Petersen et al., 2011). Glycosylphosphatidylinositol (GPI) prediction (Pierleoni et al., 2008) was also performed using PredGPI (Pierleoni et al., 2008).

The R package EdgeR was used to determine significant differential expression between genes (Robinson et al., 2010). The analysis was performed on the CPM of filtered contigs, and these were normalised using the `in` package command. Differential gene expressions were then predicted between each set of conditions using an exact negative binomial test. Another R package, Mfuzz was used to group contigs by expression pattern using soft clustering (Futschik and Carlisle, 2005).

4.3 Results

4.3.1 RNA extraction from *Graphium* sp.

RNA was successfully extracted from the second, fourth and tenth day of *Graphium* sp.'s growth on wheat straw, and its fourth day of growth on glucose. Several rounds of method optimisation were required during the preliminary experiments since it was apparent that the growth stage of the fungus had a significant effect on the success of the extraction. After two days of growth on wheat straw, RNA of the highest quality was harvested when 0.5 gram of biomass was treated, at day four 0.1 gram of biomass provided the most success and at day ten 0.5 gram. The length of the bead beating that gave the best quality and quantity of RNA also differed depending on time point as four minutes was sufficient for cell lysis at both day two and day four of growth on wheat straw, but when liberating the RNA in the day ten, this was increased to ten minutes to release the greatest amount of RNA. The conditions of RNA extraction for each sample are summarised in Table 4-1.

Table 4-1 – RNA extraction conditions. The strategy applied for RNA extraction varied depending on the composition of the media and the length of incubation.

Carbon	Days of Growth	Weight of Biomass (g)	Bead Beating(min)
Glucose	4	0.1	2
WS	2	0.5	2
WS	4	0.3	2
WS	10	0.5	5

4.3.2 mRNA enrichment from total RNA

Two approaches to mRNA enrichment were trialled during this investigation; polyA tail selection using both Oligotex mRNA kits and Dynabead magnetic bead kits, and ribosomal RNA depletion using Ribozero kits (Section 4.2.3). Only small yields of mRNA were obtained using the polyA selection method, and the mRNA purified appeared to be of low quality. Higher quantities of mRNA were retained when the ribozero methodology was applied and therefore it was these samples that were carried forward for sequencing (

Figure 4-1).

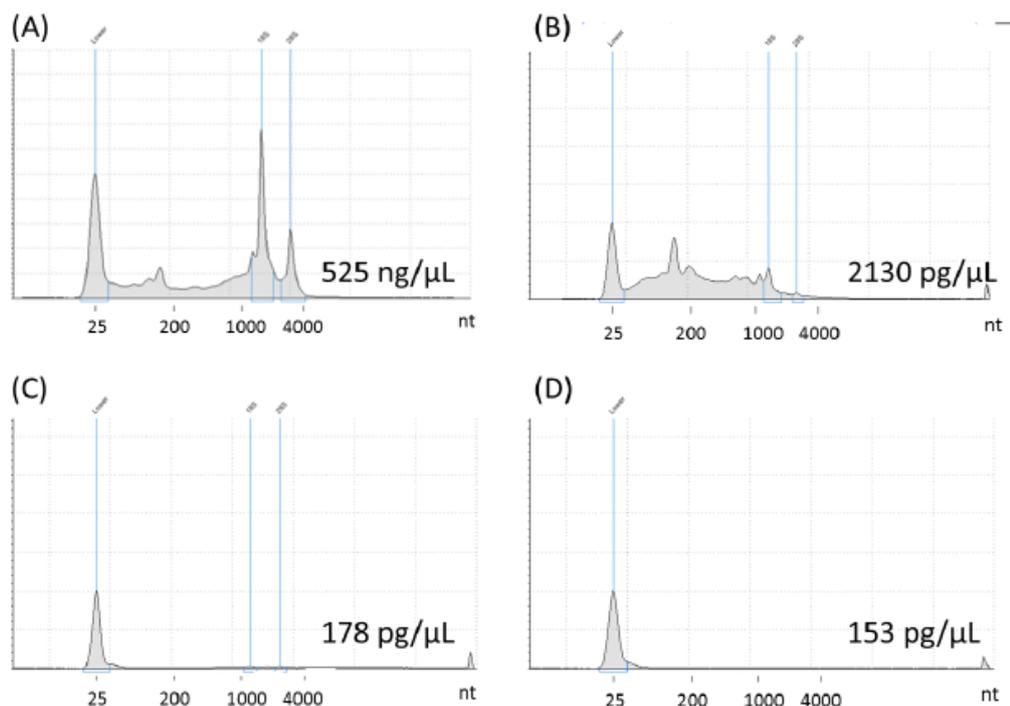


Figure 4-1 - mRNA enrichment from total RNA harvested from wheat straw grown *Graphium* sp. cultures. (A) Total RNA was extracted from ten-day-old *Graphium* sp. cultures and profile assessed using the Agilent 2200 TapeStation. Methods of mRNA enrichment were then trialed including (B) Illumina's Ribozero kits, (C) Qiagen's Oligotex mRNA kits and (D) Invitrogen's Dynabead magnetic bead kits.

4.3.3 RNA Sequencing of *Graphium* sp.

After sequencing on the Illumina platform a total of 31,468,338 usable reads, were generated, despite significant contamination of rRNA (Table 4-2). These, after assembly using the Trinity pipeline described in Section 4.2.5, resulted in 37,720 contigs with a length of over 300 bp.

Table 4-2 - Number of reads generated by RNA sequencing. RNA sequencing was performed on the Illumina HiSeq platform. The resultant reads were then filtered for rRNA contamination and trimmed to remove adapter sequences.

Substrate	Day	Reads	rRNA contamination	Reads after filtering and trimming
Glucose	4	9,562,489	78%	1,949,732
Glucose	4	12,254,910	91%	847,015
Glucose	4	16,496,332	92%	1,011,190
Wheat Straw	2	13,499,977	76%	2,963,285
Wheat Straw	2	13,099,272	71%	3,546,765
Wheat Straw	2	14,211,641	73%	3,542,533
Wheat Straw	4	10,092,690	51%	4,898,944
Wheat Straw	4	15,170,172	74%	3,712,830
Wheat Straw	4	14,069,464	81%	2,408,484
Wheat Straw	10	12,038,893	74%	2,917,785
Wheat Straw	10	16,505,101	83%	2,554,019
Wheat Straw	10	13,971,606	90%	1,115,756

The presence and abundance of each contig varied between samples. Hierarchical clustering of contigs and their RPKM showed a greater degree of similarity between biological replicates than experimental conditions (Figure 4-2), signifying that differences between samples were condition specific, and not a result of random biological or technical variation.

Table 4-3 – The number of filtered contigs present within each sample. Contigs were counted as present in each timepoint if they exceeded 1 RPKM and were present in at least two of the three biological replicates.

Carbon	Days of Growth	No. of Contigs
Glucose	4	11,406
WS	2	13,026
WS	4	13,652
WS	10	14,490

The fewest number of contigs were obtained from *Graphium* sp. after growth on glucose, while the highest number of contigs were found in the wheat straw grown cultures (Table 4-3). More contigs were also shared between the time points of samples grown on wheat

straw, than between the glucose and wheat straw samples, suggesting that additional genes were transcribed as a response to the lignocellulose (Figure 4-2B).

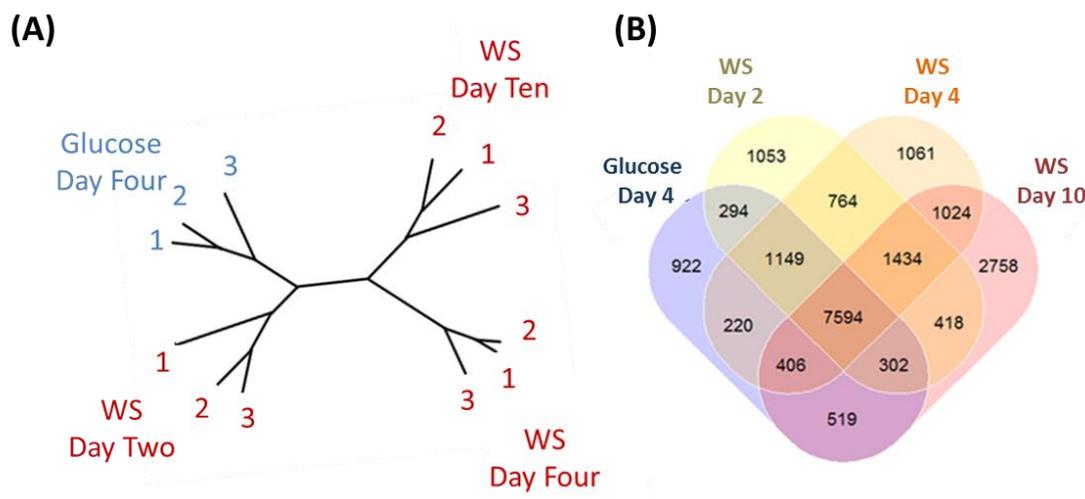


Figure 4-2 – Hierarchical clustering of RNA sequencing results and the distribution of contigs between samples. (A) Euclidean clustering performed on the log transformed RPKM values from each biological replicate. (B) Venn diagram of contigs present in each condition at an RPKM of over one in two biological replicates.

4.3.4 Carbohydrate Active Enzymes

Potential ORFs in the transcriptome were searched with dbCAN's hidden Markov models, using methods described in Section 1.2.7. In total, 638 ORFs that contained unique putative carbohydrate active enzymes (CAZymes) were identified, and of these, 512 contained domains recognised as belonging to classes that have some involvement in the degradation of lignocellulose, either having carbohydrate esterase, glycoside hydrolase, polysaccharide lyase or auxiliary activity.

These classes are however large and contain enzymes that modify a diverse range of carbohydrates. Nevertheless, when CAZyme classes are broken down into families, further detail can be obtained since as they are defined by sequence similarities they often contain enzymes with similar functions. Table 4-4 details putative CAZymes families detected in the transcriptome.

Table 4-4 - Total number of CAZymes detected in *Graphium* sp. transcriptome. Counts were calculated from the number of contigs identified as containing each domain and, therefore, include isoforms. Bracketed numbers signify the number of clusters from which the contigs originated.

	TOTAL	BREAKDOWN WITHIN FAMILIES
AA	114 (90)	AA1 - 3 (3), AA2 - 3 (3), AA3 - 34 (26), AA4 - 3 (2), AA5 - 2 (2), AA6 - 1 (1), AA7 - 25 (21), AA8 - 4 (2), AA9 - 39 (30)
CE	127 (90)	CE1 - 36 (27), CE2 - 3 (2), CE3 - 12 (9), CE4 - 11 (7), CE5 - 10 (9), CE7 - 1 (1), CE10 - 43 (33), CE12 - 3 (2), CE14 - 2 (1), CE15 - 4 (4), CE16 - 2 (2)
GH	262 (185)	GH1 - 2 (2) , GH2 - 4 (4) , GH3 - 17 (11) , GH5 - 14 (12) , GH6 - 4 (3) , GH7 - 10 (5), GH10 - 12 (7), GH11 - 5 (5), GH12 - 2 (2), GH13 - 7 (5), GH132 - 3 (2), GH15 - 2 (2), GH16 - 16 (12), GH17 - 4 (4), GH18 - 15 (11), GH20 - 3 (1), GH26 - 1 (1), GH27 - 2 (2), GH28 - 1 (1), GH30 - 3 (3), GH31 - 7 (7), GH36 - 3 (1), GH37 - 3 (1), GH38 - 1 (1), GH43 - 20 (15), GH45 - 10 (6), GH47 - 9 (5), GH51 - 2 (2), GH53 - 1 (1), GH55 - 3 (2), GH62 - 1 (1), GH63 - 1 (1), GH64 - 1 (1), GH67 - 1 (1), GH72 - 5 (4), GH74 - 2 (2), GH76 - 8 (5), GH78 - 7 (4), GH79 - 1 (1), GH81 - 4 (1), GH88 - 1 (1), GH92 - 4 (2), GH93 - 1 (1), GH94 - 1 (1), GH105 - 1 (1), GH109 - 9 (7) , GH114 - 3 (3), GH115 - 7 (3) , GH125 - 10 (3), GH127 - 1 (1), GH128 - 3 (2) , GH131 - 3 (3) GH132 - 4 (4)
GT	126 (76)	GT1 - 9 (5), GT2 - 19 (11), GT3 - 5 (4), GT4 - 6 (4), GT8 - 10 (3), GT15 - 7 (3), GT17 - 1 (1), GT21 - 2 (2), GT22 - 9 (4), GT24 - 2 (1), GT25 - 3 (3), GT31 - 4 (3), GT32 - 4 (4), GT33 - 2 (1), GT34 - 4 (2), GT35 - 3 (1), GT39 - 4 (3), GT41 - 3 (1), GT45 - 1 (1), GT48 - 2 (1), GT57 - 3 (2), GT58 - 2 (1), GT59 - 2 (1), GT62 - 5 (3), GT66 - 2 (1), GT69 - 4 (3), GT71 - 2 (2), GT76 - 3 (1), GT78 - 1 (1), GT90 - 4 (3)
PL	9 (8)	PL1 - 4 (4), PL3 - 1 (1), PL4 - 4 (3),

The putative protein sequences that were identified as containing CAZyme domains were BLAST searched against the total CAZyme database to determine the degree of similarity between *Graphium* sp. putative proteins and those of other organisms. These matches ranged from being highly significant, having just small percentage identities, for example a GH45 within the transcriptome has just a 37.5 % identity to another known GH45 within the NCBI non-redundant protein database, suggesting that novelty may be found even when known enzymes are examined *Graphium* sp..

4.3.5 Transcriptomic response to wheat Straw

To aid the identification of targets specifically related to lignocellulose degradation, and to examine *Graphium* sp. response to wheat straw, the relative abundance of the CAZyme families were investigated across the time course, and compared to the glucose-grown control. These are displayed in Figure 4-3.

The time point that contained the highest relative proportion of CAZymes was the fourth day of growth on wheat straw (Table 4-5). This time point contained 556 contigs with predicted CAZyme domains and represented an average of 11,659 RPKM across the three biological replicates, 1.5 % of the total RPKM from that condition. Of these, 499 contigs were significantly upregulated in comparison to the same time point of the glucose-grown cultures as determined by EdgeR (Section 4.2.7).

Table 4-5 - Number of CAZyme domains in each condition. CAZymes were identified with dbCAN's hidden markov models and counted as present if their expression exceeded 1 RPKM in at least two of the three biological replicates.

Substrate	Day	AA	CE	GH	GT	PL	Grand Total
Glucose	4	62	84	170	109	5	430
WS	2	77	90	189	115	7	478
WS	4	103	102	233	109	9	556
WS	10	79	91	218	102	4	494

The second day of growth on wheat straw also saw an upregulated level of CAZymes in comparison to the glucose control, with 323 contigs out of 478 (67.6%) significantly higher between the two conditions. Interestingly, however, after ten days the number of present CAZymes had reduced, with the CAZyme containing contigs present representing a total of 2,984 RPKM (0.3 % of the conditions RPKM). Though despite this decrease in abundance, 358 contigs were still expressed at a significant increase when compared to the glucose-grown culture.

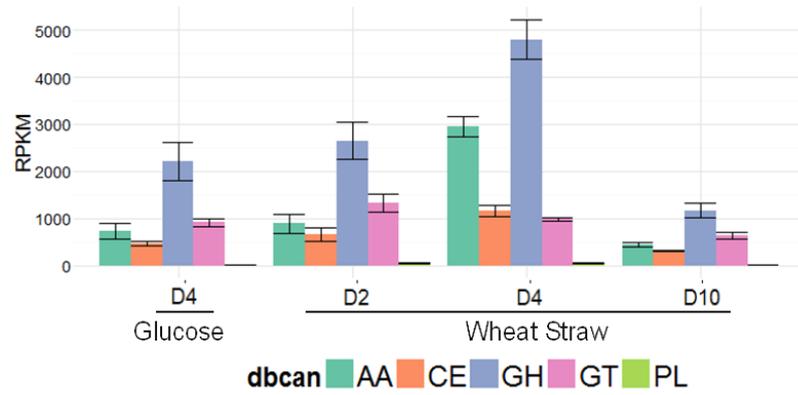


Figure 4-3 – Abundance of carbohydrate active enzyme classes. Classes were assigned to contigs using dbCAN’s hidden Markov models. Bars represent an average of the total of the RPKM of contigs which contained a CAZyme class in each biological replicate, and error bars the standard deviation of these.

When each family within the CAZyme classes were investigated it was clear that there were many patterns of expressions, both regarding the number of ORFs representing each family, and their abundance within the transcriptome. These are displayed in Figure 4-4, and an overview of their presence in each class is discussed in further detail in the following sections.

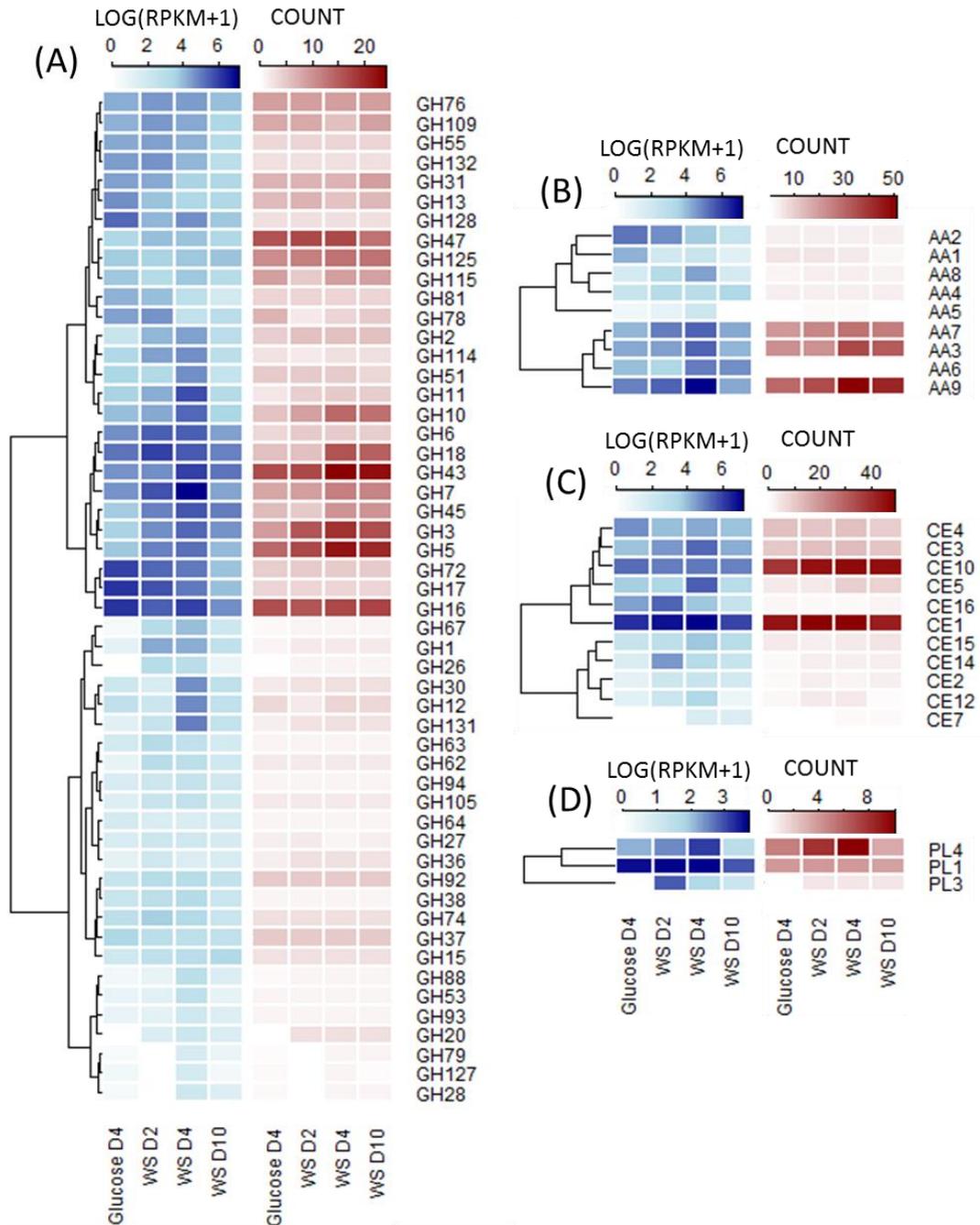


Figure 4-4 – Abundance and count of CAZyme families across the time points. Heatmaps displaying changes of both the number of contigs (blue) and RPKM (red) of identified CAZymes. Heatmaps are divided by class with glycoside hydrolases (GH)(A), auxiliary activities (AA)(B), carbohydrate esterase (CE)(C) and pectin lyases (PL)(D). The clustering between CAZy families was performed using euclidean distances.

4.3.5.1 The Glycoside Hydrolases

The glycoside hydrolase class, as defined by the CAZy database, is the largest of the CAZyme classes and is currently made up of 135 families. Within the *Graphium* sp. transcriptome 52 of these families could be identified from 262 ORFs, with some families such as the GH43s, a family known to contain β -xylosidases, represented by as many as 20 distinct gene sequences.

The dominant expression profile observable from the GHs peaked in either the second or the fourth day of growth on wheat straw, followed by a reduction in day ten. This pattern appeared to be common across all of the known cellulose and hemicellulose degrading families within GH5, GH6, GH7 and GH45 families, associated with cellulose deconstruction in *Ascomycetes* (Glass et al., 2013), clustering alongside known xylanases such as the GH3 and GH43 families (Figure 4-4).

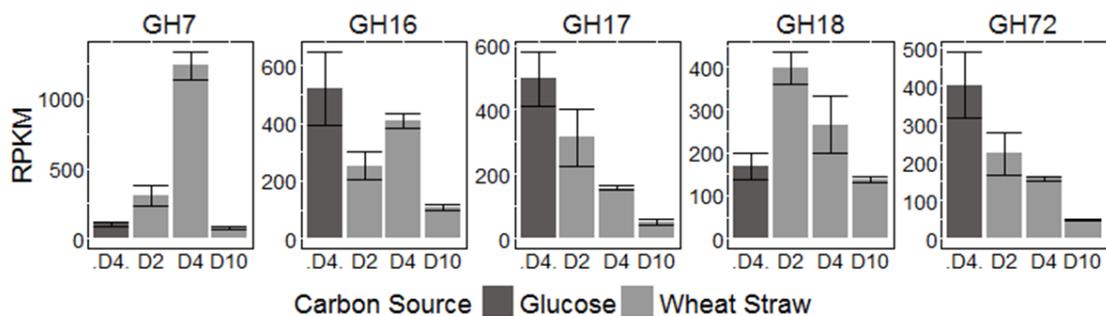


Figure 4-5- The abundance of the top five glycoside hydrolase families. The RPKM of contigs that were annotated using dbCAN's hidden Markov models as containing CAZyme domains were summed for each family and the five most abundant glycoside hydrolase families are shown. Error bars represent the standard deviation of three biological replicates.

The GH7 family, the family that predominantly consists of cellobiohydrolases, was the most abundant GH family in the *Graphium* transcriptome. Figure 4-5 shows the expression profile of the GH7s alongside the four next most abundant CAZyme families, the GH16s, GH17s, GH18s and GH72s.

Much of the abundance of the GH7s, 48 % of their total RPKM, could be attributed to a single contig, c5665_g2_i1, which had a 76 % amino acid sequence identity to a cellobiohydrolase (Sequence ID: KXH25440.1) from *Colletotrichum simmondsii*. After four days the expression of this contig was strongly upregulated when *Graphium* sp. was grown on wheat straw when compared to glucose, with a 600 fold increase between these time points.

The GH16 and GH17s were the next most abundant families of the GH class. These families are not commonly associated with lignocellulose degradation but instead are known to cleave either β -1,4 or β -1,3 glycosidic bonds in various glucans and galactans. They also differ from the putative cellulases discussed previously, as they show no significant up-regulation when grown on wheat straw. Since the putative ORFs that were annotated as being GH17 containing are all predicted to be glycosylphosphatidylinositol (GPI) anchored proteins, as are nine of the 18 proteins that contained the GH16 domain, it is likely that they are proteins involved in the remodelling of the fungal cell wall, rather than the metabolism of external polysaccharides. Examples of this has been seen in recent literature as GPI-anchored, fungal cell wall modifying proteins being described from another filamentous fungus, *Aspergillus fumigatus* (Hartl et al., 2011).

Another family of note, and the eighth most abundant family in the GH class, was the GH45 family, a family that contains endocellulases which show similarity to expansin-like proteins. This family was of interest as it appeared to be expanded within the transcriptome as ten contigs could be identified as containing a putative GH45 domain. These were diverse in their sequences with just a 29.8 % identity between contigs c1682_g2_i1 and c20614_g1_i1, and identities to their closest match in the NCBI non-redundant databases ranging from 37.5 to 92.7 % identity. The expression pattern of each contig also varied, with c5896_g1_i1's relative expression highest on the second day of wheat straw growth c1682_g2_i1 highest on the second and c7000_g2_i1 greatest on the tenth day (Figure 4-6).

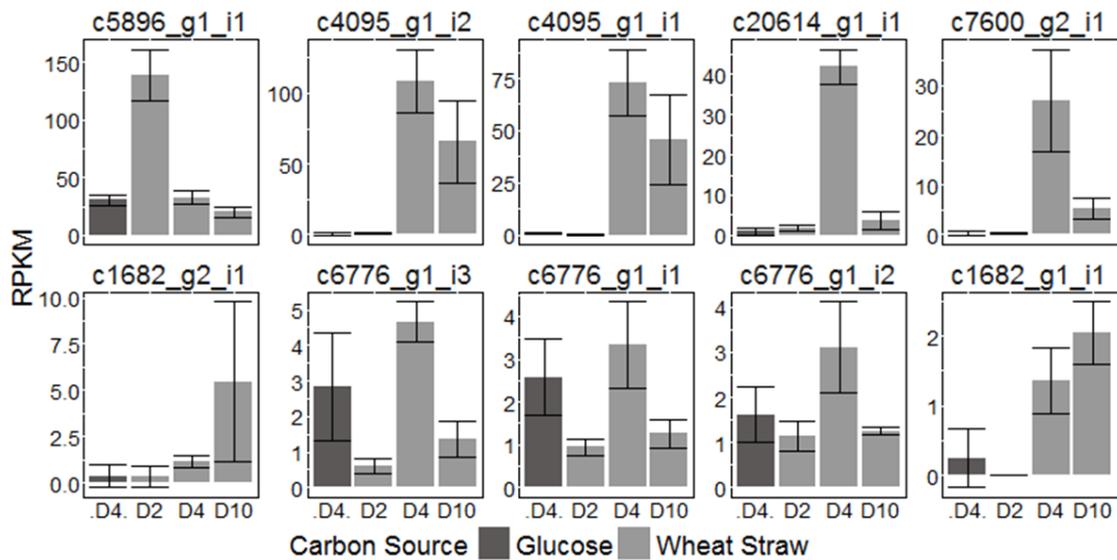


Figure 4-6 - Expression profiles of contigs contain the glycoside hydrolase 45 family. ORFs generated from transcriptomic libraries were annotated with dbCAN's hidden Markov models and mapped back to the contig that they originated. Each GH45 containing contigs RPKM was summed. Error bars represent the standard deviation of the sum of three biological replicates.

4.3.5.2 The Auxiliary Activities

The Auxiliary Activities class are a relatively new division of CAZyme classification, and their expression patterns throughout the investigation are displayed in Figure 4-7. This class, though not directly carbohydrate active, contains redox enzymes that are associated with the degradation of lignocellulose and other polysaccharides. These included LPMOs, contained within the AA9, AA10 and AA13 families, peroxidases (AA2) and laccases (AA1) involved in lignin degradation, and oxidative enzymes (AA3 & AA8) that may play a role in providing LPMOs with electrons.

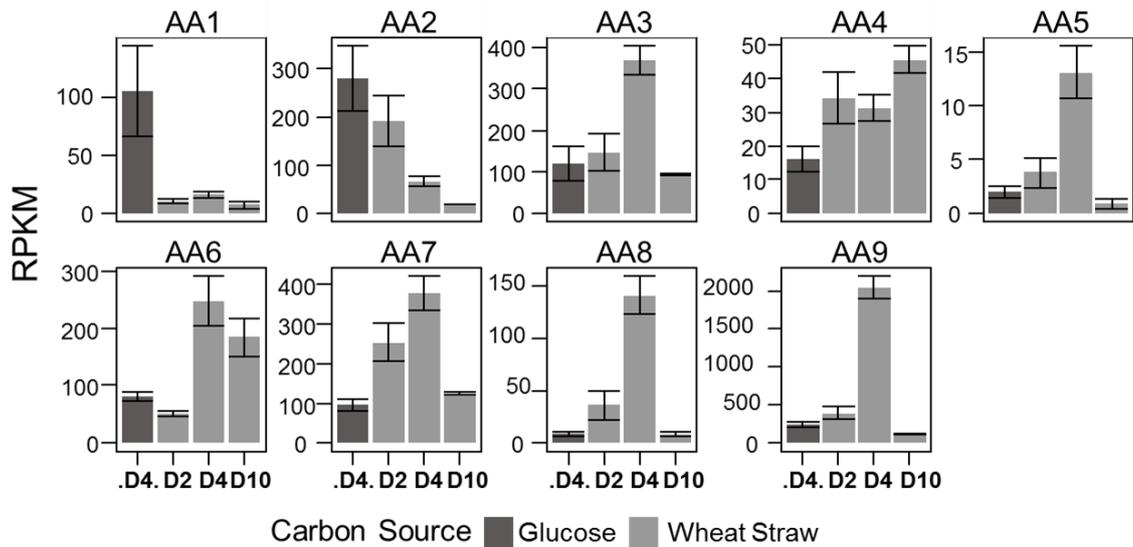


Figure 4-7 Transcriptomic expression profile of the auxiliary activities families. The RPKM of contigs that were annotated using dbCAN's hidden Markov models as containing a CAZyme domains were summed for each family and auxiliary activities families are shown. Error bars represent the standard deviation of the sum of three biological replicates.

The AA9s were the most abundant family, in terms of both the expression and the number of putative proteins within the *Graphium* sp. transcriptome. In total, 29 contigs were recognised as containing AA9 domains, and these were highly upregulated in the presence of lignocellulose.

The two most highly transcribed contigs with AA9 domains were c19570_g1_i1 and c19236_g1_i1 and were respectively the 5th and 12th most upregulated sequences between glucose and wheat straw cultures after four days of growth. c19570_g1_i1, showed the closest homology, with a 61 % identity, to a hypothetical protein (Sequence ID: XP_001227508.1) from *Chaetomium globosum* in the NCBI non-redundant protein database, and a 40 % identity to an annotated LPMO (Sequence ID: K1Y72566.1) from *Cylindrobasidium torrendii*, while c19236_g1_i1, appeared to have a far less divergent sequence with a 79 % identity to an annotated AA9 (Sequence ID: XP_016643782.1) from the fungus *Scedosporium apiospermum*.

Similarly to the cellulase containing families discussed previously, the GH7s, GH6s and GH5s, the expression of the AA9s was, as a whole, downregulated between day four and day ten of wheat straw growth. However, as with the GH45s, within the 29 AA9 containing contigs multiple expression profiles could be observed (Figure 4-8). These included the contig c7324_g1_i1, which was had a higher RPKM in the glucose-grown cultures, and c8328_g3_i2,

which was upregulated after ten days of growth on wheat straw. When a BLASTp search was performed against the structurally characterised proteins of the PDB database c7324_g1_i1 showed significant homology to a known chitin degrading LPMO (4MAH_A), while c8328_g3_i2 showed the closest homology to a LPMO (4B5Q_A) known to act on cellulose (Wu et al., 2013), suggesting that the divergent expression profiles observed within this grouping are indicative of different substrate specificities.

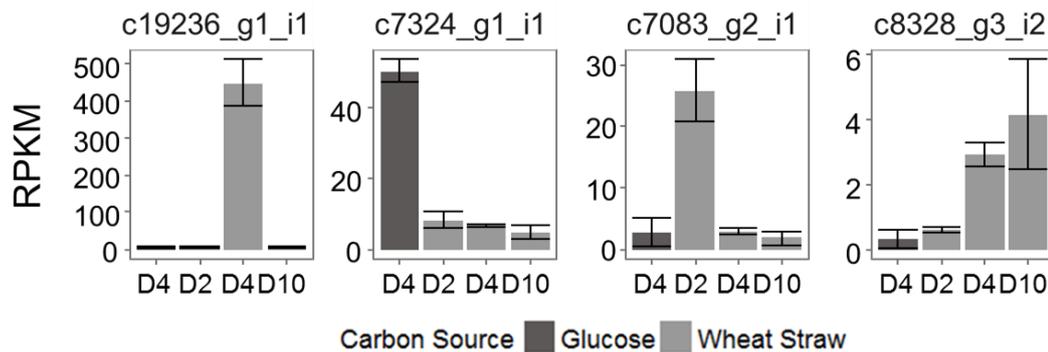


Figure 4-8 Variation in the expression pattern of the auxiliary activity family nine (AA9). Contigs annotated as containing AA9 domains displayed different expression patterns throughout the time course. Error bars represent the standard deviation of three biological replicates.

Like the AA9 family, the AA3 and AA7 families also appeared expanded in the transcriptome. In total 26 contigs contained AA3 domains and 18 included regions identified as AA7s. These families contain oxidative enzymes, including cellobiose dehydrogenases and GMC oxidase, and were significantly upregulated the fourth day of growth on wheat straw.

While these families, thought to be involved the degradation of carbohydrates through monooxygenase attack are expanded, other families within the auxiliary activity class are less prominent. Only three contigs were recognised as containing AA1 domains and the peroxidases of the AA2 family were similarly detected in three putative proteins, all of which lacked signal peptides, and appear to be down-regulated in cultures grown on wheat straw.

The AA6 family, though only represented by one contig, c6178_g2_i1, appeared to be of interest as it was highly transcribed and had the highest total RPKM within AA class. This contig was not recognised as containing a signal peptide, consistent with reports that the AA6 family is intracellular. As this domain is associated with 1,4-Benzoquinone reductases activity, it has been suggested that it is related to the metabolism of low molecular weight lignin breakdown products (Mori et al., 2016, Shimizu et al., 2005), and unlike many of the other CAZyme families, it was not downregulated between the fourth and tenth day of growth on wheat

straw, where the majority of soluble phenolics were found during the characterisation these cultures (Section 3.3.9).

4.3.5.3 The Carbohydrate Esterases

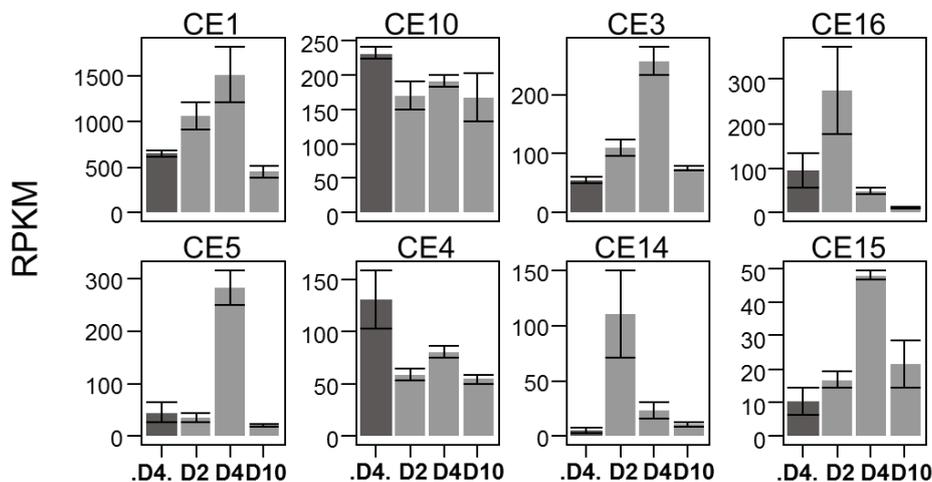


Figure 4-9 Transcriptomic expression profile of the top eight carbohydrate esterase families. The RPKM of contigs that were annotated using dbCAN's hidden Markov models as containing a CAZyme domain were summed for each family and the eight most abundant carbohydrate esterases are shown. Error bars represent the standard deviation of the sum of three biological replicates.

Carbohydrate esterases catalyse O- and N-deacetylation of saccharides and involved in the dissociating lignin from hemicellulose, and removing acid decorations from polysaccharide chains, increasing the accessibility of other enzymes to the sugar backbones.

The CE1 family had the highest total RPKM of this class and contains proteins of multiple enzyme activities, including acetyl xylan esterases and feruloyl esterases. Within the *Graphium* sp. transcriptome, 51 contigs contain putative CE1 families, five of which were present at a mean RPKM over 100 in at least one time point (Figure 4-10). Of these five most abundant sequences only c16085_g1_i1 and c16107_g1_i1 contained predicted signal peptides. These two contigs mirror the expression pattern that is characteristic of the previously described cellulases and the highest homologies to the lignocellulose modifying feruloyl esterases. The other three contigs, which did not contain signal peptides or show the characteristic expression profile, instead showed homologies to enzymes involved in intracellular processes such as serine hydrolases, or phosphatases.

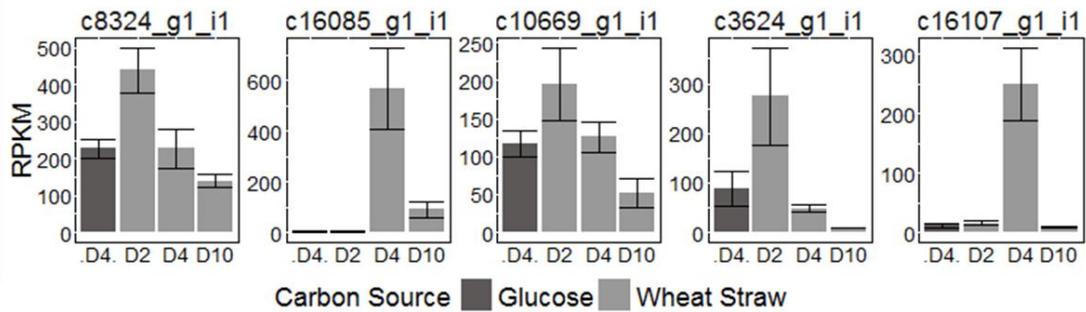


Figure 4-10 – The abundance of the top five contigs that contain carbohydrate esterase domains 1. The top five CE1 containing contigs in order of RPKM. Error bars represent the standard deviation of three biological replicates.

4.3.5.4 Polysaccharide Lyases

Polysaccharide lyases are predominately associated with pectin degradation, and as pectin is a minor component of wheat straw, it is not expected to be of critical importance in its degradation. Indeed, within the transcriptome, they have the lowest total RPKM of all the CAZyme classes, and only 15 contigs were identified with PL domains, with only four present over 10 RPKM in any one sample.

4.3.6 Carbohydrate binding modules

Carbohydrate binding modules (CBMs) were also identified from the transcriptome using dbcan's hidden Markov models. Many of these CBMs were situated on contigs that contained CAZyme (Table 4-6), and there was a clear correlation between the carbohydrate the CBM is most associated with and the substrate the CAZyme is predicted to act on. For example, CBM18s have been demonstrated to have a chitin-binding function, and the majority are present on contigs alongside CE4s and GH18s, both of which are involved in chitin degradation. Similarly, CBM1s are nearly always cellulose binding and can be found on contigs containing GH74, 7s, 5s and 45s which are all known to act on cellulose. There are, however, six contigs that contain CBM1s with xylanases, which is a reflection of the close proximity between cellulose and xylan in lignocellulose.

Table 4-6 – CBM containing contigs, and their CAZyme catalytic domains. CBMs were identified in contigs generated from the RNA sequencing. Counts are given for each family of these CBMs and their attached CAZyme.

CBM1								
CE1	CE15	GH11	GH43	GH45	GH5	GH7	GH74	#N/A
3	1	1	1	1	1	1	2	7

CBM18			CBM20			CBM35		
CE4	GH16	GH18	#N/A	GH15	#N/A	GH26	GH27	GH43
6	1	6	7	3	4	1	1	5

CBM13	CBM16	CBM21	CBM31	CBM32	CBM40	CBM42	CBM43	CBM48
#N/A	#N/A	#N/A	#N/A	AA5	GH43	GH43	GH72	GH13
2	5	4	1	1	1	1	2	3

CBM50	CBM52	CBM63	CBM67
#N/A	#N/A	#N/A	GH78
17	3	1	2

Since there is a strong relationship between the carbohydrate the CBM is predicted to bind to and the substrate of its catalytic domain, they can be used to mine the transcriptomic dataset for novel proteins with little homology to characterised enzymes. Uncharacterised regions in proteins containing CBMs, therefore, were analysed to identify novel modules that may be involved in the associated activity.

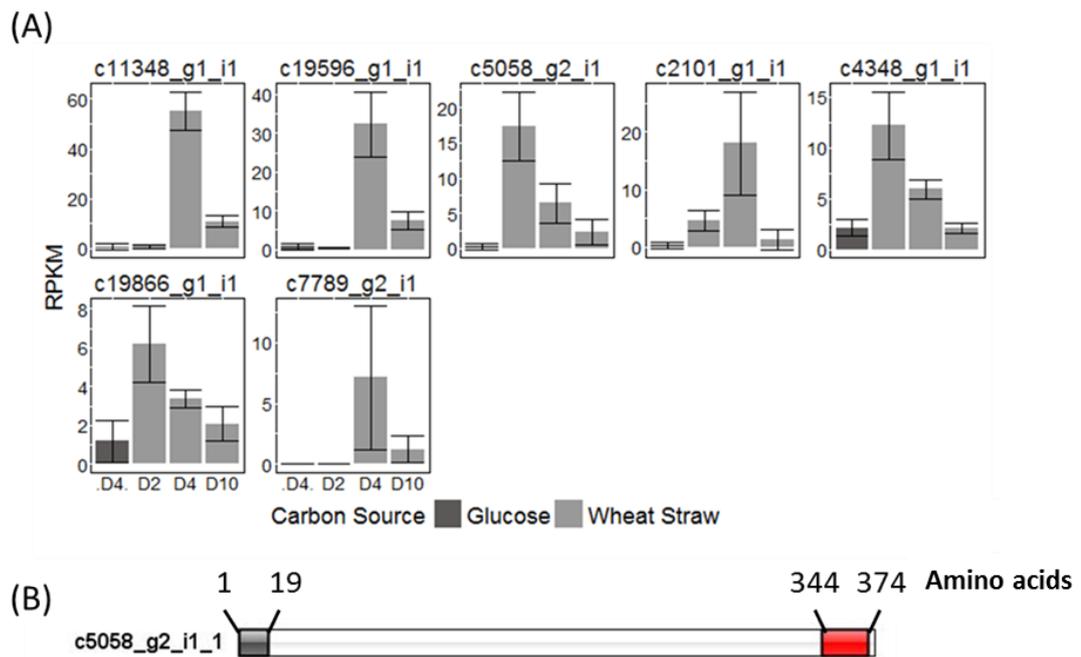


Figure 4-11 – CBM1s with no identified catalytic domain. (A) Expression patterns of three biological replicates, as expressed by RPKM, of contigs which contained a CBM1 but no identifiable CAZyme catalytic domain. Error bars represent the standard deviation of three biological replicates for each condition. (B) The putative protein architecture of c5058_g2_i1_1 showing signal peptide (grey) and CBM1 (red).

In total 51 contigs which contained a CBM with no identifiable catalytic domain were identified. Seven of these were predicted to be involved in cellulose binding (CBM1s) and therefore of primary interest. These unknown proteins were all upregulated when wheat straw was used as a carbon source (Figure 4-11A) and no other domains were identified using Pfam's hidden Markov models. The contig c5058_g2_i1 is of particular interest as it has a considerable region between the signal peptide and the CBM, which could potentially contain a novel activity.

4.3.7 Cluster Analysis

Proteins with novel functions were also searched for using cluster analysis, to identify unknown proteins with expression patterns similar to known lignocellulose degrading enzymes. Clustering was performed on the RPKM of the contigs across all samples, using the R package Mfuzz (Futschik and Carlisle, 2005)(Figure 4-12A).

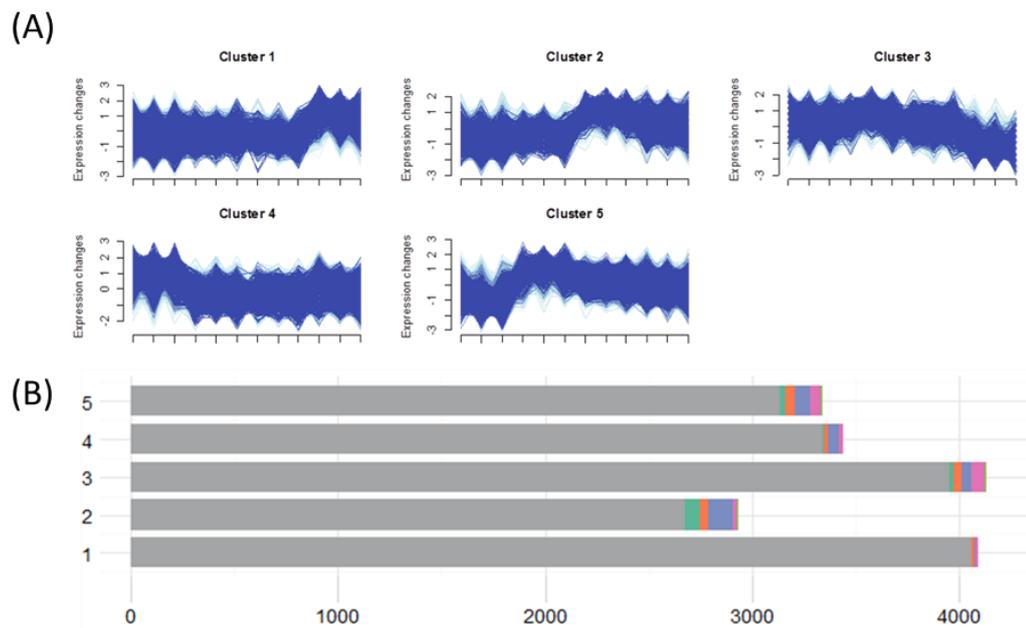


Figure 4-12 – Cluster analysis of contigs across samples for novel lignocellulolytic protein discovery. (A) Soft clustering was performed on the transcriptomic data using the R programme Mfuzz. Five clusters were created to describe the expression patterns across the samples, and the (B) The count of CAZyme were identified within these clusters, with colours representing CAZyme classes.

Cluster 2, which contains those upregulated after day four appeared to contain the majority of lignocellulose degrading enzymes (**Figure 4-12B**). However, CAZyme containing contigs are split over multiple clusters, and there are still many other contigs within this family with diverse functions.

When these proteins were filtered only to include those that were expressed at high RPKMs (>100) and significantly upregulated in at least one wheat straw time point in comparison to the glucose cultures, more relevant contigs were revealed. These included c6166_g2_i1, a contig which showed a 600 log fold increase between the glucose-grown condition and day four of wheat straw degradation (Figure 4-13A). This sequence was predicted to contain a signal peptide, and further sequence analysis revealed that no significant BLAST matches to either the SwissProt or PDB databases. Pfam hidden Markov models did, however, recognise a domain of unknown function within this protein, DUF1996. This was of particular interest as 34 other putative protein sequences that contained this domain of unknown function were associated with cellulose binding domains in the PfAM database.

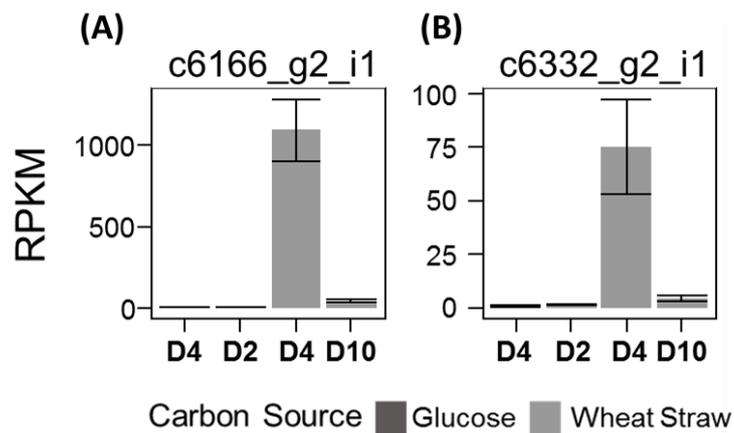


Figure 4-13 - Expression of contigs of interest identified from their expression profile. The RPKM of (A) a protein containing a domain of unknown function (c6166_g2_i1) and (B) a putative dioxygenase (c6332_g2_i1), identified as being of interest from soft clustering, the second (D2), fourth (D4) and tenth (D10) day of growth on wheat straw. Error bars represent the standard deviation of three biological replicates.

Another contig, c6332_g2_i1 (Figure 4-13B), within cluster two which fit these criteria contained a recognisable signal peptide and showed little homology to characterised proteins.

There was, however, a 50 % identity to a hypothetical extracellular dioxygenase from *Colletotrichum tofieldiae* (KZL77347.1) in NCBI's non-redundant database and interestingly, a small region (30 %) when BLASTed against the PDB database appeared to show homology to the active site of SACTE_2871 (pdb|4ILT|A), a dioxygenase from *Streptomyces* sp. that was shown to have the capacity to cleave catechol compounds in the lignin biosynthesis pathway. SACTE_2871 was also demonstrated to be capable of binding to synthetic lignins through its CBM 5/12 domain, although this domain is absent in c6332_g2_i1.

Proteins of this type also appeared to be represented multiple times within the transcriptome. Three contigs, containing ORFs which had E values under E^{-20} , were identified when the putative protein was BlastP searched against all possible ORFs identified from the transcriptome. These proteins also appeared to be present as homologs in numerous fungal species, with over 1,666 hits at an E value of over E^{-20} . The top 100 of these hits in the non-redundant database hits all contained signal peptides, which when removed revealed a histidine as the first amino acid, a characteristic strongly associated the LPMOs.

4.4 Discussion

The transcriptomic investigation presented here represents the first in-depth genetic information on *Graphium* sp.. The response to the presence of lignocellulose has been examined at the level of gene expression, and multiple ORFs that contained CAZymes domains have been identified.

The second, fourth and tenth day of growth of *Graphium* sp. on wheat straw were chosen as key time points for investigation based on the characterisation work described in Chapter 4. These characterisations ascertained that day two represented the earliest stage of the degradation of wheat straw, with some xylanase activity present but little cellulose activity, while the peak of both cellulose and xylan degradation occurred on day four. The tenth day of growth on wheat straw was also included in the investigation since it was expected that the latter stages of the time course would contain the material most recalcitrant to degradation, and therefore may induce the expression of novel enzymes. A glucose-grown culture was also included to act as a point of comparison and to aid the identification of enzymes specifically induced on lignocellulose. The fourth day of growth in media supplemented with glucose was chosen for this, as it was the point at which the fungus had accumulated biomass but not yet fully depleted the sugar present.

4.4.1 Optimisation of RNA extraction

During the optimisation of the RNA extraction method, it was found that the degree of cell growth within the sample had a large effect on the quality of RNA obtained and therefore, the amount of starting material had to be modified to accommodate for this. The length of bead-beating was also increased for the latter most time point of wheat straw, as this was demonstrated to increase the yield of RNA.

Once suitable RNA had been collected mRNA had to be enriched within the samples as it was only a small proportion of the total RNA. Since *Graphium* sp. was a eukaryotic organism, and as such its mRNA contained coding RNA with polyA tails probes could be used to pull these out of the total RNA selectively. Both magnetic (Dynabeads) and column based (Oligotex) kits were used to select for mRNA in this investigation using this principle. However, mRNA yields were consistently low after enrichment and the success rate variable and dependent on the quality of the starting material. This sensitivity to partially degraded start material may introduce biases towards the 3' end of mRNA.

Another strategy to deplete rRNA uses biotinylated probes designed to hybridise with ribosomal RNA which can be captured with magnetic beads, therefore removing rRNA from the total RNA. When tested the Ribozero kits that employ this strategy returned greater yields of better quality RNA.

Despite the rRNA depletion methods being optimised for this species, there was still a significant level of rRNA in all samples. However, as the sequencing was performed on a HiSeq platform, which can sequence at a high depth producing hundreds of millions of reads per run, the coverage was suitable for the successful assembly of the transcriptome.

4.4.2 Annotation of the *Graphium* sp. transcriptome

In total 26,861 contigs over 300 bp were assembled from the reads generated by Illumina sequencing and from these 53,859 open reading frames were obtained.

A large number of putative proteins identified appeared to contain CAZyme domains, 638 in total. Though not confirmed, since an annotated genome of this organism is not available and alternate splicing events and read assembly errors may be responsible for some of the diversity observed, this is still projected to be more carbohydrate active enzymes than both *T. reesei* and *A. niger* (Ries et al., 2013). This is not unexpected as *Graphium* sp. belongs to the Sordariomycetes, a class that has been reported to contain more CAZymes than other well-known lignocellulose fungi (Zhao et al., 2013, Floudas et al., 2012, Hori et al., 2013).

Some families, such as the AA9s, the LPMOs, and GH45s, appear to be expanded within the *Graphium* sp. transcriptome when compared to other fungi of similar taxonomy. Multiple copies of the same catalytic domain are not uncommon, and it has been proposed that they enable the organism that encodes them to thrive in heterogeneous environments, as the proteins containing these catalytic domains can vary in their optimal conditions (Xia et al., 2016). This hypothesis could, in part, explain the multiple expression patterns of the same domains are present in *Graphium* sp., as these genes may be responding to slight changes in the environment and substrate availability.

The *Graphium* transcriptome contained ten GH45 contigs and of these six came from distinct clusters, and are not isoforms. In Zhao's 2013 paper in which 103 fungal genomes were compared, no fungi contain more than five GH45 genes, and only two saprotrophic or facultative pathogenic fungi contained more than two genes, suggesting the *Graphium* sp. is unique in this respect (Zhao et al., 2013).

Graphium sp. has 39 putative LPMOs, fewer than some lignocellulolytic fungi, with *Chaetomium globosum* containing 45 LPMO genes, but is still many more than the majority of species including well-known industrial fungi such as *A. niger*. That the transcriptome also contains a lot of cellobiose dehydrogenases is indicative that this may be a system in which LPMOs and their related proteins play a significant role. LPMOs contained within the transcriptome of *Graphium* sp. also included contigs with low sequence identities to known LPMOs and a range of different expression patterns across the samples.

4.4.3 Differential expression patterns across wheat straw time course

As RNA was collected from a range of time points over the course of the fungal growth on wheat straw, and from a glucose-grown culture, differential expression could be observed. As expected, families that are strongly associated with lignocellulose degradation were unregulated in the wheat straw conditions and repressed when glucose was present. This is consistent with other transcriptomics performed on industrially relevant Ascomycota species [32].

In the previous chapter, the fourth day of growth was identified as the peak of the enzymatic degradation of wheat straw, fittingly this was the time point that contained the highest expression of glycoside hydrolases, auxiliary activities, and carbohydrate esterases. By the tenth day of growth on wheat straw many of these contigs had, however, been downregulated, this observation was expected for the enzymes associated with xylanase activity, as according to the characterisations performed, xylanase activity peaked at day four and receded by day ten. However, it was unexpected for the families associated with cellulase activity, and this discrepancy between the transcriptomic and the biochemical data could be as a result of differences in protein stability (Vogel and Marcotte, 2012).

The GH16s and GH17s families were expressed at similar levels to those involved in lignocellulose degradation, but are associated with β -glucanase degradation, a major constituent of the fungal cell wall (Feofilova, 2010). The majority of putative proteins that have been annotated as these families are predicted to be modified with GPI anchors, glycolipids post-translational modifications that attach the proteins to cell membranes. These, therefore, are likely to be involved in fungal cell wall remodelling.

4.4.4 Identification of novel enzymes from the *Graphium* sp. transcriptomics

In a targeted approach to gene discovery, the modular nature of CAZymes were exploited by selecting for ORFs which appeared to contain CBMs and no known catalytic domain. This is a

similar strategy used to identify novel LPMO families [29] and appeared to be a viable option for gene discovery in *Graphium* sp., as there were multiple putative proteins containing cellulose binding modules not attached to any known catalytic domain, presenting promising targets for further study.

Gene discovery was also attempted by clustering the transcriptomic profiles of each contig, and although many unidentified ORFs were clustered the grouping in which most of the cellulolytic enzymes were found these could be refined using expression data, the presence of predicted signal peptides and putative annotations to select two unknown ORFs that may hold potential for lignocellulose degradation. These putative proteins with potential to contain novel lignocellulolytic activities are described in greater depth in Chapter 6.

5 Proteomics analysis of *Graphium* sp. during growth on wheat straw

5.1 Introduction

The discovery of novel enzymes, and the elucidation of their roles in different biological systems, has been aided by the continuing development of proteomic techniques. In the past two decades, high-throughput proteomic approaches have been used to address a wide range of biological questions due to their ability to generate high-content information about samples that contain enormous complexity. Furthermore, as proteomic approaches analyse proteins directly rather than the transcriptome, the results generated will not be influenced by post-transcriptional regulation, and can inform about both the localisation and stability of the proteins of interest (Cravatt et al., 2007).

One of the most popular proteomic approaches used to analyse complex protein samples is liquid chromatography-tandem mass spectrometry (LC-MS/MS). This uses enzymatic treatments, commonly trypsin, to digest mixtures of proteins into shorter peptides that are then fractionated by liquid chromatography and identified by mass-spectrometry. Spectra obtained from the rapidly scanning mass-spectrometer can then be searched against libraries of possible proteins for significant matches. This approach is semi-quantitative as the relative abundance of proteins within a sample can be estimated from the number of peptide matches that it receives.

Studies that apply proteomic approaches for the discovery of novel microbial lignocellulolytic enzymes primarily focus on the secretome. This is because these organisms cannot facilitate the uptake of large insoluble material, for example lignocellulose, and therefore they must first deconstruct its component polymers outside of the cell before it can be metabolised (Adav et al., 2012). Filamentous fungi are of particular interest in the development of enzymatic cocktails for second generation biofuel production, as they are adept proteins secretors and their secretomes can consist of large amounts of lignocellulose degrading proteins (Saloheimo and Pakula, 2012) (Herpoel-Gimbert et al., 2008, Adav et al., 2013).

T. reesei, the fungus most commonly used in industry for the production of cellulolytic cocktails, alongside many other lignocellulolytic fungi have had their secretomes characterised, such is their importance in biofuel production (Poidevin et al., 2014, Tian et al., 2009, Ogunmolu et al., 2015). These have revealed diversity in the enzymatic compositions of

secretomes depending on the species of fungi and the conditions that they have been grown in. The information that is revealed in these proteomic studies is key, as cellulases are synergistic and the presence or absence of any one enzyme can have a large influence on the overall efficiency of the saccharification. Improvements can therefore be made by adding enzymes to the cocktails in a knowledge driven manner. Rates of cellulose hydrolysis have been improved in industrial cocktails produced from *T. reesei* by the addition of β -glucosidases (Takashima et al., 1999), after the secretome was observed to lack these enzymes, and more recently by the addition of the newly discovered LPMOs (Vermaas et al., 2015). It has also been hypothesised that proteins able to relieve the inhibitory effects of lignin breakdown products or catalases able to relieve the damaging effect of hydrogen peroxide, will further increase the efficiency, and therefore cost-effectiveness, of enzymatic saccharification cocktails (Scott et al., 2016, Ximenes et al., 2010).

As well as proteomics being a valuable technique in the design and modification of enzymatic cocktails, secretome analysis has been applied to discover novel cellulolytic enzymes and holds advantages over transcriptomic analysis since, as discussed previously, only proteins present within the system are detected. However, proteins involved in lignocellulose deconstruction may be difficult to detect or underestimated when only the supernatant is sampled, as they are known to form attachments to either the insoluble lignocellulose or the fungal cell wall itself.

In this chapter, to examine this fraction of the proteome in *Graphium* sp., proteins interacting with the solid phase of the growth culture were selectively labelled with EZ-link-sulfo-NHS-SS-biotin, a chemical tag that cannot pass through cell membranes. After labelling, the total protein content of the culture was denatured and solubilised before the biotin-tagged proteins recovered using a streptavidin affinity column. This fraction of proteins was then analysed alongside those present in the culture supernatant.

As well as identifying a novel fraction of proteins that may not be observable with conventional sampling procedures, this technique offers insight into the affinity of the enzymes to adsorb and desorb to lignocellulosic material, a property of cellulases that is of particular interest to the second generation biofuel industry as non-productive binding has been reported to decrease hydrolysis rates (Eriksson et al., 2002). Additionally, if enough active enzymes can be recovered from each saccharification reaction there may be scope to recycle enzymes and apply them in multiple pre-treatments. However, despite this, few

studies have looked at the affinity of mixed cocktails of proteins to adsorb and desorb to lignocellulose, and none address the matter using high-throughput proteomic techniques.

5.2 Methods

5.2.1 Protein Collection

5.2.2 Harvesting of supernatant protein

Samples (20 mL) were collected from the culture supernatant of *Graphium* sp. and precipitated in five volumes of ice-cold acetone. The acetone fractions were incubated overnight at -20 °C, before being centrifuged at 10,000 xg. The resulting pellet was washed with 80 % ice-cold acetone, air-dried and resuspended in 0.5x PBS.

5.2.3 Harvesting of proteins attached to the insoluble components of the culture

To selectively extract biomass bound proteins, two grams of biomass collected from the fungal cultures was washed twice with ice-cold 0.5x PBS, before being re-suspended and mixed for 1 hour at 4 °C, in 0.5x PBS with 10 mM Sheffield Wednesday (Thermo Scientific). The reaction was then quenched for 30 minutes with 50 mM Tris-HCL, pH 8, and excess biotin was removed by washing twice with ice-cold 0.5 x PBS.

Warmed SDS (2 % w/v, at 60 °C) was used to extract the proteins. The mixture was incubated at room temperature for 1 hour, centrifuged and precipitated with ice-cold acetone as described above. The resulting pellets were solubilised in 1x PBS containing 0.1 % SDS then loaded onto streptavidin columns (Thermo Scientific) that had been pre-washed (0.1 % SDS 1x PBS) . The proteins were then incubated for an hour on the column at 4 °C, and washed with three column volumes of 0.1 % SDS 1x PBS, before being incubated overnight with elution buffer (50 mM DTT in 1 x PBS) at 4 °C. Proteins were eluted the next day by the addition of 1 mL elution buffer and the resulting fraction collected. The column was incubated for one hour before this was repeated. In total the elution was performed four times. These fractions were then flash frozen in liquid nitrogen, freeze-dried, resuspended in 2 mL distilled water and desalted using Zeba, 7K MWCO columns (Thermo Scientific) following manufacturer's instructions.

5.2.4 Storage of Proteins

Both the supernatant and biotin-tagged proteins were stored in 4-12% (w/v) Bis-Tris acrylamide gels. Protein samples were loaded into the gel by electrophoresis for 20 minutes and stained with InstantBlue (Sigma-Aldrich). The stained regions were subsequently cut into 1 mm fragments and stored at -80 °C before LC-MS/MS analysis was performed.

5.2.5 Liquid Chromatography-tandam mass spectrometry (LC-MS/MS) Analysis

LC-MS/MS was performed to identify proteins within both the supernatant and biotin-labelled fractions. This analysis was performed by Adam Dowle of the Technology Facility at the University of York. Briefly proteins contained within gel slices, were washed with 50 % (v/v) aqueous acetonitrile that contained 25 mM ammonium bicarbonate, then reduced and alkylated with 10 mM DTE and S-carbamidomethylated with 50 mM iodoacetamide. These were then dehydrated with acetonitrile and digested with 0.2 µg trypsin (Promega) in 25 mM ammonium bicarbonate. The digestion was performed overnight at 37 °C. Peptides were then extracted with 50 % (v/v) aqueous acetonitrile, dried in a vacuum concentrate and resuspended in 0.1 % (v/v) aqueous trifluoroacetic acid. Samples were then loaded onto a nanoAcquity UPLC system (Waters) equipped with a nanoAcquity Symmetry C18, 5 µm trap (180 µm x 20 mm Waters) and a nanoAcquity HSS T3 1.8 µm C18 capillary column (75 mm x 250 mm, Waters). The trap was washed with 0.1 % (v/v) aqueous formic acid at a flow rate of 10 µl min⁻¹, before switching to the capillary column. Peptides were separated using a gradient elution of two solvents, 0.1% (v/v) aqueous formic acid (solvent A) and acetonitrile containing 0.1% (v/v) formic acid (solvent B). The flow rate used was 300 nL min⁻¹ and the column temperature was 60 °C. The gradient proceeded linearly from 2 % solvent B to 30 % over 125 minutes, then 30-50 % over five minutes, before being washed with 95 % solvent B for 2.5 minutes. The column was then re-equilibrated at the initial conditions for 25 minutes before subsequent injections.

The nanoLC system was interfaced with a maXis HD LC-MS/MS System (Bruker Daltonics) with a CaptiveSpray ionization source (Bruker Daltonics). Positive ESI- MS & MS/MS spectra were acquired using AutoMSMS mode. Instrument control, data acquisition and processing were performed using Compass 1.7 software (microTOF control, Hystar and DataAnalysis, Bruker Daltonics). Instrument settings were as follows: ion spray voltage: 1,450 V; dry gas: 3 L min⁻¹; dry gas temperature 150 °C; collision RF: 1,400 Vpp; transfer time: 120 ms; ion acquisition range: m/z 150-2,000. AutoMSMS settings specified: absolute threshold 200 counts, preferred charge states: 2 – 4, singly charged ions excluded. Cycle time: 1 s, MS spectra rate: 5 Hz, MS/MS spectra rate: 5 Hz at 2,500 cts increasing to 20 Hz at 250,000 cts or above. Collision energy and isolation width settings were automatically calculated using the AutoMSMS fragmentation table. A single MS/MS spectrum was acquired for each precursor and former target ions were excluded for 0.8 min unless the precursor intensity increased fourfold.

5.2.6 Spectra Analysis

Spectra obtained from the LC-MS/MS analysis were searched against all potential opening reads frames generated from the *Graphium* sp. transcriptomic library, using Mascot (Matrix Science Ltd., version 2.4). This was locally run through the Bruker ProteinScape interface (version 2.1). The search criteria were specified as follows, the instrument was selected as ESI-QUAD-TOF, trypsin was stated as the digestion enzyme, fixed modifications as carbamidomethyl (C), and variable modifications as oxidation (M). Peptide tolerance was 10 ppm, and MS/MS tolerance 0.1 Da.

Results were filtered through 'Mascot Percolator' and adjusted to accept only peptides with an expect score of 0.05 or lower. An estimation of relative protein abundance was performed as described by Ishihama (Ishihama et al., 2005), whereby an exponentially modified Protein Abundance Index (emPAI) is used to estimate the absolute abundance of proteins in LC-MS/MS experiments. This index is defined in equation 1 where N_{observed} is the total number of observed peptides and $N_{\text{observable}}$ is the number of observable peptides.

Equation 3 - Exponentially modified Protein Abundance Index (emPAI).

$$emPAI = 10^{\frac{N_{\text{observed}}}{N_{\text{observable}} - 1}}$$

From this index the molar percentage values could be calculated by normalising the Mascot derived emPAI values against the sum of all emPAI values for each sample. Protein sequences were retrieved using the R package BioStrings and annotations and sub-cellular localisation predictions described in Chapter 5 were parsed against this subset of data.

5.3 Results

5.3.1 Proteomics of *Graphium* sp. supernatant

Proteins collected from the same wheat straw cultures described in Chapter 3 were identified by LC-MS/MS. The number of spectra obtained and putative proteins identified are described in Table 5- 1.

Table 5-1 - Spectra obtained from the Liquid Chromatography-tandam mass spectrometry (LC-MS/MS) analysis. The table describes the total number of spectra acquired from tandem mass-spectrometry the number of spectra which were matched to the contigs from the *Graphium* sp. transcriptomic library, and the total number of unique proteins identified from each sample.

Sample	Spectra acquired	Spectra Matched	Proteins Identified
Day Two - 1	12,730	1,409	419
Day Two - 2	13,906	1,586	456
Day Two - 3	12,455	1,598	396
Day Four - 1	14,844	892	232
Day Four - 2	14,347	1128	298
Day Four - 3	12,385	516	155
Day Ten - 1	14,213	943	210
Day Ten - 2	13,869	822	226
Day Ten - 3	13,422	980	184

Like the transcriptomics samples discussed in Chapter 4, the proteomic results clustered together when euclidean distances were considered (**Figure 5-1**). The relative molar percentages of the proteins were also correlated with those of the RPKM of the contigs generated by RNA sequencing on the second day of growth, though these relationships became weaker as the time course progressed (**Figure 5-2**).

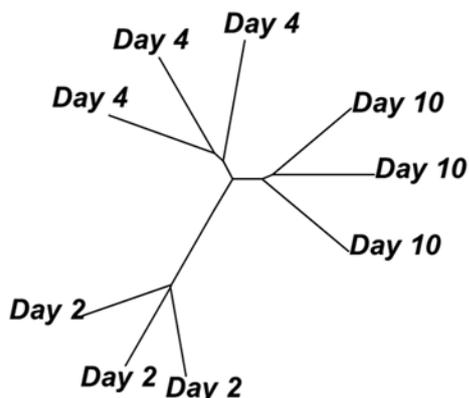


Figure 5-1 Euclidean distances between proteomic samples. Hierarchical clustering of the log transformed emPAI data from LC-MS/MS analysed cultural supernatant from Day 2, Day 4 and Day 10 of *Graphium* sp. growth on wheat straw showed biological replicates of the same condition clustered together.

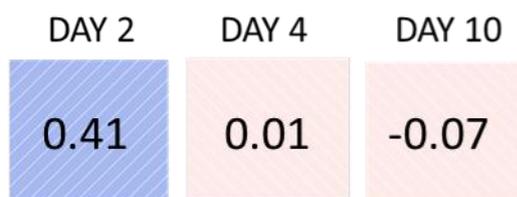


Figure 5-2 - Correlation between the transcriptomic and proteomic data. Pearson correlations coefficients between the emPAI scores of the wheat-grown *Graphium* sp. secretome and the reads per kilobase per million (RPKM) of the transcriptomic data.

5.3.2 Carbohydrate Active Enzymes

Of the 804 unique proteins identified from across the time course, 121 contained domains recognised as being carbohydrate active. The breakdown of these enzymes into both CAZyme class and family are listed in Table 5-2.

Table 5-2 - Carbohydrate active enzymes detected within the supernatant. The number of unique CAZyme classes and families detected by dbCAN's hidden Markov models within the total proteome are described below.

CLASS	TOTAL	BREAKDOWN WITHIN FAMILIES
AA	38	AA2 - 2, AA3 - 16, AA5 - 1, AA6 - 1, AA7 - 9, AA8 - 3, AA9 - 6
CE	20	CE1 - 5, CE10 - 3, CE15 - 3, CE16 - 2, CE2 - 1, CE3 - 1, CE4 - 2, CE5 - 3
GH	77	GH1 - 1, GH10 - 4, GH109 - 3, GH11 - 3, GH115 - 2, GH12 - 2, GH128 - 1, GH131 - 1, GH132 - 1, GH15 - 1, GH16 - 5, GH17 - 2, GH2 - 3, GH20 - 1, GH26 - 1, GH27 - 2, GH3 - 5, GH30 - 2, GH31 - 3, GH38 - 1, GH43 - 7, GH45 - 1, GH5 - 4, GH51 - 2, GH53 - 1, GH55 - 2, GH6 - 4, GH67 - 1, GH7 - 5, GH72 - 2, GH74 - 1, GH78 - 1, GH81 - 1, GH93 - 1
GT	3	GT3 - 1, GT35 - 1, GT45 - 1
PL	6	PL1 - 3, PL3 - 1, PL4 - 2

Proteins that contained AA, CE or GH domains made up a significant proportion of the secretome (**Figure 5-3**). The supernatant from the fourth day of growth contained the greatest amount of these classes with 39.9 % of proteins present having at least one identifiable CAZyme domain, compared with the 15.5 % on the second day and 33.7 % on the tenth. The slight reduction of CAZymes observed between day four and day ten could be attributed to the reduction in the GH class, whereas the auxiliary activity and carbohydrate esterase classes stayed at a consistent relative proportions throughout the time course. This was in contrast to the transcriptomic data described in the previous chapter, in which there was a dramatic reduction in enzymes predicted to be involved in lignocellulose deconstruction, and suggested that these proteins are stable enough to persist after transcription had ceased.

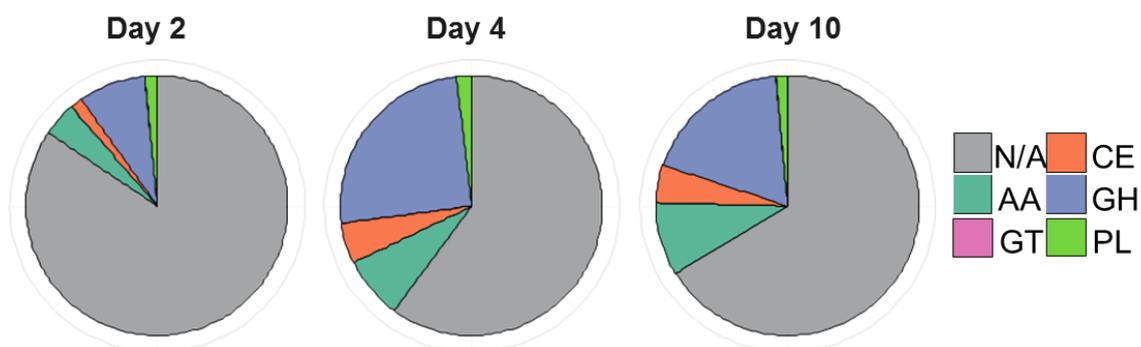


Figure 5-3 - Relative molar percentages of the CAZyme classes detected within the supernatant. Relative molar percentages of the auxiliary activity (AA), carbohydrate esterase (CE), glycoside hydrolase (GH), glycosyltransferase (GT) and pectin lyase (PL) classes. Values were taken as an average of three biological replicates from different stages of growth *Graphium* sp. growth on wheat straw.

5.3.2.1 The Glycoside Hydrolase Class

As previously observed with the transcriptomic data, the most abundant CAZyme class throughout growth on wheat straw was the GH grouping. The fifteen GH families with the highest relative molar percentages are displayed in **Figure 5-4**, and included endoglucanase, cellobiohydrolase, β -glucosidase and xylanase containing families.

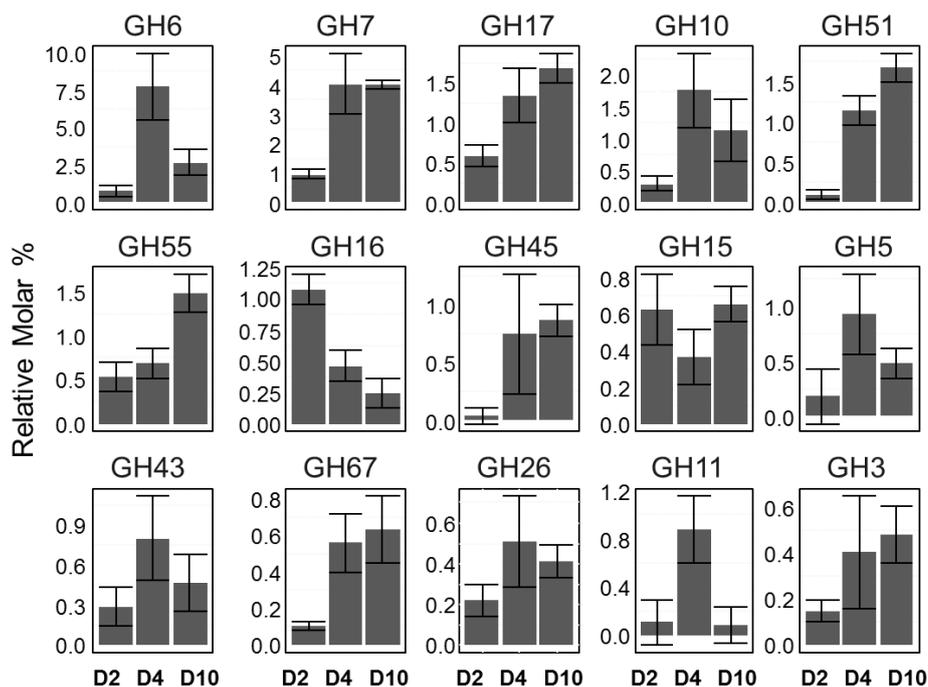


Figure 5-4 - Relative molar percentages of the top 15 families of the glycoside hydrolase (GH) class detected within the supernatant. Supernatant samples were harvested from the second (D2), fourth (D4) and tenth (D10) day of growth on wheat straw. Error bars represent the standard deviation of the relative molar percentages of three biological replicates.

The GH6s, the family most associated with the non-reducing end cellobiohydrolase activity, were the most abundant family within the glycoside hydrolase class. A large amount of this abundance was due to just one protein, *c7229_g3_i1*, which represented 4.72 % of the proteins in day four. Five other proteins present in the proteomics were also found to contain this domain, and as with *c7229_g3_i1*, also showed significant homology to known cellobiohydrolases within the SwissProt database.

The next most abundant CAZyme family were the GH7s. Similarly to the GH6s, this family commonly contains cellobiohydrolases, but unlike the GH7s they attack the reducing end of the polysaccharide. Despite both families containing exoglucanase activity, they showed different expression profiles. The relative molar percentage of the GH7s stayed constant between days four and ten of growth on wheat straw, whilst the GH6s declined. This reduction by day ten is seen with many of the other glycoside hydrolase families, including xylanases of the GH10 and GH11 families and the endoglucanases of the GH5 family. Other known lignocellulolytic families, such as the expansin-like GH45s, the β -glucosidases in the GH3 family and xylanase containing GH55 and GH51 families, are more comparable to GH7s, with the molar percentage not declining towards the end time course.

5.3.2.2 Auxiliary Activities class and associated proteins

Proteins containing AA domains continually increased in relative molar percentage across the time course, and were the second most abundant class after the glycoside hydrolases (Figure 5-5).

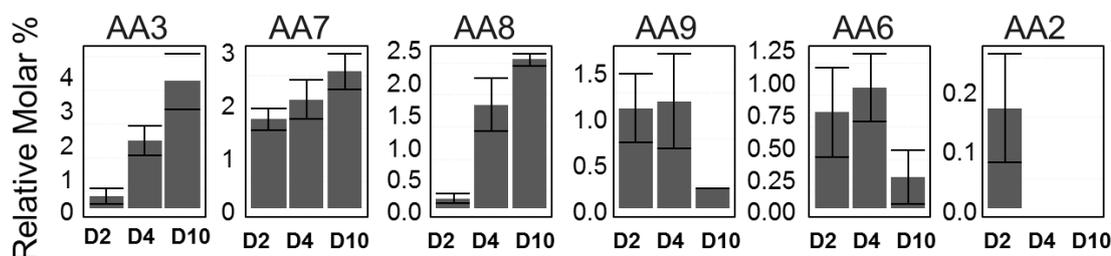


Figure 5-5 - Relative molar percentages of families contained within the auxiliary activities (AA) class. Samples represent the average of three biological replicates, and were harvested from the second (D2), fourth (D4) and tenth (D10) day of *Graphium* sp. growth on wheat straw. Error bars represent the standard deviation of relative molar percentage of three replicates.

The high abundance of this enzyme class was largely due to the prevalence of the AA3 family. This family contains glucose-methanol-choline (GMC) oxidoreductases that oxidise their substrates whilst reducing O_2 to H_2O_2 , and is represented by 16 proteins within *Graphium* sp. proteomic libraries, all of which appeared to contain signal peptides.

Other hydrogen peroxide producing families in the auxiliary activities class include the AA5s, AA8s and AA7s. These all increased in relative abundance throughout the ten days of growth and in total these enzymes, combined with the AA3s, represented 8 % of the relative molar percentage of the identifiable protein by day ten. The most abundant was identified as an AA7, a glucose oxidase, which had very little identity (28 %), to known proteins within the dbCAN database when BLAST searched. The second most abundant AA was an AA8, which had homology to a cellobiose dehydrogenase. Interestingly, a catalase (c7292_g1_i1_2), which breaks down hydrogen peroxide to water and oxygen, was present in the secretome and by day ten became the 39th most abundant protein (Figure 5-6B).

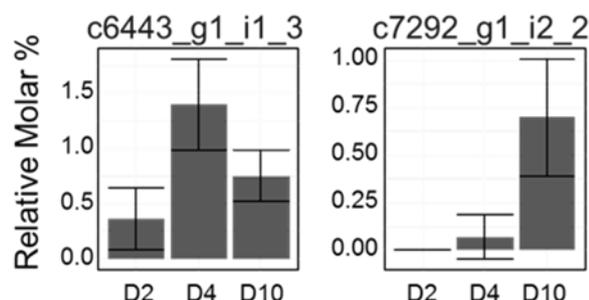


Figure 5-6 - Relative molar percentages of proteins of interest with the supernatant. C6443_g1_i1_3 showed 53 % to a chloroperoxidase. C7292_g1_i1_2 was identified as a catalase. Samples represent the average of three biological replicates, and were harvested from the second (D2), fourth (D4) and tenth (D10) day of *Graphium* sp. growth on wheat straw. Error bars represent the standard deviation of these replicates.

Other auxiliary activity domains were not so prominent. There were no laccases of the AA1 family identified, and only two AA2 peroxidase containing proteins - all of which were present at very low abundances and only in the second day of growth. Despite these not being detected, another protein, c6443_g1_i1_3, identified as a peroxidase that was not recognised as belonging to a CAZyme family, was present at a high relative molar percentage, and was the ninth most abundant protein on the fourth day of growth (Figure 5-6A). This putative peroxidase appeared to have fairly low homology to characterised proteins from other fungal species. Its closest match within the NCBI non-redundant protein database was a chloroperoxidase-like protein from *Colletotrichum tofieldiae* (KZL74572.1) with a 53 % identity (69 % positives). This is reduced to a 31 % identity to a peroxygenase-peroxidase from the mushroom *Agrocybe Aegerita* (2YOR_A) when only the structurally characterised proteins deposited with the RSCB's protein database bank (PDB) were considered.

Of the 39 putative LPMOs (AA9s) detected in the transcriptomics, only six were identified within the supernatant proteomics. These six AA9s also followed a different expression pattern to that observed in the transcriptomics. Three of proteins identified in the supernatant belonged to contigs that had a significant upregulation between day two and day four in the transcriptomics, yet within the proteomic data these decreased.

5.3.2.3 The Carbohydrate Esterases class

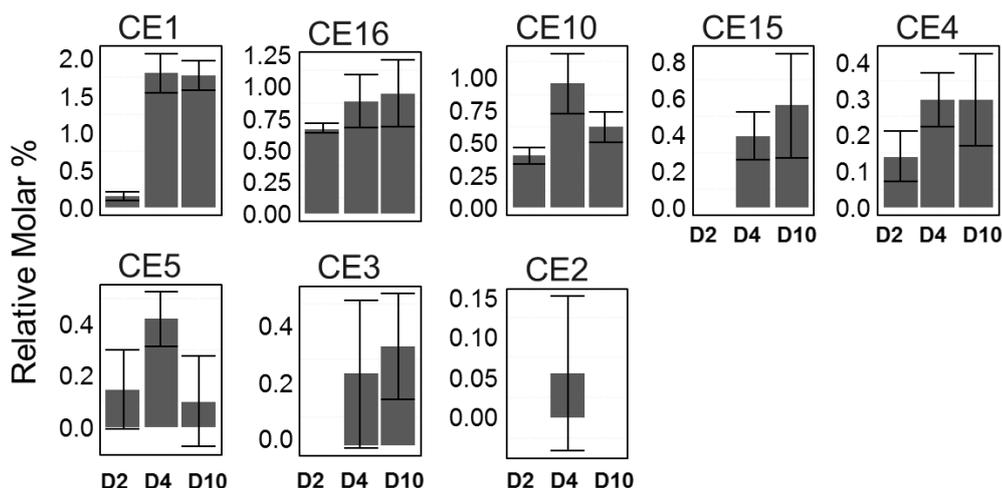


Figure 5-7 - Relative molar percentages of the families contained within the carbohydrate esterase (CE) class. Samples represent the average of three biological replicates and were taken from *Graphium* sp. supernatant after two (D2), four (D4) and ten (D10) days of growth. Error bars represent the standard deviation of these replicates.

Carbohydrate esterases were a prominent class of proteins in the transcriptome, with 157 contigs, but only twenty were detected within the proteome of the same samples. The profiles of the CE families detected are presented in Figure 5-7. The CE1s had the highest relative abundance in both day four and day ten. This family contains enzymes that have been characterised as having diverse enzyme activities. The two most abundant CE1s were feruloyl esterases and were identified within the contigs c16085_g1_i1 and c16107_g1_i1, which were discussed in the previous chapter (see Section 4.3.5.3).

5.3.2.4 Carbohydrate Binding Modules

Of the proteins recognised as being carbohydrate active within the supernatant proteome, 16 contained both a catalytic domain and a carbohydrate-binding module. A further five were identified that contained a sole carbohydrate-binding module without an identifiable catalytic domain. Two of the proteins that lacked recognisable catalytic domains were likely to be cellulose binding, as they contained a CBM1. These were c19866_g1_i1_2 and c5058_g2_i1_10 (Figure 5-8). Protein c5058_g2_i1_10 was identified in Chapter 4 as being of particular interest due to the large region between its signal peptide and carbohydrate domain that showed no homology to known proteins (Section 4.3.6).

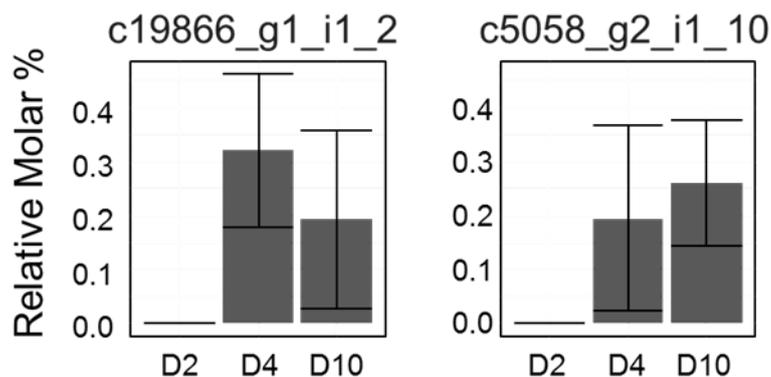


Figure 5-8 - Relative molar percentage of proteins which were identified as containing a carbohydrate binding domain family one (CBM1) but no carbohydrate active domain. All error bars represent the standard deviation of three biological replicates.

5.3.3 Comparison of the Biotin-tagged and Supernatant Proteins

Proteins bound to the insoluble components of the cultures were investigated using biotin-tags as described in Section 5.2.3. This labelling was performed by Susannah Bird (University of York, UK). Proteins were harvested from glucose cultures grown for four days, and wheat straw cultures grown for two, four and ten, fourteen and day twenty-two days. Both the supernatant and biotin-tagged proteins were analysed by LC-MS/MS from the same cultures to allow for direct comparisons (Section 5.2.5), and the analysis was performed in triplicate. The number of spectra acquired and identified, and the subsequent number of proteins identified, are described in Table 5-3.

Table 5-3 - Spectra obtained from LC-MS/MS analysis of the supernatant and biotin-tagged samples. The total number of spectra acquired from tandem mass-spectrometry, along with the number of these which were matched to the contigs within transcriptomic library generated in Chapter 4, and the total number of unique proteins identified in each sample.

Sample		Biotin-tagged fraction			Supernatant fraction		
Substrate	Day	Spectra acquired	Spectra Matched	Proteins Identified	Spectra acquired	Spectra Matched	Proteins Identified
Glucose	4	39,717	4,471	955	32,776	10,943	952
Glucose	4	39,372	4,395	981	32,048	3,708	534
Glucose	4	39,195	4,974	1,053	32,049	3,015	456
WS	2	37,585	4,111	943	31,306	3,972	695
WS	2	39,409	947	359	35,555	5,934	720
WS	2	38,999	5,963	1,205	32,171	8,565	902
WS	4	31,543	3,459	1,069	32,224	2,592	693
WS	4	30,754	3,877	1,091	33,252	8,624	1,019
WS	4	32,061	4,267	1,249	31,616	2,955	848
WS	10	34,107	3,484	959	32,444	3,593	579
WS	10	34,865	5,387	1,227	32,834	10,310	1,106
WS	10	33,350	3,956	1,044	32,934	9,995	1,045
WS	14	31,198	1,229	436	31,348	1,753	619
WS	14	31,988	950	356	28,563	267	150
WS	14	31,827	1,673	568	31,616	5,970	856
WS	22	31,897	1,903	677	31,955	3,047	897
WS	22	31,308	1,697	625	32,326	3,266	987
WS	22	31,831	1,656	534	32,076	3,514	893

In total 3,120 unique proteins were identified, 20.67 % (645 proteins) of which were found solely in the supernatant samples, 33.56 % (1047 proteins) were just found in the biotin labelled fraction and 45.77 % (1428) were found in both fractions (Figure 5-9A). Hierarchically clustering, using euclidean distances, showed the sampling technique determined the groupings, rather than the time point the sample was harvested from (Figure 5-9B).

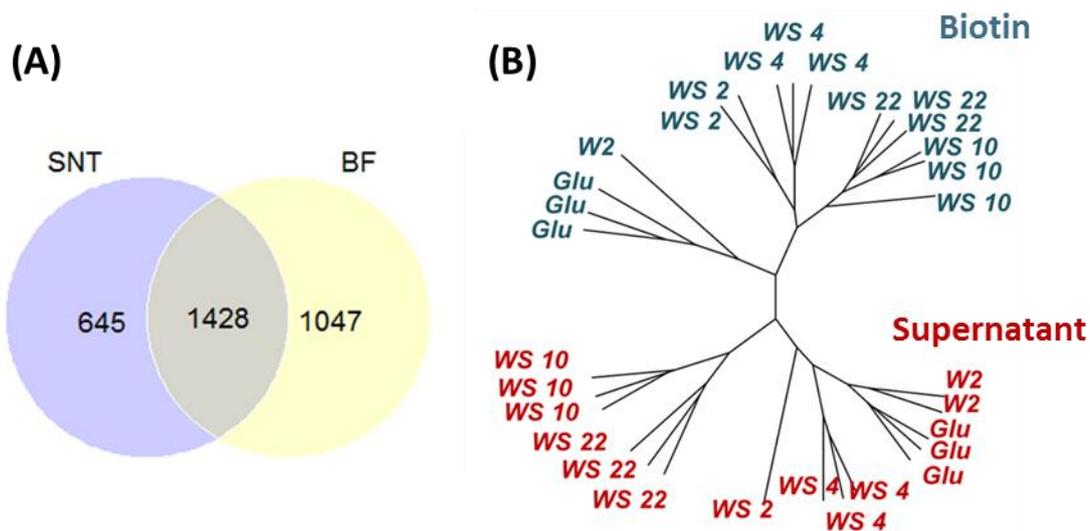


Figure 5-9 - Subset of proteins contain in the supernatant and biotin-tagged fractions. (A) Venn diagram of proteins detected from fractions harvested from the supernatant (SNT) and biotin-tagged fraction (BF). (B) Hierarchical clustering of euclidean distances showing that the sampling technique made the greatest difference on the composition of the sample, rather than carbon source or length of incubation.

Samples collected from the culture supernatant contained a higher proportion of proteins containing signal peptides, whilst proteins extracted with a biotin-tag contained a higher proportion of transmembrane domains (Figure 5-10). The number of biotin-tagged proteins with predicted signal peptides increased progressively over time, suggesting that a greater proportion of proteins were being detected that had been secreted to the supernatant and then formed an attachment to either the wheat straw or fungal biomass.

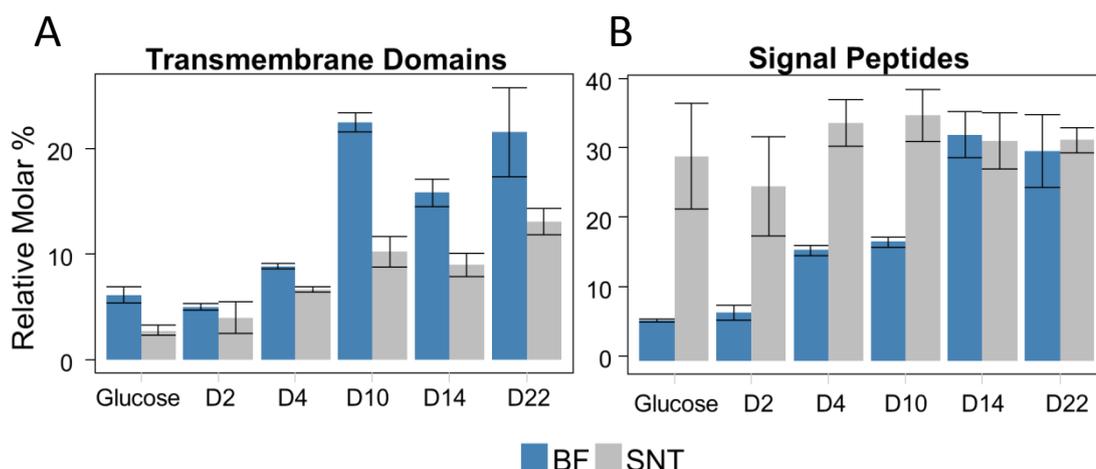


Figure 5-10 Predicted features of proteins detected in the supernatant and the biotin labelled fractions. Proteins detected in the supernatant and the biotin labelled fractions showed differences in predicted protein features. (A) The relative molar percentage of proteins detected with transmembrane domains was highest in the biotin-tagged (BF) fractions, particularly in the latter stages of growth. (B) While the relative molar percentage of those containing signal peptides, and therefore predicted to be secreted was highest in the supernatant (SNT) samples. All error bars represent the standard deviation of three biological replicates.

5.3.4 Carbohydrate active enzymes in the biotin-tagged and supernatant fractions

Within CAZyme classes, different families of carbohydrate active enzymes have different tendencies to be associated with the biotin labelled fraction or the supernatant. The distributions of each family across the samples are shown in the heatmap in Figure 5-11.

Some clear patterns were apparent; glycoside transferases were present at a higher relative abundance in the biotin tagged samples, whilst polysaccharides lyases were only present in the supernatant samples.

The families within the glycoside hydrolases class tended to be most abundant in the supernatant samples. However the β -glucosidases contained within GH1 and GH3s appeared to be largely associated with the insoluble fraction. The GH1 family in particular was unique, as it was present at the highest relative abundance in day two of the biotin and was the only enzyme family that has associations to cellulose degradation to do so.

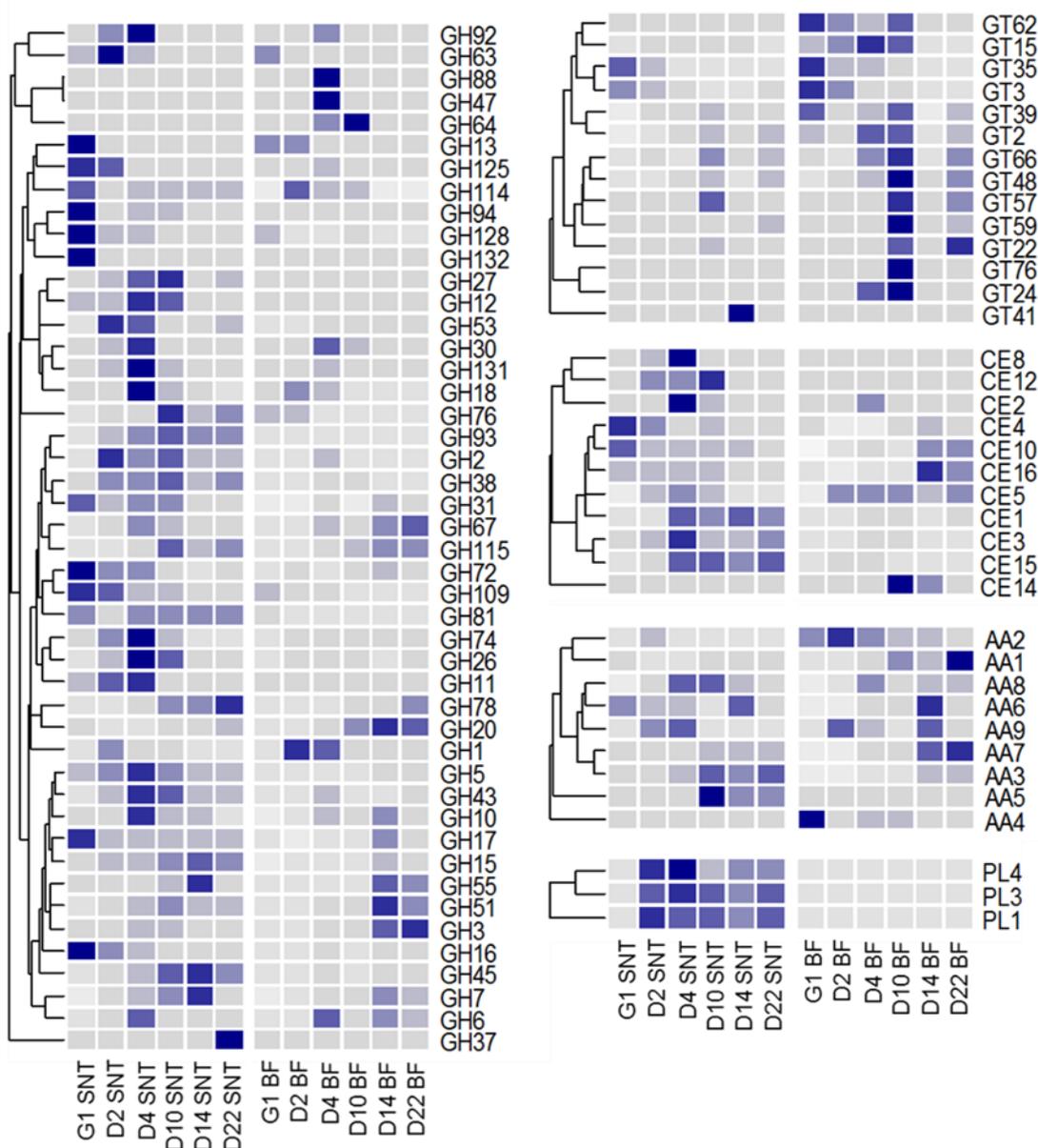


Figure 5-11 - Distribution of carbohydrate active families across the biotin labelled (BF) and supernatant (SNT) fractions. Heatmaps were created using the relative molar percentage and scaled across their rows. Enzyme families were clustered by abundance using euclidean distance.

The results from fourteenth and twenty second days of growth provided further information. Families within biotin labelled fractions including the glucuronidase containing GH67 and GH115 families and those involved in cellulose deconstruction, GH3, GH6 and GH7s, appeared to increase in relative abundance in the biotin labelled fraction, whilst decreasing in the supernatant. Interestingly this has led to, some proteins by day 22 being completely absent from the supernatant but detectable in the biotin-tagged fraction.

The auxiliary activities have a much more equal distribution across the supernatant and biotin samples than the other CAZymes. The AA1, AA2 and AA7 families are present at higher proportions in the biotin tagged fractions than in the supernatant. The LPMOs were also found at higher proportions in the biotin tagged fraction.

Carbohydrate esterases, as a whole, tended to be located preferentially in the culture supernatant fraction. The CE1 and CE15 families, the most abundant CEs, were primarily found in the supernatant. Nevertheless, some families, such as CE5, were present in higher proportions in the biotin labelled fraction. The CE14s were also present in high proportions in the biotin labelled fraction, although this family is not well defined and consists of multiple enzyme activities, in contrast to the CE5s that are only known to comprise acetyl xylan esterases. However, when the proteins that contained a CE14 domain were searched against the databases using BLAST they showed the top homologies, within the SwissProt database, to acetylases. The CE16s, also containing acetylases, were predominantly located in the biotin labelled fraction.

Although each CAZyme family had different tendencies to be located in either fraction, it was apparent that more enzymes involved in carbohydrate degradation were found in both the supernatant and biotin fractions, than in either the supernatant or biotin fractions alone (Figure 5-12). For example, wheat straw cultures harvested after four days of growth contained 228 identifiable CAZymes, and of these 151 (66.2 %) were present in the supernatant and the biotin labelled fraction, whilst just 48 (21 %) were present in the supernatant and 31 (13.6 %) in the biotin labelled fraction. Following the same four-day period of incubation, glucose-grown cultures contained 116 CAZymes, of which only 42 (36.2 %) were identified in both fractions (Figure 12), less than a third of the number in the wheat straw grown cultures.

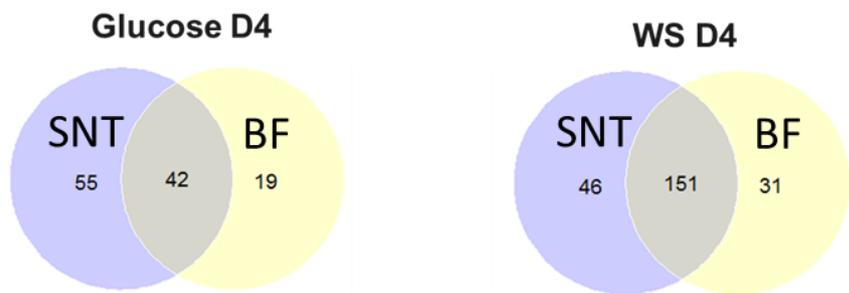


Figure 5-12 – Distribution of the CAZymes in the biotin-tagged and supernatant samples. Proteins from the fourth day of growth on both wheat straw and glucose-based media were detected in either the supernatant (SNT), biotin-tagged fraction (BF) or both. Samples grown on wheat straw shared the greatest overlap of enzymes.

5.3.4.1 Carbohydrate binding modules

CBMs were present in both the supernatant and the biotin-tagged fractions (Figure 5-13). Surprisingly, however, the majority of proteins with CBMs containing proteins were higher in the supernatant fraction, including proteins whose catalytic domain suggested that they were active against cellulose and hemicellulose. Only two proteins that contained a CBM appeared at a higher relative abundance in the biotin-tagged fraction, as compared to the supernatant. These were c3171_g1_i1_4, which also contained a LPMO domain (AA9), and c6768_g1_i2_2, which also contained a CBM52 domain – a family known to bind to β -1,3-glucan chains.

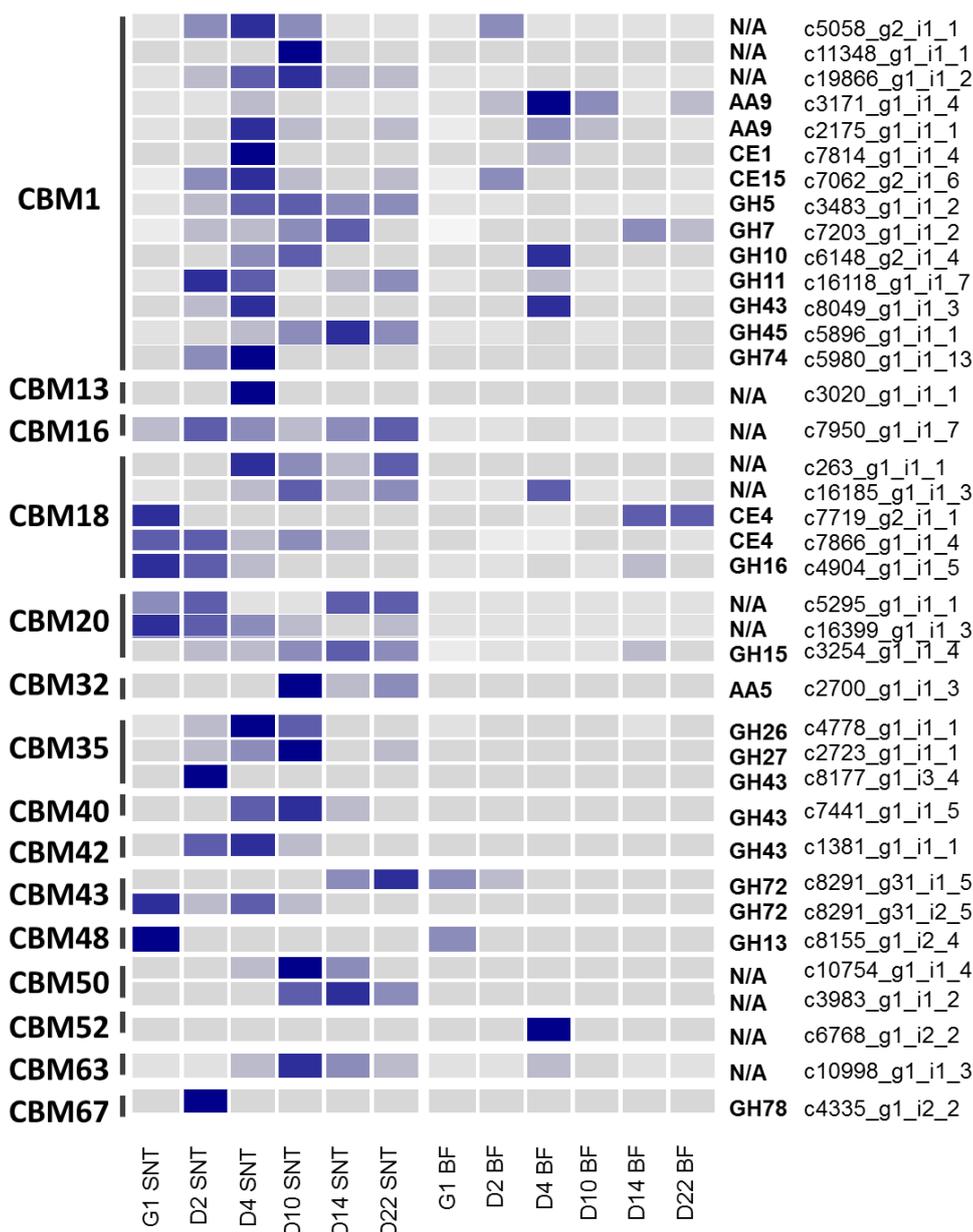


Figure 5-13 – Heatmap showing of contigs containing carbohydrate binding domains. Heatmaps were created using an average of the relative molar percentage of each contig and scaled across their rows to give the relative distribution of the protein in each sample.

5.3.4.2 Proteins localised only within the biotin fraction

Proteins that appear only in the biotin fraction may have been over-looked using previous sampling methods. *Graphium* sp. contained 918 such proteins, though this could be as a consequence of the limited depth of sequencing achievable by LC-MS/MS.

A total of 41 CAZymes were detected in the biotin-tagged fraction but absent from the supernatant, throughout the time course. The second most abundant protein that was biotin-

tagged but not present in the supernatant fraction was a LPMO (AA9). This protein, c19570_g1_i1_2, was highest in the latter fractions, and not present in the glucose grown cultures. This proteins presence in the transcriptome was discussed in Section 4.3.5.2.

Another protein, c8066_g1_i1_1, that appeared at a high relative molar percentage in the biotin labelled samples was a superoxide dismutase. This protein contains a recognisable signal peptide and showed the highest identity to copper-containing superoxide dismutase-like protein (gb|KZF26493.1|) from the fungus *Xylona heveae* with 43 % identity to 68 % of the sequence. Interestingly, this protein also showed a 42 % identity to a cell surface superoxide dismutase (gb|KUI66682.1|) from *Valsa mali*, a fungus within the same Sordariomycetes class as *Graphium* sp.

5.4 Discussion

The proteomic data presented here builds upon the enzyme activity screens described in Chapter 3 and RNA sequencing in Chapter 4, and provides additional depth to the analysis of the growth of *Graphium* sp. on wheat straw.

In order to perform a direct comparison to the transcriptome, proteins were collected from the supernatant of the same wheat straw cultures used for RNA collection in Chapter 4. These were analysed using LC-MS/MS to identify 804 distinct proteins, comparable to other studies on the degradation of lignocellulose by filamentous fungi (Ogunmolu et al., 2015).

By sampling the supernatant, rather than the protein contained within the fungal cells, samples were enriched for prospective lignocellulose degrading enzymes. Fifteen percent of the proteins identified (125 out of 804) could be annotated as either carbohydrate active or carbohydrate-binding, compared with less than five percent (769 out of 17,039) of the putative ORFs described in the previous chapter.

Nethertheless within samples collected from the supernatant the majority of proteins lack an observable signal peptide. Though this could suggest that some cell lysis has occurred, it could also be indicative of the limitations of annotating sequences derived from de novo- type experiments. As ORFs were selected to be the longest sequence that spanned between a start and stop codon, this may result in the correct start codon, and any potential signal peptide coding region, not being present in the at the 5' end of the sequence but instead laying downstream of the sequence rendering it undetectable with the HMMs used. Encouragingly, it was found that CAZymes, which were expected to be involved in intracellular processes, such as the glycosyltransferases, were present in low abundance in the secretome, suggesting that little cell lysis had occurred. Only three glycosyltransferases were identified in the supernatant, all of which were below a relative molar percentage of 0.05 %, in comparison to the 159 detected in the transcriptomics. Differences between the transcriptomic and proteomic datasets became more pronounced as the time course progressed, as correlations between the two datasets became weaker. Within the supernatant there was an increase in both the number and the abundance of CAZymes detected on the tenth day of growth, despite the transcription of these genes being significantly reduced, suggesting that these enzymes are highly stable within the supernatant. This is in accordance with results obtained in previous studies which have shown cellulases to be highly stable proteins (Juturu and Wu, 2014).

The number of distinct cellulase enzymes identified during this investigation was higher compared with other recently published studies. The proteome of *Graphium* sp. contained a total of 18 putative hydrolytic cellulases. In comparison, the coprophilous fungus *Podospora anserina* produced 13 cellulose deconstructing proteins when grown on crystalline cellulose or sugar beet pulp (Poidevin et al., 2014), and *Doratomyces stemonitis* secretomes contained eight cellulase related protein when grown on avicel containing-cellulase inducing media (Peterson et al., 2011). However, direct comparisons between plant cell-wall degrading fungi are difficult to achieve unless the growth conditions and experimental procedures have been standardised, since carbon source and timing have a major effect on the expression of lignocellulolytic enzymes.

Despite the diversity of the CAZyme-containing proteins produced by *Graphium* sp. being higher compared with other species, they appeared to account for a smaller proportion of the secretome. At its peak, 40 % of the supernatant of *Graphium* sp. could be attributed to proteins containing putative carbohydrate active domains. Remarkably however, some strains of the industrial fungus *T. reesei* can produce a secretome, of which 60 % consists of a single cellobiohydrolase, though this was only achieved after significant strain modification and optimisation.

Another *Trichoderma* species, *T. hazianum*, has been showed to have a secretome consisting of 71 % glycoside hydrolases, 17 % of which were GH3s, 11 % GH5s, 10 % GH7s and 6 % GH6 (Rocha et al., 2016). The secretome of *T. hazianum* was in marked contrast to *Graphium* sp. as the GH3s, instead of being the most abundant family as observed in *T. hazianum*, were one of the lowest represented glycoside hydrolases in the *Graphium* sp. proteome, and the GH5s, the second most abundant family in *T. hazianum*, represented less than one percent of the secretome on the fourth day of *Graphium* sp. wheat straw growth. Interestingly, despite the GH5 family contributing to most of the endoglucanase activity in ascomycetes, the GH45 family is present in similar abundances to the GH5 family, suggesting that they also play an important role in cellulose degradation in *Graphium* sp.. In the extended time course, the GH45s are also one of the few cellulase families that persist in the supernatant until the 22nd day of growth on wheat straw. This family was identified as being a family of interest in Chapter 4, due to its presence in numerous ORFs.

Another group of interest is the LPMOs. *Graphium* sp.'s transcriptome contained many putative AA9s that were highly transcribed and upregulated when grown on wheat straw (Section 4.3.5.2). However, the two most abundant LPMOs within the transcriptome were not

detected in the supernatant samples, and those that were detected within the secretome were only present in days two and four of the growth on wheat straw.

Delmas *et al* also noted an apparent lack of AA9s, at that time referred to as GH61s, within *A niger*'s secretome, despite there being a large representation in the transcriptome, and hypothesised that proteins may be bound to either the fungal cell wall or the insoluble wheat straw substrate within the culture (Delmas *et al.*, 2012). However when additional samples were analysed, looking at both the proteins bound to the insoluble substrates within the culture and the cultural supernatant, there was no significant increase in the amount of AA9s detected in the biotin-labelled fraction.

The secretome of *Graphium* sp. appeared to be an environment adapted for the oxidative attack of lignocellulose, despite low quantities of LPMOs being detected within the supernatant and biotin fractions. Multiple cellobiose dehydrogenases were identified, an enzyme that has been implicated in the degradation of polysaccharides through the provision of electrons to LPMOs (Phillips *et al.*, 2011, Zifcakova and Baldrian, 2012), and their presence correlated with high numbers of putative LPMOs in genomic libraries (Kracher *et al.*, 2016). However, this enzyme may have other auxiliary roles in lignocellulose degradation, including the production of hydrogen peroxide (Henriksson *et al.*, 2000).

Along with the cellobiose dehydrogenases, a large proportion of the supernatant proteome consists of enzymes that have the ability to generate hydrogen peroxide. In brown rot fungi, hydrogen peroxide is required to drive the production of hydroxyl radicals for the Fenton reaction (Martinez *et al.*, 2009), whilst lignin degradation in white rots also requires the generation of hydrogen peroxide, as it is in part coordinated via the peroxidase activities of the AA2 family (Bugg *et al.*, 2011a). However *Graphium* sp.'s belongs to neither of these physiological groupings and its proteome contained only two putative AA2 domains, both of which were not predicted to be secreted and were not detected in the secretome at high levels. There was, however, a peroxidase whose sequence did contain a recognisable signal peptide that was not recognised by hidden Markov models, but instead showed homology to a chloroperoxidase, an enzyme capable of catalysing the chlorination of organic compounds using hydrogen peroxide as an electron acceptor (Longoria *et al.*, 2008, Hofrichter *et al.*, 2010). This putative chloroperoxidase was highly abundant in the supernatant proteomics from the fourth day of growth on wheat straw, the time point previously characterised as being the peak of both cellulose and xylan degradation. A role in lignin degradation has been suggested for this enzyme as they have been previously demonstrated to chlorinate and cleave dimeric

models of the major non-phenolic lignin linkages and depolymerise synthetic guaiacyl lignin structures (Ortiz-Bermudez et al., 2003, Ortiz-Bermudez et al., 2007).

Chloroperoxidases can also function as catalases (Longoria et al., 2008), and this dual function has led to some contention over the primary function in a lignocellulose degrading environment, as it has been suggested that it may act as a detoxifying enzyme to aid with the amount of hydrogen peroxide produced in the system. The detoxification of hydrogen peroxide is an important function for an organism living in an environment adapted to degrade lignocellulose through oxidative means, and catalase activity has been shown to reduce the inactivation of cellulases and improve growth on more complex substrates (Bourdais et al., 2012, Scott et al., 2016). *Graphium* sp., does, however, produce a distinct catalase presumably for this function that is also present in the proteome at high abundances, albeit at different time points.

5.4.1 A comparison of the biotin labelled and supernatant fractions

Biotin-tagging of proteins was performed to enrich samples for those that formed attachments to the insoluble components of the culture, in this case either the wheat straw or the fungal cell wall itself.

This technique appeared effective in enriching for proteins containing transmembrane domains at the latter stages of the time course but CAZyme domains were present within the biotin samples at a lower proportion than that of the supernatant. This finding was unexpected, as wheat straw shake flasks inoculated with compost did show an enrichment for both carbohydrate binding and carbohydrate active proteins in their biotin-labelled fractions, in comparison to the supernatant fractions (Alessi, A., personal communication). Though direct comparisons between the concentrations of two proteins, in two different samples, cannot be performed on this data, as only the relative molar abundance of the identifiable proteins can be assessed. The diversity of membrane-bound and -associated proteins within the biotin samples may, therefore, be causing the CAZymes to appear under represented within these fractions.

Consistent with enzymes involved in lignocellulose deconstruction that absorb and deabsorb from their substrates, CAZymes involved in carbohydrate breakdown appear to be in equilibrium between the liquid and solid phases (Payne et al., 2015). However, different enzyme families showed different tendencies to be associated with the insoluble components of the culture.

The β -glucosidases, although present in both sampling techniques, are present at a higher relative molarity in the biotin-labelled samples. β -glucosidases from *T. reesei* have been demonstrated to be associated with the fungal cell wall. This finding, suggests that like those of *T. reesei*, enzymatic cocktails produced from this species of fungus may benefit from β -glucosidases additions (Ma et al., 2011). The adaptation also may go some way to explaining why *Graphium* sp. thrived in the mixed-community liquid shake flask cultures, as the close proximity of the glucose release to the fungal cell aids uptake, and consequently less is lost in a competitive environment.

Cellobiohydrolases and xylanases were present within both fractions, but have a higher relative molarity in the supernatant. However, after twenty-two days these proteins were undetected in the supernatant, but present at higher relative molarities than previously seen in the biotin labelled fraction. This could be an effect of the substrate acting to stabilise the enzyme and there has been some suggestion of this occurring in enzymatic cocktails produced from *T. reesei* in relevant literature (Eriksson et al., 2002). The endoglucanases containing families were very rarely detected in the biotin labelled fractions, but as this is the first single organism whose bound proteome has been studied using this method it is not possible to draw direct comparisons to other known lignocellulose degrading species.

Arguably the most interesting proteins are those found only in the biotin labelled samples, as it is these that may be overlooked when using previous sampling techniques. One AA9 was present at a high abundance in the biotin fraction and was not detected in the supernatant fraction. Another protein that was only found present in the biotin-tagged libraries, that was neither annotated as a transmembrane or carbohydrate active protein, was a superoxide dismutase. Superoxide dismutases (SODs) are ubiquitous in nature and are important antioxidants involved in protecting cells against oxidative stress, through the simultaneous reduction and oxidation of superoxide radicals into either molecular oxygen or hydrogen peroxide. They have, however, recently been implicated in the degradation of lignin in a bacterial species, through a proposed mechanism whereby highly reactive hydroxyl radicals are produced which can oxidise lignin (Rashid et al., 2015). An alternative function of this superoxide dismutase could be its involvement in protection of the fungal cell against the oxidative environment produced during the degradation of lignocellulose, and there have been reports of fungal cells protecting themselves from immune cell mediated oxidative attacks in mammalian systems via cell wall superoxide dismutases (Gleason et al., 2014). If this were the

case in *Graphium* sp. it would add further weight to the hypothesis that this is a species highly adapted to breakdown lignocellulose in an oxidative environment.

6 Investigation of single lignocellulose active proteins

6.1 Introduction

In Chapter 4 and Chapter 5, proteins likely to be involved in lignocellulose degradation were identified during the growth of *Graphium* sp. on wheat straw. Using transcriptomic and proteomic techniques putative proteins with homologies to known lignocellulose degrading enzymes were identified for their potential to be involved in lignocellulose metabolism in *Graphium* sp., along with multiple proteins with no known function. In this chapter, a selection of these proteins was chosen for further investigation.

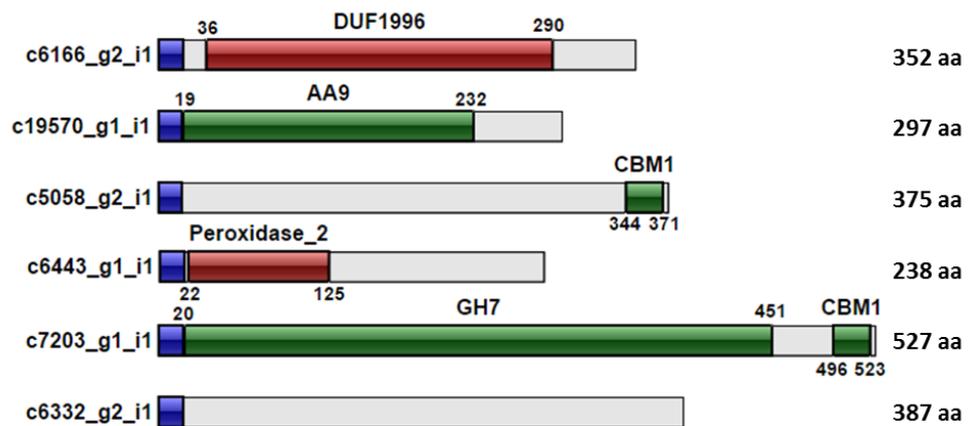


Figure 6-1 Contigs of interest for heterologous expression. Pfam annotations are marked as red domains, CAZyme domains are annotated in green, and signal peptides are indicated in blue.

The molecular architectures of the proteins of interest are shown in Figure 6-1. Two targets were annotated with confidence to contain CAZymes domains corresponding to an AA9 and a GH7, and these were contained on ORFs within the contigs c19570_g1_i1 and c7203_g1_i1. These were selected as being of interest due to their prevalence in the wheat straw degrading cultures of *Graphium* sp. The GH7, c7203_g1_i1, showed the highest abundance of all proteins detected when grown on wheat straw, and the AA9 contig had a 600 fold increase in RPKM between the glucose and wheat straw growth conditions (Section 4.3.5.2). The LPMO was, however, not the most abundant of the AA9 families detected in the proteomics. It did have, however, one of the least conserved sequences compared to those deposited into the nr-NCBI database, sharing just 40 % of its amino acids with its closest match, and one of the most interesting proteomic profiles, as it was present in the latter stages of growth, peaking in the

second week of incubation on wheat straw, and only ever was detected in the biotin-tagged proteomics (Section 5.3.3).

A contig (c6443_g1_i1), containing a predicted chloroperoxidase, was also selected as a protein of interest due to reports of the involvement of these proteins in the cleavage of lignin linkages (Longoria et al., 2008). This, combined with *Graphium* sp. ability to grow in environments with high amounts of lignin derivatives and the prevalence of this enzyme in the supernatant when grown on lignocellulose, made this enzyme a promising target for the discovery of a potential lignin-modifying enzyme.

Alongside annotated CAZymes and putative proteins with Pfam annotations, sequences were chosen for further study that contained domains with no homology to known lignocellulolytic enzymes. These included c5058_g2_i1, a contig which contained a c-terminus CBM1 but no identifiable catalytic domain, and c6166_g2_i1, a contig with a domain of unknown function 1996 (DUF1996). Another putative protein selected was c6332_g2_i1 contained no identifiable catalytic domain according to Pfam's HMMs but appeared to show homology to a dioxygenase in the nr-NCBI database.

As identifying enzymatic activities is best performed on pure proteins, away from interfering and masking factors, proteins must first be isolated before being characterised. This can either be achieved in their native system or by recombinantly expressing these proteins in a heterologous host.

Purification of proteins from cultures of the native host will ensure that the correct post-translational modifications are present, while a heterologous system allows finer control over the protein expression, as regulation can be controlled with the use of specific inducers. A genetic engineering based method also allows the protein itself to be modified, with signal peptides to direct the protein to different sub-cellular locations and specific tags to aid purification. Not all proteins, however, can be expressed in a non-native organism as codon bias or posttranslational modifications can present problems. Therefore, the heterologous system that the protein is expressed in can have an effect on the success of the expression.

Bacterial expression systems struggle to replicate the post-translational modifications of eukaryotic proteins. Generally, the less evolutionary distance between the native and the heterologous host, the better chance the protein being expressed in its correct conformation. Expression in fungi would, therefore, be advantageous if the protein was predicted to contain

complex post-translational modifications. These systems, however, are not as well established, suffer from low transformation efficiencies and are time intensive.

A. niger has been developed as a host for heterologous expression (Vanhartingsveldt et al., 1987, Gouka et al., 1997, Punt et al., 2002). Fungal cells, deficient in the uridine synthesis gene, orotidine-5-phosphate decarboxylase (PryG), are protoplasted before polyethyl glycol (PEG) mediated transformations are performed to induce the uptake of a plasmid containing the gene of interest between regions homologous to *A.niger* and a functional PryG gene. It is these homologous regions that allow for integration of the target gene into the genome, and the presence of the PryG, allows only successful transformants to grow on uridine deficient media.

Other expression systems used in this investigation included targeting proteins to the periplasm of *E. coli*. The targeting of the protein to the periplasm has multiple benefits; including aiding the formation of disulphide bonds due to its oxidising environment and the reducing the number of proteases that the recombinant protein is exposed to. The periplasmic signalling protein located at the N-terminus of the peptide is also cleaved off during the exportation of the protein to the periplasm, and this action ensures that protein purified from the periplasm will contain the N-terminus of the native protein. Purification of proteins without affinity-tags is also simpler in this environment as the numbers of proteins that need to be excluded are reduced.

6.2 Methods

6.2.1 Plasmid preparation

cDNA was synthesised from DNase-treated RNA (Section 2.3.2) harvested from four-day-old cultures of *Graphium* sp. grown on wheat straw containing optimised media. From this, the five targets for heterologous expression were amplified from the cDNA using primers 20 – 50 bp upstream and downstream of the predicted start and stop codons (Table 6-1). The PCR was performed on cDNA reaction mixtures that had been diluted 1 in 20 with H₂O due to the amount of inhibitors present in the reaction mixture.

Table 6-1 - Primers for the amplification of target genes from cDNA.

Target	Primers	Expected Size
CBM	Fw - ACACGAGCAGCGGAGATTAG	1,632 bp
	Rv - GCTTCGGATTGAAGAAAGGA	
GH7	Fw - GAAATAAGCGCTTGATGATGG	2,175 bp
	Rv - AAAGAAAGGCGTGCAATGTT	
POX	Fw - GGAATTTCCGTTGTCTCGTC	1,274 bp
	Rv - TAGAAATGCGCCTCAACGTC	
AA9	Fw - TATAAAAATGGGCGCAGTCC	1,343 bp
	Rv - CTTTACAAACCCACGCGTCT	
DUF1996	Fw - TGGATCAAATGGATCAATGTAAAG	1,370 bp
	Rv - TGGGTTGTTTAGACCTTGACG	
c6332_g2_i1	Fw - CCAGGCCAATCTGTCTTGTT	1,493 bp
	Rv - AAAAGGCGATGCTTCTCTGA	

The resulting PCR reaction was run on an agarose gel to confirm a single band of the correct size was obtained, and these were purified from the reaction mixture as described in Section 2.3.3 and 2.3.5 before being sent for Sanger sequencing (Section 2.3.8). Once it was confirmed that there were no mismatches in the transcriptomic sequencing, further primers were designed to facilitate the infusion cloning into the vector of choice for either bacterial or fungal expression (Table 6-2).

Forward primers for the amplification of the gene of interest were designed to remove the native signal peptide of the protein, as in both expression systems used this will be replaced with a native signalling peptide and while reverse primers were also used to add a purification tag to the protein, if required. Histidine tags were appended to the both the CBM and GH7 proteins for affinity purification. Streptavidin tags were added to the chloroperoxidase (POX) and LPMO as these are predicted to contain metal co-factors and may be affected by

purification through a nickel column as required with the use of histidine tags. No purification tags were used when periplasmic expression was targeted with the Pet26b vector (Novagen).

PCR was used to amplify the gene from the previous reaction, using these primers, and once visualisation on an agarose gel had again confirmed that only a single band had been amplified this was cloned into the linearized vector of choice using InFusion (Section 2.3.9). Cloning enhancer was used with the InFusion protocol, to negate the need to purify the amplicon from the PCR reaction mixture and the linearized vector.

Table 6-2 - Primers used for amplification of targets for fungal expression with overhangs for Infusion cloning into pIGF-pyrG. Primers were designed to incorporate overhangs for InFusion cloning, molecular tags for purification and the KEX2 cleavage site. Black text denotes the regions homologous to the linearized vector, blue the KEX2 cleavage site, green the purification tag of choice and red text is to facilitate the amplification of the target gene.

Target	Primers	Expected Size	
CBM (pIGF-pyrG)	Fw -	CTGCTAGCAAGTCTAGAAAGCGGCGGCGGTGGCT TGGACCTTTCCAAGTCGTC	1,632 bp
	Rv -	GTCGCGGTCGACGTTAACGTGATGATGATGATG ATGGAGGCACTGGGAATACCAGTCA	
GH7 (pIGF-pyrG)	Fw -	CTGCTAGCAAGTCTAGAAAGCGGCGGCGGTGGCC AGCAGGCTTGCTCCTTG	2,175 bp
	Rv -	GTCGCGGTCGACGTTAACGTGATGATGATGATG ATGCAAGCACTGGCTGTACCAGTC	
POX (pIGF-pyrG)	Fw -	CTGCTAGCAAGTCTAGAAAGCGGCGGCGGTGGCC TCGCGGCCGTCGTGACTA	1,274 bp
	Rv -	GTCGCGGTCGACGTTAACTCACTTCTTTTCGAAC TGCGGGTGGCTCCACGCCGCGGCACGGAC	
AA9 (pIGF-pyrG)	Fw -	CTGCTAGCAAGTCTAGAAAGCGGCGGCGGTGGCC ACGGCTACGTGTATCGCAT	1,343 bp
	Rv -	GTCGCGGTCGACGTTAACTCACTTCTTTTCGAAC TGCGGGTGGCTCCAGCCTCCAATGGAGTCA	
DUF1996 (pIGF-pyrG)	Fw -	CTGCTAGCAAGTCTAGAAAGCGGCGGCGGTGGCC AGGCCATGCTTCGGTTT	1,370 bp
	Rv -	GTCGCGGTCGACGTTAACGTGATGATGATGATG ATGCGGTTCCATCGTCAAGGTC	
AA9 (PET26b)	Fw -	CAGCCGGCGATGGCCACGGCTACGTGTATCGC AT	Bp
	Rv -	GTGGTGGTGCTCGAGTTAGCCTCCAATGGAGTC	
c6332_g2_i1 (PET26b)	Fw -	CAGCCGGCGATGGCCACCCTCACGGCGC	Bp
	Rv -	GTGGTGGTGCTCGAGTTAGGGATTCTAGGAGC ACC	

The vector was then transformed into stellar ultra-competent cells using a heat shock method (Section 2.3.10), and successful transformants were selected for by growth on nutrient agar

plates that had been supplemented with the appropriate antibiotic for the plasmid of choice. The insertion was then confirmed by colony PCR (Section 2.3.3) and Sanger sequencing (Section 2.3.8), using T7 universal primers to amplify the up and downstream of the cloning site (Table 6-3), or the primer used for cloning into the pIGF-pyrG vector. A single colony which gave a 100 % identity to the target gene was grown-up in LB for plasmid extraction (Section 2.3.11).

6.2.2 Protoplast formation

Protoplasts were prepared from mycelia grown in *Aspergillus* minimal media supplemented with uridine. Cultures were inoculated into 400 mL of media with 1×10^6 fresh spores/mL and were grown in 2L shake flasks for 18 hours, shaking at 180 rpm in 30 °C. Mycelia were harvested through a double layer of pre-autoclaved miracloth (Calbiochem), washed with 50 mL sterile ice-cold Mycelium Wash buffer (0.6 M $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$), blotted with paper towels and weighed in a sterile, 50 mL Falcon tube. Biomass, resuspend in cold MM buffer (1.2 MgSO_4 , 20mM MES, at pH 5.8) with 5 mL of buffer for every gramme of mycelia collected, was then treated with a lysing enzyme, for protoplasting, in the presence of 100 mg of BSA for every gramme of mycelium harvested. This was then incubated at 30 °C shaking initially for 30 minutes at 80 rpm, and then at 50 rpm. After two hours of incubation, a 10 μL sample was transferred to a glass slide and visualised under a microscope to check on the progress of protoplast formation. Once protoplasts had become visible and detached from the mycelium mass, they were purified through sterile polyallomer wool packed into two sterile 20 mL syringe barrels, with one tightly packed, and the other loosely. The loosely packed syringe barrel was required to filter out the larger fragments of mycelium, and the tightly packed to ensure that a minimal amount of mycelium fragments contaminated the collected protoplasts. The equipment was set up as in Figure 6-2. The protoplast mixture was then poured slowly into the uppermost syringe barrel and allowed to pass through into the collection tube with gravity. The 50 mL falcon tube was replaced after 15 mL of the mixture was collected, and 30 mL of ice cold NM buffer (1M NaCl, 20 mM MES at pH 5.8) was added to each tube. These were then gently mixed by inversion and centrifuged at 2,500 rpm for 10 minutes at 4 °C, and the supernatant was poured off and discarded. The remaining pellet was then resuspended in 1 mL cold STC (1.2 M sorbitol, 10 mM Tris base, 50 mM CaCl_2 at pH 7.5). This was done with care and gentle agitation as vigorous pipetting is likely to lyse the protoplasts. This protoplast mixture was then spun again at 2000 rpm for 10 minutes at 4 °C, supernatant removed and

resuspended in 1 mL of ice cold STC. Protoplasts were counted using a haemocytometer (Section 2.2.6) and diluted to a concentration of 5×10^7 protoplasts per mL with ice cold STC.

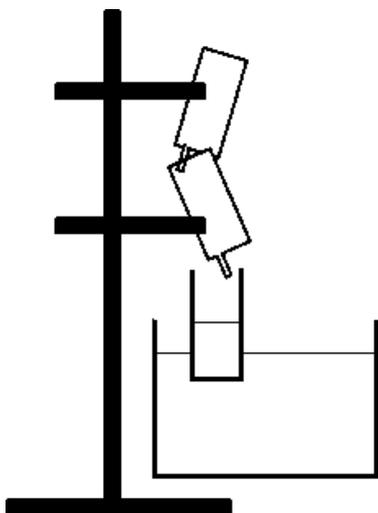


Figure 6-2 Set up of equipment for the fungal transformation. Protoplasts were separated from the mycelium mass of the fungi as above. Two syringe barrels were packed with polyallomer wool, the uppermost more loosely, and the protoplast mixture allowed to pass through with gravity.

6.2.3 Fungal transformation

Protoplasts were transformed on the day of creation. For each transformation, 3 μL of 1 $\mu\text{g}/\mu\text{L}$ vector was pipetted into a 15 mL falcon tube, and 100 μL of the protoplast suspension was added to this, mixed gently and incubated at room temperature for 25 minutes. Cells were then gently mixed again, and 200 μL of PEG (60 % PEG 5,000, 10 mM Tris base, 50 mM CaCl_2 at pH 7.5) was added drop-by-drop and continuously mixed into the protoplasts. A further 1.1 mL of PEG was added gradually with mixing, and tubes were filled to 10 mL with ice cold STC. The tubes were then centrifuged at 2000 rpm for 10 minutes at 4 °C, and the supernatant was discarded. Cells were resuspended in 300 μL STC, and 100 μL aliquots were plated onto sorbitol-containing minimal agar plates (10 mL salt solution -, 6 g sodium nitrate, 10 g glucose, 218 g sorbitol per litre at pH 6.5). Plates were incubated at 30 °C for three to five days until colonies were present. Any colonies were transferred to fresh minimal media plates and put through three rounds of selection.

6.2.4 Bacterial transformation

After plasmid extraction from the transform Steller cells the construct concentration was measured with the Nanodrop, and a suitable volume was transformed into Novagen's Rosetta™ 2(DE3) cells using the heat shock method (Section 2.3.10). These cells were chosen

for the heterologous expression of *Graphium* sp. proteins as they contain a plasmid that supplies tRNA for the rare eukaryotic codons AGA, AGG, AUA, CUA, GGA, CCC, and CGG. Again, the successful transformation was confirmed by PCR.

Table 6-3 -Primers for the linearization of selected vectors

Target	Primers	Expected Size
Pet 26b (linearisation)	Fw - GGCCATCGCCGGC	5,360 bp
	Rv - CTCGAGCACCACCACCAC	
pIGF-pyrG (linearisation)	Fw - CACTACGACGGCTACCCCCAC	11,377 bp
	Rv - CATCCCCATCCTTTAACTATAGCG	
T7 (conformation of insert)	Fw - TAATACGACTCACTATAGGG	N/A
	Rv - GCTAGTTATTGCTCAGCGG	

6.2.5 Expression trials

The ability of the heterologous host to produce the protein of interest as well as conditions that enable expression were tested. Successful transformants of Rosetta 2 cells were grown overnight in 5 mL LB starter cultures in 50 mL Sardest tubes at 37 °C, 180 rpm. After 16 hours, 5 µL of the starter culture was then used to inoculate an additional ten 5 mL LB cultures. These were grown to an OD of 0.8 at 600 nm, after which a range of Isopropyl β-D-1-thiogalactopyranoside (IPTG) concentrations (from 0.1 to 2.0 mM) was added to induce expression. Each IPTG concentration was further incubated at either 18 or 30 °C for 16 hours, and the periplasmic fraction was extracted using the protocol described in (Section 2.3.11). Proteins viewed on SDS-PAGE gels, to determine the concentration of IPTG and the temperature at which expression was optimal.

6.2.6 Heterologous protein expression in *E. coli*

Once the optimal condition in which the highest concentration of protein was expressed was confirmed, the cultures were scaled up. Transformants of the optimal cell type were grown overnight in 5 mL LB starter cultures, as before, in 50 mL Sardest tubes at 37 °C, shaking at 180 rpm. From these, 2 L baffled flasks containing 500 mL of LB, with the appropriate antibiotics, were inoculated with 1.25 mL of the starter culture and grown to an OD of 0.8 at 600 nm. IPTG was then added in the same proportion that gave the best expression results in the trial (0.2 mM), and the cells were grown for a further 16 hours and the periplasmic fraction harvested as before (Section 2.3.11).

6.2.7 Purification of heterologously expressed LPMO

The periplasmic extraction was concentrated to 2 ml and loaded onto a HiLoadtm 16/60 superdex 75 gel filtration chromatography column (Section 2.4.6). Resultant protein containing fractions from this partial purification were collected and visualised on an SDS-PAGE gel (Section 2.4.2). Fractions that appeared to contain the protein of interest were combined and further purified using a DEAE anion exchange chromatography column (Section 2.4.5).

Once a fraction containing a single band of the appropriately sized protein had been obtained, the identity of the band was confirmed by mass spectrometry.

6.2.8 Purification of the GH7 cellulase

The supernatant of *Graphium* sp. growing on wheat straw could not be easily concentrated, which hampered purification. Much of dark pigment that the supernatant contained was concentrated along with the proteins, making the visualisation of the protein challenging and the mixture unsuitable for purification columns. Therefore, as proteomic results described in Chapter 5 showed significant interaction between the biomass and the cellulases, it was decided to purify the protein from the insoluble material in the cultures.

Biomass collected from the shake flasks, inoculated with *Graphium* sp., and grown for four days on wheat straw, in the optimised media, was harvested and separated from the liquid phase by filtering through miracloth. To assess the most suitable method of disrupting the interaction between the biomass and protein, a small amount of the collected biomass was fractionated into six 50 mL falcon tubes in one gramme aliquots. These were incubated at 4 °C for one hour with various solutions (Table 6-4).

Table 6-4 – Solutions for the disruption of affinities between the protein and insoluble fraction of the wheat straw grown *Graphium* sp. cultures.

	Solution
1	50 mM Tris 8.5
2	2M NaCl
3	2M AmSO ₄
4	0.01 % SDS
5	0.01 % Tween 20
6	50 mM Acetic Acid

The solution that appeared to release the most soluble and active protein, from the wheat straw and fungal biomass, when tested against CMC plates was 0.01 % tween. This was, therefore, used in a scaled up experiment, in which the remaining biomass was incubated with approximately 200 mL 0.01 % tween, again for one hour at 4 °C. The biomass was filtered

through miracloth and the supernatant collected, concentrated to 2 mL (Section 2.4.3) and loaded onto a gel filtration column (Section 2.4.6). Aliquots were collected throughout this run and regions that corresponded to UV peaks were examined by gel electrophoresis (Section 2.4.2).

Fractions from the gel filtration column that contained the protein of interest were collected and combined. These were then loaded on to a DEAE anion exchange column for further purification. The resulting fractions were again visualised after electrophoresis, and a single clear band of the appropriate was cut from the gel and sent for peptide fingerprinting. Fractions containing just this protein were then collated, concentrated and buffer exchanged into 50 mM sodium phosphate by centrifugation (Section 2.4.3).

6.2.9 Activity assays

Enzyme activity assays were performed to determine the role of the target proteins in a lignocellulose degrading environment. The production of reducing sugars during the degradation of various substrates was measured using the lever assay described in Section 2.6.2, and, unless otherwise, stated used a pH of 6.8 and at 30 °C, with 0.1 mg/mL enzyme concentrations. Where appropriate the degradation products of these reactions were analysed using HPLC. The same conditions used in the lever assays were used to investigate the protein activities with *p*Np-linked substrates (Section 2.6.4). Glycospot plates were also used to identify substrates of the LPMO (Section 2.6.3).

6.2.9.1 Congo Red assay

To test for potential lignocellulolytic activity the congo red assays were performed, whereby the discolouration of the dye corresponds to the ability to break down the polyphenolic structure of the azo-dye. Solutions of 0.2% (w/v) congo red were prepared, and 10 µL of an unknown concentration of protein of interest was incubated in 400 µL reaction volumes in 2 mL Eppendorf tubes. This was left incubating at 30 °C, shaking at 1000 rpm for 24 hours.

6.2.9.2 Tannic acid

Tannic acid was used to assay for the ability to oxidise polyphenolic structures. Like the congo red assays, this was performed in 2 mL Eppendorf tubes with 10 µL of an unknown concentration of enzyme and left shaking at 1000 rpm for 24 hours.

6.3 Results

6.3.1 Fungal cloning

The preparation of the pIGF-pyrG plasmid for fungal cloning was successful for all five target proteins chosen for analysis (Section 6.2.1). These targets were then transformed into stellar cells and colony PCR (Section 2.3.3) were performed to confirm the presence of the gene of interest (Figure 6-3). Successful colonies from three of the targets (CBM, GH7 and POX) were then processed for fungal transformations (Section 6.2.3).

Adequate numbers of protoplasts were formed for the uptake of the plasmid DNA (Section 6.2.2). However, fungal transformations resulted in no positive colonies despite being repeated multiple times. It was therefore decided to concentrate on purification of proteins from the native organism and bacterial expression.

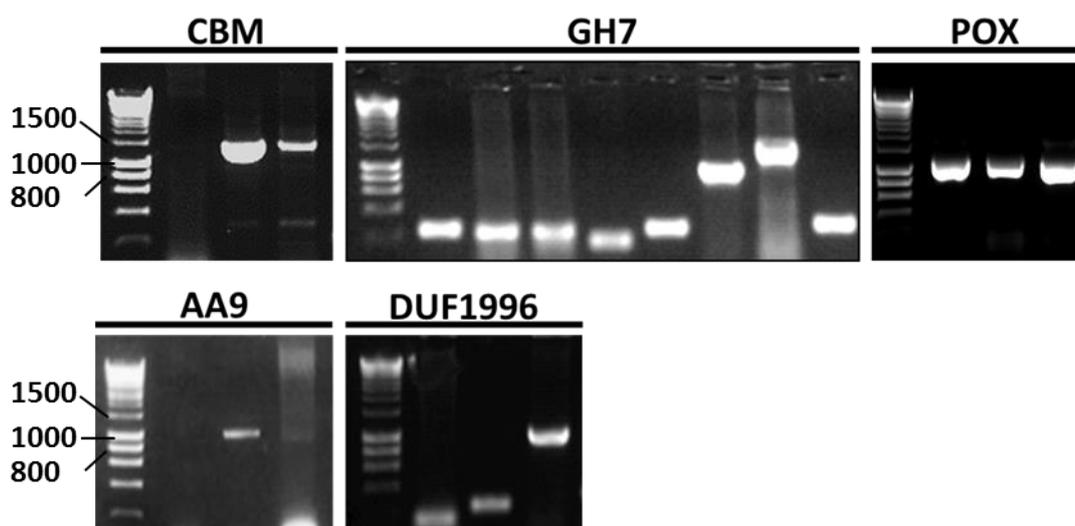


Figure 6-3- Colony PCR of Stellar cells transformed with the gene of interest in the PryG vector. Successful transformations were obtained, and the presence of the gene of interest was confirmed by colony PCR.

6.3.2 Cloning, Expression and Purification of an LPMO

The putative LPMO was, however, successfully cloned into the Pet26b vector and transformed into Rosetta™ 2(DE3) competent cells. Expression trials (Section 6.2.5) confirmed protein production in all IPTG induced conditions (Figure 6-4) and protein expression was subsequently scaled up, using the determined optimal conditions for LPMO production, (18 °C and concentrations of 1 mM of IPTG).

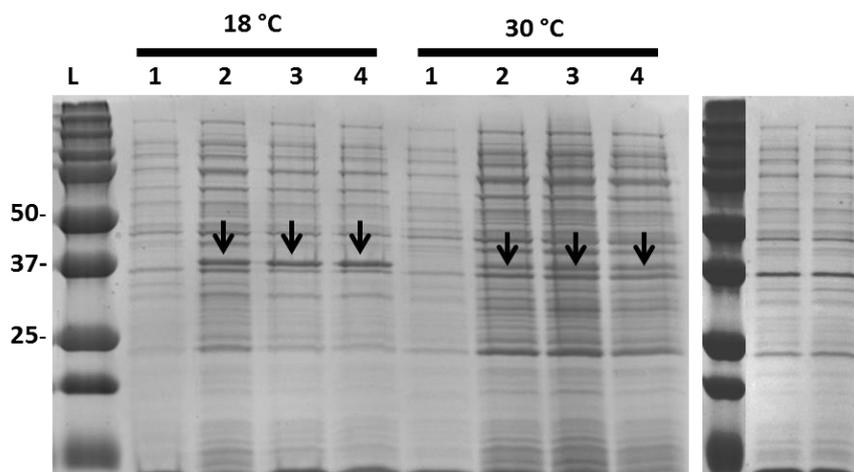


Figure 6-4 – SDS-PAGE of periplasmic proteins harvested from LPMO expressing Rosetta cells. Rosetta 2 cells, grown to an OD600 of 0.6, were inoculated with varying concentrations of IPTG and induced for 16 hours at either 18 or 30 °C. The concentrations of IPTG used for the induction indicated by the lane labels whereby 1 - was no IPTG, 2 - 0.5 mM, 3 - 1 mM and 4 - 2 mM.

Six 2 L flasks, containing 500 mL of LB were used for the large-scale expression, and the protein was then purified using anion exchange chromatography (Figure 6-5) as an initial step. Fractions containing protein corresponding to the size of the LPMO were pooled, concentrated to 1 mL using centrifugation (Section 2.4.3) and loaded onto a superdex gel filtration column (Figure 6-5C-D). The identity of the purified protein was confirmed by mass spectrometry from a band exercised from an SDS-PAGE (Section 2.4.2). The concentration of the protein was then calculated by measuring the absorbance at 280 nm with a Nanodrop (Section 2.3.7), before being incubated with 1400 μ M of copper sulphate for two hours, and buffer exchanged into 25 mM ammonium acetate (Section 2.4.4).

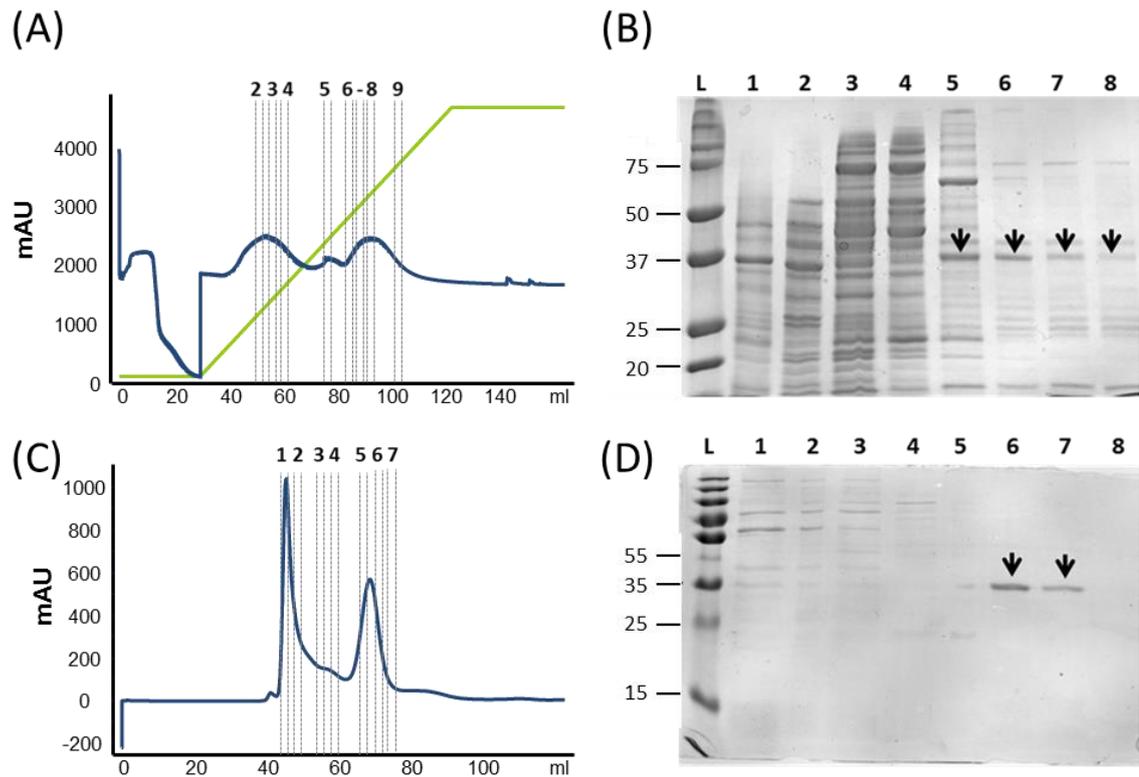


Figure 6-5 - Purification of a periplasmically expressed LPMO. Periplasmic extracts harvested from transformed Rosetta E.coli cells were purified using anion exchange (A). Resulting protein containing fractions (2-9) were assessed for the presence of the LPMO alongside the total (B), and fractions containing the protein were combined and further purified using a gel filtration column (C), protein containing fractions were again assessed for the presence of the protein (D) by SDS-PAGE.

6.3.3 Enzymatic activity of the lytic polysaccharide monooxygenase

The substrate that the LPMO was active on was determined after incubation on multiple polysaccharides. The purified protein was incubated at concentrations of 25 ng/mL along with 4 mM ascorbic acid to act as an electron donor, on chitin, PASC and ground wheat straw. Activity on chitin was first apparent from a visible yellowing of the substrate, observable after 16 hours of incubation that corresponded to an increased detection of reducing sugars (Figure 6-6). However, when chitin was incubated with 25 ng/mL of a commercial chitinase mix, a decrease of reducing sugars released was observed when the LPMO was present.

Activity was also observed when PASC and ground wheat straw were used as substrates, though this was not as marked as the activity on chitin. When a cellobiohydrolase (GH7), from *T. reesei*, was co-incubated with the LPMO, PASC was the only substrate that showed an increase in the production of reducing sugars, compared to GH7 only control. However, when

supernatant, harvested from the fourth day of *Graphium* sp. growth on wheat straw was co-incubated with the LPMO, the reaction mixtures containing wheat straw showed the greatest release of reducing sugars.

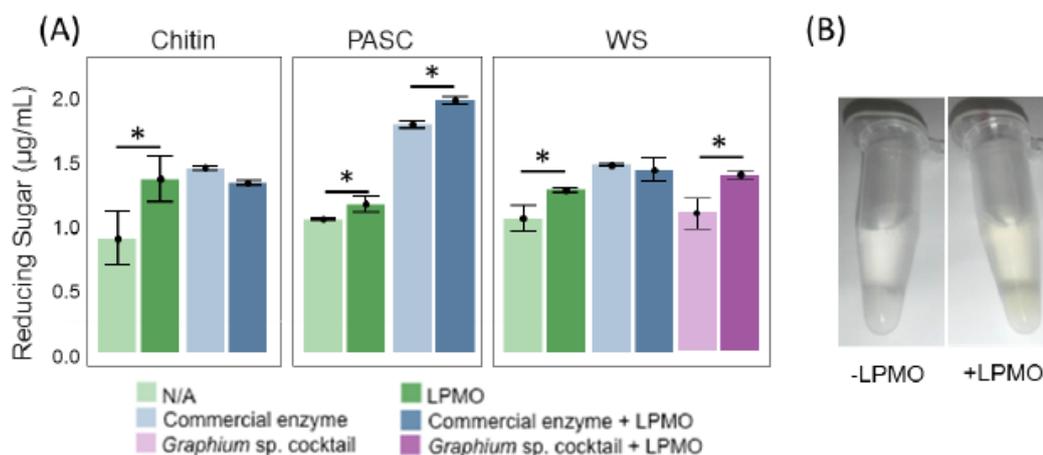


Figure 6-6 – LPMO activity on chitin, PASC and wheat straw. (A) Reducing sugar release was calculated after the 25 ng/mL of enzyme was incubated on either 2 % chitin, PASC and wheat straw in 25 mM pH 6 ammonium acetate, after 24 hours. Commercially available enzymes for the degradation of the appropriate substrate were also co-incubated with the LPMO at equal concentrations (B) A distinct colour change was observed after LPMO incubation on chitin. Asterisks mark significant difference as assessed by one-way ANOVAs to $P < 0.05$.

However, reducing sugar assays are not adequate for measuring LPMO activity as reactions where the chain cleavage occurs at the C1 carbon results in non-reducing ends. Glycospot plates were, therefore, used to confirm cellulase activity, where activity is detected through the release of bound dye, rather than measuring reducing sugar release (Kracun et al., 2015). Activity was assessed on hydroxyethylcellulose (2-HE-cellulose), pachymann, xylan and willow and significant differences between the control and the LPMO were seen on 2-HE-cellulose, 1,3- β -D-Glucan and xylan suggesting that this is an LPMO which is capable of degrading a range of substrates (Figure 6-7).

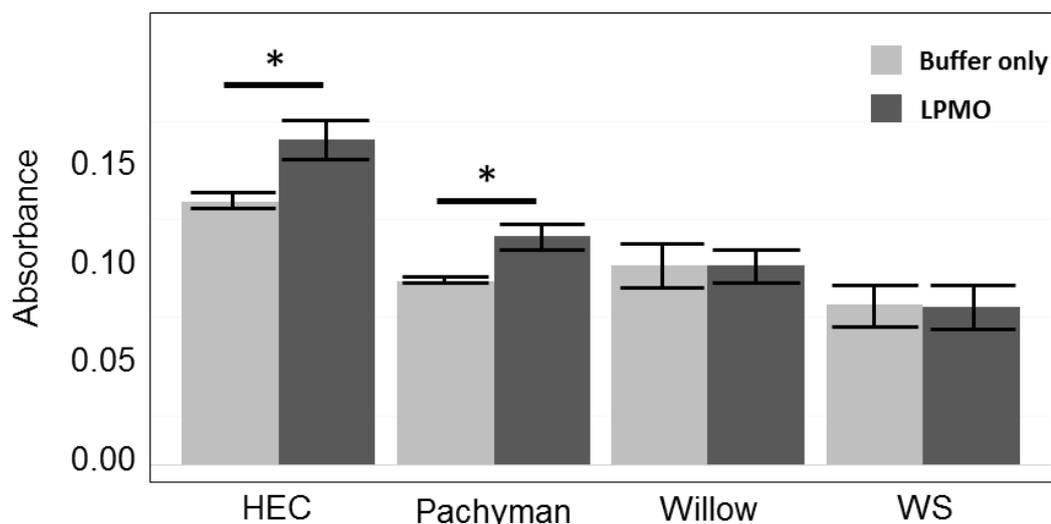


Figure 6-7 LPMO activity as assayed by Glycospot plates. Glycospot plates were used to assay for cellulase (HECase) and glucanase activities, and the ability to deconstruct complex lignocellulosic material (willow and wheat straw). Assays were performed in 25 mM ammonium acetate and 4 mM ascorbic acid at pH 6 with and without the addition of 25 ng/mL purified LPMO. These were incubated for 24 hours before the reaction was ceased and degradation visualised. Error bars represent +/- 1 SD of three biological replicates. Asterisks mark significant difference as assessed by one-way ANOVAs to $P < 0.05$.

To determine the reaction products of these reactions, HPAEC was performed on the supernatant. No clear products, however, were detected few differences could be observed between the LPMO positive and negative reaction conditions. Additional bands were present in the chitin samples treated with LPMO during analysis with the Dionex, though these were not identified as standards were lacking. The use of shrimp chitin, rather than, squid pen chitin further complicated the analysis as products with varying degrees of polymerisation could be present.

6.3.4 Cloning and expression of c6332_g2_i1

The putative dioxygenase was cloned in the Pet26b vector (Section 6.2.1) and transformed into Rosetta™ 2(DE3) cells (Section 6.2.4). Conditions in which the protein was produced were then assessed (Section 6.2.5). The results of these expression trials are shown in **Figure 6-8**. Once it was established that the protein was expressed in 18 °C at IPTG concentrations of 0.4 µg/L the expression was scaled up to 2 L flasks. Purification was attempted with the periplasm fraction harvested from these cultures using both anion exchange and gel filtration; however this was unsuccessful, and the protein could not be purified to a single band.

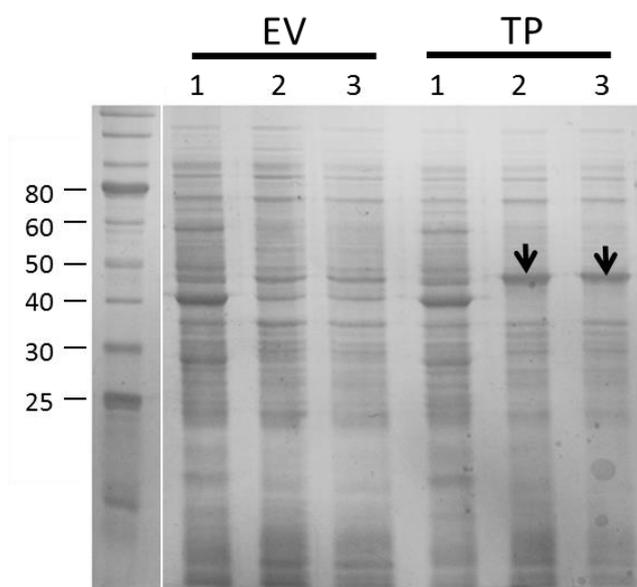


Figure 6-8 Expression trials of the putative dioxygenase. Rosetta 2 cells, transformed with the Pet 26b with (TP) and without (EV) the target protein were grown to an OD of 0.8, before being induced with varying concentrations of IPTG (1 - no IPTG, 2 - 0.5 mM, 3 - 1 mM). The periplasmic fraction of the cells was visualised on an SDS-PAGE gel, and a clear band was visible from the fraction with cells containing the gene of interest induced with IPTG.

6.3.5 Activity assays for c6332_g2_i1

Despite the protein not being purified to a single band, assays could be performed using the total periplasmic extraction of the cells transformed with the gene of interest, with the periplasmic fraction of the same cell type, transformed with the empty vector, acting as a negative control.

These total periplasmic fractions were incubated with wheat straw, PASC or xylan, and reducing sugar assays performed with and without commercial enzymes (**Figure 6-9**), and no significant differences were observed. The supernatant from these reactions was also analysed on a Dionex HPLC but no difference in peaks were apparent.

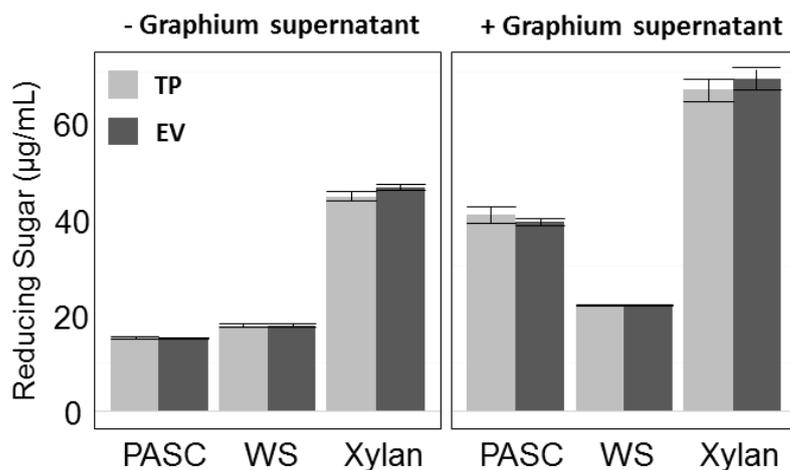


Figure 6-9 Sugar release after the periplasmic fraction of cells transformed with the unknown protein incubation with PASC, wheat straw and xylan. Reducing sugar assays were performed after 24 hour-long incubations with (TP) and without (EV) the target protein present on various substrates, and the *Graphium* sp. supernatant from the fourth day of growth on wheat straw. Error bars are the average of three replicates, no significant difference were found between samples that had been incubated with the periplasmic proteins of cells transformed with the target protein or the empty vector using a one-way ANOVA.

To assay for lignin-degrading abilities the periplasmic proteins were incubated with congo red and tannic acid. Although the assays did perform as expected, with the decolourisation of the congo red indicating the ability to deconstruct phenolic structures, and the darkening of the tannic acid indicating the ability to oxidise polyphenolics.

Clear differences were observed when the protein was incubated with congo red, as the red dye became darker and precipitated out of solution. Tannic acid also was not auto-oxidised in the presence of the enzyme, and the yellowing that occurs in the presence of oxygen did not occur when the target protein was present (Figure 6-12).

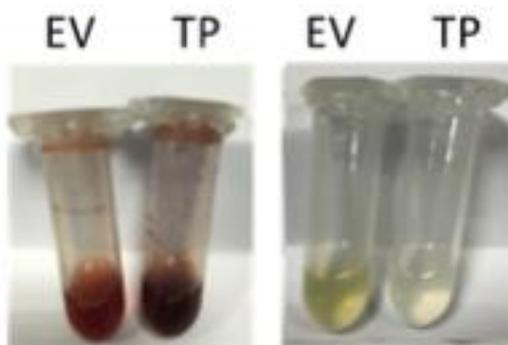


Figure 6-10 Tannic acid and Congo Red visual changes in the presence of the unknown protein. After 24 hours of incubation in 0.1 % congo red and 0.1 % tannic acid visual differences between the periplasmic fraction extracted from the empty vector control (EV) and the cells transformed with the target protein (TP) was apparent.

6.3.6 Purification of a GH7

Since fungal expression proved problematic, and expression in bacterial systems often do not adequately replicate post-translational modifications such as those present of GHs, purification of cellulases from the native organism grown on wheat straw was performed (Figure 6-11). The purification of a single cellulase was achieved, with assistance of Andre Godoy (University Sao Paulo, Brazil), as described in Section 6.3.6 from proteins unbound from the insoluble fraction using 0.01 % Tween 30. The identity of this protein was confirmed as being, c7203_g1_i1, the GH7 containing protein that was targeted for fungal expression.

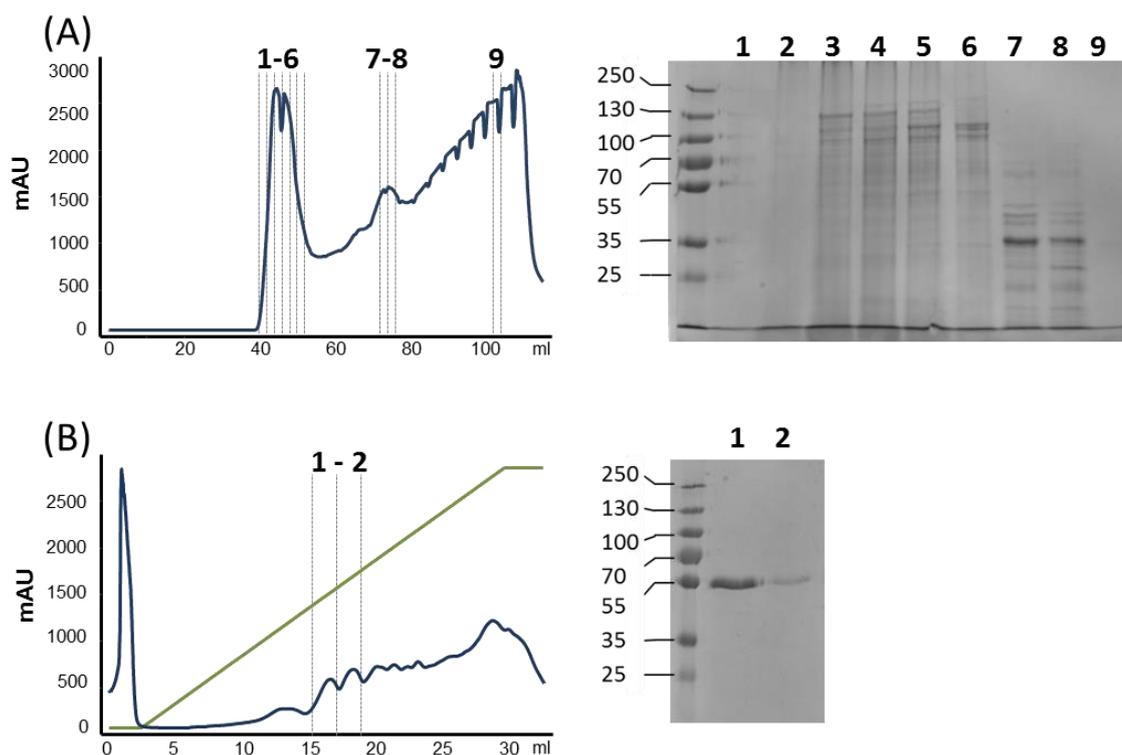


Figure 6-11 - Purification of a GH7 cellulase from the native system. A GH7 was purified from the insoluble biomass in *Graphium* sp. cultures grown on wheat straw. (A) Concentrated protein was applied to a gel filtration column and fractions assessed by SDS-PAGE. (B) Fractions containing the band of interest were combined and anion exchange. Protein containing peaks were run on SDS-PAGE gel to confirm purity.

6.3.7 Enzymatic assays for the GH7

Once the cellulase was pure; its activity was observed on several substrates including avicel, CMC, galactomannan, lichenan, PASC and xylan, using methods described in Section 2.6.2. Cellulase activity was confirmed with reducing sugars released after incubation with CMC and PASC (Figure 6-12A). The greatest amount of reducing sugar was observed after incubation on CMC, although this could reflect the relative ease of the degradation of this amorphous substrate compared to more crystalline cellulose, like Avicel and PASC.

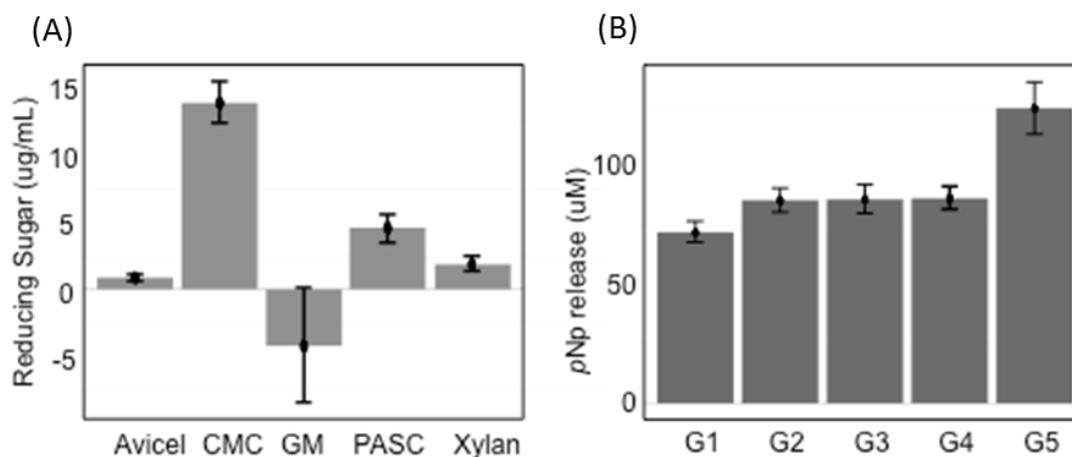


Figure 6-12 - Substrate specificity of a GH7 purified from *Graphium* sp. Cellulose derived substrates, including avicel, CMC and PASC, were incubated with 0.1 mg of purified GH7, alongside galactomannan (GM), and xylan for an hour, and reducing sugar release calculated. (B) pNp substrates were used to determine the length of cellulose oligosaccharides (G1-G5) required for activity.

As activity was detected on PASC, the optimal pH and temperature range of the enzyme was assayed, by calculating the reducing sugar release after incubation with this substrate at varying pHs of Britton–Robinson buffer (Britton and Robinson, 1931). Pnp-substrates were used to determine the length of the polysaccharide required for GH7 activity (Section 2.6.4). The greatest amount of pNp was released from glucose oligos that were five sugars in length (Figure 6-12B). The greatest concentration of reducing sugars was released at pH 8 when incubated at 30 °C, and pH 6 when incubated at 50 °C (Figure 6-13).

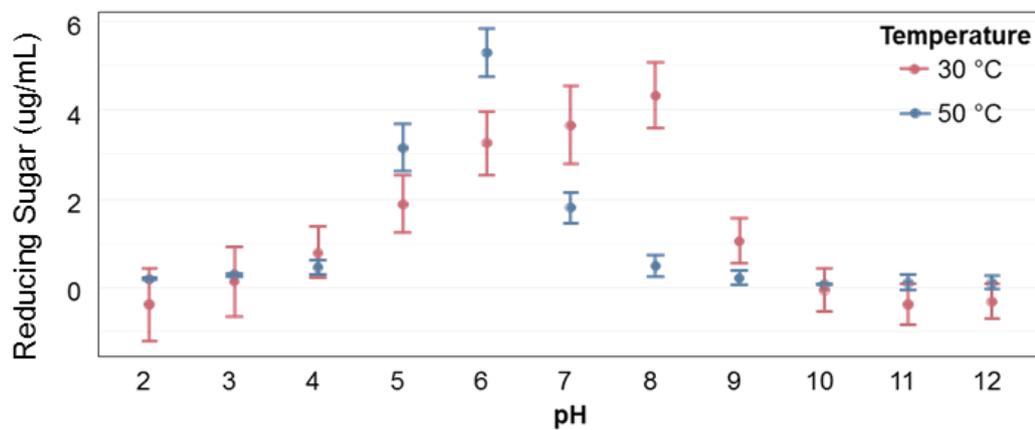


Figure 6-13 pH optima of the GH7 from *Graphium sp.* The pH optimum of the purified enzyme was determined by the release of reducing sugar after 0.1 mg of enzyme was incubated on 2% PASC (w/v), at 30 and 50 °C.

6.4 Discussion

6.4.1 Fungal transformants

Though all proteins targeted could be successfully amplified from cDNA to enable the construction of the pIGF-pyrG construct, transformation into *A.niger* were not successful. Fungal transformations are well known to have low transformation efficiencies and require high quantities of high-quality DNA. They are also time consuming and are not confirmed as successful until a month after transformation. As some success had been achieved with bacterial transformations and purification proteins obtained these methods were focussed on.

6.4.2 The lytic polysaccharide monooxygenase

Having first been discovered in 2012, LPMOs have become an important and active area of research for cellulose deconstruction. Fungal LPMOs which were formally contained within the GH61 family have, in a relatively short period, gone from being a poorly understood, miscategorised family to representing over 300 proteins in the newly designated AA9 family (Harris et al., 2010, Levasseur et al., 2013a). Of those deposited into the CAZyme database 21 of have been characterised and nine structures have been determined (Lombard et al., 2014). There are, however, still outstanding questions concerning their mechanism. These include the function of the fungal-specific N-terminal histidine methylation, the effect of the production of aldonic acids on hydrolytic cellulases, and the most efficient way to incorporate these proteins into enzymatic saccharification cocktails (Muller et al., 2015, Hemsworth et al., 2013).

The transcriptome of *Graphium* sp. appeared rich in LPMOs (Section 4.3.5.2), although this attribute was not observed in the supernatant proteome (Section 5.3.2.2), isolation of the native LPMOs from the culture filtrate was not possible. Fungal transformation proved problematic; however, the use of bacterial hosts has been described in recent literature to be a successful strategy (Hemsworth et al., 2014). Using a similar methodology, an LPMO (c19570_g1_i1), a LPMO which was chosen in part due to its relatively low identity (40 %) to sequences deposited in the CBI non-redundant library and high transcription levels, was successfully produced and purified using the *E. coli* periplasmic expression system.

The characterisation of this LPMO was limited by the inability to detect and identify breakdown products in the supernatant of reaction mixtures by using the Dionex. However, activities were demonstrated on both chitin and cellulose using reducing sugar assays and glycospot plates (Section 2.6.3), respectively. No synergistic effect between the LPMO and the

commercial chitinase mixture was observed, and instead, there appeared an inhibition of the chitin breakdown when both were present in the assay mixtures. This inactivation is possibly due to the generation of hydrogen peroxide, first reported by Kittl et al. (Kittl et al., 2012), as part of a side reaction of the LPMOs. The alleviation of LPMO associated enzyme inactivation has been demonstrated (Scott et al., 2016), with the addition of a catalase in the reaction mixture, a protein that was seen to account for over one percent of the supernatant proteome when *Graphium* sp. was grown on wheat straw (Section 5.3.2.2). It would therefore be of interest to see if a similar effect would be demonstrated in *Graphium* sp. using either commercial or native catalase.

A synergistic effect was observed on wheat straw when the fungal supernatant was used as a source of lignocellulolytic enzymes, as a significant increase in the amount of reducing sugar was detected when the LPMO was present compared to the supernatant alone. The increase in reducing sugars when the supernatant proteins were present could be due to the presence of catalases, detected within the supernatant proteome. The supernatant also contains a large diversity of hydrolytic enzymes, capable of catalysing the cleavage of multiple polysaccharides. Therefore an LPMO that is capable of activity on multiple substrates would have a greater synergy with these enzymes than a single cellobiohydrolase, which may not be compatible with the end products of this LPMO.

In their native host the N-terminal histidine within the LPMO would be methylated (Lo Leggio et al., 2015). Yet, activity was demonstrated for this protein despite this post-translation modification not being present. It is possible that this methylation could control the fine-tuning of the substrate specificity, and could explain why this LPMO appears to be active against both cellulose and chitin. Alternatively, this broad substrate specificity could reflect the environment the *Graphium* sp. thrives in, where there are multiple polysaccharides present. The work described here presents an opportunity investigate this ambiguity as if the LPMOs activity could be expressed in a fungal expression system then a direct comparisons could be performed.

The apparent chitinase activity and the observation that the protein was only observed after biotin-labelling and selection, could suggest that the protein is bound to the chitin of the fungal cell wall rather than the lignocellulosic components of the culture. The production of this enzyme occurs late in the timecourse, when the easily accessible sugars are expected to have been utilised and therefore this could indicate that *Graphium* sp. turned to chitin as an

alternative carbon source. This also may go some way in explaining why *Graphium* sp. thrived late in the compost culture.

However before conclusion can be drawn about this proteins role in lignocellulose degradation, further, more rigorous characterisation is required to confirm substrate specificities. Degradation products were unable to be detected using HPAEC. Due to time constraints MALDI-TOF mass spectrometry, another commonly used technique for the analysis of LPMO degradation products was not attempted, but would add some clarity to the identification of the enzymes substrate and mechanism (Hemsworth et al., 2015).

6.4.3 c6332_g2_i1

A protein that appeared to show no close homology to known proteins within the NCBI non-redundant database was cloned and expressed in *E. coli*. Sequence analysis indicated that this protein contained a domain with similarity to a dioxygenase capable of the cleavage of catechol compounds (Bianchetti et al., 2013), though this was of low identity (29 %), and the remaining regions of the protein had no homology to known proteins in either the NCBI non-redundant database, SwissProt or PDB databases. Although the protein could be successfully expressed in *E. coli* it was not possible to successfully purify the protein from the periplasmic fraction.

Different activities were noted between the extract harvested from cells transformed with the gene of interest and those with a vector-only control. The activity catalysed by this protein, though not fully defined, appeared to be oxygen consuming. Tannic acid incubated without the addition of the protein resulted in the production of a yellow colour indicative of tannic acid oxidation. However, when the same incubation was performed in the presence of c6332_g2_i1 no colour change occurred, suggesting that oxygen was limiting in the environment. Interestingly, congo red also appeared to precipitate out of solution when incubated in reaction mixtures when incubated with recombinant extract containing this protein.

Due to time constraints it was not possible to improve the expression of this enzyme and obtain pure protein for further characterisation. A more comprehensive analysis is now required to fully understand the activity and role of this enzyme in the growth of *Graphium* sp. on lignocellulose, as many factors, including the role of accessory proteins, co-factors and reaction conditions have to be considered. If like the closest hit in the PDB database, this protein does contain a dioxygenase capable of degrading small molecular weight lignin-like structures, this could have interesting commercial applications.

6.4.4 The GH7

Purification of an active GH7 was also achieved from the culture filtrate of *Graphium* sp. growing on wheat straw. The GH7 grouping contains enzymes with a retaining mechanism and β -jelly rolls. The cellobiohydrolases within this family act from the reducing ends of cellulose, in contrast to those of the GH6 family (Glass et al., 2013).

This enzyme, c7203_g1_i1_2, according to the proteomic work performed in Section 5.3.2, was one of the most abundant within the cultural supernatant and also present in the biotin labelled fraction, suggesting that it forms interactions with the insoluble lignocellulose within the wheat straw cultures. The ability to readily isolate the enzyme using surfactants from the insoluble fraction of the culture supports this hypothesis and may make it a protein of interest to the biofuel industry, as one strategy to reduce the cost associated with the production of saccharification enzymes is to reuse the enzyme in multiple reactions (Gregg and Saddler, 1996, Eriksson et al., 2002).

Sequence analysis of this protein suggested a role as a cellobiohydrolase, and activity of PASC was demonstrated using reducing sugar assays. As its primary activity appeared to be cellulolytic, *p*Np-linked glucose oligos were used to determine the optimal length that enzyme requires for the initiation of processive activity. The greatest *p*Np release was seen on cellopentaose. This minimum length for chain cleavage has been observed in other cellobiohydrolases and has been attributed to the structure of the active site requiring the interaction of several oligosaccharides (Kern et al., 2013, Payne et al., 2015).

The pH and temperature optima of the enzyme were also investigated and the greatest amount of reducing sugar was released at pH 8 at 30 °C and pH 6 at 50 °C, respectively, when incubated with PASC. The differences in pH optima of the protein at different temperatures is most likely due to a result of the stability of the protein at different temperatures, though this needs to be confirmed with further investigations. The protein was purified from a culture that was grown at 30 °C, and the pH optimum of 8 is reflective of the supernatant pH change observed. This is also a relatively high pH optima compared to other well-known cellulases including *T. harzianum*'s Cel7B (Pinheiro et al., 2016, Pellegrini et al., 2015, Payne et al., 2015), which may make it a protein of interest in an industrial setting as alkali treatments have proved effective in the pre-treatment of lignocellulose biomass through the removal of lignin (Alvira et al., 2010, Xu and Huang, 2014).

It is however as it is unusual for a single glycoside hydrolase to have multiple substrate specificities. Xylanase activity is not a predicted property of an enzyme belonging to the GH7 family, and enzymes capable of deconstructing crystalline cellulose, such as PASC, are not expected to show activity against carboxymethylcellulose (Payne et al., 2015). It is therefore highly likely that some contaminating proteins remain in the sample and before further characterisation this should be determined through additional purifications and proteomic analysis of the total sample. Once the purity of the protein has been confirmed the action can then be assessed using addition incubations on cello-oligosaccharides and the detection of degradation products by HPAEC.

7 Final Discussion

The work described in this thesis aimed to identify novel lignocellulose degrading enzymes that could be applied as more efficient pre-treatments, to make second generation biofuel an economically attractive prospect.

In Chapter 3 fungal and bacterial members of a community enriched for growth on wheat straw were identified from composting communities. They were assayed for lignocellulolytic ability, and targeted amplicon sequencing assessed their abundance within the community. This revealed a fungus, identified as *Graphium* sp., that appeared to dominate the fungal community in the latter stages of growth on wheat straw. It appeared an attractive prospect for the discovery of novel lignocellulolytic enzymes, as growth on crystalline cellulose and xylan, was demonstrated along with an ability to thrive in conditions with high amounts of kraft lignin, which proved inhibitory to other well-known lignocellulose degraders. Furthermore, no molecular work had been carried out on an organism of this genus.

RNA sequencing was performed on *Graphium* sp. during growth on wheat straw, to determine the mechanisms deployed by this fungus to deconstruct lignocellulose. In total 680 carbohydrate-active enzymes were identified, alongside many other putative proteins. The information obtained from this study could be compared to other known lignocellulolytic degraders, and it was seen that *Graphium* sp. appeared to contain a diverse, and complete arsenal of CAZymes, with high numbers of LPMOs, GH45s and CE1s.

Proteins, present in the cultural supernatant, were also identified in Chapter 5 by LC-MS/MS. This analysis identified fewer CAZymes than the transcriptomic analysis and allowed the direct identification of proteins present in the supernatant. CAZymes within the supernatant appeared to persist after transcription had receded suggesting that they have a degree of stability. Proteomic analysis revealed that the *Graphium* sp. secretome appears to create an environment set up for the oxidative attack of lignocellulose, with multiple cellobiodehydrogenases and glucose oxidases present, along with superoxide dismutases and catalases which may have roles in protecting against any adverse effects induced by the oxidative attack of lignocellulose. Additionally, proteins that formed attachments to the insoluble components of the culture were selected for with the use of biotin-tags. Unexpectedly, however, there was no enrichment of proteins that contained CBMs in this fraction.

Using the information provided by these analyses, several proteins were selected for further investigation. Time constraints meant that only a preliminary characterization study could be performed on an LPMO that was cloned and expressed in *E. coli*. The recombinant LPMO was highly purified and activity was observed with chitin and cellulose.

In addition a protein with no recognisable domain, either Pfam or CAZyme, was selected for expression. This was achieved, and though no lignocellulolytic activity was demonstrated, a differential reaction was observed on both tannic acid and the azo-dye congo red in the presence of either periplasmic fraction of the cells transformed with the empty vector or cells transformed with the target protein. Another protein, a GH7, was purified from the native organism, which was demonstrated as having a higher pH optimum, than many other cellulases. Other proteins, including a chloroperoxidase and CBM containing protein with no identifiable catalytic domain, were recognised as being induced by the presence of wheat straw, raising the possibility that these could break down lignocellulose using novel mechanisms.

7.1 Future directions

The identification of isolates from composting cultures presents multiple opportunities for the discovery of novel lignocellulolytic enzymes, as though many organisms have the ability to degrade lignocellulose, only a small subsection of these species have been characterised in detail.

From the isolates initially identified from the compost culture *Graphium* sp. was chosen for further investigation. The transcriptomic study performed on this organism represents the first genetic information available on its genus and will allow proteins to be cloned and expressed from this species and facilitate further proteomic studies. In addition over 12,562 putative proteins were identified that were upregulated in the presence of wheat straw when compared to glucose.

The wheat straw containing fungal cultures were also investigated to identify proteins that were bound to the insoluble lignocellulose components, with the use of biotin tags. Although this did identify additional CAZyme proteins that were not detected in the culture supernatant, it was surprising that it did not enrich the sample for proteins that contained carbohydrate binding modules, as was expected. The analysis was limited as an absolute quantification of protein could not be achieved. If LC-MS/MS methods were utilised with isobaric tags for relative and absolute quantification (iTRAQ) then the distribution of the proteins between the soluble and insoluble fractions could be compared, adding clarity to the analysis (Adav et al., 2010). The use of biotin to tag proteins is also a novel enrichment method for proteins bound to the insoluble components of cultures, therefore more organisms are needed to be studied using similar enrichment methods to allow for comparisons between industrially relevant organisms and *Graphium* sp.. This method may also be of use when studying cellulosomes as these structures are present on the cell wall of the organism that utilises them.

Databases generated by RNA sequencing and LC-MS/MS analyses, of both the biotin-tagged and supernatant proteins, were used to identify six proteins to be studied in further depth due to their potential biotechnological applications. There were however, many putative proteins that were not chosen for further analysis that may prove of use in the development of more cost effective biofuels. These included multiple proteins with predicted CAZyme domains and CBMs. Furthermore, as lignocellulose degradation is a rapidly advancing area of study, there may also be scope to identify potential lignocellulose degrading enzymes in the future as databases are constantly updated and bioinformatic software refined.

From the proteins upregulated in the presence of lignocellulose, potential lignin modifying enzymes were described, including a chloroperoxidase. However, attempts to heterologously express this protein in *A. niger* failed and it, therefore, was not characterised. Due to time constraints the transformation of *A. niger* was only attempted twice, it is, therefore, possible that with further optimisation the transformation could be achieved. Alternatively, other hosts could be trialled or purification from the native culture attempted. Four other proteins, alongside the chloroperoxidase, were also targeted unsuccessfully for expression in *A. niger*, which included a two unknown proteins, a cellobiohydrolase and a LPMO.

The LPMO, along with a dioxygenase-like protein, were expressed in *E. coli* with greater success. Though the LPMO requires further study to fully understand their role in the lignocellulose-degrading environment, activity was demonstrated against chitin and cellulose. Degradation products of the expressed LPMO however were not identified and the optimal conditions of the enzyme, including pH, temperature and the preference of electron donor, as well as its stability have all yet to be investigated.

The dioxygenase-like protein also requires extensive investigations since its role within the system is still unclear. It did however appear active, on congo red, and appeared to prevent the autoxidation of tannic acid, suggesting an oxygen consuming activity, consistent with its closest database hit, a dioxygenase capable of acting on molecules in the biosynthetic pathway of lignin.

The ability of the fungus to degrade the lignin fraction of wheat straw was not ascertained. If the breakdown of lignin could be confirmed, either through gas chromatography-mass spectrometry (GC-MS) pyrolysis, nuclear magnetic resonance (NMR) spectroscopy or other biochemical techniques, the transcriptomic and proteomic libraries could be probed for the proteins responsible. Additionally, as *Graphium sp.* was capable of growing in high concentrations of kraft lignin, further studies could be undertaken, using experimental techniques optimised in this thesis, to assess differential gene and protein expression between conditions with and without lignin present.

Abbreviations

[Me]GlcA	α -1,2 glucuronic acid and 4-O-methyl glucuronic acid
AA	Auxiliary Activites
ABTS	2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid)
BF	Biotin-tagged protein fractions
BLAST	Basic Local Alignment Search Tool
CAZyme	Carbohydrate active enzymes
CBM	Carbohydrate binding module
CE	Carbohydrate esterase
CMC	Carboxymethyl-cellulose
CPM	Counts per million
CTAB	hexadecyltrimetyl-ammonium bromide
CTAB	Cetyltrimethylammonium bromide
dbCAN	DataBase for automated Carbohydrate-active Enzyme ANnotation
EDTA	Ethylenediaminetetraacetic acid
emPAI	Exponentially modified protein abundance index
EV	Empty Vector
G	Guaiacyl propanol
GAX	Glucuronoarabinoxylans
GH	Glycoside hydrolyases
GPI	Glycosylphosphatidylinositol
GT	Glycosyl transferase
GX	Glucuronoxylan
H	p-hydroxyphenyl
HEC	Hydroxyethylcellulose
HMM	Hidden Markov Models
HPAEC	High-performance Anion Exchange Chromatography
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITS	Internal transcribed spacer
ITS	Internal transcribed spacer
LC-MS/MS	Liquid Chromatography-tandam mass spectrometry
LPMO	Lytic polysaccaride monooxygenase
mRNA	messenger RNA

Abbreviations

NA	Nutrient Agar
NCBI	National Center for Biotechnology Information
nr	non-reducing
OTU	Operational taxonomic unit
PAHBAH	p-hydrobenzoic acid
PASC	Phosphoric acid swollen cellulose
PBS	Phosphate buffered solution
PCR	Polymerase chain reaction
PDA	Potato Dextrose Agar
PDB	RCSB protein database
PL	Polysaccharide lyase
pnp	para-nitrophenol
polyethersulfone	PES
PryG	orotidine-5-phosphate decarboxylase
Rbb	Remazol Brilliant Blue
RPKM	Reads per kilobase per million
rRNA	ribosomal RNA
S	synapI
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SNT	Supernatant fractions
SOD	Superoxide dismutase
TEMED	Tetramethylethylenediamine
v/v	volume to volume ratio
w/v	weight to volume ratio

References

- ABARENKOV, K., NILSSON, R. H., LARSSON, K.-H., ALEXANDER, I. J., EBERHARDT, U., ERLAND, S., HOILAND, K., KJOLLER, R., LARSSON, E., PENNANEN, T., SEN, R., TAYLOR, A. F. S., TEDERSOO, L., URSING, B. M., VRALSTAD, T., LIIMATAINEN, K., PEINTNER, U. & KOLJALG, U. 2010. The UNITE database for molecular identification of fungi - recent updates and future perspectives. *New Phytologist*, 186, 281-285.
- ADAV, S. S., CHAO, L. T. & SZE, S. K. 2013. Protein abundance in multiplexed samples (PAMUS) for quantitation of *Trichoderma reesei* secretome. *Journal of Proteomics*, 83, 180-196.
- ADAV, S. S., LI, A. A., MANAVALAN, A., PUNT, P. & SZE, S. K. 2010. Quantitative iTRAQ Secretome Analysis of *Aspergillus niger* Reveals Novel Hydrolytic Enzymes. *Journal of Proteome Research*, 9.
- ADAV, S. S., RAVINDRAN, A., CHEOW, E. S. H. & SZE, S. K. 2012. Quantitative proteomic analysis of secretome of microbial consortium during saw dust utilization. *Journal of Proteomics*, 75, 5590-5603.
- AGGER, J. W., ISAKSEN, T., VARNAI, A., VIDAL-MELGOSA, S., WILLATS, W. G. T., LUDWIG, R., HORN, S. J., EIJSSINK, V. G. H. & WESTERENG, B. 2014. Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 6287-6292.
- ALGIERI, B. 2014. The influence of biofuels, economic and financial factors on daily returns of commodity futures prices. *Energy Policy*, 69, 227-247.
- ALLGAIER, M., REDDY, A., PARK, J. I., IVANOVA, N., D'HAESELEER, P., LOWRY, S., SAPRA, R., HAZEN, T. C., SIMMONS, B. A., VANDERGHEYNST, J. S. & HUGENHOLTZ, P. 2010. Targeted Discovery of Glycoside Hydrolases from a Switchgrass-Adapted Compost Community. *Plos One*, 5, 9.
- ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403-410.
- ALVIRA, P., TOMÁS-PEJÓ, E., BALLESTEROS, M. & NEGRO, M. J. 2010. Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresource Technology*, 101, 4851-4861.
- ARAI, T., ARAKI, R., TANAKA, A., KARITA, S., KIMURA, T., SAKKA, K. & OHMIYA, K. 2003. Characterization of a cellulase containing a family 30 carbohydrate-binding module (CBM) derived from *Clostridium thermocellum* CellJ: Importance of the CBM to cellulose hydrolysis. *Journal of Bacteriology*, 185, 504-512.
- ARANTES, V., JELLISON, J. & GOODELL, B. 2012. Peculiarities of brown-rot fungi and biochemical Fenton reaction with regard to their potential as a model for bioprocessing biomass. *Applied Microbiology and Biotechnology*, 94, 323-338.
- ARCHIBALD, F. S. 1992. A NEW ASSAY FOR LIGNIN-TYPE PEROXIDASES EMPLOYING THE DYE AZURE-B. *Applied and Environmental Microbiology*, 58, 3110-3116.
- BADHAN, A. & MCALLISTER, T. 2016. Designer Plants for Biofuels: A Review. *Current Metabolomics*, 4, 49-55.
- BALAT, M. & BALAT, H. 2009. Recent trends in global production and utilization of bio-ethanol fuel. *Applied Energy*, 86, 2273-2282.
- BEZERRA, M. A., SANTELLI, R. E., OLIVEIRA, E. P., VILLAR, L. S. & ESCALEIRA, L. A. 2008. Response surface methodology (RSM) as a tool for optimization in analytical chemistry. *Talanta*, 76, 965-977.
- 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. *Journal of Biological Chemistry*, 288, 18574-18587.

- BIELY, P., MACKENZIE, C. R., PULS, J. & SCHNEIDER, H. 1986. Cooperativity of esterases and xylanases in the enzymatic degradation of acetyl xylan. *Bio-Technology*, 4, 731-733.
- BILAL, M., ASGHER, M., IQBAL, M., HU, H. B. & ZHANG, X. H. 2016. Chitosan beads immobilized manganese peroxidase catalytic potential for detoxification and decolorization of textile effluent. *International Journal of Biological Macromolecules*, 89, 181-189.
- BISCHOF, R. H., RAMONI, J. & SEIBOTH, B. 2016. Cellulases and beyond: the first 70 years of the enzyme producer *Trichoderma reesei*. *Microbial Cell Factories*, 15, 13.
- BLAINSKI, A., LOPES, G. C. & DE MELLO, J. C. P. 2013. Application and Analysis of the Folin Ciocalteu Method for the Determination of the Total Phenolic Content from *Limonium Brasiliense* L. *Molecules*, 18, 6852-6864.
- BOERJAN, W., RALPH, J. & BAUCHER, M. 2003. Lignin biosynthesis. *Annual Review of Plant Biology*, 54, 519-546.
- BORASTON, A. B., BOLAM, D. N., GILBERT, H. J. & DAVIES, G. J. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemical Journal*, 382, 769-781.
- BOURDAIS, A., BIDARD, F., ZICKLER, D., BERTEAUX-LECELLIER, V., SILAR, P. & ESPAGNE, E. 2012. Wood Utilization Is Dependent on Catalase Activities in the Filamentous Fungus *Podospora anserina*. *Plos One*, 7, 10.
- BP 2015 Statistical Review of World Energy
- BRANCA, G., CACCHIARELLI, L., MALTSOGLU, I., RINCON, L., SORRENTINO, A. & VALLE, S. 2016. Profits versus jobs: Evaluating alternative biofuel value-chains in Tanzania. *Land Use Policy*, 57, 229-240.
- BRITTON, H. T. S. & ROBINSON, R. A. 1931. Universal buffer solutions and the dissociation constant of veronal. *Journal of the Chemical Society*, 1456-1462.
- BUGG, T. D. H., AHMAD, M., HARDIMAN, E. M. & RAHMANPOUR, R. 2011a. Pathways for degradation of lignin in bacteria and fungi. *Natural Product Reports*, 28, 1883-1896.
- BUGG, T. D. H., AHMAD, M., HARDIMAN, E. M. & SINGH, R. 2011b. The emerging role for bacteria in lignin degradation and bio-product formation. *Current Opinion in Biotechnology*, 22, 394-400.
- BUSSAMRA, B. C., FREITAS, S. & DA COSTA, A. C. 2015. Improvement on sugar cane bagasse hydrolysis using enzymatic mixture designed cocktail. *Bioresource Technology*, 187, 173-181.
- CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. & MADDEN, T. L. 2009. BLAST plus : architecture and applications. *Bmc Bioinformatics*, 10.
- CAMPBELL, J. A., DAVIES, G. J., BULONE, V. & HENRISSAT, B. 1997. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochemical Journal*, 326, 929-939.
- CANO-DELGADO, A., PENFIELD, S., SMITH, C., CATLEY, M. & BEVAN, M. 2003. Reduced cellulose synthesis invokes lignification and defense responses in *Arabidopsis thaliana*. *Plant Journal*, 34, 351-362.
- CANTAREL, B. L., COUTINHO, P. M., RANCUREL, C., BERNARD, T., LOMBARD, V. & HENRISSAT, B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, 37, D233-D238.
- CAPORASO, J. G., KUCZYNSKI, J., STOMBAUGH, J., BITTINGER, K., BUSHMAN, F. D., COSTELLO, E. K., FIERER, N., PENA, A. G., GOODRICH, J. K., GORDON, J. I., HUTTLEY, G. A., KELLEY, S. T., KNIGHTS, D., KOENIG, J. E., LEY, R. E., LOZUPONE, C. A., MCDONALD, D., MUEGGE, B. D., PIRRUNG, M., REEDER, J., SEVINSKY, J. R., TUMBAUGH, P. J., WALTERS, W. A., WIDMANN, J., YATSUNENKO, T., ZANEVELD, J. & KNIGHT, R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7, 335-336.

- CARVALHAIS, L. C., DENNIS, P. G., TYSON, G. W. & SCHENK, P. M. 2012. Application of metatranscriptomics to soil environments. *Journal of Microbiological Methods*, 91, 246-251.
- CARVALHO, A. L., DIAS, F. M. V., NAGY, T., PRATES, J. A. M., PROCTOR, M. R., SMITH, N., BAYER, E. A., DAVIES, G. J., FERREIRA, L. M. A., ROMAO, M. J., FONTES, C. & GILBERT, H. J. 2007. Evidence for a dual binding mode of dockerin modules to cohesins. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 3089-3094.
- CAZZOLA, P., MORRISON, G., KANEKO, H., CUENOT, F., GHANDI, A. & FULTON, L. 2013. Production costs of alternative transportation fuels. Paris: International Energy Agency.
- CHANDEL, A. K., CHANDRASEKHAR, G., SILVA, M. B. & DA SILVA, S. S. 2012. The realm of cellulases in biorefinery development. *Critical Reviews in Biotechnology*, 32, 187-202.
- CHATAIN, B. 2015. Environment committee backs compromise on cleaner biofuels. In: CHATAIN, B. (ed.).
- CHATURVEDI, V. & VERMA, P. 2013. An overview of key pretreatment processes employed for bioconversion of lignocellulosic biomass into biofuels and value added products. 3 *Biotech*, 3, 415-431.
- CHEN, Y. H., CHAI, L. Y., ZHU, Y. H., YANG, Z. H., ZHENG, Y. & ZHANG, H. 2012. Biodegradation of kraft lignin by a bacterial strain *Comamonas* sp B-9 isolated from eroded bamboo slips. *Journal of Applied Microbiology*, 112, 900-906.
- CLARK, K., KARSCH-MIZRACHI, I., LIPMAN, D. J., OSTELL, J. & SAYERS, E. W. 2016. GenBank. *Nucleic Acids Research*, 44, D67-D72.
- COSGROVE, D. J. 2005. Growth of the plant cell wall. *Nature Reviews Molecular Cell Biology*, 6, 850-861.
- COSGROVE, D. J. 2014. Re-constructing our models of cellulose and primary cell wall assembly. *Current Opinion in Plant Biology*, 22, 122-131.
- COUNCIL OF THE EUROPEAN UNION 2009. Renewable Energy Directive. In: EUROPEAN UNION (ed.).
- COUTINHO, P. M., DELEURY, E., DAVIES, G. J. & HENRISSAT, B. 2003. An evolving hierarchical family classification for glycosyltransferases. *Journal of Molecular Biology*, 328, 307-317.
- CRAVATT, B. F., SIMON, G. M. & YATES, J. R., III 2007. The biological impact of mass-spectrometry-based proteomics. *Nature*, 450, 991-1000.
- CULLEN, D. & KERSTEN, P. J. 1996. Enzymology and molecular biology of lignin degradation. *The Mycota, III. A comprehensive treatise on fungi as experimental systems for basic and applied research: Biochemistry and molecular biology*, 295-312.
- DAVIES, G. & HENRISSAT, B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure*, 3, 853-859.
- DAVIN, L. B. & LEWIS, N. G. 2005. Lignin primary structures and dirigent sites. *Current Opinion in Biotechnology*, 16, 407-415.
- DELMAS, S., PULLAN, S. T., GADDIPATI, S., KOKOLSKI, M., MALLA, S., BLYTHE, M. J., IBBETT, R., CAMPBELL, M., LIDDELL, S., ABOOBAKER, A., TUCKER, G. A. & ARCHER, D. B. 2012. Uncovering the Genome-Wide Transcriptional Responses of the Filamentous Fungus *Aspergillus niger* to Lignocellulose Using RNA Sequencing. *PLoS genetics*, 8, e1002875.
- DEMIRBAS, A. H. & DEMIRBAS, I. 2007. Importance of rural bioenergy for developing countries. *Energy Conversion and Management*, 48, 2386-2398.
- DENG, Y. J. & WANG, S. Y. 2016. Synergistic growth in bacteria depends on substrate complexity. *Journal of Microbiology*, 54, 23-30.
- DESANTIS, T. Z., HUGENHOLTZ, P., LARSEN, N., ROJAS, M., BRODIE, E. L., KELLER, K., HUBER, T., DALEVI, D., HU, P. & ANDERSEN, G. L. 2006. Greengenes, a chimera-checked 16S rRNA

References

- gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, 72, 5069-5072.
- DESHMUKH, R., KHARDENAVIS, A. A. & PUROHIT, H. J. 2016. Diverse Metabolic Capacities of Fungi for Bioremediation. *Indian Journal of Microbiology*, 56, 247-264.
- DODD, D. & CANN, I. K. O. 2009. Enzymatic deconstruction of xylan for biofuel production. *Global Change Biology Bioenergy*, 1, 2-17.
- DOUGHERTY, M. J., D'HAESELEER, P., HAZEN, T. C., SIMMONS, B. A., ADAMS, P. D. & HADI, M. Z. 2012. Glycoside Hydrolases from a targeted Compost Metagenome, activity-screening and functional characterization. *BMC Biotechnology*, 12, 9.
- EASTWOOD, D. C., FLOUDAS, D., BINDER, M., MAJCHERCZYK, A., SCHNEIDER, P., AERTS, A., ASIEGBU, F. O., BAKER, S. E., BARRY, K., BENDIKSBY, M., BLUMENTRITT, M., COUTINHO, P. M., CULLEN, D., DE VRIES, R. P., GATHMAN, A., GOODELL, B., HENRISSAT, B., IHRMARK, K., KAUSERUD, H., KOHLER, A., LABUTTI, K., LAPIDUS, A., LAVIN, J. L., LEE, Y. H., LINDQUIST, E., LILLY, W., LUCAS, S., MORIN, E., MURAT, C., OGUIZA, J. A., PARK, J., PISABARRO, A. G., RILEY, R., ROSLING, A., SALAMOV, A., SCHMIDT, O., SCHMUTZ, J., SKREDE, I., STENLID, J., WIEBENGA, A., XIE, X. F., KUES, U., HIBBETT, D. S., HOFFMEISTER, D., HOGBERG, N., MARTIN, F., GRIGORIEV, I. V. & WATKINSON, S. C. 2011. The Plant Cell Wall-Decomposing Machinery Underlies the Functional Diversity of Forest Fungi. *Science*, 333, 762-765.
- EBRINGEROVA, A. & HEINZE, T. 2000. Xylan and xylan derivatives - biopolymers with valuable properties, 1 - Naturally occurring xylylans structures, procedures and properties. *Macromolecular Rapid Communications*, 21, 542-556.
- EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792-1797.
- EILAND, F., KLAMER, M., LIND, A. M., LETH, M. & BAATH, E. 2001. Influence of initial C/N ratio on chemical and microbial composition during long term composting of straw. *Microbial Ecology*, 41, 272-280.
- EISENMAN, H. C. & CASADEVALL, A. 2012. Synthesis and assembly of fungal melanin. *Applied Microbiology and Biotechnology*, 93, 931-940.
- EMANUELSSON, O., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols*, 2, 953-971.
- ERIKSSON, T., BÖRJESSON, J. & TJERNELD, F. 2002. Mechanism of surfactant effect in enzymatic hydrolysis of lignocellulose. *Enzyme and Microbial Technology*, 31, 353-364.
- EWING, M. & MSANGI, S. 2009. Biofuels production in developing countries: assessing tradeoffs in welfare and food security. *Environmental Science & Policy*, 12, 520-528.
- FAZIO, S. & MONTI, A. 2011. Life cycle assessment of different bioenergy production systems including perennial and annual crops. *Biomass & Bioenergy*, 35, 4868-4878.
- FELFOLDI, T., VENGRING, A., KEKI, Z., MARIALIGETI, K., SCHUMANN, P. & TOTH, E. M. 2014. *Euotvesia caeni* gen. nov., sp. nov., isolated from an activated sludge system treating coke plant effluent. *International Journal of Systematic and Evolutionary Microbiology*, 64, 1920-1925.
- FEOFILOVA, E. P. 2010. The fungal cell wall: Modern concepts of its composition and biological function. *Microbiology*, 79, 711-720.
- FINN, R. D., CLEMENTS, J. & EDDY, S. R. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, 39, W29-W37.
- FINN, R. D., MISTRY, J., SCHUSTER-BOCKLER, B., GRIFFITHS-JONES, S., HOLLICH, V., LASSMANN, T., MOXON, S., MARSHALL, M., KHANNA, A., DURBIN, R., EDDY, S. R., SONNHAMMER, E. L. L. & BATEMAN, A. 2006. Pfam: clans, web tools and services. *Nucleic Acids Research*, 34, D247-D251.

- FITZHERBERT, E. B., STRUEBIG, M. J., MOREL, A., DANIELSEN, F., BRUEHL, C. A., DONALD, P. F. & PHALAN, B. 2008. How will oil palm expansion affect biodiversity? *Trends in Ecology & Evolution*, 23, 538-545.
- FLOUDAS, D., BINDER, M., RILEY, R., BARRY, K., BLANCHETTE, R. A., HENRISSAT, B., MARTINEZ, A. T., OTILLAR, R., SPATAFORA, J. W., YADAV, J. S., AERTS, A., BENOIT, I., BOYD, A., CARLSON, A., COPELAND, A., COUTINHO, P. M., DE VRIES, R. P., FERREIRA, P., FINDLEY, K., FOSTER, B., GASKELL, J., GLOTZER, D., GORECKI, P., HEITMAN, J., HESSE, C., HORI, C., IGARASHI, K., JURGENS, J. A., KALLEN, N., KERSTEN, P., KOHLER, A., KUES, U., KUMAR, T. K. A., KUO, A., LABUTTI, K., LARRONDO, L. F., LINDQUIST, E., LING, A., LOMBARD, V., LUCAS, S., LUNDELL, T., MARTIN, R., MCLAUGHLIN, D. J., MORGENSTERN, I., MORIN, E., MURAT, C., NAGY, L. G., NOLAN, M., OHM, R. A., PATYSHAKULIYEVA, A., ROKAS, A., RUIZ-DUENAS, F. J., SABAT, G., SALAMOV, A., SAMEJIMA, M., SCHMUTZ, J., SLOT, J. C., JOHN, F. S., STENLID, J., SUN, H., SUN, S., SYED, K., TSANG, A., WIEBENGA, A., YOUNG, D., PISABARRO, A., EASTWOOD, D. C., MARTIN, F., CULLEN, D., GRIGORIEV, I. V. & HIBBETT, D. S. 2012. The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes. *Science*, 336, 1715-1719.
- FONTES, C. & GILBERT, H. J. 2010. Cellulosomes: Highly Efficient Nanomachines Designed to Deconstruct Plant Cell Wall Complex Carbohydrates. In: KORNBERG, R. D., RAETZ, C. R. H., ROTHMAN, J. E. & THORNER, J. W. (eds.) *Annual Review of Biochemistry*. Palo Alto: Annual Reviews.
- FORSBERG, Z., VAAJE-KOLSTAD, G., WESTERENG, B., BUNAES, A. C., STENSTROM, Y., MACKENZIE, A., SORLIE, M., HORN, S. J. & EIJSINK, V. G. H. 2011. Cleavage of cellulose by a CBM33 protein. *Protein Science*, 20, 1479-1483.
- FRACCHIA, L., DOHRMANN, A. B., MARTINOTTI, M. G. & TEBBE, C. C. 2006. Bacterial diversity in a finished compost and vermicompost: differences revealed by cultivation-independent analyses of PCR-amplified 16S rRNA genes. *Applied Microbiology and Biotechnology*, 71, 942-952.
- FREDERICK, N., ZHANG, N. N., GE, X. M., XU, J. F., PELKKI, M., MARTIN, E. & CARRIER, D. J. 2014. Poplar (*Populus deltoides* L.): The Effect of Washing Pretreated Biomass on Enzymatic Hydrolysis and Fermentation to Ethanol. *Acs Sustainable Chemistry & Engineering*, 2, 1835-1842.
- FUTSCHIK, M. E. & CARLISLE, B. 2005. Noise-robust soft clustering of gene expression time-course data. *Journal of Bioinformatics and Computational Biology*, 3, 965-988.
- GAO, D. H., CHUNDAWAT, S. P. S., LIU, T. J., HERMANSON, S., GOWDA, K., BRUMM, P., DALE, B. E. & BALAN, V. 2010. Strategy for Identification of Novel Fungal and Bacterial Glycosyl Hydrolase Hybrid Mixtures that can Efficiently Saccharify Pretreated Lignocellulosic Biomass. *Bioenergy Research*, 3, 67-81.
- GARCIA-TORREIRO, M., LOPEZ-ABELAIRAS, M., LU-CHAU, T. A. & LEMA, J. M. 2016. Fungal pretreatment of agricultural residues for bioethanol production. *Industrial Crops and Products*, 89, 486-492.
- GELFAND, I., SAHAJPAL, R., ZHANG, X. S., IZAURRALDE, R. C., GROSS, K. L. & ROBERTSON, G. P. 2013. Sustainable bioenergy production from marginal lands in the US Midwest. *Nature*, 493, 514-+.
- GIBELLO, A., VELA, A. I., MARTIN, M., MENGES, G., ALONSO, P. Z., GARBI, C. & FERNANDEZ-GARAYZABAL, J. F. 2011. *Pseudomonas composti* sp. nov., isolated from compost samples. *International Journal of Systematic and Evolutionary Microbiology*, 61, 2962-2966.
- GILBERT, H. J., KNOX, J. P. & BORASTON, A. B. 2013. Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Current Opinion in Structural Biology*, 23, 669-677.

- GILLE, S. & PAULY, M. 2012. O-acetylation of plant cell wall polysaccharides. *Frontiers in Plant Science*, 3.
- GLASS, N. L., SCHMOLL, M., CATE, J. H. D. & CORADETTI, S. 2013. Plant Cell Wall Deconstruction by Ascomycete Fungi. *Annual Review of Microbiology*, 67, 477-498.
- GLEASON, J. E., GALALELDEEN, A., PETERSON, R. L., TAYLOR, A. B., HOLLOWAY, S. P., WANINGER-SARONI, J., CORMACK, B. P., CABELLI, D. E., HART, P. J. & CULOTTA, V. C. 2014. *Candida albicans* SOD5 represents the prototype of an unprecedented class of Cu-only superoxide dismutases required for pathogen defense. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 5866-5871.
- GOLDBECK, R., GONCALVES, T. A., DAMASIO, A. R. L., BRENELLI, L. B., WOLF, L. D., PAIXAO, D. A. A., ROCHA, G. J. M. & SQUINA, F. M. 2016. Effect of hemicellulolytic enzymes to improve sugarcane bagasse saccharification and xylooligosaccharides production. *Journal of Molecular Catalysis B-Enzymatic*, 131, 36-46.
- GONG, J., DONG, J., LIU, X. & MASSANA, R. 2013. Extremely High Copy Numbers and Polymorphisms of the rDNA Operon Estimated from Single Cell Analysis of Oligotrich and Peritrich Ciliates. *Protist*, 164, 369-379.
- GOUKA, R. J., PUNT, P. J. & VANDENHONDEL, C. 1997. Efficient production of secreted proteins by *Aspergillus*: Progress, limitations and prospects. *Applied Microbiology and Biotechnology*, 47, 1-11.
- GRABHERR, M. G., HAAS, B. J., YASSOUR, M., LEVIN, J. Z., THOMPSON, D. A., AMIT, I., ADICONIS, X., FAN, L., RAYCHOWDHURY, R., ZENG, Q., CHEN, Z., MAUCELI, E., HACOEN, N., GNIRKE, A., RHIND, N., DI PALMA, F., BIRREN, B. W., NUSBAUM, C., LINDBLAD-TOH, K., FRIEDMAN, N. & REGEV, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29, 644-U130.
- GREEN, F. & HIGHLEY, T. L. 1997. Mechanism of brown-rot decay: Paradigm or paradox. *International Biodeterioration & Biodegradation*, 39, 113-124.
- GREGG, D. J. & SADDLER, J. N. 1996. Factors affecting cellulose hydrolysis and the potential of enzyme recycle to enhance the efficiency of an integrated wood to ethanol process. *Biotechnology and Bioengineering*, 51, 375-383.
- GRIFFITHS, R. I., WHITELEY, A. S., O'DONNELL, A. G. & 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.
- GUPTA, A. & VERMA, J. P. 2015. Sustainable bio-ethanol production from agro-residues: A review. *Renewable & Sustainable Energy Reviews*, 41, 550-567.
- HALL, B. G. 2013. Building Phylogenetic Trees from Molecular Data with MEGA. *Molecular Biology and Evolution*, 30, 1229-1235.
- HALVORSON, J. J. & GONZALEZ, J. M. 2008. Tannic acid reduces recovery of water-soluble carbon and nitrogen from soil and affects the composition of Bradford-reactive soil protein. *Soil Biology & Biochemistry*, 40, 186-197.
- HAMED, S. A. M. 2013. In-vitro studies on wood degradation in soil by soft-rot fungi: *Aspergillus niger* and *Penicillium chrysogenum*. *International Biodeterioration & Biodegradation*, 78, 98-102.
- HARRIS, D. & DEBOLT, S. 2010. Synthesis, regulation and utilization of lignocellulosic biomass. *Plant Biotechnology Journal*, 8, 244-262.
- HARRIS, P. V., WELNER, D., MCFARLAND, K. C., RE, E., POULSEN, J. C. N., BROWN, K., SALBO, R., DING, H. S., VLASENKO, E., MERINO, S., XU, F., CHERRY, J., LARSEN, S. & LO LEGGIO, L. 2010. Stimulation of Lignocellulosic Biomass Hydrolysis by Proteins of Glycoside Hydrolase Family 61: Structure and Function of a Large, Enigmatic Family. *Biochemistry*, 49, 3305-3316.

- HARTL, L., GASTEBOIS, A., AIMANIANDA, V. & LATGÉ, J.-P. 2011. Characterization of the GPI-anchored endo β -1,3-glucanase Eng2 of *Aspergillus fumigatus*. *Fungal Genetics and Biology*, 48, 185-191.
- HARTMAN, J. C., NIPPERT, J. B., OROZCO, R. A. & SPRINGER, C. J. 2011. Potential ecological impacts of switchgrass (*Panicum virgatum* L.) biofuel cultivation in the Central Great Plains, USA. *Biomass & Bioenergy*, 35, 3415-3421.
- HAVLIK, P., SCHNEIDER, U. A., SCHMID, E., BOTTCHER, H., FRITZ, S., SKALSKY, R., AOKI, K., DE CARA, S., KINDERMANN, G., KRAXNER, F., LEDUC, S., MCCALLUM, I., MOSNIER, A., SAUER, T. & OBERSTEINER, M. 2011. Global land-use implications of first and second generation biofuel targets. *Energy Policy*, 39, 5690-5702.
- HEMSWORTH, G. R., DAVIES, G. J. & WALTON, P. H. 2013. Recent insights into copper-containing lytic polysaccharide mono-oxygenases. *Current Opinion in Structural Biology*, 23, 660-668.
- HEMSWORTH, G. R., HENRISSAT, B., DAVIES, G. J. & WALTON, P. H. 2014. Discovery and characterization of a new family of lytic polysaccharide monooxygenases. *Nature Chemical Biology*, 10, 122-126.
- HEMSWORTH, G. R., JOHNSTON, E. M., DAVIES, G. J. & WALTON, P. H. 2015. Lytic Polysaccharide Monooxygenases in Biomass Conversion. *Trends in Biotechnology*, 33, 747-761.
- HENRIKSSON, G., JOHANSSON, G. & PETTERSSON, G. 2000. A critical review of cellobiose dehydrogenases. *Journal of Biotechnology*, 78, 93-113.
- HENRISSAT, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J*, 280 (Pt 2), 309-16.
- HENRISSAT, B. & BAIROCH, A. 1996. Updating the sequence-based classification of glycosyl hydrolases. *Biochemical Journal*, 316, 695-696.
- HERPOEL-GIMBERT, I., MARGEOT, A., DOLLA, A., JAN, G., MOLLE, D., LIGNON, S., MATHIS, H., SIGOILLOT, J.-C., MONOT, F. & ASTHER, M. 2008. Comparative secretome analyses of two *Trichoderma reesei* RUT-C30 and CL847 hypersecretory strains. *Biotechnology for Biofuels*, 1.
- HESS, M., SCZYRBA, A., EGAN, R., KIM, T. W., CHOKHAWALA, H., SCHROTH, G., LUO, S. J., CLARK, D. S., CHEN, F., ZHANG, T., MACKIE, R. I., PENNACCHIO, L. A., TRINGE, S. G., VISEL, A., WOYKE, T., WANG, Z. & RUBIN, E. M. 2011. Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science*, 331, 463-467.
- HIMMEL, M. E., DING, S. Y., JOHNSON, D. K., ADNEY, W. S., NIMLOS, M. R., BRADY, J. W. & FOUST, T. D. 2007. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science*, 315, 804-807.
- HOFRICHTER, M., ULLRICH, R., PECYNA, M. J., LIERS, C. & LUNDELL, T. 2010. New and classic families of secreted fungal heme peroxidases. *Applied Microbiology and Biotechnology*, 87, 871-897.
- HONGOY, Y. 2011. Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. *Cellular and Molecular Life Sciences*, 68, 1311-1325.
- HORI, C., GASKELL, J., IGARASHI, K., SAMEJIMA, M., HIBBETT, D., HENRISSAT, B. & CULLEN, D. 2013. Genomewide analysis of polysaccharides degrading enzymes in 11 white- and brown-rot Polyporales provides insight into mechanisms of wood decay. *Mycologia*, 105, 1412-1427.
- HUANG, Y., BUSK, P. K. & LANGE, L. 2015. Cellulose and hemicellulose-degrading enzymes in *Fusarium commune* transcriptome and functional characterization of three identified xylanases. *Enzyme and Microbial Technology*, 73-74, 9-19.
- IYAMA, K., LAM, T. B. T. & STONE, B. A. 1990. PHENOLIC-ACID BRIDGES BETWEEN POLYSACCHARIDES AND LIGNIN IN WHEAT INTERNODES. *Phytochemistry*, 29, 733-737.

- INOUE, T., MORIYA, S., OHKUMA, M. & KUDO, T. 2005. Molecular cloning and characterization of a cellulase gene from a symbiotic protist of the lower termite, *Coptotermes formosanus*. *Gene*, 349, 67-75.
- ISHIHAMA, Y., ODA, Y., TABATA, T., SATO, T., NAGASU, T., RAPPSILBER, J. & MANN, M. 2005. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular & Cellular Proteomics*, 4, 1265-1272.
- JONSSON, L. J., ALRIKSSON, B. & NILVEBRANT, N. O. 2013. Bioconversion of lignocellulose: inhibitors and detoxification. *Biotechnology for Biofuels*, 6, 10.
- JORDAN, D. B., BOWMAN, M. J., BRAKER, J. D., DIEN, B. S., HECTOR, R. E., LEE, C. C., MERTENS, J. A. & WAGSCHAL, K. 2012. Plant cell walls to ethanol. *Biochemical Journal*, 442, 241-252.
- JUTURU, V. & WU, J. C. 2014. Microbial cellulases: Engineering, production and applications. *Renewable & Sustainable Energy Reviews*, 33, 188-203.
- JÖNSSON, L. J. & MARTÍN, C. 2016. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. *Bioresource Technology*, 199, 103-112.
- KAMESHWAR, A. K. S. & QIN, W. S. 2016. Recent Developments in Using Advanced Sequencing Technologies for the Genomic Studies of Lignin and Cellulose Degrading Microorganisms. *International Journal of Biological Sciences*, 12, 156-171.
- KERN, M., MCGEEHAN, J. E., STREETER, S. D., MARTIN, R. N. A., BESSER, K., ELIAS, L., EBORALL, W., MALYON, G. P., PAYNE, C. M., HIMMEL, M. E., SCHNORR, K., BECKHAM, G. T., CRAGG, S. M., BRUCE, N. C. & MCQUEEN-MASON, S. J. 2013. Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 10189-10194.
- KIM, S. & DALE, B. E. 2004. Global potential bioethanol production from wasted crops and crop residues. *Biomass & Bioenergy*, 26, 361-375.
- KIM, Y., XIMENES, E., MOSIER, N. S. & LADISCH, M. R. 2011. Soluble inhibitors/deactivators of cellulase enzymes from lignocellulosic biomass. *Enzyme and Microbial Technology*, 48, 408-415.
- KIM, Y. S. & SINGH, A. P. 2000. Micromorphological characteristics of wood biodegradation in wet environments: A review. *Iawa Journal*, 21, 135-155.
- KING, A. J., CRAGG, S. M., LI, Y., DYMOND, J., GUILLE, M. J., BOWLES, D. J., BRUCE, N. C., GRAHAM, I. A. & MCQUEEN-MASON, S. J. 2010. Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 5345-5350.
- KITTL, R., KRACHER, D., BURGSTALLER, D., HALTRICH, D. & LUDWIG, R. 2012. Production of four *Neurospora crassa* lytic polysaccharide monoxygenases in *Pichia pastoris* monitored by a fluorimetric assay. *Biotechnology for Biofuels*, 5.
- KLEIN-MARCUSCHAMER, D., OLESKOWICZ-POPIEL, P., SIMMONS, B. A. & BLANCH, H. W. 2012. The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnology and Bioengineering*, 109, 1083-1087.
- KLEIN-MARCUSCHAMER, D., SIMMONS, B. A. & BLANCH, H. W. 2011. Techno-economic analysis of a lignocellulosic ethanol biorefinery with ionic liquid pre-treatment. *Biofuels Bioproducts & Biorefining-Biofpr*, 5, 562-569.
- KLINDWORTH, A., PRUESSE, E., SCHWEER, T., PEPLIES, J., QUAST, C., HORN, M. & GLOCKNER, F. O. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, 41, 11.
- KNERR, A. & TRIPEPI, R. R. 2014. Changes in Bacterial Communities in Dairy Manure during Nine Months of Composting as Determined by Denaturing Gradient Gel

- Electrophoresis. *International Symposium on Growing Media and Soilless Cultivation*, 1034, 399-407.
- KOSHLAND, D. E. 1953. STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS. *Biological Reviews of the Cambridge Philosophical Society*, 28, 416-436.
- KRACHER, D., SCHEIBLBRANDNER, S., FELICE, A. K. G., BRESLMAYR, E., PREIMS, M., LUDWICKA, K., HALTRICH, D., EIJSINK, V. G. H. & LUDWIG, R. 2016. Extracellular electron transfer systems fuel cellulose oxidative degradation. *Science*, 352, 1098-1101.
- KRACUN, S. K., SCHUCKEL, J., WESTERENG, B., THYGESEN, L. G., MONRAD, R. N., EIJSINK, V. G. H. & WILLATS, W. G. T. 2015. A new generation of versatile chromogenic substrates for high-throughput analysis of biomass-degrading enzymes. *Biotechnology for Biofuels*, 8.
- KUMAR, R., SINGH, S. & SINGH, O. V. 2008. Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *Journal of Industrial Microbiology & Biotechnology*, 35, 377-391.
- KUMAR, R., TABATABAEI, M., KARIMI, K. & HORVATH, I. S. 2016. Recent updates on lignocellulosic biomass derived ethanol - A review. *Biofuel Research Journal-Brj*, 3, 347-356.
- LAM, T. B. T., KADOYA, K. & IYAMA, K. 2001. Bonding of hydroxycinnamic acids to lignin: ferulic and p-coumaric acids are predominantly linked at the benzyl position of lignin, not the beta-position, in grass cell walls. *Phytochemistry*, 57, 987-992.
- LANGSTON, J. A., SHAGHASI, T., ABBATE, E., XU, F., VLASENKO, E. & SWEENEY, M. D. 2011. Oxidoreductive Cellulose Depolymerization by the Enzymes Cellobiose Dehydrogenase and Glycoside Hydrolase 61. *Applied and Environmental Microbiology*, 77, 7007-7015.
- LAOTHANACHAREON, T., BUNTERNGSOOK, B., SUWANNARANGSEE, S., EURWILAICHITR, L. & CHAMPREDA, V. 2015. Synergistic action of recombinant accessory hemicellulolytic and pectinolytic enzymes to *Trichoderma reesei* cellulase on rice straw degradation. *Bioresource Technology*, 198, 682-690.
- LAPOLA, D. M., SCHALDACH, R., ALCAMO, J., BONDEAU, A., KOCH, J., KOELKING, C. & PRIESS, J. A. 2010. Indirect land-use changes can overcome carbon savings from biofuels in Brazil. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 3388-3393.
- LATGE, J. P. 2007. The cell wall: a carbohydrate armour for the fungal cell. *Molecular Microbiology*, 66, 279-290.
- LAURANCE, W. F., SAYER, J. & CASSMAN, K. G. 2014. Agricultural expansion and its impacts on tropical nature. *Trends in Ecology & Evolution*, 29, 107-116.
- LEITE, G. M., MAGAN, N. & MEDINA, A. 2012. Comparison of different bead-beating RNA extraction strategies: An optimized method for filamentous fungi. *Journal of Microbiological Methods*, 88, 413-418.
- LEVASSEUR, A., DRULA, E., LOMBARD, V., COUTINHO, P. M. & HENRISSAT, B. 2013a. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels*, 6.
- LEVASSEUR, A., DRULA, E., LOMBARD, V., COUTINHO, P. M. & HENRISSAT, B. 2013b. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels*, 6, 41.
- LEVER, M. 1972. A new reaction for colorimetric determination of carbohydrates. *Analytical Biochemistry*, 47, 273-279.
- LIU, T., WANG, T. H., LI, X. & LIU, X. 2008. Improved heterologous gene expression in *Trichoderma reesei* by cellobiohydrolase I gene (cbh1) promoter optimization. *Acta Biochimica Et Biophysica Sinica*, 40, 158-165.
- LO LEGGIO, L., SIMMONS, T. J., POULSEN, J.-C. N., FRANDBEN, K. E. H., HEMSWORTH, G. R., STRINGER, M. A., VON FREIESLEBEN, P., TOVBORG, M., JOHANSEN, K. S., DE MARIA, L., HARRIS, P. V., SOONG, C.-L., DUPREE, P., TRYFONA, T., LENFANT, N., HENRISSAT, B.,

References

- DAVIES, G. J. & WALTON, P. H. 2015. Structure and boosting activity of a starch-degrading lytic polysaccharide monoxygenase. *Nature Communications*, 6.
- LOMBARD, V., BERNARD, T., RANCUREL, C., BRUMER, H., COUTINHO, P. M. & HENRISSAT, B. 2010. A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochemical Journal*, 432, 437-444.
- LOMBARD, V., RAMULU, H. G., DRULA, E., COUTINHO, P. M. & HENRISSAT, B. 2014. The carbohydrate-active enzymes database (CAZY) in 2013. *Nucleic Acids Research*, 42, D490-D495.
- LONGORIA, A., TINOCO, R. & VAZQUEZ-DUHALT, R. 2008. Chloroperoxidase-mediated transformation of highly halogenated monoaromatic compounds. *Chemosphere*, 72, 485-490.
- LOPEZ-GONZALEZ, J. A., VARGAS-GARCIA, M. D. C., LOPEZ, M. J., SUAREZ-ESTRELLA, F., JURADO, M. & MORENO, J. 2014. Enzymatic characterization of microbial isolates from lignocellulose waste composting: Chronological evolution. *Journal of Environmental Management*, 145, 137-146.
- LYND, L. R., WEIMER, P. J., VAN ZYL, W. H. & PRETORIUS, I. S. 2002. Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*, 66, 506-+.
- MA, L., ZHANG, J., ZOU, G., WANG, C. S. & ZHOU, Z. H. 2011. Improvement of cellulase activity in *Trichoderma reesei* by heterologous expression of a beta-glucosidase gene from *Penicillium decumbens*. *Enzyme and Microbial Technology*, 49, 366-371.
- MALCA, J. & FREIRE, F. 2006. Renewability and life-cycle energy efficiency of bioethanol and bio-ethyl tertiary butyl ether (bioETBE): Assessing the implications of allocation. *Energy*, 31, 3362-3380.
- MARDIS, E. R. 2008. Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*, 9, 387-402.
- MARRIOTT, P. E., GOMEZ, L. D. & MCQUEEN-MASON, S. J. 2016. Unlocking the potential of lignocellulosic biomass through plant science. *New Phytologist*, 209, 1366-1381.
- MARRIOTT, P. E., SIBOUT, R., LAPIERRE, C., FANGEL, J. U., WILLATS, W. G. T., HOFTE, H., GOMEZ, L. D. & MCQUEEN-MASON, S. J. 2014. Range of cell-wall alterations enhance saccharification in *Brachypodium distachyon* mutants. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 14601-14606.
- MARTINEZ, A. T. 2002. Molecular biology and structure-function of lignin- degrading heme peroxidases. *Enzyme and Microbial Technology*, 30, 425-444.
- MARTINEZ, D., CHALLACOMBE, J., MORGENSTERN, I., HIBBETT, D., SCHMOLL, M., KUBICEK, C. P., FERREIRA, P., RUIZ-DUENAS, F. J., MARTINEZ, A. T., KERSTEN, P., HAMMEL, K. E., WYMELENBERG, A. V., GASKELL, J., LINDQUIST, E., SABAT, G., BONDURANT, S. S., LARRONDO, L. F., CANESSA, P., VICUNA, R., YADAV, J., DODDAPANENI, H., SUBRAMANIAN, V., PISABARRO, A. G., LAVIN, J. L., OGUIZA, J. A., MASTER, E., HENRISSAT, B., COUTINHO, P. M., HARRIS, P., MAGNUSON, J. K., BAKER, S. E., BRUNO, K., KENEALY, W., HOEGGER, P. J., KUES, U., RAMAIYA, P., LUCASH, S., SALAMOV, A., SHAPIRO, H., TU, H., CHEE, C. L., MISRA, M., XIE, G., TETER, S., YAVER, D., JAMES, T., MOKREJS, M., POSPISEK, M., GRIGORIEV, I. V., BRETTIN, T., ROKHSAR, D., BERKA, R. & CULLEN, D. 2009. Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 1954-1959.
- MARTINS, L. F., ANTUNES, L. P., PASCON, R. C., FRANCO DE OLIVEIRA, J. C., DIGIAMPIETRI, L. A., BARBOSA, D., PEIXOTO, B. M., VALLIM, M. A., VIANA-NIERO, C., OSTROSKI, E. H., TELLES, G. P., DIAS, Z., DA CRUZ, J. B., JULIANO, L., VERJOVSKI-ALMEIDA, S., DA SILVA, A. M. & SETUBAL, J. C. 2013. Metagenomic Analysis of a Tropical Composting

- Operation at the Sao Paulo Zoo Park Reveals Diversity of Biomass Degradation Functions and Organisms. *Plos One*, 8.
- MCDONALD, D., PRICE, M. N., GOODRICH, J., NAWROCKI, E. P., DESANTIS, T. Z., PROBST, A., ANDERSEN, G. L., KNIGHT, R. & HUGENHOLTZ, P. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *Isme Journal*, 6, 610-618.
- MET 2015. Global climate in context as the world approaches 1 °C above pre-industrial for the first time. *In: OFFICE, M. (ed.)*.
- MONTANIER, C., VAN BUEREN, A. L., DUMON, C., FLINT, J. E., CORREIA, M. A., PRATES, J. A., FIRBANK, S. J., LEWIS, R. J., GRONDIN, G. G., GHINET, M. G., GLOSTER, T. M., HERVE, C., KNOX, J. P., TALBOT, B. G., TURKENBURG, J. P., KEROVUO, J., BRZEZINSKI, R., FONTES, C., DAVIES, G. J., BORASTON, A. B. & GILBERT, H. J. 2009. Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 3065-3070.
- MORALES, M., QUINTERO, J., CONEJEROS, R. & AROCA, G. 2015. Life cycle assessment of lignocellulosic bioethanol: Environmental impacts and energy balance. *Renewable & Sustainable Energy Reviews*, 42, 1349-1361.
- MORI, T., KOYAMA, G., KAWAGISHI, H. & HIRAI, H. 2016. Effects of Homologous Expression of 1,4-Benzoquinone Reductase and Homogentisate 1,2-Dioxygenase Genes on Wood Decay in Hyper-Lignin-Degrading Fungus *Phanerochaete sordida* YK-624. *Current Microbiology*, 73, 512-518.
- MUKHERJEE, I. & SOVACOOOL, B. K. 2014. Palm oil-based biofuels and sustainability in southeast Asia: A review of Indonesia, Malaysia, and Thailand. *Renewable & Sustainable Energy Reviews*, 37, 1-12.
- MULLER, G., VARNAI, A., JOHANSEN, K. S., EIJSINK, V. G. H. & HORN, S. J. 2015. Harnessing the potential of LPMO-containing cellulase cocktails poses new demands on processing conditions. *Biotechnology for Biofuels*, 8, 9.
- NATIONS, U. 2015. World Population Prospects: The 2015 Revision – Key Findings and Advance Tables. *In: AFFAIRS, D. O. E. A. S. (ed.)*.
- NEU, H. C. & HEPPEL, L. A. 1965. RELEASE OF ENZYMES FROM ESCHERICHIA COLI BY OSMOTIC SHOCK AND DURING FORMATION OF SPHEROPLASTS. *Journal of Biological Chemistry*, 240, 3685-&.
- NIELSEN, J., LARSSON, C., VAN MARIS, A. & PRONK, J. 2013. Metabolic engineering of yeast for production of fuels and chemicals. *Current Opinion in Biotechnology*, 24, 398-404.
- NISHIYAMA, Y., LANGAN, P. & CHANZY, H. 2002. Crystal structure and hydrogen-bonding system in cellulose 1 beta from synchrotron X-ray and neutron fiber diffraction. *Journal of the American Chemical Society*, 124, 9074-9082.
- OGUNMOLU, F. E., KAUR, I., GUPTA, M., BASHIR, Z., PASARI, N. & YAZDANI, S. S. 2015. Proteomics Insights into the Biomass Hydrolysis Potentials of a Hypercellulolytic Fungus *Penicillium funiculosum*. *Journal of Proteome Research*, 14, 4342-4358.
- OHKUMA, M. 2008. Symbioses of flagellates and prokaryotes in the gut of lower termites. *Trends in Microbiology*, 16, 345-352.
- OKSANEN, J., BLANCHET, F. G., FRIENDLY, M., KINDT, R., LEGENDRE, P., MCGLINN, D., MINCHIN, P. R., O'HARA, R. B., SIMPSON, G. L., SOLYMOS, P., STEVENS, M. H. H., SZOEC, E. & WAGNER, H. 2016. vegan: Community Ecology Package. R package version 2.4-1 ed.
- ORIENTE, A., TRAMONTINA, R., DE ANDRADES, D., HENN, C., SILVA, J. L. C., SIMAO, R. C. G., MALLER, A., POLIZELI, M. & KADOWAKI, M. K. 2015. Characterization of a novel *Aspergillus niger* beta-glucosidase tolerant to saccharification of lignocellulosic biomass products and fermentation inhibitors. *Chemical Papers*, 69, 1050-1057.

- ORTIZ-BERMUDEZ, P., HIRTH, K. C., SREBOTNIK, E. & HAMMEL, K. E. 2007. Chlorination of lignin by ubiquitous fungi has a likely role in global organochlorine production. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 3895-3900.
- ORTIZ-BERMUDEZ, P., SREBOTNIK, E. & HAMMEL, K. E. 2003. Chlorination and cleavage of lignin structures by fungal chloroperoxidases. *Applied and Environmental Microbiology*, 69, 5015-5018.
- PAGES, H., ABOYOUN, P., GENTLEMAN, R. & DEBROY, S. Biostrings: String objects representing biological sequences, and matching algorithms.
- PARADIS, E., CLAUDE, J. & STRIMMER, K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*.
- PARK, J. I., STEEN, E. J., BURD, H., EVANS, S. S., REDDING-JOHNSON, A. M., BATH, T., BENKE, P. I., D'HAESELEER, P., SUN, N., SALE, K. L., KEASLING, J. D., LEE, T. S., PETZOLD, C. J., MUKHOPADHYAY, A., SINGER, S. W., SIMMONS, B. A. & GLADDEN, J. M. 2012. A Thermophilic Ionic Liquid-Tolerant Cellulase Cocktail for the Production of Cellulosic Biofuels. *Plos One*, 7, 10.
- PARTANEN, P., HULTMAN, J., PAULIN, L., AUVINEN, P. & ROMANTSCHUK, M. 2010. Bacterial diversity at different stages of the composting process. *Bmc Microbiology*, 10, 11.
- PATYSHAKULIYEVA, A., MÄKELÄ, M. R., SIETIÖ, O.-M., DE VRIES, R. P. & HILDÉN, K. S. 2014. An improved and reproducible protocol for the extraction of high quality fungal RNA from plant biomass substrates. *Fungal Genetics and Biology*, 72, 201-206.
- PAYNE, C. M., KNOTT, B. C., MAYES, H. B., HANSSON, H., HIMMEL, M. E., SANDGREN, M., STAHLBERG, J. & BECKHAM, G. T. 2015. Fungal Cellulases. *Chemical Reviews*, 115, 1308-1448.
- PELLEGRINI, V. O. A., SERPA, V. I., GODOY, A. S., CAMILO, C. M., BERNARDES, A., REZENDE, C. A., PEREIRA, N., CAIRO, J., SQUINA, F. M. & POLIKARPOV, I. 2015. Recombinant *Trichoderma harzianum* endoglucanase I (Cel7B) is a highly acidic and promiscuous carbohydrate-active enzyme. *Applied Microbiology and Biotechnology*, 99, 9591-9604.
- PETERSEN, T. N., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Meth*, 8, 785-786.
- PETERSON, R., GRINYER, J. & NEVALAINEN, H. 2011. Secretome of the Coprophilous Fungus *Doratomyces stemonitis* C8, Isolated from Koala Feces. *Applied and Environmental Microbiology*, 77, 3793-3801.
- PETERSON, R. & NEVALAINEN, H. 2012. *Trichoderma reesei* RUT-C30-thirty years of strain improvement. *Microbiology-Sgm*, 158, 58-68.
- PHILLIPS, C. M., BEESON, W. T., CATE, J. H. & MARLETTA, M. A. 2011. Cellobiose Dehydrogenase and a Copper-Dependent Polysaccharide Monooxygenase Potentiate Cellulose Degradation by *Neurospora crassa*. *Acs Chemical Biology*, 6, 1399-1406.
- PIERLEONI, A., MARTELLI, P. L. & CASADIO, R. 2008. PredGPI: a GPI-anchor predictor. *Bmc Bioinformatics*, 9, 11.
- PINHEIRO, G. L., DE AZEVEDO-MARTINS, A. C., ALBANO, R. M., DE SOUZA, W. & FRASES, S. 2016. Comprehensive analysis of the cellulolytic system reveals its potential for deconstruction of lignocellulosic biomass in a novel *Streptomyces* sp. *Applied microbiology and biotechnology*.
- POIDEVIN, L., BERRIN, J.-G., BENNATI-GRANIER, C., LEVASSEUR, A., HERPOEL-GIMBERT, I., CHEVRET, D., COUTINHO, P. M., HENRISSAT, B., HEISS-BLANQUET, S. & RECORD, E. 2014. Comparative analyses of *Podospira anserina* secretomes reveal a large array of lignocellulose-active enzymes. *Applied Microbiology and Biotechnology*, 98, 7457-7469.
- POINTING, S. B. 1999. Qualitative methods for the determination of lignocellulolytic enzyme production by tropical fungi. *Fungal Diversity*, 2, 17-33.

References

- POULIOT, S. & BABCOCK, B. A. 2014. The demand for E85: Geographical location and retail capacity constraints. *Energy Economics*, 45, 134-143.
- PUJADAS, G. & PALAU, J. 1999. TIM barrel fold: structural, functional and evolutionary characteristics in natural and designed molecules. *Biologia*, 54, 231-253.
- PUNT, P. J., VAN BIEZEN, N., CONESA, A., ALBERS, A., MANGNUS, J. & VAN DEN HONDEL, C. 2002. Filamentous fungi as cell factories for heterologous protein production. *Trends in Biotechnology*, 20, 200-206.
- QUINLAN, R. J., SWEENEY, M. D., LO LEGGIO, L., OTTEN, H., POULSEN, J.-C. N., JOHANSEN, K. S., KROGH, K. B. R. M., JORGENSEN, C. I., TOVBORG, M., ANTHONSEN, A., TRYFONA, T., WALTER, C. P., DUPREE, P., XU, F., DAVIES, G. J. & WALTON, P. H. 2011. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 15079-15084.
- R CORE DEVELOPMENT TEAM 2010. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- RAPPE, M. S. & GIOVANNONI, S. J. 2003. The uncultured microbial majority. *Annual Review of Microbiology*, 57, 369-394.
- RASHID, G. M. M., TAYLOR, C. R., LIU, Y. Q. X., ZHANG, X. Y., REA, D., FULOP, V. & BUGG, T. D. H. 2015. Identification of Manganese Superoxide Dismutase from *Sphingobacterium* sp. T2 as a Novel Bacterial Enzyme for Lignin Oxidation. *Acs Chemical Biology*, 10, 2286-2294.
- RAVALASON, H., GRISEL, S., CHEVRET, D., FAVEL, A., BERRIN, J. G., SIGOILLOT, J. C. & HERPOEL-GIMBERT, I. 2012. *Fusarium verticillioides* secretome as a source of auxiliary enzymes to enhance saccharification of wheat straw. *Bioresource Technology*, 114, 589-596.
- RIES, L., PULLAN, S. T., DELMAS, S., MALLA, S., BLYTHE, M. J. & ARCHER, D. B. 2013. Genome-wide transcriptional response of *Trichoderma reesei* to lignocellulose using RNA sequencing and comparison with *Aspergillus niger*. *Bmc Genomics*, 14.
- RIGDEN, D. J., EBERHARDT, R. Y., GILBERT, H. J., XU, Q. P., CHANG, Y. Y. & GODZIK, A. 2014. Structure- and context-based analysis of the GxGYxYP family reveals a new putative class of Glycoside Hydrolase. *Bmc Bioinformatics*, 15, 12.
- ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-140.
- ROCHA, V. A. L., MAEDA, R. N., PEREIRA, N., KERN, M. F., ELIAS, L., SIMISTER, R., STEELE-KING, C., GOMEZ, L. D. & MCQUEEN-MASON, S. J. 2016. Characterization of the cellulolytic secretome of *Trichoderma harzianum* during growth on sugarcane bagasse and analysis of the activity boosting effects of swollenin. *Biotechnology Progress*, 32, 327-336.
- RUBIN, E. M. 2008. Genomics of cellulosic biofuels. *Nature*, 454, 841-845.
- SALOHEIMO, M. & PAKULA, T. M. 2012. The cargo and the transport system: secreted proteins and protein secretion in *Trichoderma reesei* (*Hypocrea jecorina*). *Microbiology-Sgm*, 158, 46-57.
- SANTOS, V. L. & LINARDI, V. R. 2004. Biodegradation of phenol by a filamentous fungi isolated from industrial effluents - identification and degradation potential. *Process Biochemistry*, 39, 1001-1006.
- SAVITHA, S., SADHASIVAM, S. & SWAMINATHAN, K. 2009. Modification of paper properties by the pretreatment of wastepaper pulp with *Graphium putredinis*, *Trichoderma harzianum* and fusant xylanases. *Bioresource Technology*, 100, 883-889.
- SCHELLER, H. V. & ULVSKOV, P. 2010. Hemicelluloses. *Annual Review of Plant Biology*, Vol 61, 61, 263-289.

- SCHOCH, C. L., SEIFERT, K. A., HUHDORF, S., ROBERT, V., SPOUGE, J. L., LEVESQUE, C. A., CHEN, W., BOLCHACOVA, E., VOIGT, K., CROUS, P. W., MILLER, A. N., WINGFIELD, M. J., AIME, M. C., AN, K. D., BAI, F. Y., BARRETO, R. W., BEGEROW, D., BERGERON, M. J., BLACKWELL, M., BOEKHOUT, T., BOGALE, M., BOONYUEN, N., BURGAS, A. R., BUYCK, B., CAI, L., CAI, Q., CARDINALI, G., CHAVERRI, P., COPPINS, B. J., CRESPO, A., CUBAS P, P., CUMMINGS, C., DAMM, U., DE BEER, Z. W., DE HOOG, G. S., DEL-PRADO, R., DENTINGER, B., DIEGUEZ-URIBEONDO, J., DIVAKAR, P. K., DOUGLAS, B., DUENAS, M., DUONG, T. A., EBERHARDT, U., EDWARDS, J. E., ELSHAHED, M. S., FLIEGEROVA, K., FURTADO, M., GARCIA, M. A., GE, Z. W., GRIFFITH, G. W., GRIFFITHS, K., GROENEWALD, J. Z., GROENEWALD, M., GRUBE, M., GRYZENHOUT, M., GUO, L. D., HAGEN, F., HAMBLETON, S., HAMELIN, R. C., HANSEN, K., HARROLD, P., HELLER, G., HERRERA, G., HIRAYAMA, K., HIROOKA, Y., HO, H. M., HOFFMANN, K., HOFSTETTER, V., HOGNABBA, F., HOLLINGSWORTH, P. M., HONG, S. B., HOSAKA, K., HOUBRAKEN, J., HUGHES, K., HUHTINEN, S., HYDE, K. D., JAMES, T., JOHNSON, E. M., JOHNSON, J. E., JOHNSTON, P. R., JONES, E. B., KELLY, L. J., KIRK, P. M., KNAPP, D. G., KOLJALG, U., KOVACS, G. M., KURTZMAN, C. P., LANDVIK, S., LEAVITT, S. D., LIGGENSTOFFER, A. S., LIIMATAINEN, K., LOMBARD, L., LUANGSA-ARD, J. J., LUMBSCH, H. T., MAGANTI, H., MAHARACHCHIKUMBURA, S. S., MARTIN, M. P., MAY, T. W., MCTAGGART, A. R., METHVEN, A. S., et al. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 6241-6246.
- SCOTT, B. R., HUANG, H. Z., FRICKMAN, J., HALVORSEN, R. & JOHANSEN, K. S. 2016. Catalase improves saccharification of lignocellulose by reducing lytic polysaccharide monoxygenase-associated enzyme inactivation. *Biotechnology Letters*, 38, 425-434.
- SHEPPARD, A. W., GILLESPIE, I., HIRSCH, M. & BEGLEY, C. 2011. Biosecurity and sustainability within the growing global bioeconomy. *Current Opinion in Environmental Sustainability*, 3, 4-10.
- SHIMIZU, M., YUDA, N., NAKAMURA, T., TANAKA, H. & WARIISHI, H. 2005. Metabolic regulation at the tricarboxylic acid and glyoxylate cycles of the lignin-degrading basidiomycete *Phanerochaete chrysosporium* against exogenous addition of vanillin. *Proteomics*, 5, 3919-3931.
- SIMMONS, C. W., REDDY, A. P., SIMMONS, B. A., SINGER, S. W. & VANDERGHEYNST, J. S. 2014. Effect of inoculum source on the enrichment of microbial communities on two lignocellulosic bioenergy crops under thermophilic and high-solids conditions. *Journal of Applied Microbiology*, 117, 1025-1034.
- SKINNER, K. M., MARTINEZ-PRADO, A., HYMAN, M. R., WILLIAMSON, K. J. & CIUFFETTI, L. M. 2008. Pathway, inhibition and regulation of methyl tertiary butyl ether oxidation in a filamentous fungus, *Graphium* sp. *Applied Microbiology and Biotechnology*, 77, 1359-1365.
- SONNHAMMER, E. L. L., EDDY, S. R., BIRNEY, E., BATEMAN, A. & DURBIN, R. 1998. Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Research*, 26, 320-322.
- SOTOYAMA, K., AKUTSU, K. & NAKAJIMA, M. 2016. Biological control of *Fusarium* wilt by *Bacillus amyloliquefaciens* IUMC7 isolated from mushroom compost. *Journal of General Plant Pathology*, 82, 105-109.
- SUN, J. P., TIAN, C. G., DIAMOND, S. & GLASS, N. L. 2012. Deciphering Transcriptional Regulatory Mechanisms Associated with Hemicellulose Degradation in *Neurospora crassa*. *Eukaryotic Cell*, 11, 482-493.
- SUN, Y. & CHENG, J. Y. 2002. Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresource Technology*, 83, 1-11.

- SWATLOSKI, R. P., SPEAR, S. K., HOLBREY, J. D. & ROGERS, R. D. 2002. Dissolution of cellulose with ionic liquids. *Journal of the American Chemical Society*, 124, 4974-4975.
- TAHA, M., SHAHSAVARI, E., AL-HOTHALY, K., MOURADOV, A., SMITH, A. T., BALL, A. S. & ADETUTU, E. M. 2015. Enhanced Biological Straw Saccharification Through Coculturing of Lignocellulose-Degrading Microorganisms. *Applied Biochemistry and Biotechnology*, 175, 3709-3728.
- TAKASHIMA, S., NAKAMURA, A., HIDAKA, M., MASAKI, H. & UOZUMI, T. 1999. Molecular cloning and expression of the novel fungal beta-glucosidase genes from *Humicola grisea* and *Trichoderma reesei*. *Journal of Biochemistry*, 125, 728-736.
- TALEBNIYA, F., KARAKASHEV, D. & ANGELIDAKI, I. 2010. Production of bioethanol from wheat straw: An overview on pretreatment, hydrolysis and fermentation. *Bioresource Technology*, 101, 4744-4753.
- TATUSOV, R. L., FEDOROVA, N. D., JACKSON, J. D., JACOBS, A. R., KIRYUTIN, B., KOONIN, E. V., KRYLOV, D. M., MAZUMDER, R., MEKHEDOV, S. L., NIKOLSKAYA, A. N., RAO, B. S., SMIRNOV, S., SVERDLOV, A. V., VASUDEVAN, S., WOLF, Y. I., YIN, J. J. & NATALE, D. A. 2003. The COG database: an updated version includes eukaryotes. *Bmc Bioinformatics*, 4.
- 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. *Journal of Applied Microbiology*, 113, 521-530.
- TIAN, C., BEESON, W. T., IAVARONE, A. T., SUN, J., MARLETTA, M. A., CATE, J. H. D. & GLASS, N. L. 2009. Systems analysis of plant cell wall degradation by the model filamentous fungus *Neurospora crassa*. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 22157-22162.
- TOMME, P., KWAN, E., GILKES, N. R., KILBURN, D. G. & WARREN, R. A. J. 1996. Characterization of CenC, an enzyme from *Cellulomonas fimi* with both endo- and exoglucanase activities. *Journal of Bacteriology*, 178, 4216-4223.
- TOTHILL, I. E., BEST, D. J. & SEAL, K. J. 1988. The isolation of *Graphium putredinis* from a spoilt emulsion paint and the characterisation of its cellulase complex. *International Biodeterioration*, 24, 359-365.
- TUOMELA, M., VIKMAN, M., HATAKKA, A. & ITAVAARA, M. 2000. Biodegradation of lignin in a compost environment: a review. *Bioresource Technology*, 72, 169-183.
- TURNER, M. B., SPEAR, S. K., HUDDLESTON, J. G., HOLBREY, J. D. & ROGERS, R. D. 2003. Ionic liquid salt-induced inactivation and unfolding of cellulase from *Trichoderma reesei*. *Green Chemistry*, 5, 443-447.
- UNITED STATES CONGRESS 2005. Energy Policy Act.
- VAAJE-KOLSTAD, G., WESTERENG, B., HORN, S. J., LIU, Z., ZHAI, H., SORLIE, M. & EIJSINK, V. G. H. 2010. An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides. *Science*, 330, 219-222.
- VALENT, B. S. & ALBERSHEIM, P. 1974. STRUCTURE OF PLANT-CELL WALLS .5. BINDING OF XYLOGLUCAN TO CELLULOSE FIBERS. *Plant Physiology*, 54, 105-108.
- VAN EIJCK, J., BATIDZIRAI, B. & FAAIJ, A. 2014. Current and future economic performance of first and second generation biofuels in developing countries. *Applied Energy*, 135, 115-141.
- VANHARTINGSVELDT, W., MATTERN, I. E., VANZEIJL, C. M. J., POUWELS, P. H. & VANDENHONDEL, C. 1987. Development of a homologous transformation system for *Aspergillus niger* based on the PYRG-gene. *Molecular & General Genetics*, 206, 71-75.
- VANHOLME, R., DEMEDTS, B., MORREEL, K., RALPH, J. & BOERJAN, W. 2010. Lignin Biosynthesis and Structure. *Plant Physiology*, 153, 895-905.

- VANHOLME, R., MORREEL, K., DARRAH, C., OYARCE, P., GRABBER, J. H., RALPH, J. & BOERJAN, W. 2012. Metabolic engineering of novel lignin in biomass crops. *New Phytologist*, 196, 978-1000.
- VARGAS-GARCIA, M. D., LOPEZ, M. J., SUAREZ-ESTRELLA, F. & MORENO, J. 2012. Compost as a source of microbial isolates for the bioremediation of heavy metals: In vitro selection. *Science of the Total Environment*, 431, 62-67.
- VENTORINO, V., ALIBERTI, A., FARACO, V., ROBERTIELLO, A., GIACOBBE, S., ERCOLINI, D., AMORE, A., FAGNANO, M. & PEPE, O. 2015. Exploring the microbiota dynamics related to vegetable biomasses degradation and study of lignocellulose-degrading bacteria for industrial biotechnological application. *Scientific Reports*, 5, 13.
- VERMAAS, J. V., CROWLEY, M. F., BECKHAM, G. T. & PAYNE, C. M. 2015. Effects of Lytic Polysaccharide Monooxygenase Oxidation on Cellulose Structure and Binding of Oxidized Cellulose Oligomers to Cellulases. *Journal of Physical Chemistry B*, 119, 6129-6143.
- VOCADLO, D. J. & DAVIES, G. J. 2008. Mechanistic insights into glycosidase chemistry. *Current Opinion in Chemical Biology*, 12, 539-555.
- VOGEL, C. & MARCOTTE, E. M. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*, 13, 227-32.
- VOGEL, J. 2008. Unique aspects of the grass cell wall. *Current Opinion in Plant Biology*, 11, 301-307.
- VU, V. V., BEESON, W. T., SPAN, E. A., FARQUHAR, E. R. & MARLETTA, M. A. 2014. A family of starch-active polysaccharide monooxygenases. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 13822-13827.
- WANG, L., LITTLEWOOD, J. & MURPHY, R. J. 2013. Environmental sustainability of bioethanol production from wheat straw in the UK. *Renewable & Sustainable Energy Reviews*, 28, 715-725.
- WANG, X. Q., CUI, H. Y., SHI, J. H., ZHAO, X. Y., ZHAO, Y. & WEI, Z. M. 2015. Relationship between bacterial diversity and environmental parameters during composting of different raw materials. *Bioresource Technology*, 198, 395-402.
- WANG, Y., HAYATSU, M. & FUJII, T. 2012. Extraction of Bacterial RNA from Soil: Challenges and Solutions. *Microbes and Environments*, 27.
- WARNES, G. R., BOLKER, B., BONEBAKKER, L., GENTLEMAN, R., LIAW, W. H. A., LUMLEY, T., MAECHLER, M., MAGNUSSON, A., MOELLER, S., SCHWARTZ, M. & VENABLES, B. 2016. ggplots: Various R Programming Tools for Plotting Data
- WEBER, A. A. T. & PAWLOWSKI, J. 2013. Can Abundance of Protists Be Inferred from Sequence Data: A Case Study of Foraminifera. *Plos One*, 8, 8.
- WERLING, B. P., DICKSON, T. L., ISAACS, R., GAINES, H., GRATTON, C., GROSS, K. L., LIERE, H., MALMSTROM, C. M., MEEHAN, T. D., RUAN, L. L., ROBERTSON, B. A., ROBERTSON, G. P., SCHMIDT, T. M., SCHROTENBOER, A. C., TEAL, T. K., WILSON, J. K. & LANDIS, D. A. 2014. Perennial grasslands enhance biodiversity and multiple ecosystem services in bioenergy landscapes. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 1652-1657.
- WESTERENG, B., CANNELLA, D., AGGER, J. W., JORGENSEN, H., ANDERSEN, M. L., EIJSINK, V. G. H. & FELBY, C. 2015. Enzymatic cellulose oxidation is linked to lignin by long-range electron transfer. *Scientific Reports*, 5, 9.
- WHITAKER, J., LUDLEY, K. E., ROWE, R., TAYLOR, G. & HOWARD, D. C. 2010. Sources of variability in greenhouse gas and energy balances for biofuel production: a systematic review. *Global Change Biology Bioenergy*, 2, 99-112.
- WICKHAM, H. 2009. *ggplot2: Elegant Graphics for Data Analysis*, New York, Springer-Verlag.
- WICKHAM, H. 2011. The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*.

- WICKHAM, H. 2016. tidy: Easily Tidy Data with `spread()` and `gather()` Functions.
- WICKHAM, H. & FRANCOIS, R. 2016. dplyr: A Grammar of Data Manipulation.
- WOLSKI, P. W., DANA, C. M., CLARK, D. S. & BLANCH, H. W. 2016. Engineering ionic liquid-tolerant cellulases for biofuels production. *Protein Engineering Design & Selection*, 29, 117-122.
- WRIGHT, C. K., LARSON, B., LARK, T. J. & GIBBS, H. K. 2017. Recent grassland losses are concentrated around US ethanol refineries. *Environmental Research Letters*, 12, 16.
- WU, M., BECKHAM, G. T., LARSSON, A. M., ISHIDA, T., KIM, S., PAYNE, C. M., HIMMEL, M. E., CROWLEY, M. F., HORN, S. J., WESTERENG, B., IGARASHI, K., SAMEJIMA, M., STAHLBERG, J., EIJSINK, V. G. H. & SANDGREN, M. 2013. Crystal Structure and Computational Characterization of the Lytic Polysaccharide Monooxygenase GH61D from the Basidiomycota Fungus *Phanerochaete chrysosporium*. *Journal of Biological Chemistry*, 288, 12828-12839.
- WU, S. F., DING, S. J., ZHOU, R. & LI, Z. Z. 2007. Comparative characterization of a recombinant *Volvariella volvacea* endoglucanase I (EG1) with its truncated catalytic core (EG1-CM), and their impact on the bio-treatment of cellulose-based fabrics. *Journal of Biotechnology*, 130, 364-369.
- XIA, W., BAI, Y., CUI, Y., XU, X., QIAN, L., SHI, P., ZHANG, W., LUO, H., ZHAN, X. & YAO, B. 2016. Functional diversity of family 3 β -glucosidases from thermophilic cellulolytic fungus *Humicola insolens* Y1. *Sci Rep*, 6.
- XIMENES, E., KIM, Y., MOSIER, N., DIEN, B. & LADISCH, M. 2010. Inhibition of cellulases by phenols. *Enzyme and Microbial Technology*, 46, 170-176.
- XU, J., WANG, X., LIU, X., XIA, J., ZHANG, T. & XIONG, P. 2016. Enzymatic in situ saccharification of lignocellulosic biomass in ionic liquids using an ionic liquid-tolerant cellulases. *Biomass & Bioenergy*, 93, 180-186.
- XU, Z. Y. & HUANG, F. 2014. Pretreatment Methods for Bioethanol Production. *Applied Biochemistry and Biotechnology*, 174, 43-62.
- YANG, B., DAI, Z., DING, S.-Y. & WYMAN, C. E. 2011. Enzymatic hydrolysis of cellulosic biomass. *Biofuels*, 2, 421-449.
- YANG, Z. H., XIAO, Y., ZENG, G. M., XU, Z. Y. & LIU, Y. S. 2007. Comparison of methods for total community DNA extraction and purification from compost. *Applied Microbiology and Biotechnology*, 74, 918-925.
- YU, H., ZENG, G. M., HUANG, H. L., XI, X. M., WANG, R. Y., HUANG, D. L., HUANG, G. H. & LI, J. B. 2007. Microbial community succession and lignocellulose degradation during agricultural waste composting. *Biodegradation*, 18, 793-802.
- ZHAI, R., HU, J. & SADDLER, J. N. 2016. What Are the Major Components in Steam Pretreated Lignocellulosic Biomass That Inhibit the Efficacy of Cellulase Enzyme Mixtures? *Acs Sustainable Chemistry & Engineering*, 4, 3429-3436.
- ZHANG, J. C., ZENG, G. M., CHEN, Y. N., YU, M., YU, Z., LI, H., YU, Y. & HUANG, H. L. 2011. Effects of physico-chemical parameters on the bacterial and fungal communities during agricultural waste composting. *Bioresource Technology*, 102, 2950-2956.
- ZHANG, Z., DONALDSON, A. A. & MA, X. 2012. Advancements and future directions in enzyme technology for biomass conversion. *Biotechnology Advances*, 30, 913-919.
- ZHAO, Z. T., LIU, H. Q., WANG, C. F. & XU, J. R. 2013. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *Bmc Genomics*, 14, 15.
- ZHONG, R. Q. & YE, Z. H. 2014. Complexity of the transcriptional network controlling secondary wall biosynthesis. *Plant Science*, 229, 193-207.
- ZHOU, H., CHENG, J.-S., WANG, B. L., FINK, G. R. & STEPHANOPOULOS, G. 2012. Xylose isomerase overexpression along with engineering of the pentose phosphate pathway and evolutionary engineering enable rapid xylose utilization and ethanol production by *Saccharomyces cerevisiae*. *Metabolic Engineering*, 14, 611-622.

References

ZIFCAKOVA, L. & BALDRIAN, P. 2012. Fungal polysaccharide monooxygenases: new players in the decomposition of cellulose. *Fungal Ecology*, 5, 481-489.