Mining the secretome of the lignocellulose degrading fungus *Parascedosporium putredinis* NO1

Conor JR Scott

PhD

University of York

Biology

September 2023

Abstract

The demand for sustainable and renewable alternatives to finite and environmentally damaging fossil fuel resources is growing. Lignocellulosic biomass is available in vast amounts, at low cost, and could provide fuels, chemicals, and materials if deconstructed effectively. However, its recalcitrant nature makes the cost-effective utilisation of this substrate difficult to achieve. Despite this, wood-degrading fungi have evolved an array of powerful enzymes for the deconstruction of all components of lignocellulose including the polysaccharides and the aromatic polymer lignin.

In this work, the lignocellulose-degrading capacity of the ascomycete fungus *Parascedosporium putredinis* NO1 was explored in detail. New bioinformatic strategies were developed and employed to probe the genome of *P. putredinis* NO1, the first genome of its genus, and to isolate an *in silico* secretome to allow clearer characterisation of the biomass-degrading response. Proteomic investigations of the growth of *P. putredinis* NO1 on multiple industrially relevant lignocellulosic substrates demonstrated remarkable variation in the *P. putredinis* NO1 secretome depending on growth substrate. Molecular techniques were used to expand the demonstration of the varied secretome temporally and support the hypothesis of a tailored enzymatic response to different lignocellulosic substrates, an understanding of which will be important for the development of efficient biorefinery technology.

The tailored secretome was exploited by investigating the enzymatic response of *P. putredinis* NO1 when grown on substrates with varying lignin contents. This allowed identification of proteins with patterns of abundance suggesting roles in the breakdown of lignin. Finally, a chloroperoxidase protein was selected and investigated for its potential involvement in lignin modification. The characterisation of lignocellulose-degrading organisms from underexplored branches of the tree of life, and the identification and characterisation of new enzymes with unclear or unknown roles in lignocellulose breakdown will be vital to improving our understanding of lignocellulose deconstruction and to achieve this efficiently in biorefineries.

Disclaimer

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references. The following thesis is presented in a "journal-style" format, where the results chapters (**2**, **3**, **4**, and **5**) are in a format suitable for publication in peer-reviewed journals. The format of these manuscripts has been modified slightly and the state of publication regarding each results chapter is as follows:

Chapter 2

Title: Whole genome structural predictions reveal hidden diversity in putative oxidative enzymes of the lignocellulose degrading ascomycete *Parascedosporium putredinis* NO1

Authors: Conor JR Scott[a], Daniel R Leadbeater[a], Nicola C Oates[a], Sally R James[b], Katherine Newling[b], Yi Li[b], Nicholas GS McGregor[c], Susannah Bird[a], Neil C Bruce[a]

Affiliations:

[a] Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, United Kingdom

[b] Bioscience Technology Facility, Department of Biology, University of York, York YO10 5DD, United Kingdom

[c] York Structural Biology Laboratory, Department of Chemistry, The University of York, York, YO10 5DD, United Kingdom

State of publication: Published in Microbiology Spectrum and available at: https://doi.org/10.1128/spectrum.01035-23

Chapter 3

Title: A bioinformatic workflow for *in silico* secretome prediction with the lignocellulose degrading ascomycete fungus *Parascedosporium putredinis* NO1

Authors: Conor JR Scott[a], Daniel R Leadbeater[a], Neil C Bruce[a]

Affiliations:

[a] Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, United Kingdom

State of publication: Published in Molecular Microbiology and available at: https://onlinelibrary.wiley.com/doi/full/10.1111/mmi.15144

Chapter 4

Title: *Parascedosporium putredinis* NO1 tailors its secretome for different lignocellulosic substrates

Authors: Conor JR Scott[a], Nicholas GS McGregor[c], Nicola C Oates[a], Janina Hoßbach[a], Amira Abood[a], Alexander Setchfield[a], Daniel Leadbeater[a], Adam Dowle[b], Herman S Overkleeft[d], Gideon J Davies[c], Neil C Bruce[a]

Affiliations:

[a] Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, United Kingdom

[b] Bioscience Technology Facility, Department of Biology, University of York, York YO10 5DD, United Kingdom

[c] York Structural Biology Laboratory, Department of Chemistry, The University of York, York, YO10 5DD, United Kingdom

[d] Leiden Institute of Chemistry, Leiden University. P.O. Box 9502, 2300 RA Leiden, The Netherlands

State of publication: Manuscript prepared

Chapter 5

Title: Isolating lignin degrading enzymes from the *Parascedosporium putredinis* NO1 secretome

Authors: Conor JR Scott[a], Adam Dowle[b], Neil C Bruce[a]

Affiliations:

[a] Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, United Kingdom

[b] Bioscience Technology Facility, Department of Biology, University of York, York YO10 5DD, United Kingdom

State of publication: Manuscript prepared

Table of contents

List of figures

List of tables

Abbreviations

4MU	4-Methylumbelliferone
ABPP	Activity-based protein profiling
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
COG	Clusters of orthologous groups
CPO	Chloroperoxidase
DTT	Dithiothreitol
DUF	Domain of unknown function
EDTA	Ethylenediaminetetraacetic acid
FTIR	Fourier transformed infrared
GC-MS	Gas chromatography-mass spectrometry
GO	Gene ontology
HMM	Hidden Markov model
HPLC	High-performance liquid chromatography
KEGG	Kyoto encyclopaedia of genes and genomes
LC-MS/MS	Liquid chromatography-mass spectrometry/mass spectrometry
NCBI	National centre for biotechnology information
NMR	Nuclear magnetic resonance
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PDB	Protein data bank
PES	Polyethersulfone
SDS	Sodium dodecyl-sulphate
SEM	Scanning electron microscopy
TFA	Trifluoroacetic acid
TPM	Transcripts per million
v/v	volume to volume ratio
w/v	weight to volume ratio
WGS	Whole genome sequence

Acknowledgements

Thanks so much to Neil, Simon McQueen-Mason, Thierry, and Simon Charnock for their expert supervision, guidance, and support.

A huge thanks to Liam, Jimbo, and the boys of Double Entendre for their everlasting friendship.

Thanks to everyone in my family for their constant support.

Chapter 1

1. General introduction

1.1. The growing demand for energy, materials, and chemicals

The Anthropocene is characterised by the significant changes to the planet made by human activity. Predominantly, this is the result of the excessive use of fossil fuel resources and the distribution of the materials and chemicals derived from them (1). They were formed from the decomposing carbon-rich remains of pre-historic animals and plants which were compressed underground and heated over millions of years (2). The nature of fossil fuel deposition is also the reason why these resources are non-renewable and therefore unsustainable. Currently three major types of fossil fuels are relied on for our energy needs: coal, oil, and natural gas. In 2021, these three energy sources provided almost 80% of the global energy consumption (3). Aside from energy, fossil fuels in the form of petrochemicals are the source from which materials like gels, membranes, and plastics are developed (4). For example, 99% of plastic production relies on petrochemicals from fossil fuels (5). Fossil fuels are also the primary energy source for the industrial processes which produce staple construction materials such as concrete and steel (6, 7). The chemical industry is one of the largest in the world and is also intrinsically tied to the petrochemical industry. Petrochemicals serve as the platform for the production of 90% of bulk organic chemicals, with oil and natural gas providing 99% of the feedstocks for basic commodity chemicals (8). This production is often performed in massive processing facilities to achieve the scale needed for profit, and once again these processing facilities require huge amounts of energy derived almost entirely from fossil fuels (8).

Our reliance on fossil fuels is clear and makes the depletion of these finite resources even more alarming. Yet investment in fossil fuels is increasing once again and commitments have been made to new extraction projects (9). For example, the United Kingdom has seen recent commitment to many new extraction projects, including offshore oil projects, onshore oil and gas projects, and coal mines (10). Even if the current rates of fossil fuel extraction are maintained it is expected that coal, oil, and gas reserves are expected to be depleted within the century (11). By the year 2100, the global population is expected to exceed 10 billion (12). The uncontrolled population growth is accompanied by an increased demand for energy that has thus far been answered with an increased use of fossil fuels (13). Not only is the population increasing exponentially, living standards are also improving rapidly. China and India for example have made significant efforts to boost economic growth that are associated with massively increased fossil fuel use (14). Developing nations are facing numerous challenges such as poverty, unemployment, and low per capita income which remain the major focus instead of resource depletion. The global distribution of the demand

for energy is expanding too, with countries developing rapidly and demanding similar energy needs to which others have been privileged to for some time (15). In the current century, global electricity demand has doubled every 14.5 years and all indications point to this increasing in the future (16). In addition to the increased energy demand, an increased need for the materials and chemicals derived from petrochemicals has followed. However, it is not only the unfeasibility of maintaining a global system built upon the use of finite resources which has been well documented, but also the damage of the use of these resources to the environment, biodiversity, and to public health.

Almost 90% of human-produced carbon-dioxide emissions come from the burning of fossil fuels (17). The global warming effect of releasing carbon which was captured over millions of years back into the atmosphere is widely accepted (18). The effects of global warming on the environment are numerous and include irregular weather patterns, retreating ice sheets, and the corresponding rising sea level (19). Agricultural losses from the effects of climate change caused by global warming are predicted to occur in staple food crops globally, leading to higher prices in developed nations and the risk of hunger and famines in the less developed (20). Biodiversity losses due to climate change are clear for almost all ecosystems (21). Higher atmospheric carbon emissions have also led to acidification of oceans, resulting in habitat and biodiversity losses in ecosystems which are vital to human society and industry (22). It is not only the increased carbon emissions from fossil fuels, but also the environmentally damaging and polluting effects of extracting fossil fuels which adds to the burden of agricultural and biodiversity losses.

Fossil fuels also contribute to 85% of the airborne respirable particulate pollution (23). They are responsible for almost all atmospheric nitrogen dioxide and sulphur dioxide, and are responsible for the release of black carbon, polycyclic aromatic hydrocarbons, mercury, and volatile chemicals which lead to the formation of ground-level ozone (24). All are associated with adverse health effects in children. The Lancet reported that diseases caused by pollution were responsible for an estimated 9 million deaths in 2015, which was 16% of all deaths worldwide and three times as many deaths as AIDs, tuberculosis, and malaria combined, and 15 times as many deaths as those from all wars and forms of violence combined (25). Correlations were even demonstrated between air pollution and COVID-19 mortality and infectivity (26).

Even ignoring the environmental and public health concerns, the increasing inability of the fossil fuel based economy to sustain the projected economic growth should be warning enough to warrant change (15). The COVID-19 pandemic and geopolitical conflicts have also highlighted the volatility of the global energy trade through drastic fluctuations in supply and demand and oil price shocks (27). Never has this been more transparent than in April of

2020, where a reduced demand for liquid transportation fuel due to the COVID-19 pandemic sent oil prices crashing into the negatives at -$40 per barrel (28). This is unfortunately a symptom of a broken energy system in which the major oil producers are at war to supply the already oversupplied fuel market. This unnecessarily increased liquid fuel production serves to deplete the finite fossil fuel reserves further and prevent the establishment of cost-effective renewable alternatives.

During the COVID-19 pandemic, carbon dioxide emissions fell by 6.4 % in 2020 compared to the previous year (29). This was mainly due to reduced transportation emissions resulting from imposed lockdowns. Several other environmental benefits were observed globally such as reduced nitrogen dioxide emissions, reduced human impact on wildlife and biodiversity and reduction in noise pollution (30). This demonstrated that the hundreds of years of damage inflicted to the environment by human activity can be alleviated to some extent, and that the future quality of the environment can be improved. However, without coordinated global action these improvements will be seen as a small blip in the global environmental record. Additionally, the financial burden imposed by the pandemic could serve to undermine progress towards a renewable and sustainable energy system as countries reduce their commitment to tackling climate change (31).

Renewable energy technologies such as wind, solar, and hydroelectric power have seen significant technological advances and dramatic reductions in cost in the past decade. Many indicators suggest that there is an accelerating transition to an energy system based around the electricity produced from such technologies (32). Ambitious claims of almost two thirds of global energy supply being provided by renewable electricity by 2050 have also been made. However, currently these technologies contribute only around 10 % of the total global energy supply (3). The infrastructure for an electrical energy system is not yet well enough established either. For example, electric vehicles may be cheaper to run than internal combustion engine vehicles, but the capital cost of building these batteries remains too high (33). Another issue with renewable electricity is the failure of the technologies to meet demand regionally at times of highest demand. In the United States for example, the energy demand from air conditioners is huge and at its highest in weather conditions where little solar or wind energy is available (16). When energy supply exceeds demand, the technology to effectively store electrical energy on the scales required is not yet established. These technologies are unlikely to meet the net energy needs of a fossil-fuel based economy, they are unsuitable for many industrial processes which developed in a fossil-fuel based infrastructure, and would provide much lower rates of energy return than fossil fuels if expanded to supply the majority of global energy (34).

It is expected that instead, nations will have to diversify their energy sources across renewable technologies and other sources (35). Many other interesting renewable energy sources are developing, such as the generation of biogases from wastewater discharged by industrial plants (36, 37). But even ignoring the energy production shortcomings, they all also fail to provide the other resources we rely upon from fossil fuels. High energy-density liquid transportation fuels make up around 30 % of total current fuel demand and will continue to be needed in the future (38). The reliance on fossil fuels for chemicals and materials will only increase with the growing global population. Clearly a sustainable and renewable resource is required to meet these demands which must also be abundant, available at low-cost, and environmentally friendly.

1.2. Moving towards an alternative feedstock

In the twentieth century, the major focus of research and industry was how to exploit the then readily available fossil fuel resources. As discussed previously, extraction and utilisation of fossil fuels provided the feedstocks for many of the fuels, chemicals, and materials that have become integral to how societies and industries function around the globe. At the turn of the century and as oil prices began to rise, interest shifted to potential alternative sources of fuel. Biofuels can be simply defined as fuels produced from biomass. First-generation biofuels including bioethanol and biodiesel are predominantly derived from edible sources of biomass (39). Edible plant crops are technically renewable resources and the carbon dioxide removed by photosynthesis during growth of the feedstock potentially negates the emissions when the fuels produced from these plants are combusted (40). The process of ethanol production is relatively simple in first-generation technologies. The polysaccharide rich feedstocks are digested into simple sugars which are then fermented to ethanol (39). First-generation bioethanol is now a well-established technology around the globe. For example, in the US there are 200 corn-bioethanol plants producing around 15 billion gallons of bioethanol per year (41). Brazil is the most competitive producer of bioethanol, where 625 million tonnes of sugar cane are produced each harvest and are converted to 27 million litres of fuel (42). In Europe, biodiesel is the major biofuel and is produced predominantly from rapeseed oil but also from soybean, sunflower, and palm oil (43). These fuels are being produced and consumed in huge quantities already and the biorefinery technology to produce them is clearly mature, however the edible nature of first-generation feedstocks raises an issue of its own.

As first-generation biofuels are produced from food crops, they compete directly with food production (44). The land, water, and energy resources which could be used for growing food, and for other ecosystem services, are instead directed to fuel production. This raises ethical concerns, especially when considering the food security issues associated with a growing global population. They also compete with food production through land use changes, where arable farmland is no longer used for crops (45). This debate has led to shrinking bioethanol production in Brazil, the largest producer of first-generation biofuels, as the ethanol production is in direct competition with sugar production (39). Corn bioethanol in the US, and biodiesel in Europe obviously face the same challenge. This competition has been demonstrated to vary depending on the crops and regions in question, however in 2016 it was estimated that 200 million people could be fed by the resources used instead for bioethanol production (46). Additionally, the carbon emission benefits of first-generation biofuels have been questioned. Many assessments of greenhouse gas emissions from first-generation biofuels have reached many different conclusions, likely highlighting the

variability of fuel production regionally. In the best cases, emissions were estimated to be reduced by 80 – 90 %, but often these estimates fail to consider land use changes (47). Other life cycle assessments suggest that the intense deforestation involved in global bioethanol production compromise the carbon dioxide offset of a renewable resource (43). It has even been suggested that when considering these land use changes, biodiesel production in Brazil can produce up to 55 % more greenhouse gas emissions than traditional diesel from fossil fuels (47). Despite the clear disadvantages of first-generation biofuel technology that need to be avoided in the future, the establishment of this technology has provided a platform from which more sustainable approaches can develop. While energy demand increases, alternative feedstocks will be required to avoid further competition with food production and the application of knowledge gained from first-generation biofuels will be important in achieving this (48).

Second-generation biofuels are produced from lignocellulosic biomass, the woody residues of plant crops. These are available in vast amounts, at low cost, and theoretically do not compete with food production (49). Lignocellulose is the most abundant form of biomass on the planet with an estimated 181.5 billion tonnes produced annually, however less than 5 % of this is currently utilised (50). The biomass can be grouped into four main categories: agricultural residues, forestry residues, energy crops, and cellulosic waste. For agricultural wastes this includes components of staple food crops such as wheat straw, rice straw, sugar cane bagasse, empty fruit bunches from oil palm, and corn stover (51-55). However not all these residues are available for biorefining processes as a large proportion need to be recycled into soils for ecological function or are required for alternate uses such as animal bedding etc. (56). The forestry residues can be grouped into hardwood and softwoods which differ in structure and composition (50). Dedicated bioenergy crops include miscanthus, switchgrass, elephant grass, reed canary grass, poplar, and willow (57-61). These crops are attractive for numerous reasons, exemplified by Miscanthus the perennial C4 grass which has high nitrogen, energy, and land use efficiency and which can be harvested annually for up to 20 years (57). Another advantage is their ability to be cultivated on marginal land and therefore not compete with agricultural crops for space. Organic waste can include municipal organic wastes like paper and cardboard, garden waste, or food waste (48, 62). Many of the examples above are either low-cost resources, or no cost and negative cost wastes from other industries, representing an increased opportunity for profit (63).

In many developing countries lignocellulose poses an environmental nuisance due to its abundance and can become a pollutant if released directly into the environment (64). However, the current disposal rates of lignocellulosic substrates are no better. For example, Asia produces the vast majority of rice straw biomass and millions of tonnes of this waste

are burned every year for disposal (65). This burning generates potentially fatal levels of ground level pollution, represents huge economic losses, and decreases agricultural yields in a staple food crop for many nations (66). Therefore, utilisation of lignocellulose can reduce damage to the environment and public health while improving food security through the reduced reliance on fossil fuels and by discouraging current practices for lignocellulose disposal.

However, the process of fuel production is more complex for lignocellulosic feedstocks. A pre-treatment step is required to prepare the polysaccharides locked within lignocellulose for hydrolysis to simple sugars, which can then be fermented to fuel (67). These additional steps mean that there are higher capital costs associated with a lignocellulose biorefinery. The costs of transitioning existing first-generation biorefineries may discourage the change towards the second-generation. It is likely that financial incentives, such as a carbon penalty tax for first-generation producers, will be required to motivate this change (41). The processing costs are also higher for almost all second-generation feedstocks due mainly to lower conversion efficiencies and a higher requirement for downstream processing (68). For example, both US corn and Brazil sugarcane bioethanol production costs are up to three times higher for lignocellulosic feedstocks (67). Therefore, despite the advantages of the reduced impact on food security and the environment, second-generation technologies need to be able to compete economically with first-generation biorefineries or with fossil fuels to be adopted globally.

It is also worth briefly mentioning third-generation biofuel technologies produced from algae. In this technology, the ability of algae to grow simply with water, carbon dioxide, and sunlight and quickly accumulate biomass containing lipids is exploited (69). These lipids can then be converted to biodiesels or can even be used to produce kerosene for aviation fuels (70). The research remains currently in its infancy and faces numerous technical challenges. Algae require large volumes of water to grow, and therefore this technology will be unsuitable for countries where temperatures are below freezing for significant proportions of the year (39). The large volumes of water also present the technical issue of requiring separation of products of interest from the biomass and water on an industrial scale (71).

Lignocellulosic biomass is the most promising resource for a renewable and sustainable alternative to fossil fuels. The combination of long chain polysaccharides and branched polymers of aromatic carbon make the biomass ideal to produce not only liquid fuels, but also the materials and chemicals we also rely upon fossil fuels for. However, it is this complex network of polysaccharides and aromatics which also makes lignocellulose a difficult substrate to degrade, convert, and to generate value from.

1.3. Lignocellulose components and their potential value

All lignocellulosic feedstocks contain the same three major components: cellulose, hemicellulose, and lignin (**Figure 1.1**) (65). However, the proportions of these can vary depending on the type of biomass, the maturity, and the climate conditions (72). Cellulose and hemicellulose are the polysaccharide components of the substrate, traditionally the target for biofuel production. The lignin component confines the polysaccharides and has previously been viewed as a barrier to accessing the reservoir of sugars. However, interest in valorising the entire substrate is growing and therefore understanding the structure and composition of all components of lignocellulose is vital.

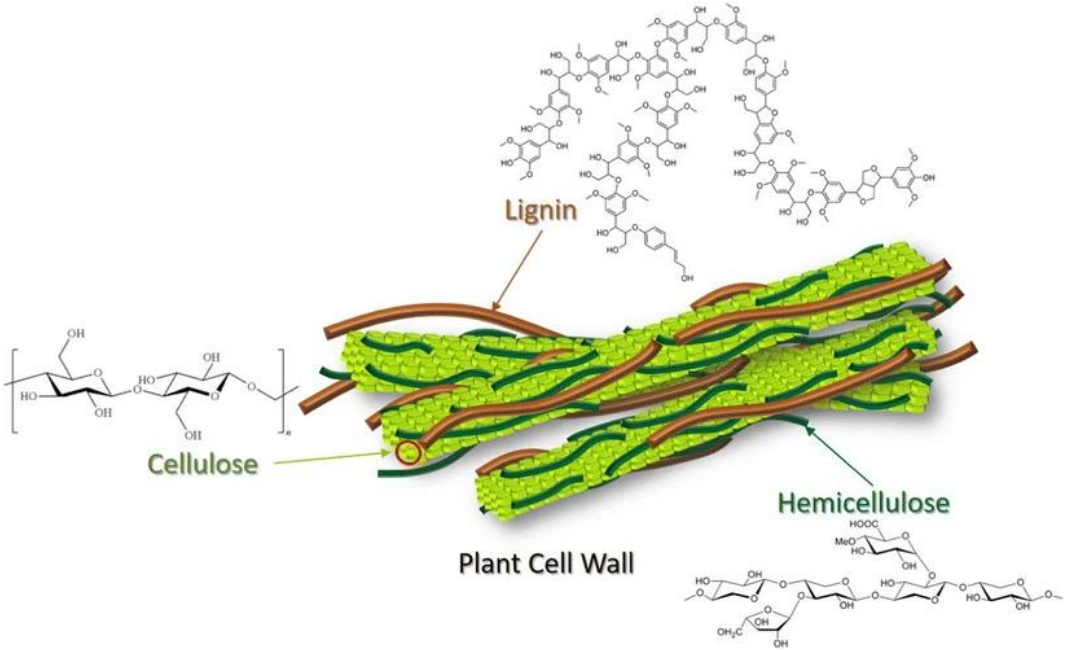


Figure 1.1 Overall structure of lignocellulosic biomass. Sourced from Jensen *et al.* 2017 (73).

1.3.1. Polysaccharides

Cellulose is usually the most abundant polysaccharide present in lignocellulose (**Figure 1.2**). It is a linear homopolysaccharide of unbranched D-glucose units which are linked together by covalent β-glycosidic bonds (64). These linear chains are arranged into bundles to form microfibrils of 3 – 4 nm in diameter, and then these microfibrils are themselves bundled further into cellulose fibres with diameters up to several tens of nanometres (74). Hydrogen bonds, hydrophobic interactions and Van der Waal's forces maintain the parallel alignment of these fibres into a crystalline structure (18). The proportions of cellulose in cereal straws can vary from around 35 – 45 % of the total composition and tends to be higher in hardwood and softwood feedstocks where cellulose can make up to 50 % of the lignocellulosic biomass (75, 76).

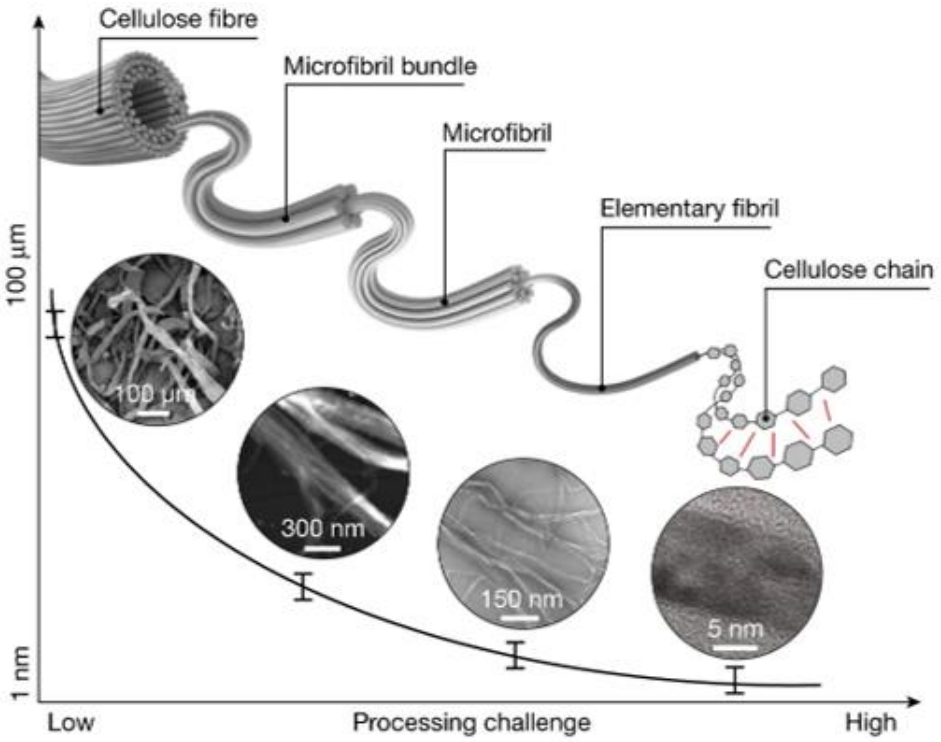


Figure 1.2 The hierarchical structure of fibrillated cellulose. Sourced from Li *et al.* 2021 (77).

In comparison to the homopolymer cellulose, hemicellulose is a heteropolymer made of repeating units of pentose sugars such as xylose and arabinose, and hexose sugars such as mannose, glucose, and galactose (**Figure 1.3**) (74). Sugar acids e.g., galacturonic acid and glucuronic acid are also present in hemicellulose. Hemicelluloses make up around 25 – 35 % of lignocellulose in cereal straws, 15 – 25 % in hardwoods, and are greatly reduced in softwood lignocellulose biomass (75, 76). The heterogenous nature of hemicellulose also gives rise to more potential for variation across different types of feedstocks. In hardwoods, the hemicellulose polysaccharides are predominantly highly acetylated heteroxylans where chains of xylan backbones have small branches of galactose, mannose, and arabinose (78). However in softwood hemicellulose the backbone is mainly partially acetylated glucomannans and galactoglucomannans with xylose and arabinose present in smaller amounts (78).

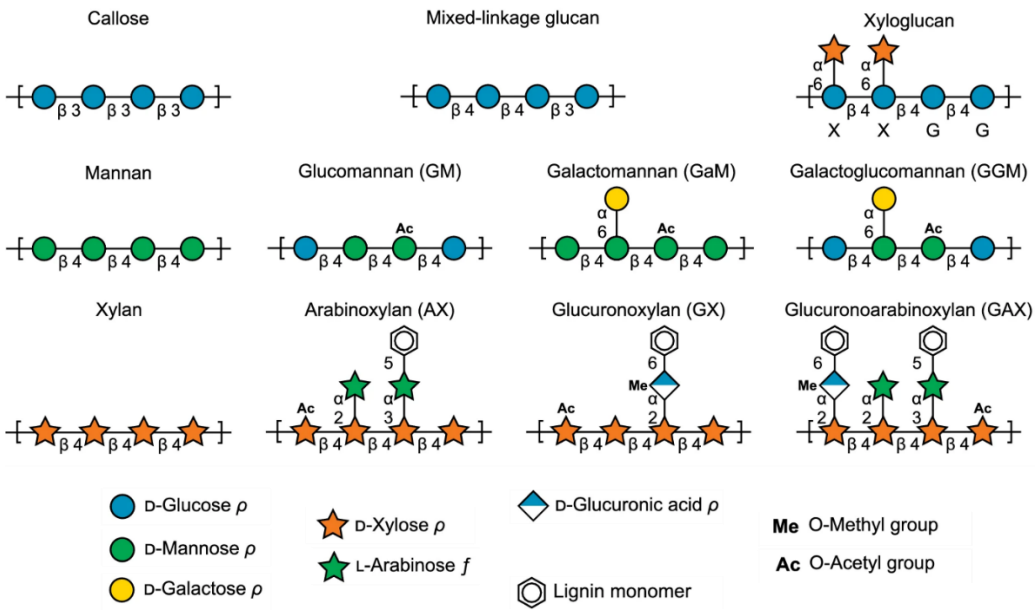


Figure 1.3 Variation in hemicellulosic sugars. Adapted from Tingley *et al.* 2021 (45).

These two polysaccharide components have traditionally been the target within lignocellulose to produce liquid fuel. There are two main routes for the conversion of these polysaccharides to fuel, thermochemical and biochemical (40). The thermochemical approach targets the entire lignocellulose substrate by heating in the presence of oxidising agents to produce solid biochar, liquid bio-oil, and syngas which can then all be utilised for energy generation (39). For the biochemical approach, the polysaccharides are instead isolated from the lignocellulosic biomass and these can then be hydrolysed to simple sugars through chemical or enzymatic hydrolysis (39). The C6 sugars produced from this hydrolysis are readily fermentable to bioethanol by yeast, usually *Saccharomyces cerevisiae*, but other microorganisms have also been investigated (53, 79). Genetically engineered strains of *S. cerevisiae*, which remains the prominent fermenting microorganism for bioethanol, have been developed with the ability to also convert C5 sugars derived from the hemicellulose to ethanol as well (80). Alternatively, these sugars could instead be fermented by *Clostridia* species to biobutanol which has high energy content, low volatility, is less corrosive, and less hygroscopic compared to ethanol (81). These advantages also mean that biobutanol can be considered a “drop-in” fuel i.e., it can be blended with current fuels at a high percentage and be used in engines without the need for modification. It has also been demonstrated that cellulosic and hemicellulosic sugars liberated from lignocellulose can provide the carbon source for the growth of microbial oil accumulating yeasts, where this oil can then be converted to biodiesel (82). This could represent a more realistic approach for the generation of renewable and sustainable biodiesel than the third-generation technologies discussed earlier as the technology required is much more well-established, and the footprint of the biorefineries would theoretically be much smaller.

The sugars which can be derived from the polysaccharide components of lignocellulose also have potential value outside the traditional view of biofuel production. The sugars can be fermented to other valuable chemicals. For example, *Lactococcus lactis* and *Bacillus coagulans* are two microorganisms capable of fermenting hexose and pentose sugars, respectively, to lactic acid which can then be used for the production of bio-based plastics or as food and cosmetic additives (83). It has been envisaged that sugars from lignocellulose could be used as the growth substrate for the production of ketones which can be used as biodiesel additives to improve fuel performance, but also have value for use in flavourings and fragrances (84). Platform chemicals, so called because of their central role in the production of multiple other high-value products, can also be generated from the polysaccharide components of lignocellulose. Platform chemicals such as 5-hydroxymethyl furfural and levulinic acid can both be generated from the hexose sugars derived from cellulose and hemicellulose (85).

Additionally, novel biomaterials can be produced from the polysaccharides within lignocellulose. The nanoscale bundles of cellulose are known as cellulose nanofibers. These have received much interest for use as a natural nanomaterial. Attempts have been made to use cellulose nanofibers in films, membranes, nanocomposites, packaging, electronics, and in medical applications (74). Cellulose has also been utilised to produce hydro-, aero-, and oleo-gels with a wide range of applications (86). The formation of nanocrystals from cellulose has been demonstrated for use in scientific process such as NMR spectroscopy and optical tagging (86). The high degree of branching but relatively low degree of polymerisation has also made hemicellulose a target for manufacturing novel biomaterials. The richness of oxygen containing functional groups such as hydroxyl, acetyl, and carboxyl groups make the polysaccharide network amenable to various chemical modifications which can fine tune the applicability of this material as a hydrogel (87). Hydrogels have many potential applications in medicine, agriculture, biosensors, and industrial adsorption (87). The low oxygen permeability gives hemicellulose antioxidant and anti-moisture properties which is of interest to produce films for food packaging (86).

1.3.2. Lignin

The polysaccharides within lignocellulose are surrounded by a complex, amorphous, and insoluble network of lignin. This network of phenylpropanoid units is formed oxidatively and deposited in the plant secondary cell walls where it provides resistance to biological and chemical attack (88). Lignin represents the second most abundant natural polymer on earth, second to cellulose, and can account for 20 – 35 % of the dry weight of plant biomass (89). The proportion of lignin, like other components of lignocellulose, varies by feedstock. In crop waste and biomass crops the lignin content can contribute to 10 – 20 % of the lignocellulose

substrate, however hardwood and softwood sources this can be up to 30 % (90). The composition of the lignin polymer changes with biomass type as well. An array of enzymes is involved in the conversion of the amino acid L-phenylalanine to the primary monolignols: coniferyl, sinapyl, and *p*-coumaryl alcohols (65). From these the *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) lignin subunits are generated (**Figure 1.4**) (90). These main aromatic components can then be cross-linked oxidatively through radical polymerisation to form the three-dimensional biopolymer (65). This radical coupling gives rise to a variety of linkages within the lignin network, predominantly ether linkages and covalent carbon-carbon bonds. Gymnosperms lack the S units, and this results in a more branched lignin structure than in angiosperms which are rich in G and S units (90). Several other phenolic compounds are also present in lignin resulting from incomplete monolignol biosynthesis (91).

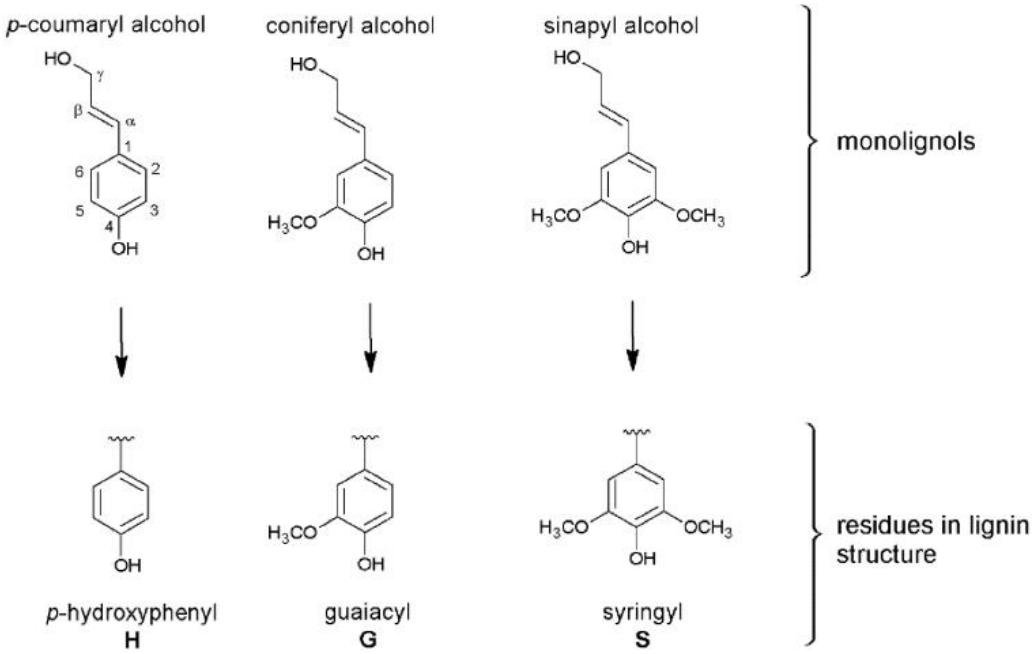


Figure 1.4 Lignin monolignol structures. Sourced from Korte *et al.* 2021 (92).

Lignin represents the most abundant reservoir of aromatic carbon on the planet and despite being seen often as a barrier to access the polysaccharides it protects, it has demonstrated its potential for generating value from an underutilised substrate (93). The paper and pulping industry generates large amounts of kraft and organosolv lignin annually. These are low molecular weight lignin containing chemical pulps with low value, and which derive their names from the chemical lignocellulose treatments that generate them (94). These lignin containing pulps have already demonstrated potential value through their use in products in such as asphalt and rubber (95, 96). Kraft lignin polymerised into high molecular weight chains also has shown potential as adhesives, carbon fibres, and thermally stable polyesters (97). Vanillin, a versatile compound with use in fragrances, pharmaceuticals, and synthetic polymers, has historically been industrially isolated from kraft lignin for decades (98).

Lignin can also be the source of a vast range of aromatic-based chemicals depending on the feedstock, depolymerisation techniques, fractionation techniques, and transformation routes (86). Low-molecular weight lignins may serve as replacements for bisphenol A in the production of epoxy resins (99). Monomeric and oligomeric aromatic compounds derived from lignin have been investigated for their potential to generate platform chemicals. Catechol for example, can be converted further into polymeric materials but is also a precursor for pesticides, flavourings and fragrances (50). Multiple aromatics currently produced from fossil fuels including benzene, toluene, xylene, and phenols have been explored for potential production from lignin (100). Various aromatic acids and aldehydes from lignin have potential in commodity chemical markets, and ring opened organic acids could be recovered from lignin as high value products (90).

Like the polysaccharides within lignocellulose, there is huge potential for the development of novel biomaterials from lignin also. The incorporation of lignin into polymers has been explored due to its UV stabilisation, flame-retardant, lubricant, and pigment-adding properties (86). The reliance on fossil fuels for plastic production was discussed earlier, and lignin has shown promise for the synthesis of thermoplastic elastomers, polymeric foams, and membranes with comparative properties to those derived from fossil fuels (90).

1.3.3. Minor components

Aside from the polysaccharide and lignin networks, other components are present in smaller proportions. Pectins account for only a small proportion of the polysaccharides in lignocellulosic material but these contribute to the biomass recalcitrance. Pectic polysaccharides are highly complex and exist in four known subclasses: homogalacturonan, rhamnogalacturonan I and II, and xylogalacturonan (101). These polysaccharides are embedded in plant cell walls, and their high level of acetylation prevents enzyme access to the cellulose and hemicellulose and limit release of these sugars during hydrolysis (102). Although present in small amounts in lignocellulose, pectic polysaccharides still represent an additional source of sugars for either biofuel production, or production of chemicals as discussed above. Pectin itself could also be a valuable product derived from lignocellulose. For example, it is currently used as a gelling agent in food items, there is evidence that treatment of fruit with pectin derived oligosaccharides can improve fruit properties, and also pectin as a supplement has multiple reported health benefits including blood cholesterol reduction and cancer cell growth inhibition (101).

The fraction of the biomass which is not combustible is known as ash and this varies depending on biomass types but generally contains silicon, potassium, calcium, sulphur, and chlorine in the most abundance (103). Proteins are present at low levels in the plant cell wall

where they may be cross-linked with lignin and polysaccharides (65). Outside of the lignin network, other phenolic compounds and polymers are also present in plant cell walls. For example, tannins are polyphenolic compounds with glucose cores and are high-value chemicals due to their role in leather making and as nutritional supplements (104).

1.4. The biorefinery goal

Biorefineries are the industrial facilities that convert renewable biological resources into products. In the context of lignocellulose these resources are the variety of woody plant biomass feedstocks discussed earlier, and the products are the range of fuels, chemicals, and materials explored also. Biorefinery development and the translation of research into biorefinery optimisation has accelerated for reasons such as energy security, environmental concerns, and the need for a sustainable and renewable supply of chemicals and materials (90). Biorefineries traditionally consist of four core sections: feedstock acquisition, pretreatment, hydrolysis, and fermentation. Lignocellulosic refineries are currently working to develop procedures that allow more complete conversion and valorisation of the entire substrate. The focus will shift instead to separation of the lignocellulosic components for various downstream conversion processes, and depending on the target products of interest these four core sections may change (50). Sustainable and profitable development of lignocellulose biorefineries has grown in the past few decades, but still faces several challenges (105).

In 2017, there were 224 biorefineries operating across Europe but only 43 of these were utilising lignocellulose (106). With the European union pushing for a more consolidated network of lignocellulosic biorefineries in the future, improvements need to be made to compensate for the lack of efficiency compared to the first-generation biorefineries which are predominant in Europe. The major barrier to achieving this shift is to improve the efficiency of conversion of such a recalcitrant substrate. This conversion needs to find a balance between being efficient, cost-effective, and environmentally friendly. This has not yet been achieved effectively at the scale required for the envisioned facilities (106). Aside from improving conversion technologies, there are several challenges to overcome to address all the issues associated with a movement away from fossil fuels and towards lignocellulose.

Feedstock selection needs to be optimised while still meeting the criteria to be considered a sustainable biomass supply. In Europe the feedstock must not be used for food or animal feed, must not be needed to maintain ecological function e.g. naturally occurring forestry residues, must be grown on marginal land or land not suitable for food production, and where possible should be available in close proximity to the proposed biorefinery (50). Using regional biomass reduces the environmental impact of transportation, and local biomass production could also provide alternative sources of income for populations in the vicinity of the biorefinery (50). Studies have demonstrated that refineries utilising Miscanthus feedstocks have a range of up to 50 kilometres, after which acquisition begins to influence economic and environmental sustainability (57). Feedstock security is an important

consideration of lignocellulosic biorefineries. Strategies such as supplementing primary feedstock supply with other sources of lignocellulosic materials will be important to maintaining this security (107). Selection of feedstocks more amenable to conversion is an option to increase biorefinery efficiency. For example, it has been demonstrated that spruce wood from trees with beetle infestations provided higher yields of fermentable sugars than from healthy trees (108). Lodgepole pine killed by mountain pine beetles was found to be easier to delignify than healthy lodgepole pine wood, however this increased digestibility lead to increased losses of fermentable sugars during lignin removal (109). This highlights once more the importance of advancing the biorefinery technology for fractionation of total lignocellulose rather than aggressive removal of lignin. Considering lignocellulose from agricultural wastes and dedicated biomass crops, it has been suggested that selective breeding or genetic engineering could be used to improve feedstock digestibility (66). Lignocellulosic biorefineries in the future will likely have to utilise multiple feedstocks, however this complicates conversion processes further (106).

As addressed earlier, lignin has historically been viewed as a barrier to accessing the polysaccharide components of lignocellulose. However, the potential value locked away within the recalcitrant lignin network is clear and it is now obvious that technology for deconstruction and conversion of lignin must improve. Development of these technologies has many advantages for biorefinery processes, it could improve the efficiency of which polysaccharides are isolated for traditional lignocellulose conversion processes, value itself could be generated from the lignin to improve overall profitability, and utilising the lignin effectively can help prevent issues with downstream processing of remaining components (105). There are also multiple environmental benefits from the improvement of lignocellulose biorefining technology. The paper and pulp industry produce several types of pollutant-containing wastewater, with the main pollutant being lignin. This results in dark coloured wastewater which prevents light from reaching lower depths in aquatic systems, leading to reduced photosynthesis and anaerobic conditions that conclude in toxic bodies of water (110). Aside from pollution reduction through utilisation, the maturation of lignocellulose conversion technologies could also lead to reduced environmental damage resulting from biorefinery operation. Less energy intensive processes and more environmentally friendly approaches would potentially limit the release of pollutants currently released in biorefinery processes. These include sulphur dioxide, nitrogen dioxide, and fine particulate matter (111). In the US it has been demonstrated that residents living within ten kilometres of biorefineries suffered significantly more from respiratory illness compared to control populations (111). An evaluation of the potential air pollution released from feasible biorefinery designs suggests that pollutant levels would exceed the US thresholds as part of the Clean Air Act and that

additional technology would be required to limit these releases (112). More efficient methods of lignocellulose separation and deconstruction will help to mitigate this.

For biorefineries to operate as effectively as possible, it is likely that a combination of cost-reduction strategies, co-production of value-added products, valorisation of waste products, and financial subsidies will be required (69). Technology to achieve more effective separation of lignin to generate additional value has been gradually improving but still requires major advancements. For example, some successful approaches have envisioned lignin-first biorefineries where the initial focus is the fractionation of lignin to leave the polysaccharide fractions unaffected for later conversion (113). In this way, the traditional view of polysaccharide utilisation first followed by valorisation of remaining lignin has been flipped. However, equally important will be the development of additional biorefinery facilities for the downstream conversion of the separated lignin which is not a practice currently adopted by the majority of biorefineries (90). The design of lignocellulosic biorefineries will have to be carefully considered to maximise efficiency. It has been suggested that a modular design will aid total conversion of feedstocks into various value streams, which is now clearly the strategy lignocellulosic biorefineries will need to adopt to compete with fossil fuels (50). In Malaysia, investigations of biorefineries found the distances between facilities within a refinery plant to be an important factor in determining overall profitability (114). However, all these considerations will be futile if the efficiency of current conversion technologies is not addressed first. Clearly the main bottleneck currently restricting the ability of biorefineries to operate effectively and achieve total substrate valorisation is the lack of efficient separation and degradation technologies for lignocellulose (105). Indeed, this has been the main research focus of improving biorefinery processes, as it is clear that current methods are not effective enough (63).

1.5. Pretreatments of lignocellulose

Pre-treatments are processes which make biomass more amenable to deconstruction. They do so by opening and rearranging the structure and changing the chemical composition of the biomass. For the total substrate valorisation described earlier in future biorefineries, these treatments need to be able to break open the rigid structure of lignocellulose so that cellulose, hemicellulose, and lignin molecules can all be accessed. However, the development of biorefinery processes including pre-treatments in the past has focused on accessing the polysaccharides for biofuel production. Therefore, the lignin has been seen as a barrier and the traditional pre-treatment approaches aim to remove this without regard to downstream lignin applications. The basic pre-treatment strategy has been focused on removing the lignin barrier, or at least rearranging it, to access the cellulosic sugars. Any retained hemicellulosic sugars are a bonus to potentially increase bioethanol yields (78). However as discussed in length previously, economic factors affecting the final product in this setup demonstrate the need to utilise and refine all fractions of lignocellulose into high value final products (78).

Separation of lignin from other components of lignocellulose is massively important due to its recalcitrance. Remaining lignin associated with the polysaccharides forms a physical barrier that restricts access to the polysaccharides, restricts fibre swelling, and non-productively binds and inactivates cellulases, all of which prevent efficient valorisation of polysaccharide components (94). Pre-treatments can be grouped into physical, chemical, or biological pre-treatments.

1.5.1. Physical pretreatment

Physical pretreatments enhance the access of enzymes to lignocellulosic components in the later hydrolysis steps by increasing porosity and swelling of lignocellulose fibres (94). They preserve virtually all components of lignocellulosic biomass which will be important in optimising the economics of biorefineries (94). However, retaining the lignin connected to the polysaccharides also presents an issue for effective hydrolysis of polysaccharides and valorisation downstream. Many forms of physical pretreatments exist, using a variety of techniques and generating different effects and results.

The most basic form of physical pretreatment is mechanical grinding, milling, or refining. In this technique the particle size of lignocellulosic biomass can be physically reduced with little change in the overall composition of substrate (94, 115). Ball milling for example, has shown promise in grinding lignocellulosic biomass into small particles. It produces low levels of fermentation inhibitors, reduces the loss of hemicellulose, and results in only minor structural changes to the lignin (115). However, longer milling times have been found to result in the

formation of globular structures that reduce enzyme accessibility in downstream processes which therefore reduces efficiency (115).

Steam explosion utilises high pressure steam at high temperatures to reduce the size of lignocellulosic biomass particles while decreasing crystallinity (116). Temperatures of 130 to 300 °C are employed for up to 10 minutes followed by a sudden release of pressure which causes the reduction in size (94). Waste steam generated in other facilities of a biorefinery could be used in this process, however the temperature, time, and pH all need to be carefully chosen based on the biomass type and the desired downstream products (117). Ammonia fibre explosion uses essentially the same method but more effectively deconstructs the lignin network (116). Uniquely, ammonia explosion can cleave ester linkages between lignin and carbohydrate complexes which achieves some separation between components (118).

Electron beam irradiation is a simple operation that relies on high energy electrons to create reactive radical species within biomass (119). These radicals take part in reactions within the lignocellulosic biomass which reduce polymerisation and crystallinity of cellulose and alter the matrix between the lignin and the polysaccharides (119). This approach has a short processing time, a large processing capacity, and no chemical requirements which reduces environmental effects (120). Photocatalysis is a similar technique which again relies on the generation of highly reactive radical species. Ultraviolet (UV) or near UV radiation is used to photoexcite semiconductor catalysts in the presence of oxygen to generate radicals then used in lignocellulose deconstruction (121).

Heating of lignocellulose is a strategy used to remove lignin and decrease the crystallinity of cellulose. Liquid hot water treatment maintains water temperature at 160 to 240 °C, resulting in lignin depolymerisation and hemicellulose solubilisation while retaining cellulose (122). Microwave-assisted pretreatments directly heat biomass via dipole polarisation and ionic conduction which is more energy efficient than traditional heating of biomass (123). The major benefit of microwave heating compared to traditional heating is the uniform heating of the substrate, which indeed leads to increased digestibility (123). The major disadvantage of heating approaches and indeed of physical pretreatments in general are their high power consumption and relatively low efficiency (49). The variety of physical techniques and their differing results suggests that an optimal physical pretreatment strategy will rely on a combination of multiple treatments. It has been demonstrated that synergistic physical approaches have beneficial effects (119).

1.5.2. Chemical pretreatment

Chemical treatments aim to achieve decomposition of lignocellulosic biomass through chemical reactions in aqueous solutions, solvents, or other liquids. Perhaps the most widely used chemical pretreatments are acid and alkaline pretreatments (124). In acid pretreatments, acids and high temperatures are used to break strong chemical bonds in hemicelluloses to expose the cellulose (51). High concentrations of acid can be used to produce highly digestible solid biomass, but this requires expensive corrosion-resistant equipment (125). High acid concentrations can also lead to increased formation of compounds which inhibit the later hydrolysis and fermentation steps, loss of carbohydrates, and promotion of condensation reactions between lignin breakdown products that make separation of lignin more difficult downstream (125). Dilute acid pretreatments can be used instead and it has been demonstrated that concentrations of 2 – 4 % sulphuric acid at 121 °C can significantly reduce the hemicellulose content in wheat straw to expose the cellulose (51). However, these dilute acid treatments require high temperatures which comes at an energy input and environmental cost. The high temperatures can also cause lignin droplets to form and deposit on the surface of the resulting cellulose which prevent enzyme access during downstream processes (125). In contrast, alkaline pretreatments use low concentrations of alkalis such as sodium hydroxide, calcium hydroxide, ammonia, or lime to remove the lignin component of a substrate (126). For example, 2 – 4 % sodium hydroxide with heating to 121 °C can significantly reduce lignin content and hemicellulose content while leaving the cellulose component more intact than in acid pre-treatment (51).

Instead of aqueous solutions, organic solvents have instead been explored for chemical pretreatment of lignocellulose. This involves cooking the woody plant biomass in a mixture of water with organic solvents to produce three distinct fractions: a cellulose rich pulp, a lignin rich solid, and a hemicellulose rich liquid (78). A diverse selection of solvents can be used such as ethanol, methanol, acetone, glycerol, ethylene glycol, organic acids, phenols, and ketones (127). The organosolv process is well established in the paper and pulp industry and reduces the use of hazardous substances compared to kraft pulping discussed earlier (78). With the view of utilising all components of lignocellulose, the organosolv process also produces a lignin which is sulphur-free, rich in functional components, more homogenous, and includes minimal carbohydrate contamination (127). Therefore, the lignin obtained from organic solvents is isolated as a pure product rather than a by-product.

Other forms of liquids have gathered popularity recently for the pretreatment of lignocellulose. Ionic liquids are ionic salts in a liquid state, for example sodium chloride is an ionic salt at room temperature but will melt to an ionic liquid at 801 °C (128). Some ionic liquids exist which form at room temperature e.g., imidazolium salts and pyrrolidinium salts

which have high thermal stabilities, low vapour pressure, and no flammability (129). The high solubility of lignocellulosic biomass in these ionic liquids means that they can be used to effectively fractionate the substrate into cellulose, hemicellulose, and lignin (129). However ionic liquids suffer from their high costs and the need to recycle pure ionic liquid (128). An alternative could be deep eutectic solvents which have characteristics of both organic solvents and ionic liquids. These solvents are prepared from hydrogen bond donors and acceptors which form a liquid driven by hydrogen bond interactions (113). Lignocellulose crystallinity can be reduced using these solvents which are cheap and easy to fabricate, do not need purification, are insensitive to water, and have been demonstrated to fractionate lignocellulose (113).

These chemical approaches more effectively fractionate the components compared to physical approaches, but also result in deconstruction or modification of lignin structures. This may be completely appropriate for biorefineries with downstream targets stemming from a modified and partially deconstructed form of lignin but may be unsuitable for refineries wishing to obtain lignin as close to the form it takes in the plant cell wall. Additionally, all physical and chemical pretreatments tend to require expensive inputs in either chemicals or machinery, have high energy demands, and often produce environmentally damaging waste streams. Therefore, there is interest in developing more economic and green approaches to lignocellulose deconstruction.

1.5.3. Biological pretreatment

Biological methods of pretreatment utilise microorganisms themselves or the enzymes they produce to deconstruct lignocellulose. These biological approaches, if realised, could represent significant advances in biorefinery development due to their low energy input requirements and the ambient conditions at which they could theoretically operate (49). Organisms exist that can mineralise lignocellulose i.e., degrade all components completely and so enzymes could be identified to degrade, potentially to convert, and therefore to valorise all components of lignocellulose. This will be vital to securing the economic future of the biorefineries.

One biological approach to is to utilise microorganisms themselves for lignocellulose treatment. The rapid growth of bacteria and their relative ease of genetic manipulation make them attractive candidates for bacteria-assisted pretreatments. For example, enzymatic hydrolysis of rice straw was improved after incubation for three days with the lignin-degrading bacteria *Cupriavidus basilensis* B-8 (125). In addition, the lignolytic bacterium *Pandoraea* sp. B-6 was found to selectively remove residual lignin remaining after acid pretreatment to give an increased sugar yield at hydrolysis of around 40 % (130). In the

same study, saccharolytic bacteria were found to selectively reduce hemicellulose content while leaving the lignin content unchanged (130). This demonstrates the potential to tailor the biological approach depending on downstream targets. Many other bacterial species have been isolated from various sources and demonstrated to have lignolytic activity (131). Some bacteria have been demonstrated to deconstruct lignin in anaerobic conditions, and there is potential for the incorporation of these strains into anaerobic digesters converting lignocellulose to biogas (131).

Fungal pretreatment offers an environmentally friendly and low-cost pretreatment for enhancing hydrolysis and fermentation steps. The fungus *Pleurotus floridanus* LIPIMC996 was able to increase the saccharification of empty fruit bunches from oil palm by 4.5 times after a four week incubation and this resulted in a higher ethanol yield obtained from simultaneous saccharification and fermentation (54). Another study screened 26 fungi for their potential to improve sugar yield after a 28-day incubation period with milled corn stover (55). Eight of these fungal strains demonstrated selective deconstruction of lignin compared to the polysaccharides, and treatment with three of these strains led to the highest sugar yield after saccharification (55). Although not technically used for pretreatment of the lignocellulose, the ability of some fungi to prepare hydrolysates after pretreatments for fermentation has been assessed. The fungus *Coniochaeta ligniaria* NRRL30616 is tolerant to inhibitor compounds often created during physical and chemical pretreatment processes and was demonstrated here to improve downstream fermentation through metabolism of these compounds (132). Fungal pretreatment has two major disadvantages: low efficiency and long residence periods (55). In the study which showed improved digestibility of empty fruit bunches from oil palm after fungal pretreatment, it was simultaneously demonstrated that phosphoric acid pretreatment was still much more effective (54).

Biological pre-treatments are often combined with physical or chemical pretreatments, and it is possible that this will be required in a biorefinery. Care needs to be taken when combining physical and chemical approaches with a biological pretreatment. Digestibility of empty fruit bunches from oil palm was higher after phosphoric acid treatment compared to a combined acid and fungal treatment due to a larger loss of material when the treatments were combined (54). Ideally, a single pretreatment step minimises complexity and reduces costs. Biological approaches could have the most potential to achieve this in an environmentally friendly manner, but a simple processing step such as milling or grinding may be required to increase the overall efficiency. Despite the capital and running costs associated with this processing, the increased efficiency may be beneficial to both the economics of a biorefinery and its environmental footprint.

While discussing microorganisms capable of deconstructing lignocellulose, it is also important to consider consolidated bioprocessing. In this process microorganisms are used which can utilise lignocellulosic feedstocks, saccharify the polysaccharides to simple sugars, and then simultaneously convert these to biofuel (79). Investigations into multiple microorganisms as candidates for consolidated bioprocessing have been undertaken. The lignocellulose degrading fungus *Fusarium oxysporum* was genetically manipulated to improve its ethanol yield when grown on lignocellulose (79, 133). The thought process behind such work is to improve upon the fermentation process which typically uses *S. cerevisiae.* The major drawbacks of this yeast are that it relies on a pretreatment step to prepare the sugars for fermentation and that it cannot naturally ferment pentose sugars such as xylose released from hemicellulose (79). Metabolic engineering of *S. cerevisiae* to impart the ability for pentose sugar fermentation, in addition to the search for novel strains with boosted ethanol production has been performed (48, 134). Other microorganisms, including bacteria such as *Bacillus subtilis, Zymomonas mobilis*, and *Clostridium thermocellum* have also been explored for their potential use in consolidated bioprocessing (135-137). However, these studies all have the view of utilising only the polysaccharides for fuel. There has been some interest in developing consolidated bioprocessing for lignin valorisation. Many bacterial strains that can depolymerise lignin and catabolise breakdown products have been identified and suggested as candidates in the development of such processes (138). The ability of *Pseudomonas putida* strain A514 to produce polyhydroxyalkanoate from lignin derivatives has been demonstrated and improved through genetic engineering (139). Again though, these processes are one-dimensional in their ability to valorise lignocellulose and inflexible in their adaptation to different feedstocks and products.

Applying a microbial consortium instead of a single species is attractive, especially for large scale industrial applications where sterility is an issue. These approaches do not require sterile conditions, they display a more diversified enzymatic arsenal with high metabolic diversity, and they remain stable and robust across a range of conditions avoiding issues such as feedback inhibition (140). Microbial consortia from a range of sources have been investigated and characterised for their ability to degrade lignocellulose, but fewer have demonstrated an ability to generate value. Perhaps the most well-characterised process utilising microbial consortia is the production of methane biogas in anaerobic digesters. Rumen-derived microbial consortia have been demonstrated to enhance anaerobic digestion of untreated wheat straw which results in increased methane yield (141). Wheat straw was also used as a feedstock for growth of a dairy cow rumen-derived microbial consortium to produce volatile fatty acids, the building blocks for polyhydroxyalkanoate which has potential use as a bioplastic (140). Cattle-stomach derived consortia have been demonstrated to

produce the commodity chemical lactic acid from the breakdown of sulphuric acid pretreated corn stover (142). The disadvantage of these consortia is that they are often complex communities with a high number of culturable and unculturable species. This makes disentangling the community interactions and functioning difficult, understanding the degradation processes complicated, and therefore scaling to an industrial level unachievable (143). It has been suggested that simpler synthetic consortia could be attractive instead, however lignin degradation has been found to improve with increased microbial diversity and therefore lignin utilisation through these approaches may be unattainable (143). Additionally, these techniques tend to focus on degradation of the entire substrate solely to produce a single product. Indeed, the consortia are often evolved or selected with a single purpose in mind.

An additional major disadvantage of any strategy using microorganisms is the loss of carbohydrates during microbial growth (144). Although selective lignin or polysaccharide degradation has been demonstrated, in the vision of an optimal biorefinery all components will require valorisation (125, 130). It is therefore more attractive to use the enzymes themselves combined in an optimal cocktail to effectively deconstruct and separate lignocellulose into different streams, all of which can be further transformed and utilised for profit. Lignocellulose degrading enzymes can be isolated from a wide range of sources. Plant biomass-degrading microorganisms, such as those discussed already, can be isolated from environments such as leaf decaying microbial communities (145), lignocellulose-containing herbivore dung (146), or from industrial waste sources (147). The lignocellulose degrading enzyme repertoires of plant pathogens have also been investigated such as food crop pathogens (133, 148, 149), or forestry pathogens (150). Enzymes from microbiomes of wood digesting arthropods like woodlice (151), termites (152), and beetles (153), or from grass-consuming ruminant microbiomes (154, 155), or even from primate microbiomes have been investigated and identified (156). The plant biomass degrading repertoire of organisms which do not require microbiomes and instead rely on endogenous lignocellulose-degrading enzymes have also been investigated such as terrestrial (157) and marine invertebrates (158, 159). Perhaps the most powerful of all these systems are the lignocellulose degrading repertoires of fungi.

1.6. Fungal lignocellulose breakdown

Much research has focused on fungal lignocellulose degradation as they are major players in the carbon cycles of forest ecosystems and efficient degraders of plant biomass (160, 161). Across the fungal kingdom, a wide range of strategies are used in the deconstruction of lignocellulose. Traditionally wood-degrading fungi have been classified into brown-rot, white-rot, and soft-rot fungi based on their growth substrate preferences and wood-decaying strategies (162). Basidiomycetes occupy various ecological niches and belong to the white- and brown-rot fungal classes (163). Brown-rot basidiomycetes effectively solubilise and degrade the polysaccharide components whereas white-rot fungi are capable of complete mineralisation of all lignocellulose components (164, 165). Brown-rot fungi are characterised by their use of non-enzymatic radical-based reactions that generate strong oxidants for the degradation of cellulose and hemicellulose with only minor changes to the lignin (166). These fungi are major invaders of forest biomass and are reported to have evolved from white-rot fungi by losing some capacity for lignin metabolism (162). Brown-rots are also characterised by their two-stage pattern of lignocellulose breakdown; firstly the lignocellulose structure is opened through oxidative chelator-mediated Fenton (CMF) reactions which generate hydroxyl radicals that attack lignocellulose, and secondly polysaccharide degrading enzymes are produced that depolymerise carbohydrates ultimately to monosaccharides (167). In contrast, white-rot fungi are considered to be the most efficient wood degraders in nature because of their array of lignin-degrading oxidative enzymes and hydrolytic enzymes for the complete mineralisation of lignocellulose (168). They are common inhabitants of fallen trees and forest litter where they play an important role in carbon recycling (169).

Despite their presence in almost all forest systems, ascomycete wood degraders have been less well investigated (170). Ascomycete lignocellulose degraders, often called soft-rot fungi, are characterised by their penetration of secondary plant cell walls by fungal hyphae where they secrete complex mixtures of degradative enzymes directly at the site of attack (162). Some ascomycetes have been well characterised for their plant biomass degrading capabilities and have been used extensively in industry. *Trichoderma reesei* is a mesophilic fungus well known for its ability to produce polysaccharide-degrading enzymes in high abundances (171). The species has benefited from decades of research and strain improvement and most cocktails of enzymes used for the industrial conversion of lignocellulose are based on enzymes secreted by *T. reesei* (172). *F. oxysporum* is a phytopathogenic fungus containing an expanded repertoire of lignocellulose degrading enzymes for its plant cell wall invading lifestyle (173). The characterisation of *F. oxysporum* is in part due to its pathogenicity of multiple important food crops but also due to industrial

interest in developing a consolidated bioprocess with the ascomycete because of its ability to ferment sugars from lignocellulose breakdown directly into ethanol (79, 174).

However, this view of classifying fungal wood-degraders is beginning to change and it has been demonstrated that the enzymatic repertoires for lignocellulose breakdown can vary greatly from species to species, and even within the same classification (162). Indeed, some basidiomycete genomes have demonstrated the capacity for degradation for all lignocellulosic components like white-rot fungi, but also lack genes encoding class II fungal peroxidases for lignin metabolism like brown-rot fungi (163). Within white-rot fungi there is variation in the strategy of decay, with some species simultaneously degrading all lignocellulosic components and others removing the lignin in advance (169). Additionally within soft-rot fungi some species have demonstrated effective cellulose degradation but a lack of lignin breakdown, but others show significant lignin depolymerisation and metabolism (164). More recently sequenced basidiomycete genomes have also demonstrated genomic capacities for lignocellulose breakdown that are indicative of both white- and brown-rot fungi (175). Modes of decay *in vivo* have also been reported to differ greatly between fungi of the same wood-degrading classification (175). The white-, brown-, and soft-rot classifications no longer represent the variety of wood decay mechanisms utilised by fungi and it is therefore becoming apparent that the diverse metabolic capabilities for lignin breakdown are no longer unique to white-rot fungi, as was once thought (169).

Around thirty years ago, the first attempts to classify enzymes involved in lignocellulose breakdown into families occurred (176). Soon after, the inception of the carbohydrate active enzyme (CAZyme) database aided the functional classification of the variety of enzymes utilised by fungi and other organisms for this process (177). The database relies on amino acid similarity of constitutive modules within an unknown protein to that of biochemically characterised enzymes (177). Primary protein sequence data can be annotated with CAZyme domains quickly using online tools to predict function based on modular sequence similarity to other entries in the database. For example, the online tool dbCAN was developed in 2012 and utilises a database of hidden Markov models (HMMs), family specific signature domains, for the quick and easy annotation of sequence data (178). CAZyme proteins can be assigned to the following catalytic classes: auxiliary activity (AA), carbohydrate esterase (CE), glycoside hydrolase (GH), glycosyl transferase (GT), and polysaccharide lyase (PL). Additionally, many proteins have carbohydrate binding modules (CBMs).

1.6.1. Auxiliary activity (AA)

The most recently developed CAZyme class is the AA class of CAZymes. This class groups together the oxidative enzymes involved in lignin and polysaccharide breakdown and was added to fully encompass plant cell wall deconstructive enzymes within the CAZyme database (179). Included within the AA class are the lytic polysaccharide monooxygenase (LPMO) enzymes. The AA9 (formerly GH61) family contains fungal LPMOs that act on polysaccharides including cellulose, xylan, glucans, starch, and pectin (180, 181). They enhance polysaccharide degradation by providing new chain ends for attack by hydrolytic CAZymes (182). The oxidative action of these copper-containing enzymes relies on exogenous electron donors such as other AA family CAZymes, small molecule reductants and even lignin (179, 182), and more recently it has been demonstrated the LPMOs utilise hydrogen peroxide (H_2O_2) as a cosubstrate (183). In the presence of such an electron donor, LPMOs catalyse oxidative cleavage of glycosidic bonds through hydroxylation of either the C1 or C4 carbon in the glycosidic linkage (**Figure 1.5**) (184). There are additional LPMO families within the AA class of CAZymes and these include the AA10 bacterial LPMOs, the AA11 fungal chitin active LPMOs, and AA13 starch degrading LPMOs (180). The AA16 LPMOs are interesting due to their recent discovery, atypical product profile, and potentially different mode of activation (172). AA3 family CAZymes are FAD-containing flavoproteins belonging to the glucose-methanol-choline (GMC) oxidoreductase family which includes activities such as cellobiose dehydrogenase, glucose-1-oxidase, aryl alcohol oxidase, alcohol oxidase, and pyranose oxidase (185). It is an attractive idea that flavin binding oxidative enzymes of this family play a role in supplying H_2O_2 to LPMOs and peroxidases or in the production of radicals that degrade lignocellulose through Fenton chemistry (183). Additionally, the AA7 glucooligosaccharide oxidases and AA12 pyrroloquinoline quinone-dependent oxidoreductases both have been implicated in LPMO systems through H_2O_2 production (186, 187).

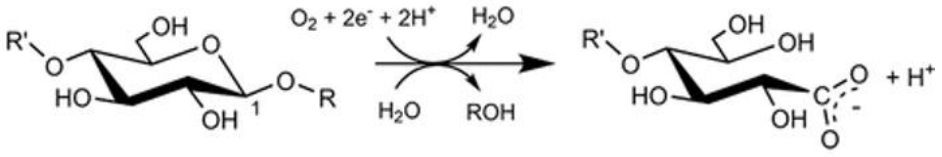


Figure 1.5 LPMO catalysed oxidation of a C1 carbon in a glycosidic linkage. Adapted from Wang *et al.* 2019 (188).

The AA1 family of CAZymes contains laccase and multi-copper oxidase enzymes which oxidatively cleave various aromatic substrates with the co-reduction of oxygen to water (189). Laccases can directly oxidise phenolic targets in lignin or utilise small molecule mediators for non-phenolic moieties (**Figure 1.6**) (190). The active radicals formed by this oxidation results in the cleavage of alkyl-aryl, C_α-C_β, and C_α-C_1 linkages within lignin or the recoupling and modification of the lignin structure through repolymerisation (190-192). The release of low molecular weight lignin polymers from biomass by laccases can also boost LPMO activity through their action as electron donors (193). These laccase-catalysed reactions are particularly efficient due to their use of readily available molecular oxygen as the final electron acceptor (191, 192).

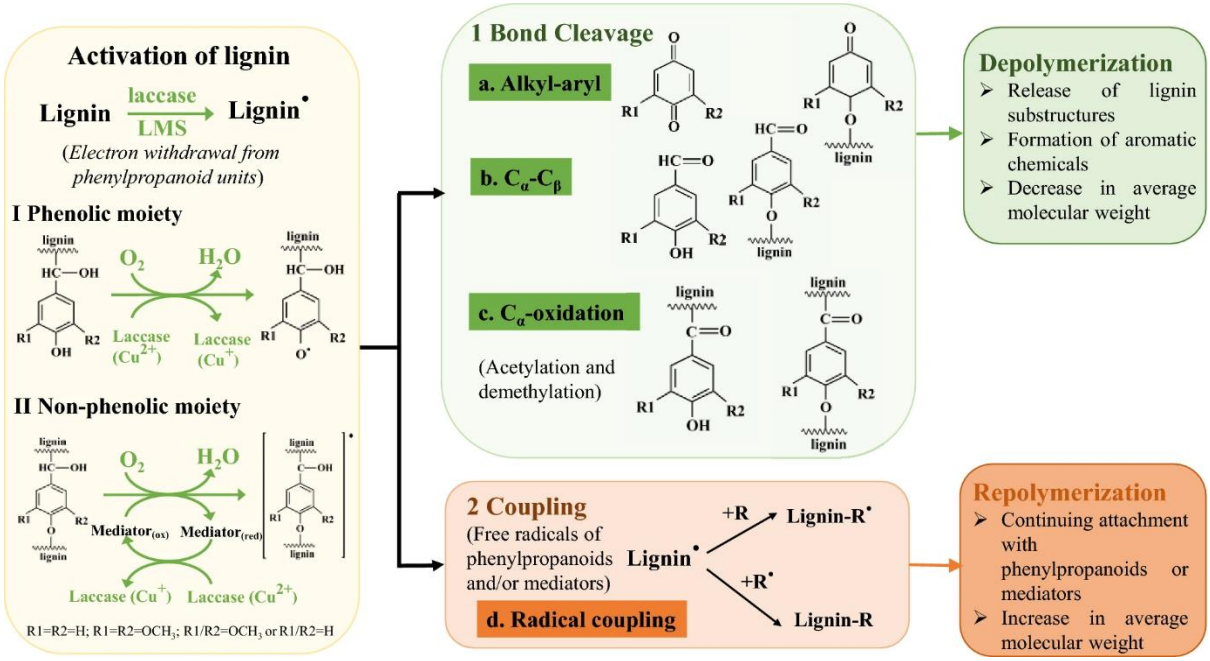


Figure 1.6. Laccase catalysed lignin breakdown and modification. Sourced from Zhou *et al.* 2023 (190).

The AA2 family of peroxidase enzymes also play a major role in oxidative lignin deconstruction. These enzymes are lacking in the traditionally classified brown-rot fungi, due to their non-lignolytic specialisation of substrate degradation (194). AA2 family peroxidases are heme-containing CAZymes that use H_2O_2 or organic peroxides as electron acceptors to catalyse various oxidative reactions (179). These class II peroxidases are divided into three lignolytic forms: lignin peroxidase (LiP), manganese peroxidase (MnP), and versatile peroxidase (VP) (**Figure 1.7**) (195). LiPs oxidise phenolic aromatic substrates and cleave carbon-carbon bonds and ether bonds in non-phenolic aromatic substrates through the utilisation of a donor substrate such as veratryl alcohol (VA), MnPs oxidise Mn(II) to Mn(III) which results in secondary oxidation of various phenolic compounds and lignin polymers, and VPs combine the catalytic properties of LiPs and MnPs (179).

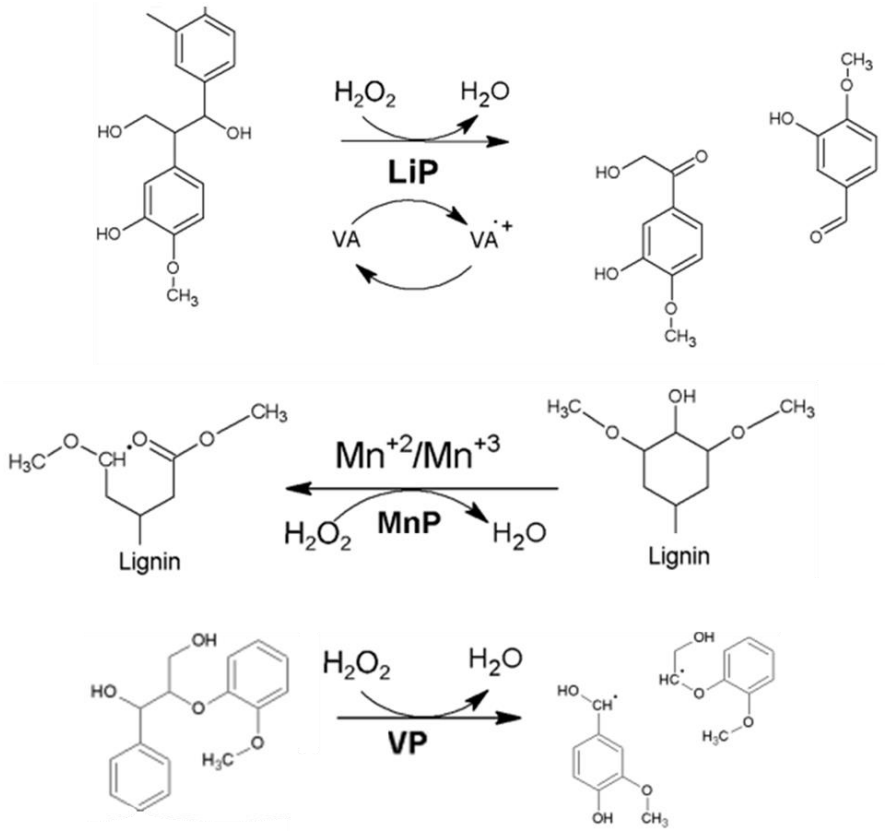


Figure 1.7 Peroxidase catalysed lignin breakdown and modification. Adapted from Janusz *et al.* 2017 (196).

AA8 family members contain iron reductase domains and are found in the N-terminal regions of cellobiose dehydrogenase enzymes but have also been discovered independently and appended to CBMs (179, 197, 198). It has been suggested that their generation of reactive hydroxyl radicals may play a role in indirect depolymerisation of lignin. The AA7 and AA12 families discussed as part of LPMO systems also potentially play an indirect role in lignin degradation through the production of H_2O_2 for Fenton reactions (199). Vanillyl-alcohol oxidases form the AA4 family and can catalyse the conversion of a wide range of phenolic oligomeric compounds (200). They may act downstream in lignin metabolism.

1.6.2. Carbohydrate binding module (CBM)

CBM domains are structurally discrete modules within an enzyme sequence that provide binding capacity to carbohydrate ligands (201). These domains are typically associated with hydrolytic CAZymes, most commonly GH class, but have also been identified in oxidative LPMOs (202-204). They are non-catalytic binding domains predominantly involved in binding plant cell wall polymers to prolong contact of catalytic domains with the substrates to enhance enzyme efficiency (205). It has also been suggested that CBMs could also aid polysaccharide deconstruction through cell-wall disruption (205). This hypothesis was supported for the chitin-binding CBM33 family which has been demonstrated disrupt the

structure of chitin and increase accessibility for chitinases (206). However it was later demonstrated that this protein was in fact a polysaccharide oxidase and support for this hypothesis remains limited (207). Many of the CBM families have been reported in fungal genomes, and variation in CBM profiles in fungal genomes likely reflects their different lifestyles (202).

Various CBM families provide binding capabilities to a diverse range of cell wall components including crystalline cellulose, non-crystalline cellulose, chitin, β-1,3-glucans, mixed linkage glycans, xylan, mannan, galactan, and starch, with some CBM families demonstrating promiscuous binding to multiple polymers (208). An alternative mode of CBM classification has been used to group these modules into three types. Type A are surface binding CBMs recognising crystalline substrates, type B bind to internal regions of glycan chains (*endo*-type), and type C bind to the ends of glycan chains (*exo*-type) (207). The specificity of a CBM may not necessarily reflect the specificity of its associated catalytic module as many cellulose binding CBMs are associated with catalytic domains active on any of the components within the plant cell wall (205). Not all CBMs provide binding capabilities for the purpose of plant cell wall deconstruction. Some CBM families such as CBM5, are involved in binding chitin and are likely involved in cell wall remodelling (209). Other CBM families like CBM43 and CBM56 are involved in binding β-1,3-glucans which are present in fungal cell walls suggesting a potential role in fungal growth processes, however these polymers are also present in plant-based substrates and could be targeted by these CBM families for breakdown (210, 211).

1.6.3. Carbohydrate esterase (CE)

As the name suggests, these CAZymes catalyse the cleavage of ester linkages to remove ester-based modifications from carbohydrates (201). The CE class of CAZymes contains a wide range of substrate specificities with families acting on xylan, chitin, peptidoglycan, feruloyl linkages, and pectin (201). Despite targeting different substrates CE class CAZymes often share high sequence similarity, and enzymes of the CE1 family have been identified with both acetyl xylan esterase and feruloyl esterase activity (212). Regarding fungal breakdown of lignocellulose, CE families which act on hemicellulose and pectin are of particular interest.

Most hemicellulosic xylans are acetylated, in addition to mannans and glucomannans also present in plant cell walls (213). Ester-based modifications along polysaccharide backbones facilitate association of hemicelluloses with the hydrophilic surfaces of cellulose (214). By removing acetyl groups from polysaccharides within plant cell walls, access is greatly improved for other lignocellulose-degrading enzymes which improves saccharification (201).

Acetyl xylan esterases are widely distributed across many CE CAZyme families and their removal of acetyl groups from xylan has been found to improve both xylan hydrolysis and cellulose accessibility (213). Additionally, ester bonds between 4-*O*-methyl-D-glucuronic acid side chain decorations and phenylpropane subunits in lignin form linkages between the polysaccharide and aromatic networks (214). These linkages are known as lignin-carbohydrate complexes (LCCs) and contribute to the recalcitrance of lignocellulose (**Figure 1.8**) (215). CE15 glucuronoyl esterases are an example of a CE family that act on these linkages, and the addition of bacterial CE15 enzymes to commercial cocktails of enzymes has been shown to significantly boost saccharification of lignocellulosic biomass (216). Therefore, there is industrial value in understanding CE class enzyme activity and function, and in identifying new CE class CAZymes.

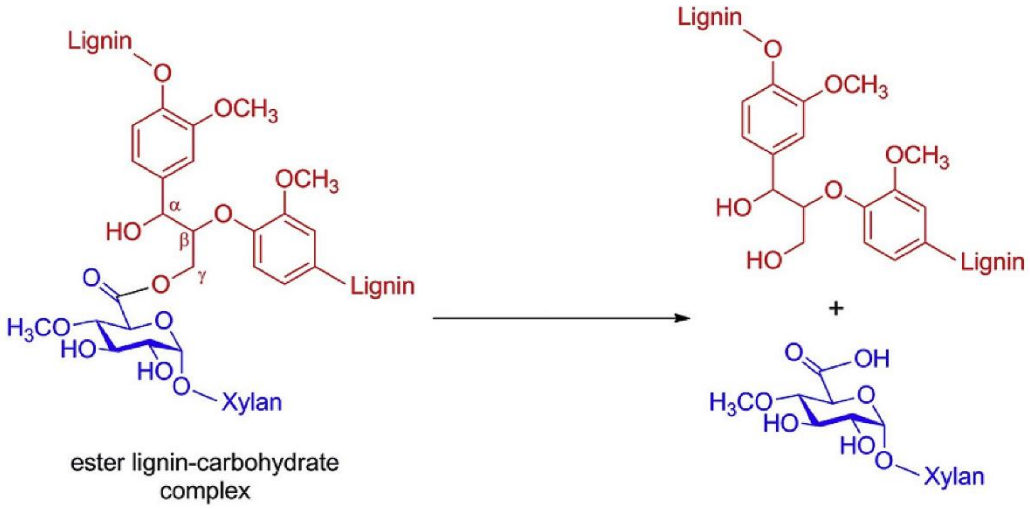


Figure 1.8 LCC cleavage. Adapted from Zhao *et al.* 2020 (217).

A CE family worth noting in the context of fungal lignocellulose breakdown is the CE10 family, as the inclusion of this family is currently challenged. The CE10 family mainly contains proteins demonstrated to act on ester-based modifications of non-carbohydrate compounds (218). However, these enzymes are widespread in fungi and have been observed in fungal secretomes when grown on lignocellulosic substrates compared to simple sugars (219, 220). Much of this family has not been demonstrated to be active on carbohydrates, and it has been suggested that they potentially play a role in lignin modification (220). When comparing CE10s across the genomes of multiple oomycetes, it was observed that CE10s were lost in the genomes of biotrophic species compared to necrotrophic species (218). Considering the lifestyle of necrotrophic species this also implies a potential role in the breakdown of plant biomass, however the exact function remains unclear.

1.6.4. Glycoside hydrolase (GH)

GH class CAZymes are abundant in the genomes of almost all living organisms and play a range of diverse roles in metabolism, antibacterial defence, and pathogenesis (221). They catalyse the hydrolysis of glycosidic bonds within carbohydrates and are perhaps the most well-characterised class of CAZymes in part due to their action on the most abundant polysaccharides on the planet (201). The history of GH characterisation is also reflected in this class having the largest number of families, 181 at the time of writing. Classification into families is based on amino acid sequence similarity, however different families may share catalytic mechanisms and structures and can also be assigned into clans (221). Another feature by which GH CAZymes can be categorised is whether they use a retaining or inverting reaction mechanism, where the configuration of the anomeric carbon of a released glycoside is either retained or inverted (221).

Another factor influencing GH family classification is substrate specificity and there are numerous GH families active against plant cell wall polymers, in addition to families that target the breakdown products from these polymers into monomers (201). For cellulose breakdown for example, fungi secrete an arsenal of endo- and exo-acting GH class cellulases in addition to β-glucosidases (222). Endo-cellulases will bind along the length of a cellulose molecule and cleave the β-1,4-glycosidic bond connecting monomers of glucose to create new chain ends, exo-cellulases bind at these new ends to release individual glucoses or short oligosaccharides, and β-glucosidases break these oligosaccharides down into cellobiose and glucose monomers (222). Similar fungal enzymes are released for the breakdown of all other polysaccharides in lignocellulosic materials, highlighting the diversity of this class of CAZyme. As the main enzymes responsible for the degradation of the main polysaccharide chains within lignocellulose, there has been much interest in discovery of new efficient GH CAZymes to improve biorefinery processes (223).

1.6.5. Glycosyl transferase (GT)

The second largest class of CAZymes are the GT class CAZymes which catalyse the formation of glycosidic linkages to form glycosides by utilising an activated donor sugar phosphate to transfer glycosyl groups to specific acceptor molecules (201). These acceptor molecules can include lipids, proteins, heterocyclic compounds, and other carbohydrate residues to create a range of potentially valuable products including glycoconjugates, oligosaccharides, and polysaccharides (221). Like GH class CAZymes they are present in almost all living organisms and predominantly play roles in biological processes outside of lignocellulose metabolism such as cell signalling, cell adhesion, cancer, and cell wall biosynthesis (221). The donor sugar substrates are most commonly activated nucleoside

diphosphate sugars, but nucleoside monophosphate sugars, lipid phosphates, and unsubstituted phosphate can also be used (224). Similarly to GH CAZymes, GTs have two main reaction mechanisms, retaining or inverting the stereochemistry of the anomeric centre (221). In contrast to the diversity of structural folds across GH families, GT class CAZymes tend to adopt one of three folds (221). GT-A and GT-B enzymes both contain two Rossmann-like domains, but these are less closely associated in GT-B enzymes, and GT-C enzymes adopt instead a mainly α-helical structure (221, 224).

The extensive catalogue of identified and characterised GT class CAZymes is impressive considering that they are typically intracellular membrane-bound proteins which makes them difficult to work with (221). This intracellular nature means that the GT class CAZymes are less relevant than other classes for extracellular lignocellulose deconstruction. However, some GT class CAZymes have been identified as extracellular proteins bound to the fungal cell surface (225). It has been suggested that in this capacity GT class CAZymes may be involved in association of fungal biomass with plant substrates to aid pathogenicity (225, 226). GT class CAZymes therefore play a role in fungal deconstruction of lignocellulose, however these enzymes are not of interest for the enzyme-based biological treatment of lignocellulose in an industrial context.

1.6.6. Polysaccharide lyase (PL)

Aside from the hydrolytic GH class CAZymes which degrade polysaccharides, there is the PL class of CAZymes specific to uronic-acid containing polysaccharides (227). PL enzymes employ a β-elimination mechanism to break the oxygen to carbon bond at the C4 position, as oppose to hydrolysis at the C1 position for hydrolases, to yield a 4,5-unsaturated sugar at the new non-reducing end of the product (227). The fewer number of PL families compared to GH families reflects the much smaller diversity in the molecular chemistry of PL substrates. The uronic acid containing substrates targeted across the PL families include pectin and oligo-galacturonans, rhamnogalacturonans, and alginates and glycosaminoglycans (227). PLs can also be classified by enzymatic mechanisms: a metal-dependent mechanism mainly utilised in pectin degradation, and the histidine/tyrosine mechanism used by families involved in *syn* and/or *anti* β-elimination (227).

PL CAZymes are widespread in fungi and present in the genomes of both ascomycete and basidiomycete wood-degrading fungi (202). However, the numbers of PL encoding genes and the families that these sequences belong to have been demonstrated to vary greatly between species (228). Together with the other CAZyme classes, these enzymes all work in concert to effectively deconstruct all components of lignocellulose with a variety of reaction mechanisms used and a variety of breakdown products created. Studying fungal

lignocellulose breakdown is therefore important for understanding deconstruction of entire lignocellulosic substrates in addition to identifying enzymes and strategies for the total valorisation of plant biomass in biorefineries. It is equally important to approach these goals through the investigation of a diverse range of fungi, and not just historically well-researched species.

1.7. Introduction to *Parascedosporium putredinis* NO1

The growth strategy of ascomycete wood-degraders, which involves penetration of plant cell walls with secretion of large quantities of enzymes from filamentous hyphae, makes them interesting candidates for the exploration of fungal lignocellulose breakdown (229). Although some ascomycetes have been extensively explored for their polysaccharide degrading capacity, as discussed earlier, their potential for lignin degradation is less clear. It has been demonstrated previously that ascomycetes can modify and degrade lignin, but the mechanisms by which this occurs still requires elucidation (146, 229).

The first fungal isolate of the genus *Parascedosporium* was initially introduced based on a strain of *Graphium tectonae* due to their similarity in conidia physiology (230). This genus belongs to the Microascaceae family of the ascomycete phylum (229). *G. tectonae* was initially isolated from the seeds of the tropical hardwood tree *Tectona grandis* and was reclassified as *Parascedosporium tectonae (230).* Later re-examination of this isolate considered this to instead be a synonym and it was reclassified again to *Parascedosporium putredinis* (231). The species has also been isolated from the fruit pods of the tropical evergreen plant Royal Poinciana or *Delonix regia* in Thailand, an environment rich in lignocellulosic residues (232).

Parascedosporium putredinis strain NO1 was isolated from a mixed microbial community growing on wheat straw enriched compost, where it thrived in the later stages of decomposition (229). Within the fungal population of this community, the abundance of internal transcribed spacer (ITS) reads assigned to the genus *Graphium* began to increase at week four and dominated the community by the eighth week of incubation, representing 84 % of the read abundances assigned to fungi. *P. putredinis* NO1 is a synamorph of *Graphium* and was readily cultivated from the community. *P. putredinis* NO1 was instantly very interesting to investigate for lignocellulose breakdown and enzyme discovery for two reasons: its dominance of the fungal population at a late stage in the culture indicated a potential utilisation of the more recalcitrant components of lignocellulose once the more readily accessible carbon sources have already been removed by other community members, and it could be selectively cultivated by growing the fungus on agar plates containing kraft lignin as the sole carbon source. Growth of *P. putredinis* NO1 alone in liquid cultures containing wheat straw was investigated and the gene expression of CAZymes belonging to the catalytic classes AA, CE, GH, and PL were found to be upregulated compared to growth with glucose. These CAZymes were identified in the extracellular proteome of these cultures through harvesting of supernatant proteins and proteins bound to the insoluble wheat straw substrate using a biotin-labelling approach (233). Within this

extracellular proteome, cellulose and hemicellulose degrading CAZymes were well represented and CE1 family CAZymes with sequence similarity to feruloyl esterases involved in breaking bonds at the lignin-carbohydrate interface were also detected. In addition to these hydrolytic enzymes, 69 AA class CAZymes were detected and included LPMO and LPMO-supporting enzymes, laccases and multi-copper oxidases, and peroxidases. A combination of the under-exploration of the lignin degradation by ascomycetes, the lifestyle of *P. putredinis* NO1 in the mixed microbial community, and the production of many oxidative lignin degrading enzymes by this fungus made it an interesting candidate to explore for new enzymes.

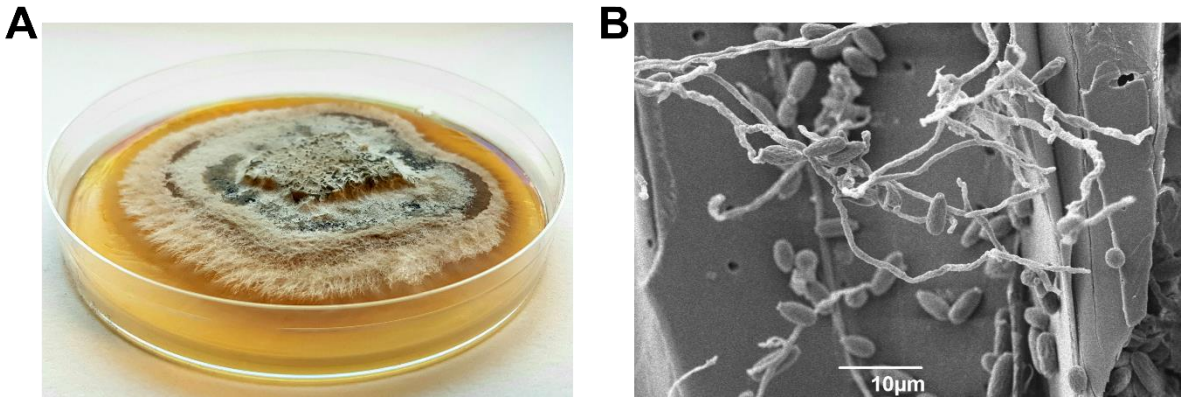


Figure 1.9 *Parascedosporium putredinis* NO1. *P. putredinis* NO1 growing on potato dextrose agar (**A**) and growing on wheat straw imaged by SEM from Oates, 2016 (234) (**B**).

Of the various bonds within the lignin network, the aryl glycerol β-*O*-4 aryl ether bond is the most abundant at around 45 – 60 % of the total linkages (235). The ability of the *P. putredinis* NO1 extracellular proteome to cleave the β-O-4 linkage in a synthetic lignin model compound was demonstrated when the fungus was grown on wheat straw but not on glucose (229). Through fractionation of the extracellular proteome a single protein was isolated with a central tyrosinase domain. Fungal tyrosinases have been implicated in promiscuous cleavage activity of ether linkages in lignin previously (236). Within monocot lignin, the flavonoid-type compound tricin is thought to act as an initiation point for lignification of plant cell walls (237). Tricin has reported antibacterial, antifungal, and insecticidal activity, and has potential pharmaceutical value due to potential anti-inflammatory and antioxidant properties (237). Incubation of this purified new oxidase enzyme with wheat straw resulted in the release of tricin and the corresponding reduced tricin levels in enzyme treated wheat straw was confirmed by NMR. Finally, incubation of the enzyme with wheat straw also boosted sugar release after subsequent digestion with cellulase enzymes. This study highlights the potential to identify new enzymes from *P. putredinis* NO1 that are active on the most recalcitrant components of lignocellulose, and which demonstrate the potential to contribute to the valorisation of lignocellulosic feedstocks.

Although not discussed in detail, an interesting aspect of the *P. putredinis* NO1 investigations is the ability of the fungus to grow on kraft lignin as the sole carbon source. The differences observed in the *P. putredinis* NO1 extracellular proteome when grown on wheat straw compared to glucose could also suggest a potential ability for this fungus to tailor its secretome to various substrates. Indeed, differences in fungal extracellular proteomes when grown on different lignocellulosic substrates have been explored extensively previously (160, 163, 229, 238). Often this is for the purpose of characterising the growth of fungi on specific substrates, or to identify proteins produced in abundance on complex lignocellulosic substrates versus simple sugars. However, understanding the level to which these differences are tailored by the fungus and how these change across multiple industrially relevant plant biomass feedstocks could be important for biorefineries. The ability of *P. putredinis* NO1 to grow on wheat straw and kraft lignin suggests the possibility of designing experiments to explore a potentially tailored enzymatic response, while identifying new enzymes with roles in lignin breakdown.

1.8. Tailored enzymatic cocktails

The complex and recalcitrant nature of lignocellulosic biomass demands multiple different enzymes to effectively deconstruct and convert the substrate. Enzymes can be mixed into "cocktails" to achieve this. The enzymes used in cocktails for the deconstruction of biomass are traditionally derived from fungi, and in particular from species of *Trichoderma* and *Aspergillus*, and strains of these species have undergone decades of research, evolution, and improvement to enhance hydrolytic enzyme production (239). Most of this research has focused on improving cellulose hydrolysis through the incorporation of cellobiohydrolases, endoglucanases, and β-glucosidases to achieve higher sugar yields for bioethanol production (240). Although, incorporation of hemicellulose degrading enzymes has been explored as well (241, 242). More recently, LPMOs for oxidative cleavage of cellulose have also been added (243). Most of the knowledge and technology for developing and producing enzyme cocktails for industrial processes is unfortunately locked away within the private sector. Although the small number of companies now have remarkable expertise and have served to rapidly accelerate the reduction in cost in enzyme cocktails, there are major drawbacks to the dominance of a handful of private companies in this field (239). Firstly, it is very difficult for any small companies with potentially novel technologies and strategies to establish a foothold in the industry, and secondly the enzyme cocktail contents are not available to the public or even included in the patent literature and this will restrain progress in enzyme cocktail development (239).

Most cocktails used for the industrial conversion of lignocellulose have been based on the enzymes produced by *T. reesei*, and this lack of diversity has become a bottleneck for cocktail improvement (172). As discussed at length earlier, for biorefineries to achieve total substrate valorisation cost-effectively, optimisation will be required. This will be determined by factors such as biorefinery location, available feedstocks, downstream conversion technology, and target products. This will introduce variation between biorefineries that demand optimisation including at the biological pretreatment step. Historically, most enzyme-cocktail development has been based on the biological digestion of dilute acid-treated corn stover for bioethanol production (239).

Single commercial mixtures cannot be optimised to cover the wide range of lignocellulosic feedstocks available. For example, hemicellulose degrading activity has found to be lacking in some commercial cocktails and these may not be appropriate for feedstocks with higher proportions of hemicellulose otherwise this reservoir of polysaccharides would not be effectively accessed (241). Lignocellulosic biomass also varies in composition depending on its growth conditions, with hardwoods showing variation when grown in temperate vs tropical

conditions (76). An alternative to commercial cocktails could be the proteins harvested from fungi grown on different substrates, where the enzyme content will be much more tailored to the substrate in question (244). Indeed, the effectiveness of enzyme-blends from *Trichoderma* isolates to release sugars from giant reed, miscanthus, and switchgrass with either acid or alkali pretreatment has been compared to commercial cocktails (244). The crude enzyme extract from one of these isolates effectively saccharified the same substrates used for growth significantly more than a commercial mix, and this effect varied with pre-treatment. A cellulose-degrading cocktail of nine enzymes from the thermophilic fungus *Myceliophthora thermophila* demonstrated remarkable differences in saccharification potential across different substrates, from 26 % conversion in spruce to 7.4 % in birch (243).

Although a solely biological approach is theoretically the most-environmentally friendly method of pretreatment, it is likely that biological pretreatments will be combined with other physiochemical methods as well. Several technologies have been discussed here earlier and show promise, and it is likely that the enzyme cocktails used will have to be optimised for each. Indeed, in the above study the ability of *Trichoderma* isolates to release sugars from the substrates which were used to grow and harvest the isolates varied depending on whether an acid or alkali pretreatment was used and whether this was a mild or a strong pretreatment (244). Saccharification of sugar cane bagasse with a simple three component cellulose degrading cocktail was demonstrated to achieve higher yields for acid-treated sugar cane bagasse compared to alkali-treated sugar cane bagasse (245). Experimenting with the ratios of these three enzymes resulted in biomass from each pretreatment requiring different ratios for optimal sugar conversion (245). In complex cocktails with large numbers of enzymes, identifying these ratios may be very difficult.

In the context of bioethanol production, good sugar yields have been achieved through the application of commercial cocktails to a wide range of lignocellulosic feedstocks (242). However purely achieving high sugar yields for bioethanol production is unlikely to be an economically feasible strategy for biorefineries. Cocktails need to appropriately prepare a substrate for complete valorisation into multiple products. There has been investigations into improving the lignin-degrading capabilities of enzyme cocktails, however these still focus on removal of lignin to increase the efficiency of polysaccharide conversion and not lignin valorisation (199, 246).

The ideal efficient enzyme cocktail should be highly active on the intended biomass type, appropriate for the valorisation strategy, operate without necessity for an earlier physiochemical pretreatment or operate alongside a mild pretreatment if completely necessary, withstand the process stress, and be cost-effective to produce and acquire (240). This will depend on the combination of individual enzymes and their ratios in the final

cocktail (240). Understanding plant biomass-degrading fungi continues to contribute to our knowledge of lignocellulose breakdown. Bioprospecting of new organisms like *P. putredinis* NO1 for new enzymes involved in lignocellulose breakdown but also understanding how enzymes act in concert *in vivo* on different substrates could help to educate the design of tailored enzymatic cocktails. Deeper exploration of enzyme combinations for breakdown of various substrates will also help to identify enzymes which are unimportant depending on the application and optimise enzyme production, which will help to reduce the cost of enzymes for biological pretreatments (239). It will also be important to understand lignocellulose breakdown temporally, and it has been suggested that different enzyme cocktails will be optimal at different stages of breakdown (242). This is an interesting concept for total substrate valorisation where an initial enzyme cocktail could be applied to fractionate a component of lignocellulose which could then be separated and valorised, while potentially making the remaining components more accessible for deconstruction by a different cocktail.

Enzymes have now been isolated from diverse organisms including fungi and have been demonstrated to improve existing cocktails or to compete with existing cocktails when combined (199, 238, 240). Again, these effects often vary depending on substrate (238). This demonstrates the value in identifying new enzymes to provide a variety of resources as these cocktails are developed and optimised to different substrates. Of particular importance will be the identification and incorporation of new lignin degrading enzymes, as it has become obvious that this underutilised, yet abundant polymer requires more focus in these cocktails.

1.9. New lignin degrading enzyme identification

To achieve the total substrate valorisation discussed in detail here already, the incorporation of new lignin-degrading activities into tailored cocktails will be vital. The ability of fungi to degrade lignin has been extensively discussed, however only a handful of lignin degrading activities from fungi have been well-characterised. These are the laccase and peroxidase enzymes discussed earlier which have been assigned to families within the AA class of CAZymes (179, 246). To improve cocktails for lignocellulose breakdown, the identification of new lignin degrading enzymes and activities which have not yet been assigned to CAZyme families are required.

The overall structure and composition in lignin and its variation across lignocellulosic feedstocks has been discussed. Within this substrate, there are multiple bonds and structures which are possible sites of attack for lignin-degrading enzymes. The most predominant is the arylglycerol β-*O*-4 aryl ether bond accounting for 45 – 60 % of the linkages, however other bonds include β-5 phenyl coumaran, β-β' pinoresinol, 5-5' biphenyl, 4-*O*-5' diaryl ether, α-aryl ether, 5-5' dibenzodioxicine and β-1 diphenyl ethane (**Figure 1.10**) (235, 247). The laccase and peroxidase enzymes non-selectively attack these bonds through their oxidative activities which generate radicals that can initiate C_{α}–C_{β} cleavage, alky-aryl cleavage, cross-linking, demethoxylation, and aromatic ring cleavage (138). For lignin valorisation the non-specific nature of these enzymes raises an issue, as the phenoxy radicals generated during lignin oxidation can result in spontaneous repolymerisation or recondensation to form new high molecular weight structures (93). It is therefore likely that a combination of these enzymes with more specific enzymes will be required for the conversion of lignin. Other less well characterised enzymes have been investigated for their potential roles in lignin breakdown.

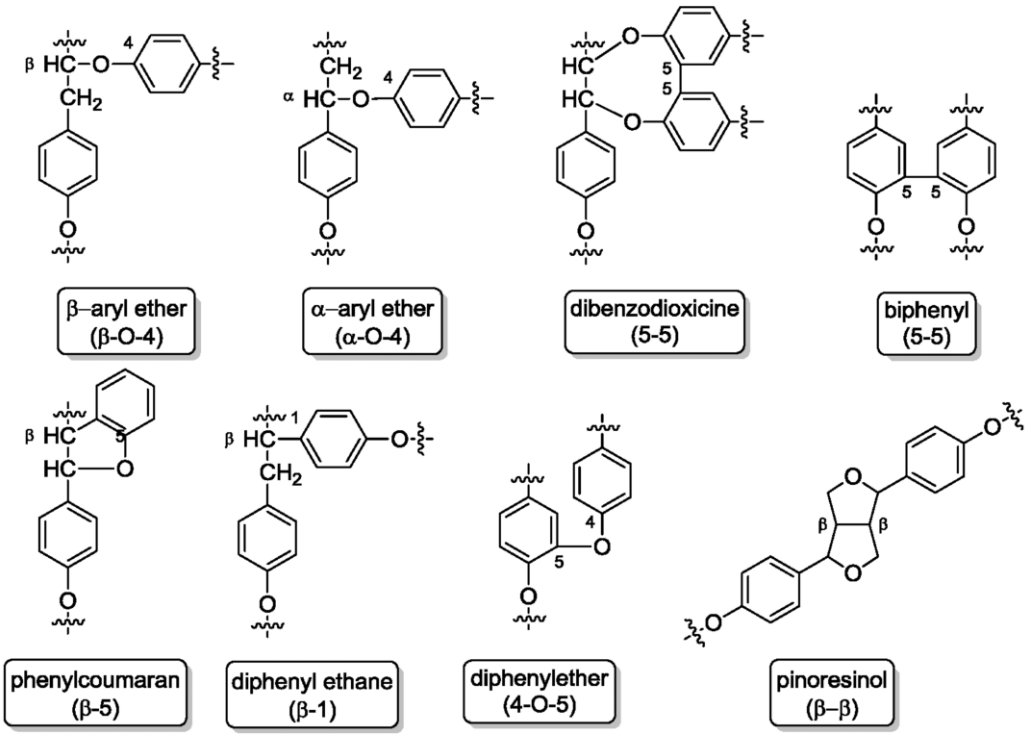


Figure 1.10 Linkages within lignin. Sourced from Witzler *et al.* 2018 (247).

Chloroperoxidases are interesting heme- or vanadium-containing peroxidase enzymes that display activities similar to halogenase, peroxidase, catalase, and cytochrome P450 enzymes *in vitro* (248). It has been suggested that during lignocellulose degradation, filamentous fungi may chlorinate lignin directly and chloroperoxidase activity has been detected in leaf litter containing soils (249). Fungal chloroperoxidases catalyse the oxidation of chloride ions in the presence of H_2O_2 to reactive chlorine electrophiles that result in the chlorination of lignin but also various aromatic structures (249). Aromatic rings of lignin in aspen wood were chlorinated when cultivated by the ascomycete wood-degrader *Curvularia inaequalis* (249). Additionally, a model lignin dimer representing the major non-phenolic structure in lignin was demonstrated to be chlorinated by heme- and vanadium-containing chloroperoxidases, but also cleaved the dimer (250). In the same study, the heme-containing CPO was demonstrated to depolymerise a synthetic guaiacyl lignin polymer (250). It is possible that this depolymerisation occurs through cleavage of ether linkages within lignin, as this activity has been reported for various aromatic substrates previously (251).

Tyrosinases have shown potential involvement in lignin breakdown through their activity on phenolic compounds. By using oxygen as the final electron acceptor tyrosinase enzymes can *o*-hydroxylate monophenols to diphenols through cresolase activity or oxidise diphenols to quinones through catecholase activity (236). Tyrosinase from *Agaricus bisporus* has been demonstrated to oxidise veratryl alcohol, a commonly used non-phenolic substrate for assaying lignolytic activity, however with much lower catalytic efficiency than a LiP from

Phanerochaete chryosporium (236). The same enzyme was demonstrated to cleave two dimeric linin model compounds at C_α–C_β and β-*O*-4 linkages. However, the oxidative activity of tyrosinase toward total lignin from wood was found to be negligible when compared to laccase and peroxidase enzymes (252). The characterisation of new lignin modifying and degrading activities is important, however the non-specificity of these enzymes remains an issue. Investigations have identified some enzymes with the potential to selectively cleave lignin bonds and these could allow the development of more tailored lignin conversion.

The phenol oxidase with β-ether linkage cleaving activity identified from *P. putredinis* NO1 and discussed in detail previously was found to have a central domain with similarity to tyrosinase enzymes (229). β-etherase enzymes belong to the glutathione-*S*-transferase superfamily, are represented in the genomes of prokaryotes and eukaryotes, and catalyse the reductive cleavage of β-*O*-4 bonds (235). The ability of a β-etherase from the basidiomycete fungus *Dichomitus squalens* has been demonstrated to selectively cleave the β-*O*-4 bond in a dimeric lignin compound and when incubated with polymeric lignin shifted the molecular weight distribution to lower molecular weights, signifying depolymerisation (235). However, this enzyme and many of the bacterial β-etherases are intracellular and therefore the biological role for these enzymes in deconstruction of lignocellulose is unclear. Despite this, treatments of lignin with these enzymes has been shown to release monomeric units from lignin and therefore the value for industrial processes is clear (93). Aside from these few examples the identification of new lignin degrading activities is extremely limited, especially for specific activities.

Despite most enzymes used in the industrial conversion of lignocellulose being sourced from fungi already, the fungal kingdom is vast, and many interesting bioactivities remain unexplored and underexploited (222). Identification of new enzymes from fungi rather than other microorganisms is valuable to industry as these proteins are often stabilised by glycosylation and can remain active even in the presence of surfactants or at high temperatures (222). Around 1200 new fungal species are identified every year and the approximately 100,000 species already described only represent less than 10 % of the total species of fungi that exist (222). From these newly discovered species, work can begin to understand their genomes, characterise their lignocellulose degrading ability, and identify target enzymes for purification and investigation.

A classic approach to the identification of enzyme activities from fungi is to harvest either fungal lysates in the case of intracellular activities, or “secretomes” for extracellular activities and screen the ability of these enzyme mixtures for activity against model compounds. Indeed, a β-etherase activity was identified from *Chaetomium* sp. 2BW-1 using this approach but the individual protein responsible was not isolated (253). This approach was

also used to identify the new phenol oxidase with β-ether cleaving ability from *P. putredinis* NO1 and through combination with fractionation and proteomics, the single protein responsible was isolated (229). Another molecular technique for enzyme identification includes cloning cDNA libraries from fungi of interest when grown in conditions which are suspected to produce a desired activity. For example, many cellulases have been identified during the growth of fungi on lignocellulosic substrates (222). However, these approaches are limited by their design as only single specific activities are revealed with these strategies.

Omics technologies can now provide a wealth of information on the biology of individual organisms or provide insights into whole communities with meta-omics. The genomic sequencing and assembly of new unexplored fungal genomes alongside the meta-genomic sequencing of DNA in whole lignocellulose-degrading communities can provide new opportunities for the identification of new lignin degrading enzyme sequences (222). A traditional sequence searching approach is to utilise the sequences of already well-characterised enzymes to search for homologous sequences in the new sequence data (254). Alternatively, the domain architectures and sequence motifs of already known enzymes can be used to build HMMs that can be used to search new sequences that match the models (255). Peptide pattern recognition is a technique that clusters protein sequences based on patterns of small peptide patterns generated from a set of proteins sharing the same activity (256). This technique has been used to identify new putative β-etherase sequences within public databases, however here it identified the same set of proteins as basic sequence homology searching (256). Unfortunately, these approaches rely on primary amino acid sequence and therefore will only identify enzymes with some sequence similarity and therefore likely some evolutionary link. Although this can be useful for finding new enzymes within a class of activities that may have altered substrate-specificity, temperature and pH optima, and activity profiles etc. it is difficult to identify completely novel activities with this approach. The recent emergence of AlphaFold has provided an accessible resource for the structural prediction of proteins from primary amino acid sequences (257). This could allow a novel approach for exploring sequence data, where structural prediction of new sequences could be used for comparisons with either experimentally determined or bioinformatically predicted structures of enzymes know to be involved in lignin breakdown.

Although genomic approaches can provide a huge amount of sequence information, this huge amount of data can make finding the relevant protein encoding genes difficult. Studies of the RNA instead provides information on the genes being actively expressed and the levels of expression correspond to the abundance of the transcripts. These studies are known as transcriptomics and are commonly used when characterising the lignocellulose-degrading lifestyles of fungi and for identifying new enzymes. For example, sequencing the

transcriptome of Antarctic ascomycete *Aspergillus sydowii* strain MS-19 revealed laccase and peroxidase enzymes with activity maintained as low as 0 °C (258). Often transcriptomic investigations into lignocellulose breakdown are designed to examine gene expression when a microorganism is grown on simple carbon sources versus a lignocellulosic substrate (167, 229, 259, 260). However, these studies often focus on characterisation of fungal physiology through analysis of upregulated predicted CAZymes already known to be involved in lignocellulose breakdown and do not explore the potential for new enzymes (167, 260). Transcriptional analysis of growth of an aquatic ascomycete *Clavariopsis aquatica* on lignocellulosic substrates demonstrated upregulation of multiple laccase and peroxidase genes but revealed many cytochrome P450 family enzymes to be upregulated also (261). However, no discussion of the potential functions of these enzymes was included.

When considering the identification of new proteins implicated in lignocellulose and lignin breakdown, it may be more pertinent to focus on identifying and characterising the proteins produced by lignocellulose-degrading fungi. Examining proteins through proteomic technologies gives direct insight into an organism or a community at a functional level, as proteins are the key players in biological processes. Proteomics has been used in attempts to characterise the plant biomass deconstructing proteomes of many industrially relevant strains of fungi (262). These proteomes have been investigated across a range of substrates, culture conditions, and strains. Proteomics has also been used to characterise newly identified fungi during lignocellulose degradation (150, 263). Indeed, the recently discovered AA16 family of LPMOs was discovered through the analysis of the secretome of *Aspergillus aculeatus* (172). However, recent reports on the identification of lignin degrading enzymes from fungi using proteomics, or any omic technology, are difficult to find.

With the wealth of data which could be obtained for lignocellulose degrading fungi using these technologies it seems wasteful to not focus efforts on new enzyme identification, particularly new lignin modifying or degrading enzymes given the clear necessity for developing biological lignin conversion technologies. One potential bottleneck is the lack of mature bioinformatic workflows for the analysis of omics data with the aim of new enzyme identification. For example, even proteomic experiments targeting extracellular proteins secreted by fungi are heavily contaminated by intracellular proteins through cell death and lysis occurring in the experimental conditions (264). This complicates identification of interesting proteins within datasets and reduces the effectiveness of comparative secretomic investigations, which will be important for understanding fungal growth responses to different lignocellulosic substrates for the development of the tailored cocktails discussed earlier.

1.10. Aims of this thesis

The lignocellulose degrading ascomycete *P. putredinis* NO1 was chosen for investigation due to the lignin utilisation lifestyle demonstrated previously, and due to the identification of a new industrially valuable lignin degrading phenol oxidase from this fungus (229). Omics technologies were used alongside the development of new bioinformatic approaches and workflows to attempt to improve the understanding of the lignocellulose-degrading potential of *P. putredinis* NO1. A newly sequenced genome provided a novel sequence dataset which could be explored using a new structural-based searching approach for potentially interesting lignocellulose degrading enzymes. From the genomic sequences, an *in silico* secretome was generated using a new bioinformatic workflow for secretome prediction. This workflow is readily applicable to other fungi and prokaryotes for secretome prediction. These bioinformatic workflows allowed more effective investigations of *P. putredinis* NO1 and simplified the identification of potentially interesting sequences, but also aimed to demonstrate how the wealth of sequence data available can be explored for any function of interest.

Utilising the genomic sequence as an effective reference, a proteomic approach was used to define the proteins produced by *P. putredinis* NO1 when grown on a range of industrially relevant substrates. Variation in the *P. putredinis* NO1 secretome was more effectively investigated using the secretome isolation workflow, and through combination with molecular techniques a tailored secretome was demonstrated. Understanding tailored enzymatic responses will be important when developing tailored enzymatic cocktails to be used industrially for the conversion of lignocellulose. The tailored secretome also suggested the ability to design investigations with *P. putredinis* NO1 to encourage differential production of lignin-degrading enzymes. This aided the identification of candidate proteins to investigate for potentially new lignin modifying or degrading activities.

Through a combination of omic investigations, bioinformatic workflows, and molecular techniques, new proteins from *P. putredinis* NO1 were explored which could be massively valuable additions to the resources available for attempts to achieve the ultimate goal of complete lignocellulose valorisation.

Chapter 2

2. Whole genome structural predictions reveal hidden diversity in putative oxidative enzymes of the lignocellulose degrading ascomycete *Parascedosporium putredinis* NO1

Authors

Conor JR Scott[a], Daniel R Leadbeater[a], Nicola C Oates[a], Sally R James[b], Katherine Newling[b], Yi Li[b], Nicholas GS McGregor[c], Susannah Bird[a], Neil C Bruce[a]

Affiliations

[a] Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, United Kingdom

[b] Bioscience Technology Facility, Department of Biology, University of York, York YO10 5DD, United Kingdom

[c] York Structural Biology Laboratory, Department of Chemistry, The University of York, York, YO10 5DD, United Kingdom

Available online at: https://doi.org/10.1128/spectrum.01035-23

2.1. Abstract

Economic valorisation of lignocellulose is paramount to realising a true circular bioeconomy; however, this requires the development of systems and processes to expand the repertoire of bioproducts beyond current renewable fuels, chemicals, and sustainable materials. *Parascedosporium putredinis* NO1 is an ascomycete that thrived at the later stages of a wheat-straw composting community culture, indicating a propensity to degrade recalcitrant lignin-enriched biomass, but exists within an underrepresented and underexplored fungal lineage. This strain has proven an exciting candidate for the identification of new enzymes targeting recalcitrant components of lignocellulose following the recent discovery of a new lignin β-ether linkage cleaving enzyme.

The first genome for the genus *Parascedosporium* for *P. putredinis* NO1 genome was sequenced, assembled, and annotated. The genome is 39 Mb in size, consisting of 21 contigs annotated to contain 9,998 protein-coding sequences. The carbohydrate-active enzyme (CAZyme) repertoire was compared to 2570 ascomycete genomes and in detail with *Trichoderma reesei*, *Fusarium oxysporum,* and sister taxa *Scedosporium boydii.* Significant expansion in the oxidative auxiliary activity class of CAZymes was observed in the *P. putredinis* NO1 genome resulting from increased sequences encoding putative lytic polysaccharide monooxygenases (LPMOs), oxidative enzymes acting within LPMO redox systems, and lignin-degrading laccases. *P. putredinis* NO1 scored above the 95th percentile for AA gene density across the ascomycete phylum, suggesting a primarily oxidative strategy for lignocellulose breakdown. Novel structure-based searching approaches were employed, revealing 17 new sequences with structural similarity to LPMO, laccase, and peroxidase sequences and which are potentially new lignocellulose-degrading enzymes.

2.2. Keywords

Parascedosporium, Ascomycete, CAZymes, Auxiliary activity, Oxidative, Lignocellulose, Lignin, AlphaFold, Structural, Structure-omics

2.3. Background

Energy consumption continues to grow rapidly alongside improvements in living standards, and fossil fuels continue to play a major role in industrial and agricultural sectors. With their widely accepted environmentally damaging effects, the need to move away from the use of fossil fuels and towards a net zero carbon fuel source is ever more pressing. Lignocellulosic residues consisting of cellulose, hemicellulose and lignin with minor amounts of pectins and nitrogen compounds offer the largest source of biomass for liquid fuel, chemicals, and energy (265). However, biorefining of lignocellulose has so far been limited by the recalcitrant nature of the intricate and insoluble lignin network (179, 266).

Fungi are exceptional wood-degraders and are predominantly used to produce an array of bioproducts, including commercial enzyme cocktails used in biological processing of lignocellulosic biomass. Ascomycetes, known as soft-rot fungi, degrade lignocellulose by penetration of plant secondary cell walls with hyphae that secrete complex enzyme cocktails in abundance at the site of attack (267). *Parascedosporium putredinis* NO1 is a soft-rot ascomycete identified previously as dominant in the later stages of a mixed microbial compost community grown on wheat straw (229). This behaviour suggests that the fungus can efficiently deconstruct and potentially metabolise the more recalcitrant carbon sources in the substrate. Indeed, the recent discovery of a new oxidase enzyme that cleaves the major β-ether units in lignin in the *P. putredinis* NO1 secretome, which releases the pharmaceutically valuable compound tricin from wheat straw while simultaneously enhancing digestibility of the biomass (229), promotes a requirement for further exploration of this taxa.

Here, an annotated reference genome for *P. putredinis* NO1 reveals a repertoire of carbohydrate-active enzymes (CAZymes) and oxidative enzymes focused on degrading the most recalcitrant components of lignocellulose. Comparisons across the ascomycete tree of life suggest an increased proportion of oxidative enzymes within the CAZyme repertoire of *P. putredinis* NO1. Further investigation through CAZyme repertoire comparison with two other industrially relevant wood-degrading ascomycetes; *Trichoderma reesei*, and *Fusarium oxysporum*, as well as sister taxa *Scedosporium boydii* reveals expansion in families of enzymes with roles in the oxidative dissolution of lignocellulose and demonstrated this fungus to be an exciting candidate for the identification of new lignocellulose degrading activities.

Novel approaches were used to search the *P. putredinis* NO1 genome for potentially unannotated enzyme sequences with relation to three types of classic oxidative lignocellulose degraders: lytic polysaccharide monooxygenases (LPMOs), laccases, and

peroxidases. Predicted structures were obtained for >96 % of the protein coding sequences in the genome. Structural searches were found to be effective at identifying multiple sequences for potentially novel proteins involved in lignocellulose breakdown which had low levels of structural similarity to the classic oxidative lignin and crystalline cellulose degrading enzymes. These sequences were also missed by sequence and domain-oriented searches. Further investigation and comparison of structures revealed varying levels of structural overlap despite the lack of sequence similarity. This strategy of combining search approaches can be adopted to identify divergent enzyme sequences which may have alternate lignocellulose degrading activity, variation in substrate-specificity, and different temperature and pH optima.

Further investigation and characterisation of such lignocellulose-degrading enzymes adds to the wealth of enzymes which can be incorporated into commercial enzyme cocktails to improve their effectiveness and boost the efficiency at which biomass is converted to renewable liquid fuel and value-added chemicals.

2.4. Results and discussion

2.4.1. The genome of *P. putredinis* NO1 suggests a strategy to degrade the most recalcitrant components of lignocellulose

The *P. putredinis* NO1 genome was sequenced using nanopore sequencing with the Oxford Nanopore Technologies' (ONT) MinION system to avoid errors in the assembly and annotation of coding regions resulting from long regions of repetitive DNA in eukaryotic genomes (268). The genome is 39 Mb in size and the assembly consists of 21 contigs, containing 9998 protein coding sequences. To investigate the lignocellulose-degrading enzyme repertoire of the *P. putredinis* NO1 genome, all protein coding sequences were annotated for CAZyme domains using the dbCAN server (178). In total, 795 CAZyme domains were predicted in the *P. putredinis* NO1 genome and the distribution of these domains across the CAZyme classes can be seen in **Figure 2.1A**. Glycoside hydrolases (GHs) are the most abundant CAZyme class with 290 identified. While auxiliary activities (AAs) also make a large contribution with 162 domains. Glycosyl transferases (GTs) contribute 113 domains, and carbohydrate esterases (CEs) contribute 51. Polysaccharide lyases (PLs) contribute the fewest with only 18 domains. In addition to these catalytic classes, 161 carbohydrate binding modules (CBMs) were also identified.

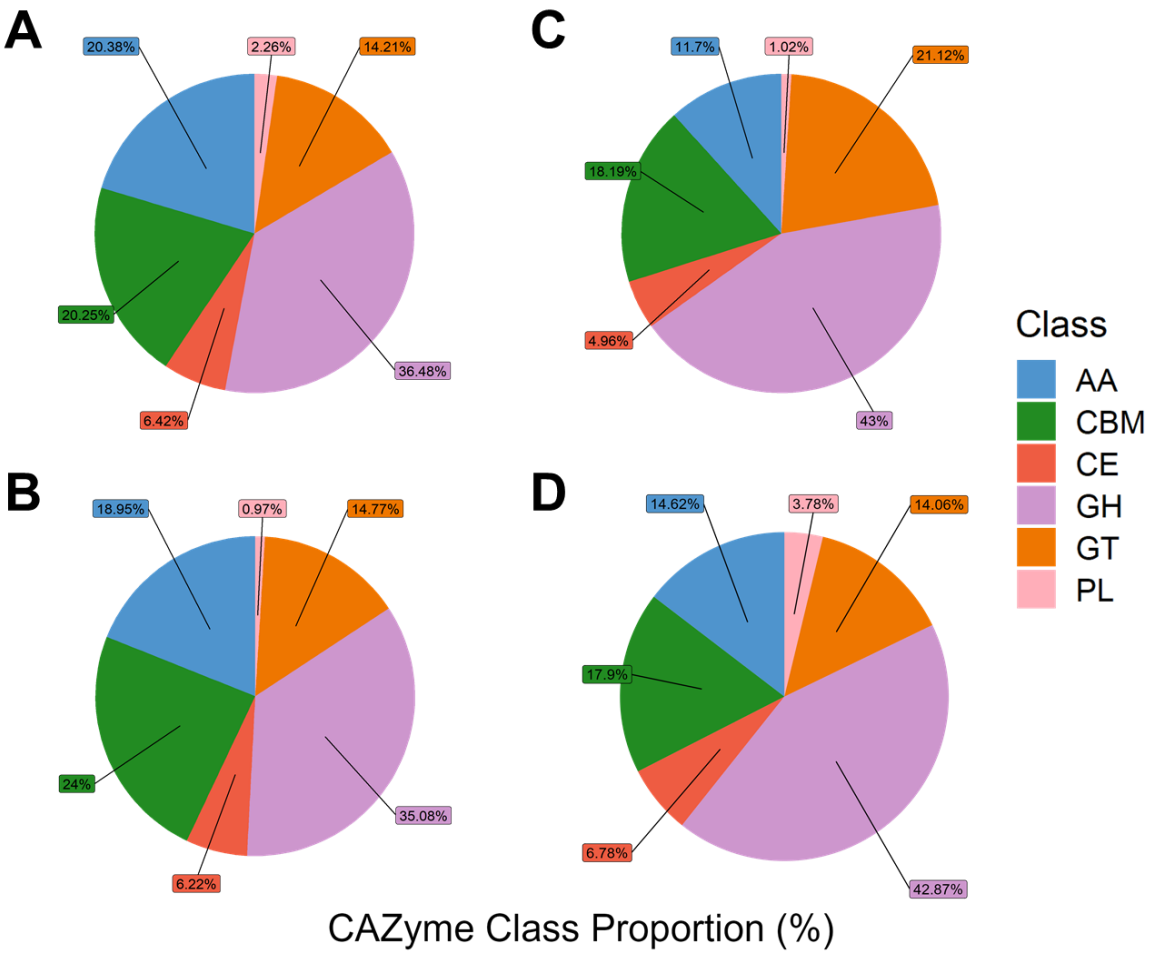


Figure 2.1 Comparison of CAZyme class repertoire. The proportions of each class of CAZyme contributing to CAZyme repertoire for *P. putredinis* NO1 (**A**), *S. boydii* (**B**), *T. reesei* (**C**), and *F. oxysporum* (**D**). Auxiliary activity (AA), carbohydrate binding module (CBM), carbohydrate esterase (CE), glycoside hydrolase (GH), glycosyl transferase (GT), polysaccharide lyase (PL).

An interesting observation was the proportionally high number of AA class CAZymes observed in the *P. putredinis* NO1 genome. To investigate this further and more broadly within the scope of the ascomycete tree of life, CAZyme profiles of all available ascomycete genomes were elucidated (**Figure 2.2**). To filter poorly represented lineages, genera and families with less than 3 and 8 species level representatives, respectively, were filtered. It was clear that *P. putredinis* NO1 has one of the highest proportions of AA class CAZymes within its repertoire relative to total CAZymes among ascomycete fungi. *P. putredinis* NO1 was a substantial outlier within the order Microascales (14.19 ± 3.15) and belonged above the 95th percentile for AA gene density among the highest AA populated genomes (20.38%), behind the genera *Diaporthe* (21.67 ± 1.3%) of the order Diaporthales, *Cladonia* (20.9 ± 1.74%) of the order Lecanorales, *Xanthoria* (20.55 ± 1.33%) of the order Teloschistales and members belonging to the highly AA enriched order *Xylariales* (19.36 ± 1.98%) containing contributions from densely populated genera *Hypoxylon* (19.78 ± 1.79%), *Annulohypoxylon*

(20.09 ± 1.1%), *Nemania* (20.25 ± 1.82%), *Neopestalotiopsis* (21.56 ± 0.66%), *Pestalotiopsis* (21.49 ± 0.97%), *Arthrinium* (20.67 ± 2.13%) and *Apiospora* (20.87 ± 1.82%). The *Parascedosporium* sister taxa *Scedosporium* (19.73 ± 1.07%) exhibited slightly lower AA density and belonged above the 90th percentile. Interestingly, *Saccharomycetales* (8.95 ± 4.63%) often associated with lignocellulose deconstruction, displayed significantly reduced AA abundance in stark contrast to neighbouring phylogenies such as Teloschistales (19.37 ± 1.6%) and Pleosporales (17.97 ± 1.98%). Members of the orders Helotiales (17.32 ± 2.36%), Botryosphaeriales (17.86 ± 1.3%), Magnaporthales (17.47 ± 0.98), and Diaporthales (18.78 ± 2.86) displayed a degree of enrichment of AAs whilst members belonging to Hypocreales (13.93 ± 4.55%) and Mycosphaerellales (15.21 ± 2.89%) displayed lower abundances. Considering how the AA class of enzymes is predominantly associated with the degradation of lignin and crystalline cellulose it highlights a potential strategy of the fungus to target these components. Indeed, in a mixed microbial community grown on wheat straw the fungus was observed to become more dominant in the later stages of the culture, potentially due to its capacity to modify the more difficult to degrade components of lignocellulose for growth (229). The gene density distribution here highlights promising candidate lineages for further exploration for additional insights into lignin and cellulose turnover.

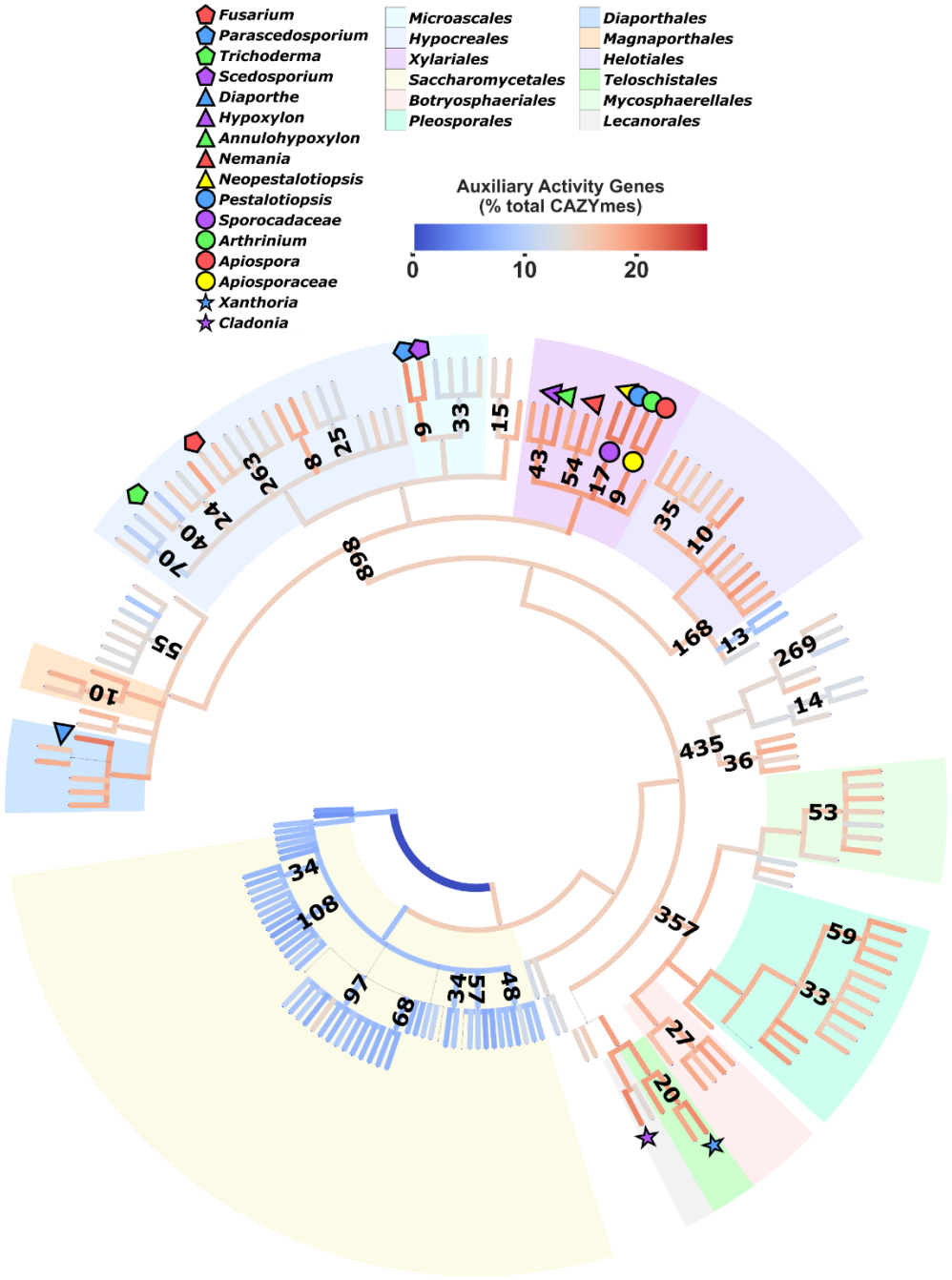


Figure 2.2 Auxiliary activity distribution and density across the ascomycete tree of life. Genes predicted for ascomycete genome assemblies were annotated for CAZymes to explore patterns in the distribution and density of auxiliary activities (n=2570) within the ascomycete phylogenetic tree. Branch colors indicate the mean proportion of auxiliary activities within only the CAZyme annotations accounted for by all descendant taxa. Numerical clade annotations represent the number of sequenced genomes available. Key taxa, including lineages above the 95th percentile for AA proportion have been highlighted. Genera and families with less than 3 and 8 species level representatives, respectively, have been pruned for clarity (n=462 taxa). Nodes of taxonomic ranks below genus have been pruned (n=93).

Within white- and brown- (basidiomycete), and soft-rot (ascomycete) fungi, it has been demonstrated that the CAZyme repertoire can vary greatly from species to species (162). To investigate the repertoire of *P. putredinis* NO1 in more detail, CAZyme domains were compared to that of three other wood-degrading ascomycetes. *Scedosporium boydii* is located within the sister taxon of *Parascedosporium* and has a genome of 43 Mb containing 1029 CAZyme domains. The genome and CAZyme complement of the soft-rot *P. putredinis* NO1 are larger than that of *Trichoderma reesei* which contains 786 domains in 34 Mb of DNA. *T. reesei* is a mesophilic soft-rot fungus known for its ability to produce high titres of polysaccharide-degrading enzymes that are used in biomass-degrading enzyme cocktails (171). The genome of *P. putredinis* NO1 is slightly smaller than that of *Fusarium oxysporum* at 47 Mb, a phytopathogenic fungus containing an expanded CAZyme repertoire of 1430 domains (173). The lignocellulose degrading activities of *F. oxysporum* have been well-investigated in part due to its pathogenicity and ability to ferment sugars from lignocellulose breakdown directly into ethanol (79, 174).

Examining the distribution of predicted CAZyme domains revealed that despite the similar overall number of CAZyme domains for *P. putredinis* NO1 and *T. reesei*, the proportion of AA class CAZyme domains is much higher in the genome of *P. putredinis* NO1 (**Figure 2.1A**). Proportionally, AA class CAZymes make the largest contribution to CAZyme repertoire in the genome of *P. putredinis* NO1 compared to the other ascomycetes (**Figure 2.1**). This again could suggest an oxidative strategy to target to lignin and crystalline cellulose. Although analysis of fungal secretomes would be required to confirm an improved ability of *P. putredinis* NO1 to deconstruct lignocellulosic components, the high potential capacity for degradation of lignin and crystalline cellulose within the genome suggests that this is an important fungus to explore for new lignocellulose-degrading enzymes. This is especially relevant considering that this is the first genome assembly of the genus *Parascedosporium*.

The increased contribution of AA class CAZymes is mirrored by a reduced proportion of GH class CAZymes in the *P. putredinis* NO1 genome compared to *T. reesei* and *F. oxysporum*. This reduced GH contribution is also visible in the genome of *S. boydii*, a close relative of *P. putredinis* NO1. Despite the reduced number of the hydrolytic GH class CAZymes, the repertoires of *P. putredinis* NO1 and *S. boydii* contain the highest proportions of CBMs, domains typically associated with hydrolytic CAZymes such as GHs (202), but which have also been observed in oxidative LPMOs (203, 204). The increased proportion of CBMs in the genome of *P. putredinis* NO1 could aid the catalytic CAZymes in accessing and binding to these substrates. Indeed, examining the CBM domains at the family level shows a high number of crystalline cellulose binding domains (CBM1) in the genome of both *P. putredinis*

NO1 and *S. boydii*, much higher than the number of domains assigned to any of the other CBM families (**Supplementary Figure 2.1**).

2.4.2. Closer investigation of the AA CAZyme repertoire reveals more about the lignocellulose degrading strategy of *P. putredinis* NO1

The high number of AA domains, a functional class that notably contains LPMOs, peroxidases, and laccases (179), in the genome of *P. putredinis* NO1 are likely to endow this fungus with the ability to degrade recalcitrant components of the plant cell wall through a primarily oxidative mechanism. LPMOs are copper-containing enzymes that enhance polysaccharide degradation by generating new sites for attack by hydrolytic CAZymes (182). LPMOs have been shown to act on all major polysaccharide components of lignocellulose. Their oxidative action relies on exogenous electron donors provided by other AA family CAZymes, small molecule reductants and even lignin (179, 182). It has recently been demonstrated that LPMOs readily utilise hydrogen peroxide (H_2O_2) as a cosubstrate also (183, 188).

Investigating the distribution of AA domains across the AA families revealed AA9 family members to be the most abundant in the *P. putredinis* NO1 genome with 35 domains, the highest in the four ascomycetes investigated here (**Supplementary Figure 2.2**). This family contains the cellulose, xylan, and glucan active LPMOs described above (181). AA3 and AA3_2 domains are the second and third most abundant families in the *P. putredinis* NO1 genome with 29 and 27 domains, respectively. These are flavoproteins of the Glucose-methanol-choline (GMC) oxidoreductase family which includes activities such as cellobiose dehydrogenase, glucose-1-oxidase, aryl alcohol oxidase, alcohol oxidase and pyranose oxidase (185). It is proposed that flavin binding oxidative enzymes of this family play a central role in spatially and temporally supplying H_2O_2 to LPMOs and peroxidases or to produce radicals that degrade lignocellulose through Fenton chemistry (183). The *P. putredinis* NO1 genome also contains 12 AA7 family domains, the family of glucooligosaccharide oxidase enzymes. These have recently been demonstrated to transfer electrons to AA9 LPMOs which boosts cellulose degradation (269). Altogether, the apparent expansion of these LPMO system families suggest a potentially increased capacity for *P. putredinis* NO1 to oxidatively target crystalline cellulose.

The genome of *P. putredinis* NO1 also contains 12 AA1 family CAZyme domains. This family includes laccase and multi-copper oxidase enzymes which catalyse the oxidation of various aromatic substrates while simultaneously reducing oxygen to water (189). It has also been demonstrated that laccases can boost LPMO activity through the release of low molecular weight lignin polymers from biomass which can in turn donate electrons to LPMOs (193).

Additionally, 7 domains belonging to the AA8 family were identified, a family of iron reductase domains initially identified as the N-terminal domain in cellobiose dehydrogenase enzymes but also found independently and appended to CBMs (179, 197, 198). These domains are believed to be involved in the generation of reactive hydroxyl radicals that can indirectly depolymerize lignin. There are 6 AA4 domains in the genome of *P. putredinis* NO1, the highest number of the four ascomycetes investigated here. These are vanillyl-alcohol oxidase enzymes with the ability to catalyse the conversion of a wide range of phenolic oligomeric compounds (200). These may act downstream of the lignin depolymerisation catalysed by other members of the AA class. There is a clear capacity in the *P. putredinis* NO1 genome for lignin depolymerisation and metabolism through the multiple domains identified belonging to these families. The *P. putredinis* NO1 genome also contains two AA16 domains, a recently identified family of LPMO proteins with an atypical product profile compared to the traditional AA9 family LPMOs and a potentially different mode of activation (172).

Gene expression of CAZymes in the *P. putredinis* NO1 genome has been explored previously during growth on glucose, compared to growth on wheat straw with samples taken at days 2, 4, and 10 (229). This transcriptomic data gives a view of the potential strategy by which *P. putredinis* NO1 utilises its expanded repertoire of AA class CAZymes. Up-regulation of AA class CAZymes during growth on wheat straw compared to growth on glucose was observed predominantly at day 4 and then gave way to up-regulation instead of mainly GH class hydrolytic CAZymes at day 10. This could represent a strategy where the recalcitrant lignin and crystalline cellulose are targeted first by LPMOs and lignin degraders such as laccases, making the polysaccharide substrates of hydrolytic GH enzymes more accessible.

2.4.3. Searching the *P. putredinis* NO1 genome for new oxidative lignocellulose-degrading enzymes with sequence-, domain-, and structural-based strategies

Due to the evidence of a strategy for *P. putredinis* NO1 to target the most recalcitrant components of lignocellulose and the recent discovery of a new oxidase with the ability to cleave the major linkage in lignin from this strain (229), it was hypothesised the genome of this fungus contains additional new enzymes for the breakdown of plant biomass. Particularly this fungus could contain new enzymes with roles in degrading the lignin and crystalline cellulose components and which have not been annotated as CAZymes in this analysis.

Traditionally, homologue searching has been performed using a sequence-based approach (254). This approach uses either the primary amino acid sequence of an example protein to

search an unknown database for similar sequences, or uses hidden Markov models (HMMs) to search for domains of interest (255). However, both techniques rely on primary amino acid sequence homology and neglect that proteins with distantly related sequences may have similar three-dimensional structures and therefore activity. The recent emergence of AlphaFold provides a resource for the fast and accurate prediction of unknown protein structures (257). Using this tool, structures were predicted for >96 % of the protein-coding regions of the *P. putredinis* NO1 genome. These structures were used to create a database of protein structures into which structures of interesting enzymes such as those for LPMOs, laccases, and peroxidases could be searched. These structural searches for new enzymes were performed alongside sequence- and domain-based searches for comparison of the ability to identify interesting new candidates.

LPMO related sequences were searched for in the *P. putredinis* NO1 genome using the sequence of an AA9 family LPMO from *Aspergillus niger* with the default E-value cut off of 1 x 10^{-5}, with the AA9 HMM from Pfam and considering domain hits that fell within the default significance inclusion threshold of 0.01 (270), and the structure of the same *A. niger* LPMO with a tailored 'lowest percentage match' parameter. In total, 49 sequences were identified across the three searching strategies and 33 of these sequences were also annotated by dbCAN as AA9 family LPMOs (**Supplementary Table 2.1**). With the objective of identifying new enzymes, the remaining 16 sequences were investigated further, and the distribution of the identification of these sequences across the three search strategies can be seen in **Table 2.1**. Two of the sequences, PutMoI and PutMoM, were identified by all three search approaches. These sequences both had conserved signal peptides with a conserved N-terminal histidine after the cleavage site, a characteristic feature of LPMOs (271).

When creating the structure database it was tempting to filter predicted structures by pLDDT score, the AlphaFold metric for prediction confidence, to create a database solely of 'high confidence' structures (257). However, pLDDT scores reflect local confidence and should instead be used for assessment of individual domains (272). The majority of the structures generated here had pLDDT score of over 60%, however pLDDT scores lower than 70% are considered low confidence (**Supplementary Figure 2.3**). Extracellular enzymes are of particular interest here, but these often have disordered N-terminal signal peptides which can reduce the overall pLDDT scores. Therefore, for secreted enzyme identification from AlphaFold structures it is inappropriate to filter by pLDDT score. Indeed, the PutMoI structure mentioned above had a pLDDT score of 62%, considered to be low confidence (257), but which had characteristic features of LPMOs and which demonstrated structural similarity to the *A. niger* AA9 LPMO used for structural searches (**Figure 2.3A** and **B**). The central beta-sheet structures align well to the *A. niger* AA9 LPMO for both PutMoI and PutMoM, but both

also have additional loops of disordered protein which likely explains the relatively low PDBefold alignment confidence scores (Q-scores) of 0.23 and 0.34 for PutMotI and PutMoM, respectively. This again highlights the unreliability of structural confidence scores alone and demonstrates how manual inspection of structural alignments may prove more useful. Despite not being annotated as AA9 LPMOs by the dbCAN server for CAZyme annotation (178), both sequences were identified using the Pfam AA9 HMM and appear to be conserved AA9 LPMOs and, therefore, are not of interest in the discovery of new enzymes.

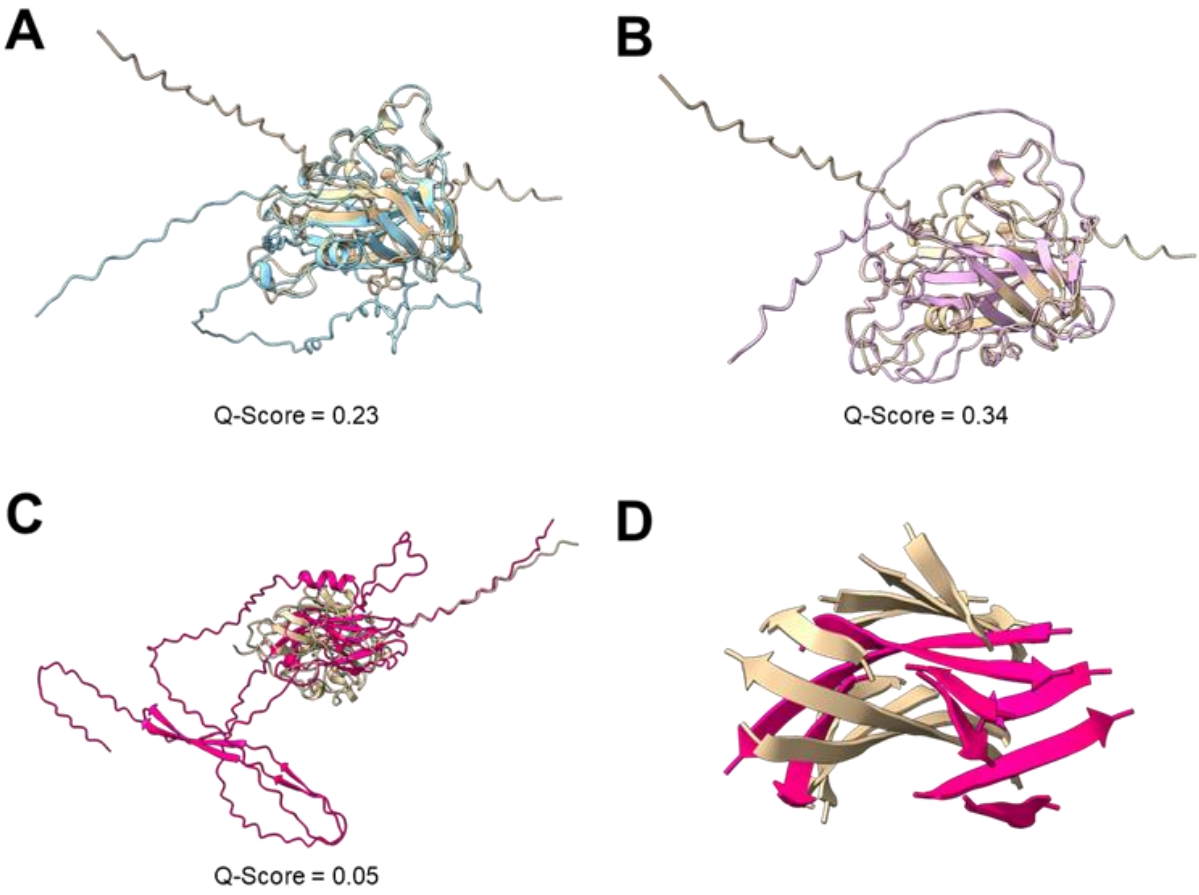


Figure 2.3 Structural comparison of LPMO related proteins. The AlphaFold predicted structures of three sequences, PutMoI (**A**), PutMoM (**B**), and PutMoP (**C** and **D**) from the *P. putredinis* NO1 genome structurally aligned to the *A. niger* AA9 LPMO used in sequence and structure-based searching (UniProt ID: A2QZE1). *A. niger* AA9 LPMO (Beige), PutMoI (Blue), PutMoM (Pink), PutMoP (Hot Pink). Q-score is a quality function of Cα alignment from PDBefold.

Table 2.1 Identifying LPMO related proteins encoded in the *P. putredinis* NO1 genome. Coding regions of proteins related to LPMOs identified through genome searching approaches with the sequence of an *A. niger* AA9 LPMO (E-value cut-off = 1 x 10^{-5}), the Pfam AA9 HMM (Significance threshold = 0.01), and the structure of the *A. niger* AA9 LPMO (Lowest percentage match = 50%) and which were not annotated as AA9 CAZymes by dbCAN. InterPro annotations were retrieved where possible.

Coding Region	GenBank Accession	Protein ID	Identified by Searching Approach			Interpro Annotation
			Sequence	Domain	Structure	
FUN_000653-T1	CAI7987917.1	PutMoA			✓	AA16 LPMO
FUN_000713-T1	CAI7987978.1	PutMoB			✓	Rho factor associated
FUN_002573-T1	CAI7991617.1	PutMoC		✓		-
FUN_002890-T1	CAI7992277.1	PutMoD			✓	AA16 LPMO
FUN_002962-T1	CAI7992399.1	PutMoE			✓	-
FUN_003190-T1	CAI7992922.1	PutMoF			✓	Ferritin-like
FUN_003535-T1	CAI7993628.1	PutMoG		✓		AA13 LPMO
FUN_003783-T1	CAI7994168.1	PutMoH			✓	-
FUN_006366-T1	CAI7999797.1	PutMoI	✓	✓	✓	AA9 LPMO
FUN_006413-T1	CAI7999893.1	PutMoJ		✓		AA9 LPMO
FUN_006553-T1	CAI8000144.1	PutMoK			✓	-
FUN_007242-T1	CAI8001774.1	PutMoL		✓		AA9 LPMO
FUN_007666-T1	CAI8002525.1	PutMoM	✓	✓	✓	AA9 LPMO
FUN_008106-T1	CAI8003467.1	PutMoN	✓	✓		AA9 LPMO
FUN_009239-T1	CAI7992001.1	PutMoO			✓	-
FUN_010012-T1	CAI8003342.1	PutMoP			✓	-

By utilising multiple searching approaches, potentially new sequences with LPMO related activities can be identified. When searching for LPMO related sequences, domain-based approaches identified all coding regions also identified by sequence-based searching as well as additional coding regions (**Supplementary Table 2.1**). This pattern of domain-based searching identifying more coding regions than sequence-based searching was also observed for the other activities investigated (**Supplementary Table 2.2** and **Supplementary Table 2.3**). For structure-based searching, parameters of the searches could be tailored to identify additional coding regions with lower overall structural similarity, but which may still be interesting. For example, searching against the *P. putredinis* NO1

genome structure database with the structure of the *A. niger* AA9 LPMO, and with the 'lowest acceptable match' parameter which is the cutoff at which secondary structures must overlap between a query and a target set at 50 %, yielded 30 coding regions (**Supplementary Table 2.1**). Of these sequences, 9 were not identified by the sequence or domain-based searching approaches and were investigated in more detail (**Table 2.1**). To investigate these further, sequences were searched against the NCBI non-redundant protein database to identify related sequences (273), conserved domains were predicted with InterPro, any CAZyme domains were annotated with dbCAN (178), the predicted structures were compared with structures in the PDB database (274), and secretion signal peptides were predicted with SignalP (275) in an attempt to elucidate the potential functions. Two of the sequences, PutMoA and PutMoD, are the two predicted AA16 LPMOs identified in the *P. putredinis* NO1 CAZyme repertoire earlier (**Supplementary Figure 2.2**). Another two sequences, PutMoH and PutMoK, were not annotated as CAZymes but had conserved BIM1-like domains. BIM1-like proteins are LPMO_auxilliary-like proteins, function in fungal copper homeostasis, and share a similar copper coordination method to the LPMOs which they are related to (276). Although not likely to be involved in lignocellulose breakdown, this highlights how structurally related proteins in terms of active site or co-factor coordination structures can be identified with structural approaches where sequence- and domain-based approaches fail. Three of the nine sequences were also identified as being upregulated when *P. putredinis* NO1 was previously grown on wheat straw compared to growth on glucose (**Supplementary File 2.1**) (229). Although this does not confirm the role of these proteins in lignocellulose breakdown, it provided another layer of information for the selection of interesting candidate sequences to investigate further. PutMoP was the most interesting sequence identified solely by the structural searching and showed upregulation during growth on wheat straw compared to glucose. It was not annotated as a CAZyme, no conserved domains were identified, and sequence homology was only observed to hypothetical proteins in the NCBI non-redundant protein database (273). Comparing the AlphaFold predicted structure of PutMoP to the *A. niger* AA9 LPMO revealed similarity at the central beta-sheet structure despite a very low Q-score of 0.05 (**Figure 2.3C** and **D**). A secretion signal peptide was also predicted for this protein, suggesting an extracellular role. This immunoglobulin-like distorted β-sandwich fold is a characteristic structural feature of LPMOs and is shared across the LPMO CAZyme families (277). The similarity of this central structure is likely the reason for identification of this sequence by structural comparison. This structural similarity at the protein centre, the lack of amino acid sequence similarity, and the conserved secretion signal makes this protein an interesting candidate for further investigation. Searching the PutMoP structure against the whole PDB structure database returned many diverse proteins not linked to lignocellulose breakdown, however the Q-score

was very low for all the structures and did not help to discern the potential activity of this protein. The sequence lacks the N-terminal histidine after the signal peptide cleavage site which is conserved in LPMOs so this protein is unlikely to be an LPMO. However, a secreted unknown protein with some central structural similarity to an important class of oxidative proteins that degrade crystalline cellulose is of definite interest.

In addition to searching for LPMO related sequences, classes of enzymes involved in the breakdown of lignin are important targets for the biorefining of plant biomass. The recalcitrance of lignin is a limiting factor hindering the industrial use of lignocellulose as a feedstock to produce biofuels. Lignin itself is also a historically underutilised feedstock for valuable chemicals (90). Laccases are multicopper oxidase family enzymes that catalyse oxidation of phenolic compounds through an electron transfer reaction that simultaneously reduces molecular oxygen to water (193). They modify lignin by depolymerisation and repolymerisation, Cα oxidation, and demethylation and are particularly efficient due to their use of readily available molecular oxygen as the final electron acceptor (191, 192).

Laccase related sequences were searched for in the *P. putredinis* NO1 genome using the sequence of an AA1 family laccase from *A. niger*, a bespoke HMM constructed from ascomycete laccase and basidiomycete multi-copper oxidase sequences downloaded from the laccase engineering database (278), and with the structure of the *A. niger* AA1 laccase. In total, 32 sequences were identified across the three searching strategies and only 9 of these were annotated by dbCAN as AA1 family CAZymes (**Supplementary Table 2.2**). The bespoke HMM allowed for more divergent sequences for these enzymes to be incorporated into the model's construction. The result was the identification of sequences that when explored further looked like laccase enzymes but were missed by traditional CAZyme annotation, highlighting how searching for CAZymes alone is a limited method for identifying lignocellulose degrading enzymes. However, for the identification of new lignocellulose degrading enzymes, more divergent sequences are of interest. A single coding sequence, PutLacJ was identified by the structural searching approach with a 30% 'lowest acceptable match' parameter that was not identified by sequence or domain-based searching (**Table 2.2**).

Table 2.2 Identifying laccase related proteins encoded in the *P. putredinis* NO1 genome. Coding regions of proteins related to laccases identified through genome searching approaches with the sequence of an *A. niger* AA1 laccase (E-value cut-off = 1 x 10^{-5}), the bespoke laccase and multicopper oxidase HMM constructed from sequences from the laccase engineering database (Significance threshold = 0.01), and the structure of the *A. niger* AA1 laccase (Lowest percentage match = 30%) and which were not annotated as AA1 CAZymes by dbCAN. InterPro annotations were retrieved where possible.

Coding Region	GenBank Accession	Protein ID	Identified by Searching Approach			Interpro Annotation
			Sequence	Domain	Structure	
FUN_000263-T1	CAI7987524.1	PutLacA		✓		-
FUN_000580-T1	CAI7987844.1	PutLacB		✓		Phospho-diesterase
FUN_000646-T1	CAI7987911.1	PutLacC		✓		-
FUN_000759-T1	CAI7988026.1	PutLacD	✓	✓		Multi-copper oxidase
FUN_000832-T1	CAI7988099.1	PutLacE		✓		-
FUN_001183-T1	CAI7988671.1	PutLacF		✓		-
FUN_001583-T1	CAI7989479.1	PutLacG		✓		Salt tolerance regulator
FUN_002249-T1	CAI7990863.1	PutLacH	✓	✓	✓	AA1 Multi-copper oxidase
FUN_002874-T1	CAI7992258.1	PutLacI		✓		-
FUN_003732-T1	CAI7994085.1	PutLacJ			✓	-
FUN_003828-T1	CAI7994234.1	PutLacK		✓		-
FUN_004259-T1	CAI7995254.1	PutLacL		✓		Nucleoside hydrolase
FUN_004616-T1	CAI7995870.1	PutLacM		✓		-
FUN_004739-T1	CAI7996089.1	PutLacN		✓		Fumaryl-acetoacetate hydrolase family
FUN_005132-T1	CAI7997298.1	PutLacO		✓	✓	AA1 Multi-copper oxidase
FUN_005520-T1	CAI7998008.1	PutLacP		✓		Diacylglycerol acyltransferase
FUN_006244-T1	CAI7999594.1	PutLacQ		✓		-
FUN_006620-T1	CAI8000270.1	PutLacR		✓		Fumaryl-acetoacetate hydrolase family

FUN_006720-T1	CAI8000684.1	PutLacS		✓		GT90
FUN_007228-T1	CAI8001746.1	PutLacT		✓		-
FUN_007508-T1	CAI8002246.1	PutLaU		✓		Pex2
FUN_008329-T1	CAI8004041.1	PutLacV		✓		ATPase-related
FUN_009491-T1	CAI7995256.1	PutLacW		✓		Helicase

PutLacJ was not annotated as a CAZyme by dbCAN but does have a predicted cupredoxin domain, a feature of laccase enzymes (279). Structural comparisons against the PDB structure database revealed alignments with moderate confidence scores to copper-containing nitrite reductases from *Neisseria gonorrhoeae* which are suggested to play a role in pathogenesis (280). In fungi, it is more likely that these are playing a role in denitrification (281). The lack of a signal peptide make it unlikely that this protein is involved in lignin depolymerisation, despite the structural similarity to the beta-sheet regions of the *A. niger* laccase (**Figure 2.4**).

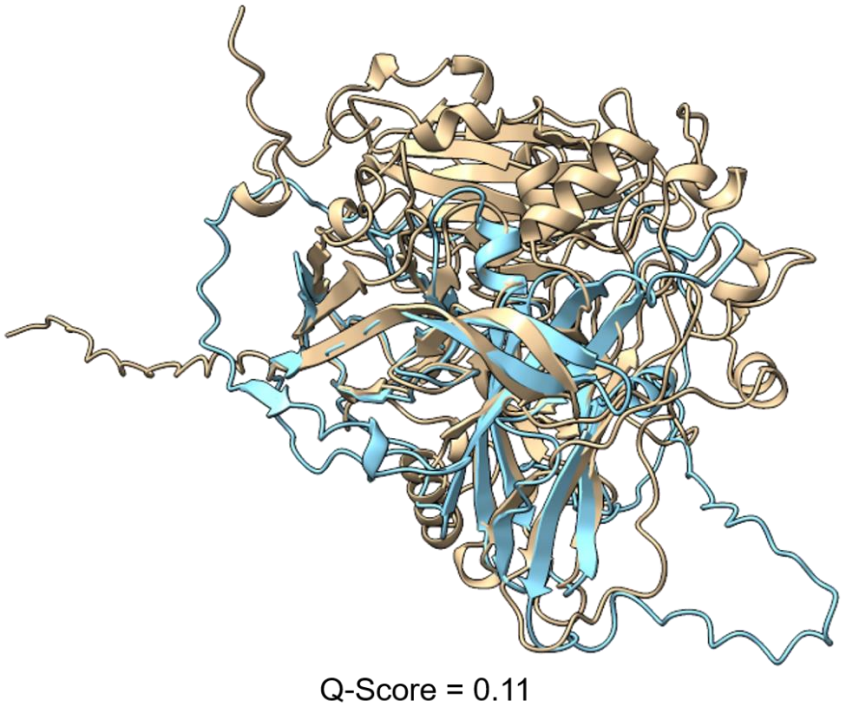


Figure 2.4 Structural comparison of PutLacJ laccase related protein. The AlphaFold predicted structures of the sequence PutLacJ from the *P. putredinis* NO1 genome structurally aligned to the *A. niger* laccase used in sequence and structure-based searching (UniProt ID: A2QB28). *A. niger* laccase (Beige), PutLacJ (Blue). Q-score is a quality function of Cα alignment from PDBefold.

Peroxidases (PODs) also play a major role in lignin deconstruction by white-rot fungi. PODs are lacking in brown-rot species, presumably due to their non-ligninolytic specialisation of substrate degradation (194). The identification of new putative peroxidases in *P. putredinis* NO1 is of interest. Fungal class II peroxidases are divided into three lignolytic forms; lignin peroxidase (LiP), manganese peroxidase (MnP), and versatile peroxidase (VP) (195).

Sequence searches into the *P. putredinis* NO1 genome using sequences of MnP from *Aureobasidium subglaciale*, LiP from *F. oxysporum*, and VP from *Pyronema confluens* only yielded 2 sequences (**Supplementary Table 2.3**). Both peroxidase related sequences were also identified by domain searching using a bespoke HMM constructed from sequences of MnPs, LiPs, and VPs downloaded from the fPoxDB database of peroxidase sequences (282). This domain-based approach only identified 3 sequences in total, all of which were annotated as AA2 family CAZymes also (**Supplementary Table 2.3**). However, structural-based searching using the structures of the same three peroxidases, and with a 'lowest acceptable match' parameter of 30% used in sequence-based searches identified 9 coding regions in total (**Supplementary Table 2.3**), 7 of which were not identified by sequence- or domain-based searching approaches and were not annotated as AA2 CAZymes (**Table 2.3**), but were all found to be upregulated previously when *P. putredinis* NO1 was grown on wheat straw compared to growth on glucose (**Supplementary File 2.1**) (229).

Table 2.3 Identifying peroxidase related proteins encoded in the *P. putredinis* NO1 genome. Coding regions of proteins related to peroxidases identified through genome searching approaches with the sequences of an MnP from *A. subglaciale*, LiP from *F. oxysporum*, and VP from *P. confluens* (E-value cut-off = 1 x 10^{-5}), the bespoke peroxidase HMM constructed from MnP, LiP, and VP sequences in the fPoxDB database (Significance threshold = 0.01), and the structure of the same three peroxidases used for sequence searches (Lowest percentage match = 30%) and which were not annotated as AA2 CAZymes by dbCAN. InterPro annotations were retrieved where possible.

Coding Region	GenBank Accession	Protein ID	Identified by Searching Approach			Interpro Annotation
			Sequence	Domain	Structure	
FUN_002995-T1	CAI7992466.1	PutPoxA			✓	DUF3632
FUN_003542-T1	CAI7993642.1	PutPoxB			✓	Arabino-furanosidase
FUN_003618-T1	CAI7993895.1	PutPoxC			✓	-
FUN_004484-T1	CAI7995643.1	PutPoxD			✓	Cell division control
FUN_008413-T1	CAI8004205.1	PutPoxE			✓	SIT4 phosphatase -associated
FUN_008923-T1	CAI7988420.1	PutPoxF			✓	-
FUN_009329-T1	CAI7993214.1	PutPoxG			✓	-

Investigating these sequences further revealed two sequences to be the most interesting, PutPoxA and PutPoxG, both with low Q-scores of 0.01 and 0.04, respectively. PutPoxA was not annotated as a CAZyme but does have a predicted domain of unknown function family 3632 (DUF3632). Genes encoding DUF3632 domains were previously found to be upregulated in the filamentous ascomycete *Neurospora crassa* when the CLR-2 transcription factor, important for growth on cellulose, was constitutively expressed (283). The protein does however lack a signal peptide and structural comparison to the *A. subglaciale* MnP shows similar helical structures, but these secondary structures do not appear to overlap very well (**Figure 2.5A**). PutPoxG was not annotated as a CAZyme and no conserved domains were identified, although the helical structures do seem to align better with the *A. subglaciale* MnP than PutPoxA (**Figure 2.5B**). Furthermore, searching of both structures against the PDB database was performed, but all alignments had very low Q-scores of less than 0.1.

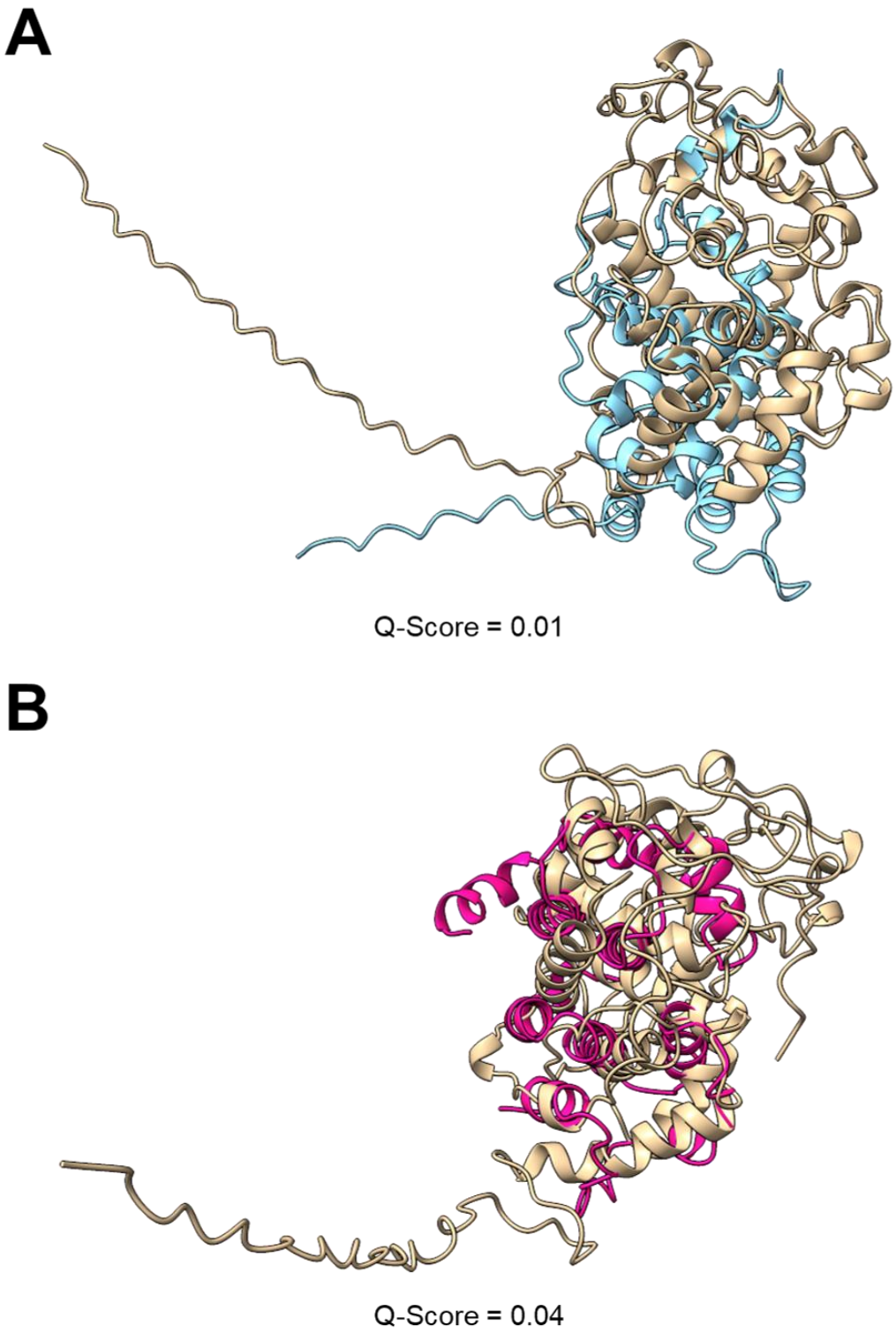


Figure 2.5 Structural comparison of peroxidase related proteins. The AlphaFold predicted structures of two sequences, PutPoxA (**A**) and PutPoxG (**B**), from the *P. putredinis* NO1 genome structurally aligned to the *A. subglaciale* MnP used in sequence and structure-based searching. A predicted structure was unavailable and so a predicted structure was generated with AlphaFold. *A. subglaciale* MnP (Beige), PutPoxA (Blue), PutPoxG (Hot Pink). Q-score is a quality function of Cα alignment from PDBefold.

As with the candidates identified by LPMO and laccase searching approaches, it is hard to be confident on sequence and structural investigation alone that these proteins are involved in lignocellulose breakdown. Although by utilising multiple searching approaches, more divergent and varied sequences with potential relation to industrially important enzymes have been identified here. This strategy of searching for new enzymes involved in the breakdown of the most recalcitrant components of lignocellulose would work well when combined with additional layers of biological data e.g., transcriptomic, or proteomic data. Many of the coding regions investigated here show structural similarity to the interesting classes of enzymes with which they were identified but lack the sequence similarity and therefore the functional annotation. Transcriptomic data showing upregulation of these genes or proteomic data showing increased abundances of these proteins when the organism in question is grown on lignocellulosic substrates would inspire more confidence in the role of these proteins in the degradation of plant-biomass. Therefore, we used sequence similarity to identify the corresponding transcripts for these coding regions in the transcriptomic time course dataset of *P. putredinis* NO1 grown for 10 days in cultures containing wheat straw published previously (229). The transcriptomic data was explored for all sequences which were identified solely by structural searches and therefore considered interesting (**Supplementary File 2.1**). For the four sequences explored in more detail, we found that three of the four: PutMoP, PutPoxA, and PutPoxG, were found to be significantly upregulated on at least one timepoint when grown on wheat straw compared to growth on glucose (**Figure 2.6**). The remaining sequence, PutLacJ, expression was found to be significantly higher during growth on glucose compared to growth on wheat straw. However structural investigation revealed that PutLacJ had similarity to copper-containing nitrite reductase proteins and it was concluded that it is unlikely to be involved in lignocellulose breakdown. Characterisation would be required to confirm the role of these candidates in lignocellulose breakdown and to understand whether these activities are new. However, the implication in lignocellulose degrading processes through the analysis of transcriptomic data provides another source of information by which candidates identified through the described strategy can be investigated. It is hoped that adoption of a similar strategy for analysis of the wealth of sequence data now publicly available will allow identification of novel enzyme sequences for many important processes to be made simpler.

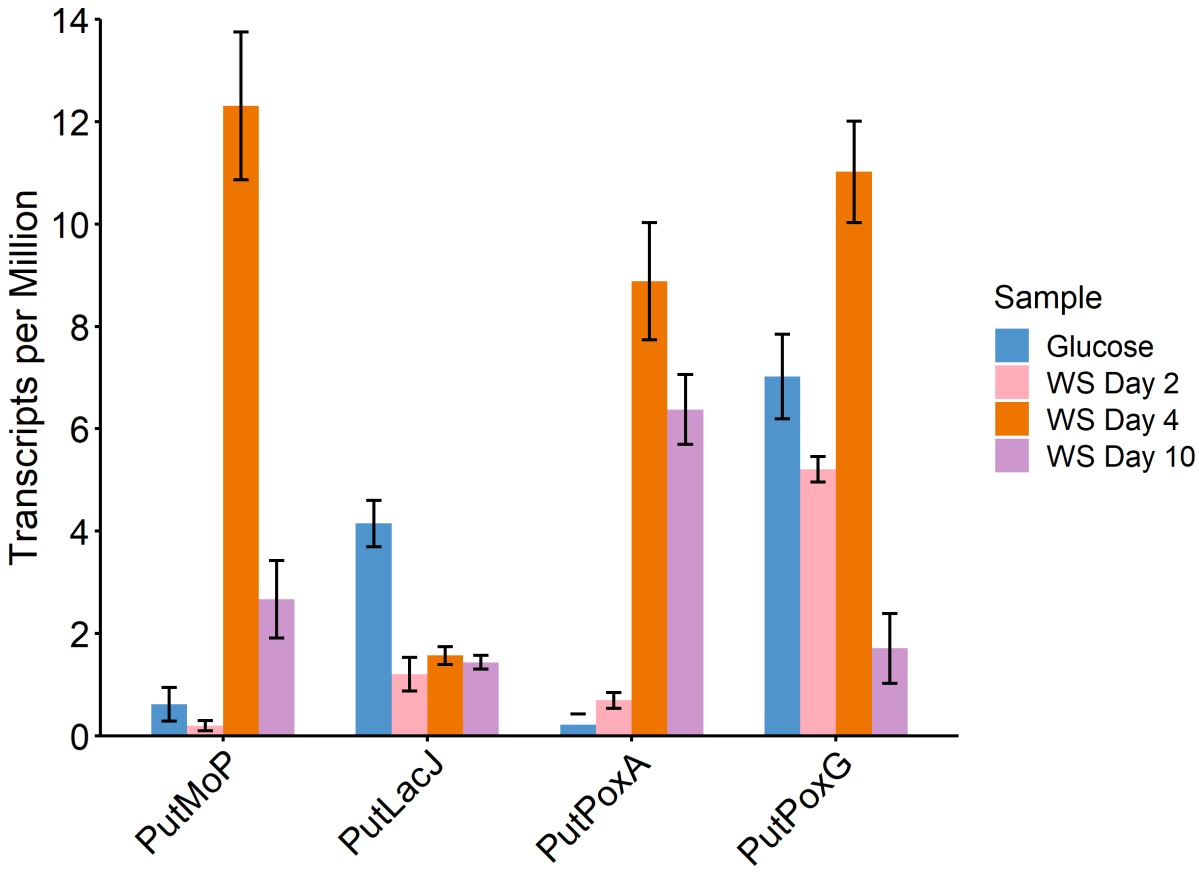


Figure 2.6 Gene expression of interesting candidates. Transcripts per Million (TPM) values for each of the four candidates explored, during growth on glucose, or on day 2, 4, and 10 of growth in liquid cultures containing wheat straw (WS).

2.5. Conclusions

P. putredinis NO1 was revealed here to contain a diverse repertoire of lignocellulose degrading enzymes in its genome. The newly annotated reference genome is a potentially useful resource, considering the potential of *P. putredinis* NO1 for the identification of industrially valuable enzymes (229). Among ascomycetes, *P. putredinis* NO1 exists within the 95th percentile for abundant auxiliary activity gene density, implying potential specialism regarding mechanisms of lignocellulose degradation and belongs to a substantially underrepresented and underexplored lineage. Investigating CAZyme families in more detail revealed an increased capacity to target the most recalcitrant components of lignocellulose when compared to three other biomass-degrading ascomycetes. For crystalline cellulose degradation, expansions were observed in families of LPMOs and in families associated with LPMO systems. Multiple domains encoding lignin-degrading laccase proteins were also identified. Considering the context in which *P. putredinis* NO1 was identified, thriving at the late stages of a mixed microbial community grown on wheat straw, it is feasible that the genome of this fungus contains new ligninolytic activities. By utilising a strategy of searching genomic data for new enzymes with simultaneous sequence-, domain-, and structural-based approaches, multiple interesting sequences were identified.

2.6. Materials and methods

2.6.1. Strain isolation

P. putredinis NO1 was isolated from a wheat straw enrichment culture and maintained as reported previously (229).

2.6.2. Genomic DNA extraction and sequencing

For DNA extraction, *P. putredinis* NO1 was grown in optimised media containing 10 % (w/v) sucrose at 30 °C with shaking at 140 rpm for 14 days. Wet fungal biomass was washed in deionised water before pelleting in 50 mL falcon tubes at 4500 rpm for 15 minutes, and ten technical replicates of 100 mg of biomass were then prepared in 1.5 mL tubes. Fungal biomass was then digested by adding 100 µL of 1 mg mL^{-1} Chitinase from *Streptomyces griseus* (Merck) and 200 µL of 50 mM EDTA and incubating at 37 °C for 3 hours. DNA extraction was then performed with the Wizard® Genomic DNA Purification Kit (Promega). Digested samples were centrifuged at 18,000 x g at 4 °C for 2 minutes and the supernatant discarded. Pellets were resuspended with 300 µL of nuclei lysis solution and 100 µL of protein precipitation solution and rotated for 5 minutes before a 5-minute incubation on ice. Samples were then centrifuged at 18,000 x g at 4 °C for 3 minutes and the supernatant transferred to fresh tubes containing 300 µL of cold isopropanol, gently mixed by inversion, and centrifuged again. The supernatant was discarded, and the pellet was washed in 70 % ice cold ethanol before centrifugation followed by air drying the DNA pellet. The pellet was then resuspended in 50 µL of DNA rehydration solution with the addition of 1.5 µL of RNase solution. Samples were than incubated at 37 °C for 15 minutes followed by rehydration at 4 °C overnight. Replicate DNA samples were run on 0.75 % agarose TAE gel alongside GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific) at 120V for 40 minutes. The gel was then visualised in the Uvitec Gel-Documentation system to confirm the presence of long strand DNA.

Genomic DNA was subject to an additional clean up step using a 0.6:1 ratio of AMPure XP beads:sample prior to long read sequencing using the Oxford Nanopore Technologies' (ONT) MinION system. The sequencing library was prepared using ONT's ligation sequencing kit SQK-LSK109, as per the manufacturer's guidelines with modifications as follows: Incubation times for end repair steps were increased from 5 minutes to 30 minutes; ligation reactions were performed at room temperature for 1 hour, and elution steps were performed at 37 °C for 15 minutes. The resulting DNA libraries were sequenced on MinION R9.4.1 flow cells with a 48-hour run time. Basecalling was performed using Guppy V 3.5.2 software.

2.6.3. Genome assembly and annotation

Oxford Nanopore Technologies reads were filtered to those of length over 5 kb with SeqKit 0.11.0 (284) before being assembled with Canu 2.0 (285). The resulting genome assembly was filtered with Tapestry 1.0.0 (286) to 39 Mb, 21 contigs, before being polished with Medaka 0.11.3. Previously obtained Illumina reads were used to polish the assembly. Short read Illumina sequencing libraries were prepared using the NEBNext Ultra DNA library prep kit for Illumina (New England Biolabs), and sequenced on an Illumina HiSeq 2500, with paired end 100 bp reads, by the University of Leeds Next Generation Sequencing Facility. The Illumina reads were quality-checked with FastQC 0.11.7 (287) and adapter trimmed with Cutadapt 2.10 (288) and used for three rounds of Pilon 1.23 (289) polishing of the genome assembly. A previously obtained transcriptome assembly from NO1 grown on six lignocellulosic substrates (wheat straw, empty fruit bunches from palm oil, wheat bran, sugar cane bagasse, rice straw and kraft lignin) was used for genome annotation with FunAnnotate 1.8.1 and InterproScan 5.46 (290, 291).

2.6.4. Ascomycete genome annotation and CAZyme prediction

All available genome assemblies (n= 2635) of ascomycota origin were retrieved from the NCBI genome assembly database. Genome assemblies with N50 values > 1000 were retained and gene prediction was performed with FunAnnotate v1.8.1 (60), BUSCO (61), and AUGUSTUS (62), generating a final dataset of 2570 genomes. Predicted genes for each genome were annotated with the CAZyme database (v.09242921) and mean gene densities were then calculated for each taxonomic level for comparative analysis. Unique taxonomy identifiers (taxid) for each genome were retrieved from the NCBI taxonomy database using the Entrez NCBI API (292). No filtering was undertaken and a phylogenetic tree was reconstructed using ETE3 to retrieve the tree topology (get_topology) without intermediate nodes at a rank limit of genus (63) (**Figure 2.2**). Gene densities from annotations were mapped to the corresponding genomes on the tree. Genome metadata and annotations are available in **Supplementary File 2.2**.

The number and proportion of CAZyme domains in the genomes of *P. putredinis* NO1 (GCA_949357655.1), *S. boydii* (GCA_002221725.1), *T. reesei* (GCA_016806875.1), and *F. oxysporum* (GCA_023628715.1) were plotted using the 'ggplot2' package of R studio 3.6.3 (293, 294).

2.6.5. Sequence-based searches for LPMOs, laccases, and peroxidases

The sequences for an ascomycete AA9 family LPMO and for an AA1 family laccase were obtained from the CAZy database (177). An AA9 LPMO from *Aspergillus niger* (GenBank: CAK97151.1) and an AA1 Laccase from *A. niger* (GenBank: CAK37372.1) were used. For

peroxidase sequences, individual sequences for three types of reported lignin degrading peroxidases were obtained from the fPoxDB database (282). A manganese peroxidase from *Aureobasidium subglaciale* (GenBank: EJD50148.1), a lignin peroxidase from *F. oxysporum* f. sp. *lycopersici* (NCBI RefSeq: XP_018248194.1), and a versatile peroxidase from *Pyronema confluens* (Locus: PCON_11254m.01) only available from the fPoxDB database were used.

These sequences were searched against the *P. putredinis* NO1 genome protein sequences through command line BLAST with an E-value cut off of 1×10^{-5} (295). Results were compiled for the three classes of peroxidase.

2.6.6. Domain-based searches for LPMOs, laccases, and peroxidases

Due to the lack of online databases for LPMO sequences, the genome was searched for LPMO related sequences using the Pfam AA9 HMM (270).

Sequences for basidiomycete laccases and ascomycete Multicopper oxidases were downloaded from the Laccase Engineering Database 7.1.11 (278). These were aligned using Kalign 3.0 and this alignment subsequently used to generate a bespoke hidden Markov model (HMM) using the HMMER 3.2.1 programme (296, 297).

Sequences for manganese peroxidases, lignin peroxidases and versatile peroxidases were downloaded from the fPoxDB database (282). These were aligned and used to construct a bespoke HMM model as before.

These models were used to search the *P. putredinis* NO1 genome using HMMER 3.2.1 (297) and domain hits falling within the default significance inclusion threshold of 0.01.

2.6.7. Structure-based searches for LPMOs, laccases, and peroxidases

Predicted structure for >96 % (n=9611) of coding regions in the *P. putredinis* NO1 genome were modelled using AlphaFold v2.0.0 on the VIKING computer cluster (257).

The 9611 models of coding sequences were compiled into ‘tarball’ databases and compressed into ‘.tar.gz’ files on the VIKING cluster. These files were uploaded to the PDBefold online server to search against (298). Structures for the same sequences used in sequence-based searching were obtained from UniProt database if available (299), or modelled using AlphaFold v.2.00 on the VIKING computing cluster. These structures were searched against the *P. putredinis* NO1 structure database using PDBefold to identify similar structures in the *P. putredinis* NO1 genome. The ‘lowest acceptable match’ parameter was adjusted depending on the activity being searched with until coding regions not identified using sequence- or domain-based searching strategies were identified.

2.6.8. *In silico* investigation of candidate sequences

Sequences which were identified by structural searching solely were considered potentially interesting and warranted further investigation to attempt to elucidate function. Sequences were searched against the NCBI non-redundant protein database with default search parameters and an E-value cut off of 1×10^{-5} to investigate proteins with similar sequence (273). Domains were predicted using the primary amino acid sequence with the InterPro tool for domain prediction with default parameters (300). CAZyme domains were predicted with the online dbCAN prediction tool with default search parameters (178).Interesting candidate structures were further investigated with PDBefold by searching the structures against the whole PDB database to identify structurally similar proteins using a 'lowest acceptable match' parameter of 70% (274, 298). Secretion signals were predicted using SignalP 6.0 with default parameters (275). Altogether, this annotation information was used to investigate the potential functions of interesting sequences.

2.6.9. Transcriptomic data for interesting sequences

A previously published transcriptomic dataset for *P. putredinis* NO1 was used to validate expression of sequences of interest identified here during growth on lignocellulosic substrates (229). Gene expression data in transcripts per million (TPM) for all sequences identified solely by structural approaches and not by sequence- or domain-based searching and therefore considered to be interesting for all three activities explored here: LPMO, laccase, and peroxidase. Gene expression data is available in **Supplementary File 2.1**.

2.7. Declarations

2.7.1. Availability of data and materials

The sequence data generated and analysed during the current study are available in the European Nucleotide Archive, project code PRJEB60285, secondary accession ERP145344 (https://www.ebi.ac.uk/ena/browser/view/PRJEB60285). The WGS Sequence Set for the genome assembly is available in the European Nucleotide Archive, Accession CASHTG010000000.1 (https://www.ebi.ac.uk/ena/browser/view/CASHTG010000000). The assembly is also available through the NCBI database, Accession GCA_949357655.1 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_949357655.1).

2.7.2. Competing interests

The authors declare that they have no competing interests.

2.7.3. Funding

This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC), UK (Grant BB/1018492/1, BB/P027717/1, and BB/W000695/1). CS was supported by a CASE studentship from the BBSRC Doctoral Training Programme (BB/M011151/1) with Prozomix Ltd.

2.7.4. Author contributions

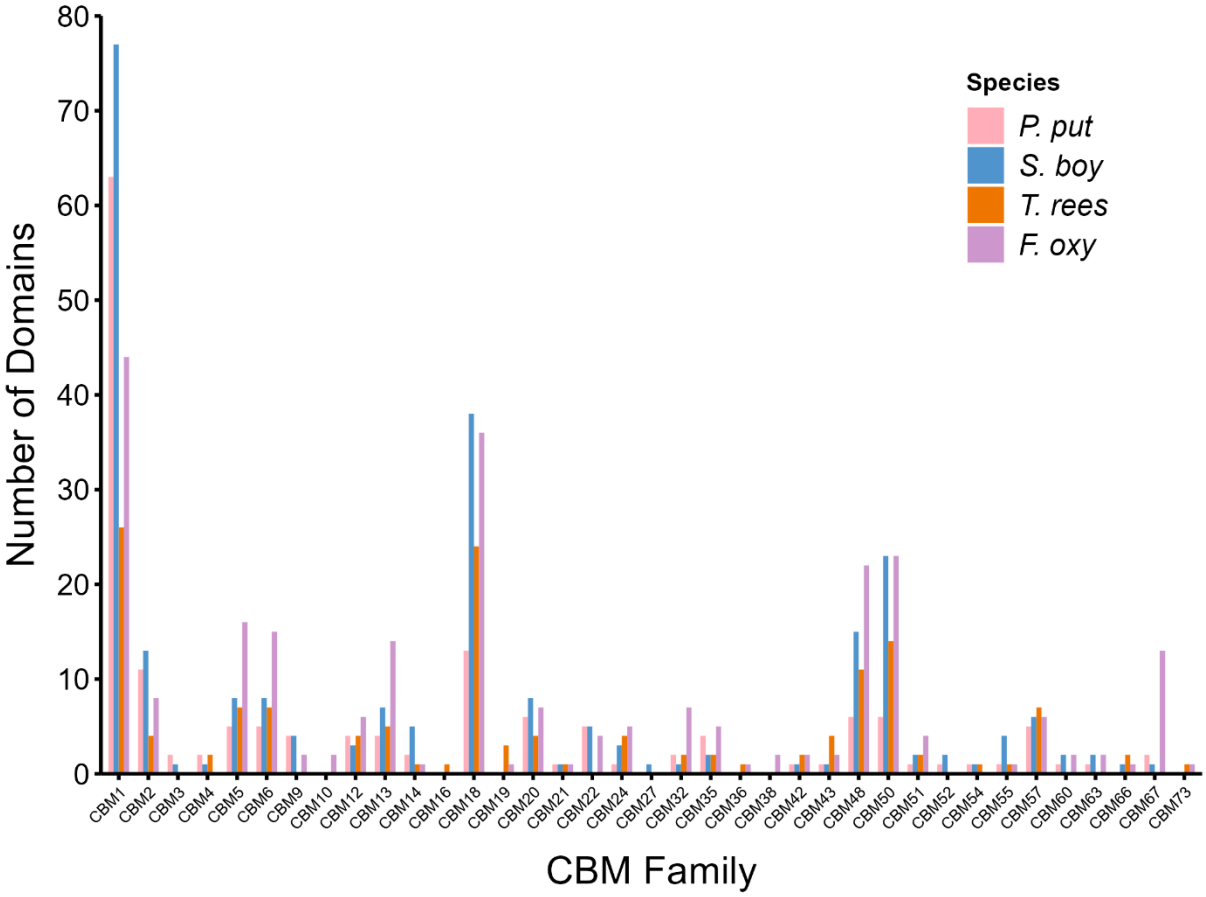
CJRS conceptualised the investigation carried out in this paper, extracted the genomic DNA from *P. putredinis* NO1, performed CAZyme repertoire comparison analysis, structurally modelled the *P. putredinis* NO1 genome, performed sequence-, domain-, and structure-based searches of the genome, analysed the search strategy results and was the major contributor in writing the manuscript. DRL carried out annotation of ascomycete genomes and CAZyme repertoire comparison analysis and was a major contributor to the writing of the manuscript. NCO was involved in maintaining *P. putredinis* NO1 and extraction of genomic DNA. SRJ library prepped and sequenced the *P. putredinis* NO1 genomic DNA. KN assembled the *P. putredinis* NO1 genome, performed initial annotation and aided deposition of sequence data. YL assembled the transcriptome that was used for annotation of the *P. putredinis* NO1 genome. NGSM was a contributor to the writing of the paper. SB carried out the Illumina sequencing which was used to polish the *P. putredinis* NO1 genome assembly. NCB was a major contributor to the conceptualisation and supervision of the study in addition to making a major contribution to the writing of the manuscript.

2.7.5. Acknowledgements

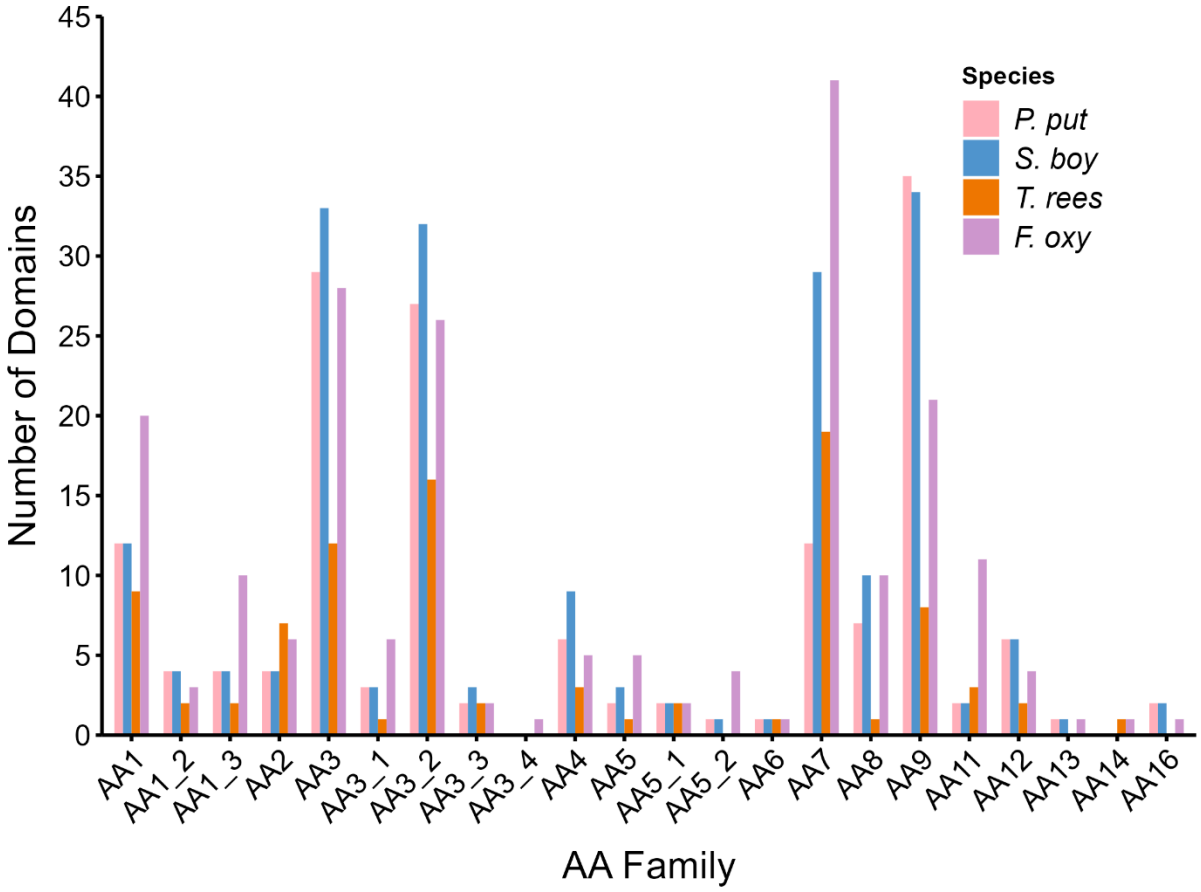
Special thanks to Sally James for performing the Nanopore sequencing in her kitchen in the first weeks of the COVID-19 pandemic, and to Katherine Newling for her immense help with all bioinformatic work and my endless questions.

2.8. Supplementary material

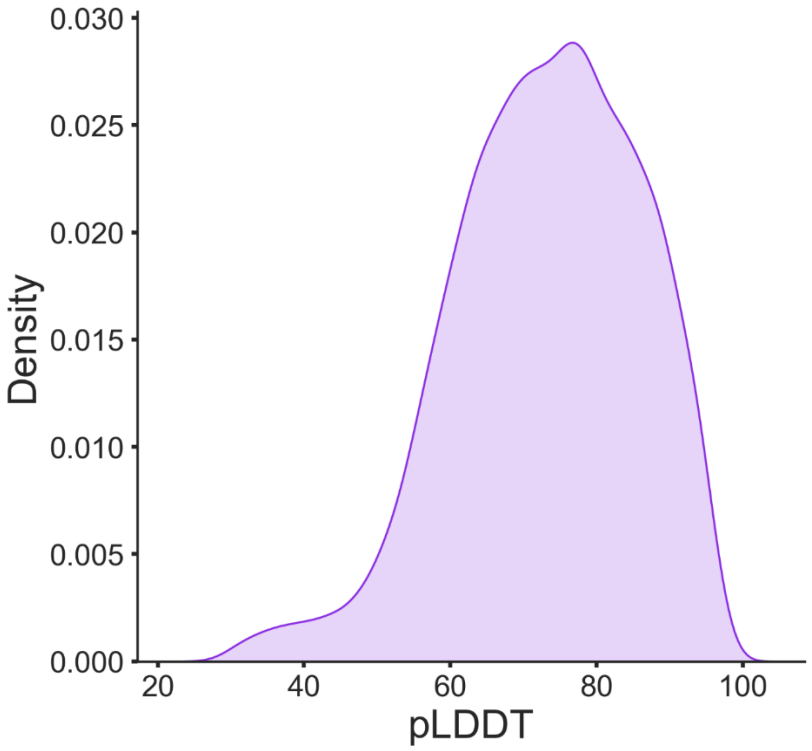
2.8.1. Supplementary figures



Supplementary Figure 2.1 Comparison of CBM class CAZyme repertoire. The number of CBM class CAZyme domains of each family for four lignocellulose degrading ascomycetes; *P. putredinis* NO1, *S. boydii*, *T. reesei*, and *F. oxysporum*.



Supplementary Figure 2.2 Comparison of AA class CAZyme repertoire. The number of AA class CAZyme domains of each family for four lignocellulose degrading ascomycetes; *P. putredinis* NO1, *S. boydii*, *T. reesei*, and *F. oxysporum*.



Supplementary Figure 2.3 Structural database pLDDT score distribution. Density plot for structural confidence predicted local distance difference test (pLDDT) score distribution. Data shown is from 9611 structural predictions.

2.8.2. Supplementary tables

Supplementary Table 2.1 Coding regions of LPMO related proteins identified through genome searching approaches with the sequence of an *A. niger* AA9 LPMO (E-value cut-off = 1 x 10^{-5}), the Pfam AA9 HMM (Significance threshold = 0.01), and the structure of the *A. niger* AA9 LPMO (Lowest percentage match = 50%).

Coding Region	GenBank Accession	Identified by Searching Approach			AA9 CAZyme
		Sequence	Domain	Structure	
FUN_000316-T1	CAI7987579.1	✓	✓	✓	✓
FUN_000653-T1	CAI7987917.1			✓	
FUN_000713-T1	CAI7987978.1			✓	
FUN_001321-T1	CAI7988928.1	✓	✓		✓
FUN_001635-T1	CAI7989580.1	✓	✓		✓
FUN_001667-T1	CAI7989636.1	✓	✓	✓	✓
FUN_001939-T1	CAI7990224.1	✓	✓		✓
FUN_002327-T1	CAI7991101.1	✓	✓		✓
FUN_002573-T1	CAI7991617.1		✓		
FUN_002628-T1	CAI7991711.1	✓	✓	✓	✓
FUN_002887-T1	CAI7992274.1	✓	✓		✓
FUN_002890-T1	CAI7992277.1			✓	
FUN_002962-T1	CAI7992399.1			✓	
FUN_003076-T1	CAI7992608.1	✓	✓	✓	✓
FUN_003190-T1	CAI7992922.1			✓	
FUN_003419-T1	CAI7993442.1		✓		✓
FUN_003436-T1	CAI7993459.1	✓	✓		✓
FUN_003437-T1	CAI7993460.1	✓	✓	✓	✓
FUN_003535-T1	CAI7993628.1		✓		
FUN_003783-T1	CAI7994168.1			✓	
FUN_004209-T1	CAI7995165.1	✓	✓	✓	✓
FUN_004243-T1	CAI7995223.1	✓	✓	✓	✓
FUN_004290-T1	CAI7995316.1		✓	✓	✓
FUN_004866-T1	CAI7996285.1	✓	✓	✓	✓
FUN_005222-T1	CAI7997468.1	✓	✓	✓	✓
FUN_006366-T1	CAI7999797.1	✓	✓	✓	
FUN_006413-T1	CAI7999893.1		✓		
FUN_006553-T1	CAI8000144.1			✓	
FUN_006658-T1	CAI8000562.1		✓	✓	✓
FUN_006983-T1	CAI8001138.1	✓	✓	✓	✓
FUN_007242-T1	CAI8001774.1		✓		
FUN_007243-T1	CAI8001776.1		✓		✓
FUN_007537-T1	CAI8002298.1	✓	✓		✓
FUN_007636-T1	CAI8002470.1	✓	✓	✓	✓
FUN_007666-T1	CAI8002525.1	✓	✓	✓	
FUN_008023-T1	CAI8003320.1	✓	✓	✓	✓

FUN_008106-T1	CAI8003467.1	✓	✓		
FUN_008107-T1	CAI8003469.1		✓		✓
FUN_008454-T1	CAI8004268.1	✓	✓	✓	✓
FUN_008468-T1	CAI8004296.1	✓	✓		✓
FUN_008688-T1	CAI8004812.1	✓	✓	✓	✓
FUN_008818-T1	CAI8005031.1	✓	✓		✓
FUN_009030-T1	CAI7989078.1	✓	✓	✓	✓
FUN_009239-T1	CAI7992001.1			✓	
FUN_009799-T1	CAI7999895.1		✓		✓
FUN_009919-T1	CAI8001775.1	✓	✓	✓	✓
FUN_009920-T1	CAI8001777.1		✓		✓
FUN_010012-T1	CAI8003342.1			✓	
FUN_010091-T1	CAI8004298.1	✓	✓	✓	✓

Supplementary Table 2.2 Coding regions of laccase related proteins identified through genome searching approaches with the sequence of an *A. niger* AA1 laccase (E-value cut-off = 1 x 10^{-5}), the bespoke laccase and multicopper oxidase HMM constructed from sequences from the laccase engineering database (Significance threshold = 0.01), and the structure of the *A. niger* AA1 laccase (Lowest percentage match = 30%).

Coding Region	GenBank Accession	Identified by Searching Approach			AA1 CAZyme
		Sequence	Domain	Structure	
FUN_000263-T1	CAI7987524.1		✓		
FUN_000580-T1	CAI7987844.1		✓		
FUN_000646-T1	CAI7987911.1		✓		
FUN_000758-T1	CAI7988025.1	✓	✓	✓	✓
FUN_000759-T1	CAI7988026.1	✓	✓		
FUN_000832-T1	CAI7988099.1		✓		
FUN_001183-T1	CAI7988671.1		✓		
FUN_001583-T1	CAI7989479.1		✓		
FUN_001846-T1	CAI7990063.1	✓	✓	✓	✓
FUN_002227-T1	CAI7990821.1		✓	✓	✓
FUN_002249-T1	CAI7990863.1	✓	✓	✓	
FUN_002296-T1	CAI7991047.1	✓	✓	✓	✓
FUN_002408-T1	CAI7991234.1	✓	✓	✓	✓
FUN_002874-T1	CAI7992258.1		✓		
FUN_003566-T1	CAI7993680.1	✓	✓	✓	✓
FUN_003732-T1	CAI7994085.1			✓	
FUN_003828-T1	CAI7994234.1		✓		
FUN_004259-T1	CAI7995254.1		✓		
FUN_004508-T1	CAI7995689.1	✓	✓	✓	✓
FUN_004577-T1	CAI7995802.1	✓	✓	✓	✓
FUN_004616-T1	CAI7995870.1		✓		
FUN_004739-T1	CAI7996089.1		✓		
FUN_005070-T1	CAI7996980.1	✓	✓	✓	✓
FUN_005132-T1	CAI7997298.1		✓	✓	
FUN_005520-T1	CAI7998008.1		✓		
FUN_006244-T1	CAI7999594.1		✓		
FUN_006620-T1	CAI8000270.1		✓		
FUN_006720-T1	CAI8000684.1		✓		
FUN_007228-T1	CAI8001746.1		✓		
FUN_007508-T1	CAI8002246.1		✓		
FUN_008329-T1	CAI8004041.1		✓		
FUN_009491-T1	CAI7995256.1		✓		

Supplementary Table 2.3 Coding regions of peroxidase related proteins identified through genome searching approaches with the sequences of an MnP from *A. subglaciale*, LiP from *F. oxysporum*, and VP from *P. confluens* (E-value cut-off = 1 x 10^{-5}), the bespoke peroxidase HMM contricted from MnP, LiP, and VP seqeunces in the fPoxDB database (Significance threshold = 0.01), and the structure of the same three peroxidases used for seqeunce searches (Lowest percentage match = 30%).

Coding Region	GenBank Accession	Identified by Searching Approach			AA2 CAZyme
		Sequence	Domain	Structure	
FUN_002995-T1	CAI7992466.1			✓	
FUN_003542-T1	CAI7993642.1			✓	
FUN_003618-T1	CAI7993895.1			✓	
FUN_004484-T1	CAI7995643.1			✓	
FUN_004903-T1	CAI7996357.1		✓		✓
FUN_004941-T1	CAI7996699.1	✓	✓	✓	✓
FUN_005340-T1	CAI7997684.1	✓	✓	✓	✓
FUN_008413-T1	CAI8004205.1			✓	
FUN_008923-T1	CAI7988420.1			✓	
FUN_009329-T1	CAI7993214.1			✓	

2.8.3. Supplementary files

Supplementary files are available through the pre-print of this manuscript:
https://doi.org/10.1128/spectrum.01035-23

Supplementary File 2.1 Gene expression of interesting sequences. Sequences identified solely by structural searching approaches were considered interesting and were searched for in transcriptomic data from triplicate cultures of *P. putredinis* NO1 grown on glucose or grown on wheat straw with samples taken at days 2, 4, and 10. Gene expression values are presented in transcripts per million (TPM).

Supplementary File 2.2 Ascomycete genome annotations. Annotations of all ascomycete genomes used in this analysis (n = 2570).

Chapter 3

3. A bioinformatic workflow for *in silico* secretome prediction with the lignocellulose degrading ascomycete fungus *Parascedosporium putredinis* NO1

Authors

Conor JR Scott[a], Daniel R Leadbeater[a], Neil C Bruce[a]

Affiliations

[a] Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, United Kingdom

Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/mmi.15144

3.1. Abstract

The increasing availability of microbial genome sequences provides a reservoir of information for the identification of new microbial enzymes. Genes encoding proteins engaged in extracellular processes are of particular interest as these mediate the interactions microbes have with their environments. However, proteomic analysis of secretomes is challenging and often captures intracellular proteins released through cell death and lysis. Secretome prediction workflows from sequence data are commonly used to filter proteins identified through proteomics but are often simplified to a single step and are not evaluated bioinformatically for their effectiveness. Here, a workflow to predict a fungal secretome was designed and applied to the coding regions of the *Parascedosporium putredinis* NO1 genome. This ascomycete fungus is an exceptional lignocellulose degrader from which a new lignin degrading enzyme has previously been identified. The “secretome isolation” workflow is based on two strategies of localisation prediction and secretion prediction each utilising multiple available tools. The workflow produced three final secretomes with increasing levels of stringency. All three secretomes showed increases in functional annotations for extracellular processes and reductions in annotations for intracellular processes. Multiple sequences isolated as part of the secretome lacked any functional annotation and make exciting candidates for novel enzyme discovery.

3.2. Keywords

Parascedosporium, Ascomycete, Secretome, Bioinformatics, Genome, CAZymes.

3.3. Background

Proteomics is the study of proteins within a sample and involves the use of techniques that provide high molecular specificity for a broad range of peptides in a single measurement (3). Proteins are predominantly responsible for biological functions, and therefore acquiring qualitative and quantitative data on proteins can help us understand microbiological processes, such as lignocellulose breakdown (4).

Fungi are exceptional wood degraders and produce an array of bioproducts, including secreted enzymes used in commercial enzyme cocktails for the valorisation of lignocellulosic biomass. *Parascedosporium putredinis* NO1 is an ascomycete fungus from which new extracellular lignin degrading enzymes have previously been identified (229). With an available reference genome containing 9998 protein coding sequences for *P. putredinis* NO1, more effective proteomic experiments can now be carried out to help to identify the full repertoire of enzymes secreted by the fungus to facilitate growth on lignocellulose (301).

A “secretome” is defined as the set of proteins secreted by a cell or an organism at a given time (163). However, extracellular protein studies are not straightforward and often unavoidably identify contaminant intracellular proteins in abundance because of cell-death and lysis (6). For secretomic investigations this adds an additional layer of complexity and redundancy, especially for novel enzyme identification. Bioinformatic techniques could instead be used to filter proteomic data to predict in silico secretomes, allowing extracellular processes to be more clearly and accurately understood, and simplifying new enzyme identification.

Previous attempts to predict *in silico* fungal secretomes have used single prediction tools, such as the prediction of secretion signal peptides (8-10). However, proteins with signal peptides may be targeted to secretory pathways, but not necessarily secreted (302). Additionally, fungal protein secretion is more complex as proteins can be secreted via conventional or unconventional pathways (163). For example, it has been demonstrated that various metabolic enzymes are secreted by fungi despite the absence of secretion signals (303). Perhaps the most well investigated method of fungal unconventional protein secretion is through vesicles, which are utilised by fungi as efficient vehicles for the release of proteins into the extracellular environment, along with polysaccharides and pigments (304). Other secretomic investigations attempt to create basic workflows for *in silico* secretome prediction (163, 305). However, these often lack diversity in the tools used in each step and fail to confirm their effectiveness bioinformatically. As such, many available datasets are considered (meta-) exo-proteomes as they include contaminant intracellular proteins due to the lack of secretome identification pipelines.

Here, a bioinformatic workflow was designed to isolate sequences of the *P. putredinis* NO1 genome predicted to be secreted. The workflow is built around two initial strategies of prediction: localisation prediction and secretion prediction. In both strategies, more than one tool is used to capture secretome sequences which may be missed by a single tool alone. The effectiveness of each tool to predict a subset of sequences enriched in sequences encoding enzymes known to be extracellular and secreted was evaluated through annotation of sequences for COG category, CAZyme class, and KEGG metabolic pathways. Three resulting secretomes were produced with increasing levels of stringency on the sequences included: relaxed, strict, and super strict. All secretomes contained greatly reduced numbers of sequences compared to the total number of sequences in the *P. putredinis* NO1 genome and showed increases in annotations for extracellular functions. Subsets containing 1933, 812, or 509 sequences were produced for the relaxed, strict, or super strict secretomes, respectively. This will allow comparative investigations with proteomic data to be more accurate, and identification of enzymes and other new proteins to be made much simpler.

3.4. Results and discussion

3.4.1. Designing a workflow to isolate the *P. putredinis* NO1 secretome

The first genome of the genus *Parascedosporium* provides a unique resource to explore for potentially new enzymes. The genome assembly is 39 Mb, consists of 21 contigs, and contains 9998 protein coding sequences. *P. putredinis* NO1 belongs to the Microascaceae family of ascomycete fungi and is the sister taxon of *Scedosporium* species, of which genomes for four species are available. The genomic repertoires of predicted lignocellulose degrading enzymes were compared previously between *P. putredinis* NO1 and *Scedosporium boydii* and were found to be very similar (301). To identify and isolate sequences in the *P. putredinis* NO1 genome predicted to encode proteins which are actively secreted into the extracellular space, a secretome isolation workflow was designed (**Figure 3.1**). Localisation prediction was performed using the tools DeepLoc and BUSCA to identify sequences predicted to encode extracellular proteins. SignalP, TargetP, and SecretomeP were used to identify sequences predicted to contain proteins with secretion signal peptides. SecretomeP was simultaneously used to attempt to predict sequences encoding non-classically secreted proteins. All sequences were submitted to TMHMM for the prediction of sequences encoding proteins containing transmembrane helices and therefore transmembrane proteins. Multiple annotation strategies were used to build information on the potential functions of all sequences of the *P. putredinis* NO1 genome and were used for evaluation of the workflow.

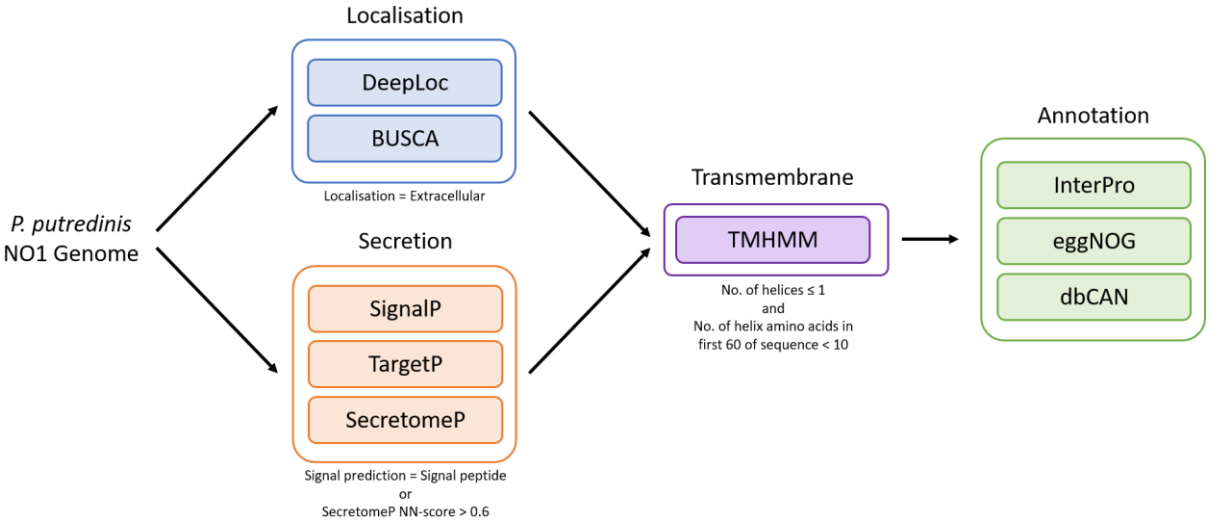


Figure 3.1 Secretome isolation workflow. The bioinformatic workflow developed to identify and isolate coding regions of the *P. putredinis* NO1 genome predicted to encode actively secreted extracellular proteins.

3.4.2. Investigating discrepancies in prediction tools of the secretome isolation workflow

Differences in the sequences captured by each tool were observed for both localisation and secretion branches of the workflow, highlighting the importance of utilising multiple tools during secretome prediction. This also reflects the differences in conventional and unconventional release of extracellular proteins in fungi. Proteins favouring vesicle-mediated release and without the presence of secretion signal peptides would be missed if only using tools for conventionally secreted protein prediction (303, 304). This is especially relevant here in the context of lignocellulose breakdown as vesicle-mediated secretion of lignocellulose-degrading enzymes has been demonstrated for *Trichoderma reesei*, another ascomycete degrader of plant biomass (306). In total, 769 coding regions of the *P. putredinis* NO1 genome were predicted to be extracellular by DeepLoc. Considerably more sequences at 1588 were predicted to be extracellular by BUSCA and 622 of these sequences were identified by both tools (**Figure 3.2A**). Comparing prediction of classical secretion signal peptides across the three secretion tools revealed 798 protein sequences predicted to contain signal peptides by all three tools, 975 by at least two tools, and 411 exclusively by SecretomeP (**Figure 3.2B**). SecretomeP also predicted an unusually large number of sequences, 3226, to encode non-classically secreted proteins. These predictions showed little overlap with sequences predicted to encode classically secreted proteins by SignalP and TargetP (**Supplementary Figure 3.1**). This suggested overprediction for non-classically secreted proteins with this tool.

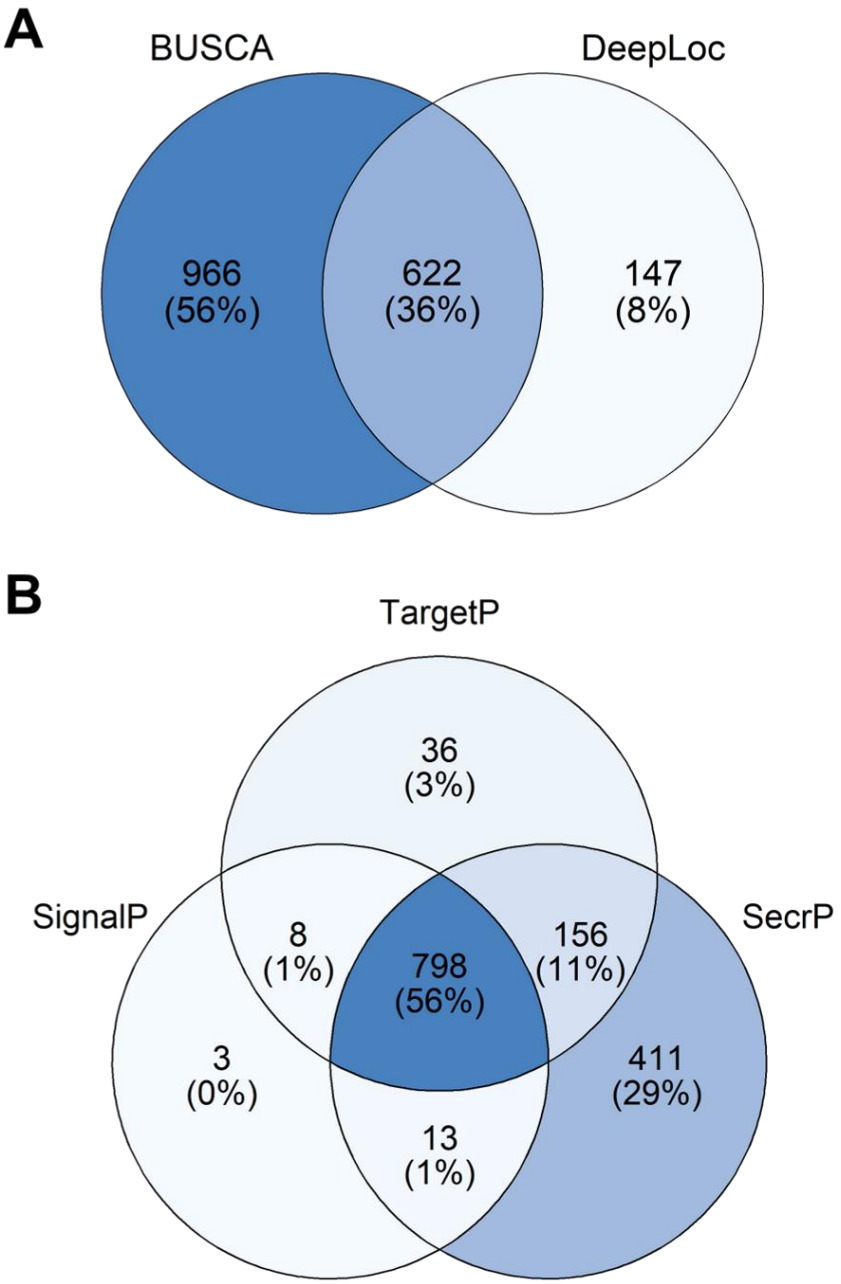


Figure 3.2 Visualising discrepancies between prediction tools. The differences in sequences predicted encode extracellular proteins by each localisation prediction tool (**A**), and sequences predicted to encode classically secreted proteins by each secretion prediction tool (**B**).

3.4.3. Evaluating individual tools of the secretome isolation workflow

COG category assignments and KEGG metabolic pathway terms from eggNOG mapper, and CAZyme class annotations from dbCAN were used to evaluate the effectiveness of the workflow. Proportions of each annotation in the whole genome were compared to the subsets of proteins from each of the prediction tools. Both localisation prediction tools, and all classical secretion prediction tools demonstrated expected increased proportions of functional annotations associated with proteins of fungal secretomes. These tools demonstrated increased proportions of sequences assigned to COG category G for carbohydrate metabolism, which is expected as carbohydrate breakdown begins outside of the cell. The tools showed reductions in the proportion of intracellular Glycosyl Transferase (GT) class CAZymes (264). Finally, all tools except for classical secretion prediction by SecretomeP showed increased proportions of assignments to the KEGG metabolic pathway for carbohydrate metabolism. Other increases and reductions were observed for all validation methods respective to the whole genome annotation, and these patterns varied by tool (**Supplementary Figure 3.2**, **Supplementary Figure 3.3**, and **Supplementary Figure 3.4**). This is likely a reflection of the different sequences captured by each tool due to the different prediction methods. For new enzyme identification, this is important for capturing as much of the secretome as possible. Sequences predicted to encode non-classically secreted proteins by SecretomeP did not show the expected increases in proportions. SecretomeP is designed for bacterial or mammalian sequences but has been used for fungal secretome prediction before and so was investigated here (163). However, its inability to accurately isolate sequences encoding secretome proteins, and due to the unusually large number of sequences captured by SecretomeP, non-classical secretion was omitted from secretome isolation. The tools used in this workflow could be readily applied to any sequence data from other fungal species or other microorganisms and the tools contain options for either eukaryotic or prokaryotic prediction. The removal or incorporation of transmembrane proteins could be altered based on the purpose of the in silico secretome isolation or based on the microorganism of interest. For example, in the context of lignocellulose breakdown some anaerobic fungi and bacteria have been demonstrated to assemble extracellular cell surface tethered constructs known as cellulosomes to enhance lignocellulose dissolution and product uptake (307). Therefore, the membrane-associated proteins involved in these structures may be of interest for investigations of these microorganisms.

3.4.4. Filtering the *P. putredinis* NO1 genome to isolate the secretome

Discrepancies between prediction tools inspired the creation of multiple “secretome” subsets with different levels of stringency. Sequences predicted to encode extracellular proteins and

sequences predicted to encode secretion signals were merged into three final subsets: relaxed, strict, and super strict. For each subset, proteins predicted to encode transmembrane proteins by TMHMM were removed to give the final “secretomes”. Conventionally, fungal secretomes consider membrane bound extracellular proteins (1), however for the purpose of investigating lignocellulose breakdown by aerobic fungi these were omitted as the free extracellular proteins are more likely to be involved in depolymerisation reactions. If membrane bound proteins are of interest, then this step could be removed. The “relaxed” secretome subset contained coding regions predicted to encode extracellular proteins by at least one localisation prediction tool or predicted to encode secreted proteins by at least one secretion prediction tool, totalling 1933 sequences. The “strict” secretome subset contained coding regions predicted to encode extracellular proteins by both localisation tools or sequences predicted to encode secreted proteins by all three secretion signal prediction tools, totalling 812 sequences. Finally, the ”super strict” secretome contains sequences predicted to encode extracellular proteins by both localisation prediction tools and which were also predicted to encode secretion signals by all secretion prediction tools, totalling 509 sequences.

To evaluate the secretomes; COG category, CAZyme class, and KEGG pathway proportions were again compared for each secretome against the whole genome. All secretome subsets demonstrated increased proportions of protein sequences assigned to COG categories G (carbohydrate metabolism), O (Post-translational modification / turnover / chaperone functions), and X (Unassigned) (**Figure 3.3**). For category G, the degree of increase correlated with the strictness of the secretome subset (Genome: 5.8 %, Relaxed: 12.7 %, Strict: 20.0 %, Super Strict: 23.0 %), the same pattern was observed for category O (Genome: 5.1 %, Relaxed: 6.4 %, Strict: 10.3 %, Super Strict: 12.5 %), whereas for category X it was the strict secretome that had the largest increase (Genome: 24.5 %, Relaxed: 34.0 %, Strict: 37.8 %, Super Strict: 37.1 %). Protein sequences with no clear functional annotation were abundant in the *P. putredinis* NO1 genome and were assigned to category X for comparison with other functional categories. The large proportion of these proteins in the *P. putredinis* NO1 genome reflects the novelty of this organism with this being the first genome assembly of the *Parascedosporium* genus. These proteins also represent a reservoir of potentially interesting new sequences and even new activities. For many of the other COG categories, reductions were observed compared to the genome for all secretome subsets, and again the degree of reduction increased with how strictly the secretomes were filtered. Regarding some COG categories, differences between the secretomes were observed. For COG category E (Amino acid metabolism and transport), a slight increase in proportions of assignments was observed for the relaxed and strict secretomes but a

reduction was seen in the super strict secretome (Genome: 4.5 %, Relaxed: 5.9 %, Strict: 4.7 %, Super Strict: 3.9 %). Only the relaxed secretome showed an increase in the proportion of assignments to category Q (Secondary structure) (Genome: 3.7 %, Relaxed: 4.0 %, Strict: 1.8 %, Super Strict: 1.4 %). The reasons for the increases in these COG categories which are not expected to include extracellular enzymes were not clear but may be due to mis-assignment of secretome proteins to these categories. Importantly, all secretomes showed increased proportions of proteins which lacked any functional annotations (i.e., assigned to category X). All secretomes also contained large proportions of protein sequences assigned to COG Category S (Function Unknown) (Genome: 19.9 %, Relaxed: 20.2 %, Strict: 18.3 %, Super Strict: 17.2 %). Protein sequences assigned to this category were found to have predicted putative domains but lacked an overall functional annotation. Altogether, the sequences belonging to S and X categories represent an important subset of sequences for new enzyme identification and the persistence of these proteins in all secretomes demonstrates how such a bioinformatic workflow can isolate a subset of protein sequences which potentially contains new enzymes and activities. Considering the lignocellulose-degrading lifestyle of *P. putredinis* NO1 and the previous identification of a new secreted phenol oxidase enzyme involved in lignocellulose breakdown, it can be hypothesised that this fungus may contain other new lignocellulose-degrading enzyme activities (229). Indeed, the protein sequence encoding this new enzyme was isolated in all secretomes and assigned to COG category S, inspiring confidence that other new enzymes belonging to S and X categories have also been isolated.

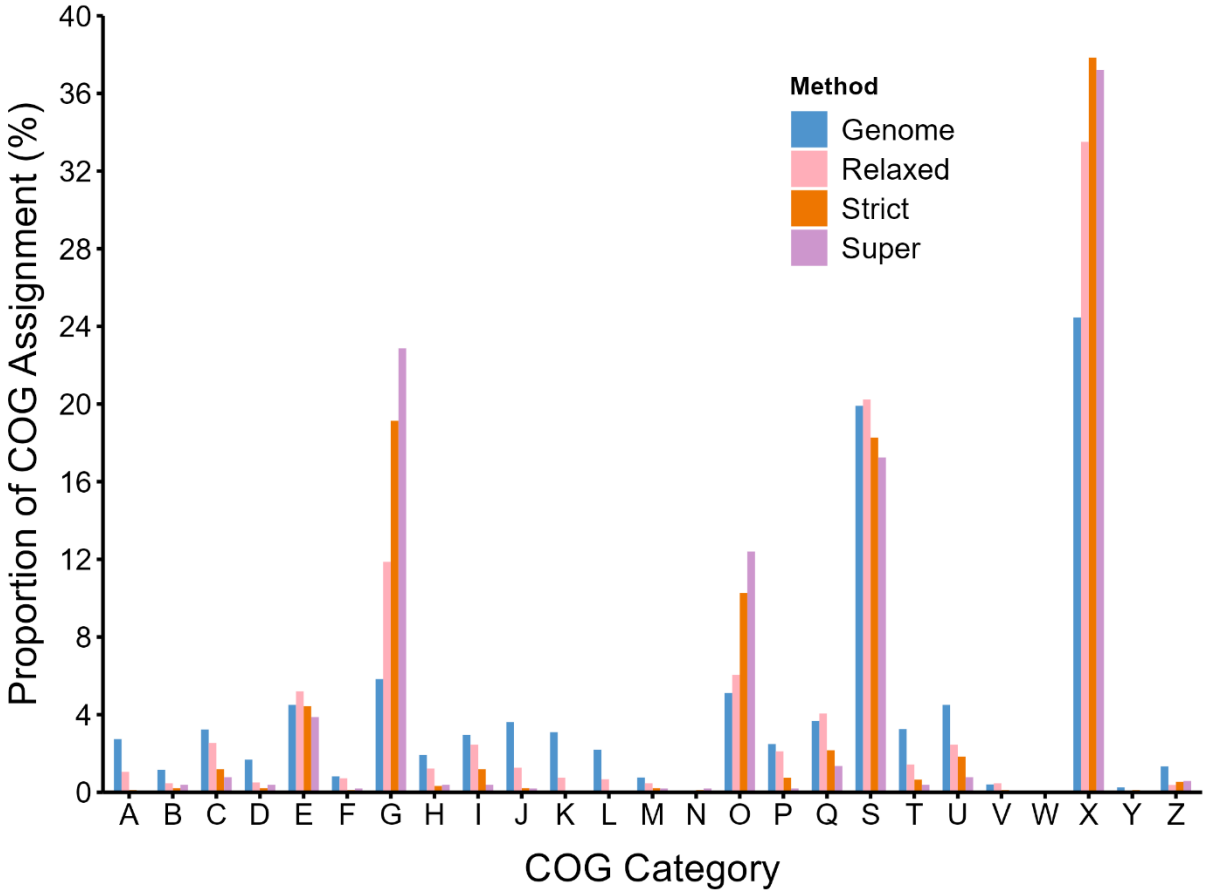


Figure 3.3 Investigating COG annotation proportions for secretome protein sequences. The proportion of each COG Category in the whole genome COG annotation compared to the relaxed, strict, and super strict secretome subsets. A = RNA processing and modification, B = Chromatin structure, C = Energy production and conversion, D = Cell cycle control and mitosis, E = Amino acid metabolism and transport, F = Nucleotide metabolism and transport, G = Carbohydrate metabolism and transport, H = Coenzyme metabolism, I = Lipid metabolism, J = Translation, K = Transcription, L = Replication and repair, M = Cell wall/membrane/envelope biogenesis, N = Cell motility, O = Post-translational modification/turnover/chaperone functions, P = Inorganic ion transport and metabolism, Q = Secondary structure, T = Signal transduction, U = Intracellular trafficking and secretion, Y = Nuclear structure, Z = Cytoskeleton, S = Function Unknown, X = Unassigned.

When comparing CAZyme class annotation proportions across the secretomes and the whole genome, all secretomes had largely reduced proportions of intracellular glycosyl transferase (GT) class CAZymes (**Figure 3.4**) (264). In particular, the super strict secretome contained no GT class CAZymes (Genome: 14.2 %, Relaxed: 3.1 %, Strict: 1.3 %, Super Strict: 0.0 %). Mirroring this GT reduction, all secretomes had increased proportions of carbohydrate esterase (CE) (Genome: 6.4 %, Relaxed: 8.5 %, Strict: 10.0 %, Super Strict: 11.8 %), carbohydrate binding module (CBM) (Genome: 20.3 %, Relaxed: 25.6 %, Strict: 29.8 %, Super Strict: 32.3 %), and polysaccharide lyase (PL) class CAZyme sequences (Genome: 2.3 %, Relaxed: 3.3 %, Strict: 3.6 %, Super Strict: 3.9 %). These increases

correlated with the strictness of the filtering used to obtain each secretome. All secretomes had increased proportions of auxiliary activity (AA) class CAZymes (Genome: 20.4 %, Relaxed: 23.6 %, Strict: 25.4 %, Super Strict: 21.1 %), however the super strict secretome had the smallest increase. As these CAZyme classes all predominantly act on extracellular substrates, this was expected. A reduction in glycoside hydrolase (GH) class proportions was seen for all secretomes with the strict subset showing the largest reduction (Genome: 36.5 %, Relaxed: 35.8 %, Strict: 29.6 %, Super Strict: 30.8 %). This is the largest class of CAZymes, and many members of this class have been suggested to have intracellular activities previously (201). Therefore, a reduction in the proportion of GH CAZymes may be expected.

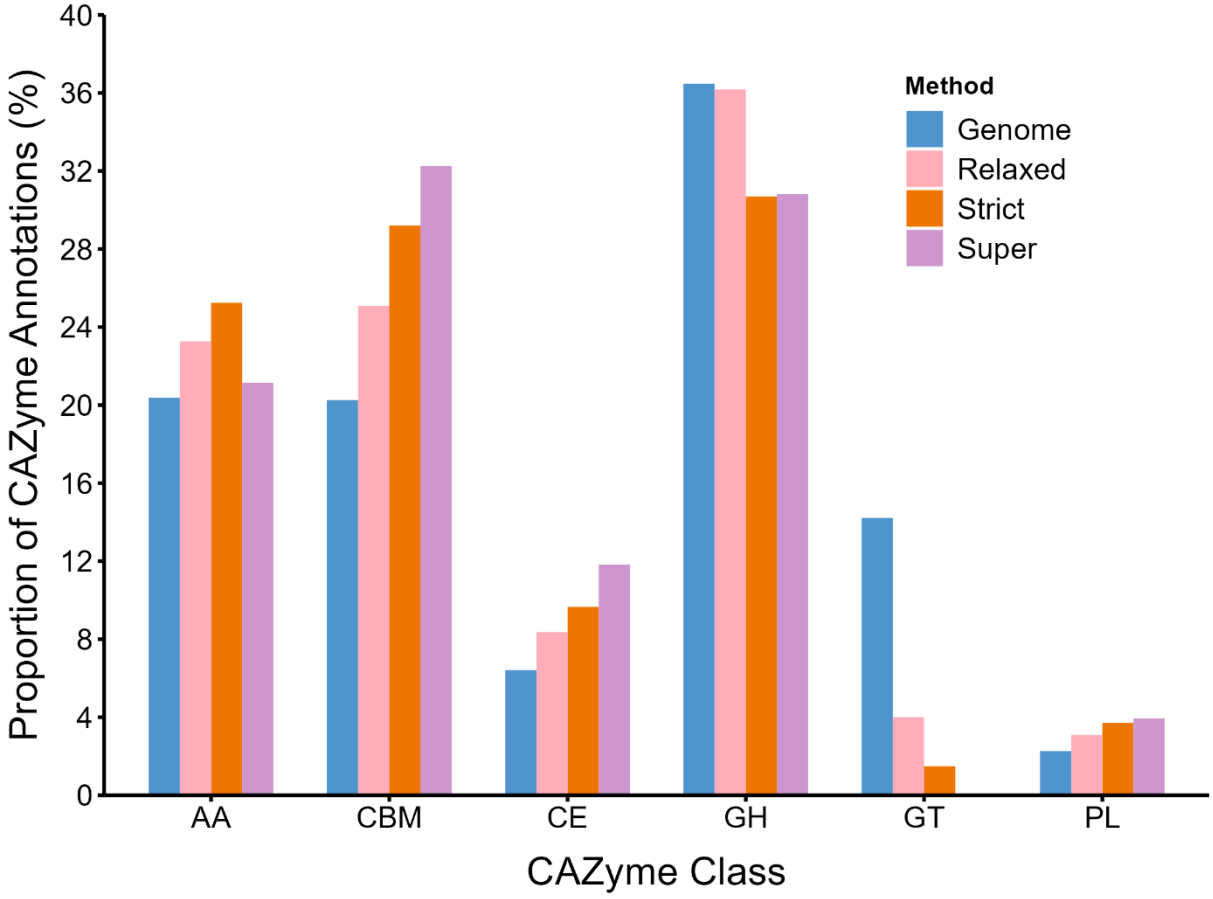


Figure 3.4 Investigating CAZyme annotation proportions for secretome protein sequences. The proportion of each CAZyme Class in the whole genome CAZyme annotation compared to the relaxed, strict, and super strict secretome subsets. AA = Auxiliary activity, CBM = Carbohydrate binding module, CE = Carbohydrate esterase, GH = Glycoside hydrolase, GT = Glycosyl transferase, PL = Polysaccharide lyase.

KEGG metabolic pathway assignments were found to be less clear than COG categories and CAZyme class annotations and tended to differ more between tools. However, expected increases in the proportions of assignments to metabolic classes were still observed. All

secretomes show increased proportions of assignments to the carbohydrate metabolism pathway (Genome: 21.2 %, Relaxed: 21.5 %, Strict: 28.8 %, Super Strict: 38.9 %), however this is only a slight increase for the relaxed secretome (**Figure 3.5**). A similar pattern was observed for the metabolism of other amino acids pathway (Genome: 4.2 %, Relaxed: 4.6 %, Strict: 6.3 %, Super Strict: 9.3 %). For other pathways, differences were observed between the secretomes. For example, no assignments to the xenobiotics biodegradation and metabolism pathway were present in the super strict secretome, although the strict secretome only showed a slight reduction in proportion of assignments to this pathway compared to the genome, and the relaxed secretome even showed an increase (Genome: 5.5 %, Relaxed: 9.7 %, Strict: 4.5 %, Super Strict: 0.0 %). Only the strict secretome showed an increased proportion of assignments to the metabolism of terpenoids and polyketides pathway (Genome: 2.1 %, Relaxed: 1.7 %, Strict: 3.6 %, Super Strict: 1.9 %). The strict and super strict secretomes were observed to have increased proportions of assignments to the glycan biosynthesis and metabolism pathway (Genome: 6.3 %, Relaxed: 4.3 %, Strict: 9.0 %, Super Strict: 9.3 %). These assignments may be the result of mis-assignment due to the action of many extracellular fungal enzymes on the β-glycosidic linkages which are present in glycans (308). The super strict secretome had a large increase in the proportion of assignments to the biosynthesis of other secondary metabolites, whereas the relaxed and strict secretomes only showed slight increases (Genome: 3.4 %, Relaxed: 4.6 %, Strict: 4.5 %, Super Strict: 9.3 %). The relaxed and strict secretome showed increases in assignments to amino acid metabolism pathways compared to a reduction in assignments to these pathways for the super strict secretome (Genome: 19.6 %, Relaxed: 21.8 %, Strict: 20.7 %, Super Strict: 14.8 %).

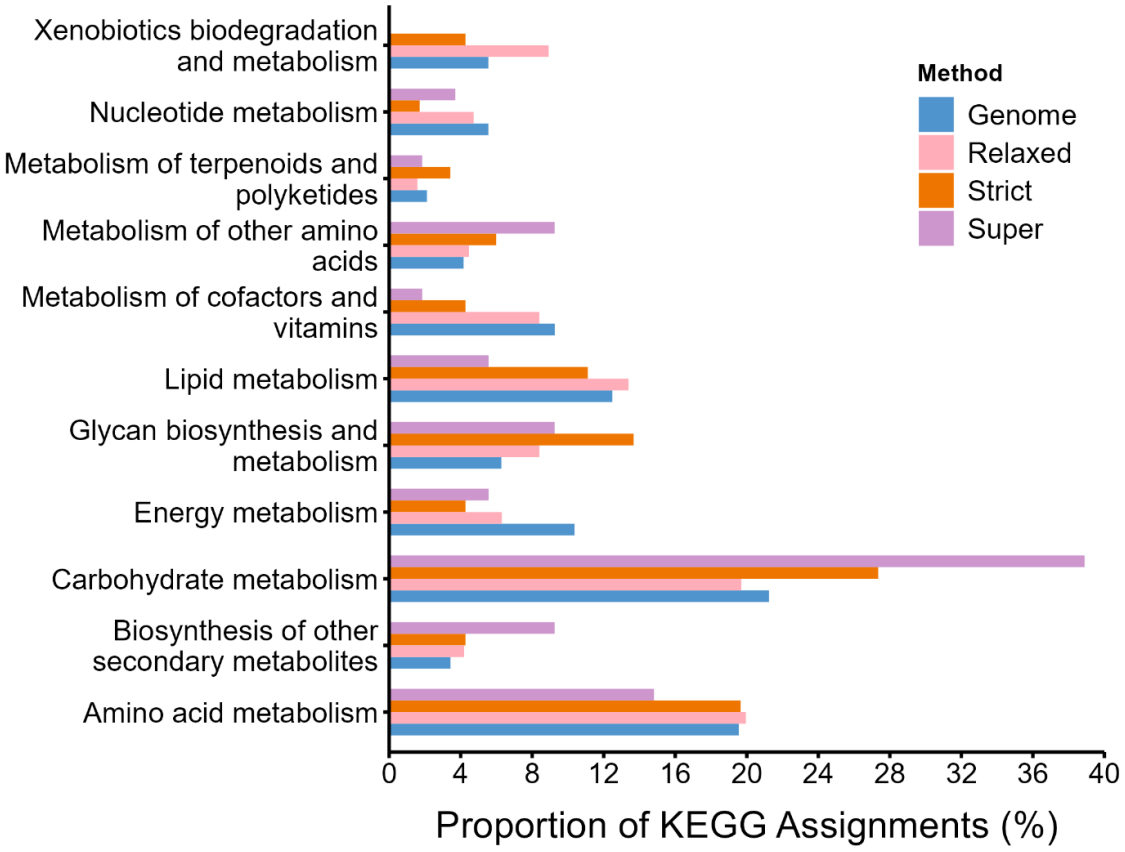


Figure 3.5 Investigating KEGG pathway annotation proportions for secretome protein sequences. The proportion of KEGG assignments to metabolic pathways in the whole genome compared to the relaxed, strict, and super strict secretome subsets.

To determine and quantify the effect of the workflow on the *P. putredinis* NO1 protein dataset, principal component analysis of K-Means clustered outputs for each of the tools was performed (**Figure 3.6**). The first, second and third components explained 40.25%, 20.69% and 11.29% of the total variance respectively. Cluster 1 (pink; N=3583; SecretomeP outlier subset) contained proteins exclusively with positive identifications from SecretomeP which were proteins omitted from the original workflow. Cluster 2 (blue; N=1224; secretome subset) clustered distinctly separately from the other clusters and consisted of proteins with positive secretion results from all tools and as such represents the identified secretome. Cluster 3 (orange; N=1169; transmembrane subset) contained proteins weighted toward both positive DeepLoc outputs and a higher number of TMHMM helices suggesting this cluster contained predominantly transmembrane proteins. Cluster 4 (purple; N=4023; intracellular) contained proteins with little to no positive identifications from any tools and as such represent the intracellular fraction. While further resolution in subcellular localizations could be achieved with this data, these results support the identification of the putative secretome with this workflow. Overall, the expected patterns were observed for all three secretomes which suggests an ability of this workflow to capture secretome proteins.

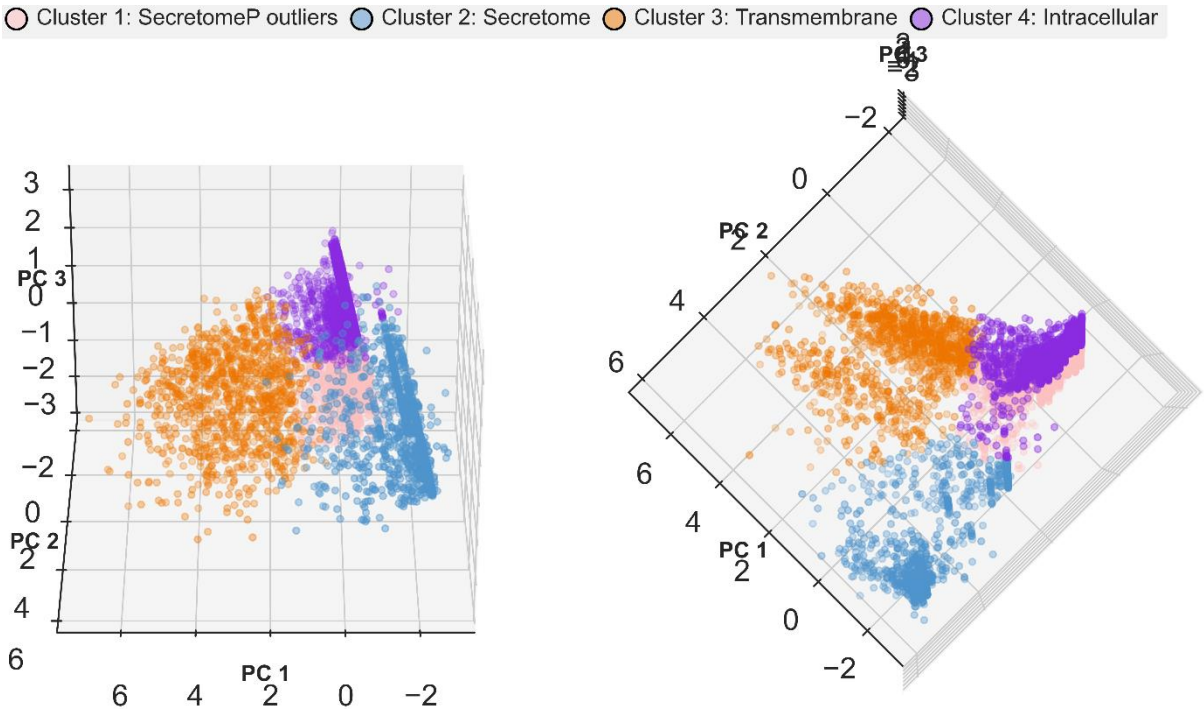


Figure 3.6 Principal component analysis of K-means clustered predictions for protein localisation. K-Means clustering was performed using protein localisation predictions generated for each protein (n=9998) from workflows tools; BUSCA, DeepLoc, SignalP, SecretomeP, TargetP and TMHMM. Clusters (n=4) are coloured by K-Means cluster. Subplots are ordinated for clarity. PC; Principal component.

3.5. Conclusions

Here, a workflow for predicting a fungal *in silico* secretome from genomic sequences was presented and evaluated. The workflow incorporated two strategies of localisation and secretion prediction both using multiple tools. Each tool demonstrated patterns in proportions of annotations associated with secretome proteins compared to the whole genome annotation. This included increased proportions of assignments to pathways for carbohydrate metabolism and reduced proportions of intracellular GT class CAZymes. There were also differences in patterns of proportional increases and reductions across the tools for each of the evaluation methods. This demonstrated the importance of incorporating multiple tools into secretome prediction workflows. The expected patterns were also observed in the final filtered secretomes. Again, differences were observed between the secretomes, however the main patterns were observed for all secretomes, and intensity correlated with the stringency of filtering.

Depending on the purpose of the *in silico* secretome, it is envisaged that different secretomes could be used. For example, for new enzyme identification the relaxed secretome would be more appropriate to maximise captured sequences where novel sequences are unlikely to be captured by all secretome prediction tools. For the identification of new extracellular enzymes where functional annotation is impossible due to high sequence or structural divergence, bioinformatic workflows like that presented here can quickly and simply allow these enzymes sequences to be isolated. Indeed, sequences assigned to COG categories for unknown functions were present in all secretomes alongside the new phenol oxidase enzyme identified previously from this fungus (229). This suggested that this workflow can effectively isolate subsets of sequences encoding potentially new enzymes and activities. Combination with proteomic data would reduce the number of sequences further. In contrast, for comparative secretomic studies the stricter secretomes would be favourable as confidence is increased that most of the sequences in these secretomes are truly secreted and extracellular. The workflow is readily adaptable across eukaryotic and prokaryotic organisms, as all tools used here have options for predictions from bacterial sequences. Modifications can be made through decisions on incorporation or removal of protein sequences at each stage of the workflow, for example whether to include extracellular transmembrane protein sequences as part of the predicted secretome. As easily as extracellular proteins can be identified they can also be removed if interest is instead focused on intracellular proteins. As new tools are developed and become popularised, they can easily be incorporated into the workflow.

3.6. Materials and methods

3.6.1. Localisation prediction

Localisation prediction was carried out for all predicted coding regions of the *P. putredinis* NO1 genome using the online tool BUSCA (309), and with DeepLoc v1.0 on the high performance computing cluster at the University of York (310).

3.6.2. Secretion prediction

Secretion signal prediction was carried out for all predicted coding regions of the *P. putredinis* NO1 genome using the online tools SignalP v6.0 (275) and TargetP v2.0 (311). Secretion signal prediction and simultaneous non-classical secretion prediction was performed using the online tool SecretomeP v2.0, where sequences not predicted to encode signal peptides but with an NN-score > 0.6 were predicted to be non-classically secreted (312).

3.6.3. Transmembrane helices prediction

Transmembrane helix prediction for the identification of transmembrane proteins was performed for all predicted coding regions of the *P. putredinis* NO1 genome using the online tool TMHMM v2.0 (313). Sequences predicted to encode more than one transmembrane helix were assumed to be transmembrane proteins. Sequences predicted to encode a single transmembrane helix, but with less than 10 amino acids of this helix occurring in the first 60 amino acids of the protein sequence (indicating a signal peptide) were also assumed to be transmembrane proteins and were also removed.

3.6.4. Sequence annotation

Sequences of all predicted coding regions in the *P. putredinis* NO1 genome were annotated for COG categories and KEGG pathway annotations using the online tool eggNOG mapper v2 (314). CAZyme domain annotation with dbCAN (178) of all predicted coding regions was performed using the CAZyme database v09242921 as described previously (301).

3.6.5. Prediction clustering

K-Means cluster principal component analysis (PCA) was performed with scikit-learn (SKlearn) (315) on output data from each of the tools. Categorical variables were factorized into secretion positive results (1) and secretion negative results (0). Numerical outputs (TMHMM First60 and helices and SecretomeP NN values) were unchanged. Within-Cluster Sum of Squares indicated a four cluster solution was optimal.

3.6.6. Secretome isolation

The database of localisation prediction, secretion signal prediction, transmembrane helix prediction, and annotation for all predicted coding regions of the *P. putredinis* NO1 genome was assembled in R v4.2.0 (293). Evaluation of each tool used was performed using annotation information and plotted in R using the ggplot2 package (294).

3.7. Declarations

3.7.1. Availability of data and materials

The sequence data generated and analysed during the current study are available in the European Nucleotide Archive, project code PRJEB60285, secondary accession ERP145344 (https://www.ebi.ac.uk/ena/browser/view/PRJEB60285). The WGS Sequence Set for the genome assembly is available in the European Nucleotide Archive, Accession CASHTG010000000.1 (https://www.ebi.ac.uk/ena/browser/view/CASHTG010000000). The assembly is also available through the NCBI database, Accession GCA_949357655.1 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_949357655.1).

3.7.2. Competing interests

The authors declare that they have no competing interests.

3.7.3. Funding

This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC), UK (BB/W000695/1). CS was supported by a CASE studentship from the BBSRC Doctoral Training Programme (BB/M011151/1) with Prozomix Ltd.

3.7.4. Author contributions

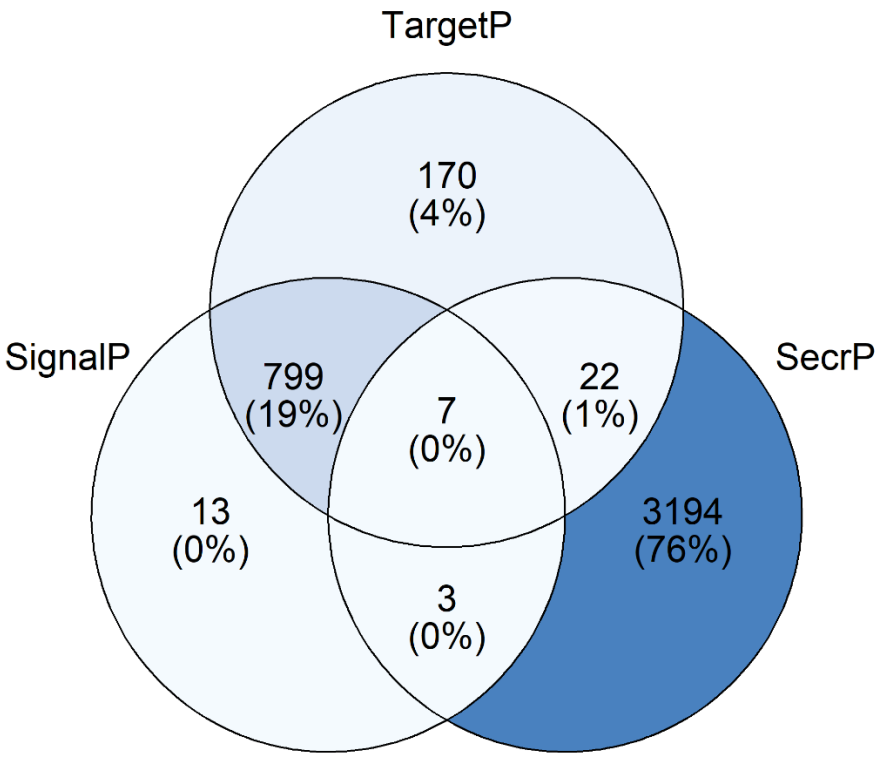
CJRS conceptualised the investigation carried out in this paper, developed the methodology, carried out the majority of formal analysis, visualised the analysis and was the major contributor to the writing of the manuscript. DRL contributed to the formal analysis, visualisation, and writing of the manuscript. NCB contributed to the supervision of this work and assisted writing the manuscript.

3.7.5. Acknowledgements

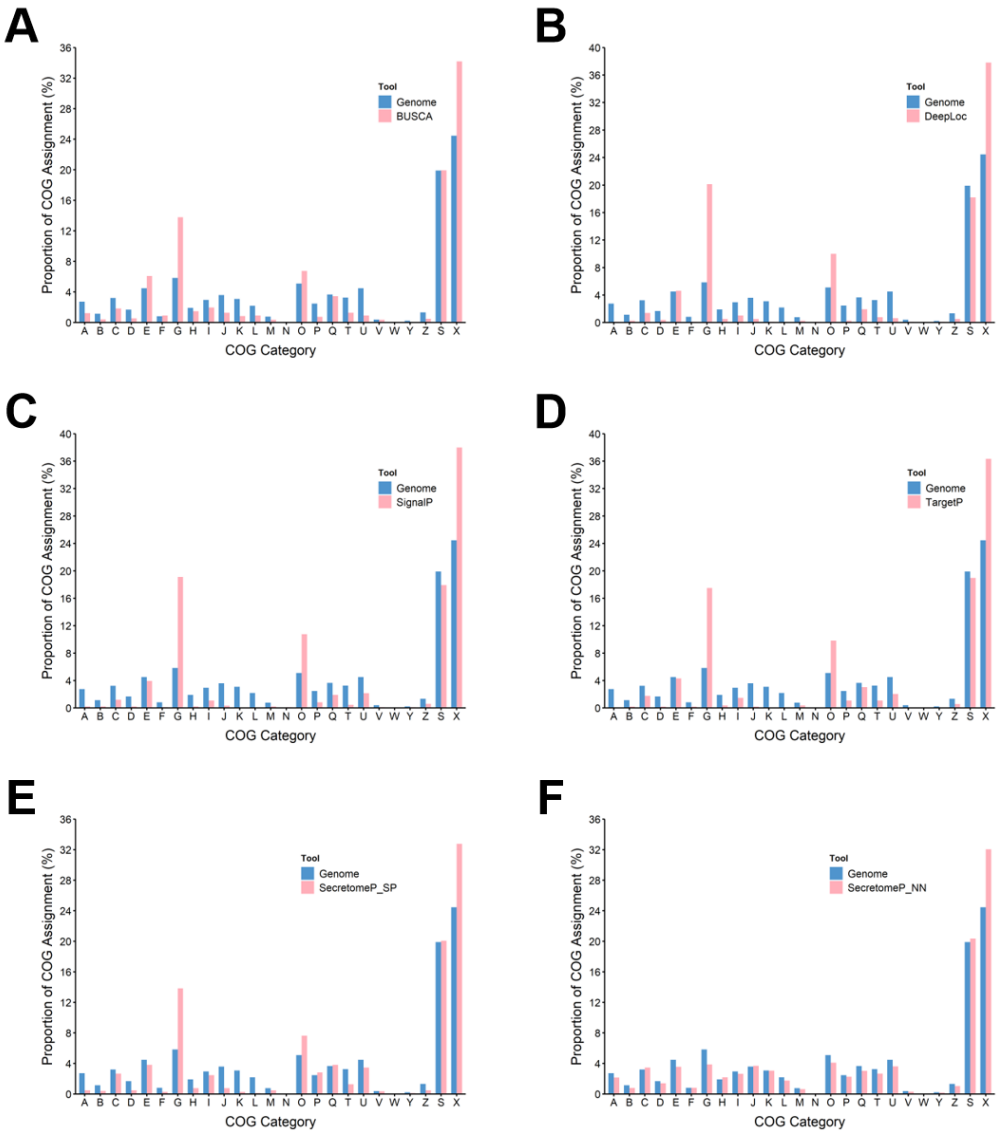
This project was undertaken on the Viking Cluster, which is a high-performance computing facility provided by the University of York. We are grateful for computational support from the University of York High Performance Computing service, Viking and the Research Computing team.

3.8. Supplementary material

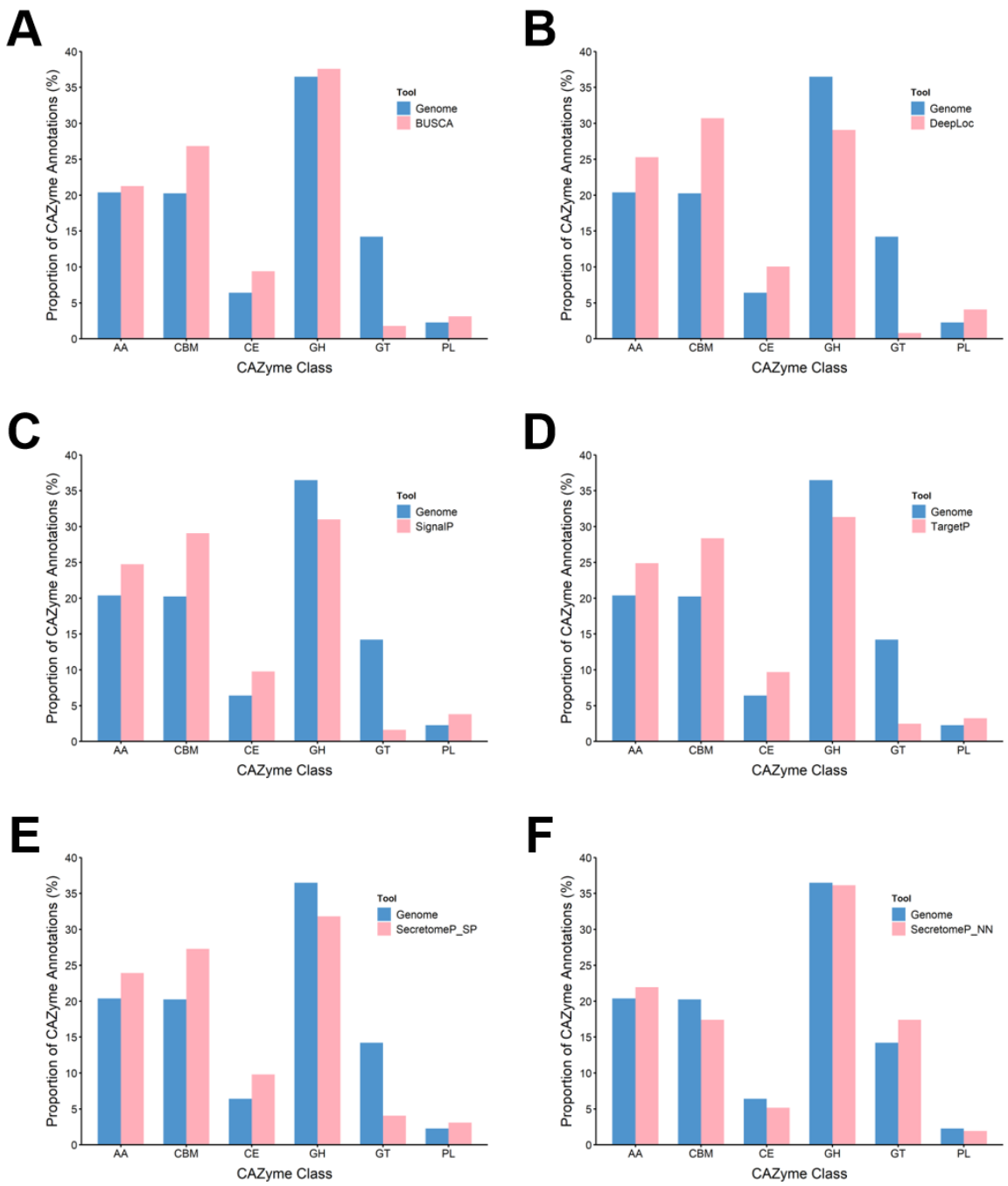
3.8.1. Supplementary figures



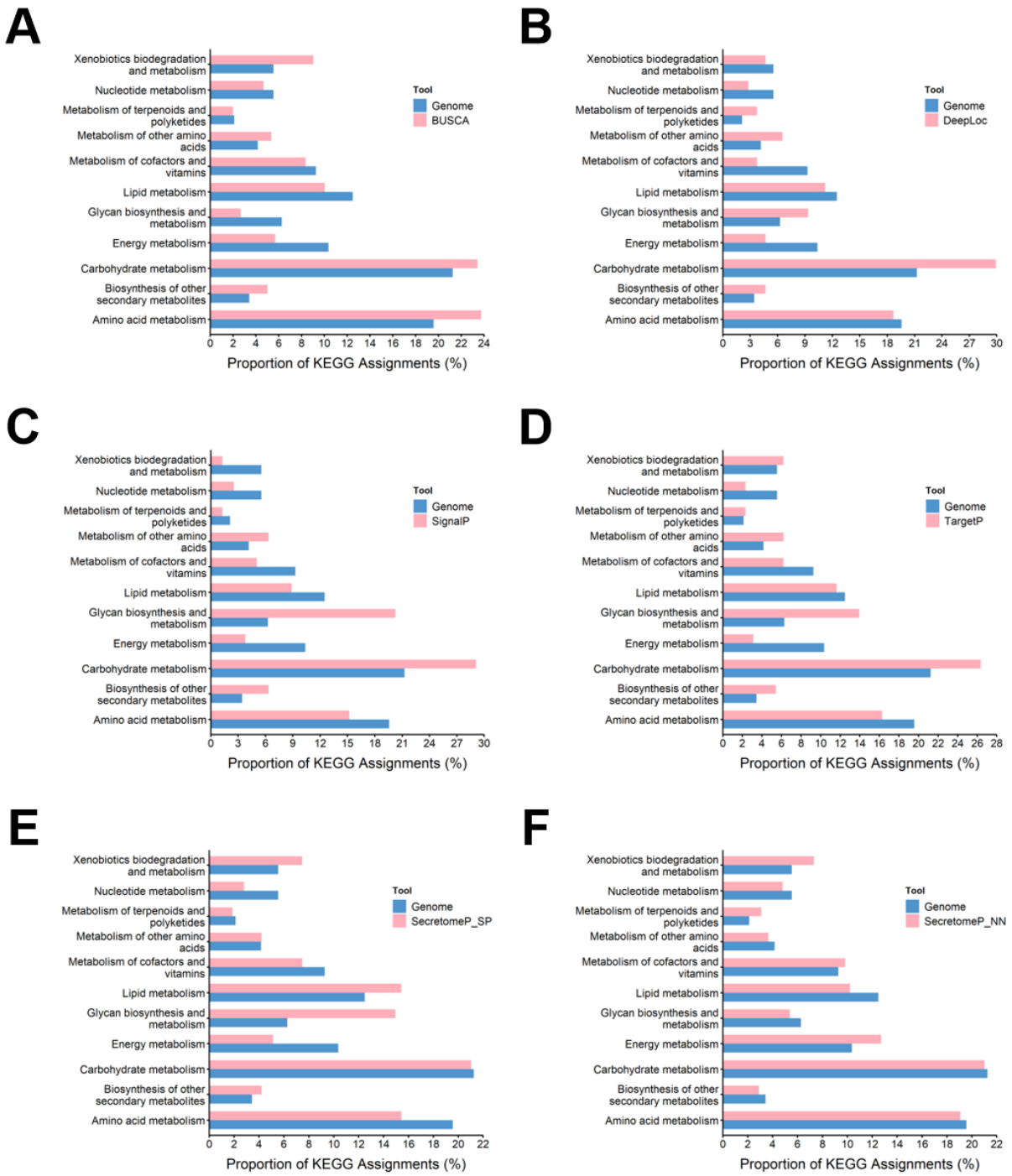
Supplementary Figure 3.1 Investigating SignalP and TargetP secretion signal and SecretomeP non-classical secretion prediction. The number of sequences predicted to encode secretion signal peptides by SignalP and TargetP, and the number of sequences predicted to encode non-classically secreted proteins by SecretomeP visualised.



Supplementary Figure 3.2 Investigating COG annotation proportions. The proportion of each COG Category in the whole genome COG annotation compared to the subsets of protein sequences predicted to encode extracellular proteins by BUSCA (**A**) and DeepLoc (**B**), predicted to encode classically secreted proteins by SignalP (**C**), TargetP (**D**), and SecretomeP (**E**), and predicted to encode non-classically secreted proteins by SecretomeP (**F**). A = RNA processing and modification, B = Chromatin structure, C = Energy production and conversion, D = Cell cycle control and mitosis, E = Amino acid metabolism and transport, F = Nucleotide metabolism and transport, G = Carbohydrate metabolism and transport, H = Coenzyme metabolism, I = Lipid metabolism, J = Translation, K = Transcription, L = Replication and repair, M = Cell wall/membrane/envelope biogenesis, N = Cell motility, O = Post-translational modification/turnover/chaperone functions, P = Inorganic ion transport and metabolism, Q = Secondary structure, T = Signal transduction, U = Intracellular trafficking and secretion, Y = Nuclear structure, Z = Cytoskeleton, S = Function Unknown, X = Unassigned.



Supplementary Figure 3.3 Investigating CAZyme annotation proportions. The proportion of each CAZyme Class in the whole genome CAZyme annotation compared to the subsets of protein sequences predicted to be extracellular by BUSCA (**A**) and DeepLoc (**B**), predicted to encode classically secreted proteins by SignalP (**C**), TargetP (**D**), and SecretomeP (**E**), and predicted to encode non-classically secreted proteins by SecretomeP (**F**). AA = Auxiliary activity, CBM = Carbohydrate binding module, CE = Carbohydrate esterase, GH = Glycoside hydrolase, GT = Glycosyl transferase, PL = Polysaccharide lyase.



Supplementary Figure 3.4 Investigating KEGG pathway annotations for BUSCA predicted extracellular protein coding sequences. The proportion of KEGG assignments to metabolic pathways in the whole genome compared to subsets of sequences predicted to encode extracellular proteins by BUSCA (**A**) and DeepLoc (**B**), predicted to encode classically secreted proteins by SignalP (**C**), TargetP (**D**), and SecretomeP (**E**), and predicted to encode non-classically secreted proteins by SecretomeP (**F**).

Chapter 4

4. *Parascedosporium putredinis* NO1 tailors its secretome for different lignocellulosic substrates

Authors

Conor JR Scott[a], Nicholas GS McGregor[c], Nicola C Oates[a], Janina Hoßbach[a], Amira Abood[a], Alexander Setchfield[a], Daniel Leadbeater[a], Adam Dowle[b], Herman S Overkleeft[d], Gideon J Davies[c], Neil C Bruce[a]

Affiliations

[a] Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, United Kingdom

[b] Bioscience Technology Facility, Department of Biology, University of York, York YO10 5DD, United Kingdom

[c] York Structural Biology Laboratory, Department of Chemistry, The University of York, York, YO10 5DD, United Kingdom

[d] Leiden Institute of Chemistry, Leiden University. P.O. Box 9502, 2300 RA Leiden, The Netherlands

4.1. Abstract

Parascedosporium putredinis NO1 is a wood degrading ascomycete with a propensity to target the most recalcitrant components of lignocellulose. Here we applied proteomics and activity-based protein profiling (ABPP) to investigate the ability of *P. putredinis* NO1 to tailor its secretome for growth on different lignocellulosic substrates.

Proteomic analysis of soluble and insoluble culture fractions following growth of *P. putredinis* NO1 on six lignocellulosic substrates, highlights the adaptability of the response of the *P. putredinis* NO1 secretome to different substrates. Differences in protein abundance profiles were maintained and observed across substrates after bioinformatic filtering of the data to remove intracellular protein contamination to more accurately identify the components of the secretome. These differences across substrates extended to carbohydrate active enzymes (CAZymes) at both class and family level. Investigation of abundant activities in the secretomes for each substrate revealed similar variation but also the high abundance of 'unknown' proteins in all conditions investigated. Fluorescence-based ABPP of secreted cellulases, xylanases and β-glucosidases confirmed highly adaptive time- and substrate-dependent glycoside hydrolase production by this fungus.

P. putredinis NO1 is a promising new candidate for the identification of enzymes suited to the degradation of recalcitrant lignocellulosic feedstocks. The investigation of proteomes from the biomass bound and culture supernatant fractions provides a more complete picture of a fungal lignocellulose-degrading response. In-depth understanding of this varied response will enhance efforts towards the development of tailored enzyme systems for use in biorefining.

4.2. Keywords

Parascedosporium putredinis NO1, Lignocellulose, Proteomics, CAZymes, Activity-based protein profiling

4.3. Background

Lignocellulosic residues from the waste components of established food crops (55, 263, 316) and from dedicated biomass crops (244) are a large and underutilised source of biomass for biorefinery-based production of chemicals and biofuels. The major constituents of lignocellulose are cellulose, hemicelluloses, lignin, and minor amounts of pectins and nitrogen compounds (265). However, this intricate and insoluble network of polysaccharides and aromatic polymers make lignocellulose a difficult substrate to degrade, convert, and to generate value from. But if deconstruction of this complex substrate can be understood, then it can offer a source of fuels, chemicals, and materials for a wide range of applications (80, 83, 90, 91).

Biological pre-treatments in the form of enzyme cocktails can deconstruct plant biomass as an environmentally friendly alternative to traditional physiochemical methods. Deployment of these enzymatic treatments has been limited by the hydrophobic and recalcitrant nature of lignocellulose. The heterogeneity of the substrate demands a diverse array of enzymes, and treatment can be limited by the release of inhibitor compounds during biomass breakdown (317). This process is further complicated by the differing structures and compositions between lignocellulose sources, each ideally requiring a tailored enzyme cocktail, something which current commercial cocktails do not address (243).

Much research has focused on natural wood degrading microorganisms for enzymatic solutions to these issues. (160, 161). Of particular significance, fungi are widely recognised as efficient degraders of plant biomass and major contributors to the carbon cycle of forest ecosystems. Across the fungal kingdom, a broad range of strategies are used in the deconstruction of lignocellulose. This includes the secretion of arrays of oxidative and hydrolytic enzymes to attack the polysaccharide components of plant cell walls, the production of high levels of lignin-degrading peroxidases and laccase enzymes to attack the lignin directly, and generation of strong oxidants, *via* non-enzymatic radical based reactions to degrade lignocellulose indirectly (165, 168). Ascomycete soft-rot fungi are interesting sources of lignocellulose degrading enzymes because they often rely on a strategy of penetrating plant secondary cell walls where they can secrete large amounts of enzymes directly at the site of attack (267).

Parascedosporium putredinis NO1 is a soft-rot ascomycete identified from a mixed microbial community grown on wheat straw (229). *P. putredinis* NO1 dominated the fungal population in the later stages of the community culture, suggesting an ability to metabolise the more recalcitrant carbon sources in the substrate. This makes *P. putredinis* NO1 an interesting candidate for the identification of new lignocellulose degrading enzymes. Indeed, a new

oxidase enzyme was recently identified in the *P. putredinis* NO1 secretome which cleaves the major β-ether units in lignin to enhance the digestibility of wheat straw releasing tricin, coumaric acid and vanillic acid in the process (229).

An annotated reference genome for this fungus now provides a resource for more effective investigation of the *P. putredinis* NO1 secretome (301). Proteomic analysis revealed an extensive and diverse array of carbohydrate active enzymes (CAZymes) produced by NO1 during growth. The ability of *P. putredinis* NO1 to grow on various industrially relevant lignocellulosic substrates or kraft lignin indicates that it is an exceptional lignocellulose, and particularly lignin, degrader.

To understand what genes are expressed during complex biomass degradation and to what degree these vary depending on lignocellulosic substrate, the secretome during the growth of *P. putredinis* NO1 on six diverse lignocellulosic substrates was investigated using both proteomic and activity-based protein profiling (ABPP) techniques. Proteins were harvested from both the supernatant and substrate-bound fractions of liquid cultures to give a more complete picture of the lignocellulose degrading proteome. The variability of this lignocellulose-degrading protein complement across substrates evidences a highly dynamic and tailored response to substrates by *P. putredinis* NO1. The enzymes identified here represent the vanguard of biomass-degrading capacity produced by a fungus that is highly effective at degrading lignin enriched biomass. Understanding this complex enzyme-system will aid the design of feedstock-matched enzyme cocktails to enhance biorefining efficiency.

4.4. Results and discussion

4.4.1. Analysis of the *P. putredinis* NO1 secretome on different lignocellulosic substrates

To investigate whether the secretome of *P. putredinis* NO1 varies with lignocellulosic substrates, the fungus was grown on six substrates: oil palm empty fruit bunch, kraft lignin, rice straw, sugar cane bagasse, wheat bran, and wheat straw, and the resulting proteomes were sampled and analysed. Proteins were harvested from culture supernatants and via a surface protein labelling approach with biotin that harvests proteins bound to substrates (233). It should be noted that kraft lignin is more soluble than the other substrates and therefore a larger portion of the kraft lignin bound fraction likely reflects proteins associated to the fungal cell wall. After harvesting and identifying proteins through label-free LC-MS, molar percentage values were generated for each protein. Quality control was performed by investigating protein count, performing principal component analysis, and clustering for all replicates. Through this analysis a bound fraction sugar cane bagasse replicate, a supernatant fraction sugar cane bagasse replicate, a supernatant kraft lignin replicate, and a supernatant fraction wheat bran replicate were identified as outliers and removed from further analysis.

Across all samples 2014 proteins were identified in at least one replicate, 1890 proteins were identified among the substrate-bound fractions and 973 were identified among the supernatant fractions, with 849 proteins shared across both fractions (**Supplementary Figure 4.1**). Many intracellular proteins which were likely the results of cell death and lysis were present in the proteomic dataset despite efforts to target the extracellular fractions (309). The proteomic data was therefore filtered to remove the contaminating intracellular proteins according to the bioinformatics workflow presented previously (318). Secretome proteins were predicted to be extracellular using both BUSCA and DeepLoc localisation prediction tools (309, 310), or were predicted to encode a secretion signal by SignalP, TargetP and SecretomeP tools (275, 311, 312). Proteins which contained more than one predicted transmembrane domain by TMHMM or contained a single predicted transmembrane helix with more than 10 amino acids of this helix occurring in the first 60 amino acids of the protein sequence were then removed (313). The final secretome contained 228 proteins, a significant reduction from the total proteome dataset. In the substrate bound fraction of the secretome 165 proteins were identified in at least a single replicate, in the culture supernatant fraction 185 proteins were identified, and 122 proteins were shared between both fractions (**Figure 4.1**). The now much smaller number of proteins identified exclusively in the bound and supernatant fractions were 43 and 63 proteins, respectively. The difference between the substrate bound and culture supernatant fractions

highlights the importance of capturing both fractions, in order to fully understand the secretome response.

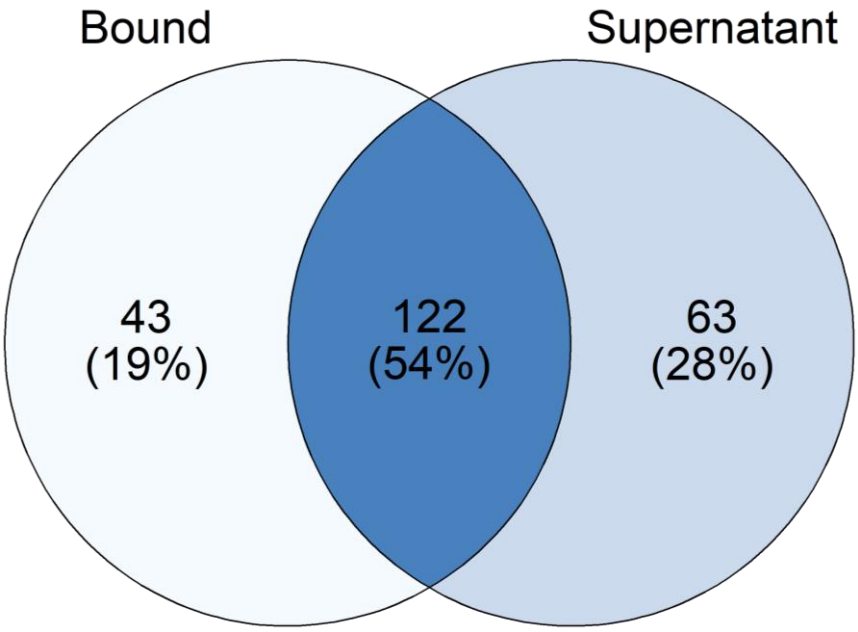


Figure 4.1 Investigating the distribution of proteins across bound and supernatant fractions in the *P. putredinis* NO1 secretome. The number of proteins identified in at least one replicate across all substrates for the bound fraction compared to the supernatant fraction within the predicted *P. putredinis* NO1 secretome from growth on six lignocellulosic substrates for four days.

The overall protein abundance profiles of the *P. putredinis* NO1 secretome on each substrate investigated were visualised by scaling MolPct abundance for each protein across substrates and clustering (**Figure 4.2**). The substrate dependent variation in the *P. putredinis* NO1 secretome was clearly visible, with all substrates displaying substrate exclusive protein profiles to varying extents in both fractions. The differences between substrates were also investigated through principal component analysis, and clustering (**Supplementary Figure 4.2** and **Supplementary Figure 4.3**). In the substrate bound fraction, replicates grouped very well within substrates in PCA plots and clustered well within substrates in dendrograms, while demonstrating good separation between substrates in PCA plots and dendrograms also. In the supernatant fraction the differences between substrates were less clear, with single replicates from some substrates grouping away from other replicates of the same substrate; however, this is likely due to the lower identified protein count for these replicates in the supernatant fraction. The heatmaps in **Figure 4.2** still demonstrate that clear differences are visible across substrates in both fractions.

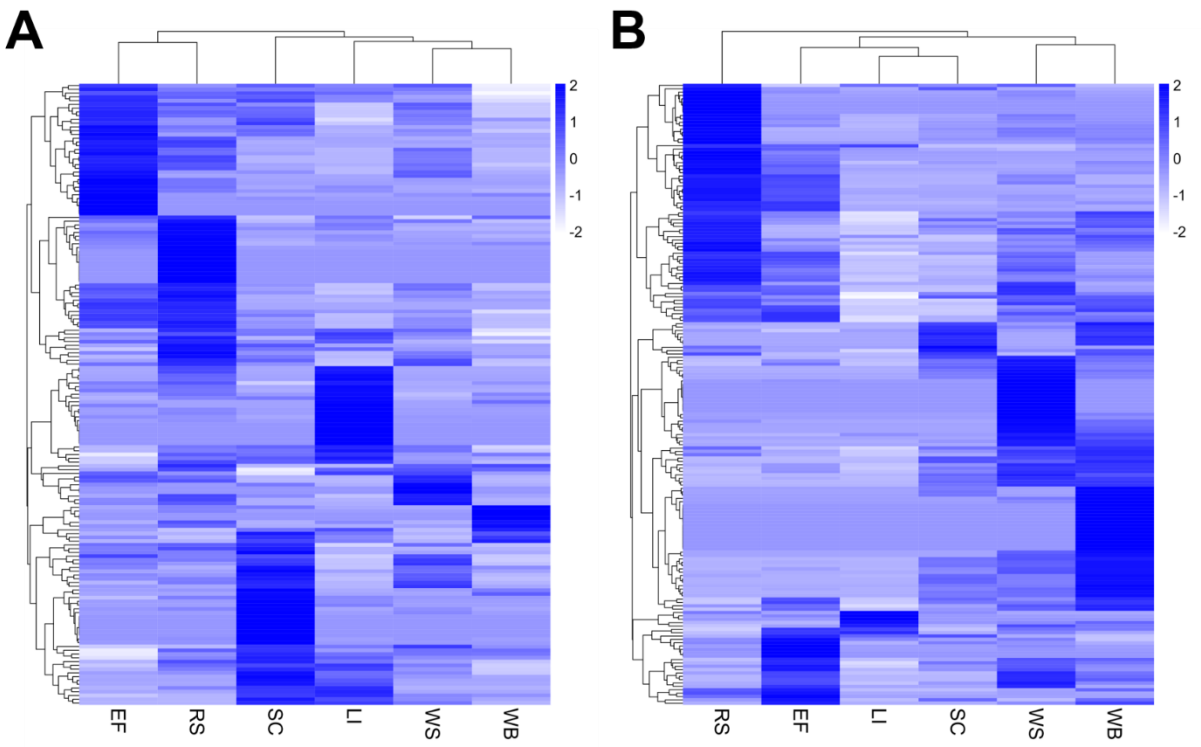


Figure 4.2 Differences in the *P. putredinis* NO1 secretome across lignocellulosic substrates. Molar percentage values for proteins identified on at least one substrate in the bound (**A**) and supernatant (**B**) fractions of the *P. putredinis* NO1 secretome scaled to Z-scores across substrates. EF = Empty fruit bunch, LI = Kraft lignin, RS = Rice straw, SC = Sugar cane bagasse, WB = Wheat bran, WS = Wheat straw.

4.4.2. An investigation of the functional profile of the *P. putredinis* NO1 secretome on different lignocellulosic substrates

To investigate the functional profiles of the *P. putredinis* NO1 secretome across lignocellulosic substrates, proteins were annotated with CAZyme domains and GO categories and terms. The overall functional profiles were investigated by comparing the proportional abundance of proteins annotated with CAZyme domains, if not annotated as CAZymes then assigned to GO categories, or not annotated with either (**Figure 4.3**). Clear variation in the proportions of functional category assignments were observed across substrates in both bound and supernatant fractions. In the bound fraction (**Figure 4.3A** and **B**), CAZymes contribute the most to total abundance for all substrates. This contribution is slightly reduced for the kraft lignin substrate, which is to be expected due to the low abundance of polysaccharides on which many CAZymes act. In the supernatant fraction (**Figure 4.3C** and **D**), CAZymes again contribute the most to total abundance except for rice straw. Surprisingly, CAZymes contributed the most to total abundance in the substrate bound fraction for kraft lignin, which perhaps reflects the higher solubility of kraft lignin and the relatively lower amount of insoluble material available for enzyme binding compared to

the other cultures. Despite bioinformatic filtering of proteomic data to produce the *P. putredinis* NO1 secretome some intracellular and membrane proteins may have persisted, as evidenced by the contribution of “cellular component” GO category proteins to abundance for all substrates in both bound and supernatant fractions. In both the bound and supernatant fractions, “unknown” proteins contribute similar proportions of abundance as the functional GO categories and again the proportion of these unknown proteins vary across substrates in both fractions investigated. The persistence of these proteins in the *P. putredinis* NO1 secretome is exciting for the prospect of identifying new lignocellulose degrading enzymes.

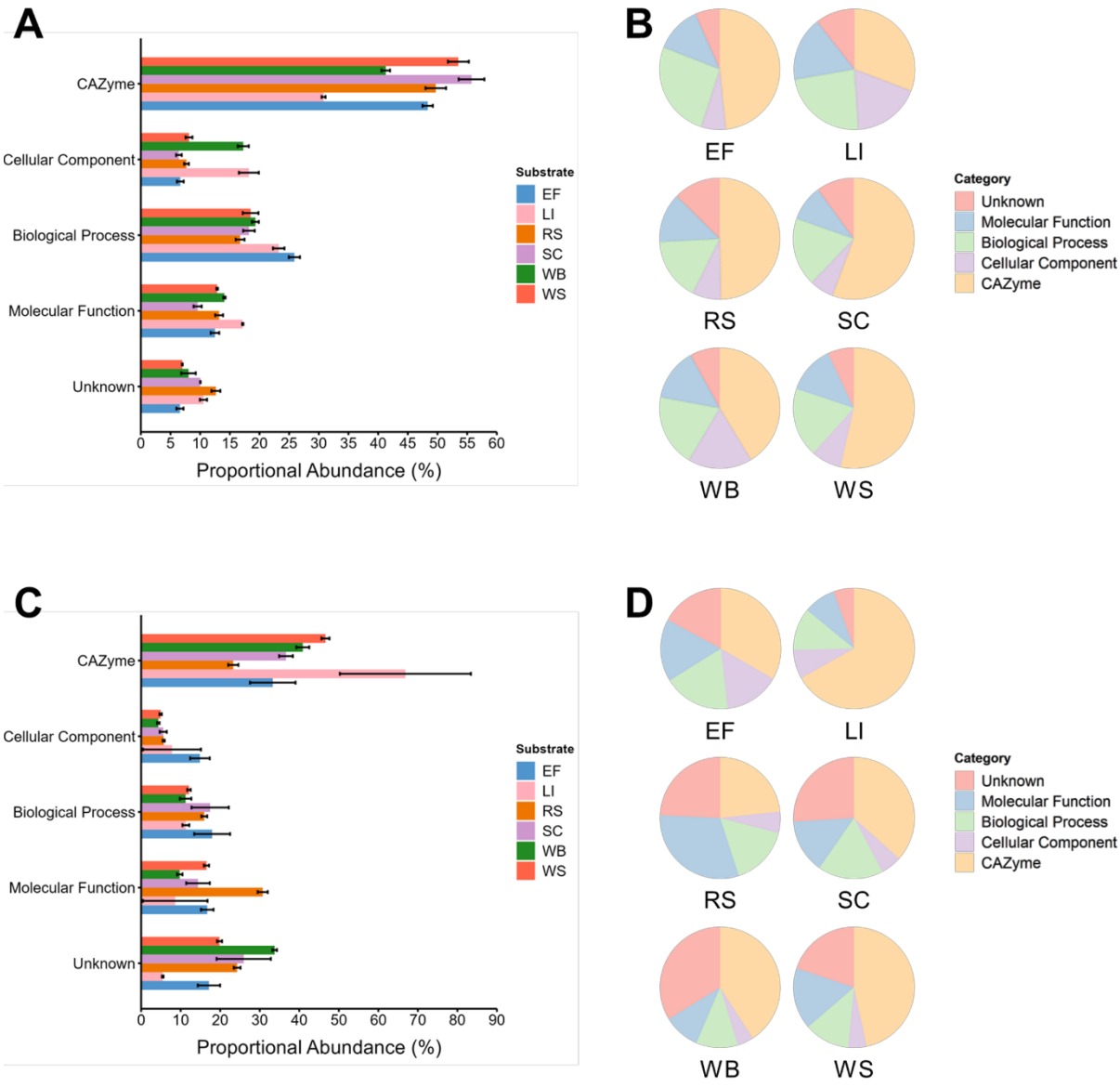


Figure 4.3 Differences in functional categories of the *P. putredinis* NO1 secretome across lignocellulosic substrates. The molar percentage abundance of proteins identified as CAZymes, that were not identified as CAZymes but assigned to GO categories, or that were not annotated as CAZymes or with GO categories were calculated proportionally and compared across substrates for

each category for the bound (**A**) and supernatant fractions (**C**) of the *P. putredinis* NO1 secretome. The proportions of each functional category for each substrate were also visualised for the bound (**B**) and supernatant (**D**) fractions separately. EF = Empty fruit bunch, LI = Kraft lignin, RS = Rice straw, SC = Sugar cane bagasse, WB = Wheat bran, WS = Wheat straw.

To understand the functional profiles of the *P. putredinis* NO1 secretome in more detail, the proportional abundance of the ten most abundant GO terms were plotted for each substrate for the bound and supernatant fractions (**Figure 4.4** and **Figure 4.5**). For all substrates in both the bound (**Figure 4.4**) and supernatant (**Figure 4.5**) fractions, the top GO term annotations in terms of contribution to total abundance were predominantly extracellular activities associated with hydrolytic breakdown of lignocellulose. In both fractions, many substrates also included membrane annotations in the top ten most abundant GO terms, possibly the result of misinterpretation by SignalP, TargetP and SecretomeP of membrane insertion signals as secretion signals or from the binding and extraction of extracellular cell surface proteins by the biotin labelling technique. In the bound fraction, membrane GO annotations contributed the highest abundance of any GO term for kraft lignin (**Figure 4.4B**) and wheat bran (**Figure 4.4E**), which may reflect the reduced CAZyme proportions seen previously for these substrates in this fraction. Aside from cell surface proteins, abundant intracellular annotations were largely absent from the secretomes. Unknown proteins i.e., with no GO annotations, were also present in the top ten most abundant annotations for all substrates in both fractions except for the supernatant fraction of kraft lignin. In the supernatant fraction, unknown proteins made the largest contribution of any annotation for empty fruit bunch (**Figure 4.5A**), rice straw (**Figure 4.5C**), sugar cane bagasse (**Figure 4.5D**), wheat bran (**Figure 4.5E**), and wheat straw (**Figure 4.5F**). The most abundant annotations for most substrates in both fractions appear to be associated with the breakdown of cellulose and xylan, two abundant polysaccharides in these lignocellulosic substrates. However, there is a clear difference in profile for the kraft lignin substrate in both the bound (**Figure 4.4B**) and supernatant (**Figure 4.5B**) fractions. In the bound fraction, there appears to be more membrane targeted GO annotations in the ten most abundant annotations. Finally, the profile of the most abundant GO annotations looks unique for the bound fraction of wheat bran (**Figure 4.4E**). Wheat bran has consistently appeared different in abundance and functional profiles to other substrates and may reflect its difference in composition from the other lignocellulosic substrates such as its high starch, hemicellulose, and protein content, and reduced lignin content (319).

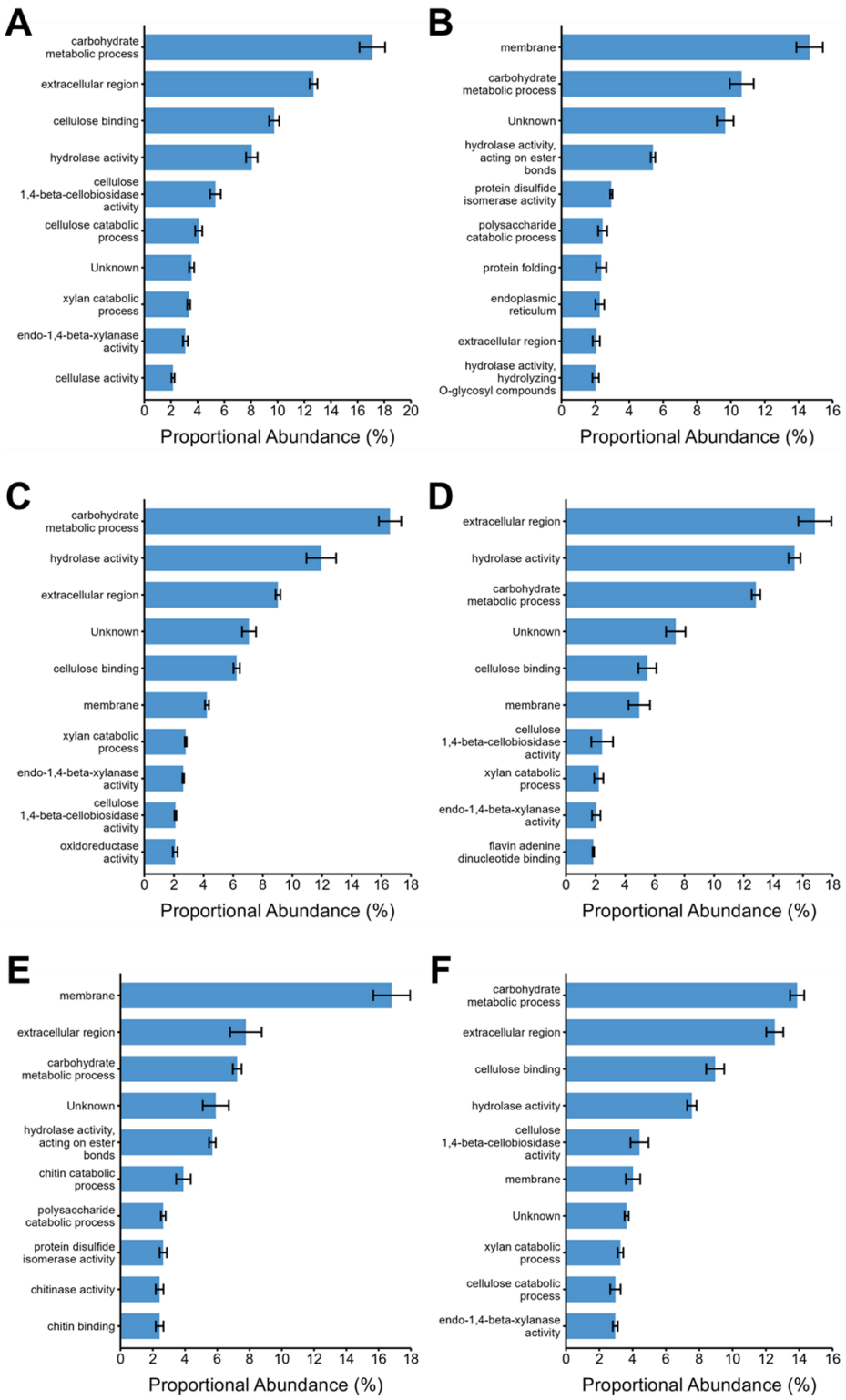


Figure 4.4 Proportional abundances of enzyme activities in the bound fraction of the *P. putredinis* NO1 secretome. The molar percentage abundance of proteins assigned to GO terms were calculated proportionally for the bound fraction of the *P. putredinis* NO1 secretome after four days of growth on six lignocellulosic substrates. EF = Empty fruit bunch (**A**), LI = Kraft lignin (**B**), RS = Rice straw (**C**), SC = Sugar cane bagasse (**D**), WB = Wheat bran ©, WS = Wheat straw (**F**).

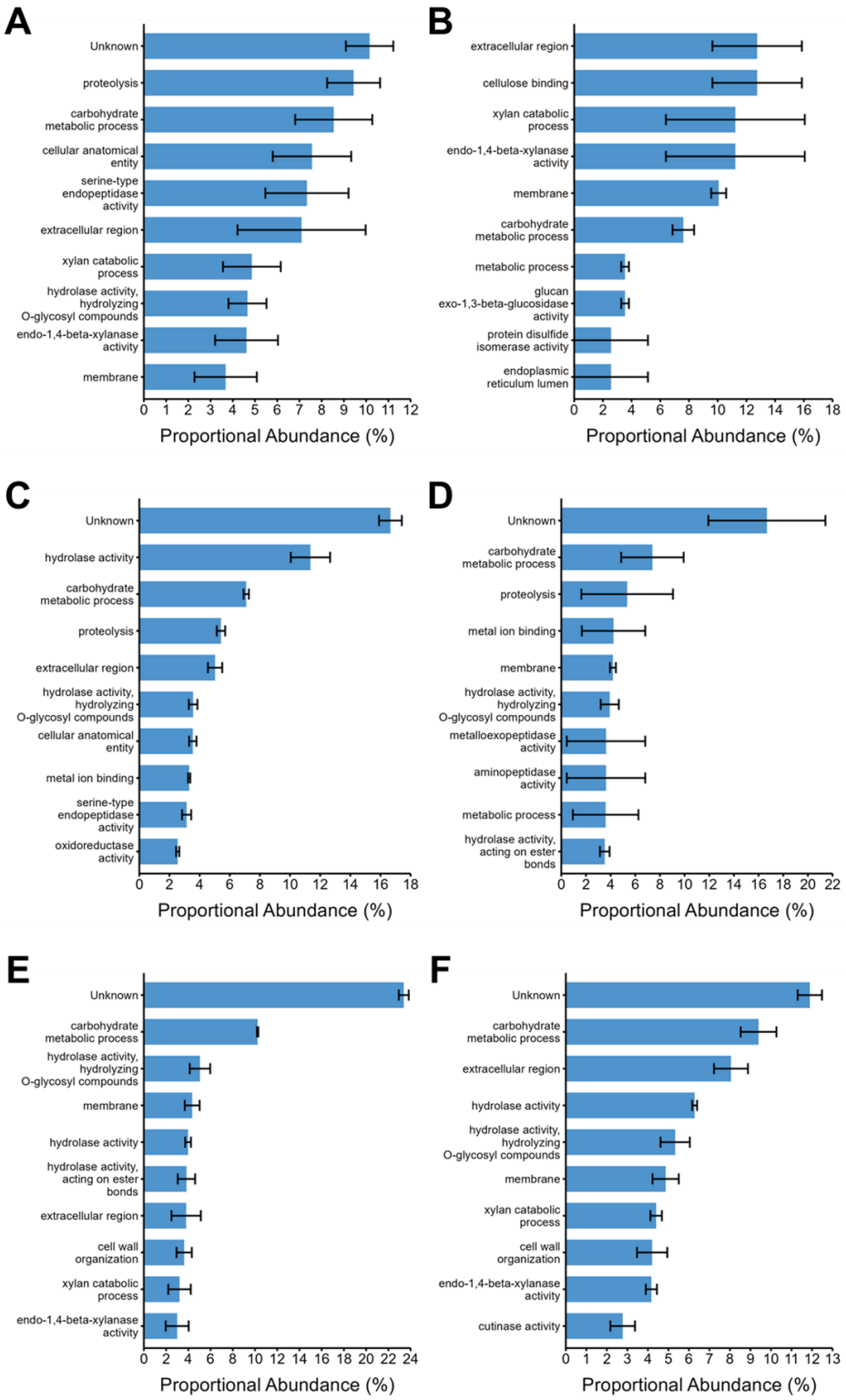


Figure 4.5 Proportional abundances of enzyme activities in the supernatant fraction of the *P. putredinis* NO1 secretome. The molar percentage abundance of proteins assigned to GO terms were calculated proportionally for the supernatant fraction of the *P. putredinis* NO1 secretome after four days of growth on six lignocellulosic substrates. EF = Empty fruit bunch (**A**), LI = Kraft lignin (**B**), RS = Rice straw (**C**), SC = Sugar cane bagasse (**D**), WB = Wheat br©(**E**), WS = Wheat straw (**F**).

4.4.3. The lignocellulose degrading secretome of *P. putredinis* NO1 varies depending on the growth substrate

To investigate the lignocellulose-degrading enzyme repertoire of the *P. putredinis* NO1 secretome, protein sequences were annotated for CAZyme domains using the dbCAN server (178). CAZyme proteins are assigned to the following catalytic classes: auxiliary activity (AA), carbohydrate esterase (CE), glycoside hydrolase (GH), glycosyl transferase (GT), and polysaccharide lyase (PL). Additionally, many CAZymes have carbohydrate binding modules (CBMs) which are also assigned.

Comparing the contribution of proteins assigned to these catalytic classes to total CAZyme abundance across substrates demonstrates similar variation seen for the total secretome and for functional categories in both bound and supernatant fractions (**Figure 4.6**). In the bound fraction AA class CAZymes contribute a high proportion of total CAZyme abundance for the empty fruit bunch, rice straw, sugar cane bagasse, and wheat straw substrate but are low in abundance for wheat bran and kraft lignin substrates (**Figure 4.6A** and **Figure 4.6B**). Crystalline cellulose is present in lower levels or is absent from wheat bran and kraft lignin substrates, and this may reflect a reduced production of AA class CAZymes involved in lytic polysaccharide monooxygenase (LPMO) systems which degrade crystalline cellulose oxidatively (179, 182, 319, 320). Potentially, the oxidative enzymes responsible for lignin breakdown are still present but in low abundances. The reductions in AA abundances for wheat bran and kraft lignin substrates corresponds with increased proportions of all other CAZyme classes in the bound fraction, including GT class CAZymes (224). Some GT enzymes are extracellular and bound to the fungal cell surface which is likely why these proteins have been isolated in the *P. putredinis* NO1 secretome and they are more likely to play a role in fungal growth, cell wall remodelling, or potentially associating the fungus with the growth substrate (225, 226). PL class enzymes were absent from both fractions of the *P. putredinis* NO1 secretome during growth on kraft lignin. They were either absent or present in low abundances for all substrates in the bound fraction but were present with varying contributions to total CAZyme abundance in the supernatant fraction. This could be explained by the soluble pectin substrates on which PL enzymes act, which are more likely to be present in the culture supernatant (321). In the supernatant fraction, GH class CAZymes dominated the CAZyme abundance profiles for all substrates (**Figure 4.6C** and **Figure 4.6D**). It can be hypothesised that soluble components of the substrates are present initially and are also released into the supernatant fraction during degradation. These soluble components would then be accessible to the wide array of hydrolytic GH class CAZyme activities which act on diverse substrates, and this potentially explains the abundance of these enzymes in this fraction (221).

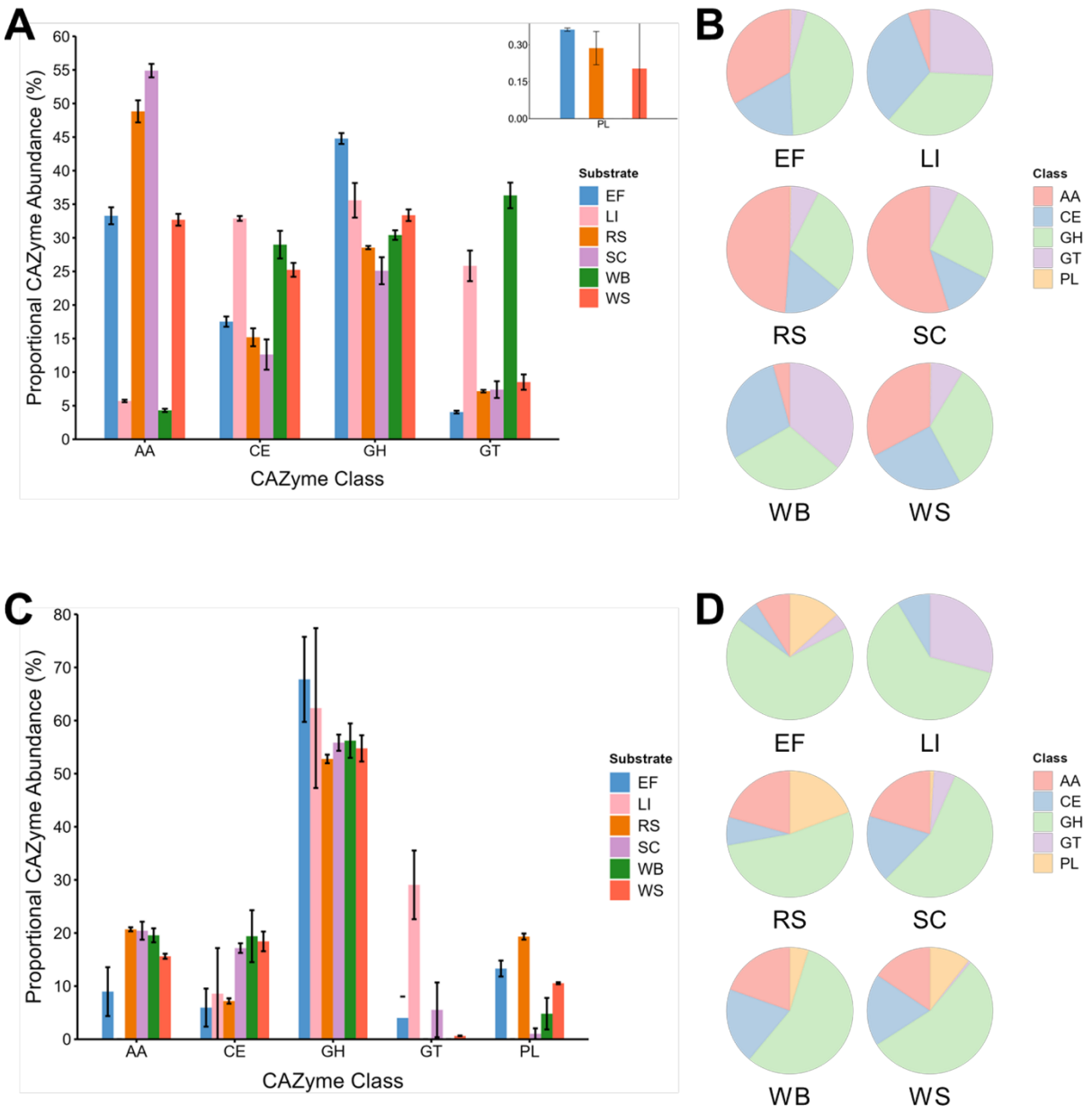


Figure 4.6 Differences in proportional catalytic CAZyme class abundance of the *P. putredinis* NO1 secretome across lignocellulosic substrates. The molar percentage abundance of proteins belonging to each catalytic class of CAZyme were calculated proportionally to the total abundance of CAZymes for each substrate and compared across substrates for the bound (**A**) and supernatant fractions (**C**) of the *P. putredinis* NO1 secretome. The proportions of each CAZyme class for each substrate were also visualised for the bound (**B**) and supernatant (**D**) fractions separately. EF = Empty fruit bunch, LI = Kraft lignin, RS = Rice straw, SC = Sugar cane bagasse, WB = Wheat bran, WS = Wheat straw.

To investigate the CAZyme classes in more detail and to address some of the questions raised by comparing class profiles, the CAZyme family abundances were compared across substrates for the bound and supernatant fraction separately (**Figure 4.7**). Variation across substrates was again clear at the CAZyme family level in both fractions even for substrates which looked similar at the class level. This highlights how similarities at the class level masked these underlying differences at the level of the CAZyme family, which defines enzyme activities more specifically. As expected, the relative abundances of AA9 LPMOs and other AA families e.g., AA3, AA7, and AA12 which act in LPMO systems are reduced compared for kraft lignin and wheat bran substrates in the bound fraction (185, 186, 269). However, the AA1 laccase family enzymes responsible for the oxidative cleavage of lignin structures are present in increased abundance in the bound fraction for the kraft lignin substrate (269). All AA families were muted in abundance in the bound fraction for wheat bran except for the AA13 family of starch degrading LPMOs, likely resulting from the increased starch content of the substrate (184, 319). Variation in the abundance profiles of proteins with CBMs belonging to different families can also be seen in (**Figure 4.7**). CBMs are non-catalytic domains which predominantly bind to plant cell wall polymers to prolong contact of catalytic CAZyme domains with their substrates to enhance efficiency (205). A wide range of CBM families provide binding capabilities for different enzymes to a range of cell wall components, and the variation in CBM containing protein abundance observed here may reflect the varying compositions and structures of the substrates. Indeed some expected patterns can be seen in the bound fraction, such as the relatively lower abundance of CBM1 proteins, involved in crystalline cellulose binding, to kraft lignin and wheat bran substrates which have reduced proportions of cellulose in comparison to the other substrates investigated (260, 319, 320). For wheat bran, in the bound fraction there is a relatively high abundance of CBM43, CBM5, and CBM56 containing proteins. CBM5 proteins are involved in binding chitin and likely to be involved in cell-wall remodelling here (209). CBM43 and CBM56 are involved in binding β-1,3-glucans, which are present in fungal cell walls, but also present in the wheat bran substrate and are potentially being targeted for breakdown here (210, 211, 319). The hydrolytic families of CAZymes belonging to GH, CE, and PL classes show variation in abundance profiles across all substrates in both fractions and similarly likely reflect a tailored hydrolytic response to the different polysaccharide compositions of each substrate. Generally, the abundances of these hydrolytic enzymes are lower for the kraft lignin substrate in both fractions, agreeing with its lack of polysaccharides. In the supernatant fraction the relatively high abundance of some polysaccharide hydrolysing or binding CAZyme family domains during growth on kraft lignin may reflect enzymes produced irrespective of growth substrate or in response to lignin breakdown products. Indeed, it has been observed in *T. reesei* that spore germination alone can lead to massive

upregulation of polysaccharide degrading CAZymes to prepare the fungus for a habitat containing plant biomass (322). Perhaps the GH10, GH26, GH55, and CE15 families produced in relatively high abundance for kraft lignin represent a similar response. Investigating CAZyme family profiles in the supernatant fraction also demonstrates reduced relative abundances for all CAZyme domain containing proteins for the sugar cane bagasse substrate, and variation across wheat straw, rice straw, and empty fruit bunches from palm oil, despite all these substrates having similar overall biomass compositions (54, 86). Therefore, this variation represents the complexity of the fungal plant biomass-degrading response to different substrates which is still poorly understood.

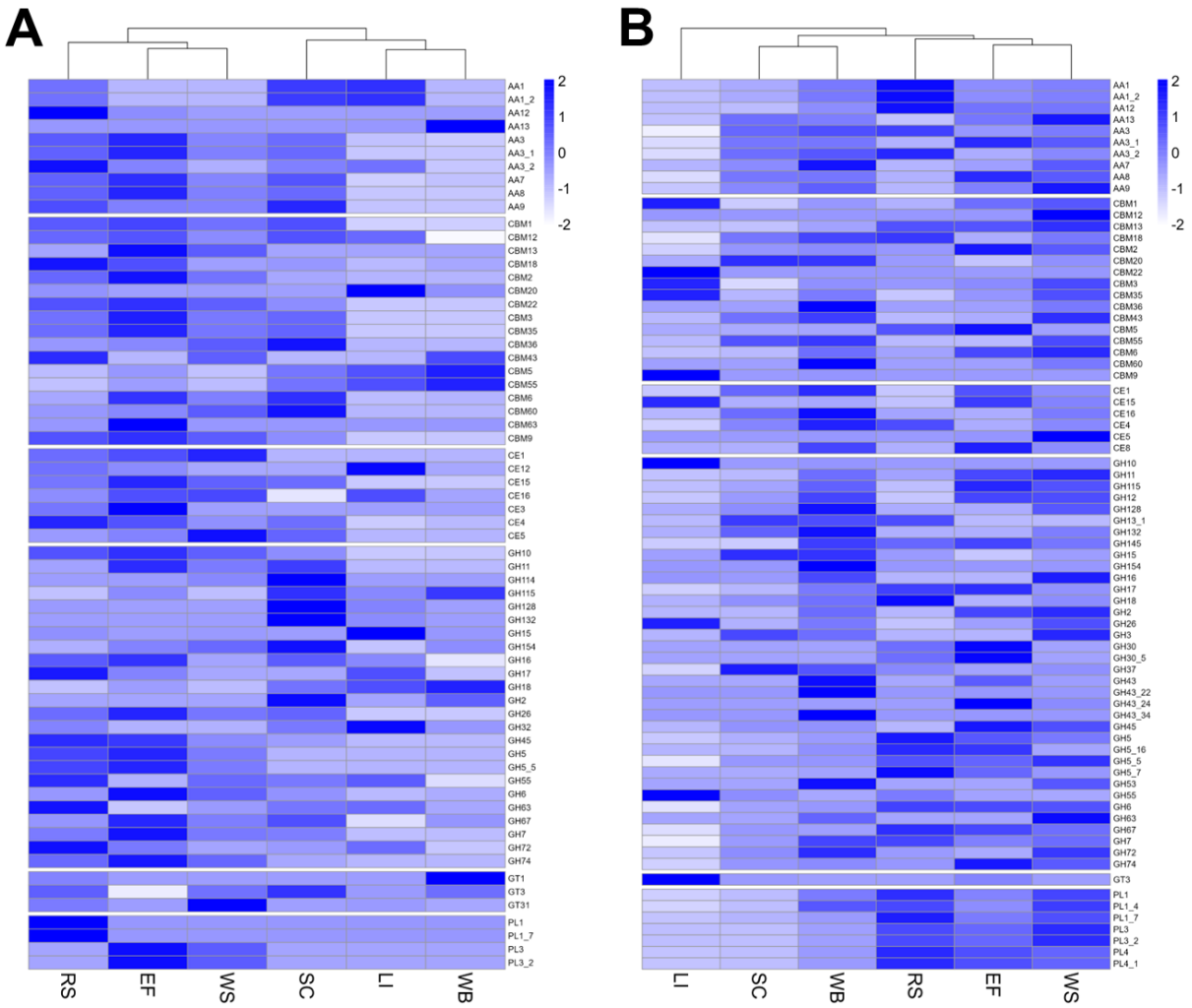


Figure 4.7 Differences in CAZyme family abundance of the *P. putredinis* NO1 secretome across lignocellulosic substrates. Molar percentage values for proteins annotated as CAZymes and identified on at least one substrate of the *P. putredinis* NO1 secretome scaled to Z-scores across substrates for the bound (**A**) and supernatant (**B**) fractions separately. EF = Empty fruit bunch, LI = Kraft lignin, RS = Rice straw, SC = Sugar cane bagasse, WB = Wheat bran, WS = Wheat straw.

These differences all suggest a potential ability of *P. putredinis* NO1 to tailor its secretome dependent on the growth substrate. Empty fruit bunch, rice straw, sugar cane bagasse, and wheat straw all show clear difference between the substrates as well, although the overall compositions are more similar. It was possible that the observed differences in CAZyme families were a product of different fungal growth stage on each substrate, as all data was harvested at a single timepoint. An alternative approach was therefore taken to confirm that differences in the *P. putredinis* NO1 secretome are maintained over time.

4.4.4. Activity-based protein profiling shows how lignocellulose-degrading enzyme profiles vary over time

Fluorescence-based activity-based protein profiling (ABPP) with cyclophellitol-derived activity-based probes (ABPs) for retaining β-glucosidases (323), cellulases (324), and xylanases (325) have been used to screen fungal secretomes previously (325, 326). Cellulases, xylanases, and β-glucosidases are broadly distributed and highly expressed in plant-biomass degrading fungi, making them good candidates for ABPP techniques and for comparison across lignocellulolytic systems (326). These probes were also used to display induction of an array of GH enzymes in basidiomycete secretomes when grown on lignocellulosic compared to simple substrates like maltose (326). Therefore, this is an appropriate and effective technique to explore the potentially tailored secretome of the ascomycete *P. putredinis* NO1 during growth on multiple industrially relevant lignocellulosic substrates.

Activity-based probes for glycoside hydrolases (GHs) were employed to screen samples of supernatants harvested from cultures of *P. putredinis* NO1 grown on the same substrates as the proteomics experiment at days 3, 4, 5, 6, 7, and 10 (327). The supernatants were used to screen for, and to determine the relative levels of lignocellulose-degrading glycoside hydrolases. The activities were present in the empty fruit bunch from palm oil, rice straw, sugar cane bagasse, wheat bran, and wheat straw samples (**Figure 4.8**). Unfortunately, the high aromatic content of the kraft lignin meant that this substrate was incompatible with this technique (**Supplementary Figure 4.4**). Fluorescent gel visualisation was performed for triplicate samples from all substrates and good agreement in hydrolase abundance pattern over time was observed within substrates (**Supplementary Figure 4.4**). For simplicity, single replicate time course fluorescence gels are presented in **Figure 4.8**, and integrated band intensity values resolved in **Figure 4.9**.

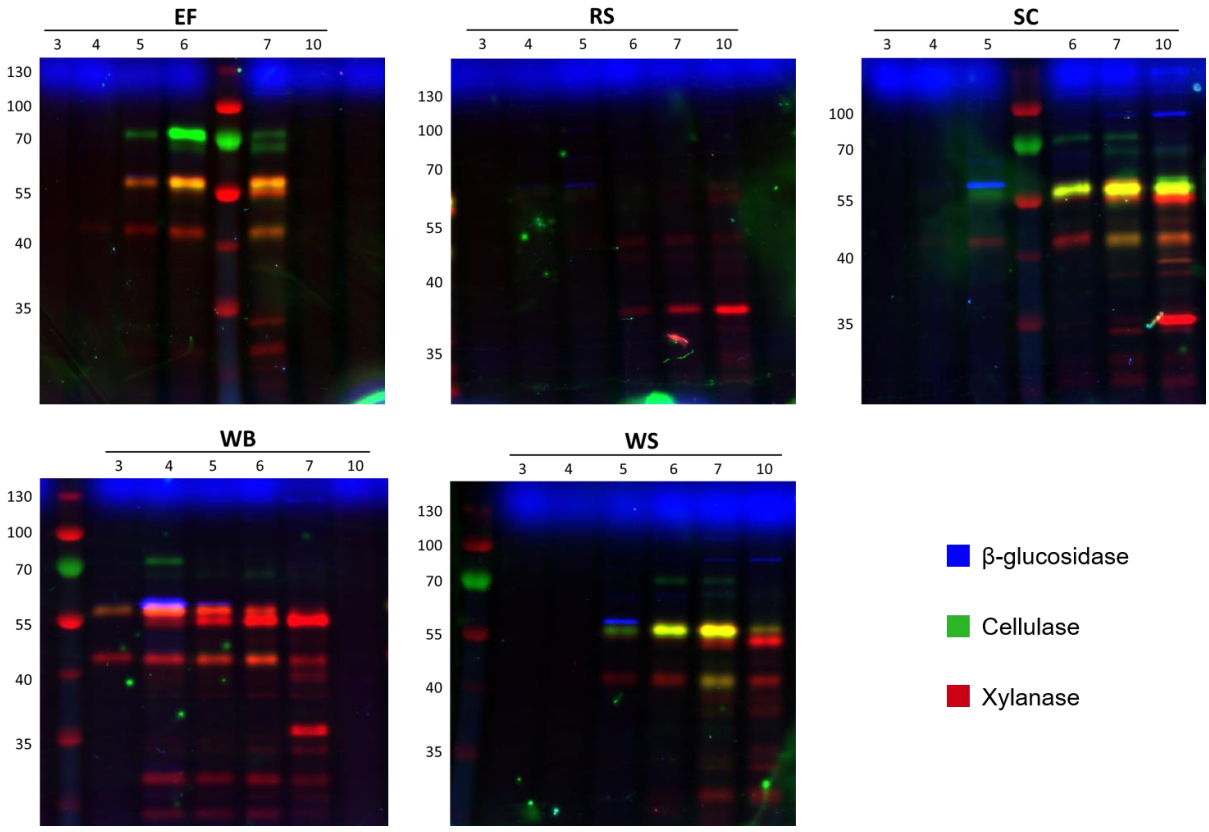


Figure 4.8 Differences in *P. putredinis* NO1 glycoside hydrolase production over time visualised with activity-based probes. Fluorescence imaging following SDS-PAGE is shown for single replicates of samples of culture supernatants taken at days 3, 4, 5, 6, 7, and 10 of growth treated with a triplex probe mixture targeting cellulases, xylanases, and retaining β-glucosidases. EF = Empty fruit bunch, RS = Rice straw, SC = Sugar cane bagasse, WB = Wheat bran, WS = Wheat straw. Secretomes were stained in triplicate, a single replicate is shown here.

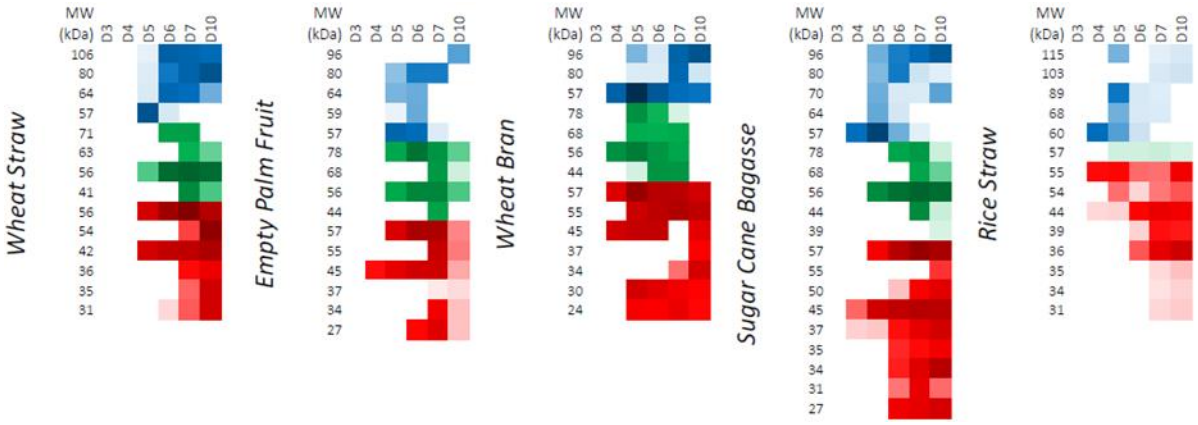


Figure 4.9 ABPP-determined variation in relative active enzyme levels over time during *P. putredinis* NO1 growth on various substrates. Resolved bands running at different apparent MW values (left column in each block) were integrated in the Cy2 (β-glucosidase probe), Cy3 (cellulase probe), or Cy5 (xylanase probe) channels. Average band integration values (n=3) are shown as colour intensity varying from white (not detected) to full colour (~1,000,000 counts) to black (saturation) on a logarithmic scale. Secretomes were prepared and stained in biological triplicate following different culture times (labels above columns).

As seen previously for screening of basidiomycete secretomes, the *P. putredinis* NO1 secretome shows differential GH production over time dependent on growth substrate (**Figure 4.8** and **Figure 4.9**) (326). *P. putredinis* NO1 demonstrated muted production of GHs at day 3 for all substrates. For rice straw, this production remained low for the entire time course with low levels of xylanase and β-glucosidase and only a single cellulase detected at low abundances. Although the production profiles of these GHs was demonstrated to vary by species, it is worth noting that transcriptomic analysis of another ascomycete, *Thielavia terrestris,* only found a single cellulase gene to be in the most highly expressed genes when grown on rice straw compared to 11 cellulases for growth on Avicel (259, 326). It was also observed that compared to growth on glucose, predominantly oxidative AA LPMO family CAZymes and hemicellulose-active enzymes were up-regulated during *T. terrestris* growth on rice straw (259). The reason for the lack of cellulose targeting GHs also observed here is unclear but may suggest a predominantly oxidative approach to rice straw deconstruction instead. Relatively high abundances of AA3 and AA12 family CAZymes was observed in the rice straw secretome supernatant fraction, where the probes are deployed (**Figure 4.7**). These oxidative enzymes act within LPMO systems, however relatively low abundances of the AA9 LPMOs were also observed. Further investigation specifically into growth on rice straw would be required to fully understand the strategy of lignocellulose breakdown adopted here by *P. putredinis* NO1, but is warranted considering the abundance and environmental issues posed by the vast amounts of rice straw generated annually (65).

Substrate-specific xylanase production patterns have been observed previously using the same fluorescent probe to investigate the growth of *Aspergillus niger* on beechwood xylan (325). The xylanase probe detected similar variation across substrates here, most noticeably for the wheat bran substrate where xylanase production was clearly the most dominant of the three types of GH screened here (**Figure 4.8**). Cellulase production was detected despite being obscured on the gel images by the high xylanase signals (**Figure 4.9**). However, cellulase production was not sustained until day 10 as it was for wheat straw, sugar cane bagasse, and empty fruit bunch from palm oil substrates. This reflects the pattern of reduced cellulose degrading CAZyme family abundance observed for wheat bran in the proteomic investigation, a substrate with reduced cellulose content (319). Although, wheat bran has been explored as a potential substrate for cellulase production from the ascomycete *T. reesei,* where the nitrogen-rich substrate induces cellulase production (328, 329). This demonstrates how varied fungal lignocellulose-degrading enzyme responses can be for the same substrate and how characterisation of new lignocellulose-degrading systems, such as that of *P. putredinis* NO1, are important for understanding different approaches which may adopted industrially for the conversion of biomass.

GH production patterns looked similar for wheat straw and sugar cane bagasse substrates aside from some additional low molecular weight xylanases detected for the latter. All substrates except rice straw also showed the presence of a 57 kDa glucosidase at high levels on day 5 which then became undetectable by day 7 for wheat straw, empty fruit bunches from palm oil, and sugar can bagasse substrates, often giving way to a series of higher MW glucosidases. This two-stage pattern of glucosidase production may represent distinct responses to water-soluble glucosides vs. polysaccharide-derived glucosides. It has also been observed previously in microbial communities that β-glucosidase patterns can shift depending on other available carbon sources (330). Perhaps, the two-stage pattern of β-glucosidase production observed here also reflects a change in available carbon after day 5 of growth on these substrates.

Time-dependent analysis of the oil palm empty fruit bunch secretomes showed that induction was slow, with no enzymes detected at day 4. Day 5 showed a strong induction across all samples of the production of cellulases, glucosidases, and xylanases. Especially obvious was the sustained high production of the highest molecular weight cellulase detected across all substrates investigated (**Figure 4.8** and **Figure 4.9**). The production of hemicellulose-degrading enzymes by *A. niger* on this substrate has been explored previously, however exploration of cellulase production has not been performed for fungal isolates (331). The growth of the bacterium *Paenibacillus macerans* on empty fruit bunches from palm oil has been explored where degradation of the substrate was demonstrated (332). However,

cellulase activity from this microorganism failed to be demonstrated as it was here for *P. putredinis* NO1. Overall, this time-dependent analysis demonstrates the time-dependent adaptability of the *P. putredinis* NO1 secretome to different lignocellulosic substrates and illuminates the different CAZyme specificities deployed by this fungus.

4.5. Conclusions

Proteomic analysis of the secretome of *P. putredinis* NO1 grown on multiple lignocellulosic substrates showed substrate dependent variation of the CAZyme complement both bound to the biomass and in the culture supernatant. This variation was maintained when comparing the abundances of CAZymes at both the class and family level. Patterns of abundance of important lignocellulose degrading enzymes were observed, for example the lack of crystalline cellulose targeting enzymes during growth on substrates with reduced cellulose contents. When investigating CAZymes at the family level, which more specifically defines enzyme activities, the abundance profiles were found to vary greatly across substrates in both fractions of the lignocellulose degrading environment explored. This likely reflects the varying structures and compositions of the industrially relevant lignocellulosic substrates investigated here and suggests a potential ability for *P. putredinis* NO1 to actively tailor its enzymatic response.

It could be argued that differences in proteomic data across substrates in fact reflected different stages of fungal growth on the different substrates. However, by utilising an ABPP-based approach to visualise secretomes it was demonstrated that the diversity and titres of active cellulases, xylanases, and β-glucosidases in the secretomes varied both with substrate and over time. Therefore, fungal growth stage is not solely responsible for differences observed between the secretomes. Expected patterns were also observed here, such as the sustained muted cellulase production on wheat bran which has a reduced cellulose content (319). Patterns of production of these GH class CAZymes were also found which both agreed and contradicted previous reports on fungal GH expression on lignocellulosic substrates. This highlights how varied and complex lignocellulose degrading systems can be and demonstrates the value in exploring new systems like *P. putredinis* NO1 to better understand lignocellulose breakdown and to identify novel enzymes.

Understanding the adaptability of fungal secretomes will allow increased efficiency in the depolymerisation and biorefining of lignocellulosic substrates. Furthermore, the significant number of unknown proteins identified in the secretome after filtering through the bioinformatics pipeline suggests that there are still new lignocellulosic activities to be discovered.

4.6. Materials and methods

4.6.1. Strain isolation

P. putredinis NO1 was isolated from a wheat straw enrichment culture and maintained as reported previously (229).

4.6.2. *P. putredinis* NO1 cultures for proteomics

Triplicate 500 mL solutions of media containing 1.5 % (w/v) rice straw, wheat straw, sugarcane bagasse, wheat bran, empty fruit bunches from palm oil or containing 5 % (w/v) kraft lignin were inoculated to a final concentration of 10^5 spores/mL of *Parascedosporium putredinis* NO1. Cultures were incubated at 30 °C at 150 rpm for 4 days before wet mycelia was harvested for transcriptomic and proteomic investigation. The optimised media 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, Yeast Extract 8.85 g/L and Hutner's trace elements and was based on *Aspergillus niger* media (333).

4.6.3. Harvesting the *P. putredinis* NO1 proteomes across substrates

To harvest supernatant proteins, culture supernatants were centrifuged at 12,000 x g for 20 minutes at 4 °C and were then filter sterilised through 0.22 µm PES filter units. Triplicate 5 mL technical replicates for each culture were combined with 5 volumes of ice cold 100 % (v/v) acetone and mixed by inverting before incubation overnight at -20 °C. Samples were then centrifuged at 4500 rpm for 20 minutes at 4 °C and the acetone supernatant discarded. The pellets were then washed twice by the addition of ice cold 80 % (v/v) acetone, vortexing and centrifuging at 4500 rpm for 10 minutes at 4 °C. Pellets were air dried and resuspended in 1 mL of 0.5x PBS buffer before transferring to Eppendorfs and snap freezing in liquid nitrogen and storing at -80 °C.

To extract proteins bound to substrates, triplicate samples of 2 g of biomass from each culture were washed twice through the addition of 25 mL ice-cold 0.5x PBS and centrifugation at 4500 rpm for 5 minutes at 4 °C. The supernatant was discarded, and pellets were resuspended in 19 mL of 0.5x PBS with 1 mL of biotin (EZ-link-Sulfo-NHS-SS-biotin, Thermo Scientific) with rotation at 4 °C for 1 hour. Samples were then centrifuged at 4500 rpm for 10 minutes at 4 °C and the supernatant discarded. The reaction was quenched by the addition of 25 mL of 50 mM Tris-HCl pH 8.0 with rotation at 4 °C for 30 minutes. The biomass was pelleted as before, and the supernatants discarded. Biomass pellets were washed twice with 20 mL 0.5x PBS and were resuspended in 10 mL of 2 % (w/v) SDS pre-heated to 60 °C with rotation at 20 °C for 1 hour. Samples were centrifuged at 4500 rpm for 10 minutes and the supernatant collected. Five volumes of ice-cold 100 % (v/v) acetone was

added, and samples were incubated at -20 °C overnight. Samples were pelleted and washed in the same way as before for the supernatant fraction but were resuspended after air drying in 1 mL of 0.1 % (w/v) SDS solution before filtering through 0.22 µm PES filters.

Each replicate was loaded onto its own individual HiTrap Streptavidin HP 1 mL column (GE Healthcare) at a flow rate of 0.5 mL/min. Columns were left to incubate for 1 hour at 4 °C and were then washed with 10 mL of 0.1 % (w/v) SDS in PBS solution at a flow rate of 1 mL/min. Proteins were eluted by loading 1 mL of 50 mM DTT in PBS solution and incubating columns overnight at 4 °C. Another 1 mL of 50 mM DTT/PBS was added, and 1 mL of protein eluted, the column was incubated for 1 hour at 4 °C and then another 1 mL 50 mM DTT/PBS was added and another 1 mL of protein collected.

Both supernatant and bound fraction proteins were then desalted by spinning the samples through 5 mL Zeba™ spin columns 7k MWCO. Desalted samples were then lyophilised overnight and resuspended in NuPAGE loading buffer before being loaded onto NuPAGE gels (Invitrogen). Gels were run for 6 minutes at 180 V.

4.6.4. Peptide identification by LC-MS/MS

Protein samples in gels were then prepared for and subjected to label-free LC-MS by the Metabolomics and Proteomics department in the Technology Facility at the University of York.

In-gel tryptic digestion was performed post reduction with DTE and S-carbamidomethylation with iodoacetamide. Extracted peptides were loaded onto an mClass nanoflow UPLC system (Waters) equipped with a nanoEaze M/Z Symmetry 100 Å C 18, 5 µm trap column (180 µm x 20 mm, Waters) and a PepMap, 2 µm, 100 Å, C 18 EasyNano nanocapillary column (75 mm x 500 mm, Thermo). The trap wash solvent was aqueous 0.05 % (v/v) trifluoroacetic acid and the trapping flow rate was 15 µL/min. The trap was washed for 5 min before switching flow to the capillary column. Separation used gradient elution of two solvents: solvent A, aqueous 0.1 % (v/v) formic acid; solvent B, acetonitrile containing 0.1 % (v/v) formic acid. The flow rate for the capillary column was 300 nL/min and the column temperature was 40 °C. The linear multi-step gradient profile was: 3-10 % B over 8 mins, 10-35 % B over 115 mins, 35-99 % B over 30 mins and then proceeded to wash with 99 % solvent B for 4 min. The column was returned to initial conditions and re-equilibrated for 15 min before subsequent injections.

The nanoLC system was interfaced with an Orbitrap Fusion Tribrid mass spectrometer (Thermo) with an EasyNano ionisation source (Thermo). Positive ESI-MS and MS 2 spectra were acquired using Xcalibur software (version 4.0, Thermo). Instrument source settings were: ion spray voltage, 1900-2100 V; sweep gas, 0 Arb; ion transfer tube temperature; 275

°C. MS 1 spectra were acquired in the Orbitrap with: 120,000 resolution, scan range: m/z 375-1,500; AGC target, 4e5; max fill time, 100 ms. Data dependent acquisition was performed in top speed mode using a 1 s cycle, selecting the most intense precursors with charge states >1. Easy-IC was used for internal calibration. Dynamic exclusion was performed for 50 s post precursor selection and a minimum threshold for fragmentation was set at 5e3. MS 2 spectra were acquired in the linear ion trap with: scan rate, turbo; quadrupole isolation, 1.6 m/z; activation type, HCD; activation energy: 32 %; AGC target, 5e3; first mass, 110 m/z; max fill time, 100 ms. Acquisitions were arranged by Xcalibur to inject ions for all available parallelizable time.

Peak picking, database searching and quantification of proteomic data from Thermo raw files was performed using FragPipe (v19.1). Data were searched against a custom database of all coding regions of the *P. putredinis* NO1 genome appended with common contaminants and reversed sequences. The default LFQ-MBR workflow was used with the following modifications: precursor mass tolerance = +/- 3 ppm; fragment mass tolerance = 0.5 Da; IonQuant, feature detection m/z tolerance = 3 ppm; MBR RT tolerance = 7.5 min; add MaxLFQ, MBR FDR = 0.01, MBR min ions = 2. Final protein-level data were filtered to 1 % FDR, a minimum protein probability of 0.99 and a minimum of 2 peptides. Molar percentage values were calculated for each protein in each sample as a percentage of the sum of MaxLFQ values for each sample.

4.6.5. Quality control of proteomic data

Identified protein count, principal component analysis, and hierarchical clustering with a Canberra distance matrix and wardD2 clustering were used to investigate proteomic data for all biological replicates for all substrates investigated. This analysis was performed in R studio 4.2.3 using the ‘ggplot2’, ‘FactoMineR’, ‘factoextra’, ‘ggdendro’, and ‘dendextend’ packages (293, 294, 334-337). From this analysis, outliers were removed to provide a final dataset for comparative analysis.

4.6.6. Isolating the *P. putredinis* NO1 secretome

Proteins were filtered to produce the *P. putredinis* NO1 secretome using the ‘strict’ filtering workflow developed previously (318). Secretome proteins were predicted to be extracellular by both BUSCA and DeepLoc localisation prediction tools (309, 310), or which were predicted to encode a secretion signal by SignalP, TargetP and SecretomeP tools (275, 311, 312), and which lacked more than one predicted transmembrane domain by TMHMM or contained a single predicted transmembrane helix with more than 10 amino acids of this helix occurring in the first 60 amino acids of the protein sequence (313).

4.6.7. Comparing the *P. putredinis* NO1 secretome across substrates

All comparative analysis was performed in R studio v 4.2.3 (293), and analysis was repeated for the total proteome and the filtered secretome.

The number of proteins identified in at least one replicate across all substrates was compared between the bound and supernatant fraction using the 'ggVennDiagram' package (338). To visualise differences across substrates, heatmaps were created to compare the average molar percentage of proteins across substrates using the 'pheatmap' package (339). Principal component analysis was carried out to investigate replicate grouping within and between substrates using the 'FactoMineR' and 'factoextra' packages (334, 335). Canberra distance matrix calculation with wardD2 clustering was used to plot dendrograms to investigate replicate clustering within and between substrates using the 'ggdendro', and 'dendextend' packages (336, 337). GO category and GO term annotation was performed using Blast2GO in the OmicsBox package (340, 341) and abundances visualised and compared using the 'ggplot2' package (294). CAZyme domain annotation was performed using the dbCAN server (178), and abundances visualised and compared using the 'ggplot2' package (294).

4.6.8. Fluorescence-based activity-based protein profiling

For ABPP, 2 µL of triplex probe mixture (60 µM JJB376, CB644, SYF230) was added to 18 µL of secretome without added buffer (measured pH ~7.5). The reactions were incubated for 1 hour at 30 °C then quenched by the addition of 8 µL of 4x Laemmli buffer (Bio-Rad) and heating to 95 °C for 2 minutes. Then, 10 µL of the resulting solution was separated over either a 10 % SDS-PAGE gel or a 4-20 % gradient gel (Bio-Rad) at 200 V. The resulting gels were transferred to a Typhoon 5 scanner without fixation and imaged using the Cy2, Cy3, and Cy5 laser/filter sets sequentially. The resulting images were integrated using ImageQuant (GE Healthcare).

4.7. Declarations

4.7.1. Availability of data and materials

All proteomic data generated during this research are available at MassIVE MSV000092129 https://massive.ucsd.edu/ProteoSAFe/private-dataset.jsp?task=ec1445b7b6d04a858b30fcdf5522b046

4.7.2. Competing interests

The authors declare that they have no competing interests.

4.7.3. Funding

This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC), UK (BB/P027717/1, BB/W000695/1, BB/W003309/1, BB/S01196X/1). CJRS was supported by a CASE studentship from the BBSRC Doctoral Training Programme (BB/M011151/1) with Prozomix Ltd. GJD is funded by the Royal Society Ken Murray Research Professor and GJD/HO though ERC-2020-SyG-951231 ”Carbocentre”.

4.7.4. Author contributions

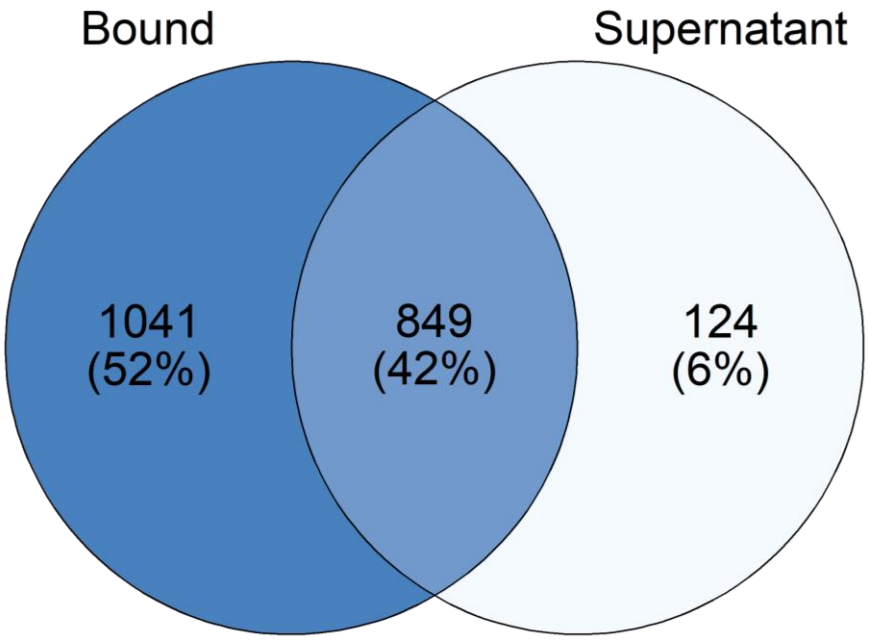
CJRS conceptualised the investigation carried out in this paper, performed all proteomic data quality control, secretome isolation, all comparative analysis, supernatant harvest for fluorescent labelling, and was the major contributor in writing the manuscript. NGSM carried out fluorescent labelling of culture supernatants, fluorescent imaging, and gel integration, and was a major contributor in writing the manuscript. NCO aided conceptualisation of the investigation carried out in this paper and was involved in proteome harvest. JH was involved in proteome harvest. AA, DL, and AS were involved in proteome harvest. AD performed proteomic identification of protein samples and supported proteomic data quality control and comparative analysis. HO supervised development of fluorescent probes. GD helped with supervision of the study. NCB was a major contributor to the conceptualisation and supervision of the study in addition to making a major contribution to the writing of the manuscript.

4.7.5. Acknowledgements

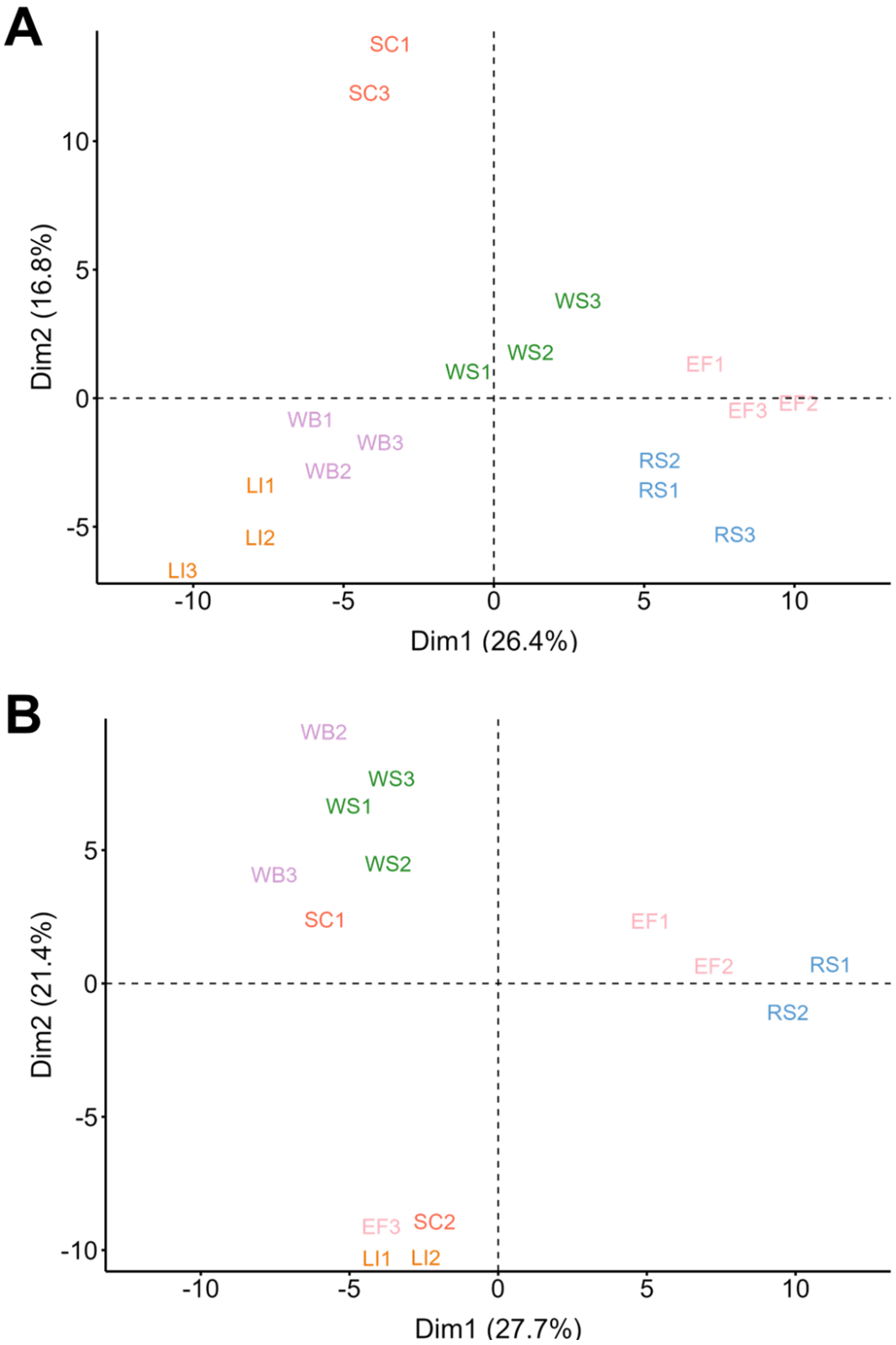
The York Centre of Excellence in Mass Spectrometry was created thanks to a major capital investment through Science City York, supported by Yorkshire Forward with funds from the Northern Way Initiative, and subsequent supports from ESPRC (EP/K039660/1; EP/M028127/1).

4.8. Supplementary material

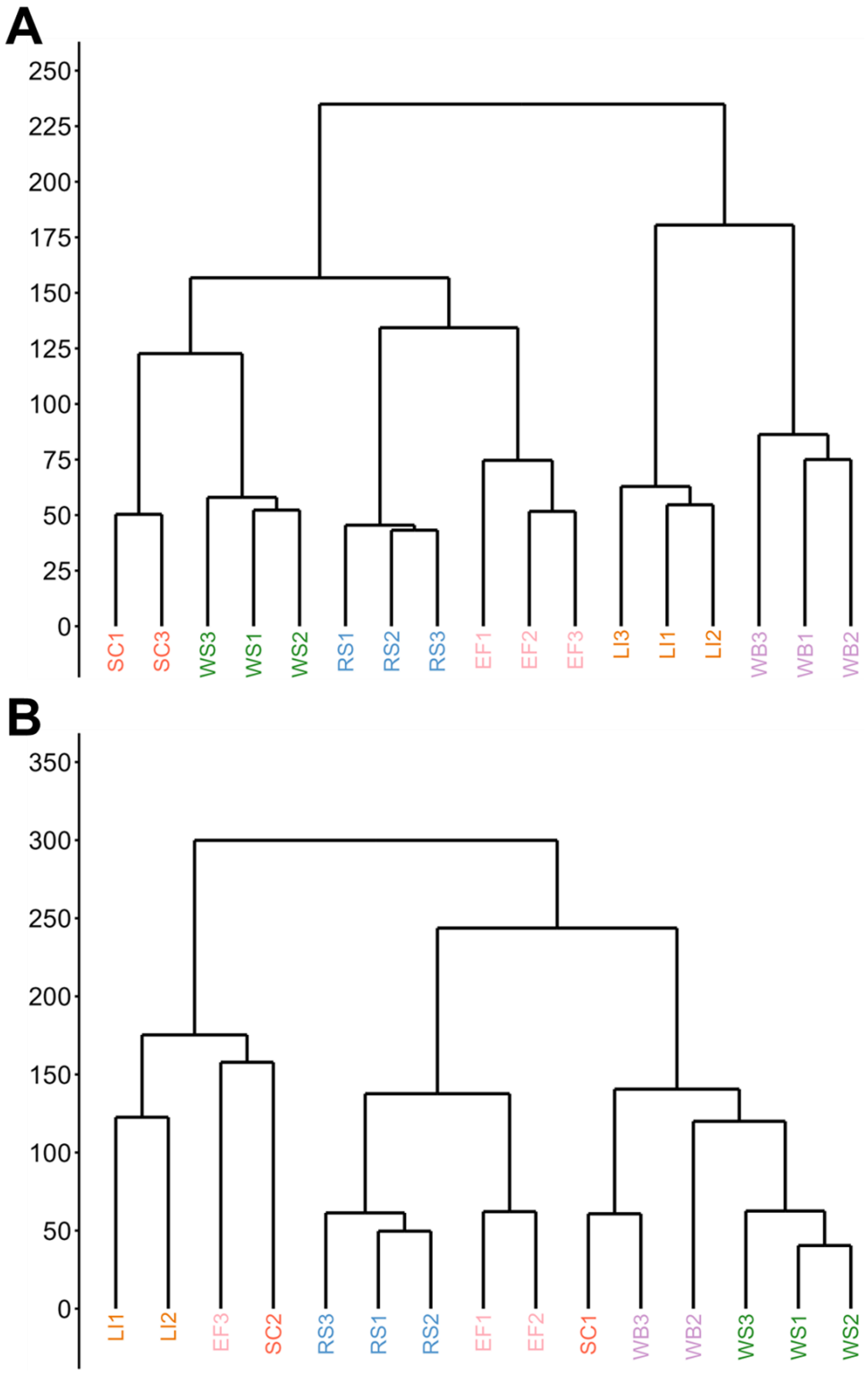
4.8.1. Supplementary figures



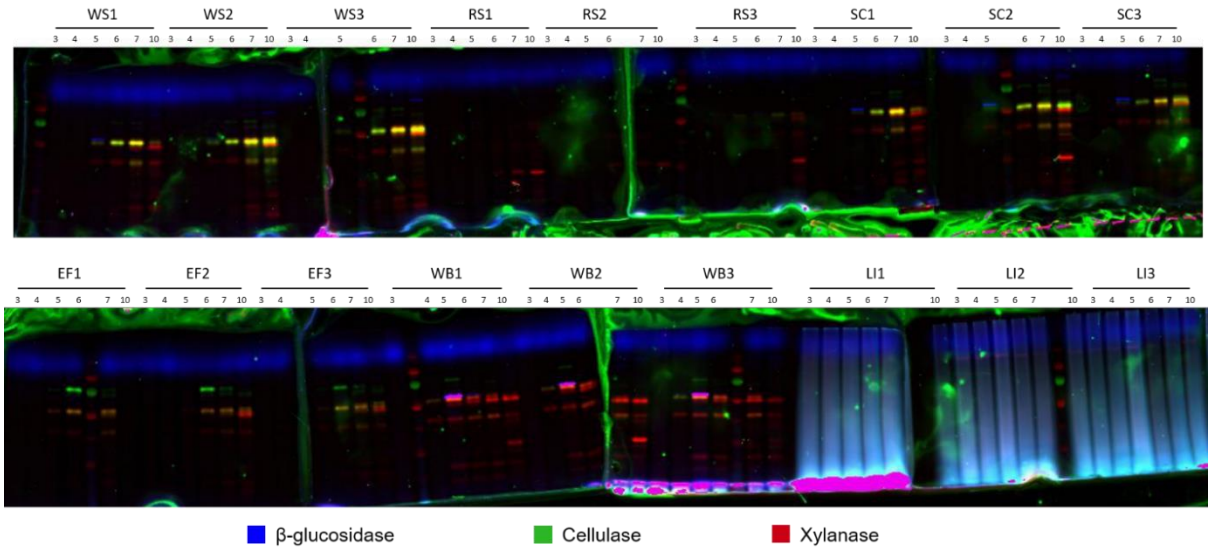
Supplementary Figure 4.1 Investigating the distribution of proteins across bound and supernatant fractions in the *P. putredinis* NO1 proteome. The number of proteins identified in at least one replicate across all substrates for the bound fraction compared to the supernatant fraction when *P. putredinis* NO1 was grown on six lignocellulosic substrates for 4 days.



Supplementary Figure 4.2 Variation in the *P. putredinis* NO1 secretome across lignocellulosic substrates. A PCA was calculated from the MolPct values of secretome proteins identified in the bound (**A**) and supernatant (**B**) fractions for cultures of *P. putredinis* NO1 grown on six substrates. EF = Empty fruit bunch, LI = Kraft lignin, RS = Rice straw, SC = Sugar cane bagasse, WB = Wheat bran, WS = Wheat straw.



Supplementary Figure 4.3 Clustering of replicates of the *P. putredinis* NO1 secretome across lignocellulosic substrates. A Canberra distance matrix was calculated and hierarchically clustered by Ward's minimum variance method for molar percentage values for proteins in the bound (**A**) and supernatant (**B**) fractions of the *P. putredinis* NO1 secretome after growth on six substrates. EF = Empty fruit bunch, LI = Kraft lignin, RS = Rice straw, SC = Sugar cane bagasse, WB = Wheat bran, WS = Wheat straw.



Supplementary Figure 4.4 Differences in *P. putredinis* NO1 glycoside hydrolase production over time visualised with activity-based probes. Fluorescence imaging following SDS-PAGE is shown for three replicates of samples of culture supernatants taken at days 3, 4, 5, 6, 7, and 10 of growth treated with a triplex probe mixture targeting cellulases, xylanases, and retaining β-glucosidases. EF = Empty fruit bunch, LI = Kraft lignin, RS = Rice straw, SC = Sugar cane bagasse, WB = Wheat bran, WS = Wheat straw.

Chapter 5

5. Isolating lignin degrading enzymes from the *Parascedosporium putredinis* NO1 secretome

Authors

Conor JR Scott[a], Adam Dowle[b], Neil C Bruce[a]

Affiliations

[a] Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, United Kingdom

[b] Bioscience Technology Facility, Department of Biology, University of York, York YO10 5DD, United Kingdom

5.1. Abstract

Effective utilisation of lignin, the most abundant aromatic polymer on earth, is required to achieve total substrate valorisation of lignocellulose to fuels, chemicals, and materials. For this to be realised in an environmentally friendly manner, biological approaches for lignin conversion need to be expanded beyond the use of traditional lignin degrading enzymes. *Parascedosporium putredinis* NO1 is an ascomycete within an underexplored genus, which has demonstrated the ability to tailor its secretome to different substrates, and from which a new lignin-degrading enzyme has been previously identified.

Here, proteomic identification and bioinformatic secretome prediction was used to investigate the growth of *P. putredinis* NO1 on wheat straw, delignified wheat straw, and dioxane lignin. Remarkable variability of the secretome was demonstrated across growth substrates, and expected patterns of variation for predicted lignin-degrading enzymes were observed. Isolation of proteins with significantly higher abundance on lignin-enriched substrates identified 85 proteins, 36 of which were implicated in lignocellulose breakdown. *Pp*CPO, a chloroperoxidase-like protein within this set of proteins, was recombinantly expressed and demonstrated to exhibit traditional haloperoxidase activity and the ability to cleave β-*O*-4 linkages in a model lignin compound. Incubation of *Pp*CPO with dioxane lignin demonstrated a reduction in vanillic acid in the soluble product profile.

Profiling the lignin-degrading potential of fungal secretomes, especially for underexplored lineages of fungi, will help to identify new enzymes with roles in lignin breakdown. Through the investigation of the *P. putredinis* NO1 secretome when grown on substrates with varying lignin contents, a new chloroperoxidase enzyme was identified. By exploring this new enzyme, the role of chloroperoxidases in lignin modification has been supported and the ability to identify new enzymes which are not traditionally implicated in lignin breakdown through such investigations has been demonstrated.

5.2. Keywords

Parascedosporium putredinis NO1, Lignocellulose, Lignin, Proteomics, CAZymes, Chloroperoxidase

5.3. Background

Lignocellulose is an abundant resource to produce renewable and sustainable fuels, chemicals, and materials. Traditionally lignocellulose has been viewed as a reservoir of polysaccharides to be degraded and converted to second-generation biofuels. However, for biorefineries to operate competitively with fossil fuels it has become clear that the whole substrate needs to be valorised (90). Lignocellulose consists primarily of three major components, the cellulose and hemicellulose polysaccharide components, and the lignin network (265). Lignin is an aromatic heteropolymer formed oxidatively and deposited into plant secondary cell walls, giving them their resistance to biological and chemical attack (88). Often the lignin network is seen as nothing but a barrier to access the polysaccharides, however as the most abundant aromatic polymer on earth it is a valuable reservoir of carbon (93).

The value of lignin can be seen in the incorporation of kraft lignin, generated from the kraft paper pulping process, in products such as asphalt and rubber, and of vanillin extracted from kraft lignin in fragrances (95, 96, 98). However the kraft process is an energy intensive and chemically aggressive treatment utilising high pH and temperatures exceeding 160 °C (320). To avoid the high capital cost associated with such treatments, and to avoid environmentally damaging practices moving forward, biological breakdown of lignin is an attractive process. Biological depolymerisation of lignin utilises microorganisms or isolated enzymes from lignocellulose degrading organisms to specifically degrade polymeric lignin while avoiding the stringent conditions required for physiochemical approaches (89). One major advantage is the use of cocktails of enzymes to degrade all components of lignocellulose in tandem to provide multiple valuable products and realise the biorefinery goal of total substrate valorisation. Although the intricate and insoluble nature of lignin makes enzymes access and therefore conversion efficiency challenging. Oxidative lignin degrading enzymes such as peroxidases and laccases are important enzymes for the conversion of lignin and lignocellulose and are well characterised (342-345). But to generate the large variety of high value products which lignin has the potential for, more than just these enzymes will be required.

Parascedosporium putredinis NO1 is an ascomycete fungus identified and investigated previously due to the potential to identify new lignin degrading enzymes (229, 301, 346). Comparison of the *P. putredinis* NO1 genome across the ascomycete tree of life revealed a high proportion of oxidative auxiliary activity class carbohydrate active enzymes (CAZymes) in the genome of this fungus, many of which are implicated in the breakdown of lignin (301). Additionally, the variable secretome produced by *P. putredinis* NO1 dependent on

lignocellulosic substrate suggested that carefully designed functional investigations could be used to isolate enzyme classes of interest, such as lignin degrading enzymes (346). Finally, the identification of a new β-ether cleaving oxidase enzyme from *P. putredinis* NO1 suggested that this fungus could harbour further lignin degrading enzymes with novel activities and industrial value (229).

To investigate the lignin degrading enzyme repertoire of *P. putredinis* NO1, the fungus was grown on three substrates with varying lignin content. Wheat straw, delignified wheat straw, and dioxane lignin extracted from wheat straw were used as substrates in cultures for the growth of *P. putredinis* NO1 to harvest both the soluble and substrate-bound fractions of the extracellular proteome (229, 233). Similar to a previous proteomic investigation of *P. putredinis* NO1, the secretome was remarkably variable across substrates and this variation was maintained when investigating CAZymes with predicted roles in lignocellulose breakdown. Expected changes in abundance of lignin-degrading protein families were observed across the substrates in both fractions suggesting that potentially new lignin degrading enzymes may demonstrate related patterns. The secretome was analysed to identify a subset of proteins with significantly higher abundance on substrates with higher lignin contents. Within this subset many proteins with unclear potential roles in lignin breakdown were identified, for example a chloroperoxidase-like protein. Chloroperoxidases have been implicated in the modification and degradation of model lignin compounds previously (249, 250). Here, this chloroperoxidase-like protein (*Pp*CPO) was recombinantly expressed, purified, and the potential for its involvement in modification of polymeric lignin from an industrially relevant substrate was investigated.

5.4. Results and discussion

5.4.1. The *P. putredinis* NO1 secretome varies across substrates with different lignin contents

To develop a proteomics experiment from which the lignin degrading enzymes of the *P. putredinis* NO1 secretome could be isolated, three substrates for growth were obtained. Delignified wheat straw was produced through sodium chlorite and glacial acetic acid treatment of wheat straw. Delignification of lignocellulosic biomass using the acidified sodium chlorite method has been well-established previously (347-349). This approach has been demonstrated to effectively remove lignin while minimising degradation and solubilisation of the polysaccharide components (348, 350). The proportions of lignin, cellulose, hemicellulose, and ash were quantified and investigated for the delignified and normal wheat straw to confirm a reduction in lignin content (**Table 5.1**). Clear differences were seen between normal and delignified wheat straw, with lignin content greatly reduced from 34.2 to 14.2 %. Some lignin remains, likely due to the strong covalently linked lignin-carbohydrate complexes present in lignocellulosic biomass (348). However, the large reduction in lignin content was still expected to increase accessibility to the polysaccharide components of the substrates and therefore elicit a change in the *P. putredinis* NO1 secretome. The increased lignin content was investigated further by Fourier transformed infrared spectroscopy (FTIR), which demonstrated decreased absorbance around 2900, and 1700 to 1200 cm^{-1} wavelengths that could correspond to bonds within structural groups of lignin (**Supplementary Figure 5.1**) (350-353).

Table 5.1 Biomass composition. Proportions of acetyl bromide soluble lignin, cellulose, hemicellulose sugars, and ash content were quantified and calculated for triplicate samples of normal and delignified wheat straw.

Component	Normal wheat straw (%)	Delignified wheat straw (%)
Lignin	34.2	14.2
Cellulose	40.5	67.6
Hemicellulose	11.2	11.4
Ash	14.1	6.8

The increased cellulose proportion mirrors the reduced proportion of lignin in the delignified wheat straw compared to untreated wheat straw. Acidified sodium chlorite treatment has been shown to minimally effect the structure of cellulose previously (350). The lignin content could have potentially been lowered further through a more aggressive sodium chlorite treatment utilising higher temperatures for longer periods of time. However, the increased

lignin removal would likely result in a higher loss of polysaccharides from cellulose and alteration of the polysaccharide component structures (349). An increase in proportion was not observed for hemicellulose, likely due to the removal of some hemicellulose by solubilisation during the delignification process (349). By examining the monosaccharide content of the hemicellulose within both normal and delignified wheat straw, it was clear that the monosaccharide proportional profiles were still very similar (**Supplementary Table 5.1**). The ash content was reduced in the delignified compared to normal wheat straw, possibly due to the removal of inorganic residues from the surface of the biomass (354). Regardless, the major difference in biomass composition here is the largely reduced lignin content. The biomass was finally investigated through SEM imaging of normal and delignified wheat straw and differences were clearly seen on the surfaces of the biomass. The surface of fibres of delignified material appeared to be much cleaner and smoother, potentially from where the lignin and any surface components were stripped away by delignification (**Supplementary Figure 5.2**). A similar change was observed for the delignification of silvergrass and was attributed to the loss of waxes on the surface of lignocellulosic fibres (347). In the same study the removal of lignin also resulted in fibres with coarse and uneven surfaces, and this has also been observed for sugar cane bagasse however was not observed here for wheat straw (347, 350).

To harvest the lignin degradation enzymes contained in the *P. putredinis* NO1 repertoire, the ascomycete was grown with three substrates. *P. putredinis* NO1 was grown in liquid cultures containing wheat straw, delignified wheat straw, and dioxane lignin generated from the wheat straw. Dioxane is one of many organic solvents which can be used for the solubilisation of lignin from lignocellulosic biomass. Utilisation of organic solvents and water to create a cooking liquor that solubilises lignin components is often referred to as the “organosolv” process (355). The product is a more homogeneous lignin product that is rich in functional groups and includes minimal carbohydrate contamination (127). After growth of *P. putredinis* NO1 on these substrates for 7 days, proteins were harvested from culture supernatants, and from the insoluble substrates utilising a biotin surface protein labelling technique developed previously and which has been used in a similar experimental setup to explore the *P. putredinis* NO1 secretome previously (233, 346). Quality control was performed through investigation of protein count, performing principal component analysis, and clustering for all replicates. After outlier removal, three biological replicates were maintained for all substrates in both bound and supernatant fractions for analysis.

In total, 1423 proteins were identified across bound and supernatant fractions with abundance in at least one replicate. In the bound fraction 1357 proteins were identified and 394 proteins were identified in the supernatant fraction, with 328 proteins shared between

both (**Supplementary Figure 5.3**). As observed previously, many intracellular proteins were identified within this proteomic dataset which were likely released by cell lysis occurring in these cultures and therefore the dataset was filtered bioinformatically to isolate the "secretome" proteins. A bioinformatic workflow presented previously was utilised which uses both localisation prediction and secretion signal prediction to reflect the conventional and unconventional pathways by which fungi release proteins (163, 318). Proteins predicted to be extracellular but also cell-surface bound through transmembrane anchors were omitted from this predicted secretome as these are unlikely to be involved in the lignin degrading or modifying activity of interest here. The resulting secretome contained 212 proteins in total, with 178 proteins identified in the bound fraction, 153 proteins identified in the supernatant fraction, and 119 proteins shared between both (**Figure 5.1**). The discrepancy in proteins isolated in the secretome across the bound and supernatant fractions highlights the importance of capturing both fractions in investigations into fungal lignocellulose degrading response.

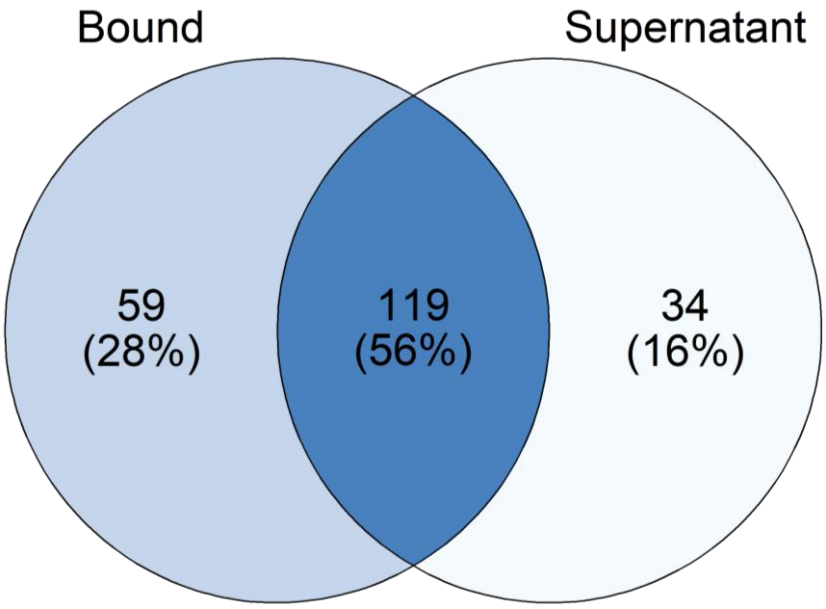


Figure 5.1 Investigating the distribution of proteins across bound and supernatant fractions in the *P. putredinis* NO1 secretome. The number of proteins identified in at least one replicate across all substrates for the bound fraction compared to the supernatant fraction within the predicted *P. putredinis* NO1 secretome from growth on three lignocellulosic substrates for 7 days.

Clear variation was visible in the abundance profiles of the *P. putredinis* NO1 proteome dependent on the growth substrate (**Supplementary Figure 5.4**), and this variation was maintained after bioinformatic filtering to produce the *in silico* secretome (**Figure 5.2**). Variation dependent on substrate was also demonstrated through principal component analysis, where replicates grouped well within substrate and showed very good separation

between substrates in both bound and supernatant fractions (**Supplementary Figure 5.5**). A similar pattern was also seen when investigating the secretome through clustering and plotting on dendrograms, where replicates clustered within substrate but showed distance between substrates (**Supplementary Figure 5.6**).

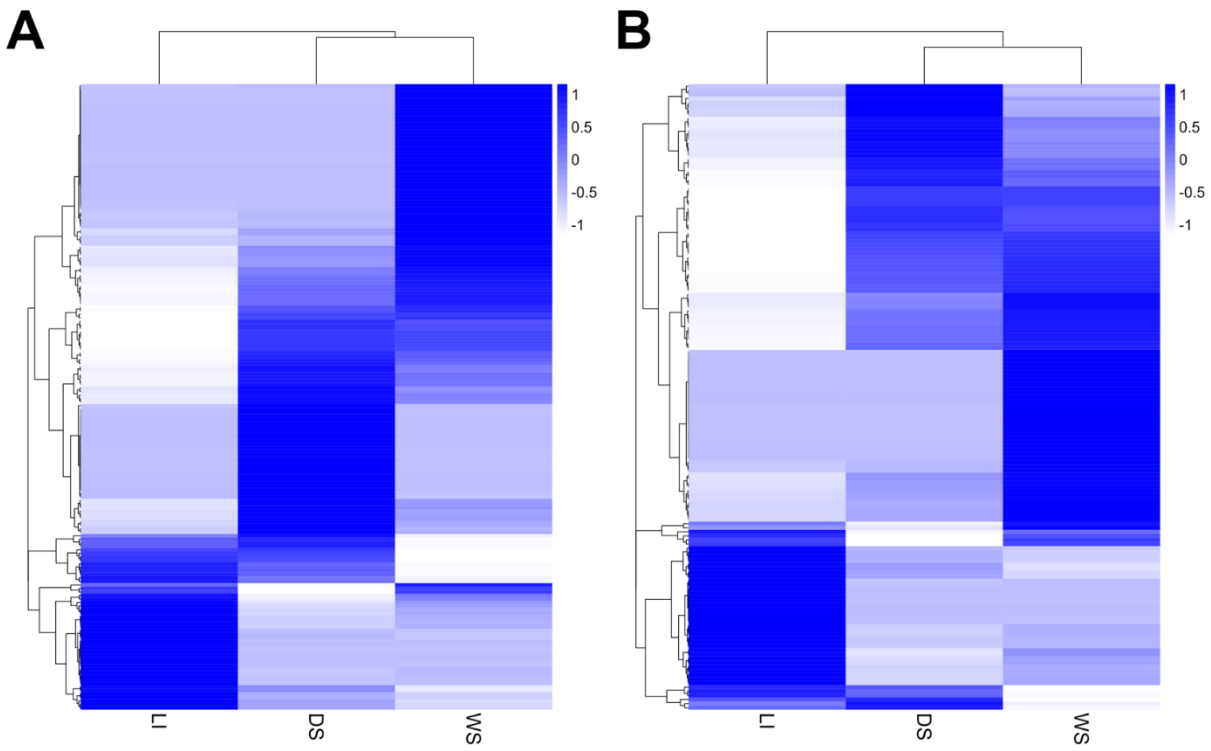


Figure 5.2 Differences in the *P. putredinis* NO1 secretome across lignocellulosic substrates. Molar percentage values for proteins identified on at least one substrate in the bound (**A**) and supernatant (**B**) fractions of the *P. putredinis* NO1 secretome scaled to Z-scores across substrates. DS = Delignified straw, WS = Wheat straw, LI = Dioxane lignin.

The maintenance of this variation when looking solely at proteins which are more confidently predicted to be actively released into the extracellular space aligns with the tailored secretome observed previously for this fungus and is exciting for identification of lignin degrading enzymes. Indeed, the tailoring of fungal secretomes to different lignocellulosic substrates has been explored previously to enable the tailoring of industrial enzyme cocktails for different feedstocks (244). For example, when *Trichoderma reesei* is grown on different lignocellulosic substrates it produces a secretome which can more effectively release sugars from the same substrate used for growth compared to a broad-spectrum commercial cocktail (244). The secretome of *P. putredinis* NO1 was demonstrated to vary remarkable when grown on industrially relevant lignocellulosic substrates (346). Through activity-based labelling of hydrolase enzymes, differences in the lignocellulose-degrading secretome response of *P. putredinis* NO1 were found to be maintained temporally, suggesting a tailored response for plant biomass breakdown (346).

5.4.2. The *P. putredinis* NO1 secretome contains unknown enzyme activities

The functional profiles of the *P. putredinis* NO1 secretome were found to vary across substrates and contain unknown activities in high proportions. The proportional abundance of proteins annotated with CAZyme domains, if not annotated as CAZymes then assigned to GO categories, or not annotated with either (Unknown) were compared across substrates (**Figure 5.3**). Variation was observed in the functional profiles of the secretome across substrates for the bound and supernatant fractions. CAZymes and “molecular function” proteins contribute the most to abundance for most substrates in both fractions. Unknown proteins also contribute to abundance in both fractions for all substrates, although the contribution is much larger for the lignin substrate in the supernatant fraction. The occurrence of unknown proteins in the secretome, especially those which are abundant during growth on dioxane lignin was promising for identification of new lignin degrading enzymes. All substrates in both fractions also contain contributions of “cellular component” annotated proteins, which demonstrates that despite filtering of proteomic data to produce the *P. putredinis* NO1 secretome that some intracellular proteins persist. This could result from mis-annotation of the protein sequences as secreted proteins by bioinformatic tools used in the secretome isolation workflow (318), or from the mis-annotation of secretome proteins as intracellular during the functional annotation performed here. Either way, the secretome isolation workflow was evaluated and demonstrated to effectively remove most proteins annotated with intracellular functions previously (318).

Differences in the functional profiles of the *P. putredinis* NO1 secretome were also observed at the level of the GO term annotation (**Figure 5.4**). Extracellular activities were predominantly seen as the most abundant contributors to protein abundance for all substrates, and the patterns of abundances of these activities vary between bound and supernatant fractions, and between substrates. The bound fraction of the dioxane lignin secretome contains more intracellular activities, likely reflecting the disproportionate acquisition of intracellular and fungal membrane bound proteins by the biotin labelling approach for a more soluble substrate (**Figure 5.4C**). Excitingly, the *P. putredinis* NO1 secretome contains proteins with no functional annotation, or “Unknown” proteins, in the top ten most abundant secretome activities during growth on all three substrates and in both soluble and insoluble fractions of the cultures.

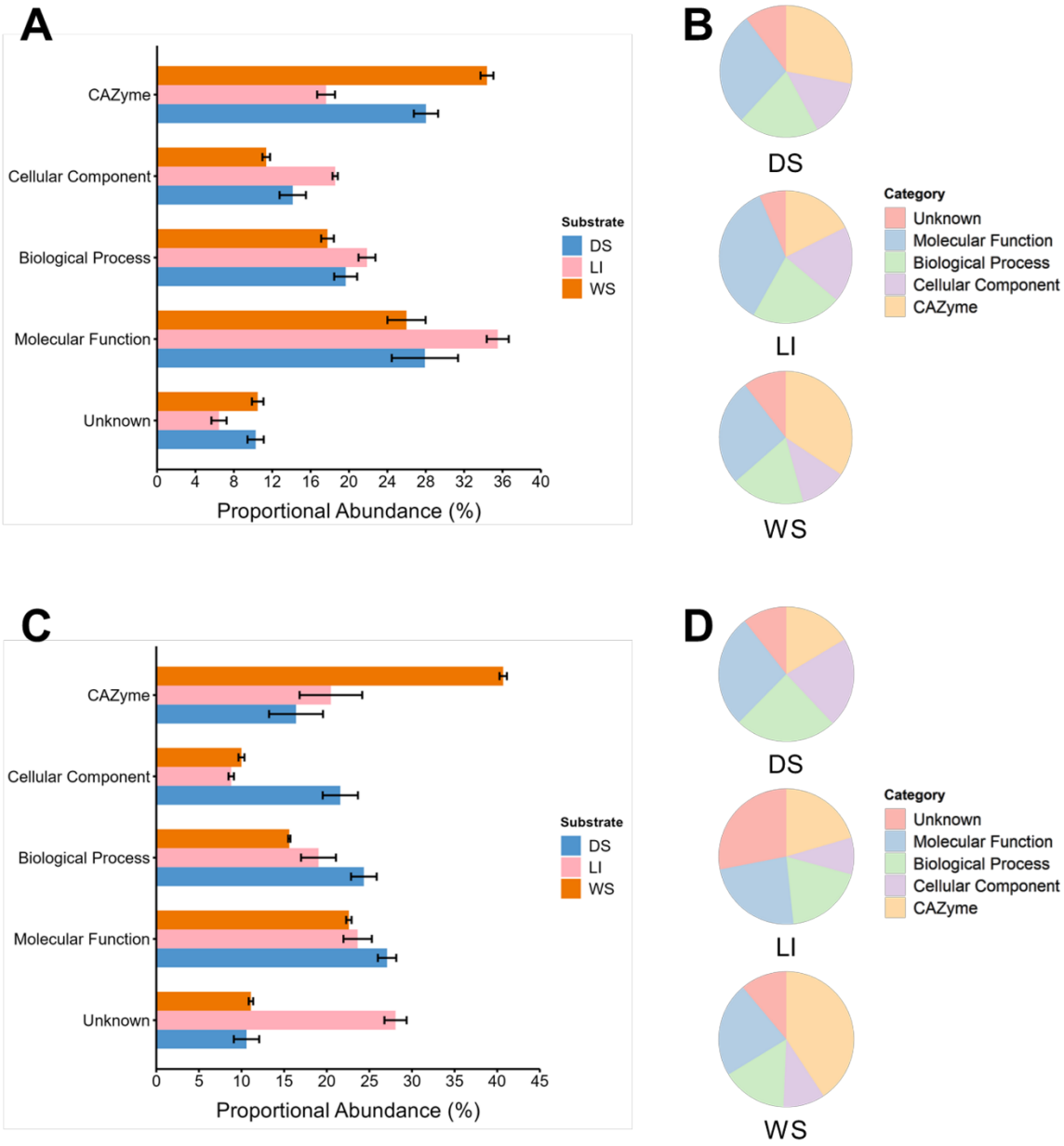


Figure 5.3 Differences in functional categories of the *P. putredinis* NO1 secretome across lignocellulosic substrates. The molar percentage abundance of proteins identified as CAZymes, that were not identified as CAZymes but assigned to GO categories, or that were not annotated as CAZymes or with GO categories were calculated proportionally and compared across substrates for each category for the bound (**A**) and supernatant fractions (**C**) of the *P. putredinis* NO1 secretome. The proportions of each functional category for each substrate were also visualised for the bound (**B**) and supernatant (**D**) fractions separately. DS = Delignified straw, WS = Wheat straw, LI = Dioxane lignin.

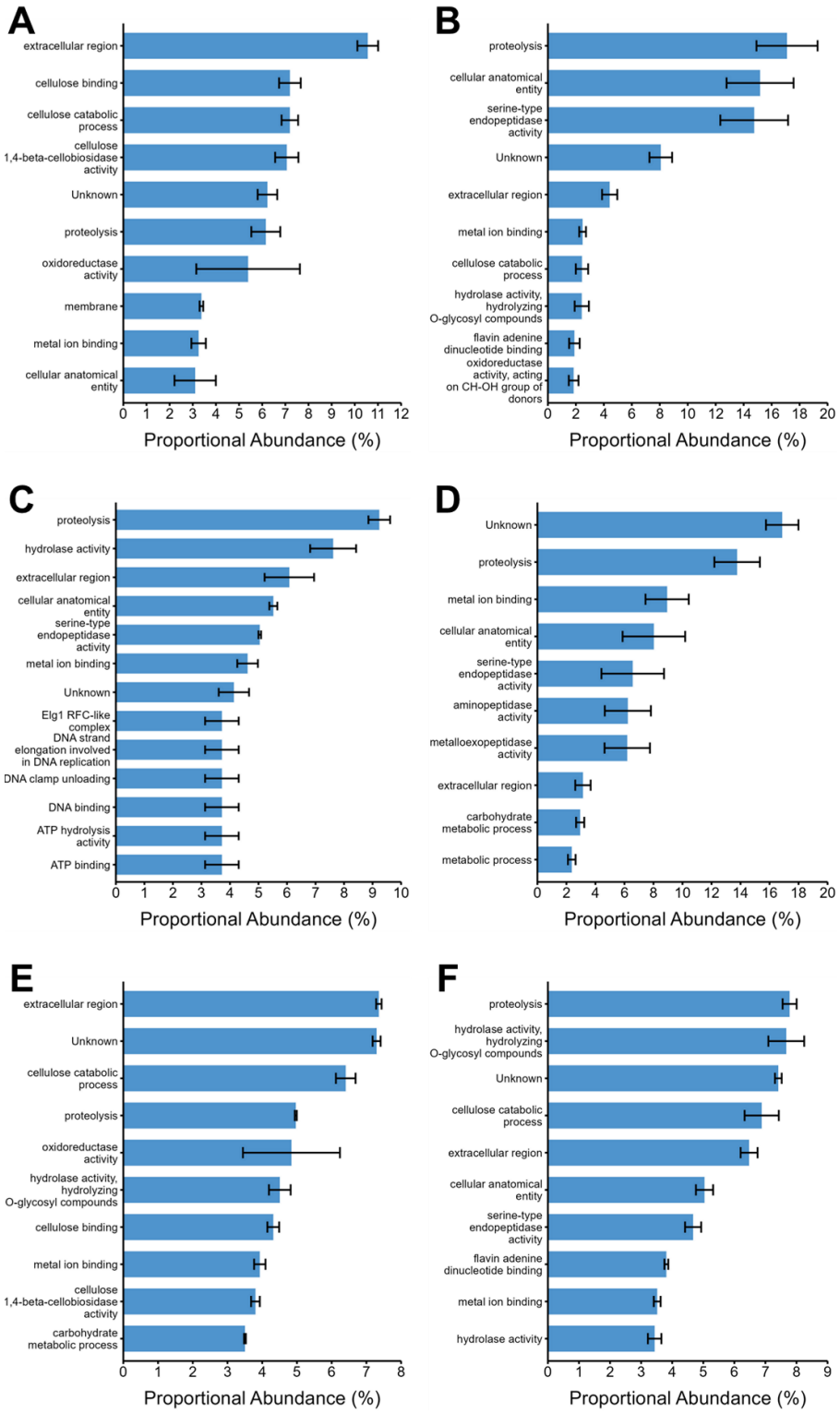


Figure 5.4 Proportional abundances of enzyme activities in the *P. putredinis* NO1 secretome. The molar percentage abundance of proteins assigned to GO terms were calculated proportionally for the bound (**A**, **C**, and **E**) and supernatant (**B**, **D**, and **F**) fractions of the *P. putredinis* NO1 secretome after seven days of growth on three substrates. DS = Delignified straw (**A** and **B**), WS = Wheat straw (**C** and **D**), LI = Dioxane lignin (**E** and **F**).

5.4.3. The CAZyme profile of the *P. putredinis* NO1 secretome varies across substrate

The abundance profiles of CAZymes within the *P. putredinis* NO1 secretome were compared across substrates at both the class and family level. The class profiles were found to appear quite similar across substrate (**Figure 5.5**), particularly for the wheat straw and delignified straw substrates and particularly in the supernatant fraction. The dioxane lignin substrate looks unique in the bound fraction, where it has a much larger contribution of carbohydrate esterase (CE) class CAZymes to total CAZyme abundance. Lignin obtained from wheat straw by organosolv processes has previously been reported to contain minimal amounts of hemicellulosic polysaccharides and it has been suggested that *p*-coumaric and ferulic acid linkages between lignin and hemicellulose may be more resistant to this process (355). Potentially the presence of these linkages in a more exposed manner for the dioxane lignin substrate elicits the production of CE class CAZymes which may act on these linkages. The increased proportional abundance of CE class CAZymes in the bound fraction during growth on dioxane lignin is mirrored by a reduction in glycoside hydrolase (GH) class CAZyme abundance. This reduction was expected considering the polysaccharide bonds upon which these enzymes act were likely to be present in minimal amounts in the lignin substrate.

Despite the somewhat similar overall CAZyme class profiles, remarkable variation was observed when investigating CAZyme abundance profile at the family level (**Figure 5.6**). A pattern of higher relative abundance of many CAZyme families for wheat straw, less for delignified straw, and even fewer for dioxane lignin was seen in both bound and supernatant fractions. This was expected considering wheat straw has all the lignocellulosic components upon which CAZymes act, delignified straw has predominantly polysaccharide components but a reduced lignin component, and the dioxane lignin substrate is suspected to be predominantly composed of lignin with minimal polysaccharides (355). It was also observed that there were promising patterns of abundance for known lignin degrading enzymes. For example, AA1 family CAZymes were present in higher relative abundance on the wheat straw and dioxane lignin substrates but were reduced on the delignified biomass in both the bound and supernatant fractions. The AA1 family contains the laccase and multi-copper oxidase enzymes which efficiently cleave, oxidise, and demethylate lignin structures (191, 192). Therefore, it is understandable that these enzymes are produced in relatively higher abundances on substrates where lignin is present in higher proportions. An increased abundance of AA9 family CAZymes on the delignified substrate compared to other substrates in both bound and supernatant fractions was observed. These lytic polysaccharide monooxygenase (LPMO) enzymes oxidatively cleave bonds in crystalline cellulose, of which there is an increased proportion in this substrate (179, 182). Aside from the catalytic classes of CAZymes, many CAZyme proteins also contain additional domains

called carbohydrate binding modules (CBMs). These structurally discrete domains provide binding capacity to carbohydrate ligands to bring the catalytic domains in close proximity to their target substrates to increase catalytic efficiency (201, 205). Reduced abundance of all CBM families except for CBM5 was observed for the dioxane lignin substrate in both fractions. This was expected due to the absence of the polysaccharide substrates which these CAZyme domains are involved in binding (202). CBM5 domains are involved in the binding of chitin and are more likely to be involved in fungal cell wall processes here (209). Overall, these differences inspired confidence that novel lignin degrading enzymes could have been following the same tailored patterns of abundance.

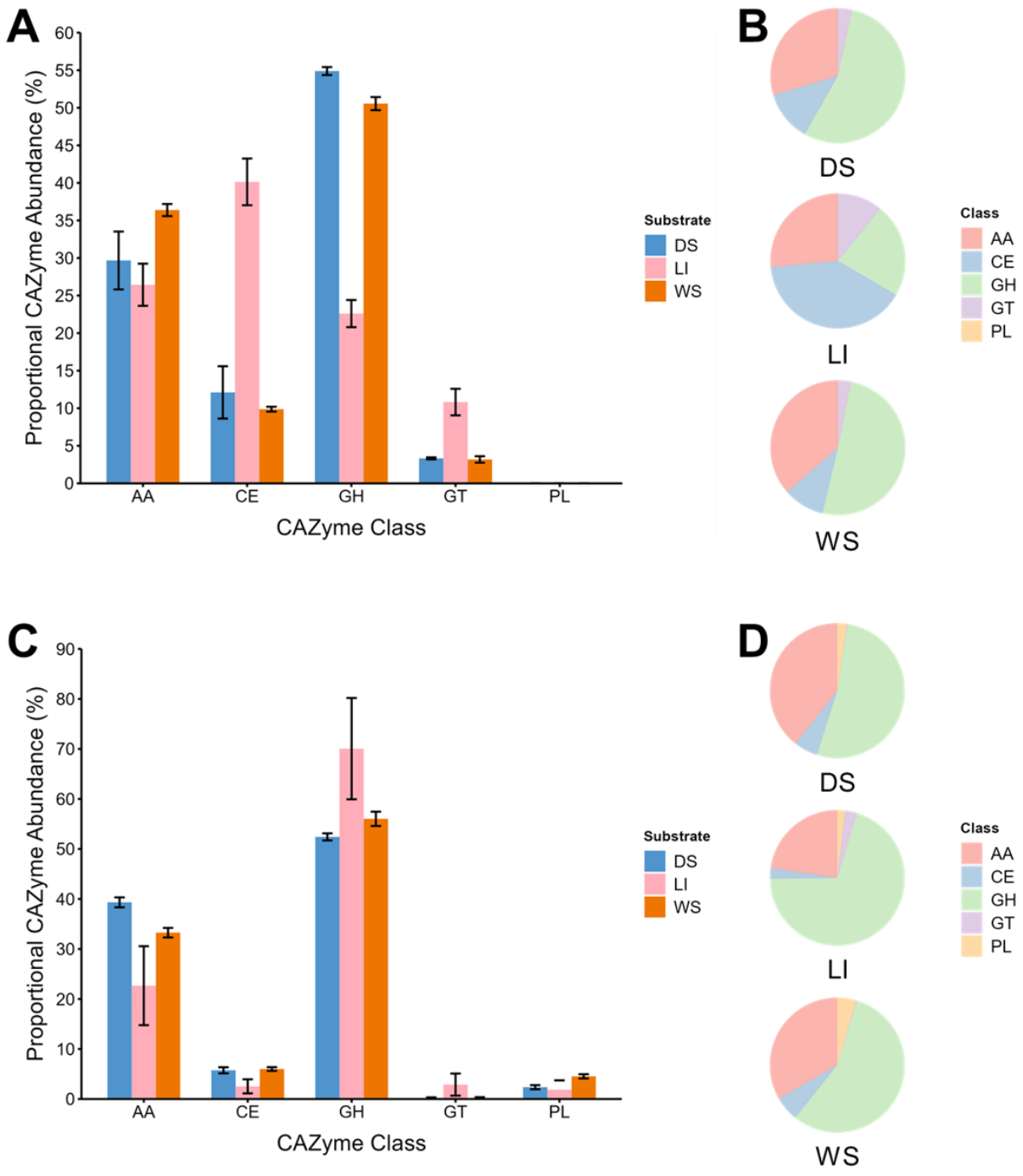


Figure 5.5 Differences in proportional CAZyme class abundance of the *P. putredinis* NO1 secretome across lignocellulosic substrates. The molar percentage abundance of proteins belonging to each class of CAZyme were calculated proportionally to the total abundance of CAZymes for each substrate and compared across substrates for the bound (**A**) and supernatant fractions (**C**) of the *P. putredinis* NO1 secretome. The proportions of each CAZyme class for each substrate were also visualised for the bound (**B**) and supernatant (**D**) fractions separately. DS = Delignified straw, WS = Wheat straw, LI = Dioxane lignin.

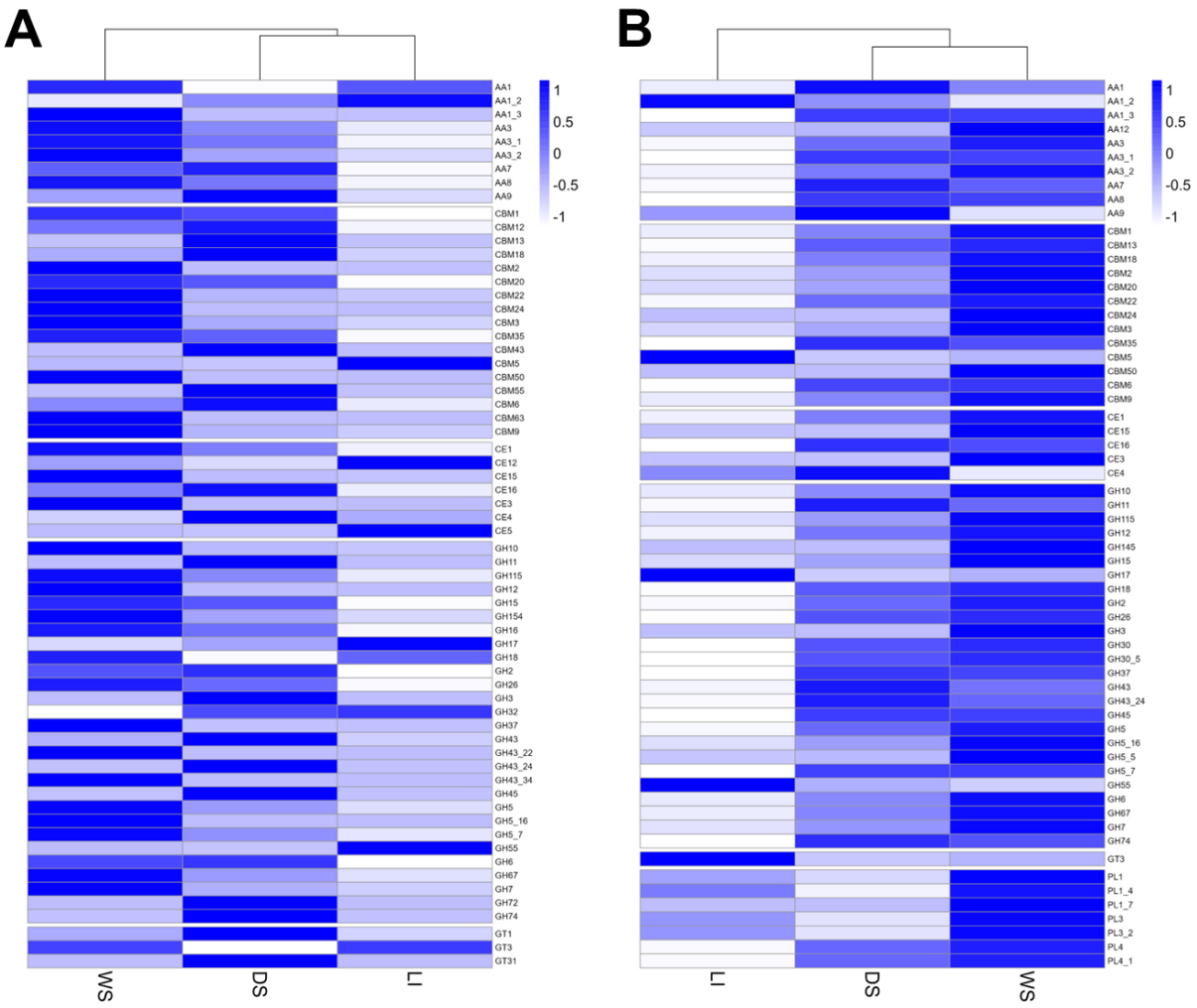


Figure 5.6 Differences in CAZyme family abundance of the *P. putredinis* NO1 secretome across lignocellulosic substrates. Molar percentage values for proteins annotated as CAZymes and identified on at least one substrate of the *P. putredinis* NO1 secretome scaled to Z-scores across substrates for the bound (**A**) and supernatant (**B**) fractions separately. DS = Delignified straw, WS = Wheat straw, LI = Dioxane lignin.

5.4.4. The *P. putredinis* NO1 secretome is a resource for new lignin degrading enzyme identification

Within the *P. putredinis* NO1 secretome, it was hypothesised that lignin degrading enzymes could be identified within two subsets of proteins. Firstly, proteins which are more abundant on the dioxane lignin substrate compared to the delignified straw and wheat straw substrates. Secondly, proteins which are more abundant on wheat straw compared to delignified straw. Through statistical testing by pairwise comparisons between substrates with multiple test correction, 85 proteins were identified which fell into these categories. Investigating these proteins broadly at the functional level reveals that 49 of these proteins

were annotated as CAZymes. The remaining 36 proteins were not annotated as CAZymes and therefore were not clearly implicated from functional annotation alone as involved in lignocellulose breakdown. However, the presence of these proteins in the proteomic data when *P. putredinis* NO1 is grown on lignocellulosic substrates, the isolation of these proteins as part of the *P. putredinis* NO1 secretome, and the higher relative abundances of these proteins on substrates containing lignin were all factors that made these proteins interesting candidates to investigate further.

Within this subset of potential "lignin degrading proteins" was a protein with sequence similarity to chloroperoxidases and aromatic peroxygenases. It has a conserved secretion signal peptide predicted by all three tools of the secretion branch of the secretome isolation bioinformatic workflow and was predicted to be extracellular by both tools of the localisation branch of the workflow (318). The AA2 CAZyme family consists of heme-containing peroxidase enzymes implicated in the breakdown of lignin through oxidative reactions utilising hydrogen peroxide (H_2O_2) and organic peroxides as electron acceptors (179). Chloroperoxidase enzymes however are not classified as CAZymes. Instead, these are heme- or vanadium-containing enzymes which have been demonstrated to display a range of *in vitro* activities including halogenase, peroxidase, catalase, and cytochrome P450 activity (248). These enzymes have been implicated in lignocellulose breakdown previously due to the presence of chloroperoxidase encoding sequences in the genomes of wood-degrading fungi and the detection of lignocellulose chlorination in leaf litter containing soils (249). Aromatic lignin structures in aspen wood were demonstrated to be chlorinated when cultivated by the ascomycete wood-degrader *Curvularia inaequalis* (249). It has been suggested that this occurs through the oxidation of chloride ions to produce chlorine electrophiles which subsequently react with aromatic structures (249).

The structure of this chloroperoxidase-like protein was predicted with AlphaFold and investigated (**Figure 5.7**) (257). The predominantly helical structure of the protein is similar to other peroxidase and cytochrome P450 enzymes (356). There is a clear N-terminal signal peptide extending away from the main structure. Structural comparisons to the protein data bank (PDB) database revealed structural similarity to heme-containing peroxidases and chloroperoxidases (274). The structure was compared to the crystal structure of a chloroperoxidase from *Caldariomyces fumago* (2CPO) and clear similarity was observed between the structures (**Figure 5.7B**). One unique feature of chloroperoxidases compared to other peroxidases is the utilisation of a distal glutamic acid residue, instead of a traditional histidine, as the acid-base catalytic group (356). This glutamic acid residue is visible in the *P. putredinis* NO1 chloroperoxidase-like protein located close to an N-terminal cysteine residue that serves to coordinate iron (**Figure 5.7C**). The coordination of the heme group in relation

to these two important residues can be seen for the *C. fumago* chloroperoxidase crystal structure in comparison with the *P. putredinis* NO1 chloroperoxidase-like protein in **Figure 5.7D**.

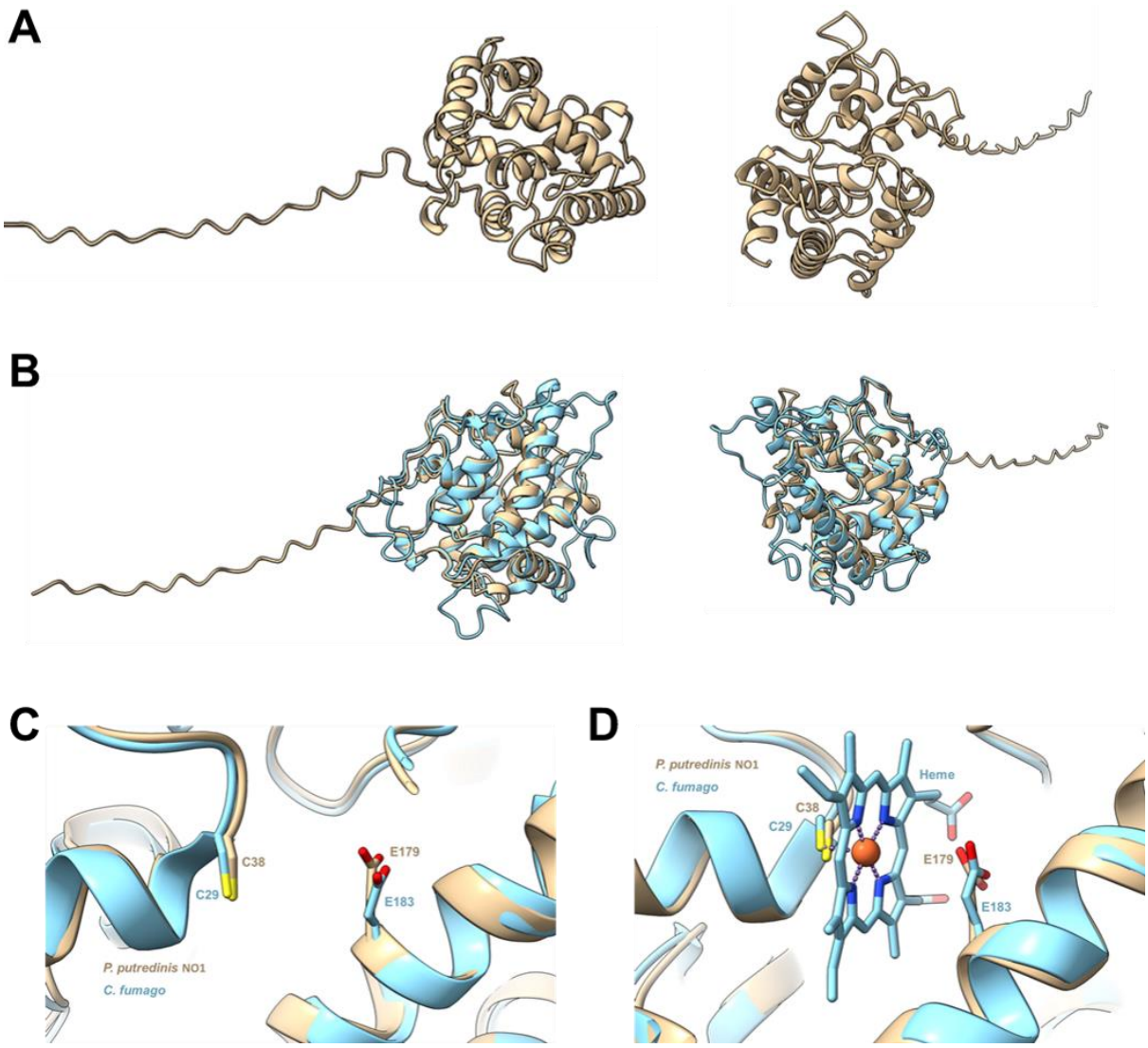


Figure 5.7 Investigation of the structure of the *P. putredinis* NO1 chloroperoxidase-like protein. The structure of the chloroperoxidase-like protein from the *P. putredinis* NO1 secretome was predicted (Beige) (**A**) and compared to the crystal structure of a chloroperoxidase (Blue) from *C. fumago* (PDB ID: 2CPO) (**B**). The heme coordination cysteine and active site glutamic acid residues (**C**), and the coordination of the heme group in relation to these residues (**D**) was investigated.

To investigate the chloroperoxidase-like protein identified in the *P. putredinis* NO1 secretome, the coding sequence was codon optimised, synthesised, and expressed heterologously in *Pichia pastoris*. The recombinant protein was purified to isolate a 30 kDa protein, around 5 kDa higher than the predicted molecular weight of the protein which suggests post-translational modification (**Supplementary Figure 5.7**) (357). The protein was isolated to high-purity and haloperoxidase activity was assayed using an o-dianisidine gel

staining assay reported previously to detect active chloroperoxidase proteins (249). Samples of bovine serum albumin (BSA), the *P. putredinis* NO1 chloroperoxidase-like protein (*Pp*CPO), and a commercial chloroperoxidase from *C. fumago* (*Cf*CPO) were prepared for SDS-PAGE gels in denaturing and non-denaturing conditions. Samples in denaturing conditions contained 1% β-mercaptoethanol and were boiled for 5 minutes before resolution on duplicate SDS-PAGE gels. One gel was Coomassie stained to confirm no major differences in sample resolution between denaturing and non-denaturing conditions (**Figure 5.8A**). The second gel was stained in a solution containing o-dianisidine in the presence of potassium bromide and H_2O_2. Both *Pp*CPO and *Cf*CPO catalysed the transfer of bromide ions to o-dianisidine to form a stain at the correct molecular height of the protein (**Figure 5.8B**), indicating haloperoxidase activity. This activity was lost under denaturing conditions suggesting the purified protein was initially active.

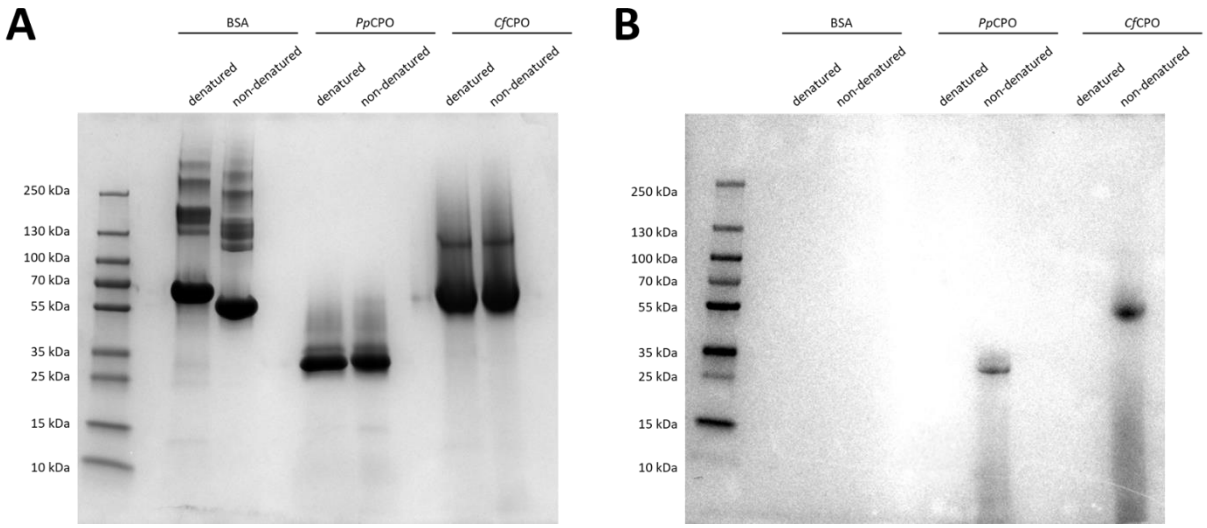


Figure 5.8 Investigating the haloperoxidase activity of the recombinant *Pp*CPO protein. Duplicate SDS-PAGE gels were resolved for denatured and non-denatured samples of BSA, *Pp*CPO, and *Cf*CPO. Gels were either stained in Coomassie (**A**), or in o-dianisidine in the presence of potassium bromide and H_2O_2 (**B**).

The role of chloroperoxidases in lignin breakdown has been investigated previously through the use of model lignin compounds (250). Chlorination of lignin model compounds and cleavage of bonds within a synthetic guaiacyl lignin polymer have been demonstrated previously for heme- and vanadium-containing peroxidases, with the suggestion made that cleavage of ether bonds occurred (250). Arylglycerol β-*O*-4 ether linkages are the most predominant linkages in lignin accounting for 45 – 60 % of the linkages (235). To investigate the ether cleaving ability of the recombinant *Pp*CPO enzyme the protein was incubated with a synthetic substrate mimicking the β-*O*-4 linkage in lignin, which when cleaved releases the fluorescent substrate 4-Methylumbelliferone (4MU) (**Figure 5.9A**). Reactions were

performed at pH 3.0 and in the presence of 20 mM KCl and 20 mM H_2O_2 as reported previously for the investigation of heme-containing chloroperoxidases (250). Reactions were initiated by the addition of H_2O_2 and measurements of 4MU fluorescence were taken immediately and after 16 hours of incubation at 30 °C with shaking at 600 rpm, the difference between 4MU fluorescence was then used to calculate the nanomoles of 4MU released (**Figure 5.9B**). *Pp*CPO was found to release significantly more 4MU than boiled *Pp*CPO and buffer only controls ($F = 14.44$, $d.f. = 2$, $p < 0.05$). The *Cf*CPO was omitted from this study as the dark colour of the enzyme suspension interfered with measurements of 4MU fluorescence. The increased release of 4MU by *Pp*CPO suggested the potential for this enzyme, and chloroperoxidases in general, to play a biological role in lignin degradation or at least modification, something which has been suggested previously (249, 250).

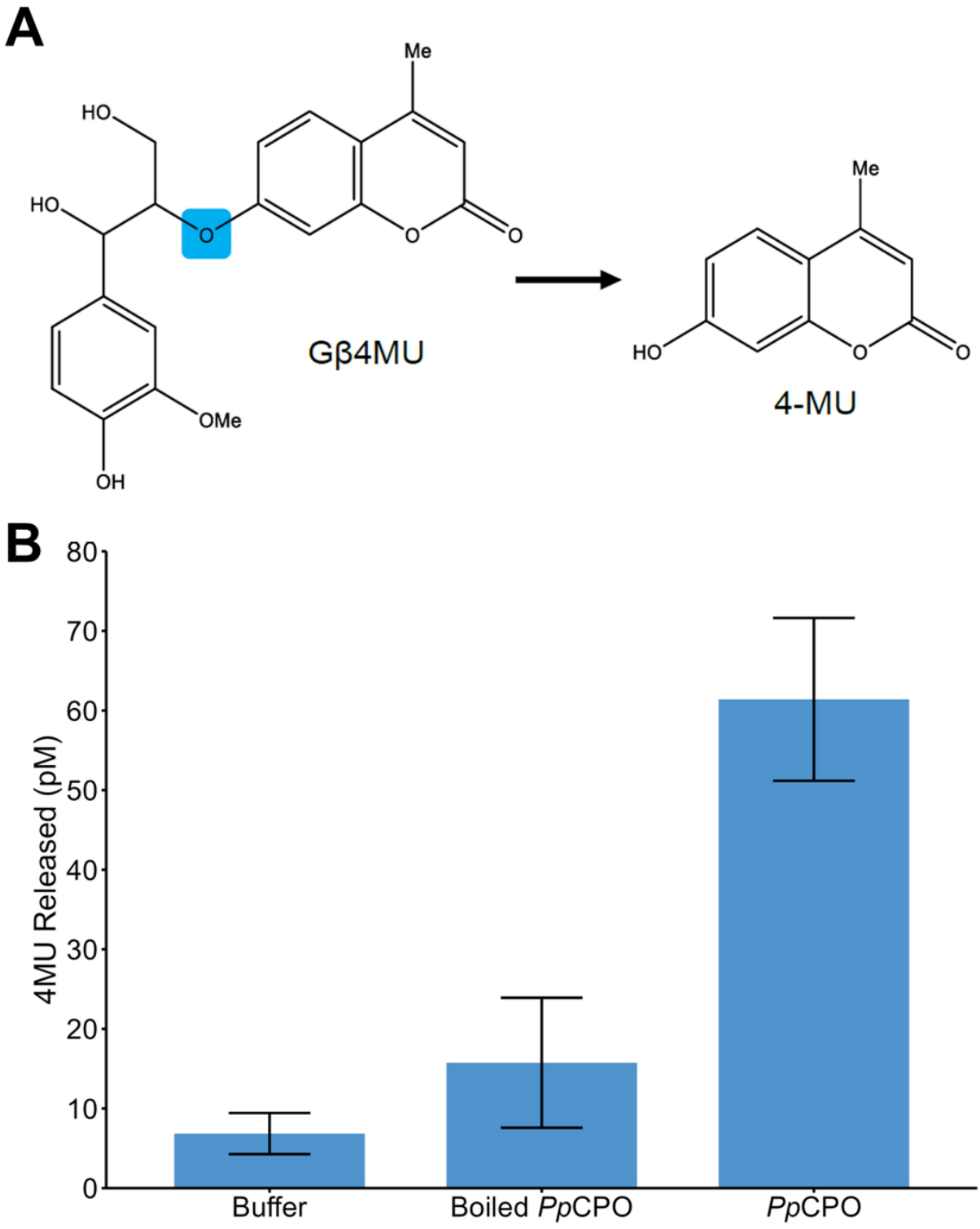


Figure 5.9 Investigating the activity of *Pp*CPO towards a synthetic lignin β-O-4 linkage. The synthetic VT221 substrate (**A**) was incubated in triplicate with 25 mM potassium phosphate buffer pH3.0, 20 mM KCl, 20 mM H_2O_2 either alone (Buffer), with 50 µg of boiled *Pp*CPO, or 50 µg *Pp*CPO. Fluorescence of released 4MU was measured immediately after the addition of H_2O_2 to initiate the reaction and after 16 hours of incubation. The 4MU released in this time was calculated (**B**).

To investigate the potential for lignin modification further, the recombinant *Pp*CPO was incubated in similar conditions (25 mM potassium phosphate buffer pH 3.0, 20 mM H_2O_2, 20 mM KCl), as reported previously for investigations into heme-containing chloroperoxidases (250), in the presence of the same dioxane lignin from wheat straw used in proteomic experiments. Incubations without enzyme, with *Cf*CPO enzyme, or with boiled enzyme controls were also performed. The soluble fraction was then harvested and investigated by HPLC analysis. Standard solutions of caffeic, *p*-coumaric, sinapic, trans-cinnamic, trans-ferulic, and vanillic acid in the same buffer used for assays were analysed alongside samples. Although only qualitative, some differences were observed between the active enzyme incubations and the buffer only or boiled enzyme controls (**Figure 5.10**). Principal component analysis of the chromatogram data for all five conditions demonstrated good grouping of replicates within conditions and separation between conditions, with active *Pp*CPO and *Cf*CPO conditions forming a distinct group of their own (**Supplementary Figure 5.9**). Of particular interest, a small peak with a retention time of just below 14 minutes was observed to be slightly reduced in conditions with active enzymes *Pp*CPO or *Cf*CPO (**Figure 5.10B**). This peak aligns with the same retention time observed for caffeic acid and vanillic acid standards (**Supplementary Figure 5.8**). Soluble fractions for the buffer, *Pp*CPO, and boiled *Pp*CPO conditions were also analysed by gas chromatography-mass spectrometry. Vanillic acid was identified with a retention time of 14.329 minutes in all conditions and again appeared to be reduced in the *Pp*CPO treated condition (**Figure 5.10C**).

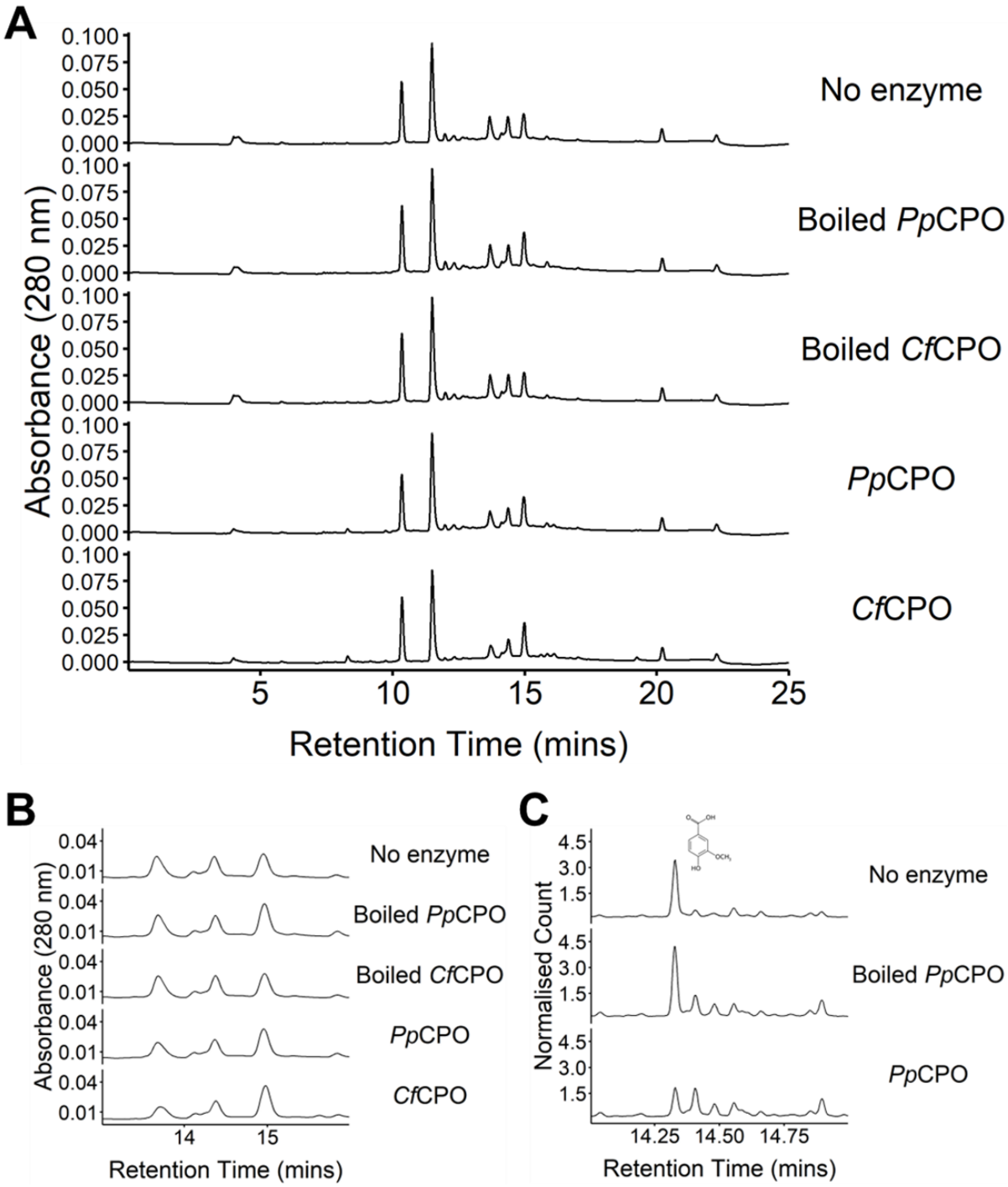


Figure 5.10 Investigating dioxane lignin modification. Dioxane lignin from wheat straw was incubated in triplicate with 25 mM potassium phosphate buffer pH 3.0, 20 mM H_2O_2, and 20 mM KCl alone (No enzyme), or with 100 µg of boiled *Pp*CPO, boiled *Cf*CPO, *Pp*CPO, or *Cf*CPO for 16 hours at 30 °C with shaking at 600 rpm. The soluble fraction of these incubations was subjected to HPLC analysis (**A** and **B**). The soluble fractions for conditions with no enzyme or with incubation with boiled *Pp*CPO or *Pp*CPO were also subjected to GCMS analysis and vanillic acid identified at 14.329 minutes (**C**). Chromatograms are representative of three replicates.

Vanillic acid can be derived from organosolv-obtained lignin and it has been suggested that such processes can produce low molecular weight lignin products (358). The presence of vanillic acid in samples treated only in buffer with no enzyme suggests that either the dioxane lignin contains vanillic acid initially, or that the acidic buffer conditions used in this assay release vanillic acid from the dioxane lignin (**Figure 5.10C**). However, the isolation of vanillic acid from technical lignins, such as organosolv process derived lignins, usually relies on alkaline treatments at temperatures greater than 150 °C (359). Therefore, it is unlikely that the relatively ambient conditions used in this assay liberated vanillic acid from the dioxane lignin, instead it is more likely the vanillic acid was present in the substrate before enzyme treatments. The apparent reduction in soluble vanillic acid in conditions treated with active enzymes *Pp*CPO and *Cf*CPO suggests that the vanillic acid was potentially modified or degraded. In aspen wood decayed for 24 weeks by the chloroperoxidase producing fungus *C. inaequalis* chlorinated, vanillin structures were identified as major chlorinated lignin breakdown products (249). If indeed the CPO enzymes were modifying vanillic acid structures, then appearance of other compounds in both HPLC and GCMS analysis for these conditions may be expected, but this was not observed. However, *Cf*CPO has previously demonstrated specific activity for chlorinated phenolic compounds where the enzyme catalysed the formation of insoluble polymeric materials (248). It could be proposed that the enzymes investigated here chlorinated vanillic acid to form chlorinated phenolic compounds and then oxidatively polymerised these products to insoluble material, which could explain the reduction of vanillic acid in the soluble fraction. To investigate this further, the insoluble material could be depolymerised into fragments, for example via alkaline treatment, to produce a profile of lignin breakdown products which could then be characterised. This could also be used to investigate whether these enzymes are active on polymeric lignin substrates, something which is yet to be achieved. Although the work presented here supports the potential for lignin modification by chloroperoxidases, the role of these enzymes in lignocellulose breakdown remains unclear.

5.5. Conclusions

The previously demonstrated ability of *P. putredinis* NO1 to tailor its secretome depending on growth substrate was reinforced here. Bioinformatic filtering of proteomic data allowed a more effective demonstration of the *P. putredinis* NO1 secretome variability when grown on wheat straw, delignified straw, and dioxane lignin from wheat straw. This variation was investigated at a functional level, and importantly this variation was maintained when looking at the profiles of predicted lignocellulose degrading enzymes. Within the CAZyme family abundance profiles of the different secretomes, expected patterns of abundance for lignin-degrading families of enzymes were observed. For example, the increased abundance of laccase family enzymes on substrates enriched in lignin content. This inspired confidence that other proteins with potential involvement with lignin breakdown, and which are contained within the enzymatic repertoire of *P. putredinis* NO1, may be following similar patterns. To isolate a subset of potentially lignin-degrading enzymes, proteins with significantly higher abundance on lignin enriched substrates were identified. Within these 85 proteins, 36 were not annotated as CAZymes and therefore were not predicted to be implicated in the deconstruction of either lignocellulose or lignin. One such protein was a chloroperoxidase-like protein which was structurally similar to a crystallised chloroperoxidase, and which had the same active site residues. The suggestion of a role for chloroperoxidases in lignocellulose breakdown has been made previously and the ability to modify or cleave model lignin compounds has been demonstrated. However, a clear implication in lignin depolymerisation is yet to be achieved. Here, the recombinant *Pp*CPO enzymes was produced to high purity and demonstrated traditional haloperoxidase activity alongside the ability to cleave β-*O*-4 ether linkages in a model lignin compound. As this is the major structural linkage within lignin, this supports a role for these enzymes in lignin breakdown. To investigate this further, dioxane lignin was incubated with the *Pp*CPO enzyme, and the profile of soluble products characterised. Surprisingly, the emergence of new soluble products was not observed and instead a reduction in a soluble product identified as vanillic acid was seen. Although initially unexpected, it is possible that *Pp*CPO is chlorinating the phenolic compound vanillic acid and then catalysing the oxidative polymerisation into an insoluble product as both activities have been demonstrated previously. Unfortunately, the role of chloroperoxidases in lignin modification or degradation remains unclear and further characterisation of insoluble lignin treated with chloroperoxidases may be required to help elucidate this. However, the work performed here has supported the implication of these enzymes in lignin breakdown through the identification of a new chloroperoxidase from *P. putredinis* NO1 and has added to the limited knowledge of the potential activities which chloroperoxidases may exhibit toward this substrate.

5.6. Materials and methods

5.6.1. Strain isolation

P. putredinis NO1 was isolated from a wheat straw enrichment culture and maintained as reported previously (229).

5.6.2. Delignification of wheat straw biomass

Wheat straw was milled to < 2 mm particles using the Retsch cyclone mill twister. Two flasks were prepared containing 10 g of milled wheat straw and 480 mL of water and incubated in a water bath at 75 °C for 120 mins total. The delignification process was charged by the addition of 3 g of sodium chlorite (80% purity, Merck) and 1.5 mL glacial acetic acid (Merck) to each flask at 0, 40 and 80 mins. After the incubation was complete, flasks were cooled to room temperature and biomass was washed three times in water until the filtrate ran neutral and was then washed once in acetone before air drying.

5.6.3. Biomass content analysis

To quantify lignin content, the acetyl bromide soluble lignin assay was performed in triplicate for both untreated and delignified wheat straw. Triplicate samples of 4 mg of biomass was prepared for each substrate and incubated with 250 µL of acetyl bromide:glacial acetic acid (25% : 75%, v/v) at 50 °C for 2 hours. Samples were then incubated for an additional hour with vortexing every 15 minutes before being combined with 1 mL 2M NaOH and 175 µL 0.5 M hydroxylamine in 5 mL volumetric flasks. Flasks were made up to 5 mL with glacial acetic acid and then the absorbance of a 1 in 10 dilution in glacial acetic acid was measured in quartz cuvettes at 280 nm using glacial acetic acid as the blank. The lignin content was then calculated using the following formula and with a coefficient of 17.75:

$$\frac{A280}{coefficient * pathlength} * \frac{total\ volume * 100}{sample\ weight}$$

To quantify the matrix polysaccharide content, triplicate samples of 4 mg of biomass was prepared for both untreated and delignified wheat straw alongside dried sugar standards containing fucose, arabinose, rhamnose, galactose, glucose, xylose, mannose, galacturonic acid and glucuronic acid. Each sample and standard were hydrolysed by adding 500 µL of 2 M trifluoroacetic acid (TFA) followed by flushing tubes with argon to prevent oxidation of matrix polysaccharides components. Tubes were then incubated at 100 °C for 4 hours and then cooled to room temperature before evaporating TFA. Samples were then resuspended in 200 µL of dH_2O and centrifuged at 12,700 rpm for 10 minutes to pellet the cellulose. The matrix monosaccharide containing supernatant was transferred to HPLC vials through 0.45 µM PTFE filters for analysis on the Dionex™ ICS-3000.

The remaining cellulose pellets were then washed once in 1.5 mL of dH_2O and three times in 1.5 mL of acetone and then air dried overnight. Dried cellulose pellets were then hydrolysed to glucose by adding 90 µL of 72% (v/v) sulphuric acid to each pellet and incubating at room temperature for 4 hours before adding 1890 µL of water to dilute the sulphuric acid to 3.2% (v/v) and incubating at 120 °C for 4 hours. Samples were then centrifuged at 12,700 rpm for 10 minutes and the glucose concentration of the supernatant measured using the anthrone assay. Cellulose content was calculated from the glucose concentration in samples determined from a standard curve of glucose standards.

Finally, ash content was determined by incinerating triplicate samples of 0.5 g of untreated and delignified wheat straw at 600 °C for 24 hours. Samples were weighed before and immediately after incineration to calculate the % ash content.

5.6.4. Fourier transformed infrared (FTIR) spectroscopy

Normal and delignified wheat straw was further milled to <0.5 mm particle size in the Retsch cyclone mill twister before loading five replicate samples onto a Perkin Elmer FTIR spectrophotometer with the following conditions: Abscissa units = Wavenumber, Ordinate units = Absorbance (A), Start = 4000 cm^{-1}, End = 700 cm^{-1}, Resolution = 4 cm^{-1}, Accumulations = 4. Absorbance values were normalised based on maximum absorbance value within each replicate. Outliers were removed based on principal component analysis of absorbance data was performed in R studio 4.2.3 using the 'FactoMineR', 'factoextra' packages (293, 334, 335). Average absorbances for normal and delignified wheat straw were plotted for comparison of biomass structural bond content in R studio 4.2.3 using the 'ggplot2' package (294).

5.6.5. Scanning electron microscopy (SEM) of wheat straw substrates

SEM of normal and delignified wheat straw was performed at 400x, 750x, and 2000x magnifications by the Microscopy department of the Technology Facility at the University of York.

5.6.6. Obtaining dioxane mild acidolysis lignin

Wheat straw was milled to < 2 mm in a Retsch cyclone mill twister and was washed in hexane to remove extractives before air drying. Dioxane lignins were isolated by the Biorenewables Development Centre (BDC) as reported previously (360). Per 10 g of milled and washed wheat straw, 120 mL of dioxane:water mixture (9:1, v/v) was added containing 200 mM hydrochloric acid. The suspension was refluxed under nitrogen gas bubbling and magnetic stirring for 30 minutes. The cooled mixture was then vacuum-filtered through a porcelain Büchner funnel lined with filter paper. The residue was washed three times with 50

mL of dioxane and water mixture. Filtrates were pooled and the pH was adjusted to 3.0 to 4.0 using a saturated $NaHCO_3$ aqueous solution. The solutions were then concentrated to approximately 50 mL by rotary evaporation at 45 °C. Isolated lignins were precipitated by combining the concentrated solution with 500 mL of cold water and the precipitant was subsequently recovered by centrifugation at 1200 x g for 30 minutes at 4 °C. The precipitant was washed through the addition of 100 mL pure water and repeating the centrifugation. The final precipitant was then dried by lyophilization.

5.6.7. *P. putredinis* NO1 cultures for proteomics

Triplicate 500 mL solutions of optimised media were prepared containing 1.5 % (w/v) wheat straw, delignified wheat straw, or dioxane lignin extracted from wheat straw. Chloramphenicol was added to a final concentration of 25 µg/mL and cultures were inoculated with 2.5 mL of mycelial inoculum from a seed culture of *P. putredinis* NO1 grown from 10^5 spores/mL. These cultures were incubated at 30 °C with shaking at 150 rpm for 7 days before wet mycelia was harvested for proteomic investigation. The optimised media contained KCl 0.52 g/L, KH2PO4 0.815 g/L, K2HPO4 1.045 g/L, MgSO4 1.35 g/L, NaNO3 1.75 g/L, yeast extract 8.85 g/L and Hutner's trace elements and was based on *Aspergillus niger* media (333).

5.6.8. Harvesting the *P. putredinis* NO1 proteomes across substrates

To harvest supernatant proteins, culture supernatants were centrifuged at 12,000 x g for 20 minutes at 4 °C and then supernatants filter sterilised through 0.22 µm PES filter units. Triplicate 5 mL technical replicates for each culture were combined with 5 volumes of ice cold 100% (v/v) acetone and mixed by inverting before incubation overnight at -20 °C. Samples were then centrifuged at 4500 rpm for 20 minutes at 4 °C and the acetone supernatant discarded. Pellets were then washed twice by the addition of ice cold 80% (v/v) acetone, vortexing and centrifuging at 4500 rpm for 10 minutes at 4 °C. Pellets were air dried and resuspended in 1 mL of 0.5x PBS buffer before transferring to Eppendorfs and snap freezing in liquid nitrogen and storing at -80 °C.

To extract proteins bound to substrates, triplicate samples of 2 g of biomass from each culture were washed twice through the addition of 25 mL ice-cold 0.5x PBS and centrifugation at 4500 rpm for 5 minutes at 4 °C. The supernatant was discarded, and pellets were resuspended in 19 mL of 0.5x PBS with 1 mL of biotin (EZ-link-Sulfo-NHS-SS-biotin, Thermo Scientific) before rotating at 4 °C for 1 hour. Samples were then centrifuged at 4500 rpm for 10 minutes at 4 °C and the supernatant discarded. The reaction was then quenched by the addition of 25 mL of 50 mM Tris-HCl pH 8.0 and rotating at 4 °C for 30 minutes. The biomass was pelleted by centrifugation at 4500 rpm for 10 minutes at 4 °C and the

supernatants discarded. Biomass pellets were washed twice again with 20 mL of 0.5x PBS with centrifugation at 4500 rpm for 10 minutes at 4 °C. Pellets were then resuspended in 10 mL of 2 % (w/v) SDS pre-heated to 60 °C and rotated at room temperature for 1 hour. Samples were centrifuged at 4500 rpm for 10 minutes and the supernatant transferred to fresh tubes. Five volumes of ice-cold 100% (v/v) acetone was added, and samples were incubated at -20 °C overnight. Samples were pelleted and washed in the same way as before for the supernatant fraction but were resuspended after air drying in 1 mL of 0.1% SDS solution before filtering through 0.22 µm PES filters.

Each sample was then loaded onto its own HiTrap Streptavidin HP 1 mL column (GE Healthcare) at a flow rate of 0.5 mL/min. Columns were left to incubate for 1 hour at 4 °C and were then washed with 10 mL of 0.1% (w/v) SDS in PBS solution at a flow rate of 1 mL/min. Proteins were eluted by loading 1 mL of 50 mM DTT in PBS solution and incubating columns overnight at 4 °C. Another 1 mL of 50 mM DTT/PBS was added, and 1 mL of protein eluted, the column was incubated for 1 hour at 4 °C and then another 1 mL 50 mM DTT/PBS was added and another 1 mL of protein collected.

Both supernatant and bound fraction proteins were then desalted by spinning the samples through 5 mL Zeba™ spin columns 7k MWCO. Desalted samples were then freeze dried overnight and resuspended in NuPAGE loading buffer before being loaded onto NuPAGE gels (Invitrogen). Gels were run for 6 minutes at 180 V.

Due to low protein concentration for dioxane lignin replicates, four additional lignin cultures were repeated to give seven replicates in total.

5.6.9. Peptide identification by LC-MS/MS

Protein samples in gels were then prepared for and subjected to label-free LC-MS by the Metabolomics and Proteomics department in the Technology Facility at the University of York.

Peptides were loaded onto an mClass nanoflow UPLC system (Waters) equipped with a nanoEaze M/Z Symmetry 100 Å C 18, 5 µm trap column (180 µm x 20 mm, Waters) and a PepMap, 2 µm, 100 Å, C 18 EasyNano nanocapillary column (75 mm x 500 mm, Thermo). The trap wash solvent was aqueous 0.05 % (v/v) trifluoroacetic acid and the trapping flow rate was 15 µL/min. The trap was washed for 5 min before switching flow to the capillary column. Separation used gradient elution of two solvents: solvent A, aqueous 0.1 % (v/v) formic acid; solvent B, acetonitrile containing 0.1% (v/v) formic acid. The flow rate for the capillary column was 300 nL/min and the column temperature was 40°C. The linear multi-step gradient profile was: 3-10 % B over 8 mins, 10-35 % B over 115 mins, 35-99 % B over

30 mins and then proceeded to wash with 99 % solvent B for 4 min. The column was returned to initial conditions and re-equilibrated for 15 min before subsequent injections.

The nanoLC system was interfaced with an Orbitrap Fusion Tribrid mass spectrometer (Thermo) with an EasyNano ionisation source (Thermo). Positive ESI-MS and MS 2 spectra were acquired using Xcalibur software (version 4.0, Thermo). Instrument source settings were: ion spray voltage, 1900-2100 V; sweep gas, 0 Arb; ion transfer tube temperature; 275 °C. MS 1 spectra were acquired in the Orbitrap with: 120,000 resolution, scan range: m/z 375-1,500; AGC target, 4e5; max fill time, 100 ms. Data dependent acquisition was performed in top speed mode using a 1 s cycle, selecting the most intense precursors with charge states >1. Easy-IC was used for internal calibration. Dynamic exclusion was performed for 50 s post precursor selection and a minimum threshold for fragmentation was set at 5e3. MS 2 spectra were acquired in the linear ion trap with: scan rate, turbo; quadrupole isolation, 1.6 m/z; activation type, HCD; activation energy: 32 %; AGC target, 5e3; first mass, 110 m/z; max fill time, 100 ms. Acquisitions were arranged by Xcalibur to inject ions for all available parallelizable time.

Peak picking, database searching and quantification of proteomic data from Thermo raw files was performed using FragPipe (v19.1). Data were searched against a custom database of all coding regions of the *P. putredinis* NO1 genome appended with common contaminants and reversed sequences. The default LFQ-MBR workflow was used with the following modifications: precursor mass tolerance = +/- 3 ppm; fragment mass tolerance = 0.5 Da; IonQuant, feature detection m/z tolerance = 3 ppm; MBR RT tolerance = 7.5 min; add MaxLFQ, MBR FDR = 0.01, MBR min ions = 2. Final protein-level data were filtered to 1 % FDR, a minimum protein probability of 0.99 and a minimum of 2 peptides. Molar percentage values were calculated for each protein.

5.6.10. Quality control of proteomic data

Identified protein count, principal component analysis, and hierarchical clustering with a Canberra distance matrix and wardD2 clustering were used to investigate proteomic data for all biological replicates for all substrates investigated. This analysis was performed in R studio 4.2.3 using the ‘ggplot2’, ‘FactoMineR’, ‘factoextra’, ‘ggdendro’, and ‘dendextend’ packages (293, 294, 334-337). From this analysis, four dioxane lignin replicates were identified as outliers and removed from the proteomic dataset for final analysis to give three replicates for all substrates in the final proteomic dataset.

5.6.11. Isolating the *P. putredinis* NO1 secretome

Proteins were filtered to produce the *P. putredinis* NO1 secretome using the ‘strict’ filtering workflow developed previously (318). Secretome proteins were predicted to be extracellular

by both BUSCA and DeepLoc localisation prediction tools (309, 310), or which were predicted to encode a secretion signal by SignalP, TargetP and SecretomeP tools (275, 311, 312), and which lacked more than one predicted transmembrane domain by TMHMM or contained a single predicted transmembrane helix with more than 10 amino acids of this helix occurring in the first 60 amino acids of the protein sequence (313).

5.6.12. Comparing the *P. putredinis* NO1 secretome across substrates

All comparative analysis was performed in R studio 4.2.3 (293), and analysis was repeated for the total proteome and the filtered secretome.

The number of proteins identified in at least one replicate across all substrates was compared between the bound and supernatant fraction using the 'ggVennDiagram' package (338). To visualise differences across substrates, heatmaps were created to compare the average molar percentage of proteins across substrates using the 'pheatmap' package (339). Principal component analysis was carried out to investigate replicate grouping within and between substrates using the 'FactoMineR' and 'factoextra' packages (334, 335). Canberra distance matrix calculation with wardD2 clustering was used to plot dendrograms to investigate replicate clustering within and between substrates using the 'ggdendro', and 'dendextend' packages (336, 337). GO category and GO term annotation was performed using Blast2GO in the OmicsBox package (340, 341) and abundances visualised and compared using the 'ggplot2' package (294). CAZyme domain annotation was performed using the dbCAN server (178), and abundances visualised and compared using the 'ggplot2' package (294).

5.6.13. Structural investigation of *Pp*CPO

Structural prediction of the *Pp*CPO genomic coding sequence (CAI7994069.1) was performed using AlphaFold 2.0.0 (257) on the VIKING computer cluster. The predicted structure was structurally compared to crystal structures in the PDB database (274) using the online tool PDBefold with a 'lowest acceptable match' parameter set at 70 % (298). Visualisation of the predicted structure of the *Pp*CPO enzyme and comparisons to the crystal structure of the chloroperoxidase enzyme from *C. fumago* (PDB ID: 2CPO) were performed in ChimeraX 1.5 (361).

5.6.14. Cloning the *Pp*CPO sequence for *P. pastoris* X-33 recombinant expression

The coding region for *Pp*CPO (CAI7994069.1) was codon optimised for *P. pastoris* in Geneious Prime 2022.1.1 and the sequence was synthesised by Twist Biosciences and cloned into pTwist high copy plasmids with chloramphenicol selection markers. Amplification of the coding sequences was performed with the following primers (Forward: 5'–

CATCACCACCACCACGGTGTTGTTGATACTCACCCATGG–3'; Reverse: 5'–GAGTTTTTGTTCTAGAATTCAAGCAGCAGCTCTTACCTTGTC–3') to provide homology for In-Fusion (Takara Bio) cloning into the linearised pPICZaB-NHIS plasmid provided by the Protein Production department of the Technology Facility at the University of York. This expression plasmid was integrated into the genome of *P. pastoris* X-33 for recombinant expression and transformants were plated on yeast extract peptone dextrose (YPD) agar plates containing 100 µg/mL Zeocin. Transformation plates were incubated at 30 °C for 2-3 days until colonies formed.

5.6.15. Screening for *Pp*CPO expressing *P. pastoris* X-33 colonies

Re-streaks of eight colonies were each used to inoculate 2 mL aliquots of BMGY media containing the following: yeast extract 1 % (w/v), peptone 2 % (w/v), potassium phosphate buffer pH 6.0 100 mM, yeast nitrogen base 1.34 % (w/v), biotin 0.0004 % (w/v), casamino acids 1 % (w/v), glycerol 1 % (v/v). These pre-cultures were incubated at 30 °C with shaking at 660 rpm overnight and the optical density (OD) at 600 nm was confirmed to be between 10 and 20. Then, 100 µL of pre-cultures were used to inoculate 2 mL of BMMY media containing yeast extract 1 % (w/v), peptone 2 % (w/v), potassium phosphate buffer pH 6.0 100 mM, yeast nitrogen base 1.34 % (w/v), biotin 0.0004 % (w/v), casamino acids 1 % (w/v), methanol 0.5 % (v/v). The cultures were incubated at 30 °C with shaking at 660 rpm for 72 hours and were fed with BMMY with 10 % (v/v) methanol every 24 hours, with samples taken at t = 0, 24, 48, and 72 hours. The supernatant of these samples was investigated by western blotting with anti-His antibody at a concentration of 1 in 5000 to confirm the expression of the recombinant *Pp*CPO enzyme. The best expressing colony was selected for scale-up of expression.

5.6.16. Expression and purification of *Pp*CPO

The re-streak of the selected *Pp*CPO expressing colony was used to inoculate four 7.5 mL pre-cultures of BMGY and were incubated at 30 °C with shaking at 220 rpm for 24 hours, after which the OD was confirmed to be between 10 and 20. These pre-cultures were pooled and used to inoculate 600 mL BMMY which was incubated for 72 hours at 30 °C with shaking at 180 rpm and was fed with BMMY containing 10 % (v/v) methanol every 24 hours. Samples were taken of pre-cultures and of the expression culture at t = 24, 48, and 72 (end). The supernatant was harvested by centrifugation of the culture at 20,000 x g for 20 minutes at 4 °C. The pH was adjusted to around 7.5 and a sample was taken (sup), before purification over a 5 mL HisTrap FF column (Cytiva) equilibrated in binding buffer of 50 mM Tris-HCl, 100 mM NaCl, pH 7.5 at a flow rate of 5 mL/min. The column was washed with 10 CVs of binding buffer. Samples of flowthrough (FT) and wash (W) were taken. The

recombinant protein was eluted by switching to an elution buffer of 50 mM Tris-HCl, 100 mM NaCl, 500 mM imidazole pH 7.5 and collecting eluted protein. The eluted protein was pooled and desalted into binding buffer through a 10 mL 7k MWCO Zeba™ Spin desalting column (Thermo Fisher Scientific) before concentration through a Biomax 10 kDa ultrafiltration membrane (Merck) to around 5 mg/mL. Samples throughout the expression and purification process were prepared for duplicate SDS-PAGE gels, for Coomassie staining and western blotting with anti-His antibody at a concentration of 1 in 5000.

5.6.17. Gel-staining assay for haloperoxidase activity

Duplicate samples of 40 µg of bovine serum albumin (BSA) (Thermo Fisher Scientific), recombinant *Pp*CPO, and of a commercial chloroperoxidase form *C. fumago* (Sigma 25810) (*Cf*CPO) in binding buffer were combined with 4x loading dye (362). β-mercaptoethanol was added to a final concentration of 1 % (v/v) to a single sample for each protein and these were boiled at 100 °C for 5 minutes. The second sample for each protein condition was left non-denatured. Both samples were run on duplicate SDS-PAGE gels to resolve protein bands. One gel was Coomassie stained, and the second gel was stained in a solution of 1 mM o-dianisidine, 100 mM KBr, 100 mM KH_2PO_4, 1 mM H_2O_2 overnight at room temperature.

5.6.18. Assaying for β-O-4 linkage cleavage activity

Triplicate cultures of 500 µL total volume were prepared containing 25 mM potassium phosphate buffer pH 3.0, 20 mM KCl, 20 mM H_2O_2, and 100 µM VT221 substrate (**Figure 5.9A**) with either nothing, 50 µg of *Pp*CPO, or 50 µg boiled *Pp*CPO. H_2O_2 was added to initiate the reaction and 100 µL was immediately (t = 0) combined with 50 µL of 200 mM glycine-NaOH pH 10.0 in a black 96-well plate. Standards of 0 – 20 µM 4MU in the same buffer as samples were also prepared in the plate in the same way. 4MU fluorescence was measured using an excitation of 360-320 nm and measuring emission at 450-430 nm. The samples were incubated at 30 °C with shaking at 600 rpm for around 16 hours and released 4MU fluorescence was measured in the same way. Standards were used to calculate the concentration of free 4MU at each time point and the 4MU released over 16 hours was calculated by subtracting the t = 0 time point from the t = 16 time point.

5.6.19. Investigating lignin modification by *Pp*CPO

Triplicate reactions containing 15 mg of dioxane lignin were prepared to 300 µL total volumes containing 25 mM potassium phosphate buffer pH 3.0, 20 mM KCl, 20 mM H_2O_2 and with the addition of either nothing (no enzyme), 150 µg *Pp*CPO, 150 µg boiled *Pp*CPO, 150 µg *Cf*CPO, or 150 µg boiled *Cf*CPO. Reactions were incubated at 30 °C with shaking at 600 rpm for 16 hours. The insoluble material was pelleted and the soluble fractions

harvested by centrifugation at 12,000 x g for 10 minutes at room temperature. Samples were analysed by HPLC (2695 Separations Module and 2996 Photodiode Array Detector; Waters, Milford, MA, USA) with a Waters X-Bridge C18 column (300 mm x 4.5 mm, 5 µM). The mobile phases for the gradient conditions were mobile phase A, 0.1 % (v/v) formic acid; mobile phase B, 100 % (v/v) methanol. The gradient ran: 0 min 100 % A 0 % B, 6 min 100 % A 0 % B, 16 min 0 % A 100 % B, 22 min 100 % A 0 % B, 25 min 100 % A 0 % B. Chromatogram data at 280 nm was analysed in R 4.2.0 (293).

For GCMS, 100 µL of no enzyme, *Pp*CPO, and boiled *Pp*CPO samples were dried by rotary evaporation for around 2 hours and samples were resuspended in 50 µL MSTFA + 1 % TMCS (Thermo Fisher Scientific) and 50 µL pyridine (Thermo Fisher Scientific). Samples were subjected to GCMS using the Agilent GC-MSD instrument with ultra-inert liner and the Agilent J&W HP-5MS column (30m, 0.250 mm diameter, 0.25 µm film). Inlet temperature was maintained at 280 °C and the transfer line at 290 °C with a flow rate of 1.2 mL/min with runs held at 90 °C for 3 minutes and then increased to 290 °C at 12 °C/min with a 10-minute final hold. Injection volumes of 1 µL were used. The resulting data was integrated in the Agilent Mass Hunter Qualitative Analysis software, and the chromatograms plotted in R 4.2.0 (293).

5.7. Declarations

5.7.1. Availability of data and materials

All proteomic data generated during this research are available at MassIVE MSV000092143.

5.7.2. Competing interests

The authors declare that they have no competing interests.

5.7.3. Funding

This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC), UK (Grant BB/1018492/1, BB/P027717/1). CS was supported by a CASE studentship from the BBSRC Doctoral Training Programme (BB/M011151/1) with Prozomix Ltd.

5.7.4. Author contributions

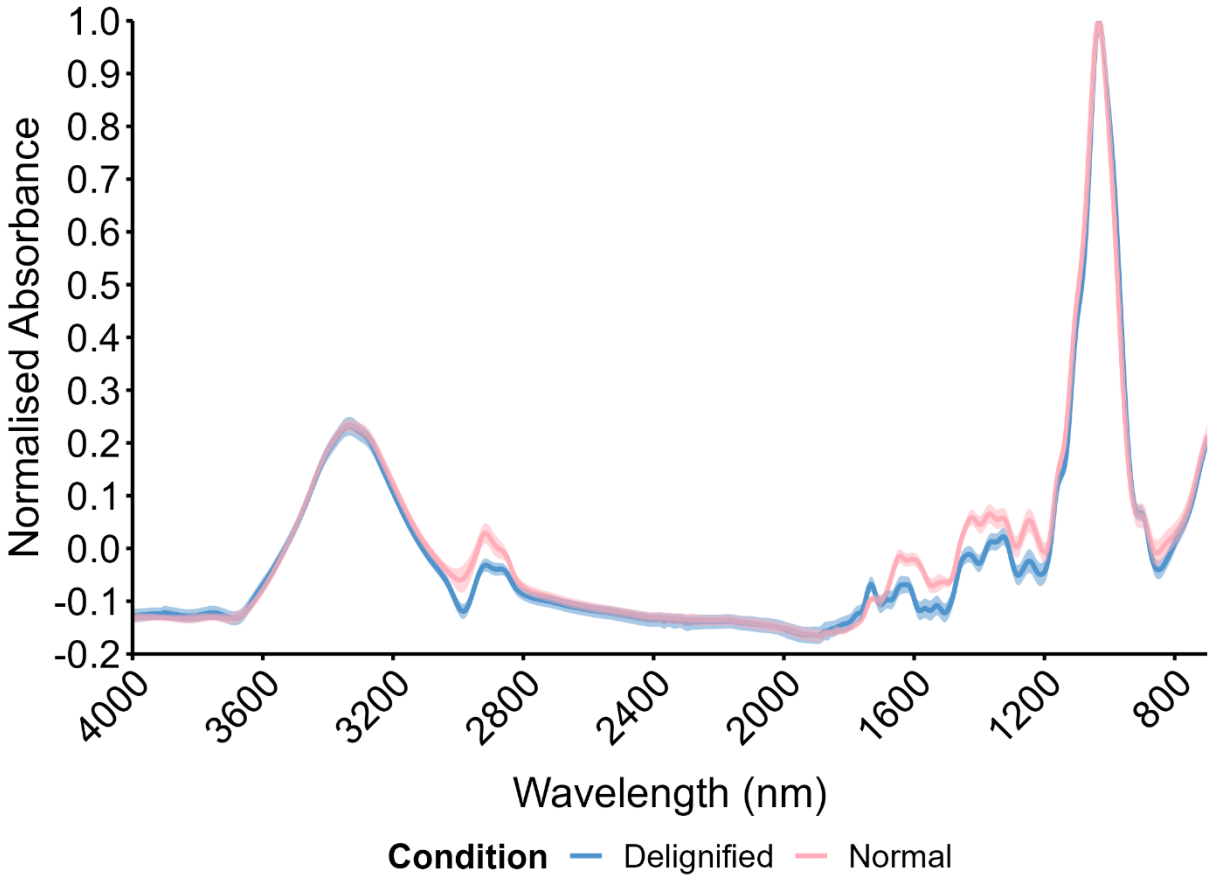
CJRS conceptualised the investigation carried out in this paper, performed the proteomic experiment, performed all proteomic analysis, cloned and expressed the recombinant protein, purified the protein, and performed all characterisation. AD performed the proteomic identification and supported the proteomic analysis. NCB was a major contributor to the conceptualisation of the study, supervised the work, and contributed to the writing of the manuscript.

5.7.5. Acknowledgements

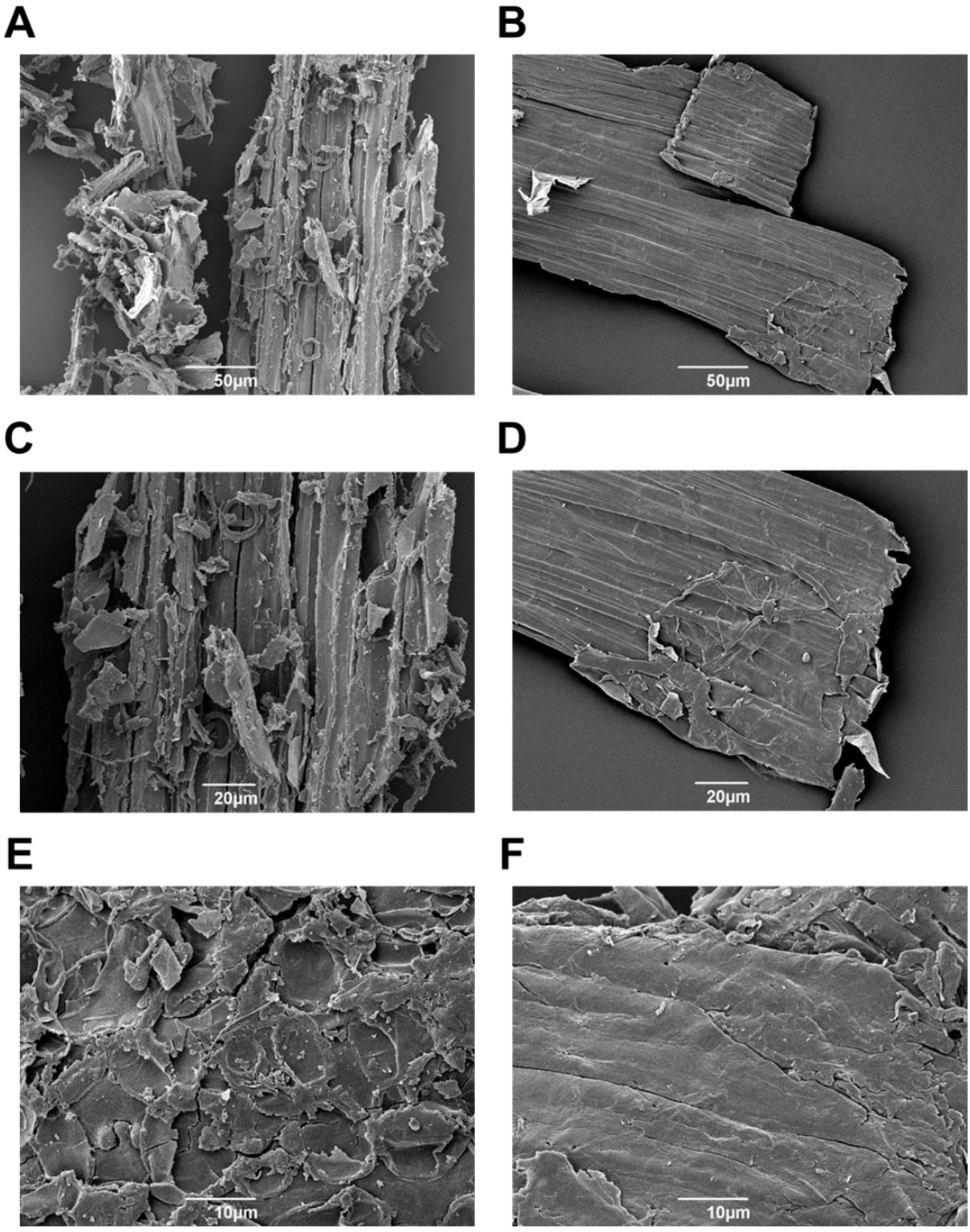
The York Centre of Excellence in Mass Spectrometry was created thanks to a major capital investment through Science City York, supported by Yorkshire Forward with funds from the Northern Way Initiative, and subsequent supports from ESPRC (EP/K039660/1; EP/M028127/1).

5.8. Supplementary material

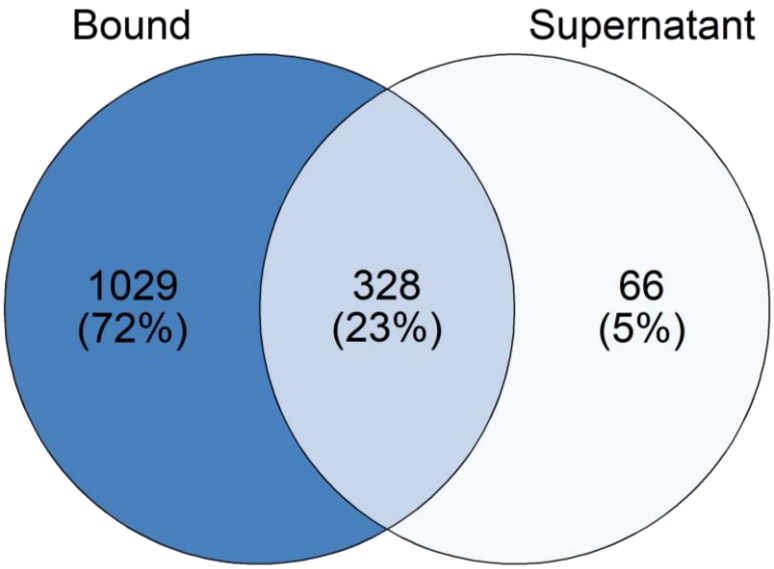
5.8.1. Supplementary figures



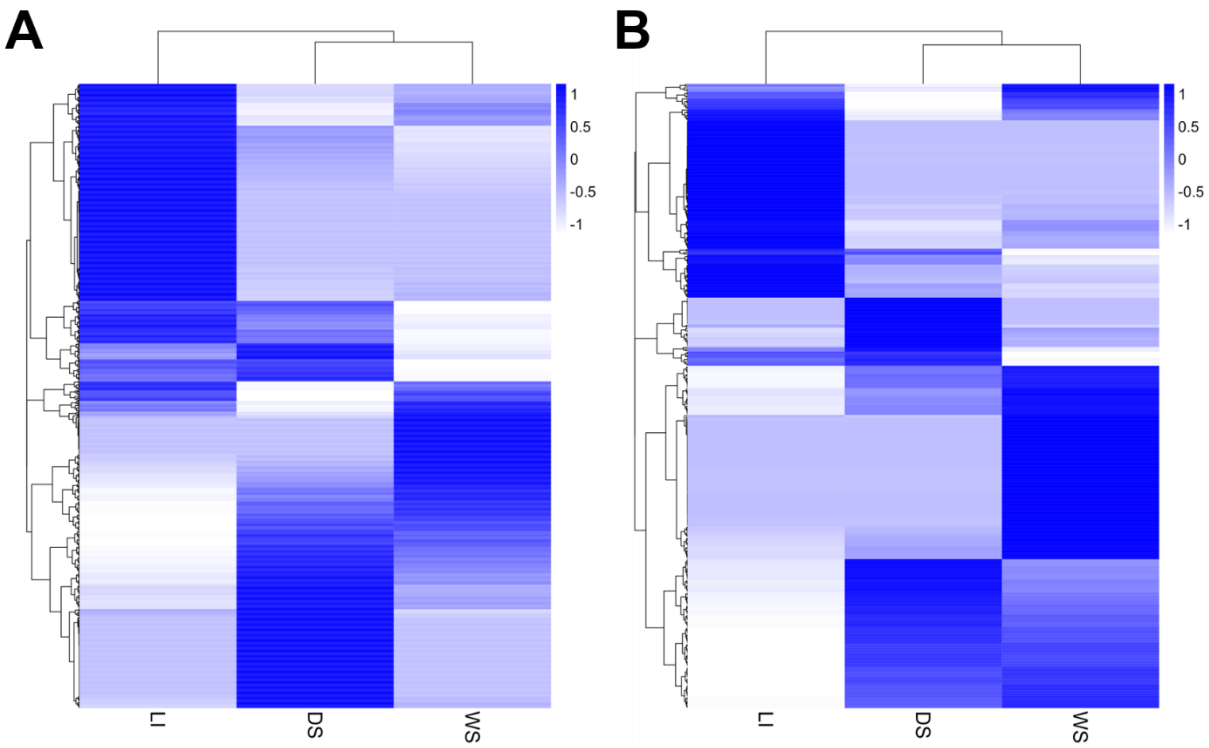
Supplementary Figure 5.1 Fourier transformed infrared spectroscopy of normal and delignified wheat straw. FTIR spectroscopy was used to measure absorbances from 4000 to 700 cm^{-1} for triplicate samples of normal and delignified wheat straw. Absorbance was normalised for each replicate by maximum absorbance value.



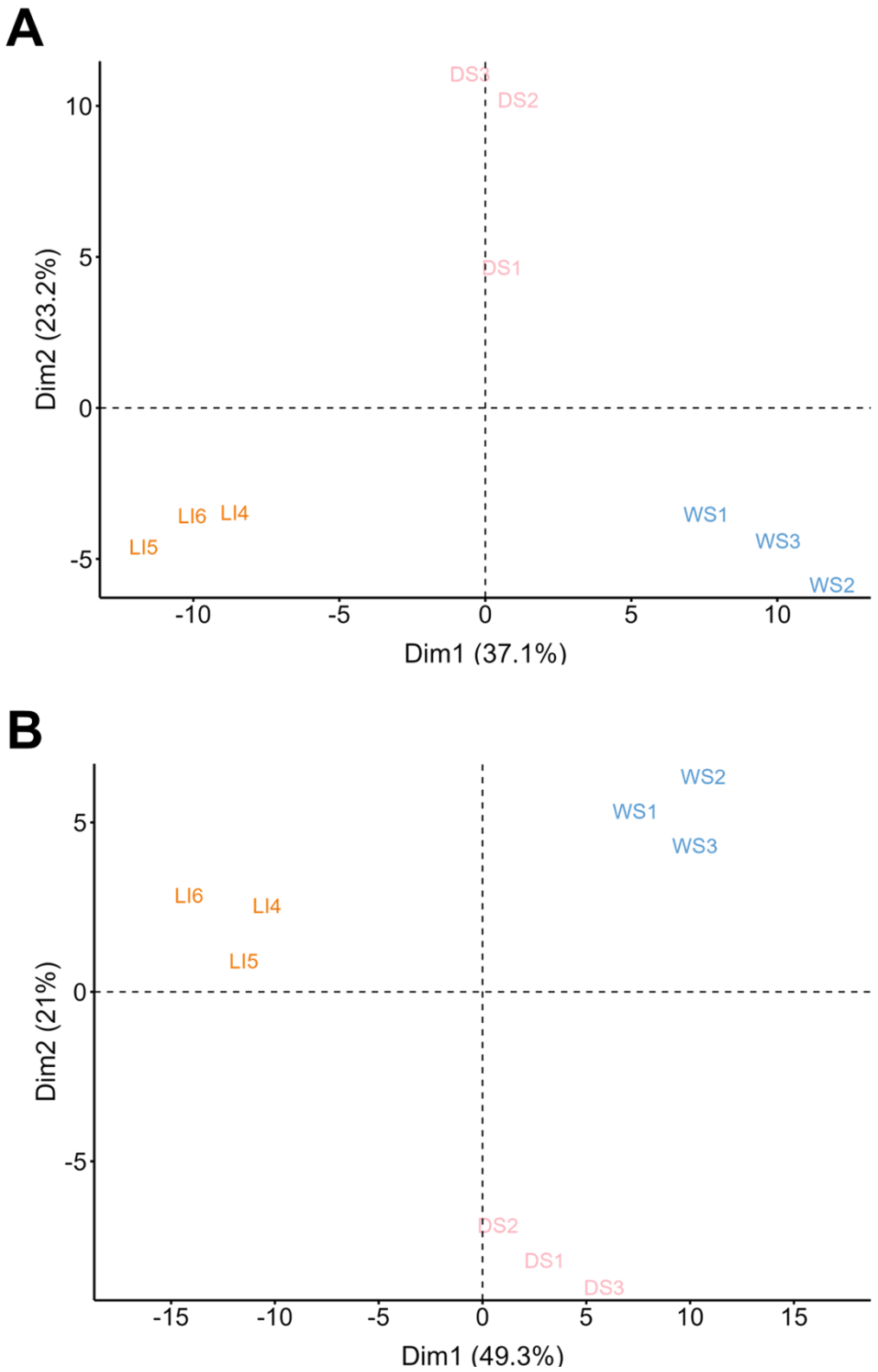
Supplementary Figure 5.2 Comparing the surfaces of normal and delignified wheat straw by scanning electron microscopy. Scanning electron microscopy (SEM) was used to image normal (**A**, **C**, and **E**) and delignified (**B**, **D**, and **F**) wheat straw at x400 (**A** and **B**), x750 (**C** and **D**), and x2000 (**E** and **F**) magnification to visually compare the substrates.



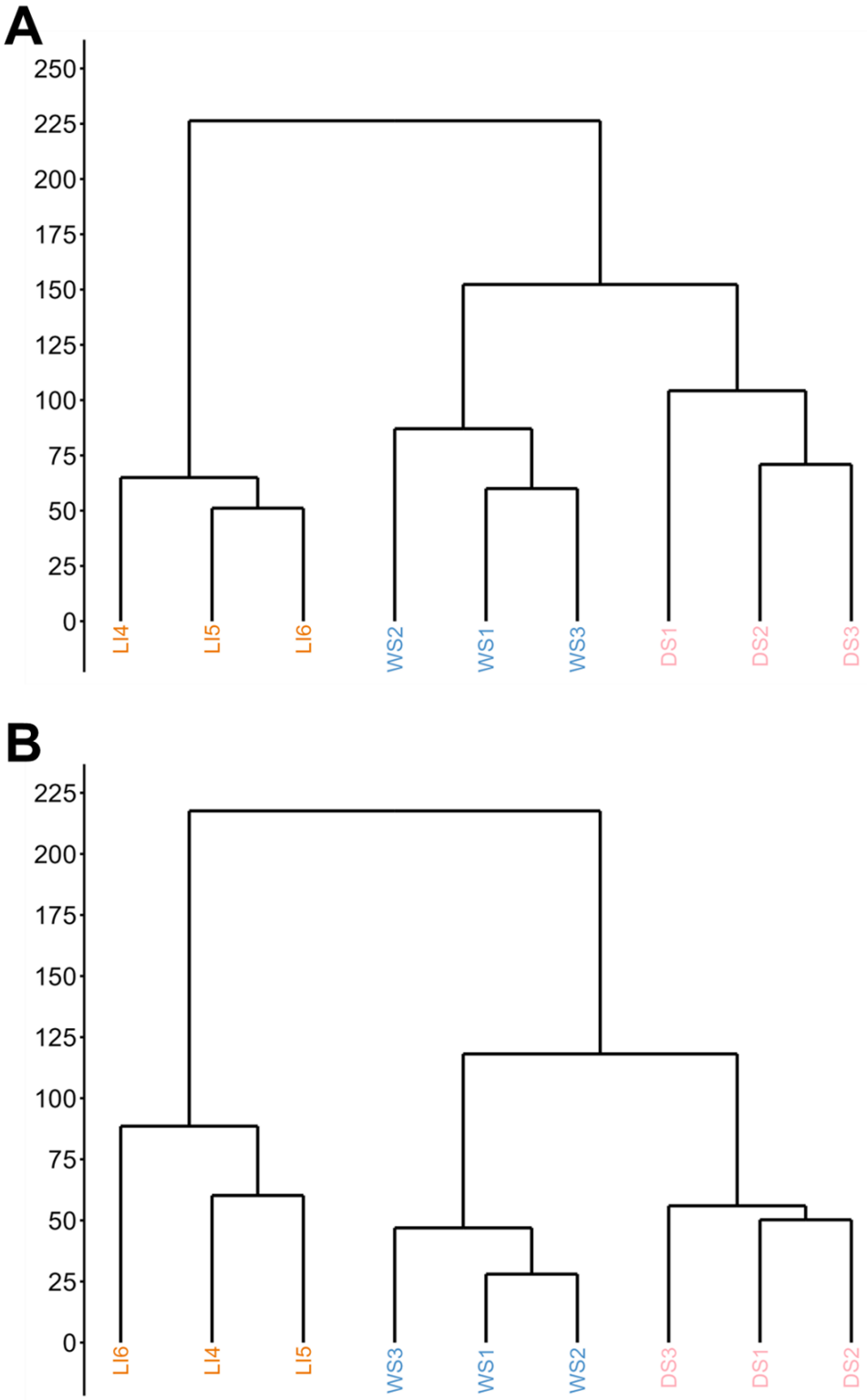
Supplementary Figure 5.3 Investigating the distribution of proteins across bound and supernatant fractions in the *P. putredinis* NO1 lignin degrading proteome. The number of proteins identified in at least one replicate across all three substrates for the bound fraction compared to the supernatant fraction within the *P. putredinis* NO1 lignin degrading proteome harvested from growth on three lignocellulosic substrates for 7 days.



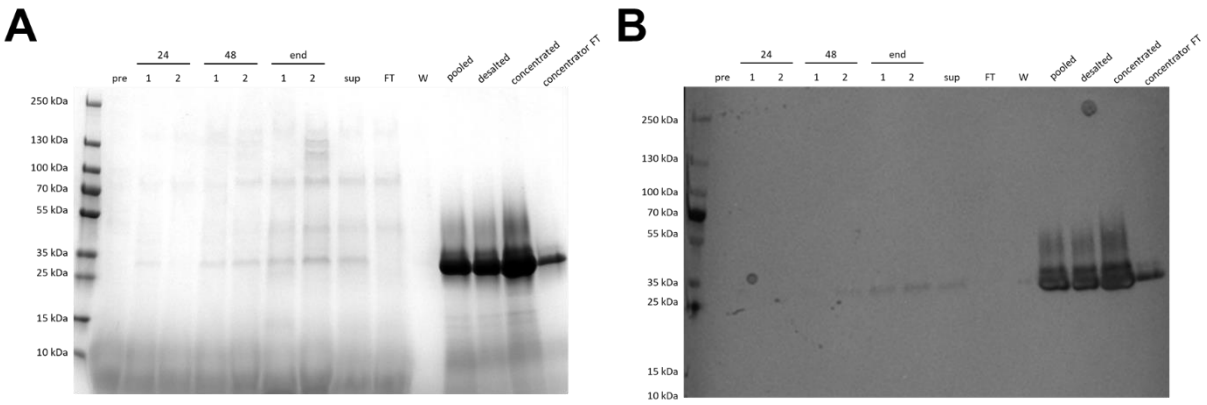
Supplementary Figure 5.4 Differences in the *P. putredinis* NO1 proteome across substrates. Molar percentage values for proteins identified on at least one substrate in the bound (**A**) and supernatant (**B**) fractions of the *P. putredinis* NO1 proteome scaled to Z-scores across substrates. DS = Delignified straw, WS = Wheat straw, LI = Dioxane lignin.



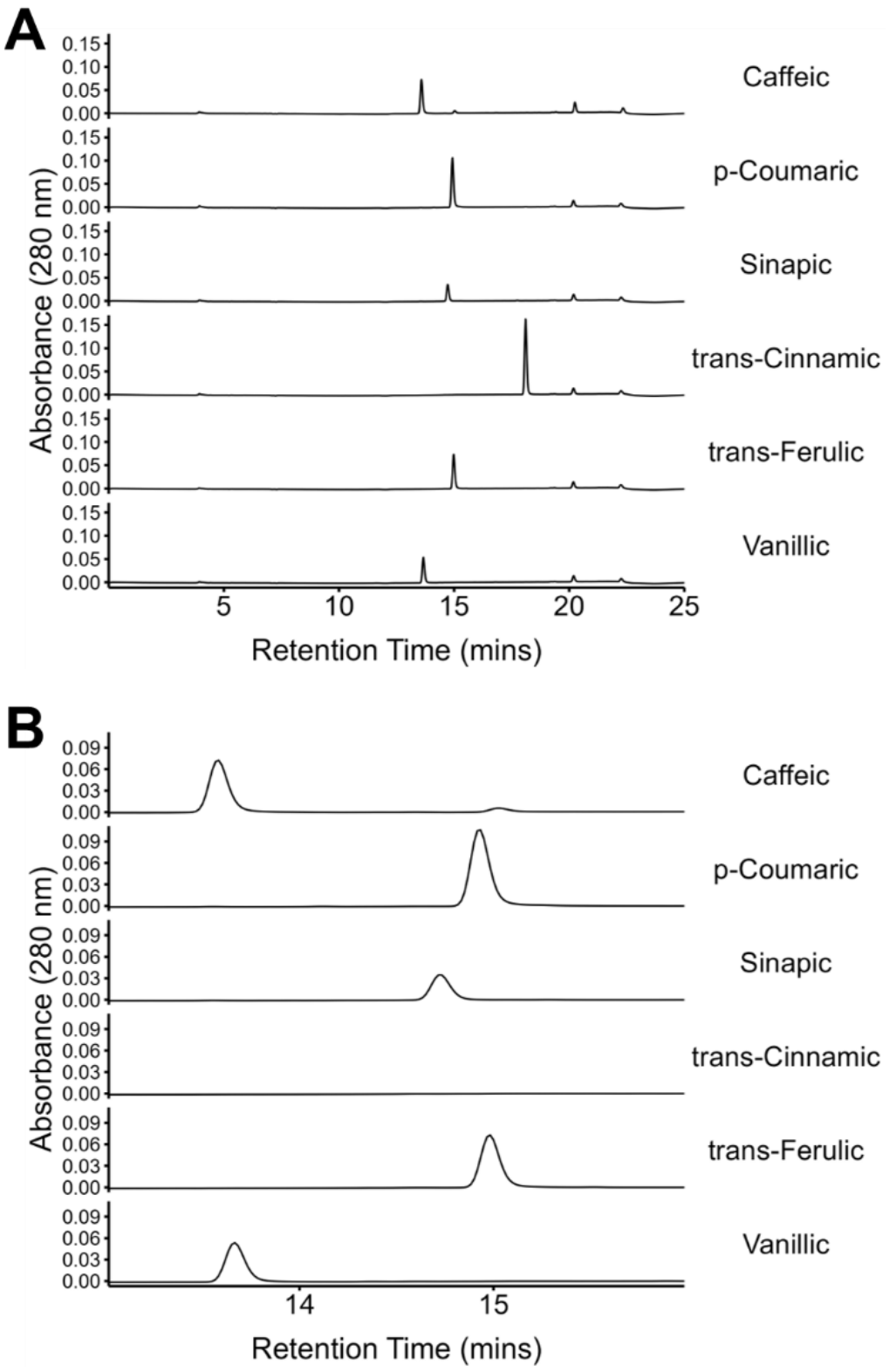
Supplementary Figure 5.5 Variation in the *P. putredinis* NO1 secretome across lignocellulosic substrates. A PCA was calculated from the MolPct values of secretome proteins identified in the bound (**A**) and supernatant (**B**) fractions for cultures of *P. putredinis* NO1 grown on three substrates. DS = Delignified straw, WS = Wheat straw, LI = Dioxane lignin.



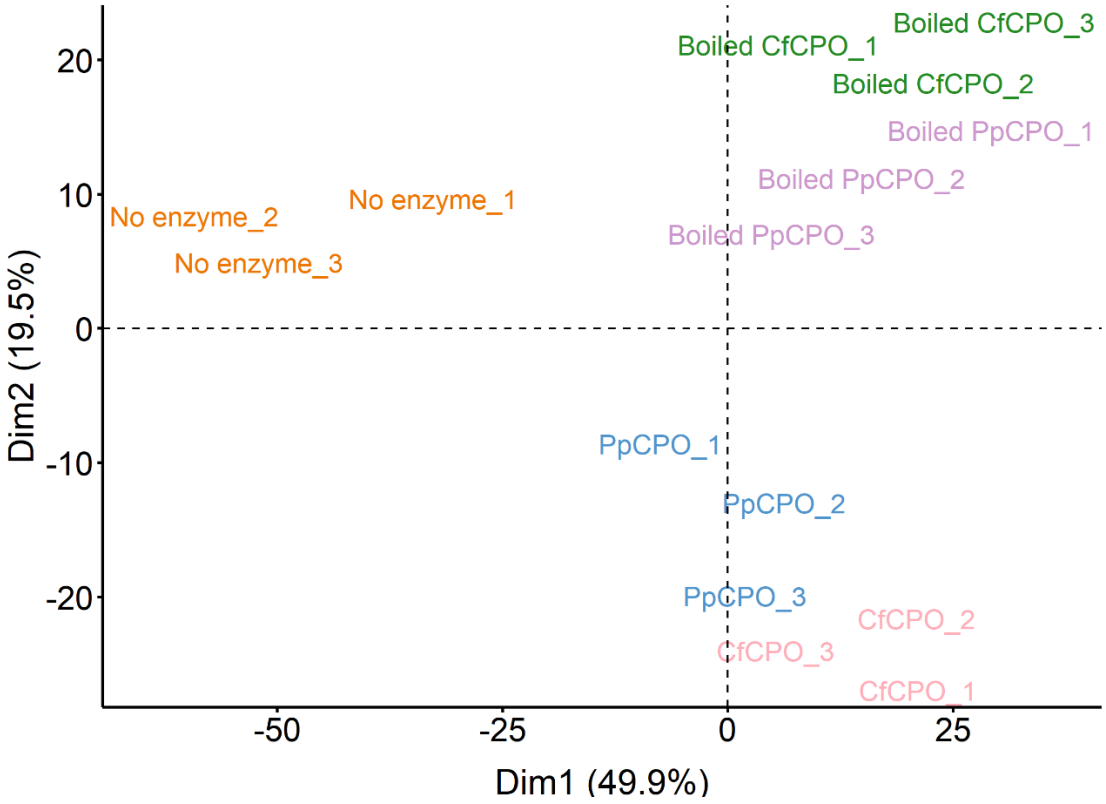
Supplementary Figure 5.6 Clustering of replicates of the *P. putredinis* NO1 secretome across lignocellulosic substrates. A canberra distance matrix was calculated and hierarchically clustered by Ward's minimum variance method for molar percentage values for proteins in the bound (**A**) and supernatant (**B**) fractions of the *P. putredinis* NO1 secretome after growth on three substrates. DS = Delignified straw, WS = Wheat straw, LI = Dioxane lignin.



Supplementary Figure 5.7 Purification of the recombinant *P. putredinis* NO1 chloroperoxidase-like protein. Samples of pre-cultures (pre), of 24-hour, 48-hour, and end time points from two 600 mL cultures (1 and 2) of recombinant *P. putredinis* NO1 chloroperoxidase-like expressing *P. pastoris* X-33, the combined soluble fraction (sup), the flowthrough (FT) and wash (W) fractions from nickel-affinity purification, the pooled eluted protein, the desalted protein, the concentrated protein and the concentrator flowthrough were prepared for duplicate SDS-PAGE gels for Coomassie staining (**A**) and western blotting (**B**) with anti-His antibody.



Supplementary Figure 5.8 Characterisation of lignin derivative standards by HPLC analysis. Standards of lignin derivatives were prepared to 100 µM concentration in 25 mM potassium phosphate buffer with 20 mM KCl, 20 mM H_2O_2 and subjected to HPLC analysis. Total chromatograms (**A**) and chromatograms focused on 13 to 16 minutes (**B**) are representative of three replicates.



Supplementary Figure 5.9 Principal component analysis of lignin modification HPLC data. Dioxane lignin from wheat straw was incubated in triplicate with 25 mM potassium phosphate buffer pH 3.0, 20 mM H_2O_2, and 20 mM KCl alone (No enzyme), or with 100 µg of boiled *Pp*CPO, boiled *Cf*CPO, *Pp*CPO, or *Cf*CPO for 16 hours at 30 °C with shaking at 600 rpm. The soluble fraction of these incubations was subjected to HPLC analysis, and the chromatogram data used to calculate a PCA.

5.8.2. Supplementary tables

Supplementary Table 5.1 Relative proportion of monosaccharides in hemicellulosic sugars. The proportions of hemicellulosic sugars were quantified for triplicate samples of normal and delignified wheat straw.

Monosaccharide	Delignified wheat straw (%)	Normal wheat straw (%)
Arabinose	14.82126	14.06978
Fucose	0.745018	3.443761
Galactose	11.48155	12.32325
Galacturonic acid	4.251331	3.80574
Glucose	16.08821	16.26114
Glucuronic acid	1.136435	1.564497
Mannose	1.36668	3.386076
Rhamnose	10.21066	9.309209
Xylose	39.89886	35.83655

Chapter 6

6. Discussion

The unsustainability of fossil fuels combined with the environmental and public health concerns associated with their extraction and use has driven attention towards possible alternatives. The development of first-generation biofuel technology provided the platform to transition to the use of lignocellulosic feedstocks which are available in vast amounts, at low cost, and do not compete directly with food production. However, these feedstocks are greatly varied and complex, and significant advances are still required to effectively utilise this intricate and insoluble network of polysaccharides and aromatic polymers. Unfortunately, the current technologies for the conversion of lignocellulose have been historically developed with a view to produce fuel alone. Therefore, many of these technologies are not suitable for achieving total substrate valorisation. Biological treatments of woody plant biomass may offer an efficient and environmentally friendly approach to lignocellulose conversion, which may also be suitable for total substrate valorisation. The ability to degrade, convert, and utilise lignocellulosic residues exists throughout nature, but fungi are perhaps the most powerful organisms in terms of their ability to produce enzymes to deconstruct all components of plant biomass.

There is, therefore, growing interest in understanding the capabilities of a wider range of the fungal kingdom, to improve the current understanding of fungal lignocellulose breakdown, and to identify new enzymes to help generate value from these substrates. *P. putredinis* NO1 is a recently identified strain of an underexplored genus of ascomycete fungi. Its ability to efficiently utilise lignocellulose as a carbon source for growth and the previous identification of a new oxidase enzyme which cleaves the major structural bond in lignin from this fungus make it an exciting candidate to explore. Through a combination of omics techniques and bioinformatic analysis, the lignocellulose-degrading capacity of this fungus was explored in detail. The ability of *P. putredinis* NO1 to tailor its secreted enzyme response to different industrially relevant lignocellulosic substrates provides a platform for a deeper understanding of fungal conversion of plant biomass while also demonstrating the importance of developing tailored enzymatic cocktails for industrial biorefinery processes. The novelty of *P. putredinis* NO1, the in-depth characterisation of its enzymatic repertoire, and the omics and bioinformatic techniques used here, also generates a reservoir of data to mine for new lignocellulose-degrading enzymes and activities.

6.1. Understanding the genomic capacity of *P. putredinis* NO1 for lignocellulose breakdown

An important step in improving the understanding of *P. putredinis* NO1 was obtaining an annotated and assembled genome for this fungus. As the first genome for the *Parascedosporium* genus it provided a novel reservoir of sequence data to explore. The lignocellulose-degrading potential of *P. putredinis* NO1 was investigated at the genomic level, and a high proportion of oxidative AA class CAZymes was observed in comparison to other ascomycetes. It was hypothesised that this may suggest a strategy in a wood-degrading community where *P. putredinis* NO1 targets the more recalcitrant crystalline cellulose and lignin components as carbon sources for growth. However, CAZyme repertoire alone may not encompass the complete set of genes with potential lignocellulose-degrading activity within the *P. putredinis* NO1 genome. With the recent advancement of structural prediction techniques, then predicted structures of almost all coding sequences in the *P. putredinis* NO1 genome could be compared to the structures of enzymes of interest. By combining this with sequence- and domain-based approaches, predicted structures with similarity to the structures of important oxidative lignocellulose degrading enzymes, but which lack any sequence similarity could be isolated and explored.

This strategy allowed the identification of many interesting sequences which would not have been isolated using sequence or structural searching approaches individually. The strategy would benefit from validation of the sequences identified through characterisation of the proteins they encode. As tools to handle the growing wealth of structural data mature, it would be interesting to investigate the use of structural comparison tools that can more effectively search for important structural motifs. For example, if predicted structures encoded within the *P. putredinis* NO1 genome could be identified with structural similarity to oxidative enzymes at the active site or at metal co-ordination sites, but which lacked structural similarity outside of these regions. This strategy could also become more powerful when combined with other bioinformatic techniques to further reduce the requirement for manual investigation of sequences.

6.2. Effectively isolating the *in silico* secretome of *P. putredinis* NO1

Although the new genome for *P. putredinis* NO1 provided a more effective reference for future experiments such as proteomic investigations, the inherent intracellular contamination of proteomic investigations into extracellular processes makes analysis difficult. Therefore, a bioinformatic workflow was developed to isolate sequences encoding proteins predicted to be involved in extracellular processes and actively released into the extracellular space. The workflow was applied here to all coding regions of the *P. putredinis* NO1 genome but is readily adaptable to any sequence data from microorganisms. An advantage of the workflow is its flexibility to incorporate new tools as they emerge. The advancing techniques for structural prediction allowed the genome of *P. putredinis* NO1 to be explored using a new strategy, however the tools to investigate structural data still require improvement. As these tools improve, their incorporation into the workflow can be explored. This could be through prediction of extracellular proteins based on structure or the use of predicted structures for functional annotation. This could improve the accuracy of the secretome isolation workflow and allow more effective evaluations of its ability to remove intracellular proteins. Although improvements to this workflow will be made as new tools and strategies develop, the ability of this workflow to effectively isolate a subset of sequences of interest for an extracellular process was demonstrated here. By achieving this, the clarity of proteomic investigations into the lignocellulose-degrading ability of *P. putredinis* NO1 was greatly improved.

6.3. Revealing the complex tailored secretome response of *P. putredinis* NO1 to different lignocellulosic substrates

The bioinformatic analysis of the newly obtained sequence data in the *P. putredinis* NO1 genome allowed the potential lignocellulose degrading capacity of the genome to be investigated, and potentially new enzymes to be identified. However, to investigate lignocellulose breakdown by *P. putredinis* NO1 at a functional level, proteomic analysis of the growth of this fungus on multiple industrially relevant substrates was performed. By harvesting proteins from both soluble and insoluble fractions a more complete picture of the lignocellulose degrading response was captured and utilising the secretome isolation data generated through analysis of the genome allowed clearer comparisons between conditions. The result was the demonstration of a remarkably varied secretome produced by *P. putredinis* NO1 in response to the different substrates investigated. Understanding the adaptability of fungal secretomes to different lignocellulosic substrates could be industrially important. Currently, commercial cocktails used in the deconstruction of lignocellulose by biorefineries lack diversity. This results from a focus on cellulose solely for fuel production and not complete valorisation, a strategy that will need to change and will require a range of enzymatic combinations depending on the feedstocks available for biorefineries and their target products.

An activity-based labelling approach was employed to visualise the changing patterns of lignocellulose-degrading enzyme production over time and across substrates in the *P. putredinis* NO1 secretome. Again, variation in the plant biomass degrading response was observed across substrates and importantly this variation was maintained over time. Unfortunately, this technique is currently only available for a small number of hydrolytic CAZymes. Although it further demonstrates variation in lignocellulose degrading enzymes produced when grown on different lignocellulosic substrates and highlights the importance of reflecting this in industry, it would be ideal to include other classes of enzymes in this labelling approach. Particularly, it would be interesting to explore the time-dependent variation in the production of the oxidative enzymes which have already been explored within the genome of *P. putredinis* NO1 and which are involved in degrading lignin and crystalline cellulose. Improving the understanding of the roles of these enzymes in the breakdown of different industrially relevant substrates would be of major value for the improvement of commercial cocktails.

6.4. Utilising the tailored *P. putredinis* NO1 secretome for the identification of new lignin-degrading enzymes

Another issue with the limited commercial cocktails for the deconstruction of lignocellulose is their focus on the polysaccharide fractions of biomass. These cocktails need to be expanded to incorporate new enzymes for the depolymerisation, degradation, and modification of lignin which is currently hugely underutilised despite its clear potential value as the most abundant aromatic polymer on earth. The tailored secretome of *P. putredinis* NO1 suggested that experiments could be designed to investigate the production of enzymes with functions of interest, in this case lignin degradation. When the growth of *P. putredinis* NO1 was explored across three substrates with varying lignin content, a subset of proteins was isolated based on their increased abundance on lignin-enriched substrates.

Within this subset were many proteins not annotated as CAZymes and therefore not traditionally implicated in lignocellulose or lignin breakdown. One such protein was the chloroperoxidase-like protein which showed structural similarity to crystallised chloroperoxidases and the same active site residues. Through recombinant expression and purification, the potential for this protein's involvement in lignin degradation or modification was investigated, something which has been suggested for chloroperoxidases previously. Evidence for the ability of the *P. putredinis* NO1 chloroperoxidase enzyme to cleave a model compound mimicking the major structural linkage in lignin was demonstrated, and changes to the soluble product profile after incubation of the enzyme with a polymeric lignin substrate was observed. However, the exact role of this chloroperoxidase in lignin breakdown remains unclear. Further characterisation of lignin substrates, but also industrially relevant lignocellulosic feedstocks, after incubation with this enzyme may be required to understand this further. Particularly, the chlorination of polymeric lignin substrates will be an important step in supporting the role of chloroperoxidases in lignin modification. The difficulty in understanding the interaction of these enzymes with lignin is partly due to the insolubility of the substrate. Characterisation of the structure and composition of insoluble lignin may be more effective in understanding the activity of chloroperoxidases towards lignin than investigating the soluble products. However, the investigation of this new chloroperoxidase enzyme from *P. putredinis* NO1 has added to the very limited knowledge of these widespread fungal enzymes regarding lignin breakdown.

6.5. Final conclusions

The work presented here has extensively explored the newly identified ascomycete wood-degrading fungus *P. putredinis* NO1. Exploration of the first genome for this genus provided insight into the ability for *P. putredinis* NO1 to target the most recalcitrant components of lignocellulose as carbon sources for growth and highlighted the potential to identify new enzymes from this fungus. Particularly, the identification of new sequences and proteins with lignin degrading or modifying activity was of major interest. This was due to the shift in biorefinery goals for lignocellulose conversion where the importance of total substrate valorisation, including the generation of products from the historically underutilised lignin component, is now being realised. The characterisation of new microorganisms such as *P. putredinis* NO1 with the ability to degrade lignocellulose will be vital to expand the repertoire of enzymes available to achieve this in an environmentally friendly and cost-effective manner.

However, it is not just the identification of enzymes alone which will be important in achieving this. An improved understanding of total lignocellulose conversion will be required, and the flexibility of these systems in nature to different substrates. For this to be achieved, the lignocellulose-degrading strategies of new microorganisms will need to be characterised, new bioinformatic workflows to investigate sequence data and simplify analyses will be required, and experiments will have to be carefully designed to explore these processes at a functional level. In this work, all of this has been attempted for *P. putredinis* NO1 to understand its tailored secretome and to identify new enzymes. Perhaps more importantly though, this work has provided a platform of workflows and analysis strategies that can be utilised for the future investigation of other lignocellulose-degrading systems to expand our knowledge of lignocellulose breakdown and to enhance biorefinery processes for the development of sustainable and renewable fuels, chemicals, and materials.

References

1. Koch S, Klitzman R. Reliance on fossil fuels: ethical implications for intensivists. Intens Care Med. 2023;49(3):330-3.

2. Environmental, Energy Study I. Fossil Fuels [Available from: https://www.eesi.org/topics/fossil-fuels/description.

3. Ritchie H, Roser M, Rosado P. Energy. Our World in Data [Internet]. 2022 2022/10/27. Available from: https://ourworldindata.org/energy

https://ourworldindata.org/energy-production-consumption.

4. Wang D, Cui JH, Gan MZ, Xue ZH, Wang J, Liu PF, et al. Transformation of Biomass DNA into Biodegradable Materials from Gels to Plastics for Reducing Petrochemical Consumption. J Am Chem Soc. 2020;142(22):10114-24.

5. Singh N, Ogunseitan OA, Wong MH, Tang Y. Sustainable materials alternative to petrochemical plastics pollution: A review analysis. Sustainable Horizons. 2022;2:100016.

6. Concrete needs to lose its colossal carbon footprint. Nature. 2021;597(7878):593-4.

7. Olmez GM, Dilek FB, Karanfil T, Yetis U. The environmental impacts of iron and steel industry: a life cycle assessment study. J Clean Prod. 2016;130:195-201.

8. Tickner J, Geiser K, Baima S. Transitioning the Chemical Industry: The Case for Addressing the Climate, Toxics, and Plastics Crises. Environment. 2021;63(6):4-15.

9. World Energy Investment 2023. Available from: https://www.iea.org/reports/world-energy-investment-2023.

10. Tip of the iceberg: The future of fossil fuel extraction. Available from: https://policy.friendsoftheearth.uk/insight/tip-iceberg-future-fossil-fuel-extraction.

11. Henrich E, Dahmen N, Dinjus E, Sauer J. The Role of Biomass in a Future World without Fossil Fuels. Chem Ing Tech. 2015;87(12):1667-85.

12. Roser M, Rodés-Guirao L. Future Population Growth. Our World in Data [Internet]. 2013 2013/5/9. Available from: https://ourworldindata.org/future-population-growth.

13. Dhanya BS, Mishra A, Chandel AK, Verma ML. Development of sustainable approaches for converting the organic waste to bioenergy. Sci Total Environ. 2020;723:138109.

14. Hanif I, Faraz Raza SM, Gago-de-Santos P, Abbas Q. Fossil fuels, foreign direct investment, and economic growth have triggered CO2 emissions in emerging Asian economies: Some empirical evidence. Energy. 2019;171:493-501.

15. Kreps BH. The Rising Costs of Fossil-Fuel Extraction: An Energy Crisis That Will Not Go Away. Am J Econ Sociol. 2020;79(3):695-717.

16. Leonard MD, Michaelides EE, Michaelides DN. Energy storage needs for the substitution of fossil fuel power plants with renewables. Renewable Energy. 2020;145:951-62.

17. Goli A, Shamiri A, Talaiekhozani A, Eshtiaghi N, Aghamohammadi N, Aroua MK. An overview of biological processes and their potential for CO2 capture. J Environ Manage. 2016;183:41-58.

18. Dashtban M, Schraft H, Qin W. Fungal bioconversion of lignocellulosic residues; opportunities & perspectives. Int J Biol Sci. 2009;5(6):578-95.

19. Abbass K, Qasim MZ, Song HM, Murshed M, Mahmood H, Younis I. A review of the global climate change impacts, adaptation, and sustainable mitigation measures. Environ Sci Pollut R. 2022;29(28):42539-59.

20. Parry ML, Rosenzweig C, Iglesias A, Livermore M, Fischer G. Effects of climate change on global food production under SRES emissions and socio-economic scenarios. Global Environ Chang. 2004;14(1):53-67.

21. Butchart SHM, Walpole M, Collen B, van Strien A, Scharlemann JPW, Almond REA, et al. Global Biodiversity: Indicators of Recent Declines. Science. 2010;328(5982):1164-8.

22. Hoegh-Guldberg O, Mumby PJ, Hooten AJ, Steneck RS, Greenfield P, Gomez E, et al. Coral reefs under rapid climate change and ocean acidification. Science. 2007;318(5857):1737-42.

23. (IEA) IEA. Weo-2016 Special Report Energy and Air Pollution. Paris, France: International Energy Agency; 2016.

24. Perera F. Pollution from Fossil-Fuel Combustion is the Leading Environmental Threat to Global Pediatric Health and Equity: Solutions Exist. Int J Environ Res Public Health. 2017;15(1).

25. Landrigan PJ, Fuller R, Acosta NJR, Adeyi O, Arnold R, Basu N, et al. The Lancet Commission on pollution and health. Lancet. 2018;391(10119):462-512.

26. Travaglio M, Yu YZ, Popovic R, Selley L, Leal NS, Martins LM. Links between air pollution and COVID-19 in England. Environ Pollut. 2021;268.

27. Zakeri B, Paulavets K, Barreto-Gomez L, Echeverri LG, Pachauri S, Boza-Kiss B, et al. Pandemic, War, and Global Energy Transitions. Energies. 2022;15(17).

28. Slav I. The Real Reason Oil Prices Crashed | OilPrice.com2020 2020/3/23. Available from: https://oilprice.com/Energy/Oil-Prices/The-Real-Reason-Oil-Prices-Crashed.html.

29. Tollefson J. COVID curbed carbon emissions in 2020 - but not by much. Nature. 2021;589(7842):343.

30. Paital B. Nurture to nature via COVID-19, a self-regenerating environmental strategy of environment in global context. Sci Total Environ. 2020;729:139088.

31. Barbier EB, Burgess JC. Sustainability and development after COVID-19. World Dev. 2020;135:105082.

32. Gielen D, Boshell F, Saygin D, Bazilian MD, Wagner N, Gorini R. The role of renewable energy in the global energy transformation. Energy Strateg Rev. 2019;24:38-50.

33. Kåberger T. Progress of renewable electricity replacing fossil fuels. Global Energy Interconnection. 2018;1(1):48-52.

34. Kreps BH. Energy Sprawl in the Renewable-Energy Sector: Moving to Sufficiency in a Post-Growth Era. Am J Econ Sociol. 2020;79(3):719-49.

35. Alizadeh R, Soltanisehat L, Lund PD, Zamanisabzi H. Improving renewable energy policy planning and decision-making through a hybrid MCDM method. Energy Policy. 2020;137:111174.

36. Vaez E, Zilouei H. Towards the development of biofuel production from paper mill effluent. Renewable Energy. 2020;146:1408-15.

37. Wang S, Yu S, Lu Q, Liao Y, Li H, Sun L, et al. Development of an alkaline/acid pre-treatment and anaerobic digestion (APAD) process for methane generation from waste activated sludge. Sci Total Environ. 2020;708:134564.

38. Shylesh S, Gokhale AA, Ho CR, Bell AT. Novel Strategies for the Production of Fuels, Lubricants, and Chemicals from Biomass. Accounts Chem Res. 2017;50(10):2589-97.

39. Lee RA, Lavoie J-M. From first- to third-generation biofuels: Challenges of producing a commodity from a biomass of increasing complexity. Anim Fron. 2013;3(2):6-11.

40. Naik SN, Goud VV, Rout PK, Dalai AK. Production of first and second generation biofuels: A comprehensive review. Renew Sust Energ Rev. 2010;14(2):578-97.

41. Esmaeili SAH, Szmerekovsky J, Sobhani A, Dybing A, Peterson TO. Sustainable biomass supply chain network design with biomass switching incentives for first-generation bioethanol producers. Energy Policy. 2020;138.

42. Amorim HV, Lopes ML, Oliveira JVD, Buckeridge MS, Goldman GH. Scientific challenges of bioethanol production in Brazil. Appl Microbiol Biot. 2011;91(5):1267-75.

43. Havlík P, Schneider UA, Schmid E, Böttcher H, Fritz S, Skalský R, et al. Global land-use implications of first and second generation biofuel targets. Energy Policy. 2011;39(10):5690-702.

44. Izah SC, Ohimain EI, Others. Bioethanol production from cassava mill effluents supplemented with solid agricultural residues using bakers' yeast [*Saccharomyces cerevisiae*]. Journal of Environmental Treatment Techniques. 2015;3(1):47-54.

45. Tingley JP, Low KE, Xing XH, Abbott DW. Combined whole cell wall analysis and streamlined in silico carbohydrate-active enzyme discovery to improve biocatalytic conversion of agricultural crop residues. Biotechnol Biofuels. 2021;14(1).

46. Rulli MC, Bellomi D, Cazzoli A, De Carolis G, D'Odorico P. The water-land-food nexus of first-generation biofuels. Sci Rep-Uk. 2016;6.

47. Gasparatos A, Stromberg P, Takeuchi K. Sustainability impacts of first-generation biofuels. Anim Fron. 2013;3(2):12-26.

48. Gronchi N, Favaro L, Cagnin L, Brojanigo S, Pizzocchero V, Basaglia M, et al. Novel Yeast Strains for the Efficient Saccharification and Fermentation of Starchy By-Products to Bioethanol. Energies. 2019;12(4):1-13.

49. Rezania S, Oryani B, Cho J, Talaiekhozani A, Sabbagh F, Hashemi B, et al. Different pretreatment technologies of lignocellulosic biomass for bioethanol production: An overview. Energy. 2020;199:117457.

50. Dahmen N, Lewandowski I, Zibek S, Weidtmann A. Integrated lignocellulosic value chains in a growing bioeconomy: Status quo and perspectives. Glob Change Biol Bioenergy. 2019;11(1):107-17.

51. Zheng Q, Zhou TT, Wang YB, Cao XH, Wu SQ, Zhao ML, et al. Pretreatment of wheat straw leads to structural changes and improved enzymatic hydrolysis. Sci Rep-Uk. 2018;8.

52. Liu D, Yan X, Zhuo SN, Si MY, Liu MR, Wang S, et al. *Pandoraea* sp B-6 assists the deep eutectic solvent pretreatment of rice straw via promoting lignin depolymerization. Bioresource Technol. 2018;257:62-8.

53. Jugwanth Y, Sewsynker-Sukai Y, Kana EBG. Valorization of sugarcane bagasse for bioethanol production through simultaneous saccharification and fermentation: Optimization and kinetic studies. Fuel. 2020;262.

54. Ishola MM, Isroi, Taherzadeh MJ. Effect of fungal and phosphoric acid pretreatment on ethanol production from oil palm empty fruit bunches (OPEFB). Bioresource Technol. 2014;165:9-12.

55. Saha BC, Qureshi N, Kennedy GJ, Cotta MA. Biological pretreatment of corn stover with white-rot fungus for improved enzymatic hydrolysis. Int Biodeter Biodegr. 2016;109:29-35.

56. Daioglou V, Stehfest E, Wicke B, Faaij A, van Vuuren DP. Projections of the availability and cost of residues from agriculture and forestry. Glob Change Biol Bioenergy. 2016;8(2):456-70.

57. Wagner M, Mangold A, Lask J, Petig E, Kiesel A, Lewandowski I. Economic and environmental performance of miscanthus cultivated on marginal land for biogas production. Glob Change Biol Bioenergy. 2019;11(1):34-49.

58. Ssegane H, Zumpf C, Cristina Negri M, Campbell P, Heavey JP, Volk TA. The economics of growing shrub willow as a bioenergy buffer on agricultural fields: A case study in the Midwest Corn Belt. Biofuels Bioprod Biorefin. 2016;10(6):776-89.

59. Clifton-Brown J, Harfouche A, Casler MD, Dylan Jones H, Macalpine WJ, Murphy-Bokern D, et al. Breeding progress and preparedness for mass-scale deployment of perennial lignocellulosic biomass crops switchgrass, miscanthus, willow and poplar. Glob Change Biol Bioenergy. 2019;11(1):118-51.

60. Toscan A, Morais ARC, Paixão SM, Alves L, Andreaus J, Camassola M, et al. High-pressure carbon dioxide/water pre-treatment of sugarcane bagasse and elephant grass: Assessment of the effect of biomass composition on process efficiency. Bioresour Technol. 2017;224:639-47.

61. Dien BS, Casler MD, Hector RE, Iten LB, Nichols NN, Mertens JA, et al. Biochemical processing of reed canarygrass into fuel ethanol. Int J Low-Carbon Tech. 2012;7(4):338-47.

62. Safarian S, Unnthorsson R. An Assessment of the Sustainability of Lignocellulosic Bioethanol Production from Wastes in Iceland. Energies. 2018;11(6).

63. Hassan SS, Williams GA, Jaiswal AK. Moving towards the second generation of lignocellulosic biorefineries in the EU: Drivers, challenges, and opportunities. Renewable Sustainable Energy Rev. 2019;101:590-9.

64. Adewuyi A. Production of Biodiesel from Underutilized Algae Oil: Prospects and Current Challenges Encountered in Developing Countries. Biology (Basel). 2022;11(10).

65. Saini JK, Saini R, Tewari L. Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments. 3 Biotech. 2015;5(4):337-53.

66. Nguyen DT, Gomez LD, Harper A, Halpin C, Waugh R, Simister R, et al. Association mapping identifies quantitative trait loci (QTL) for digestibility in rice straw. Biotechnol Biofuels. 2020;13:165.

67. Stephen JD, Mabee WE, Saddler JN. Will second-generation ethanol be able to compete with first-generation ethanol? Opportunities for cost reduction. Biofuels Bioprod Biorefin. 2012;6(2):159-76.

68. Macrelli S, Mogensen J, Zacchi G. Techno-economic evaluation of 2nd generation bioethanol production from sugar cane bagasse and leaves integrated with the sugar-based ethanol process. Biotechnol Biofuels. 2012;5:22.

69. Rafa N, Ahmed SF, Badruddin IA, Mofijur M, Kamangar S. Strategies to Produce Cost-Effective Third-Generation Biofuel From Microalgae. Front Energy Res. 2021;9.

70. Said Z, Nguyen TH, Sharma P, Li CH, Ali HM, Nguyen V, et al. Multi-attribute optimization of sustainable aviation fuel production-process from microalgae source. Fuel. 2022;324.

71. Toor M, Kumar SS, Malyan SK, Bishnoi NR, Mathimani T, Rajendran K, et al. An overview on bioethanol production from lignocellulosic feedstocks. Chemosphere. 2020;242:125080.

72. Heidari M, Dutta A, Acharya B, Mahmud S. A review of the current knowledge and challenges of hydrothermal carbonization for biomass conversion. J Energy Inst. 2019;92(6):1779-99.

73. Jensen CU, Guerrero JKR, Karatzos S, Olofsson G, Iversen SB. Fundamentals of Hydrofaction((TM)): Renewable crude oil from woody biomass. Biomass Conversion and Biorefinery. 2017;7(4):495-509.

74. Kumagai A, Endo T. Comparison of the surface constitutions of hemicelluloses on lignocellulosic nanofibers prepared from softwood and hardwood. Cellulose. 2018;25(7):3885-97.

75. Espinosa E, Sanchez R, Otero R, Dominguez-Robles J, Rodriguez A. A comparative study of the suitability of different cereal straws for lignocellulose nanofibers isolation. Int J Biol Macromol. 2017;103:990-9.

76. Popescu CM, Singurel G, Popescu MC, Vasile C, Argyropoulos DS, Willfor S. Vibrational spectroscopy and X-ray diffraction methods to establish the differences between hardwood and softwood. Carbohyd Polym. 2009;77(4):851-7.

77. Li T, Chen CJ, Brozena AH, Zhu JY, Xu LX, Driemeier C, et al. Developing fibrillated cellulose as a sustainable technological material. Nature. 2021;590(7844):47-56.

78. Nitsos C, Rova U, Christakopoulos P. Organosolv Fractionation of Softwood Biomass for Biofuel and Biorefinery Applications. Energies. 2017;11(1):50.

79. Anasontzis GE, Kourtoglou E, Villas-Boas SG, Hatzinikolaou DG, Christakopoulos P. Metabolic Engineering of *Fusarium oxysporum* to Improve Its Ethanol-Producing Capability. Front Microbiol. 2016;7.

80. Demeke MM, Dietz H, Li YY, Foulquie-Moreno MR, Mutturi S, Deprez S, et al. Development of a D-xylose fermenting and inhibitor tolerant industrial *Saccharomyces cerevisiae* strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering. Biotechnol Biofuels. 2013;6.

81. Tsai TY, Lo YC, Dong CD, Nagarajan D, Chang JS, Lee DJ. Biobutanol production from lignocellulosic biomass using immobilized *Clostridium acetobutylicum*. Appl Energ. 2020;277.

82. Ananthi V, Prakash GS, Chang SW, Ravindran B, Nguyen DD, Vo DVN, et al. Enhanced microbial biodiesel production from lignocellulosic hydrolysates using yeast isolates. Fuel. 2019;256.

83. Ou MS, Ingram LO, Shanmugam KT. l(+)-Lactic acid production from non-food carbohydrates by thermotolerant *Bacillus coagulans*. J Ind Microbiol Biot. 2011;38(5):599-605.

84. Hanko EKR, Denby CM, Nogue VSI, Lin WY, Ramirez KJ, Singer CA, et al. Engineering beta-oxidation in *Yarrowia lipolytica* for methyl ketone production. Metab Eng. 2018;48:52-62.

85. Zheng XJ, Gu XC, Ren Y, Zhi ZH, Lu XB. Production of 5-hydroxymethyl furfural and levulinic acid from lignocellulose in aqueous solution and different solvents. Biofuel Bioprod Bior. 2016;10(6):917-31.

86. Borrero-Lopez AM, Valencia C, Franco JM. Lignocellulosic Materials for the Production of Biofuels, Biochemicals and Biomaterials and Applications of Lignocellulose-Based Polyurethanes: A Review. Polymers-Basel. 2022;14(5).

87. Xu Y, Liu K, Yang Y, Kim M-S, Lee C-H, Zhang R, et al. Hemicellulose-based hydrogels for advanced applications. Front Bioeng Biotechnol. 2022;10:1110004.

88. Kumar A, Anushree, Kumar J, Bhaskar T. Utilization of lignin: A sustainable and eco-friendly approach. J Energy Inst. 2020;93(1):235-71.

89. Zhou NN, Thilakarathna WPDW, He QS, Rupasinghe HPV. A Review: Depolymerization of Lignin to Generate High-Value Bio-Products: Opportunities, Challenges, and Prospects. Front Energy Res. 2022;9.

90. Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, et al. Lignin Valorization: Improving Lignin Processing in the Biorefinery. Science. 2014;344(6185):709-+.

91. Schutyser W, Renders T, Van den Bosch S, Koelewijn SF, Beckham GT, Sels BF. Chemicals from lignin: an interplay of lignocellulose fractionation, depolymerisation, and upgrading. Chem Soc Rev. 2018;47(3):852-908.

92. Korte I, Kreyenschmidt J, Wensing J, Broring S, Frase JN, Pude R, et al. Can Sustainable Packaging Help to Reduce Food Waste? A Status Quo Focusing Plant-Derived Polymers and Additives. Appl Sci-Basel. 2021;11(11).

93. Rashid GMM, Bugg TDH. Enhanced biocatalytic degradation of lignin using combinations of lignin-degrading enzymes and accessory enzymes. Catal Sci Technol. 2021;11(10):3568-77.

94. Chandra RP, Chu QL, Hu JG, Zhong N, Lin M, Lee JS, et al. The influence of lignin on steam pretreatment and mechanical pulping of poplar to achieve high sugar recovery and ease of enzymatic hydrolysis. Bioresource Technol. 2016;199:135-41.

95. Moretti C, Corona B, Hoefnagels R, van Veen M, Vural-Gürsel I, Strating T, et al. Kraft lignin as a bio-based ingredient for Dutch asphalts: An attributional LCA. Sci Total Environ. 2022;806(Pt 1):150316.

96. Mohamad Aini NA, Othman N, Hussin MH, Sahakaro K, Hayeemasae N. Lignin as Alternative Reinforcing Filler in the Rubber Industry: A Review. Frontiers in Materials. 2020;6.

97. Wells T, Jr., Kosa M, Ragauskas AJ. Polymerization of Kraft lignin via ultrasonication for high-molecular-weight applications. Ultrason Sonochem. 2013;20(6):1463-9.

98. Maeda M, Hosoya T, Yoshioka K, Miyafuji H, Ohno H, Yamada T. Vanillin production from native softwood lignin in the presence of tetrabutylammonium ion. J Wood Sci. 2018;64(6):810-5.

99. Asada C, Basnet S, Otsuka M, Sasaki C, Nakamura Y. Epoxy resin synthesis using low molecular weight lignin separated from various lignocellulosic materials. Int J Biol Macromol. 2015;74:413-9.

100. Huang X, Ludenhoff JM, Dirks M, Ouyang X, Boot MD, Hensen EJM. Selective Production of Biobased Phenol from Lignocellulose-Derived Alkylmethoxyphenols. ACS Catal. 2018;8(12):11184-90.

101. Xiao CW, Anderson CT. Roles of pectin in biomass yield and processing for biofuels. Front Plant Sci. 2013;4.

102. Latarullo MBG, Tavares EQP, Maldonado GP, Leite DCC, Buckeridge MS. Pectins, Endopolygalacturonases, and Bioenergy. Front Plant Sci. 2016;7.

103. Sorek N, Yeats TH, Szemenyei H, Youngs H, Somerville CR. The Implications of Lignocellulosic Biomass Chemical Composition for the Production of Advanced Biofuels. Bioscience. 2014;64(3):192-201.

104. Mattonai M, Ribechini E. Fast screening for hydrolysable and condensed tannins in lignocellulosic biomass using reactive Py-GC/MS with in situ silylation. J Anal Appl Pyrol. 2018;135:242-50.

105. Wang H, Peng XD, Li H, Giannis A, He C. Recent Biotechnology Advances in Bio-Conversion of Lignin to Lipids by Bacterial Cultures. Front Chem. 2022;10.

106. Hassan SS, Williams GA, Jaiswal AK. Lignocellulosic Biorefineries in Europe: Current State and Prospects. Trends Biotechnol. 2018.

107. Pandiyan K, Singh A, Singh S, Saxena AK, Nain L. Technological interventions for utilization of crop residues and weedy biomass for second generation bio-ethanol production. Renewable Energy. 2019;132:723-41.

108. Berlin A, Munoz C, Gilkes N, Alamouti SM, Chung P, Kang KY, et al. An evaluation of British Columbian beetle-killed hybrid spruce for bioethanol production. Appl Biochem Biotech. 2007;137:267-80.

109. Pan XJ, Xie D, Yu RW, Lam D, Saddler JN. Pretreatment of lodgepole pine killed by mountain pine beetle using the ethanol organosolv process: Fractionation and process optimization. Ind Eng Chem Res. 2007;46(8):2609-17.

110. Abhishek A, Dwivedi A, Tandan N, Kumar U. Comparative bacterial degradation and detoxification of model and kraft lignin from pulp paper wastewater and its metabolites. Appl Water Sci. 2017;7(2):757-67.

111. Lee EK, Romeiko XX, Zhang WJ, Feingold BJ, Khwaja HA, Zhang XS, et al. Residential Proximity to Biorefinery Sources of Air Pollution and Respiratory Diseases in New York State. Environ Sci Technol. 2021;55(14):10035-45.

112. Eberle A, Bhatt A, Zhang YM, Heath G. Potential Air Pollutant Emissions and Permitting Classifications for Two Biorefinery Process Designs in the United States. Environ Sci Technol. 2017;51(11):5879-88.

113. Liu YZ, Chen WS, Xia QQ, Guo BT, Wang QW, Liu SX, et al. Efficient Cleavage of Lignin-Carbohydrate Complexes and Ultrafast Extraction of Lignin Oligomers from Wood Biomass by Microwave-Assisted Treatment with Deep Eutectic Solvent. Chemsuschem. 2017;10(8):1692-700.

114. Salleh SF, Gunawan MF, Zulkarnain MFB, Halim A. Modelling and optimization of biomass supply chain for bioenergy production. Journal of Environmental Treatment Techniques. 2019;7(4):689-95.

115. Vaidya AA, Donaldson LA, Newman RH, Suckling ID, Campion SH, Lloyd JA, et al. Micromorphological changes and mechanism associated with wet ball milling of *Pinus radiata* substrate and consequences for saccharification at low enzyme loading. Bioresource Technol. 2016;214:132-7.

116. Mokomele T, Sousa LD, Balan V, van Rensburg E, Dale BE, Gorgens JF. Incorporating anaerobic co-digestion of steam exploded or ammonia fiber expansion pretreated sugarcane residues with manure into a sugarcane-based bioenergy-livestock nexus. Bioresource Technol. 2019;272:326-36.

117. Santo MCD, Cardoso EB, Guimaraes FEG, DeAzevedo ER, da Cunha GP, Novotny EH, et al. Multifaceted characterization of sugarcane bagasse under different steam explosion severity conditions leading to distinct enzymatic hydrolysis yields. Ind Crop Prod. 2019;139.

118. Gollapalli LE, Dale BE, Rivers DM. Predicting digestibility of ammonia fiber explosion (AFEX)-treated rice straw. Appl Biochem Biotech. 2002;98:23-35.

119. Leskinen T, Kelley SS, Argyropoulos DS. E-beam irradiation & steam explosion as biomass pretreatment, and the complex role of lignin in substrate recalcitrance. Biomass Bioenerg. 2017;103:21-8.

120. Fei XH, Jia WB, Wang JQ, Chen T, Ling YS. Study on enzymatic hydrolysis efficiency and physicochemical properties of cellulose and lignocellulose after pretreatment with electron beam irradiation. Int J Biol Macromol. 2020;145:733-9.

121. Corro G, Pal U, Cebada S. Enhanced biogas production from coffee pulp through deligninocellulosic photocatalytic pretreatment. Energy Sci Eng. 2014;2(4):177-87.

122. Zhuang XS, Wang W, Yu Q, Qi W, Wang Q, Tan XS, et al. Liquid hot water pretreatment of lignocellulosic biomass for bioethanol production accompanying with high valuable products. Bioresource Technol. 2016;199:68-75.

123. Amini N, Haritos VS, Tanksale A. Microwave assisted pretreatment of eucalyptus sawdust enhances enzymatic saccharification and maximizes fermentable sugar yield. Renewable Energy. 2018;127:653-60.

124. Kucharska K, Rybarczyk P, Holowacz I, Lukajtis R, Glinka M, Kaminski M. Pretreatment of Lignocellulosic Materials as Substrates for Fermentation Processes. Molecules. 2018;23(11).

125. Yan X, Wang ZR, Zhang KJ, Si MY, Liu MR, Chai LY, et al. Bacteria-enhanced dilute acid pretreatment of lignocellulosic biomass. Bioresource Technol. 2017;245:419-25.

126. Bali G, Meng XZ, Deneff JI, Sun QN, Ragauskas AJ. The Effect of Alkaline Pretreatment Methods on Cellulose Structure and Accessibility. Chemsuschem. 2015;8(2):275-9.

127. Guragain YN, Bastola KP, Madl RL, Vadlani PV. Novel Biomass Pretreatment Using Alkaline Organic Solvents: A Green Approach for Biomass Fractionation and 2,3-Butanediol Production. Bioenerg Res. 2016;9(2):643-55.

128. Zhang QG, Hu JJ, Lee DJ. Pretreatment of biomass using ionic liquids: Research updates. Renewable Energy. 2017;111:77-84.

129. Zhang P, Dong SJ, Ma HH, Zhang BX, Wang YF, Hu XM. Fractionation of corn stover into cellulose, hemicellulose and lignin using a series of ionic liquids. Ind Crop Prod. 2015;76:688-96.

130. Si MY, Liu D, Liu MR, Yan X, Gao CJ, Chai LY, et al. Complementary effect of combined bacterial-chemical pretreatment to promote enzymatic digestibility of lignocellulose biomass. Bioresource Technol. 2019;272:275-80.

131. Rashid GMM, Duran-Pena MJ, Rahmanpour R, Sapsford D, Bugg TDH. Delignification and enhanced gas release from soil containing lignocellulose by treatment with bacterial lignin degraders. J Appl Microbiol. 2017;123(1):159-71.

132. Nichols NN, Sharma LN, Mowery RA, Chambliss CK, van Walsum GP, Dien BS, et al. Fungal metabolism of fermentation inhibitors present in corn stover dilute acid hydrolysate. Enzyme Microb Tech. 2008;42(7):624-30.

133. Anasontzis GE, Zerva A, Stathopoulou PM, Haralampidis K, Diallinas G, Karagouni AD, et al. Homologous overexpression of xylanase in *Fusarium oxysporum* increases ethanol productivity during consolidated bioprocessing (CBP) of lignocellulosics. J Biotechnol. 2011;152(1-2):16-23.

134. Bettiga M, Bengtsson O, Hahn-Hagerdal B, Gorwa-Grauslund MF. Arabinose and xylose fermentation by recombinant *Saccharomyces cerevisiae* expressing a fungal pentose utilization pathway. Microb Cell Fact. 2009;8.

135. Maleki F, Changizian M, Zolfaghari N, Rajaei S, Noghabi KA, Zahiri HS. Consolidated bioprocessing for bioethanol production by metabolically engineered *Bacillus subtilis* strains. Sci Rep-Uk. 2021;11(1).

136. He MX, Wu B, Qin H, Ruan ZY, Tan FR, Wang JL, et al. *Zymomonas mobilis*: a novel platform for future biorefineries. Biotechnol Biofuels. 2014;7.

137. Lin PP, Mi L, Moriok AH, Yoshino MM, Konishi S, Xu SC, et al. Consolidated bioprocessing of cellulose to isobutanol using *Clostridium thermocellum*. Metab Eng. 2015;31:44-52.

138. Salvachua D, Karp EM, Nimlos CT, Vardon DR, Beckham GT. Towards lignin consolidated bioprocessing: simultaneous lignin depolymerization and product generation by bacteria. Green Chem. 2015;17(11):4951-67.

139. Wang XP, Lin L, Dong JD, Ling J, Wang WP, Wang HL, et al. Simultaneous Improvements of *Pseudomonas* Cell Growth and Polyhydroxyalkanoate Production from a Lignin Derivative for Lignin-Consolidated Bioprocessing. Appl Environ Microb. 2018;84(18).

140. Lazuka A, Auer L, Bozonnet S, Morgavi DP, O'Donohue M, Hernandez-Raquet G. Efficient anaerobic transformation of raw wheat straw by a robust cow rumen-derived microbial consortium. Bioresource Technol. 2015;196:241-9.

141. Kong XP, Du J, Ye XM, Xi YL, Jin HM, Zhang M, et al. Enhanced methane production from wheat straw with the assistance of lignocellulolytic microbial consortium TC-5. Bioresource Technol. 2018;263:33-9.

142. Sun YQ, Li XY, Wei CX, Qi WB, Xiu ZL. An aptly industrialized bioprocess for lactic acid production from corn stover using thermotolerant microbial consortia. Bioproc Biosyst Eng. 2021;44(11):2445-54.

143. Puentes-Tellez PE, Salles JF. Construction of Effective Minimal Active Microbial Consortia for Lignocellulose Degradation. Microb Ecol. 2018;76(2):419-29.

144. Li GH, Chen HZ. Synergistic mechanism of steam explosion combined with fungal treatment by *Phellinus baumii* for the pretreatment of corn stalk. Biomass Bioenerg. 2014;67:1-7.

145. Voriskova J, Baldrian P. Fungal community on decomposing leaf litter undergoes rapid successional changes. Isme J. 2013;7(3):477-86.

146. van Erven G, Kleijn AF, Patyshakuliyeva A, Di Falco M, Tsang A, de Vries RP, et al. Evidence for ligninolytic activity of the ascomycete fungus *Podospora anserina*. Biotechnol Biofuels. 2020;13:75.

147. Kumar M, Verma S, Gazara RK, Kumar M, Pandey A, Verma PK, et al. Genomic and proteomic analysis of lignin degrading and polyhydroxyalkanoate accumulating β-proteobacterium *Pandoraea* sp. ISTKB. Biotechnol Biofuels. 2018;11:154.

148. Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature. 2010;464(7287):367-73.

149. Sabbadin F, Urresti S, Henrissat B, Avrova AO, Welsh LRJ, Lindley PJ, et al. Secreted pectin monooxygenases drive plant infection by pathogenic oomycetes. Science. 2021;373(6556):774-+.

150. Matsumoto R, Mehjabin JJ, Noguchi H, Miyamoto T, Takasuka TE, Hori C. Genomic and Secretomic Analyses of the Newly Isolated Fungus *Perenniporia fraxinea* SS3 Identified CAZymes Potentially Related to a Serious Pathogenesis of Hardwood Trees. Appl Environ Microb. 2023;89(5).

151. Bredon M, Dittmer J, Noel C, Moumen B, Bouchon D. Lignocellulose degradation at the holobiont level: teamwork in a keystone soil invertebrate. Microbiome. 2018;6.

152. Marynowska M, Goux X, Sillam-Dussès D, Rouland-Lefèvre C, Halder R, Wilmes P, et al. Compositional and functional characterisation of biomass-degrading microbial communities in guts of plant fibre- and soil-feeding higher termites. Microbiome. 2020;8(1):96.

153. Scully ED, Geib SM, Hoover K, Tien M, Tringe SG, Barry KW, et al. Metagenomic profiling reveals lignocellulose degrading system in a microbial community associated with a wood-feeding beetle. PLoS One. 2013;8(9):e73827.

154. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, et al. Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. Science. 2011;331(6016):463-7.

155. Chang L, Ding M, Bao L, Chen Y, Zhou J, Lu H. Characterization of a bifunctional xylanase/endoglucanase from yak rumen microorganisms. Appl Microbiol Biotechnol. 2011;90(6):1933-42.

156. Xu B, Xu W, Li J, Dai L, Xiong C, Tang X, et al. Metagenomic analysis of the *Rhinopithecus bieti* fecal microbiome reveals a broad diversity of bacterial and glycoside hydrolase profiles related to lignocellulose degradation. BMC Genomics. 2015;16:174.

157. Sabbadin F, Hemsworth GR, Ciano L, Henrissat B, Dupree P, Tryfona T, et al. An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion. Nat Commun. 2018;9(1):756.

158. Sabbadin F, Pesante G, Elias L, Besser K, Li Y, Steele-King C, et al. Uncovering the molecular mechanisms of lignocellulose digestion in shipworms. Biotechnol Biofuels. 2018;11:59.

159. Besser K, Malyon GP, Eborall WS, Paro da Cunha G, Filgueiras JG, Dowle A, et al. Hemocyanin facilitates lignocellulose digestion by wood-boring marine crustaceans. Nat Commun. 2018;9(1):5125.

160. Pellegrin C, Morin E, Martin FM, Veneault-Fourrey C. Comparative Analysis of Secretomes from Ectomycorrhizal Fungi with an Emphasis on Small-Secreted Proteins. Front Microbiol. 2015;6:1278.

161. Miao Y, Chen X, Li T, Zhu H, Tang S, Liu D, et al. Proteomic analysis reflects an environmental alkalinization-coupled pH-dependent mechanism of regulating lignocellulases in *Trichoderma guizhouense* NJAU4742. Biotechnol Biofuels. 2020;13:6.

162. Sista Kameshwar AK, Qin W. Comparative study of genome-wide plant biomass-degrading CAZymes in white rot, brown rot and soft rot fungi. Mycology. 2018;9(2):93-105.

163. Alfaro M, Castanera R, Lavin JL, Grigoriev IV, Oguiza JA, Ramirez L, et al. Comparative and transcriptional analysis of the predicted secretome in the lignocellulose-degrading basidiomycete fungus *Pleurotus ostreatus*. Environ Microbiol. 2016;18(12):4710-26.

164. Lopez MJ, Vargas-García MdC, Suárez-Estrella F, Nichols NN, Dien BS, Moreno J. Lignocellulose-degrading enzymes produced by the ascomycete *Coniochaeta ligniaria* and related species: Application for a lignocellulosic substrate treatment. Enzyme Microb Technol. 2007;40(4):794-800.

165. Presley GN, Zhang JW, Schilling JS. A genomics-informed study of oxalate and cellulase regulation by brown rot wood-degrading fungi. Fungal Genet Biol. 2018;112:64-70.

166. Arantes V, Milagres AMF. Degradation of cellulosic and hemicellulosic substrates using a chelator-mediated Fenton reaction. J Chem Technol Biotechnol. 2006;81(3):413-9.

167. Umezawa K, Niikura M, Kojima Y, Goodell B, Yoshida M. Transcriptome analysis of the brown rot fungus *Gloeophyllum trabeum* during lignocellulose degradation. Plos One. 2020;15(12).

168. Li F, Ma F, Zhao H, Zhang S, Wang L, Zhang X, et al. A Lytic Polysaccharide Monooxygenase from a White-Rot Fungus Drives the Degradation of Lignin by a Versatile Peroxidase. Appl Environ Microbiol. 2019;85(9).

169. Fernandez-Fueyo E, Ruiz-Duenas FJ, Ferreira P, Floudas D, Hibbett DS, Canessa P, et al. Comparative genomics of *Ceriporiopsis subvermispora* and *Phanerochaete chrysosporium* provide insight into selective ligninolysis. P Natl Acad Sci USA. 2012;109(14):5458-63.

170. Nghi DH, Bittner B, Kellner H, Jehmlich N, Ullrich R, Pecyna MJ, et al. The Wood Rot Ascomycete *Xylaria polymorpha* Produces a Novel GH78 Glycoside Hydrolase That Exhibits alpha-L-Rhamnosidase and Feruloyl Esterase Activities and Releases Hydroxycinnamic Acids from Lignocelluloses. Appl Environ Microb. 2012;78(14):4893-901.

171. Qian Y, Zhong L, Sun Y, Sun N, Zhang L, Liu W, et al. Enhancement of Cellulase Production in *Trichoderma reesei* via Disruption of Multiple Protease Genes Identified by Comparative Secretomics. Front Microbiol. 2019;10:2784.

172. Filiatrault-Chastel C, Navarro D, Haon M, Grisel S, Herpoel-Gimbert I, Chevret D, et al. AA16, a new lytic polysaccharide monooxygenase family identified in fungal secretomes. Biotechnol Biofuels. 2019;12.

173. Demers JE, Gugino BK, Jimenez-Gasco MD. Highly Diverse Endophytic and Soil *Fusarium oxysporum* Populations Associated with Field-Grown Tomato Plants. Appl Environ Microb. 2015;81(1):81-90.

174. Nirmaladevi D, Venkataramana M, Srivastava RK, Uppalapati SR, Gupta VK, Yli-Mattila T, et al. Molecular phylogeny, pathogenicity and toxigenicity of *Fusarium oxysporum* f. sp *lycopersici*. Sci Rep-Uk. 2016;6.

175. Riley R, Salamov AA, Brown DW, Nagy LG, Floudas D, Held BW, et al. Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. Proc Natl Acad Sci U S A. 2014;111(27):9923-8.

176. Henrissat B, Claeyssens M, Tomme P, Lemesle L, Mornon JP. Cellulase Families Revealed by Hydrophobic Cluster-Analysis. Gene. 1989;81(1):83-95.

177. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. 2022;50(D1):D571-D7.

178. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018;46(W1):W95-W101.

179. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnol Biofuels. 2013;6(1):41.

180. Jagadeeswaran G, Gainey L, Prade R, Mort AJ. A family of AA9 lytic polysaccharide monooxygenases in *Aspergillus nidulans* is differentially regulated by multiple substrates and at least one is active on cellulose and xyloglucan. Appl Microbiol Biot. 2016;100(10):4535-47.

181. Monclaro AV, Petrovic DM, Alves GSC, Costa MMC, Midorikawa GEO, Miller RNG, et al. Characterization of two family AA9 LPMOs from *Aspergillus tamarii* with distinct activities on xyloglucan reveals structural differences linked to cleavage specificity. Plos One. 2020;15(7).

182. Kracher D, Scheiblbrandner S, Felice AKG, Breslmayr E, Preims M, Ludwicka K, et al. Extracellular electron transfer systems fuel cellulose oxidative degradation. Science. 2016;352(6289):1098-101.

183. Bissaro B, Varnai A, Rohr AK, Eijsink VGH. Oxidoreductases and Reactive Oxygen Species in Conversion of Lignocellulosic Biomass. Microbiol Mol Biol R. 2018;82(4).

184. Nekiunaite L, Arntzen MO, Svensson B, Vaaje-Kolstad G, Abou Hachem M. Lytic polysaccharide monooxygenases and other oxidative enzymes are abundantly secreted by *Aspergillus nidulans* grown on different starches. Biotechnol Biofuels. 2016;9.

185. Sutzl L, Laurent CVFP, Abrera AT, Schutz G, Ludwig R, Haltrich D. Multiplicity of enzymatic functions in the CAZy AA3 family. Appl Microbiol Biot. 2018;102(6):2477-92.

186. Várnai A, Umezawa K, Yoshida M, Eijsink VGH. The Pyrroloquinoline-Quinone-Dependent Pyranose Dehydrogenase from *Coprinopsis cinerea* Drives Lytic Polysaccharide Monooxygenase Action. Appl Environ Microbiol. 2018;84(11).

187. Kont R, Bissaro B, Eijsink VGH, Valjamae P. Kinetic insights into the peroxygenase activity of cellulose-active lytic polysaccharide monooxygenases (LPMOs). Nat Commun. 2020;11(1).

188. Wang BJ, Walton PH, Rovira C. Molecular Mechanisms of Oxygen Activation and Hydrogen Peroxide Formation in Lytic Polysaccharide Monooxygenases. Acs Catal. 2019;9(6):4958-69.

189. Ferraroni M, Westphal AH, Borsari M, Tamayo-Ramos JA, Briganti F, de Graaff LH, et al. Structure and function of *Aspergillus niger* laccase McoG. Biocatalysis. 2017;3(1):1-21.

190. Zhou M, Fakayode OA, Ren MN, Li HX, Liang JK, Yagoub AEA, et al. Laccase-catalyzed lignin depolymerization in deep eutectic solvents: challenges and prospects. Bioresources and Bioprocessing. 2023;10(1).

191. Lassouane F, Ait-Amar H, Amrani S, Rodriguez-Couto S. A promising laccase immobilization approach for Bisphenol A removal from aqueous solutions. Bioresource Technol. 2019;271:360-7.

192. Hilgers R, Vincken JP, Gruppen H, Kabel MA. Laccase/Mediator Systems: Their Reactivity toward Phenolic Lignin Structures. Acs Sustain Chem Eng. 2018;6(2):2037-46.

193. Brenelli L, Squina FM, Felby C, Cannella D. Laccase-derived lignin compounds boost cellulose oxidative enzymes AA9. Biotechnol Biofuels. 2018;11.

194. Zhu Y, Plaza N, Kojima Y, Yoshida M, Zhang JW, Jellison J, et al. Nanostructural Analysis of Enzymatic and Non-enzymatic Brown Rot Fungal Deconstruction of the Lignocellulose Cell Wall. Front Microbiol. 2020;11.

195. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, et al. The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes. Science. 2012;336(6089):1715-9.

196. Janusz G, Pawlik A, Sulej J, Swiderska-Burek U, Jarosz-Wilkolazka A, Paszczynski A. Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. FEMS Microbiol Rev. 2017;41(6):941-62.

197. Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, Aerts A, et al. The Plant Cell Wall-Decomposing Machinery Underlies the Functional Diversity of Forest Fungi. Science. 2011;333(6043):762-5.

198. Henriksson G, Johansson G, Pettersson G. Is Cellobiose Oxidase from *Phanerochaete chrysosporium* a One-Electron Reductase. Biochim Biophys Acta. 1993;1144(2):184-90.

199. Brenelli LB, Persinoti GF, Cairo JPLF, Liberato MV, Goncalves TA, Otero IVR, et al. Novel redox-active enzymes for ligninolytic applications revealed from multiomics analyses of *Peniophora* sp. CBMAI 1063, a laccase hyper-producer strain. Sci Rep-Uk. 2019;9.

200. Xu CF, Su X, Wang JH, Zhang FZ, Shen GN, Yuan Y, et al. Characteristics and functional bacteria in a microbial consortium for rice straw lignin-degrading. Bioresource Technol. 2021;331.

201. Park YJ, Jeong YU, Kong WS. Genome Sequencing and Carbohydrate-Active Enzyme (CAZyme) Repertoire of the White Rot Fungus *Flammulina elastica*. Int J Mol Sci. 2018;19(8).

202. Zhao ZT, Liu HQ, Wang CF, Xu JR. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. Bmc Genomics. 2013;14.

203. Hansson H, Karkehabadi S, Mikkelsen N, Douglas NR, Kim S, Lam A, et al. High-resolution structure of a lytic polysaccharide monooxygenase from *Hypocrea jecorina* reveals a predicted linker as an integral part of the catalytic domain. J Biol Chem. 2017;292(46):19099-109.

204. Bennati-Granier C, Garajova S, Champion C, Grisel S, Haon M, Zhou S, et al. Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by *Podospora anserina*. Biotechnol Biofuels. 2015;8.

205. Hervé C, Rogowski A, Blake AW, Marcus SE, Gilbert HJ, Knox JP. Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. Proc Natl Acad Sci U S A. 2010;107(34):15293-8.

206. Vaaje-Kolstad G, Horn SJ, van Aalten DMF, Synstad B, Eijsink VGH. The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. J Biol Chem. 2005;280(31):28492-7.

207. Gilbert HJ, Knox JP, Boraston AB. Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. Curr Opin Struc Biol. 2013;23(5):669-77.

208. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J. 2004;382(Pt 3):769-81.

209. Manjeet K, Madhuprakash J, Mormann M, Moerschbacher BM, Podile AR. A carbohydrate binding module-5 is essential for oxidative cleavage of chitin by a multi-modular lytic polysaccharide monooxygenase from *Bacillus thuringiensis* serovar *kurstaki*. Int J Biol Macromol. 2019;127:649-56.

210. Hurtado-Guerrero R, Schuttelkopf AW, Mouyna I, Ibrahim AFM, Shepherd S, Fontaine T, et al. Molecular Mechanisms of Yeast Cell Wall Glucan Remodeling. J Biol Chem. 2009;284(13):8461-9.

211. Krishnan B, Srivastava SS, Sankeshi V, Garg R, Srivastava S, Sankaranarayanan R, et al. beta gamma-Crystallination Endows a Novel Bacterial Glycoside Hydrolase 64 with CA(2+)-Dependent Activity Modulation. J Bacteriol. 2019;201(23).

212. Dilokpimol A, Verkerk B, Li XX, Bellemare A, Lavallee M, Frommhagen M, et al. Screening of novel fungal Carbohydrate Esterase family 1 enzymes identifies three novel dual feruloyl/acetyl xylan esterases. Febs Lett. 2022;596(15):1932-43.

213. Sista Kameshwar AK, Qin W. Understanding the structural and functional properties of carbohydrate esterases with a special focus on hemicellulose deacetylating acetyl xylan esterases. Mycology. 2018;9(4):273-95.

214. Raji O, Baath JA, Vuong TV, Larsbrink J, Olsson L, Master ER. The coordinated action of glucuronoyl esterase and alpha-glucuronidase promotes the disassembly of lignin-carbohydrate complexes. Febs Lett. 2021;595(3):351-9.

215. Agger JW, Busk PK, Pilgaard B, Meyer AS, Lange L. A New Functional Classification of Glucuronoyl Esterases by Peptide Pattern Recognition. Front Microbiol. 2017;8.

216. Arnling Bååth J, Mazurkewich S, Knudsen RM, Poulsen J-CN, Olsson L, Lo Leggio L, et al. Biochemical and structural features of diverse bacterial glucuronoyl esterases facilitating recalcitrant biomass conversion. Biotechnol Biofuels. 2018;11:213.

217. Zhao Y, Shakeel U, Rehman MSU, Li HQ, Xu X, Xu J. Lignin-carbohydrate complexes (LCCs) and its role in biorefinery. J Clean Prod. 2020;253.

218. de Vries S, de Vries J. A Global Survey of Carbohydrate Esterase Families 1 and 10 in Oomycetes. Front Genet. 2020;11.

219. Islam MS, Haque MS, Islam MM, Emdad EM, Halim A, Hossen QMM, et al. Tools to kill: Genome of one of the most destructive plant pathogenic fungi *Macrophomina phaseolina*. Bmc Genomics. 2012;13.

220. Arntzen MØ, Bengtsson O, Várnai A, Delogu F, Mathiesen G, Eijsink VGH. Quantitative comparison of the biomass-degrading enzyme repertoires of five filamentous fungi. Sci Rep. 2020;10(1):20267.

221. Amin K, Tranchimand S, Benvegnu T, Abdel-Razzak Z, Chamieh H. Glycoside Hydrolases and Glycosyltransferases from Hyperthermophilic Archaea: Insights on Their Characteristics and Applications in Biotechnology. Biomolecules. 2021;11(11).

222. Seppala S, Wilken SE, Knop D, Solomon KV, O'Malley MA. The importance of sourcing enzymes from non-conventional fungi for metabolic engineering and biomass breakdown. Metab Eng. 2017;44:45-59.

223. Murphy C, Powlowski J, Wu M, Butler G, Tsang A. Curation of characterized glycoside hydrolases of Fungal origin. Database-Oxford. 2011.

224. Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem. 2008;77:521-55.

225. King R, Urban M, Lauder RP, Hawkins N, Evans M, Plummer A, et al. A conserved fungal glycosyltransferase facilitates pathogenesis of plants by enabling hyphal growth on solid surfaces. Plos Pathog. 2017;13(10).

226. Bandi CK, Agrawal A, Chundawat SPS. Carbohydrate-Active enZyme (CAZyme) enabled glycoengineering for a sweeter future. Curr Opin Biotech. 2020;66:283-91.

227. Garron ML, Cygler M. Uronic polysaccharide degrading enzymes. Curr Opin Struc Biol. 2014;28:87-95.

228. van den Brink J, de Vries RP. Fungal enzyme sets for plant polysaccharide degradation. Appl Microbiol Biotechnol. 2011;91(6):1477-92.

229. Oates NC, Abood A, Schirmacher AM, Alessi AM, Bird SM, Bennett JP, et al. A multi-omics approach to lignocellulolytic enzyme discovery reveals a new ligninase activity from *Parascedosporium putredinis* NO1. Proc Natl Acad Sci U S A. 2021;118(18).

230. Gilgado F, Gene J, Cano J, Guarro J. Reclassification of *Graphium tectonae* as *Parascedosporium tectonae* gen. nov., comb. nov., *Pseudallescheria africana* as *Petriellopsis africana* gen. nov., comb. nov and *Pseudallescheria fimeti* as *Lophotrichus fimeti* comb. nov. Int J Syst Evol Micr. 2007;57:2171-8.

231. Lackner M, de Hoog GS. *Parascedosporium* and its relatives: phylogeny and ecological trends. Ima Fungus. 2011;2(1):39-48.

232. Perera RH, Hyde KD, Persoh D, Jones EBG, Liu JK, Liu ZY. Additions to wild seed and fruit fungi 1: The sexual morph of *Diaporthe rosae* on *Magnolia champaca* and *Senna siamea* fruits in Thailand. Mycosphere. 2018;9(2):256-70.

233. Alessi AM, Bird SM, Oates NC, Li Y, Dowle AA, Novotny EH, et al. Defining functional diversity for lignocellulose degradation in a microbial community using multi-omics studies. Biotechnol Biofuels. 2018;11.

234. Oates N. Mining microbial compost communities for lignocellulose degrading proteins.: University of York; 2016.

235. Marinovic M, Nousiainen P, Dilokpimol A, Kontro J, Moore R, Sipila J, et al. Selective Cleavage of Lignin beta-O-4 Aryl Ether Bond by beta-Etherase of the White-Rot Fungus *Dichomitus squalens*. Acs Sustain Chem Eng. 2018;6(3):2878-82.

236. Min K, Yum T, Kim J, Woo HM, Kim Y, Sang BI, et al. Perspectives for biocatalytic lignin utilization: cleaving 4-O-5 and C-alpha-C-beta bonds in dimeric lignin model compounds catalyzed by a promiscuous activity of tyrosinase. Biotechnol Biofuels. 2017;10.

237. Li M, Pu YQ, Yoo CG, Ragauskas AJ. The occurrence of tricin and its derivatives in plants. Green Chem. 2016;18(6):1439-54.

238. Poidevin L, Berrin JG, Bennati-Granier C, Levasseur A, Herpoel-Gimbert I, Chevret D, et al. Comparative analyses of *Podospora anserina* secretomes reveal a large array of lignocellulose-active enzymes. Appl Microbiol Biot. 2014;98(17):7457-69.

239. Banerjee G, Scott-Craig JS, Walton JD. Improving Enzymes for Biomass Conversion: A Basic Research Perspective. Bioenerg Res. 2010;3(1):82-92.

240. Contreras F, Pramanik S, Rozhkova AM, Zorov IN, Korotkova O, Sinitsyn AP, et al. Engineering Robust Cellulases for Tailored Lignocellulosic Degradation Cocktails. Int J Mol Sci. 2020;21(5).

241. Qing Q, Wyman CE. Hydrolysis of different chain length xylooliogmers by cellulase and hemicellulase. Bioresource Technol. 2011;102(2):1359-66.

242. Van Dyk JS, Pletschke BI. A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes-Factors affecting enzymes, conversion and synergy. Biotechnol Adv. 2012;30(6):1458-80.

243. Karnaouri A, Matsakas L, Topakas E, Rova U, Christakopoulos P. Development of Thermophilic Tailor-Made Enzyme Mixtures for the Bioconversion of Agricultural and Forest Residues. Front Microbiol. 2016;7.

244. Cianchetta S, Bregoli L, Galletti S. Microplate-Based Evaluation of the Sugar Yield from Giant Reed, Giant Miscanthus and Switchgrass after Mild Chemical Pre-Treatments and Hydrolysis with Tailored *Trichoderma* Enzymatic Blends. Appl Biochem Biotech. 2017;183(3):876-92.

245. Kim IJ, Lee HJ, Kim KH. Pure enzyme cocktails tailored for the saccharification of sugarcane bagasse pretreated by using different methods. Process Biochem. 2017;57:167-74.

246. Su YJ, Yu XX, Sun Y, Wang G, Chen H, Chen G. Evaluation of Screened Lignin-degrading Fungi for the Biological Pretreatment of Corn Stover. Sci Rep-Uk. 2018;8.

247. Witzler M, Alzagameem A, Bergs M, El Khaldi-Hansen B, Klein SE, Hielscher D, et al. Lignin-Derived Biomaterials for Drug Release and Tissue Engineering. Molecules. 2018;23(8).

248. Longoria A, Tinoco R, Vazquez-Duhalt R. Chloroperoxidase-mediated transformation of highly halogenated monoaromatic compounds. Chemosphere. 2008;72(3):485-90.

249. Ortiz-Bermudez P, Hirth KC, Srebotnik E, Hammel KE. Chlorination of lignin by ubiquitous fungi has a likely role in global organochlorine production. P Natl Acad Sci USA. 2007;104(10):3895-900.

250. Ortiz-Bermudez P, Srebotnik E, Hammel KE. Chlorination and cleavage of lignin structures by fungal chloroperoxidases. Appl Environ Microb. 2003;69(8):5015-8.

251. Kinne M, Poraj-Kobielska M, Ralph SA, Ullrich R, Hofrichter M, Hammel KE. Oxidative Cleavage of Diverse Ethers by an Extracellular Fungal Peroxygenase. J Biol Chem. 2009;284(43):29343-9.

252. Gronqvist S, Viikari L, Niku-Paavola ML, Orlandi M, Canevali C, Buchert J. Oxidation of milled wood lignin with laccase, tyrosinase and horseradish peroxidase. Appl Microbiol Biot. 2005;67(4):489-94.

253. Otsuka Y, Sonoki T, Ikeda S, Kajita S, Nakamura M, Katayama Y. Detection and characterization of a novel extracellular fungal enzyme that catalyzes the specific and hydrolytic cleavage of lignin guaiacylglycerol beta-aryl ether linkages. Eur J Biochem. 2003;270(11):2353-62.

254. Pearson WR. An introduction to sequence similarity ("homology") searching. Curr Protoc Bioinformatics. 2013;Chapter 3:3.1.-3.1.8.

255. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. Bmc Bioinformatics. 2010;11.

256. Voss H, Heck CA, Schallmey M, Schallmey A. Database Mining for Novel Bacterial beta-Etherases, Glutathione-Dependent Lignin-Degrading Enzymes. Appl Environ Microb. 2020;86(2).

257. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021.

258. Cong BL, Wang NF, Liu SH, Liu F, Yin XF, Shen JH. Isolation, characterization and transcriptome analysis of a novel Antarctic *Aspergillus sydowii* strain MS-19 as a potential lignocellulosic enzyme source. Bmc Microbiol. 2017;17.

259. Tolgo M, Huttner S, Rugbjerg P, Thuy NT, Thanh VN, Larsbrink J, et al. Genomic and transcriptomic analysis of the thermophilic lignocellulose-degrading fungus *Thielavia terrestris* LPH172. Biotechnol Biofuels. 2021;14(1).

260. Miyauchi S, Navarro D, Grisel S, Chevret D, Berrin JG, Rosso MN. The integrative omics of white-rot fungus *Pycnoporus coccineus* reveals co-regulated CAZymes for orchestrated lignocellulose breakdown. Plos One. 2017;12(4).

261. Heeger F, Bourne EC, Wurzbacher C, Funke E, Lipzen A, He GF, et al. Evidence for Lignocellulose-Decomposing Enzymes in the Genome and Transcriptome of the Aquatic Hyphomycete *Clavariopsis aquatica*. J Fungi. 2021;7(10).

262. Guo HL, Wang XD, Lee DJ. Proteomic researches for lignocellulose-degrading enzymes: A mini-review. Bioresource Technol. 2018;265:532-41.

263. Salvachua D, Martinez AT, Tien M, Lopez-Lucendo MF, Garcia F, de los Rios V, et al. Differential proteomic analysis of the secretome of *Irpex lacteus* and other white-rot fungi during wheat straw pretreatment. Biotechnol Biofuels. 2013;6.

264. Moremen KW, Haltiwanger RS. Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. Nat Chem Biol. 2019;15(9):853-64.

265. Andlar M, Rezić T, Marđetko N, Kracher D, Ludwig R, Šantek B. Lignocellulose degradation: An overview of fungi and fungal enzymes involved in lignocellulose degradation. Eng Life Sci. 2018;18(11):768-78.

266. Yamamoto M, Tomiyama H, Koyama A, Okuizumi H, Liu S, Vanholme R, et al. A century-old mystery unveiled: Sekizaisou is a natural lignin mutant. Plant Physiol. 2020.

267. Kameshwar AKS, Qin W. Molecular Networks of *Postia placenta* Involved in Degradation of Lignocellulosic Biomass Revealed from Metadata Analysis of Open Access Gene Expression Data. Int J Biol Sci. 2018;14(3):237-52.

268. Tørresen OK, Star B, Mier P, Andrade-Navarro MA, Bateman A, Jarnot P, et al. Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. Nucleic Acids Res. 2019;47(21):10994-1006.

269. Momeni MH, Fredslund F, Bissaro B, Raji O, Vuong TV, Meier S, et al. Discovery of fungal oligosaccharide-oxidising flavo-enzymes with previously unknown substrates, redox-activity profiles and interplay with LPMOs. Nat Commun. 2021;12(1).

270. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2019;47(D1):D427-D32.

271. Petrovic DM, Bissaro B, Chylenski P, Skaugen M, Sorlie M, Jensen MS, et al. Methylation of the N-terminal histidine protects a lytic polysaccharide monooxygenase from auto-oxidative inactivation. Protein Sci. 2018;27(9):1636-50.

272. Varadi M, Velankar S. The impact of AlphaFold Protein Structure Database on the fields of life sciences. Proteomics. 2022.

273. Pruitt KD, Tatusova T, Maglott DR. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2005;33(Database issue):D501-4.

274. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-42.

275. Teufel F, Armenteros JJA, Johansen AR, Gislason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40(7):1023-+.

276. Garcia-Santamarina S, Probst C, Festa RA, Ding C, Smith AD, Conklin SE, et al. A lytic polysaccharide monooxygenase-like protein functions in fungal copper import and meningitis. Nat Chem Biol. 2020;16(3):337-+.

277. Arora R, Bharval P, Sarswati S, Sen TZ, Yennamalli RM. Structural dynamics of lytic polysaccharide monoxygenases reveals a highly flexible substrate binding region. J Mol Graph Model. 2019;88:1-10.

278. Sirim D, Wagner F, Wang L, Schmid RD, Pleiss J. The Laccase Engineering Database: a classification and analysis system for laccases and related multicopper oxidases. Database. 2011;2011:bar006.

279. Pardo I, Rodriguez-Escribano D, Aza P, de Salas F, Martinez AT, Camarero S. A highly stable laccase obtained by swapping the second cupredoxin domain. Sci Rep-Uk. 2018;8.

280. Boulanger MJ, Murphy MEP. Crystal structure of the soluble domain of the major anaerobically induced outer membrane protein (AniA) from pathogenic *Neisseria*: A new class of copper-containing nitrite reductases. J Mol Biol. 2002;315(5):1111-27.

281. Matsuoka M, Kumar A, Muddassar M, Matsuyama A, Yoshida M, Zhang KYJ. Discovery of Fungal Denitrification Inhibitors by Targeting Copper Nitrite Reductase from *Fusarium oxysporum*. J Chem Inf Model. 2017;57(2):203-13.

282. Choi J, Detry N, Kim KT, Asiegbu FO, Valkonen JPT, Lee YH. fPoxDB: fungal peroxidase database for comparative genomics. Bmc Microbiol. 2014;14.

283. Craig JP, Coradetti ST, Starr TL, Glass NL. Direct Target Network of the *Neurospora crassa* Plant Cell Wall Deconstruction Regulators CLR-1, CLR-2, and XLR-1. Mbio. 2015;6(5).

284. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One. 2016;11(10):e0163962.

285. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722-36.

286. Davey JW, Catta-Preta CMC, James S, Forrester S, Motta MCM, Ashton PD, et al. Chromosomal assembly of the nuclear genome of the endosymbiont-bearing trypanosomatid *Angomonas deanei*. G3. 2021;11(1).

287. Andrews S. FastQC: a quality control tool for high throughput sequence data. Cambridge, UK: Babraham Institute. 2011.

288. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17(1):10-2.

289. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9(11):e112963.

290. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236-40.

291. Palmer JMS, Jason. Funannotate v.1.8.1: Eukaryotic genome annotation. 2020.

292. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the National Center for Biotechnology Information in 2023. Nucleic Acids Res. 2022.

293. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2022.

294. Villanueva RAM, Chen ZJ. ggplot2: Elegant Graphics for Data Analysis, 2nd edition. Meas-Interdiscip Res. 2019;17(3):160-7.

295. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST plus : architecture and applications. Bmc Bioinformatics. 2009;10.

296. Lassmann T. Kalign 3: multiple sequence alignment of large data sets. Bioinformatics. 2019.

297. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic Acids Res. 2018;46(W1):W200-W4.

298. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D. 2004;60:2256-68.

299. UniProt C. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47(D1):D506-D15.

300. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, et al. InterPro in 2022. Nucleic Acids Res. 2022.

301. Scott CJR. Whole genome structural predictions reveal hidden diversity in putative oxidative enzymes of the lignocellulose degrading ascomycete *Parascedosporium putredinis* NO1. biorRxiv. 2023.

302. Nielsen H. Predicting Secretory Proteins with SignalP. In: Kihara D, editor. Protein Function Prediction: Methods and Protocols. New York, NY: Springer New York; 2017. p. 59-73.

303. Miura N, Ueda M. Evaluation of Unconventional Protein Secretion by *Saccharomyces cerevisiae* and other Fungi. Cells-Basel. 2018;7(9).

304. Rodrigues ML, Nosanchuk JD, Schrank A, Vainstein MH, Casadevall A, Nimrichter L. Vesicular transport systems in fungi. Future Microbiol. 2011;6(11):1371-81.

305. Gogleva A, Drost HG, Schornack S. SecretSanta: flexible pipelines for functional secretome prediction. Bioinformatics. 2018;34(13):2295-6.

306. de Paula RG, Antonieto ACC, Nogueira KMV, Ribeiro LFC, Rocha MC, Malavazi I, et al. Extracellular vesicles carry cellulases in the industrial fungus *Trichoderma reesei*. Biotechnol Biofuels. 2019;12.

307. Artzi L, Bayer EA, Morais S. Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. Nat Rev Microbiol. 2017;15(2):83-95.

308. Krautter F, Iqbal AJ. Glycans and Glycan-Binding Proteins as Regulators and Potential Targets in Leukocyte Recruitment. Front Cell Dev Biol. 2021;9.

309. Savojardo C, Martelli PL, Fariselli P, Profiti G, Casadio R. BUSCA: an integrative web server to predict subcellular localization of proteins. Nucleic Acids Res. 2018;46(W1):W459-W66.

310. Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics. 2017;33(21):3387-95.

311. Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A, et al. Detecting sequence signals in targeting peptides using deep learning. Life Sci Alliance. 2019;2(5).

312. Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel. 2004;17(4):349-56.

313. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567-80.

314. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021;38(12):5825-9.

315. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12(85):2825-30.

316. Manavalan T, Manavalan A, Thangavelu KP, Heese K. Secretome analysis of *Ganoderma lucidum* cultivated in sugarcane bagasse. J Proteomics. 2012;77:298-309.

317. Salvachúa D, Prieto A, López-Abelairas M, Lu-Chau T, Martínez AT, Martínez MJ. Fungal pretreatment: An alternative in second-generation ethanol from wheat straw. Bioresour Technol. 2011;102(16):7500-6.

318. Scott CJR, Leadbeater DR, Bruce NC. A bioinformatic workflow for *in silico* secretome prediction with the lignocellulose degrading ascomycete fungus *Parascedosporium putredinis* NO1. Molecular Microbiology. 2023.

319. Merali Z, Collins SRA, Elliston A, Wilson DR, Käsper A, Waldron KW. Characterization of cell wall components of wheat bran following hydrothermal pretreatment and fractionation. Biotechnol Biofuels. 2015;8:23.

320. Crestini C, Lange H, Sette M, Argyropoulos DS. On the structure of softwood kraft lignin. Green Chem. 2017;19(17):4104-21.

321. Mayans O, Scott M, Connerton I, Gravesen T, Benen J, Visser J, et al. Two crystal structures of pectin lyase A from *Aspergillus* reveal a pH driven conformational change and striking divergence in the substrate-binding clefts of pectin and pectate lyases. Structure. 1997;5(5):677-89.

322. Metz B, Seidl-Seiboth V, Haarmann T, Kopchinskiy A, Lorenz P, Seiboth B, et al. Expression of Biomass-Degrading Enzymes Is a Major Event during Conidium Development in *Trichoderma reesei*. Eukaryot Cell. 2011;10(11):1527-35.

323. Jiang JB, Beenakker TJM, Kallemeijn WW, van der Marel GA, van den Elst H, Codee JDC, et al. Comparing Cyclophellitol N-Alkyl and N-Acyl Cyclophellitol Aziridines as Activity-Based Glycosidase Probes. Chem-Eur J. 2015;21(30):10861-9.

324. de Boer C, McGregor NGS, Peterse E, Schroder SP, Florea BI, Jiang JB, et al. Glycosylated cyclophellitol-derived activity-based probes and inhibitors for cellulases. Rsc Chem Biol. 2020;1(3):148-55.

325. Schroder SP, de Boer C, McGregor NGS, Rowland RJ, Moroz O, Blagova E, et al. Dynamic and Functional Profiling of Xylan-Degrading Enzymes in *Aspergillus* Secretomes Using Activity-Based Probes. Acs Central Sci. 2019;5(6):1067-78.

326. McGregor NGS, de Boer C, Santos M, Haon M, Navarro D, Schroder S, et al. Activity-based protein profiling reveals dynamic substrate-specific cellulase secretion by saprotrophic basidiomycetes. Biotechnol Biof Biop. 2022;15(1).

327. Wu L, Armstrong Z, Schroder SP, de Boer C, Artola M, Aerts JMFG, et al. An overview of activity-based probes for glycosidases. Curr Opin Chem Biol. 2019;53:25-36.

328. Singhania RR, Sukumaran RK, Pandey A. Improved cellulase production by *Trichoderma reesei* RUT C30 under SSF through process optimization. Appl Biochem Biotech. 2007;142(1):60-70.

329. Verma N, Kumar V. Impact of process parameters and plant polysaccharide hydrolysates in cellulase production by *Trichoderma reesei* and *Neurospora crassa* under wheat bran based solid state fermentation. Biotechnol Rep (Amst). 2020;25:e00416.

330. Zhang XY, Chen XH, Li SS, Bello A, Liu JW, Gao LY, et al. Mechanism of differential expression of beta-glucosidase genes in functional microbial communities in response to carbon catabolite repression. Biotechnol Biof Biop. 2022;15(1).

331. Ottenheim C, Verdejo C, Zimmermann W, Wu JC. Hemicellulase production by *Aspergillus niger* DSM 26641 in hydrothermal palm oil empty fruit bunch hydrolysate and transcriptome analysis. J Biosci Bioeng. 2014;118(6):696-701.

332. Uke A, Sornyotha S, Baramee S, Tachaapaikoon C, Pason P, Waeonukul R, et al. Genomic analysis of *Paenibacillus macerans* strain I6, which can effectively saccharify oil palm empty fruit bunches under nutrient-free conditions. J Biosci Bioeng. 2023;136(1):1-6.

333. Hutner SP, L; Schatz, A; Haskins, C. Some approaches to the study of the role of metals in the metabolism of microorganisms. Proceedings of the American Philosophical Society. 1950;94:152-70.

334. Le S, Josse J, Husson F. FactoMineR: An R package for multivariate analysis. J Stat Softw. 2008;25(1):1-18.

335. Kassambra A, Mundt F. Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R Package Version 1.0.7 ed2020.

336. de Vries A, Ripley BD. ggdendro: Create Dendrograms and Tree Diagrams Using 'ggplot2'. R package version 0.1.23 ed2022.

337. Galili T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics. 2015;31(22):3718-20.

338. Gao CH, Yu GC, Cai P. ggVennDiagram: An Intuitive, Easy-to-Use, and Highly Customizable R Package to Generate Venn Diagram. Front Genet. 2021;12.

339. Kolde R. Pheatmap: pretty heatmaps. R package version. 2012.

340. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008;36(10):3420-35.

341. Bioinformatics B. OmicsBox - Bioinformatics Made Easy. 2019.

342. Zhang ST, Dong ZJ, Shi J, Yang CR, Fang Y, Chen G, et al. Enzymatic hydrolysis of corn stover lignin by laccase, lignin peroxidase, and manganese peroxidase. Bioresource Technol. 2022;361.

343. Sánchez-Ruiz MI, Ayuso-Fernández I, Rencoret J, González-Ramírez AM, Linde D, Davó-Siguero I, et al. *Agaricales* Mushroom Lignin Peroxidase: From Structure-Function to Degradative Capabilities. Antioxidants (Basel). 2021;10(9).

344. Fernández-Fueyo E, Ruiz-Dueñas FJ, Martínez MJ, Romero A, Hammel KE, Medrano FJ, et al. Ligninolytic peroxidase genes in the oyster mushroom genome: heterologous expression, molecular structure, catalytic and stability properties, and lignin-degrading ability. Biotechnol Biofuels. 2014;7(1):2.

345. Hilgers R, Vincken J-P, Gruppen H, Kabel MA. Laccase/Mediator Systems: Their Reactivity toward Phenolic Lignin Structures. Acs Sustain Chem Eng. 2018;6(2):2037-46.

346. Scott CJR. *Parascedosporium putredinis* NO1 tailors its secretome for different lignocellulosic substrates [Manuscript submitted for publication]. 2023.

347. Nan Y, Jia L, Yang M, Xin D, Qin Y, Zhang J. Simplified sodium chlorite pretreatment for carbohydrates retention and efficient enzymatic saccharification of silvergrass. Bioresour Technol. 2018;261:223-31.

348. Jin K, Liu X, Jiang Z, Tian G, Yang S, Shang L, et al. Delignification kinetics and selectivity in poplar cell wall with acidified sodium chlorite. Ind Crops Prod. 2019;136:87-92.

349. Siqueira G, Várnai A, Ferraz A, Milagres AMF. Enhancement of cellulose hydrolysis in sugarcane bagasse by the selective removal of lignin with sodium chlorite. Appl Energy. 2013;102:399-402.

350. Han Y, Bai Y, Zhang J, Liu D, Zhao X. A comparison of different oxidative pretreatments on polysaccharide hydrolyzability and cell wall structure for interpreting the greatly improved enzymatic digestibility of sugarcane bagasse by delignification. Bioresources and Bioprocessing. 2020;7(1):24.

351. Lu Y, Lu Y-C, Hu H-Q, Xie F-J, Wei X-Y, Fan X. Structural Characterization of Lignin and Its Degradation Products with Spectroscopic Methods. Journal of Spectroscopy. 2017;2017.

352. Kang X, Zhang Y, Li L, Sun Y, Kong X, Yuan Z. Enhanced methane production from anaerobic digestion of hybrid *Pennisetum* by selectively removing lignin with sodium chlorite. Bioresour Technol. 2020;295:122289.

353. Zhang J, Wang Y, Du X, Qu Y. Selective removal of lignin to enhance the process of preparing fermentable sugars and platform chemicals from lignocellulosic biomass. Bioresour Technol. 2020;303:122846.

354. Hansted ALS, Cacuro TA, Nakashima GT, Costa VE, Yamamoto H, Yamaji FM. Use of a lignocellulosic residue as solid fuel: The effect of ash content in the energy potential. Ind Crop Prod. 2018;116:209-14.

355. Xu F, Sun JX, Sun RC, Fowler P, Baird MS. Comparative study of organosolv fignins from wheat straw. Ind Crop Prod. 2006;23(2):180-93.

356. Sundaramoorthy M, Terner J, Poulos TL. The crystal structure of chloroperoxidase: A heme peroxidase-cytochrome P450 functional hybrid. Structure. 1995;3(12):1367-77.

357. Safder I, Khan S, Islam I-U, Ali MK, Bibi Z, Waqas M. *Pichia pastoris* expression system: a potential candidate to express protein in industrial and biopharmaceutical domains. Biomed Lett. 2018;4(1):1-14.

358. Erdocia X, Prado R, Corcuera MA, Labidi J. Effect of different organosolv treatments on the structure and properties of olive tree pruning lignin. J Ind Eng Chem. 2014;20(3):1103-8.

359. Klein J, Kupec R, Stockl M, Waldvogel SR. Degradation of Lignosulfonate to Vanillic Acid Using Ferrate. Adv Sustain Syst. 2023;7(4).

360. Hibino T, Shibata D, Ito T, Tsuchiya D, Higuchi T, Pollet B, et al. Chemical properties of lignin from *Aralia cordata*. Phytochemistry. 1994;37(2):445-8.

361. Pettersen EF, Goddard TD, Huang CRC, Meng EEC, Couch GS, Croll TI, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 2021;30(1):70-82.

362. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970;227(5259):680-5.