

**Bioprospecting halotolerant
lignocellulolytic enzymes from salt marsh
ecosystems**

Daniel Leadbeater

PhD

University of York

Biology

September 2018

Abstract

Lignocellulose is an abundant agricultural waste product presenting an attractive renewable feedstock for ethanol production. Currently, lignocellulosic bioprocessing is economically constrained requiring expensive pretreatments, high enzyme loadings and large volumes of freshwater imposing water security concerns. Seawater is an abundant and inexpensive prospective replacement solvent. The discovery of highly active halotolerant enzymes with novel tolerances may enable seawater incorporation in bioprocessing; sustainably improving process economics and mitigating water security concerns. Salt marshes are intertidal ecosystems where lignocellulose is remineralised within the predominantly marine organic rich sediments; representing model ecosystems for targeted bioprospecting of highly active, halotolerant and inhibitor resistant enzymes.

In this thesis, salt marsh carbon cycling is explored using an *in situ* degradation experiment. Metatranscriptomics in conjunction with metasecretome proteomics enabled the identification of the lignocellulose-associated enzymatic profile at the ecosystem level and their most likely taxonomic origin. Amplicon sequencing was conducted to reveal the abundance profile for key taxa.

Investigations revealed 11,268 identifiable biomass associated proteins, 410 of which were annotated as putative carbohydrate active domains within 307 proteins. The ecosystem level profile displayed a preference for cellulose over matrix polysaccharide deconstruction. Carbohydrate esterase family 1 enzymes were highly abundant; however, lignin modifying enzymes were not, suggesting a decoupling mechanism for accessibility in lieu of oxidative modification. Fungi made no contribution to the lignocellulolytic library; the dominant lignocellulolytic producing taxa were bacteria, identified as the *Proteobacteria* and *Bacteroidetes* families *Flavobacteriaceae*, *Alteromonadaceae*, *Cytophagaceae*, *Saccharospirillaceae*, *Marinilabiaceae*, *Flammeovirgaceae*, *Bacteroidaceae* and *Vibrionaceae* and *Sphingobacteriaceae*.

A spectrum of potential enzymes were selected for expression in a heterologous system; two enzymes were characterised in depth and both demonstrated halotolerance. This included a putative GH6; identified as an exo- β -1,4-glucanase with significant activity on crystalline cellulose. A second protein displayed a product profile representative of a β -1,4-glucosidase with significant transglycosylase activity on oligosaccharide substrates.

Table of Contents

Table of Contents

| | |
|--|----|
| Abstract..... | 2 |
| Table of Contents..... | 3 |
| List of figures..... | 9 |
| List of tables..... | 13 |
| List of equations..... | 14 |
| Acknowledgements..... | 15 |
| Author’s declaration | 16 |
| Chapter 1 Introduction | 17 |
| 1.1 Lignocellulosic biomass as a renewable resource..... | 17 |
| 1.2 Products from lignocellulosic biomass..... | 19 |
| 1.3 Lignocellulose – the plant cell wall..... | 20 |
| 1.3.1 Cellulose | 21 |
| 1.3.2 Hemicellulose | 21 |
| 1.3.3 Lignin | 24 |
| 1.4 Second generation biofuels..... | 25 |
| 1.5 Economic constraints to bioprocessing..... | 27 |
| 1.6 Seawater as an economically attractive replacement media | 30 |
| 1.7 Halotolerance | 32 |
| 1.8 Salt marshes | 33 |
| 1.9 Salt marsh carbon cycling..... | 36 |
| 1.10 Microbial ecology of salt marshes..... | 38 |
| 1.11 Salt marshes as an underexplored model ecosystem with biotechnological potential..... | 42 |
| 1.12 Enzymatic lignocellulose deconstruction | 42 |
| 1.12.1 Enzymatic classifications | 43 |
| 1.12.2 Cellulose deconstruction..... | 44 |
| 1.12.4 Oxidative polysaccharide deconstruction | 49 |
| 1.12.5 Carbohydrate binding modules..... | 50 |
| 1.12.6 Lignin deconstruction..... | 51 |
| 1.13 Aims..... | 52 |
| Chapter 2 Materials and methods | 53 |

Table of Contents

| | | |
|--|--|----|
| 2.1 | Chemical reagents | 53 |
| 2.1.1 | Buffers | 53 |
| 2.1.2 | Organisms..... | 53 |
| 2.1.3 | Phosphoric acid swollen cellulose (PASC) | 54 |
| 2.2 | Molecular biology techniques | 54 |
| 2.2.1 | Plasmid extraction and purification (PCR)..... | 54 |
| 2.2.2 | Polymerase chain reaction | 54 |
| 2.2.3 | Agarose gel electrophoresis | 56 |
| 2.2.4 | PCR product purification | 56 |
| 2.2.5 | Nucleic acid quantification | 56 |
| 2.2.6 | DNA sequencing | 57 |
| 2.2.7 | pH measurements | 57 |
| 2.2.8 | Media preparation | 57 |
| 2.2.9 | Plate preparation..... | 57 |
| 2.3 | Proteomic methods..... | 58 |
| 2.3.1 | Bradford Assay for protein quantification..... | 58 |
| 2.3.2 | Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) | 58 |
| Chapter 3 Time series <i>in situ</i> decomposition of lignocellulose in a salt marsh | | 60 |
| 3.1 | Introduction..... | 60 |
| 3.2 | Materials and methods | 63 |
| 3.2.1 | Nomenclature rationalisation: (meta)secretome or (meta)exoproteome | 63 |
| 3.2.2 | Field experiment..... | 64 |
| 3.2.2.1 | Field experiment synopsis | 64 |
| 3.2.2.2 | Biomass collection and preparation..... | 65 |
| 3.2.2.3 | Deployment | 66 |
| 3.2.2.4 | Sampling harvesting | 68 |
| 3.2.2.5 | Sample processing..... | 71 |
| 3.2.2.5.1 | Simultaneous DNA and RNA extraction | 71 |
| 3.2.2.5.2 | Protein processing..... | 72 |
| 3.2.2.5.3 | Biomass preparation | 73 |
| 3.2.2.5.4 | Mass balance | 73 |
| 3.2.2.5.5 | Salinity and pH measurements | 73 |
| 3.2.3 | Enzyme assays | 74 |
| 3.2.4 | Biomass composition analysis..... | 74 |
| 3.2.4.1 | Analysis of matrix polysaccharides..... | 75 |

Table of Contents

| | | |
|---|---|-----|
| 3.2.4.2 | Analysis of crystalline cellulose | 75 |
| 3.2.4.3 | Analysis of acetyl bromide soluble lignin | 76 |
| 3.2.4.4 | Analysis of total ash..... | 76 |
| 3.2.4.5 | Absolute quantification of lignocellulose fractions..... | 77 |
| 3.3 | Results | 77 |
| 3.3.1 | Field observations | 77 |
| 3.3.2 | Nucleic acid extractions..... | 78 |
| 3.3.3 | Mass balance | 79 |
| 3.3.4 | Protein extraction..... | 79 |
| 3.3.4.1 | Metasecretome activity | 82 |
| 3.3.5 | Biomass composition analysis..... | 87 |
| 3.3.5.1 | Absolute composition values | 87 |
| 3.3.5.1.1 | Matrix polysaccharide composition | 87 |
| 3.3.5.1.2 | Crystalline cellulose..... | 88 |
| 3.3.5.1.3 | Acetyl bromide soluble lignin..... | 89 |
| 3.3.5.1.4 | Ash..... | 90 |
| 3.3.5.1.5 | Integrated biomass composition..... | 92 |
| 3.3.5.2 | Extrapolated composition values | 94 |
| 3.4 | Discussion | 96 |
| Chapter 4 Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome . | | 100 |
| 4.1 | Introduction..... | 100 |
| 4.2 | Methods | 102 |
| 4.2.1 | Transcriptomic analysis | 102 |
| 4.2.1.1 | DNA depletion | 102 |
| 4.2.1.2 | mRNA enrichment..... | 103 |
| 4.2.1.3 | RNA sequencing..... | 103 |
| 4.2.1.4 | Transcriptome assembly (de novo)..... | 103 |
| 4.2.1.5 | Filtering and processing | 104 |
| 4.2.2 | Proteomic analysis..... | 105 |
| 4.2.2.1 | Liquid chromatography tandem mass spectrometry (LS_MS/MS) | 105 |
| 4.2.2.2 | Peptide to protein identification pipeline..... | 106 |
| 4.2.3 | Bioinformatic techniques in the analysis of the metasecretome | 108 |
| 4.3 | Results | 109 |
| 4.3.1 | Transcriptomic filtering and processing..... | 109 |
| 4.3.1.1 | zFPKM and size exclusion filtering..... | 109 |

Table of Contents

| | | |
|--|--|-----|
| 4.3.1.2 | zFPKM and size exclusion filtering..... | 110 |
| 4.3.2 | LC-MS/MS acquired data..... | 111 |
| 4.3.3 | Protein profile..... | 112 |
| 4.3.3.1 | Carbohydrate active enzyme profile | 113 |
| 4.3.3.1.1 | Carbohydrate esterases | 116 |
| 4.3.3.1.2 | Auxiliary activities..... | 117 |
| 4.3.3.1.3 | Polysaccharide lyases | 117 |
| 4.3.3.1.4 | Cellulosome associated domains | 118 |
| 4.3.3.2 | Carbohydrate binding domain profile..... | 118 |
| 4.3.3.2.1 | Catalytic associations to carbohydrate binding domains..... | 119 |
| 4.3.4 | Phylogenetic and functional classification of the metasecretome | 121 |
| 4.4 | Discussion..... | 130 |
| 4.4.1 | Functional assignment of the metasecretome | 130 |
| 4.4.2 | Functional phylogeny of the lignocellulose associated salt marsh microbiome..... | 133 |
| Chapter 5 Decomposition associated microbial community profile | | 137 |
| 5.1 | Introduction..... | 137 |
| 5.2 | Materials and methods | 139 |
| 5.2.1 | DNA Sequencing..... | 139 |
| 5.2.1.1 | Genomic DNA preparation and RNA depletion..... | 139 |
| 5.2.1.2 | Target gene selection and primer selection and development | 140 |
| 5.2.1.3 | Amplification of conserved regions..... | 143 |
| 5.2.1.4 | Illumina sequencing..... | 144 |
| 5.2.2 | Bioinformatic analysis | 146 |
| 5.2.2.1 | Sequence to community profile analysis pipeline..... | 146 |
| 5.3 | Results | 149 |
| 5.3.1 | Database processing..... | 149 |
| 5.3.2 | Prokaryotic profiles | 151 |
| 5.3.2.1 | Archaea..... | 151 |
| 5.3.2.2 | Bacteria..... | 153 |
| 5.3.2.2.1 | Gammaproteobacteria..... | 154 |
| 5.3.2.2.2 | Bacteroidetes | 156 |
| 5.3.2.2.3 | Deltaproteobacteria | 158 |
| 5.3.2.2.4 | Firmicutes..... | 159 |
| 5.3.2.2.5 | Verrucomicrobia..... | 159 |
| 5.3.3 | Fungal profile..... | 165 |

Table of Contents

| | | |
|-----------|---|-----|
| 5.4 | Discussion | 167 |
| 5.4.1 | Bacteria..... | 167 |
| 5.4.2 | Fungi | 172 |
| Chapter 6 | Expression and characterisation of putatively lignocellulolytic enzymes..... | 175 |
| 6.1 | Introduction..... | 175 |
| 6.2 | Methods | 177 |
| 6.2.1 | cDNA synthesis | 177 |
| 6.2.2 | genomic DNA preparation..... | 177 |
| 6.2.3 | Plasmid preparation | 177 |
| 6.2.4 | Cell preparation..... | 181 |
| 6.2.5 | Bacterial transformation | 182 |
| 6.2.6 | Expression screening | 182 |
| 6.2.7 | Protein purification | 183 |
| 6.2.8 | Thermal shift assay..... | 183 |
| 6.2.9 | Chemiluminescent immunoblotting..... | 184 |
| 6.2.10 | Assays | 184 |
| 6.2.10.1 | Reducing sugar assay..... | 185 |
| 6.2.10.2 | Substrate affinities | 185 |
| 6.2.10.3 | Bivariate pH, temperature optima | 185 |
| 6.2.10.4 | Salinity tolerance..... | 186 |
| 6.2.10.5 | Thermostability | 186 |
| 6.2.10.6 | Rate of activity..... | 187 |
| 6.2.10.7 | Processivity..... | 187 |
| 6.2.10.8 | Product profile..... | 187 |
| 6.2.10.9 | Signal peptide prediction | 188 |
| 6.3 | Results | 188 |
| 6.3.1 | Target amplification | 188 |
| 6.3.2 | Target cloning..... | 188 |
| 6.3.3 | Expression screening | 190 |
| 6.3.4 | Purification and characterisation of a putative GH6 (a_c570829_g1_i1_1) | 193 |
| 6.3.4.1 | Purification | 193 |
| 6.3.4.2 | Substrate specificity | 194 |
| 6.3.4.3 | Bivariate pH, temperature optima | 195 |
| 6.3.4.4 | Salinity tolerance..... | 196 |
| 6.3.4.5 | Thermostability | 197 |

Table of Contents

| | | |
|---------------|---|-----|
| 6.3.4.6 | Rate of activity..... | 199 |
| 6.3.4.7 | Product profile..... | 200 |
| 6.3.5 | Purification and characterisation of a putative GH5 (c_c892255_g1_i1_1) | 202 |
| 6.3.5.1 | Purification | 202 |
| 6.3.5.2 | Substrate specificity | 203 |
| 6.3.5.3 | Bivariate pH, temperature optima | 205 |
| 6.3.5.4 | Salinity tolerance..... | 206 |
| 6.4 | Discussion | 207 |
| 6.4.1 | Cloning approach..... | 207 |
| 6.4.2 | The GH6 (a_c570829_g1_i1_1) | 208 |
| 6.4.3 | The GH5 (c_c892255_g1_i1_1) | 210 |
| Chapter 7 | Final discussion..... | 213 |
| 7.1 | Future perspectives..... | 215 |
| Appendix | | 220 |
| Abbreviations | | 221 |
| References | | 223 |

List of figures

List of figures

| | |
|--|----|
| Figure 1.1 Lignocellulosic feedstocks potentially available by geographical location..... | 19 |
| Figure 1.2 Molecular structure of cellulose..... | 21 |
| Figure 1.3 Molecular structure of xylan..... | 22 |
| Figure 1.4 Simplified molecular architecture of glucuronoarabinoxylan..... | 23 |
| Figure 1.5 Hydroxycinnamyl alcohols and derivative monolignols..... | 24 |
| Figure 1.6 Potential ethanol yields from globally abundant grass-derived feedstocks..... | 27 |
| Figure 1.7 General overview of lignocellulosic bioprocessing from primary production to ethanol distillation..... | 28 |
| Figure 1.8 Potential global distributions of salt marsh and mangroves ecosystems by sea surface temperature (SST) and 20°C isotherm..... | 34 |
| Figure 1.9 Lignocellulose cycling, deposition, redistribution and exportation..... | 37 |
| Figure 1.10 Synergistic action of cellulase targeting of a cellulose polymer..... | 45 |
| Figure 1.11 Structures of endo- and exo- β -1,4 glucanases..... | 46 |
| Figure 1.12 Synergistic action of xylan targeting of an undecorated xylan backbone..... | 47 |
| Figure 1.13 Schematic of major hydrolytic enzymes targeting glucoarabinoxylan..... | 48 |
| Figure 3.1 Schematic overview of the <i>in situ</i> degradation experiment..... | 65 |
| Figure 3.2 Experiment location..... | 67 |
| Figure 3.3 Deployment of the <i>in situ</i> degradation experiment..... | 68 |
| Figure 3.4 Measured pH and salinity throughout the time course experiment..... | 77 |
| Figure 3.5 Tidal and sampling regime..... | 78 |
| Figure 3.6 Genomic DNA and total RNA extractions..... | 78 |
| Figure 3.7 Temporal mass loss and decomposition rates..... | 79 |
| Figure 3.8 SDS-PAGE of metasecretome extraction and storage..... | 80 |
| Figure 3.9 Dry weight biomass and recovered protein..... | 81 |
| Figure 3.10 Analysis of protein recovered as a function of dry weight biomass..... | 82 |
| Figure 3.11 Regression analysis of normalised protein yields per dry weight unit biomass..... | 82 |
| Figure 3.12 Lignocellulolytic activity of the pooled metasecretome across all time points..... | 83 |
| Figure 3.13 Metasecretome xylanase activity by time point..... | 84 |
| Figure 3.14 Xylanase activity of the metasecretome..... | 85 |
| Figure 3.15 Regression analysis of xylanase activity concentration..... | 86 |
| Figure 3.16 Relative metasecretome activity rates for AZCL substrates..... | 86 |
| Figure 3.17 Temporal composition of matrix polysaccharide constituent sugars..... | 87 |

List of figures

| | |
|--|-----|
| Figure 3.18 Total matrix polysaccharide temporal profile..... | 88 |
| Figure 3.19 Temporal <i>in situ</i> biomass crystalline cellulose profile. | 89 |
| Figure 3.20 Temporal <i>in situ</i> biomass lignin profile..... | 90 |
| Figure 3.21 Temporal <i>in situ</i> biomass ash profile. | 91 |
| Figure 3.22 Regression analysis of inter week ash deviation and mean tide height..... | 92 |
| Figure 3.23 Lignocellulose composition of <i>in situ</i> biomass..... | 93 |
| Figure 3.24 Rate of change of the lignocellulose fractions..... | 94 |
| Figure 3.25 Composition of <i>in situ</i> remaining biomass..... | 95 |
| Figure 3.26 Normalised composition of remaining biomass inclusive of ash..... | 95 |
| Figure 3.27 Normalised composition of remaining biomass excluding ash. | 96 |
| Figure 4.1 Protein identification pipeline. | 108 |
| Figure 4.2 Analysis pipeline for the annotation and analysis of the PMO database. PMO; peptide matching open reading frame. | 109 |
| Figure 4.3 zFPKM and size exclusion filtering of metatranscriptomic databases..... | 111 |
| Figure 4.4 Protein and CAZy distribution..... | 113 |
| Figure 4.5 Hierarchal clustering of the temporal enzyme class profile (Euclidean). | 114 |
| Figure 4.6 Temporal profile of the 15 most abundant enzyme classes..... | 115 |
| Figure 4.7 Temporal profile of the 20 most abundant ORFs. | 116 |
| Figure 4.8 The temporal profile of the observed carbohydrate esterase (CE) families. | 117 |
| Figure 4.9 Temporal profile of the observed auxiliary families (AA). | 117 |
| Figure 4.10 Temporal profile of the observed polysaccharide lyase (PL) families. | 118 |
| Figure 4.11 Temporal profile of the observed cellulosome associated domains..... | 118 |
| Figure 4.12 Hierarchal clustering of the temporal carbohydrate binding domain profile. | 119 |
| Figure 4.13 Association of observed carbohydrate binding domains (CBM) to catalytic domains. ... | 120 |
| Figure 4.14 Temporal profile of the 15 most abundant CBM families. | 121 |
| Figure 4.15 Metasecretome distribution at the superkingdom level..... | 122 |
| Figure 4.16 Most likely taxonomic origin of carbohydrate active enzymes at phylum level. | 123 |
| Figure 4.17 Most likely taxonomic origin at class level resolution of carbohydrate active enzymes (CAZy)..... | 123 |
| Figure 4.18 Phylogenetic distribution of CAZy ORFs at phylum resolution for all time points. | 124 |
| Figure 4.19 Phylogenetic distribution of CAZy ORFs at class level resolution. | 124 |
| Figure 4.20 Phylogenetic distribution of CAZy ORFs at class-order level resolution..... | 125 |
| Figure 4.21 Phylogenetic distribution of CAZy ORFs at class-family level resolution..... | 125 |
| Figure 4.22 Taxonomic CAZy diversity of the metasecretome. | 127 |

List of figures

| | |
|--|-----|
| Figure 4.23 Phylogenetic distribution of CAZy classes at class resolution..... | 128 |
| Figure 4.24 Sum of mean molar percentage CAZy profile by most likely taxonomic origin by time point..... | 129 |
| Figure 5.1 General ribosomal domain architecture..... | 140 |
| Figure 5.2 Hypervariable regions of the ribosomal 18S rRNA gene. | 141 |
| Figure 5.3 Hypervariable regions of the ribosomal 16S rRNA gene. | 141 |
| Figure 5.4 Analysis pipeline for amplicon sequence databases. | 148 |
| Figure 5.5 A selection of amplicon profiles for teach target region. | 150 |
| Figure 5.6 Read distribution of the amplicon profile databases, pre- and post-assignment. | 150 |
| Figure 5.7 Prokaryotic profile from 16S rRNA sequence homology at Kingdom resolution. | 152 |
| Figure 5.8 Archaeal profiles elucidated from 16S rRNA sequence homology. | 152 |
| Figure 5.9 Temporal OTU richness for bacterial and archaea lineages. | 153 |
| Figure 5.10 Phylum level bacterial profile elucidated from 16S rRNA sequence homology. | 154 |
| Figure 5.11 Bacterial profiles elucidated from 16S rRNA sequence homology..... | 154 |
| Figure 5.12 Normalised <i>Gammaproteobacterial</i> fraction. | 155 |
| Figure 5.13 CAZy producing <i>Gammaproteobacteria</i> normalised within the <i>Gammaproteobacterial</i> profile..... | 156 |
| Figure 5.14 Normalised <i>Bacteroidetes</i> fraction. | 157 |
| Figure 5.15 CAZy producing <i>Bacteroidetes</i> normalised within the <i>Bacteroidetes</i> profile..... | 157 |
| Figure 5.16 Normalised <i>Deltaproteobacteria</i> fraction..... | 158 |
| Figure 5.17 CAZy producing <i>Deltaproteobacteria</i> normalised within the <i>Bacteroidetes</i> profile. | 159 |
| Figure 5.18 Normalised <i>Deltaproteobacteria</i> fraction..... | 159 |
| Figure 5.19 Normalised <i>Verrucomicria</i> fraction..... | 160 |
| Figure 5.20 Unfiltered CAZy producing families at the microbiome level..... | 160 |
| Figure 5.21 Unfiltered CAZy producing taxa at the microbiome level at family level taxonomic resolution..... | 161 |
| Figure 5.22 Family level resolution of CAZy producers and their respective CAZy contributions..... | 162 |
| Figure 5.23 CAZy producing families..... | 163 |
| Figure 5.24 Productivity index for CAZy producing taxa at family level resolution..... | 165 |
| Figure 5.25 Fungi profile at kingdom and phylum levels..... | 165 |
| Figure 5.26 Fungi profile at class and order levels..... | 166 |
| Figure 5.27 Fungi profile at the microbiome level at family level taxonomic resolution. | 166 |
| Figure 6.1 Domain architectures of a_c570829_g1_i1_1 and c_c892255_g1_i1_1..... | 177 |

List of figures

| | |
|---|-----|
| Figure 6.2 T7 colony PCRs. Transformed Stellar™ cells with the putative gene of interest in pET-52b(+). | 190 |
| Figure 6.3 Representative SDS-PAGE of the screening process..... | 192 |
| Figure 6.4 Fluorescent immunoblots of a random subset of protein targets. | 193 |
| Figure 6.5 Purification of a putative GH6..... | 194 |
| Figure 6.6 A representative thermal shift profile of a_c570829_g1_i1_1 (GH6). | 194 |
| Figure 6.7 Substrate affinities of a_c570829_g1_i1_1 (GH6). | 195 |
| Figure 6.8 Bivariate pH and temperature optima of a_c570829_g1_i1_1 (GH6)..... | 196 |
| Figure 6.9 Salinity tolerance of a_c570829_g1_i1_1 (GH6)..... | 197 |
| Figure 6.10 Thermal shift profiles of a_c570829_g1_i1_1 (GH6) post heat treatment in different treatments. | 198 |
| Figure 6.11 Thermostability of a_c570829_g1_i1_1 (GH6) in different salt concentrations. | 199 |
| Figure 6.12 Rate of activity of a_c570829_g1_i1_1 (GH6). | 200 |
| Figure 6.13 Product profile a_c570829_g1_i1_1 after incubation with oligosaccharides. | 201 |
| Figure 6.14 Product profile and processivity of a_c570829_g1_i1_1..... | 201 |
| Figure 6.15 Purification of a putative GH5..... | 202 |
| Figure 6.16 A representative thermal shift profile of c_c892255_g1_i1_1 (T8)..... | 203 |
| Figure 6.17 Activity of c_c892255_g1_i1_1 on pNP-R substrates. | 204 |
| Figure 6.18 Visible reduction in Avicel™ co-incubated with c_c892255_g1_i1_1..... | 204 |
| Figure 6.19 Product profile of c_c892255_g1_i1_1 after incubation with oligosaccharides. | 205 |
| Figure 6.20 Bivariate pH and temperature optima of c_c892255_g1_i1_1 (GH5)..... | 206 |
| Figure 6.21 Salinity tolerance of c_c892255_g1_i1_1 (GH5)..... | 207 |

List of tables

List of tables

| | |
|---|-----|
| Table 1.1 Major components in the chemical composition of seawater. | 31 |
| Table 2.1 PCR reaction components. | 55 |
| Table 2.2 General PCR thermocycling conditions. | 55 |
| Table 3.1 Transect position of the five biological replicates in Welwick salt marsh. | 68 |
| Table 3.2 Sampling and tidal regime. | 70 |
| Table 4.1 Sequence reads throughout RNA sequence processing and assembly. | 110 |
| Table 4.2 Protein sample origin, loading and spectra acquired from LC-MS/MS analysis. | 112 |
| Table 5.1 Primer selection for salt marsh community profiling. | 142 |
| Table 5.2 Primer architecture for direct amplification to indexing. | 142 |
| Table 5.3 Primer pairs tested that were deemed inadequate for salt marsh samples. | 143 |
| Table 5.4 PCR conditions for amplification of hypervariable regions. | 143 |
| Table 5.5 Thermocycling conditions for ribosomal 16S amplification using 515F-Y and 806R. | 144 |
| Table 5.6 Thermocycling conditions for ITS2 amplification using fITS7 and ITS4ngs. | 144 |
| Table 5.7 Thermocycling conditions for ribosomal 18S amplification using 574F and 952R. | 144 |
| Table 5.8 AMPure XP bead concentrations at each purification stage for each amplicon profile. | 145 |
| Table 5.9 Thermocycling conditions for Nextera XT™ indexing. | 145 |
| Table 5.10 Commands for the replication of the 16S rRNA amplicon database processing pipeline. | 149 |
| Table 5.11 Statistics for the amplicon database processing pipeline. | 151 |
| Table 6.1 Primers utilised for plasmid linearisation and confirmation of transformation. | 178 |
| Table 6.2 Primers for amplification of selected proteins. | 179 |
| Table 6.3 Nested primers for amplification of selected sequences. | 181 |
| Table 6.4 PCR conditions for the successful amplification of ORF targets. | 189 |
| Table 6.5 An overview of successfully cloned targets into pET-52(b)+ in Stellar™ cells. | 190 |
| Table 6.6 Optimal expression conditions of screened protein targets. | 191 |

List of equations

List of equations

| | |
|--|-----|
| Equation 3.1 Mass balance partial conversion factor..... | 73 |
| Equation 3.2 Determination of the proportion of acetyl bromide soluble lignin (ABSL). | 76 |
| Equation 4.1 zFPKM. | 104 |
| Equation 4.2 Exponentially modified protein abundance index (emPAI)..... | 107 |
| Equation 5.1 Taxonomic productivity index (PI)..... | 164 |
| Equation 6.1 Processivity. | 187 |

Acknowledgements

Firstly, I am sincerely grateful to my supervisors Prof. Neil Bruce and Prof. Simon McQueen-Mason for initially giving me the opportunity to undertake this project and who have provided their expertise, guidance and unwavering support throughout this project. I truly could not have asked for better mentors.

Besides my supervisors, I would like to thank the members of my Thesis Advisory Panel; Dr. Thorunn Helgason and Dr. James Chong whose diverse perspectives provided insightful feedback, encouragement and often tough questions that helped reinforce and shape the direction of the research.

I would like to express gratitude towards all researchers in the Centre for Novel Agricultural Products who have provided assistance; but above all, made these last four years an enjoyable experience. I thank Dr. Joe Bennett and Dr. Luisa Elias for taking the time to pass on their experience with protein expression and purification. I also thank Dr. Joe Taylor for taking considerable time in explaining community analysis. I would also like to thank Dr. Nicola Oates who has provided invaluable input throughout this project.

I would like to thank Dr. Deborah Rathbone and Susan Heywood at the Biorenewables Development Centre, who gave me access to their laboratory and research facilities; providing input and assistance during sequencing.

I would also like to acknowledge the landowners and representatives of Natural England and the Yorkshire Wildlife Trust for the pleasant application (and negotiation) process and final granting of permission to use Welwick Salt Marsh as an experimental location.

Author's declaration

Author's declaration

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

Chapter 1 Introduction

1.1 Lignocellulosic biomass as a renewable resource

Woody plant biomass is composed of lignocellulose and represents a large reservoir of stored global carbon. Lignocellulose is a composite of complex polysaccharides (cellulose and hemicellulose) and an aromatic polymer; lignin, that form an interconnected network in the secondary plant cell wall that confers structural and mechanical strength. It is an agricultural waste product that represents a promising feedstock for biofuel production where stored carbon in the form of sugars and aromatic compounds can be released and fermented or transformed to desirable products. Lignocellulose exists in various forms; it is prevalent throughout many industries including agriculture, food and forestry.

Waste lignocellulosic material is produced annually from green wastes, forestry residues and agricultural by-products. Whilst these waste streams have uses, there is a disparity between yield and use, resulting in a surplus of material that is wasted and often burnt on the field contributing to pollution and health problems. The annual yield of wasted lignocellulosic material in the UK is almost 16 million tonnes (Davide and David, 2014). This consists of forestry residues, green wastes and agricultural by-products.

Agricultural waste residues in the UK exist in the form of grass (wheat, barley, oil seed rape) straw. Greater than 4 million of the 11.7 million tonnes of straw in the UK is uncollected (Davide and David, 2014). Globally, other sources of waste lignocellulosic material include rice straw, corn stover and sugarcane bagasse which are also present in abundance (Figure 1.1). Sugarcane is the most intensely cultivated crop globally, yielding 2.165×10^{10} tonnes of sugarcane annually, which is primarily produced in Brazil (34.1%), India (15.8%), China (5.8%) and Thailand (4.6%); however, the lignocellulosic fraction of this crop is relatively minor (Zhao and Li, 2015). Maize is the second largest agricultural crop in the world (in terms of utilisable mass produced yielding 1.06×10^9 tonnes per year) and is predominantly produced in the United States (Zhao et al., 2018). Maize production creates residual lignocellulosic material colloquially referred to as corn stover (residual stems, leaves and cobs), of which 150 million tonnes is produced and wasted annually (Zhao et al., 2018, Kadam and McMillan, 2003). Asia produces 6.86×10^8 tonnes of rice grain annually (with 7.59×10^8 produced globally) and over 1 billion tonnes of straw, 20% of which utilised, and the remaining is burnt or wasted (Sattlewal et al., 2018). Wheat is the third most abundant cereal crop yielding 7.4×10^8 globally, which generates 3.54×10^8 tonnes of putatively available wheat straw distributed

Introduction

throughout Asia (41%), Europe (37.4%) America (17.7%) (Sarkar et al., 2012, Jaisamut et al., 2016). Oil palm is experiencing significant growth and currently produces in excess of 4.55×10^7 tonnes of palm oil (2013) in Malaysia (41%) and Indonesia (44%); the lignocellulosic fraction mainly exists as empty fruit bunch (0.32 per unit fruit) which is incinerated (no longer in Malaysia due to regulatory control) or used as mulch, but also includes mesocarp (0.12 per unit fruit), shell (0.05 per unit fruit), fronds and trunks (Suksong et al., 2017, Mohammad et al., 2012, Mukherjee and Sovacool, 2014). The straw to grain ratio for rice (1.5:1) is higher but comparable to that of wheat straw (1.3:1), both of which are significantly higher than that of maize (stover to corn; 1:1), sugarcane (bagasse to sugarcane 0.28:1) and oil palm (empty fruit bunch:fruit, 0.2-0.32:1); conferring rice as the most productive lignocellulosic waste residue crop globally, followed closely by wheat (Zhao et al., 2018, van der Weijde et al., 2013, Cerqueira et al., 2007, Talebnia et al., 2010, Mohammad et al., 2012).

A large proportion of these agricultural wastes are openly burnt, degraded on site or recycled within soils to maintain soil quality and nutrient levels. The open burning of agricultural wastes results in incomplete combustion and the release of undesirable pollutants. At local scales combustion results in the emission of particulate black carbon which directly impacts human health (respiratory complications and carcinogens) and has a local net-warming effect; it also releases toxic gases (carbon monoxide), volatile organic compounds and harmful polycyclic aromatic hydrocarbons (Galdos et al., 2013). The potential global impacts are equally concerning, with CO₂, NO_x, CH₄ and SO₂ emissions and ozone depletion associated with lignocellulose combustion (Gadde et al., 2009, Arai et al., 2015). Complete removal of this lignocellulosic material from agricultural systems (by burning or for bioprocessing) is thought to have deleterious consequences for soil quality; potentially limiting availability (Blanco-Canqui and Lal, 2009); however, no consensus has been reached (Cherubin et al., 2018). Despite this, lignocellulosic materials are considered renewable and represent vast reservoirs of feedstocks available globally with promising potential for the conversion to valuable products; disincentivising burning and mitigating the associated deleterious effects.

Introduction

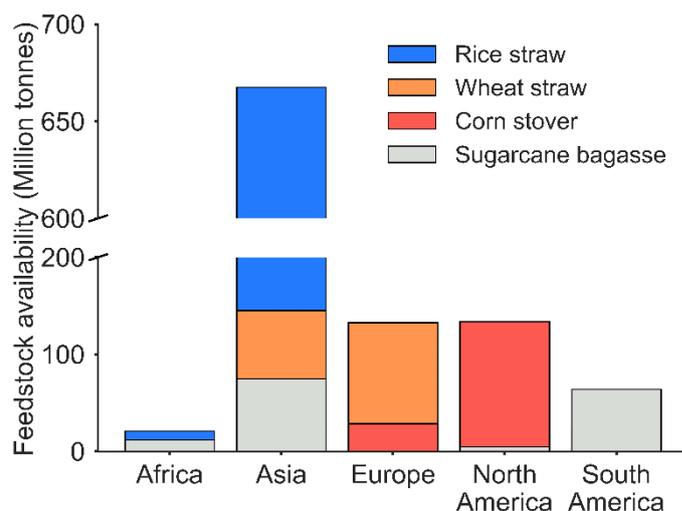


Figure 1.1 Lignocellulosic feedstocks potentially available by geographical location. Data sourced from (Kim and Dale, 2004).

1.2 Products from lignocellulosic biomass

Lignocellulose is a complex composite material and its inherent complexity confers its transformation into a plethora of products. Using lignocellulose as a source material a wide range of products have been developed; bio-derived plastics (bioplastics), additives for composite bioplastics, platform chemicals with broad downstream applicability, potential food additives and general chemicals. Despite an array of products, there is currently no consensus schematic as to the optimal processing pathway to utilise lignocellulose to generate these renewable products.

Lignocellulose derived products have been used to improve the cost effectiveness of bacterial polyesters. Polyhydroxyalkanoates (PHAs) are organic bioplastics produced by bacteria, the most common of which is poly-3-hydroxybutyrate (P3HB). Bioplastics are of interest as they are biodegradable, however current economic conditions and manufacturing processes mean they are priced at a premium to conventional equivalents, which has inhibited commercial uptake (Li et al., 2016b). PHAs can be blended with lignocellulose derived products (such as cellulose and ethyl cellulose) to confer versatility in the resulting composite bioplastic (Li et al., 2016b, Bilal et al., 2017, Cai et al., 2012). Other bioplastics arising indirectly from lignocellulose is the fermentation of lignocellulose derived sugars to lactic acid, which then produces the biodegradable polylactic acid (PLA) (Singhvi and Gokhale, 2013).

Platform chemicals can be derived from lignocellulosic material including xylitol, vanillin and furfural (200,000 tonnes is required annually at a value of over \$300 million), the latter of which is

utilised in the production of phenol-based plastics, pesticides and varnishes (Pothiraj et al., 2006). Other platform chemicals include a variety of sugars and aromatics that can be fermented toward more desirable products. The most documented fermented product is (bio)ethanol.

The complexity of lignocellulosic biomass confers significant plasticity in its use; in the generation of a variety of products. As waste lignocellulosic biomass is generated in abundance globally, it is apparent that this is a currently undervalued and underexploited potential resource.

1.3 Lignocellulose – the plant cell wall

Lignocellulose is a complex interwoven polymer composite comprised of two major polysaccharide groups (cellulose and hemicellulose) and an aromatic polymer (lignin). It is present in the secondary cell wall of plant cells and it differs from the primary cell wall in that it is only formed in mature tissues that have ceased growing, to provide mechanical strength and rigidity (Cosgrove and Jarvis, 2012). It is also compositionally different to primary cell walls with xyloglucans and pectins in the primary cell wall substituted for glucomannans, xylans and lignin in the secondary cell wall, conferring strength over elasticity (Mellerowicz and Sundberg, 2008). Despite a wealth of knowledge surrounding lignocellulose structure and composition, the full extent of its structures, linkages and composition has yet to be fully elucidated.

Secondary cell walls are significantly thicker than primary cell walls (Kumar et al., 2016). Therefore, residual biomass is predominantly composed of secondary cell walls and its associated polymers, known collectively as lignocellulose. The structure and composition of lignocellulose varies significantly within the kingdom *Plantae* (specifically between soft woods, hard woods and grasses) (Anwar et al., 2014). This thesis considers the biotechnological potential for agricultural waste lignocellulose, which is often derived from grasses. Therefore, grass lignocellulose will be described in the following sections. Wheat straw is the dominant lignocellulosic feedstock in the UK of which the largest component is cellulose (33-40%), followed by hemicellulose (20-25%) and finally lignin (15-20%) (McKendry, 2002).

Introduction

1.3.1 Cellulose

Cellulose is not only the dominant constituent of lignocellulose but also the most abundant biopolymer on Earth (Cosgrove, 2005). Cellulose is a linear polysaccharide composed of repeating β -(1-4) *D*-glucose units with species dependent degrees of polymerisation (DP) of 925-5500, with grasses exhibiting lower DP to softwoods and hardwoods (Hallac and Ragauskas, 2011). Successive glucose residues are rotated by 180° around the β -(1-4) glycosidic linkages, this stereochemistry confers cellobiose (a disaccharide composed of two β -(1-4) linked glucose units) as the basic repeating unit (Cocinero et al., 2009).

Linear cellulose chains pack tightly in parallel and crystallise to form cellulose microfibrils through intrachain and interchain hydrogen bonding of hydroxyl groups (Shen and Gnanakaran, 2009). This crystallinity confers stability and recalcitrance to degradation through core chain inaccessibility. This is further exacerbated by exposed end heterogeneity; the cellulose chain has a reducing end in which the terminal glucose unit has an anomeric carbon at position 1 (which usually forms the β -(1-4) linkage) and a non-reducing end, in which the anomeric carbon is involved in a glycosidic linkage (Figure 1.2).

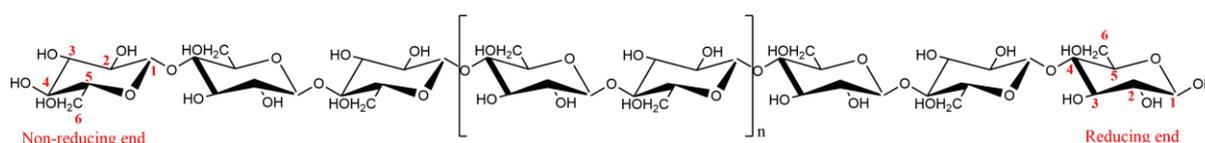


Figure 1.2 Molecular structure of cellulose. Cellulose consists of repeating β -(1-4) *D*-glucose units, termini are displayed as reducing and non-reducing ends. Repeating unit indicates a cellobiose subunit, adapted after (Zhang et al., 2011c).

1.3.2 Hemicellulose

Hemicellulose encompasses a diverse suite of polysaccharides that form a complex and interconnected network known as the matrix polysaccharides or cross-linking glycans. These polysaccharides consist of a spectrum of pentose (5C) and hexose (6C) sugars and in contrast to cellulose, are often branching and contain various decorations (acetyl, methyl or sugars) (Scheller and Ulvskov, 2010). The inherent form complexity and spatial heterogeneity confers matrix polysaccharides as amorphous. Matrix polysaccharides form an interconnected network that

interacts with the cellulose microfibrils through hydrogen bonding or physical entrapment during biosynthesis (Scheller and Ulvskov, 2010). The matrix polysaccharides therefore form an interconnected coating to cellulose microfibrils and ostensibly function to enhance and ensure physical properties and rigidity in the cell wall, by tethering cellulose microfibrils together (Scheller and Ulvskov, 2010, Park and Cosgrove, 2015).

The predominant matrix polysaccharides in the secondary cell walls of grasses (family *Poaceae*) are glucuronoarabinoxylan (GAX) which can comprise up to 50% of the total cell wall polysaccharides (Scheller and Ulvskov, 2010). Minor polysaccharides include xyloglucan, glucomannan, mixed linkage glucans β -(1-3,1-4)-glucans and pectins (Wang et al., 2016b, Scheller and Ulvskov, 2010, Vogel, 2008). GAX are structurally and compositionally diverse, but all belong to a family of polysaccharides known as xylans as they are all supported by a xylan backbone of linear β -(1-4) linked xylosyl residues (Figure 1.3) (Zhong et al., 2018). GAX contains single length side chains of arabinose and glucuronic acid at O-3 and O-2 positions respectively (Figure 1.4) (Vogel, 2008).

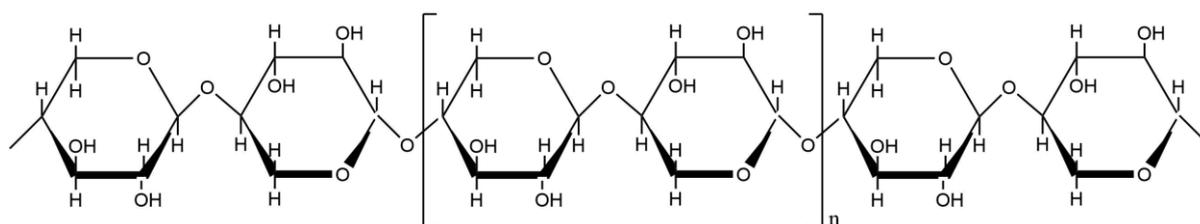


Figure 1.3 Molecular structure of xylan. Model of the linear β -(1-4) linked xylosyl residues comprising the xylan backbone.

GAX is highly decorated with many functional groups and sugars at various locations (Figure 1.4). The function of these decorations has yet to be fully elucidated but preventing crystalline formations has been suggested, with the exception of acetylation which facilitates the binding of the xylan backbone to cellulose (Busse-Wicher et al., 2014, Gomez et al., 2008). Up to 55% of xylosyl units in GAX are acetylated, suggesting cellulose binding as an important function (Gille and Pauly, 2012). The magnitude of acetylation and abundance of GAX is thought to be the main contributor to the formation of acetate, a major inhibitor of hydrolysis in commercial cocktails (Selig et al., 2009). Other decorations include galactosyl, methylated and non-methylated glucuronic acid ([Me]GlcA), xylose and importantly in grasses, ferulic acid (Marriott et al., 2016, Vogel, 2008).

Introduction

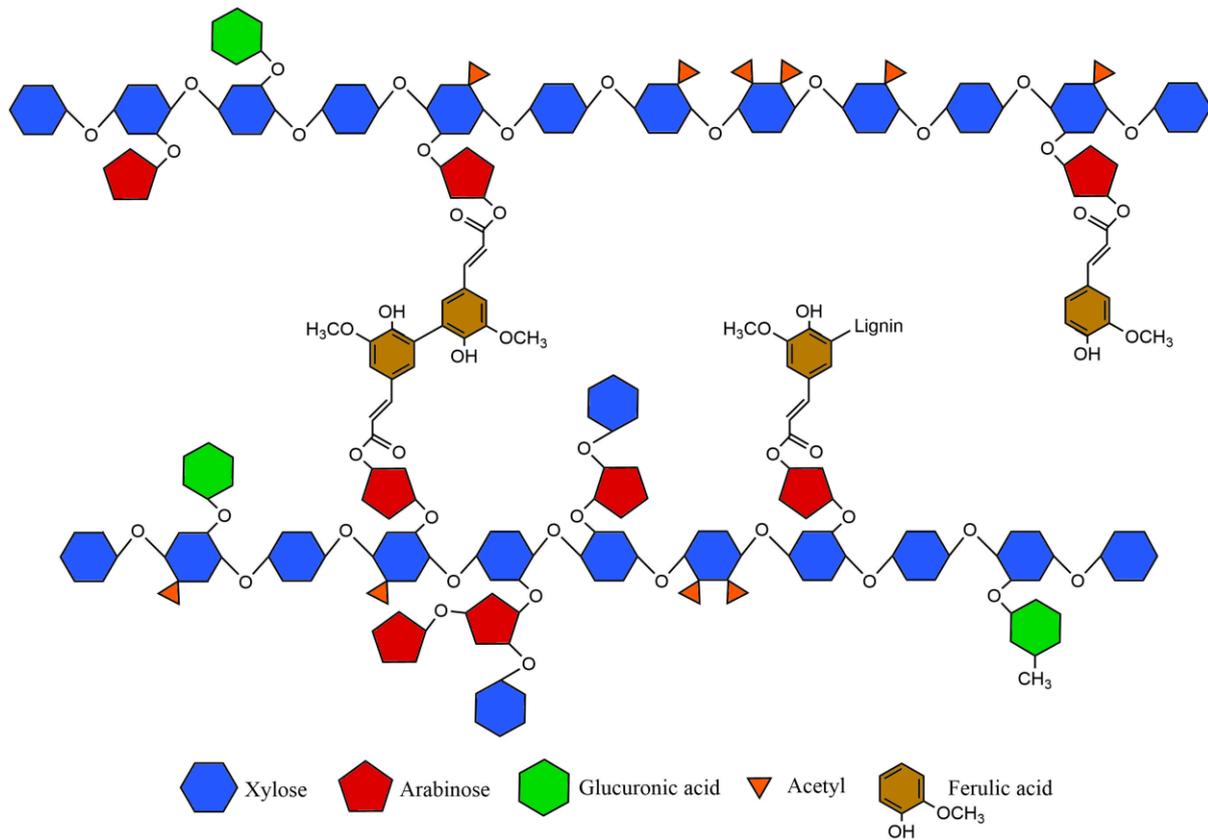


Figure 1.4 Simplified molecular architecture of glucuronoarabinoxylan. Inter and intra cross linkages in grasses are displayed, schematic assimilated from (Gomez et al., 2008, Buanafina, 2009, Marriott et al., 2016).

Not only does GAX bind to cellulose but ferulic acid also interfaces the GAX branches with lignin polymers and other GAX polysaccharides. The point at which lignin and hemicellulose converge is known as the lignin-carbohydrate complex (LCC). In grasses, LCC linkages are thought to consist mainly of aryl ester (ferulic linkages between ferulic acid and arabinose side chains in GAX) and aryl ether bonds (Buanafina, 2009). These linkages have been shown to negatively affect xylan hydrolysis (Grabber et al., 1998). Additionally, these linkages have been identified as potentially important for simplifying biorefining processes (Humbird et al., 2011). Therefore, not only do these bonds provide a covalent linkage to the most recalcitrant lignocellulose component, lignin (at the LCC), but also confer inherent recalcitrance in the structure itself.

1.3.3 Lignin

Lignin is the third most abundant polymer in the secondary cell wall of grasses, and in contrast to cellulose and hemicellulose, it is an aromatic polymer with a quasi-random arrangement and no repeating pattern. Lignin is thought to be the structural and UV protective component enabling the colonisation of plants on land at least 450 million years ago (Weng and Chapple, 2010). Lignin is composed mainly of phenylpropanoid units *p*-hydroxyphenyl (H), syringyl (S) and guaiacyl (G) residues (collectively known as monolignols), derived from the hydroxycinnamyl alcohols *p*-coumaryl, coniferyl and sinapyl precursors (Figure 1.5) (Lu et al., 2017). While the diversity of lignin components is less than matrix polysaccharides, it is the mechanism of lignin biosynthesis that confers the greatest level of spatial heterogeneity in the plant cell wall. Lignin biosynthesis occurs through free radical mediated coupling to form C-C or ether bonds between monolignols (favouring coupling at β - β , β -O-4 and β -5 positions), therefore the bond formed is dependent on the proximity and presence of monolignol units and their respective orientation at the time of incorporation (Boerjan et al., 2003, Vanholme et al., 2010). This feature means lignin biopolymers are racemic, which confers near infinite levels of spatial and compositional heterogeneity. Lignin polymerisation is processive, as opposed to a chain reaction, as the free radical is derived from monolignol oxidation (dehydration), which is quenched during the coupling of two dehydrated monolignols (Vanholme et al., 2010).

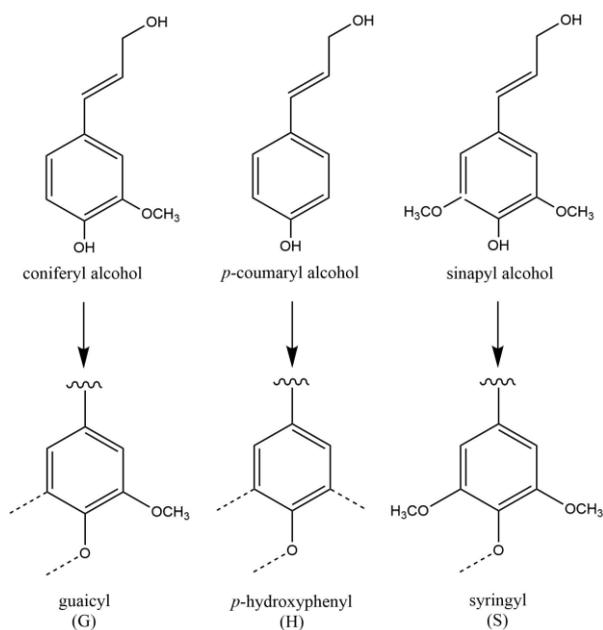


Figure 1.5 Hydroxycinnamyl alcohols and derivative monolignols. Adapted after (Gibson et al., 2013).

Lignin composition is species dependent and monocot lignin (which includes grasses) exhibit a higher proportions of H residues, which are present at negligible levels in dicots (Vogel, 2008). However, across all angiosperms the G and S residues are consistently dominant, but in varying proportions (Karlen et al., 2018). Lignin is distributed throughout the secondary cell wall matrix as a function of the mechanism of polymerisation where it structurally protects core polysaccharides through covalent (ferulate and diferulate esters) linkages to matrix polysaccharides and non-covalent linkages to cellulose microfibrils (Buanafina, 2009, Zhang et al., 2015). Lignin complexity and distribution throughout the secondary cell wall confers extreme recalcitrance and environmental persistence; to such an extent that deposition during the Carboniferous period and subsequent fossilisation to coal facilitated human industrialisation (Weng and Chapple, 2010, Hayatsu et al., 1979).

1.4 Second generation biofuels

The interest in biofuels has arisen from increasing global population and energy demands with a desire to implement carbon neutral strategies to sustainably sequester the demands of these forthcoming increases. Increasing global population (projected to reach or exceed 8.5 billion by 2030) indicates an additional 1.3 billion people will utilise and require energy. Considered together with globalisation in the form of increased industrialisation, urbanisation, and motorisation, the growth in population translates to a projected net increase in global energy consumption of 36% by 2030 (IEA, 2010). The European Commission, under the directive 2009/28/EC, mandates that by 2030, 20% of the total energy consumption in Europe must be derived from renewable sources. At the 2015 annual G7 summit, the IPCC recommendations to reduce the impact of climate change were discussed and industrialised nations declared a commitment to a zero-carbon economy. The aim stated is to phase out fossil derived fuels by the end of the century. Additionally, in the UK, the 2008 Climate Change Act was introduced to reduce greenhouse gas emissions by a minimum of 80% by 2050 through multiple decarbonisation schemes.

Currently over 91.32 million barrels of oil (MBO) are produced daily (in 2016) against a consumption rate of 95.36 MBO per day (Davis et al., 2018). Current reserves (excluding new discoveries post 2007 and assuming a constant consumption rate) and extraction rates are only capable of supplying this demand for an additional 40-50 years (Davis et al., 2011). Due to the extent of motorised infrastructure that is dependent on liquid fuel, liquid alternatives must be sought to address these issues. Potential renewable alternatives include biofuels (such as bioethanol,

Introduction

biodiesel, butanol) which are viewed as a renewable subsidy (rather than a direct replacement) for conventional fossil-derived petroleum.

The underlying principle of biofuels is to use photosynthetic organisms (plants or algae) to fix atmospheric CO₂ and water into monomeric sugar, where they are fermented or stored as polysaccharides, which can then be deconstructed into its constituent parts and chemically or biologically transformed into a useable form of fuel. As such, biofuels utilise biological feedstocks to generate (via various processes) monosaccharide sugars or lipids (di- and triglycerides) which are then fermented to bioethanol or trans-esterified to fatty acid methyl esters (biodiesel) respectively (Coppola et al., 2009). First generation biofuels largely utilise the edible fraction of agricultural products (e.g. maize starch) but the feedstock is deemed unsuitable as a result of food security concerns. Second generation biofuels were developed to overcome the perceived limitations of the first generation. Second generation biofuels are derived from the nutrient poor, non-food biomass fraction, which encompasses the lignocellulosic wastes described in section 1.2. This feedstock, in contrast to first generation biofuels, mitigates concern over both food (by not utilising the edible crop products) and land (food crops are not displaced by dedicated crops, new agricultural land is not required and low value marginal land becomes suitable) security (Naik et al., 2010, Havlik et al., 2011). The major benefit of bio-based fuels is that they exhibit significantly lower carbon emissions to their fossil-derived counterparts and in some cases approach carbon neutrality, however this is situation specific, hugely variable and no consensus assessment methodology exists (Johnson, 2009).

First generation biofuels are being produced and deployed successfully in the United States (using corn starch as a feedstock) and Brazil (using sugarcane sucrose). These two countries produce 84% of the global bioethanol production (over 27,000 million gallons) (RFA, 2018). Despite first generation biofuel production being well established, second generation processing is significantly more complex. The comparative heterogeneity of lignocellulose as opposed to first generation crop feedstocks (starch and sucrose) requires an additional pre-treatment step and a more complex saccharification phase to release fermentable products. Despite these caveats, significant ethanol yields can be achieved (Figure 1.6).

Introduction

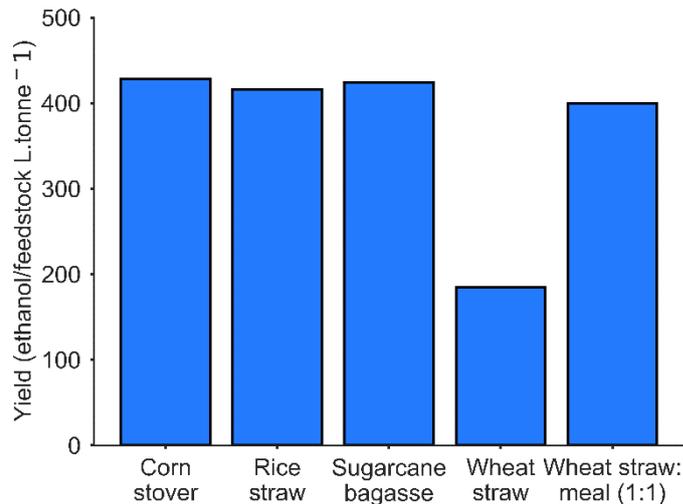


Figure 1.6 Potential ethanol yields from globally abundant grass-derived feedstocks. Data sourced from (Erdei et al., 2010, Naik et al., 2010).

Currently second generation biofuel production consists of a biomass pre-treatment phase (polysaccharide exposure), enzymatic hydrolysis phase (saccharification; monosaccharide release), fermentation phase (ethanol production) and distillation phase (ethanol extraction) (Agbor et al., 2011). During this process a bottleneck exists whereby pre-treatment and saccharification are rate limiting steps (Alvira et al., 2010). The processes for pre-treatments vary but industrially involve acid and base reagents, large volumes of water and high temperatures to strip away the recalcitrant fractions of lignocellulose to expose cellulose microfibrils. The enzymatic hydrolysis of pre-treated biomass is conducted with commercially available cocktails which hydrolyses the remaining biomass into its constituent sugars. These two processes present an opportunity where applied biotechnology has the potential to increase the efficiency of these processes to improve sustainability and process economics.

1.5 Economic constraints to bioprocessing

Despite promising technologies and an overabundance of lignocellulosic feedstock; achieving economic feasibility has proven challenging for second generation bioprocessing. This is largely a result of lignocellulose complexity inhibiting product isolation, purification and deconstruction, which are easily achievable with first generation feedstocks. To overcome this complexity, additional processing stages have been introduced (Figure 1.7). Firstly, exposing the core crystalline cellulose requires pre-treatments to remove the lignin and hemicellulose fractions. This incurs additional

Introduction

capital expenditure and operating costs in pretreatment processing (acids, alkali, ionic liquids which require subsequent neutralisation prior to hydrolysis) and expensive resistant reactors or large energy inputs for the various thermal, thermochemical, or physical pretreatments (Chen et al., 2017a). These pretreatments also generate unavoidable by-products that partially inhibit subsequent hydrolysis and are carried through to also inhibit fermentation (this is exacerbated further as the inhibitors require dilution which reduces ethanol yield) (Chen et al., 2016, Jonsson and Martin, 2016).

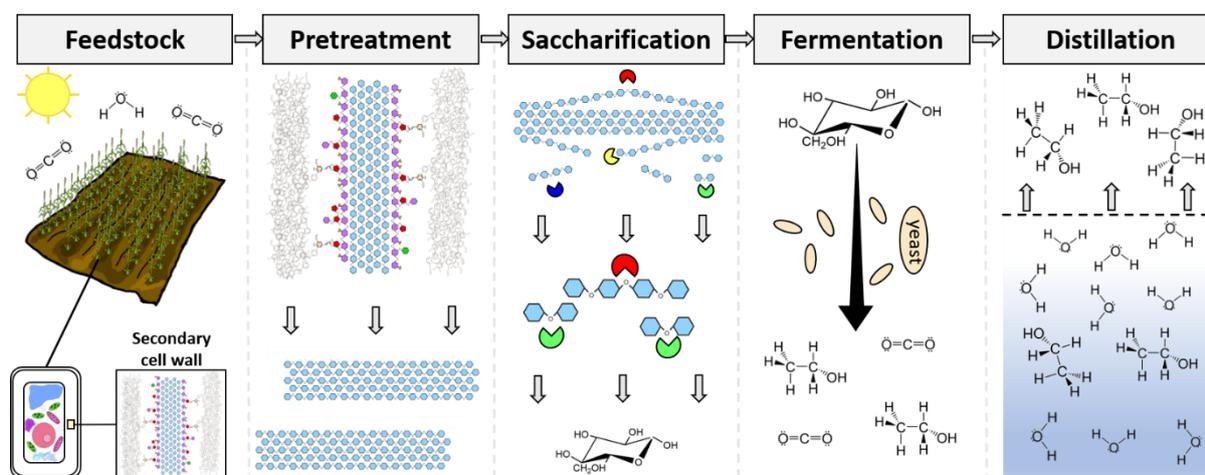


Figure 1.7 General overview of lignocellulosic bioprocessing from primary production to ethanol distillation.

In contrast to first generation biofuels, where starch can be depolymerised with α -amylase (and often glucoamylase), cellulosic polymers require a broader spectrum of enzymes (Kumar and Singh, 2016). Enzyme associated operating costs in lignocellulosic biorefineries are significant and reported to be in the order of up to 30% (Valdivia et al., 2016) and 23-34% (at the time of writing based on 0.34-0.5 USD per gallon) (Humbird et al., 2011). It is hypothesised that economic feasibility can be achieved at 0.1 USD per litre (Rocha-Martin et al., 2017). However, an estimated a cost of 1.47 USD per gallon has been reported which is greater than the cost of ethanol (1.39 USD per gallon at the time of writing, with a three year average of around 1.45 USD per gallon) (Klein-Marcuschamer et al., 2012). These discrepancies are likely the result of disparities in lignocellulosic feedstocks (which differ greatly) and processing methodologies (e.g. pre-treatments) which require case specific analysis. The leading lignocellulolytic cocktail (Cellic[®] CTec2 and more recently CTec3) is produced by Novozymes with a market price of 1.341 USD/10⁶ FPU for CTec2 (Mesa et al., 2017). In addition to this, the commercial cocktails are poorly optimised (they are concentrated culture supernatant

Introduction

containing the secreted lignocellulolytic enzymes of *Trichoderma reesei* and exogenous additives) (Billard et al., 2012).

Water consumption is also significantly increased in second generation processes (while first generation water use is predominantly agriculture related) as a result of the additional complexed pretreatment, hydrolysis and dilution phases. Water intensity (L water/L ethanol) of first generation biofuels require 10-17 L/L (corn starch), whilst switchgrass derived lignocellulosic ethanol (second generation) has a water intensity of 1.9-9.8 L/L (the high range is due to different processing methods), these are comparatively higher than crude oil derived petroleum at 2.8-6.6 L/L (Wu et al., 2009). Other second generation water intensity estimates include 1.24-1.9 L/L (Dutta and Phillips, 2009) and to 5.9 L/L (Fang et al., 2015a). These figures discount the even larger water footprints (500-9,812 L/L) with respect to crop growth (complete bioethanol life cycle) as it references true biomass waste products and not food products (for wheat the footprint is an estimated 4,946 L/L) (Fingerman et al., 2010, Gerbens-Leenes et al., 2009).

The large volumes of fresh water required for bioethanol production imposes water security concerns and constrains economics. In temperate regions such as the UK, where business water rates are in the range of £1.2-1.7/m³, translating to around 0.24p GBP and 1.2p GBP per litre of ethanol based on the water intensities above (Yorkshire Water, UK, small business tariffs), the economic burden is insignificant. However, in water scarce or consistently arid regions water must be purified by reverse osmosis incurring an economic premium. Furthermore, such large volumes of water may not be available (underdeveloped infrastructure to support a biorefineries requirements). Contextually, these volumes of water in regards to processing facilities with a turnover of millions of litres of ethanol are to have excessive impacts locally, or perhaps in the case of arid regions (which is a large proportion of the planet) operationally prohibitive (Liu et al., 2017b, Service, 2009).

Various engineering and structural modifications to biorefining processes have developed to minimise water consumption, with predicted reductions to 0.6-1.6 litres of water per litre of ethanol, however these modifications likely require inhibitory capital expenditure and are unproven at commercial scale (Martin et al., 2011). To completely circumvent the need for water use mitigation, recovery or process trade-offs that may result from resource limitations, the simplest solution is to exchange water as the reaction media for a more suitable alternative. The alternative is required to be inexpensive, renewable and highly abundant, such as seawater.

1.6 Seawater as an economically attractive replacement media

An attractive alternative would be the use of seawater, an abundant and inexpensive solvent, which would liberate fresh water for human consumption and agricultural use, whilst simultaneously acting to enhance sustainability and process economics and ensure water security. This is of particular interest for the UK, with a proposed optimal geographic location for a second generation biorefinery along the Humber estuary, with a near limitless supply of estuarine to marine water (LBnet and E4tech, 2016).

In order to incorporate untreated seawater as the solvent in biorefineries it must be compatible with all other processes. Seawater is a complex chemical substance which has an average pH of 8.15, contains roughly 3.5% salt (84% of which is sodium chloride, and 14% potassium chloride) and over 82 trace elements, complicated further by their prevalence as often diverse chemical species (Elderfield et al., 2006, Turner et al., 1981, Byrne et al., 1988). This chemical complexity is problematic for enzymatic reactions where inhibition can occur, and the ionic strength incurs instability, precipitation and biological inactivation of enzymes.

Commercially available cellulase cocktails include Celluclast[®] (Novozymes), Accellerase[®] 1500 (Genecor) and the Cellic[®] CTec cocktails (Novozymes), all of which perform uniformly (with Celluclast[®] exhibiting lower thermostability), which is unsurprising as they all originate from the same genus of fungi (*Trichoderma*) (Gomes et al., 2018). Accellerase[®] 1500 has been shown to be effective at degrading cellulose in seawater, performing at ca. 90% relative activity (Grande and de Maria, 2012). Similar results have been observed with Cellic[®] CTec2 exhibiting minor inhibition (4.4%) in seawater media, which was hypothesised to be as a result of destabilising cation and anion interactions in the ionic solution (Fang et al., 2015a). This reduction in activity further burdens the economic viability of bioethanol processing.

Additionally, a major flaw in these cocktails is their susceptibility to inhibitors (such as acetic acid, formic acid, furfural, xylo-oligosaccharides and 5-hydroxymethylfurfural (HMF), and an array of aromatic lignin degradation products) produced during pre-treatment processes (Rajan and Carrier, 2014). Seawater pre-treatments also produce higher concentrations of inhibitors (Fang et al., 2015a). This further constrains process economics (also requiring dilution and larger volumes of water).

Introduction

Table 1.1 Major components in the chemical composition of seawater. Adapted after (Atkinson and Bingman, 1997).

| Solute | Composition |
|--|-------------|
| Major Cations (mmol.kg ⁻¹) | |
| Na ⁺ | 470 |
| K ⁺ | 10.2 |
| Mg ⁺² | 53 |
| Ca ⁺² | 10.3 |
| Sr ⁺¹ | 0.09 |
| Sum | 607 |
| Major Anions (mmol.kg ⁻¹) | |
| Cl ⁻ | 550 |
| SO ₄ ⁻² | 28 |
| TCO ₂ | 1.9 |
| TB | 0.42 |
| Sum | 608 |
| Trace (μmol.kg ⁻¹) | |
| Li | 20 |
| Si | 5 |
| Mo | 0.1 |
| Ba | 0.04 |
| V | 0.04 |
| Ni | 0.004 |
| Cr | 0.003 |
| Al | 0.002 |
| Cu | 0.001 |
| Zn | 0.001 |
| Mn | 0.0004 |
| Fe | 0.0001 |
| Cd | 0.0001 |
| Pb | 0.000006 |
| Co | 0.000005 |
| Ag | 0.000001 |
| Ti | 0.000001 |

In consideration of these factors, an opportunity is presented whereby targeted enzymatic bioprospecting of highly active, halotolerant, inhibitor-resistant enzymes can facilitate the construction of unique, highly efficient cocktails. These novel cocktails would be able to operate in untreated seawater, at lower loadings, enhancing prospects for economic and industrial feasibility whilst maintaining water security at local scales.

1.7 Halotolerance

Many enzymes perform poorly in ionic media as they aggregate, precipitate and lose biological functioning. The mechanism of enzymatic functioning under saline conditions are not fully known, however halotolerant enzymes exhibit more acidic surface residues (aspartic acid and glutamic acid) and reduced lysine compared to non-halophilic counterparts, allowing them to have a negative electrostatic surface potential (Enache and Kamekura, 2010). The Hofmeister series describes electro-selectivity of ions and their ion-protein interactions causing proteins to salt in (where ion-protein interaction shields the protein from protein-protein interactions, increasing solubility) and salt out proteins (high ion concentrations cause osmotic dehydration of the protein, creating an ion depleted zone, facilitating protein-protein interactions and precipitation) at increasing ion concentration (Zhang and Cremer, 2006).

Therefore, it is apparent that an important factor affecting protein solubility and stability in ionic solutions is the initial resistance to dehydration. This is putatively a result of a greater negative surface charge in halotolerant proteins as characterised by acidic residues; which form salt-bridges (defined as an interaction between two oppositely charged groups, one of which is within hydrogen bonding distance) between their anionic carboxylate and dissolved cationic salt (which is often hydrated), forming an ordered hydration shell, facilitating the dissolution, stabilisation and functioning of proteins at higher salt concentrations and preventing protein aggregation (Madern et al., 2000, Nayek et al., 2014, Warden et al., 2015). In the context of seawater and the Hofmeister series; Ca^{2+} , Mg^{2+} , Na^+ , K^+ , Br^- and Cl^- are highly abundant destabilisers (Table 1.1) (Fang et al., 2015a). Recently, experimental evidence has been provided by substituting up to 18 surface residues (for aspartic acid and glutamic acid) on a carbonic anhydrase, with the engineered protein displaying significantly greater halotolerance and thermostability than the wild type, but also significantly reduced catalytic activity (Warden et al., 2015).

In addition, halotolerant proteins exhibit a reduced abundance of hydrophobic, and increased abundance of borderline hydrophobic amino acids (glycine, alanine and valine), potentially

alleviating the issues of salt concentration-induced osmotic dehydration by forming a solvation shell on the surface of the proteins maintaining hydration (Delgado-Garcia et al., 2012, Liu et al., 2014). Furthermore, salts appear to play a protective role against denaturants, likely by maintaining critical secondary and tertiary structures, however the exact mechanism remains elusive (Sinha and Khare, 2014). An additional benefit of halotolerance appears to be thermostability, with halotolerant enzymes exhibiting higher optimal temperatures and higher optimal pH than their counterparts (Bonugli-Santos et al., 2015, Sinha and Khare, 2014).

There is evidence that halotolerant proteins also contain a suite of industrially desirable traits (denaturant resistance, thermostability, pH plasticity) as well as functioning in ionic (seawater) media. While evidence suggests replacing surface amino acids with acidic residues may enhance halotolerance, the practicality of such engineering strategies is restricted by the lack of understanding of the mechanisms of halotolerance and the number of enzymes required to hydrolyse lignocellulose efficiently. A more suitable option is to bioprospect novel enzymes from suitable natural environments where selection pressures have driven halotolerant adaptations in lignocellulolytic enzymes.

While lignocellulose is a terrestrial biopolymer not commonly found in the marine (saline) environment, there are three known ecosystems with reliable lignocellulose inputs; *Zostera* seagrass meadows, mangroves and salt marshes. The lignocellulose in mangroves and salt marsh environments is remineralised within the marine domain, suggesting the microbiome contains a spectrum of halotolerant lignocellulolytic enzymes. As such these are model environments to bioprospect for halotolerant and lignocellulolytic enzymes.

1.8 Salt marshes

Salt marshes exist on sheltered coastal sites where sediments accumulate. They are generally found in temperate regions below the position of the 20°C isotherm (sea surface temperature) (Figure 1.8), above this temperature mangroves become the dominant intertidal ecosystem (Saintilan et al., 2014, Mcowen et al., 2017). They are highly diverse intertidal ecosystems which transition from terrestrial to marine environments characterised by halotolerant flora and fauna. Salt marshes are also one of the most productive ecosystems on the planet; with net primary productivity (NPP) estimates exceeding 1 kg C.m⁻².y⁻¹, comparable to intensive agriculture (Morris et al., 2013) (Li et al., 2014).

Introduction

Salt marshes are structurally and biologically heterogeneous with moderate levels of disturbance underpinned by tidal inundation (almost twice daily) from the diurnal cycle (in the UK). Diurnal inundation is the dominant factor inducing temporal fluctuations in pH (7.2-8.3), dissolved oxygen (3-13.5 mg.L⁻¹), temperature (sub zero to 27°C and above), salinity (20-42ppt), pCO₂ and nutrient availability (Baumann et al., 2015, Alldred et al., 2017). It is this abiotic heterogeneity that causes the characteristic plant zonation patterns whereby competitive displacement by greater stress tolerant species creates floral zones perpendicular to the elevation gradient, with competitive subordinates refuged at the less favourable lower marsh (Emery et al., 2001, Schoolmaster and Stagg, 2018).

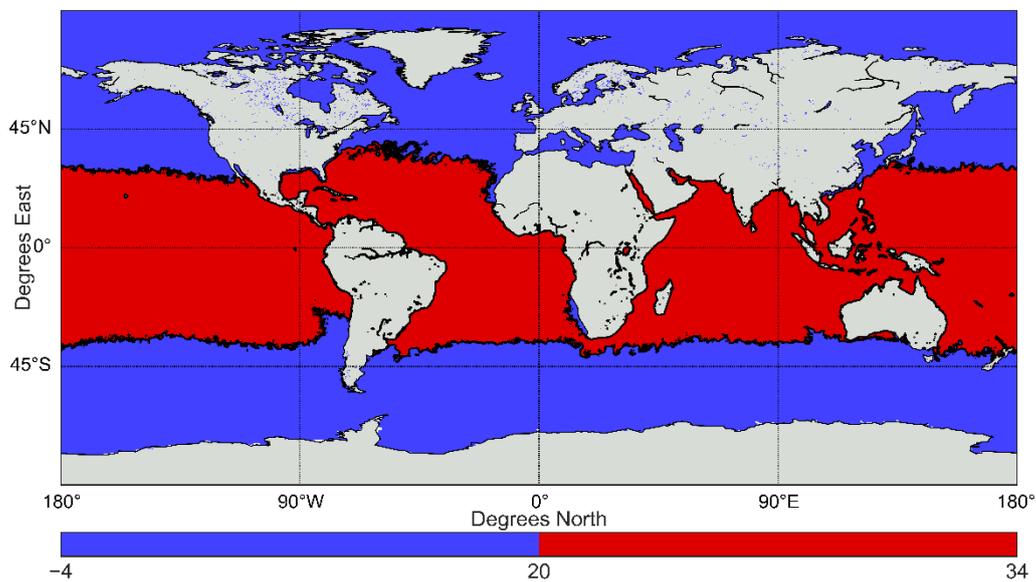


Figure 1.8 Potential global distributions of salt marsh and mangroves ecosystems by sea surface temperature (SST) and 20°C isotherm. Shaded areas represent potential distribution of salt marsh (blue) and mangrove (red) ecosystems. Data plotted is SST on 15/7/2018. Data retrieved from the Physical Oceanography Distributed Active Archive Center and was generated from the Moderate Resolution Imaging Spectroradiometer instrument on the Aqua and Terra satellites (GHRSS, 2010).

Zonation patterns are predominantly the result of a stress gradient that increases with decreasing elevation. Salinity and nutrient availability are highest in the low marsh (increased inundation frequency and length), and lowest in the high marsh. In the UK, halotolerant species, such as *Salicornia europaea* and *Spartina spp.* often dominate the low and low to mid marshes respectively. However, there is significant variation in salt marsh biotope and dominant flora.

The high primary productivity of salt marshes is intrinsically linked to their carbon sequestration capability, with organic carbon burial rates in salt marshes estimated to be 57-218 g m⁻² year⁻¹, which

is significantly higher than any terrestrial ecosystem (Chmura, 2013, Hopkinson et al., 2012). This level of productivity confers salt marshes as the dominant carbon sequestering ecosystem per unit area on the planet, responsible for the global sequestration of 4.8-87.2 billion tonnes of carbon annually (McLeod et al., 2011). Material retention (lignocellulosic biomass) in salt marshes is imperative to their survival (enabling high the levels of carbon sequestration), which is dependent upon maintaining an accretion rate greater than sea level rise (including native biomass deposition and exogenous sedimentation) (Chmura, 2013). Given the predicted rate of increase of sea level rise, understanding the mechanisms of salt marsh accretion and particularly carbon retention and remineralisation is critical (Church et al., 2013). This is particularly pertinent in the context of ocean warming invoking latitudinal encroachment of mangroves onto salt marsh domains (Kelleway et al., 2017).

Salt marsh survival is critically important as they provide a plethora of valuable ecosystem services that directly and indirectly benefit surrounding environments and human populations (Himes-Cornell et al., 2018). Salt marshes provide habitats, refugia and nursery grounds for fisheries (crustaceans, molluscs and fish). Primary production and outwelling exports nutrients to support adjacent environments (Sousa et al., 2017). Regulatory services such as detoxification (sequestration of heavy metals from the water column and nitrogen interception) orchestrated by toxin tolerant halophytes and sedimentation rates (Nelson and Zavaleta, 2012, Andrews et al., 2008, Nikalje and Suprasanna, 2018, Chen et al., 2017b). Carbon sequestration (see preceding text and section 1.9) as a service has also received appropriate attention. However, the most highly regarded service that has become driving factor inspiring salt marsh protection, management and realignment programs are the defensive provisioning by salt marsh ecosystems in preserving and maintaining coastlines. Salt marsh topography and flora are responsible for energetic attenuation, preventing coastal erosion; proving salt marshes as a natural, manageable and sustainable biological barrier between unpredictable and often deleterious tidal effects (Andrews et al., 2008). Therefore, salt marshes act as a naturally occurring analogue to economically constraining engineered coastal defenses which are transient.

To ascribe value to these ecosystem services, monetary values have been estimated based on their propensity to enrich human welfare. An intangible value of \$24.8 trillion (2011) has been ascribed to the collective value of the global ecosystem services provided by salt marshes (de Groot et al., 2012, Costanza et al., 2014). Despite this value, the lack of molecular characterisation of these ecosystems has meant the biotechnological potential of salt marshes to benefit human welfare and industry has been overlooked.

The clear constructive role of salt marshes has been continually disregarded, historically and currently the occupied area of salt marshes is declining, with a global loss estimated at 25% since the 1800s and an annual rate of loss of 1-2% (McLeod et al., 2011). This reduction is more pronounced in the UK with a 50% loss observed since Roman Britain (Airoidi and Beck, 2007). This rate of loss is perpetuated by land reclamation, predominantly for agricultural purposes and coastal development at a global scale. Understanding carbon cycling in salt marshes with respect to organic matter, of which a major constituent is lignocellulose, is of great ecological importance to ensure the security of these ecosystems and subsequent persistence of their invaluable ecosystem services at a time where a growing human population will become increasingly dependent upon them (Chmura, 2013, Morris et al., 2016).

1.9 Salt marsh carbon cycling

Often the dominant plants are perennial grasses (such as *Spartina spp.*), dieback of these plants introduces vast amounts of lignocellulosic biomass into the marine environment. It is thought the majority of remineralisation occurs within the salt marsh (only 0.05% net aerial primary productivity (NAPP) exported); with up to 89% of the native lignocellulosic biomass redistributed to the mid and upper zones (which act as carbon sinks) or deposited on the strandline (Figure 1.9) (Bouchard and Lefevre, 2000, Jackson et al., 1986). The fate of biomass has yet to be fully elucidated but it is highly likely that the majority is degraded on site and insignificant fractions are exported to the surrounding waters with the magnitude of export variable both within and between marshes (Gordon et al., 1985, Hopkinson et al., 2012, Montemayor et al., 2015, Chmura, 2013, Hemminga et al., 1993).

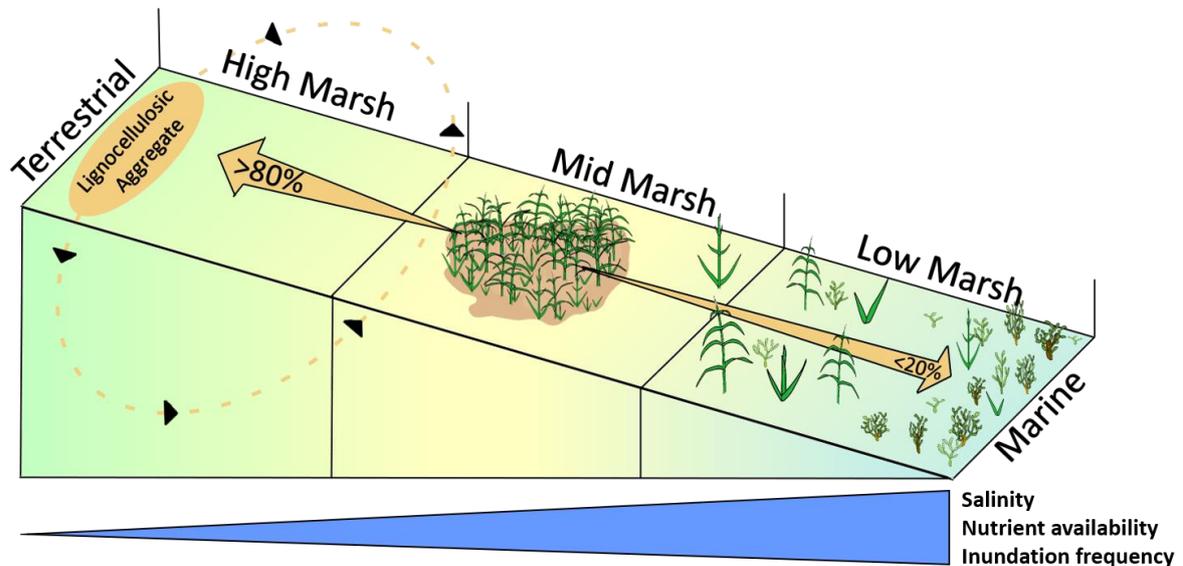


Figure 1.9 Lignocellulose cycling, deposition, redistribution and exportation.

The fallen biomass lies on the surface and becomes buried within the sediments. High organic carbon loadings within salt marsh sediments of $0.73 - 16.2 \text{ kg C m}^{-2}$ have been reported and interestingly these loadings appear to decrease with depth (Drake et al., 2015, Bai et al., 2016, Zhang et al., 2018a). Not only do the carbon loadings decrease with depth, but the composition of the organic carbon also likely changes with depth, with refractory lignin derivatives becoming enriched and polysaccharides being lost (Fogel et al., 1989, Benner et al., 1987). This reduction and profile discontinuity with depth suggests *in situ* decomposition, which is likely orchestrated by microbial consortia (which appear to preferentially target the less recalcitrant hemicellulose and cellulose fractions).

Microbial remineralisation of lignocellulose is an important part of the global carbon cycle whereby lignocellulose is degraded into its constituent sugars, utilised as an energy source and released as CO_2 which may then sequestered by primary producers or persist in the biosphere. However, little is known about carbon cycling within salt marshes (particularly the associated microbiome) beyond productivity and import/export studies. The role of microbial communities has only been elucidated for terrestrial plant pathogens involved in the dieback phase (Torzilli et al., 2006, Benner et al., 1984b).

1.10 Microbial ecology of salt marshes

Salt marshes are considered diverse ecosystems but are relatively unexplored with respect to microbial communities and microbial carbon cycling. Salt marsh microbiomes contain bacterial, archaeal and eukaryotic (including fungal) constituents (Bowen et al., 2012, Weber et al., 2011, Seyler et al., 2014); however, in contrast to terrestrial systems, it is apparent that bacteria dominate salt marsh sediments and are orders of magnitude more abundant than fungi or archaea (Chaudhary et al., 2018). Concordant between studies is the dominance of *Proteobacteria* and *Bacteroidetes* with minor contributions by *Actinobacteria*, *Acidobacteria*, *Chloroflexi*, *Planctomycetes* and *Verrucomicrobia* (Bolhuis and Stal, 2011, Bowen et al., 2012).

Few microbial profiles have been generated for salt marshes, despite this, the few that exist are insightful (Bowen et al., 2012, Bolhuis and Stal, 2011, Mohamed and Martiny, 2011). The most intriguing observation by Bowen et al. (2012) is that the microbial community is homogenous throughout the upper salt marsh sediments and independent of the within marsh location. This is unexpected given the topographically complex and heterogeneous nature of the salt marsh predicated by moderate disturbance (mediated by the temporally variable tidal regime) that creates distinct microhabitats. It was also observed that the suspended community in the water column (drainage creeks) contained a significantly different community. This suggests that these microorganisms have adapted to withstand outwelling (retention by biofilm), and that retention to salt marshes is auspicious. This is unlikely to be the case for organisms originating from terrestrial or marine systems and implicates the community as salt marsh adapted (not interchangeable between systems).

The stable state of the microbiome is synonymous with the observation that various salt marsh microbial profiles are insignificantly effected by various treatments (elevated CO₂ and nitrogen) and marginally effected by salinity gradients (despite a notable decline in freshwater taxa with an increase in salinity) suggesting the salt marsh microbiome is robust in response to factors usually considered deterministic; however, some changes in community structure were observed (Lee et al., 2017, Campbell and Kirchman, 2013). Despite the stable state, seasonal variations have been alluded to (Dini-Andreote et al., 2016a). Recently, microbiome stability has been demonstrated experimentally with resilience post microbiome translocation (Angermeyer et al., 2018). The taxonomic diversity within salt marsh sediments likely hosts a plethora of functions capable of offsetting environmental changes, conferring resilience.

Introduction

The microbial consortia vary marginally with elevation (between high and low marsh), but varies significantly with depth between 0-10cm (Cleary et al., 2017, Bowen et al., 2012). Profile change with depth is likely a function of oxygen depletion, with anoxic sediments hosting a suite of alternative respiratory modes characteristic of specific taxonomic lineages, whose distribution is correlated to the stratification of their required terminal electron acceptor. This trend extends to oceanic sediments which exhibit similar microbial profiles to salt marsh sediments where *Bacteroidetes*, *Firmicutes*, *Chloroflexi* and *Proteobacteria* become enriched at moderate and extreme depths, though no core-depth profiles exist for salt marshes (Petro et al., 2017). Salt marshes are topographically complex and moderately disturbed ecosystems; however, only the upper layer (1-5mm) of sediment is consistently oxygenated. It is, therefore, likely that different microbial consortia interact with lignocellulose at surface levels (containing pristine lignocellulose) and subsurface (containing partially degraded lignocellulose) sediments.

The carbon cycling associated microbial profiles are largely understudied (Vigneron et al., 2014), with current research focusing on nitrogen cycling (Dini-Andreote et al., 2016a, Salles et al., 2017, Kearns et al., 2015), iron and sulfur cycling (Moreau et al., 2010, Lambais et al., 2008, Angermeyer et al., 2016, Thomas et al., 2014), bioremediation/pollution (Sanni et al., 2015, Beazley et al., 2012), succession dynamics and abiotic determinants (Dini-Andreote et al., 2014, Bowen et al., 2009), or in-depth holistic assessments (Bowen et al., 2012). Interestingly, whilst the compositions may vary slightly, salt marsh microbiomes appear analogous with those associated with marine organic matter rich sediment (sapropel), reflecting the density of organic matter in this environment; and implicating some of the groups contained within these taxa as having roles within the carbon cycle, however taxonomic associations to carbon cycling have yet to be demonstrated (Oni et al., 2015, Andrei et al., 2017).

While high taxonomic resolution salt marsh microbiome data is scarce, at class levels; *Acidimicrobiia*, *Alteromonadales*, *Flavobacteriia*, *Rhodobacterales* and *Thiotrichales* appear to be more abundant in surface sediments, while *Anaerolineae* and *Deltaproteobacteria* were more abundant at depth (Cleary et al., 2017). The function of these highly abundant groups within salt marshes has yet to be established. Environmental microbial carbon cycling studies are lacking; however, *in vitro* studies using ¹³C lignocellulose have revealed *Gammaproteobacteria* (*Kangiella*), *Bacteroidetes* (*Flavobacteria*) and *Spirochaeta* as benefactors of degradation products (Darjany et al., 2014). In consideration with their respective abundance in salt marsh sediments, these findings would implicate a large proportion of the sediment microbiome is lignocellulose responsive; however, this only provides evidence for the utilisation of degradation products by these taxa and

not direct involvement in the degradation process. Additionally, *in vitro* studies are divorce from reality and ascribing ecological significance from their findings requires prudence.

Proteobacteria are one of the largest prokaryotic phyla consisting of seven classes; Alpha-, Beta-, Gamma-, Delta-, Epsilon-, Zeta-*proteobacteria* and *Acidithiobacillia*; however, the former-most five are consistently observed in abundance in marine samples (Williams and Kelly, 2013). Whilst a diversity of functions likely exists within each class, attempts have been made to assign putative functions to these taxa. In intertidal settings, *Alphaproteobacteria* and *Betaproteobacteria* have been associated with nitrogen cycling (Kearns et al., 2015) and rhizosphere sediment (Wang et al., 2007), with *Alphaproteobacteria* more abundant in aerobic conditions (Lambais et al., 2008). *Betaproteobacteria* are known ammonia oxidisers (*Nitrosomonas* and *Nitrospira*) together with *Gammaproteobacteria* (*Nitrosococcus*), with the former ostensibly more abundant (Ma et al., 2018, Nold et al., 2000). The *Alphaproteobacterial* order *Rhodobacterales* are associated with sulfur oxidation in salt marsh sediments (Thomas et al., 2014).

Gammaproteobacteria are a diverse group of bacteria, with a diverse suite of ecological functions including C, N and S cycling. Within salt marshes this group is predominantly associated with carbon cycling, with over 50% of the dark carbon fixation (carbon fixation in the absence of light) in salt marshes attributed to this group (Dyksma et al., 2016). Dark carbon fixation is poorly understood but is thought to utilise available reduced and oxidised compounds to liberate energy for carbon fixation independent of light (Santoro et al., 2013).

Abundant salt marsh dwelling *Gammaproteobacteria* include the orders *Alteromonadales*, *Chromatiales*, *Thiotrichales* and *Oceanospirillales* (Cleary et al., 2017). The former are considered oligotrophic, capable utilising a diversity of compounds, therefore it is unsurprising that they have been observed in abundance in coastal microbiomes, rich in chemical diversity, alleviating direct competition for more valuable substrates (Cho and Giovannoni, 2004, Sun et al., 2015b). *Chromatiales* and *Thiotrichales* are interesting as these groups are putatively involved in sulfur cycling; however, unlike the related sulfur-reducing *Deltaproteobacteria*, are photosynthetic sulfur-oxidising organisms (*Chromatiales* are commonly known purple-sulfur oxidising bacteria) (Pokorna and Zabranska, 2015, Thomas et al., 2014, Cleary et al., 2017). Groups closely related to *Oceanospirillales* have been observed as dominant in oil plume associated microbiomes, representatives of this group are known hydrocarbon degraders, capable of utilising many lignocellulosic breakdown products as a carbon source (Hazen et al., 2010, Yakimov et al., 2003, Beazley et al., 2012, Wang et al., 2015a).

Introduction

The *Deltaproteobacteria* contain established sulfate reducing bacteria (SRB) (Karlin et al., 2006). In salt marshes the *Desulfobacteraceae* and *Desulfovibrionaceae* are reported as highly abundant (Cui et al., 2017). These organisms are capable of respiring anaerobically, reducing sulfate (SO_4^{2-}) to hydrogen sulfide (H_2S). This action facilitates the production of insoluble iron sulfides (which give anoxic salt marsh sediments the characteristic dark-grey hue) by reduction of iron oxyhydroxides by hydrogen sulfide (Coleman et al., 1993). Additionally, Fe(III) reduction to Fe(II) by iron reducing bacteria (FeRB) has been associated with *Deltaproteobacterial* members within salt marshes (FeRB are putatively present throughout the kingdoms bacteria and archaea) (Lowe et al., 2000).

Bacteroidetes are often one of the most abundant phyla in salt marshes, particularly the classes *Bacteroides*, *Flavobacteriia* and *Sphingobacteriia* (Bolhuis and Stal, 2011, Lydell et al., 2004). These classes are often cited as capable of complex polymer degradation and as such are known as heterotrophic organic matter decomposers (including lignocellulosic material) in salt marsh ecosystems (Cleary et al., 2017, Lydell et al., 2004). *Flavobacteriia* (Bauer et al., 2006), *Sphingobacteriia* (Zhou et al., 2017) and *Bacteroides* (Jimenez et al., 2015a) are all known to be equipped with carbon deconstructing hydrolytic enzymes.

High abundances of the phylum *Chloroflexi* have been observed in salt marsh and marine sediments but at higher abundances at depth in aromatic compound rich sediments (Oni et al., 2015, Vigneron et al., 2014, Petro et al., 2017). The putative propensity of *Chloroflexi* in degrading aromatic compounds, which likely originate from the recalcitrant lignin fraction in lignocellulose, would implicate them as carbon cyclers, however this group is understudied in the context of salt marshes.

Fungal isolates originating from salt marshes are thought to be important contributors to salt marsh carbon cycling with lignocellulolytic capability (Torzilli et al., 2006, Buchan et al., 2003). Many salt marsh fungi are plant pathogens that act upon standing senescent flora, which target lignocellulose, to gain internal access to nutritionally rich cell contents (Walker and Campbell, 2010). The standing matter is not fully inundated with seawater, and can not be considered truly marine, as such, the fungal representatives described in the studies above may be of terrestrial origin. Despite the lack of characterisation, more recent research rebukes the role of halotolerant fungi as saprotrophs in salt marshes determined *in vitro* (Cortes-Tolalpa et al., 2018). The fungal profile within salt marsh sediments is dominated by *Ascomycota*, with minor contributions by *Basidiomycota*, *Glomeromycota* and *Chytridiomycota*, similar to bacteria, the profile does not change with the salinity gradient, but is altered by the presence of plants; however, the function of these lineages in this ecosystem is unknown (Mohamed and Martiny, 2011).

The representative archaea in salt marshes have not been studied in depth; however, considering their abundance relative to either bacteria or fungi, their contribution to ecosystem functioning is likely minor. They are thought to be involved in nitrogen (nitrification) and carbon (methane oxidation) cycling (Moin et al., 2009, Nelson et al., 2009). *Crenarchaeota* appear to be dominant; however, our understanding of archaeal profiles is still rudimentary (Li et al., 2012).

1.11 Salt marshes as an underexplored model ecosystem with biotechnological potential

Salt marshes present an understudied marine ecosystem, where terrestrial biomass is produced, enters the intertidal zone and becomes remineralised whilst exposed to saline and abiotic stresses. It is likely that native organisms have evolved and adapted to produce halotolerant lignocellulolytic enzymes. The organic carbon rich sediments, rich in lignin derived aromatic compounds has likely evolved enzymes resistant to the inhibition that is industrially problematic (pretreatment derived inhibitors). Therefore, salt marshes are a model environment to bioprospect not only halotolerant enzymes, facilitating the use of seawater (alleviating water security concerns), but also inhibitor tolerant enzymes, that may require no dilutive measures, reducing enzyme loadings and improving operational expenditure. The high levels of diversity and productivity in salt marshes likely instigates intense competition that may have selected for high rates of activity, driving the evolution of highly active enzymes; biotechnological exploitation of which could reduce enzyme loadings during bioprocessing and improve process economics. The identification of a spectrum of enzymes will enable the construction of a synthetic cocktail with optimisable plasticity that is not achievable in current commercial cocktails.

1.12 Enzymatic lignocellulose deconstruction

Lignocellulose has evolved to resist biological and physical attack and confer structural rigidity for over 450 million years, the result of which is an innately complex heteropolymer with significant variations in composition and structure across the kingdom *Plantae* (Weng and Chapple, 2010). This complexity has cultivated a spectrum of strategies for deconstruction across the tree of life (Cragg et al., 2015). The strategies are so vast that novel enzymes and families are continually being discovered (Sabbadin et al., 2018). As such it is likely the extent of identified lignocellulolytic enzymes is outweighed by the magnitude of current unknowns.

Introduction

Lignocellulolytic microbes and enzymes are ubiquitous and have been identified from many terrestrial and marine environments including; seawater (Hirasawa et al., 2006), deep sea sediments (Mai et al., 2013), algae (Trivedi et al., 2011, Trivedi et al., 2013, Martin et al., 2014), salt lakes (Li and Yu, 2013, Wejse et al., 2003), salterns (Prakash et al., 2012), hydrothermal vents (Hung et al., 2011), mangroves (Niladevi et al., 2008), termites (Kashyap et al., 2014), soils (Li et al., 2015), leaf litter (Akintola et al., 2017), invertebrates (Guo et al., 2013) and rumen (Pang et al., 2017, Wang et al., 2015b). The identification and discovery of these enzymes has accelerated due to the introduction of next generation sequencing technologies. This has facilitated the use of novel techniques such as metatranscriptomics, metaproteomics and metagenomics which can be integrated to enable the identification of organisms, genes and enzymes independent of cultivation and isolation/purification methodologies (Guo et al., 2018). These novel techniques often provide orders of magnitude more information than conventional approaches and in conjunction with curated reference databases, facilitate the probing of a microbiome or individual organisms functioning at a resolution not otherwise achievable. However, the value of the information generated is determined by the depth of knowledge stored in curated databases, which often poorly capture novelty. To address this, various curated databases with a focus of lignocellulolytic enzymes have been developed to aid in the identification and characterisation of this data through homologous classification.

1.12.1 Enzymatic classifications

To capture and categorise the diverse suite of lignocellulolytic enzymes an archive database (Carbohydrate Active Enzyme database; CAZy) has been constructed (Lombard et al., 2014). In this database enzymes are classified based on homologies within their amino acid sequence, which correlates with their mechanism of action rather than enzymatic specificity (Cantarel et al., 2009, Henrissat, 1991). These classifications are also combined with a nomenclature often descriptive of their function when experimentally characterised enzyme information is available (Cantarel et al., 2009). As a result, these databases assist in the rapid conservative assignment of potential enzymatic homology and function to newly generated sequence data, while definitive characterisation still requires experimental evidence.

The classifications can be grouped into classes, of which there are six; Glycoside Hydrolases (GH) are the most diverse and include glycosidases and transglycosidases (Henrissat and Davies, 1997), Glycosyltransferases (GT) (Coutinho et al., 2003), Polysaccharide Lyases (PL) (Lombard et al., 2010), Carbohydrate Esterases (CE), Auxiliary Activities (AA) which indirectly catalyse carbohydrate

deconstruction (Levasseur et al., 2013) and Carbohydrate Binding Domains (CBM) (Boraston et al., 2004). Further resolution has been achieved by subdivision of classes into families, of which for each family there must be an experimentally biochemically characterised enzyme published in peer-reviewed scientific literature (Cantarel et al., 2009). Currently there are 153 GH families, 106 GT families, 28 PL families, 16 CE families, 15 AA families and 83 CBM families described. The databases are continuously maturing and frequently updated by merging and splitting of families (dependent on newly available information). Further resolution is available at the sub-family level where enzymes groups are derived from a putative common ancestor, and in the form of clans, which consider structure in the form of folds and catalytic symmetry (presently only available for the GH class) (Lombard et al., 2014, Cantarel et al., 2009).

1.12.2 Cellulose deconstruction

As cellulose is a linear polysaccharide composed of repeating β -(1,4) *D*-glucose units, cellulases are predominantly β -(1,4) glycoside hydrolases, that differ in how they interact with cellulose and the size of the products produced. Cellulose is deconstructed by a group of enzymes collectively known as cellulases. The comparative simplicity of cellulose polymers to hemicellulose and lignin exhibits a disparately large diversity of enzyme families, with cellulases present in classes GH1, GH3, GH5-9, GH12, GH44, GH45, GH48 and GH74 (Talamantes et al., 2016). This likely reflects the ubiquitous utilisation of cellulosic glucose as a carbon source across the tree of life. The majority of cellulases are hydrolytic and catalyse the cleavage of glycosidic ether linkages by incorporation of a water molecule, however a novel species is oxidative (see section 1.11.3). They are broadly categorised into endo- acting; which act within the cellulose chain, exo- acting; that act at chain termini and β -glucosidases; that hydrolyse variable chain length oligosaccharides into glucose (Figure 1.10). These groups act synergistically to depolymerise crystalline and amorphous cellulose regions into soluble glucose and gluco-oligosaccharides.

Introduction

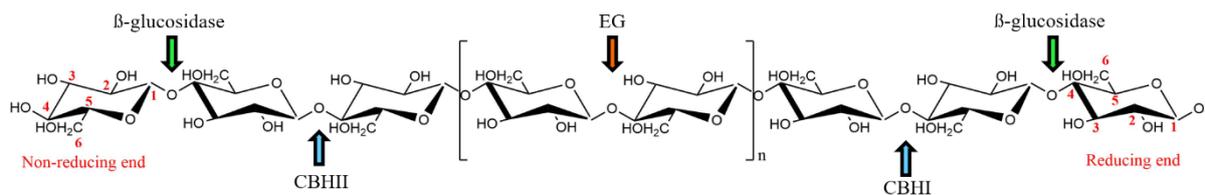


Figure 1.10 Synergistic action of cellulase targeting of a cellulose polymer. Cellulose is displayed with termini as reducing and non-reducing ends. EG; endo- β -(1,4)-glucanase, CBH; exo- β -(1,4)-glucanase (cellobiohydrolase).

Endo- β -(1,4)-glucanases are considered integral to efficient cellulose deconstruction, such that optimised cellulose cocktails consist of up to 20-40% endoglucanases (Banerjee et al., 2010, Billard et al., 2012). This is due to their unique ability to cleave glycosidic linkages at random internal positions within the cellulose polymer, this subsequently depolymerises the chain, multiplying accessible initiation sites that are rate limiting for exo-acting enzymes. This category of endoglucanases acts more efficiently on amorphous regions of cellulose (Reyes-Sosa et al., 2017, Billard et al., 2012). Endoglucanases exhibit an architectural cleft (Figure 1.11C-D) containing their active site that facilitates interaction with regions of cellulose polymers and their subsequent catalytic cleavage (Naas et al., 2015).

Exo- β -(1,4)-glucanases are the dominant component in commercial and optimised cellulose cocktails, constituting up to 40-60% (Billard et al., 2012). They are collectively known as cellobiohydrolases (CBH) for their production of cellobiose and cellotriose from cellulose polymers and oligosaccharides. CBH act at cellulose polymer termini, from either the non-reducing end (GH6) or the reducing end. In contrast to endoglucanases that display an exterior binding groove, CBH exhibit an internal active site tunnel (active tunnel, Figure 1.11A-B) (Colussi et al., 2015). A single cellulose strand becomes threaded through the active tunnel, where a processive mechanism occurs cleaving soluble sugars as the enzyme moves along the strand (mediated by complex catalytic interactions between the strand and enzyme domains) (Payne et al., 2013). These enzymes are capable of hydrolysing amorphous cellulose regions, but also putatively capable of stripping and hydrolysing strands from the surface of the crystalline cellulose structure, and as such display greater levels of activity than endoglucanases (Payne et al., 2013, Billard et al., 2012).

1,4- β -glucosidases are another important component in the cellulose deconstruction process. These enzymes are responsible for catalysing the hydrolysis of glycosidic linkages in oligosaccharides (released by endo- and exo-acting enzymes) to release glucose, the most desirable product for downstream bioprocessing (Singh et al., 2015). They are the least abundant cellulase component in

commercial and optimised cocktails but are crucial in creating a uniform profile of products during cellulose deconstruction (Banerjee et al., 2010).

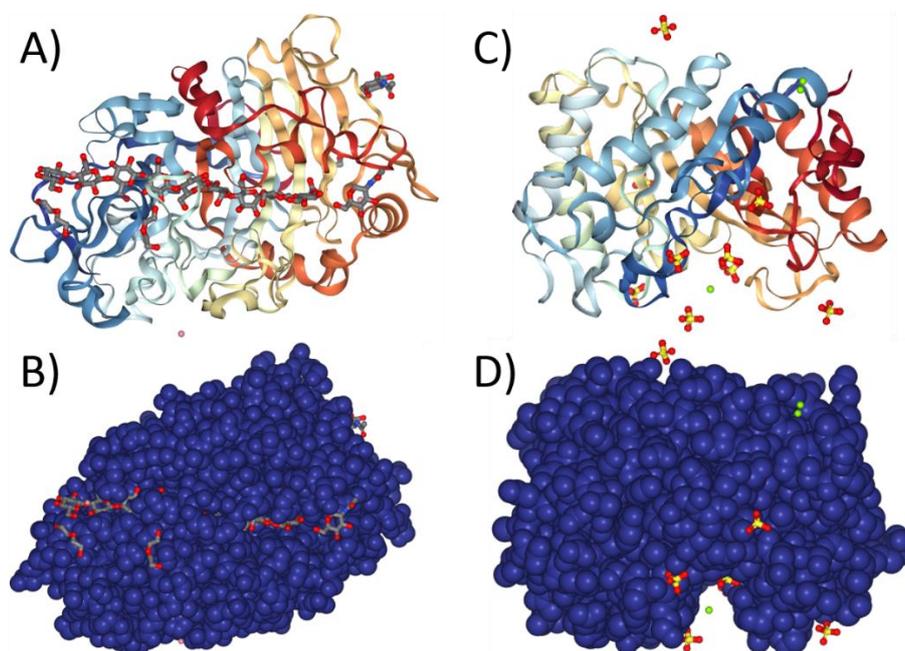


Figure 1.11 Structures of endo- and exo- β -1,4 glucanases. A-B) *Trichoderma reesei* GH7 Cel7A E217Q exo- β -1,4 glucanase (cellobiohydrolase) with bound cellononaose spanning the active tunnel, PDB accession; 4C4C (Knott et al., 2014). C-D) *Trichoderma reesei* GH5 Cel5A endo- β -1,4 glucanase displaying the large open cleft, PDB accession; 3QR3 (Lee et al., 2011).

The optimal composition of a cellulose cocktail is dependent on the feedstock, origin of the feedstock and the pre-treatment which affects the cellulose crystallinity, degree of polymerisation and type and concentration of inhibitor. Current commercial cocktails are biologically derived from *Trichoderma reesei*, which, whilst equipped with a comparatively low amount of cellulases (seven endoglucanases, two CBH and one oxidative acting cellulase at the genome level), is highly productive with yields of 100g.L (Martinez et al., 2008). However, this form of production limits plasticity and hinders efforts to optimise cocktails as the cocktail foundation is inalterable (excluding genome engineering approaches). Therefore, the only avenue for improvement is additives, which is the path chosen to improve Cellic® cocktails, with the addition of missing synergistic families; AA9, β -glucosidases and xylanases (Sun et al., 2015a, Druzhinina and Kubicek, 2017, Takashima et al., 1999). A bottom up approach would be the identification of promising enzymes that could be used interchangeably to construct optimised synthetic cocktails for each unique application.

1.12.3 Hemicellulose deconstruction

Hemicellulose is more heterogeneous than cellulose by incorporating both hexose and pentose sugars with variable degrees of branching which is complexed further by the presence of decorations. Similarly, to cellulose, the xylan backbone polysaccharide, which is common feature of dominant polysaccharides in grasses, consist of linear β -(1-4) linked xylosyl residues is targeted by endo- acting, exo- acting, terminal- acting hydrolases and oxidative enzymes.

The xylan backbone is deconstructed by endo- β -(1-4)-xylanases, exo- β -(1,4)-xylanases (xylobiohydrolases), and 1,4- β -xylosidases, which are the xylan degrading equivalents of endoglucanases, cellobiohydrolases and β -glucosidases respectively (Figure 1.11) (Walia et al., 2017, Ravanal et al., 2010). These enzymes act synergistically to deconstruct the xylan polysaccharide into smaller xylo-oligosaccharides and xylose monomers. They belong to the families GH1, GH3, GH5, GH8, GH10, GH11, GH30, GH39, GH43, GH51-52, GH54, GH98, GH116 and GH120.

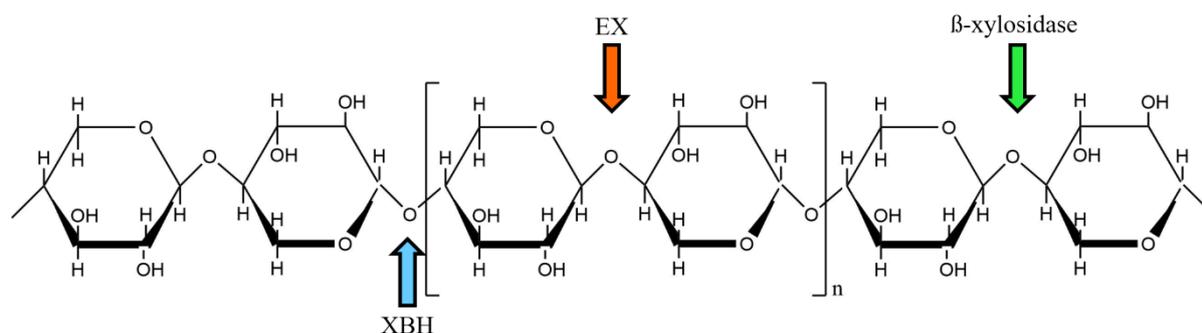


Figure 1.12 Synergistic action of xylan targeting of an undecorated xylan backbone. EX; endo- β -(1,4)-xylanase, XBH; exo- β -(1,4)-xylanase (xylobiohydrolase).

Grass xylan is extensively decorated with acetyl, arabinosyl, galactosyl, xylosyl and, importantly in grasses, ferulic acid residues (methylated and non-methylated glucuronic acid is not an abundant decoration in grasses) (Marriott et al., 2016, Vogel, 2008, Pauly et al., 2013). This level of heterogeneity requires an equal diversity of destructive enzymes to accommodate deconstruction. The acetyl decorations are thought to contribute to the structural coupling of xylan to cellulose microfibrils, while the other decorations and branched substituents are thought to retain an amorphous structuring (Gomez et al., 2008, Busse-Wicher et al., 2014). These decorations act to prevent accessibility to the backbone and inhibit degradation, therefore a spectrum of hemicellulases are required to act synergistically to efficiently depolymerise matrix polysaccharides (Figure 1.12).

Introduction

The most abundant decoration is the O-acetylation of xylosyl residues within the backbone polysaccharide. Deacetylation is catalysed by acetyl xylan esterases (AXE) which belong to the carbohydrate esterase (CE) families CE1-7 and CE12. The proximity of xylan to cellulose creates a physical barrier that inhibits the action of cellulases. Enzymatic mediated dissociation (solubilisation) of xylan from cellulose as a result of deacetylation has improved cellulose hydrolysis, likely by virtue of increased accessibility (Selig et al., 2008, Zhang et al., 2011a). A similar effect is observed for xylanases where acetylated xylosyl residues inhibit their activity (as a function of structural hinderance) (Selig et al., 2009). Therefore, potential exists to alter xylan backbone structure through deacetylation and subsequent solubilisation to promote cellulose hydrolysis for substrates with residual matrix polysaccharides post pretreatment.

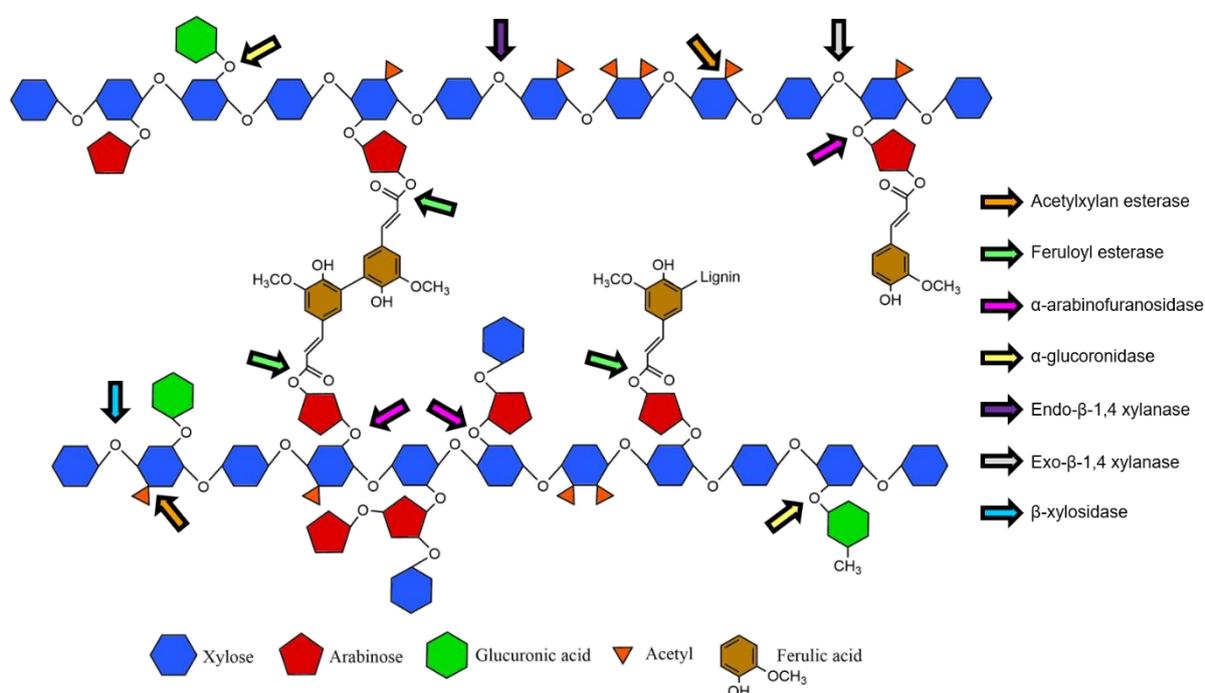


Figure 1.13 Schematic of major hydrolytic enzymes targeting glucoarabinoxylan. Predominant polysaccharide linkages in grasses are displayed with a simplified molecular architecture of glucoarabinoxylan.

Other xylan substituents such as arabinosyl and galactosyl residues which branch off the acetylated backbone polymer. The former is a minor component in grass xylans and is deconstructed by α -glucuronidases belonging to families GH4, GH67 and GH115, however their propensity to promote the hydrolysis of cellulose of matrix polysaccharide hydrolysis is unknown (Gilbert, 2010, Rogowski et al., 2014). Arabinosyl side chain residues are cleaved by α -arabinosidase (Ohta et al., 2010) however, the ability of the arabinosyl residues to cross link to ferulic acid has provided a more

appropriate focal point. The side chains (Araf) can be cleaved by α -L-arabinofuranosidases to simplify the xylan backbone and (for ferulic acid cross linked Araf chains) destabilise the lignin-carbohydrate complex anchor, freeing previously inaccessible polysaccharides and increasing both matrix polysaccharide (Raweesri et al., 2008) and cellulose hydrolysis (Alvira et al., 2011). These enzymes belong to the families; GH2, GH3, GH43, GH51, GH54 and GH62.

The incorporation of ferulic acid (and *p*-coumaroyl) arabinose cross linkages are an important structure in the secondary plant cell walls in grasses, forming the major covalent intra- cross linkages to other polysaccharides and inter- linkages between matrix polysaccharides and lignin (Vogel, 2008, de Oliveira et al., 2015). The ester linkage between arabinosyl and ferulic acid can be hydrolysed by a class of carbohydrate esterases known as feruloyl esterases for their ability to catalyse the hydrolysis of ferulate cross linked to polysaccharides (Li et al., 2011) . These enzymes belong to the CE1 class and have been shown to improve cellulose hydrolysis from both acid and alkaline pretreated straw (Badhan et al., 2014, Tabka et al., 2006).

1.12.4 Oxidative polysaccharide deconstruction

Recently, another group of enzymes known as lytic polysaccharide monooxygenases (LPMOs), previously catalogued as GH61, have been shown to act upon cellulose. Common to all LPMOs is a structure known as the histidine brace, comprising two histidine residues that coordinate a copper atom and a single tyrosine residue which form the active site (Hemsworth et al., 2013). In contrast to hydrolytic enzymes, LPMOs are metalloenzymes that requires molecular oxygen and electron donors (two electrons per cleavage event) or hydrogen peroxide (Bertini et al., 2018). Electron donors include cellobiose dehydrogenase, ascorbate, gallic acid and lignin derived phenolics (Courtade et al., 2016, Frommhagen et al., 2017). Glycosidic cleavage is orchestrated at the copper containing active site (captured by the N terminal histidine situated within) that cycles between resting Cu (II) and active Cu (I) states that captures molecular oxygen to catalyse a peroxygenase type reaction, however the mechanism has yet to be fully elucidated (Wang et al., 2018a, Frandsen et al., 2016, Westereng et al., 2015, Hemsworth et al., 2013).

LPMOs currently span six classes; AA9 (exclusively fungal and active on cellulose and hemicellulose), AA10 (exclusively bacterial, active on chitin and cellulose), AA11 (exclusively fungal and active on chitin) and AA13 (exclusively fungal and active on starch), AA14 (active on xylan) and AA15 (active on cellulose and chitin), of which only AA9, AA10 and AA15 have been observed to be active on cellulose (Bertini et al., 2018, Couturier et al., 2018, Sabbadin et al., 2018). Significant

interest around LPMOs has stemmed from their ability to act synergistically with GH enzymes to increase rates of activity and reduce enzyme loadings (Harris et al., 2010). This is concurrent with the finding that LMPO addition to commercial enzyme cocktails release up to 60% more glucose, translating to reduced enzyme loadings or incubation times (Muller et al., 2015).

1.12.5 Carbohydrate binding modules

Carbohydrate binding modules (CBMs) are accessory domains that are often associated with catalytic domains through a flexible polypeptide linker with a high density of serine, proline and threonine residues (Shoseyov et al., 2006, Jensen et al., 2018). CBMs are thought to interact with substrates, retaining prolonged intimate catalytic proximity and therefore improving reaction rates, particularly for soluble enzymes targeting insoluble substrates such as lignocellulose (Guillen et al., 2010, Duan et al., 2016). Increased reaction rates are energetically favourable for the secreting organism. CBM truncations have experimentally revealed the benefit of these domains to catalytic efficiency (Meng et al., 2015, Jensen et al., 2018).

Despite increased rates of activity for catalytic domains conjugated to CBMs, the majority of cellulases (greater than 60%) appear to lack CBMs (Varnai et al., 2013). CBMs are considered industrially problematic as they increase protein size and reduce yields during production. Reportedly, increased substrate concentrations can replicate the benefit of CBMs for truncated mutants (Jensen et al., 2018). This is likely a function of adaptation to high and low substrate concentration environments, such as terrestrial leaf litter solids and an inundated salt marsh respectively. The effectiveness of CBMs likely require assessment on an individual enzyme and situation basis. Alternative advantageous CBM usage has been demonstrated and exploited through constructing synthetic architectures not found in nature, this method has successfully increased the catalytic efficiency of various enzymes (Crouch et al., 2016, Reyes-Ortiz et al., 2013).

Additionally, CBMs are thought to disrupt and moderate the structure of the substrates it binds to; however, this is a relatively unexplored and contentious feature (Din et al., 1991, Shoseyov et al., 2003, Wang et al., 2008). There are currently 83 CBM families (CBM33 was reclassified to AA10) with known binding associations to cellulose, most hemicellulose polysaccharides and chitin. At present, there are no known lignin binding modules, however recently lignin binding peptides were discovered which act similarly to peptide aptamers (Yamaguchi et al., 2016, Isozaki et al., 2018). This diversity provides a library for the construction of synthetic architectures to target selected substrates.

1.12.6 Lignin deconstruction

As the most recalcitrant cell wall polymer, lignin is usually removed during pretreatment processes. The effectiveness of these pretreatments and the lack of high value lignin derived products until recently has reduced attention toward enzymatic lignin depolymerisation, however modification of lignin by various enzyme families is known. These enzymes are produced by fungi and bacteria and are oxidoreductases known as laccases or peroxidases (Abdelaziz et al., 2016).

Laccases (benzenediol: oxygen oxidoreductases) are often multi-copper containing enzymes capable of oxidising an array of electron rich substrates through abstraction; when molecular oxygen is the final electron acceptor water is produced as a by-product (Mate and Alcalde, 2017). Laccase activity is therefore depending on the redox potential of the copper core containing active site, with greater redox potential cores capable of abstraction from greater redox potential substrates (Gunne et al., 2014). Laccases have historically been thought to be restricted to phenolic lignin substituents, due to their low redox potential, however their mechanistic nature confers a diversity of functions such as decarboxylation, demethylation and demethoxylation of phenolic substituents (Christopher et al., 2014). The development of cyclic laccase-mediator systems has overcome the limitations of low redox potentials and expanded the laccase substrate repertoire to include non-phenolic lignin substituents. Mediators are capable of acting as an electron shuttle, that, after laccase catalysed oxidation, abstracts an electron from the substrate, diminishing the direct role of the laccase (Baiocco et al., 2003). However, this system has considerable caveats, the dominant of which is unstable mediator intermediates (Christopher et al., 2014). Despite potential lignin degrading applications, laccases are considered useful in purifying phenolic hydrolysis inhibitors from prehydrolysate generated during pretreatment (Fang et al., 2015b). Recently, enzymatic lignocellulose pretreatment with laccases was successfully reported with an 80% hydrolysis yield, however laccases inactivate the hydrolytic cocktail requiring an additional laccase-inactivation phase (Giacobbe et al., 2018).

Lignocellulose acting peroxidases are class II heme-peroxidases (Zamocky et al., 2015). Class II peroxidases are diverse and include lignin- (LP), manganese- (MnP) and versatile peroxidases (VP) which are powerful oxidants that exhibit great redox potentials, and therefore a larger range of phenolic substituents as substrates than laccases (Christopher et al., 2014, Abdelaziz et al., 2016). LPs target non-phenolic constituents whilst MnPs and VPs target phenolic and non-phenolic substituents. Peroxidases abstract electrons from lignin substituents similarly to laccases to oxidise and depolymerise lignin. Despite these strengths, laccases are favoured due to the requirement of

peroxidases for H₂O₂ as an electron acceptor, which they are also inhibited by (Nicell and Wright, 1997).

All known mechanisms of enzymatic lignin modification require oxygen which is known to be rate limiting (Ortner et al., 2015). This requires an exogenous oxygen supply considered a processing burden (Huber et al., 2016). As a result, incorporation of these enzymes in bioprocessing has been understudied until recently due to their additional benefit of detoxifying the prehydrolysate of inhibitors. There is evidence of anaerobic lignin deconstruction from microbial consortia harvested from salt marsh ecosystems (Benner et al., 1984a). This suggests depolymerisation in the absence of oxygen; however, the mechanisms have yet to be elucidated and further research into this phenomenon could alleviate these industrial constraints.

1.13 Aims

This thesis aims to apply an integrated approach encompassing the consolidation of metatranscriptomics, metasecretome proteomics and metagenomics to explore salt marsh lignocellulose deconstruction. The study aims to use the generated data to characterise the decomposition process by identifying lignocellulolytic enzymes from the native salt marsh microbiome and identify key functional carbon cycling associated taxonomic groups. In addition, this information is to be interpreted at an ecosystem level to provide insight into the mechanistic strategies employed by the native microbiome. The enzymes identified are to be assessed for potential biotechnological translation through recombinant expression and biochemical validation with a focus on halotolerance. This approach has the potential to probe an unknown process in an uncharacterised environment and address current constraints in lignocellulose bioprocessing and enable the use of seawater within biorefining processes, alleviating water security concerns.

An *in situ* degradation experiment was conducted to replicate lignocellulose deconstruction in a natural salt marsh during a time course. This platform is used in conjunction with high throughput DNA sequencing and proteomic techniques to identify the microorganisms and secreted enzymes throughout the composting process. The process was analysed at the ecosystem metasecretome level to construct a temporal profile of lignocellulolytic activity during *in situ* decomposition. The most likely taxonomic origin of identified enzyme facilitated the identification of key functional carbon cycling taxa. A diverse spectrum of lignocellulolytic enzymes was identified and a subset of enzymes that were integral at the ecosystem level were recombinantly expressed, characterised and assessed for halotolerance.

Chapter 2 Materials and methods

In this Chapter, general methodologies that were applied across all Chapters will be covered. Chapter specific materials and methods are presented in the materials and methods section of their respective Chapters (sections 3.2, 4.2, 5.2 and 6.2).

2.1 Chemical reagents

Chemicals and reagents utilised throughout the work presented in this thesis originated from Agilent technologies, Cambio, Cambridge Biosciences, GE Healthcare, Illumina, Merck, New England BioLabs, Promega, Qiagen and Sigma-Aldrich and Thermo Fisher Scientific.

2.1.1 Buffers

Water in all experiments unless explicitly stated, was obtained using an Elga PureLab Ultra water polisher. All water was purified to a resistivity value of 18 M Ω -cm.

Phosphate buffered saline (PBS) was prepared to a 1x final concentration using 8g NaCl, 0.2g KCl, 1.44g Na₂HPO₄ and 0.24g KH₂PO₄ in 1L H₂O. The final concentrations of reagents are; 10mM Na₂HPO₄, 280mM NaCl and 6mM KCl. The final solution was adjusted to pH 7.4 unless stated otherwise. Dilute concentrations were prepared by diluting with H₂O, concentrated variations were prepared using the ingredients above at specified ratios.

Artificial seawater (SeaChem) was prepared using 34g salt mixture per 1L H₂O. The solution was heated in a 100°C water bath for 30 min and allowed to cool. The pH was adjusted to 8 to allow dissolution of residual salts and re-adjusted to pH 8 if required unless stated otherwise.

2.1.2 Organisms

Escherichia coli strains Rosetta-gami™ 2 (DE3) competent cells (Novagen) and Stellar™ ultra-competent cells (ClonTech) were utilised for heterologous expression and vector cloning respectively.

Materials and methods

2.1.3 Phosphoric acid swollen cellulose (PASC)

Phosphoric acid swollen cellulose (PASC) was generated using Avicel® (Sigma 11365) and the PASC in this study was kindly generated and provided by Dr. Nicola Oates (Centre for Novel Agricultural Products, University of York). PASC was prepared by incubating 1g Avicel with 30mL 85% phosphoric acid on ice for 1 hour. The PASC was then precipitated and cleaned with 100mL acetone, funnel filtered (glass-filter) and washed an additional three times with acetone. A final wash was conducted with 500mL water and the PASC was size fractionated to homogeneity with a blender.

2.2 Molecular biology techniques

2.2.1 Plasmid extraction and purification (PCR)

Plasmids were purified from previously transformed cells using Qiaprep Miniprep Kits (Qiagen) according to the manufacturer's protocol. Cells were grown at 37°C at 175rpm in 5mL LB for 16 hours prior to extraction.

2.2.2 Polymerase chain reaction

Polymerase chain reactions (PCR) were performed with thermal gradient cycling equipment; Tetrad Z, Pelteir DNA Engine (BioRad), Pelteir PTC 200 (MJ Research) or Prime (Techne). Primers were purchased from Integrated DNA technologies or Eurofins. Annealing temperatures were calculated using an open source T_m calculator (New England BioLabs, <https://tmcalculator.neb.com/#!/main>) or the SnapGene software package (version 4.2). Reactions were predominantly performed with Phusion® High-Fidelity DNA polymerase (Thermo Scientific F530S) (for error sensitive cases), for general purposes DreamTaq DNA polymerase was used (Thermo Scientific K1081). General reaction components and thermocycling conditions are displayed in Table 2.1-2.

Materials and methods

Table 2.1 PCR reaction components.

| Polymerase | | | |
|---------------------------------|------------------|---------------------|--|
| Phusion | | | |
| Component | Volume(μ L) | Final concentration | |
| dH ₂ O to 25 μ L | 15.75 | - | |
| HF/GC Buffer (5x) | 5 | 1x | |
| dNTP (10mM) | 0.5 | 200 μ M | |
| pF (10 μ M) | 1.25 | 0.5 μ M | |
| pR (10 μ M) | 1.25 | 0.5 μ M | |
| DNA | 1 | 0.1-2ng/ μ L | |
| Polymerase (2000U/mL) | 0.25 | 1U/50 μ L | |
| Taq | | | |
| dH ₂ O to 25 μ L | 9 | - | |
| Master mix (2x) | 12.5 | 1x | |
| pF (10 μ M) | 1.25 | 0.5 μ M | |
| pR (10 μ M) | 1.25 | 0.5 μ M | |
| DNA | 1 | 0.1-2ng/ μ L | |

Table 2.2 General PCR thermocycling conditions.

| Polymerase | | | | |
|----------------------|-----------------------------|--|------------------|--|
| Phusion | | | | |
| Step | Temperature ($^{\circ}$ C) | Time | Cycles | |
| Initial denaturation | 98 | 30s | | |
| Denaturation | 98 | 10s | Repeat x26-34 | |
| Annealing | * | 30s | | |
| Extension | 72 | 15s per kb | | |
| Final extension | 72 | 10 min | | |
| Hold | 4 | - | | |
| Taq | | | | |
| Initial denaturation | 98 | 30s | | |
| Denaturation | 98 | 10s | Repeat x26-34 | |
| Annealing | * | 30s | | |
| Extension | 72 | 1 min up to 2kb extended by 1min per additional kb | | |
| Final extension | 72 | 10 min | | |
| Hold | 4 | - | | |

PCR enhancer was used to increase primer specificity during initial cloning amplifications. A 5x concentration PCR enhancer solution contained 2.7M betaine (Sigma B2629) and 6.7% DMSO, which was used at a final concentration in PCR reactions of 1x (0.54M betaine and 1.34% DMSO). This was

Materials and methods

used in conjunction with primer specific buffers for all polymerases. Betaine and DMSO are PCR additives that have isostabilising effects. They are thought to form hydrogen bonds within the major and minor grooves with DNA templates destabilizing the helical structure and preventing the formation of secondary structures. This is useful for guanine and cytosine rich DNA templates which often form strong secondary structures through self-complementarity within the sequence that are obstructive for polymerases.

2.2.3 Agarose gel electrophoresis

DNA profiles were assessed using agarose gel electrophoresis. Agarose gels were prepared by dissolving with 1% (w/v) molecular grade agarose (Sigma A9539) in Tris-acetate-EDTA buffer (TAE; 40mM Tris, 20mM acetic acid and 1mM EDTA) by microwave heating. Ethidium bromide (Sigma) was added to a final concentration of 0.2µg/mL upon cooling. The gels were moulded using agarose gel moulds (Bio-Rad) and electrophoresis was conducted in sub-cell GT tanks (Bio-Rad) with a Powerpac 300 power source (Bio-Rad) at 120V for 30-60 min (dependent upon gel size), or until the negatively charged DNA had migrated 70% of the distance toward the anode. Gels were visualised using a UV imaging station (UVItec) with the UVIproMW1 software package (version 11.03).

2.2.4 PCR product purification

DNA products from PCR reactions were isolated and purified using Wizard® SV Gel and PCR clean up kit (Promega, A9281). For products with downstream cloning related applications NucleoSpin® Gel and PCR clean up kit (ClonTech) were used according to the manufacturer's protocol with a single deviation; the loading NTI buffer was diluted to 13% to exclude small (primer dimer) products. Chapter specific methods are also described in their respective Chapters.

2.2.5 Nucleic acid quantification

Nucleic acid pools (DNA and/or RNA) were quantified using a NanoDrop™ 1000 (Thermo) spectrophotometer. Chapter specific methods are also described in their respective Chapters.

Materials and methods

2.2.6 DNA sequencing

Samples containing 5µL of 80-100ng purified DNA and 5µL of 5µM *a priori* selected sequencing primer (or synonymous concentration equivalents) were sent to GATC Biotech (Germany) using the LIGHTRUN sanger sequencing service. Sequencing results were quality checked using BioEdit (version 7.0.4), Sequence Scanner (Applied Biosystems, version 1.0) or using Python (version 2.7, 3.2, 3.4, 3.6).

2.2.7 pH measurements

Solution pH was measured and altered using a Microprocessor pH meter 210 (HANNA Instruments) with a Refillable pH probe (Fisherbrand FB68793).

2.2.8 Media preparation

Luria Broth (LB) was utilised unless explicitly stated for most microbial activities. LB Broth (Sigma L3522) was prepared according to the manufacturer's instructions (25g/L), the final media contained 10g/L tryptone, 10g/L NaCl and 5g/L yeast extract.

2.2.9 Plate preparation

LB Broth (see section 2.2.8) was utilised unless explicitly stated. LB Broth with agar (Sigma L2897) was prepared according to the manufacturer's protocol (35g/L), the final solution contained 15g/L agar, 10g/L tryptone, 5g/L yeast extract and 5g/L NaCl.

2.3 Proteomic methods

2.3.1 Bradford Assay for protein quantification

Protein concentrations were quantified using the Quick Start™ Bradford Protein Assay (Bio-Rad 5000201). 5µL of protein sample was incubated for 10 min with 245µL the Quick Start™ solution and the absorbance at 595nm was measured using a Tecan Sunrise™ microplate spectrophotometer. Protein concentration as determined using a standard curve of known protein concentrations constructed using bovine serum albumin (BSA, Sigma, A2153).

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

Proteins were analysed by observing their size separation using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Gels consisted of an upper stacking layer and lower separating layer and were created using glass casts provided in the Mini-Protean® Tetra cell system kit (Bio-Rad). The lower separating (10% acrylamide w/v) was prepared using 3.3mL acrylamide (30% w/v), 2.5mL 1.5M Tris-HCl (pH 8.8), 100µL of 10% sodium dodecyl sulfate (SDS) and 4mL H₂O. Immediately before loading into the 0.75mm gel casing for moulding, 10µL tetramethylethylenediamine (TEMED) and 100µL of 10% (w/v) ammonium persulfate (APS) was added to initiate polymerisation. After loading, the casings were filled with 50% (v/v) isopropanol to flatten the gel. The gel was allowed to set for 10-20 mins and the isopropanol was removed before the stacking (upper) layer (4% acrylamide w/v) was added, which was prepared with 1.3mL acrylamide (30% w/v), 2.5mL 0.5M Tris-HCl (pH 6.8), 50µL of 10% SDS and 6.2mL H₂O. Immediately before loading 5µL TEMED and 50µL of 10% (w/v) APS were added. The stacking gel was pipetted on top of the lower separating layer and a comb was placed in the cast. The gel was incubated for 10-20 mins to polymerise before removing the comb.

Polymerised gels were placed in Mini-Protean tanks (Bio-Rad) which were filled with SDS running buffer (0.1% SDS, 25mM Tris, 192 mM glycine). Protein samples were mixed 4:1 with protein loading buffer (0.4% (w/v) bromophenol blue, 8% (w/v) SDS, 40% (v/v) glycerol, 200mM Tris pH 6.8, 400mM β-mercaptoethanol) and incubated at 95°C for 5 mins to denature. Gels were loaded alongside commercial markers and electrophoresis was performed for 30-60 min at 220V using a Powerpac 300 (Bio-Rad). Proteins were stained by incubating the gel with InstantBlue (Sigma ISB1L).

2.4 Data analysis

Data processing and analysis was performed using python (versions 2.7, 3, 3.2, 3.4, 3.5, 3.6), with open source tools such as Pandas and NumPy (Oliphant, 2006, McKinney, 2010). Many of the analyses were custom built. Biopython was utilised to rapidly parse, manipulate and write sequence data (Cock et al., 2009). Graphs were generated using matplotlib (version 2.2-2.2.2) and seaborn (Hunter, 2007, Waskom et al., 2014). SciPy was used for statistical analyses (Jones et al., 2014).

Chapter 3 Time series *in situ* decomposition of lignocellulose in a salt marsh

3.1 Introduction

Lignocellulose is a terrestrially derived complex biopolymer composite produced by land plants. It is usually introduced into the marine environment via riverine transport (Syvitski et al., 2005). Up to 30% of organic carbon in oceanic marine sediments is of terrestrial origin (Burdige, 2005, Burdige, 2007). The majority of accumulated organic carbon can be found in deltaic or marginal sediments (Hedges and Keil, 1995). A disparity exists between the large influx of terrestrial lignocellulose into the marine environment, than can be observed in sediments (Bianchi, 2011). This suggests that the recalcitrant lignocellulose is degraded by marine microbiomes containing lignocellulolytic consortia.

Few marine environments contain native lignocellulose (with exceptions being salt marshes, seagrass meadows and mangroves). Salt marshes are highly diverse intertidal ecosystems which transition from terrestrial to marine environments characterised by halotolerant fauna and flora and high primary productivity. Salt marshes vary greatly in their form (structure), biotope (fauna and flora) and salinity (brackish to marine). Often salt marshes exist in sheltered intertidal areas which are energetically limited (low wind and tide energy). They are characterised by plant zonation whereby competitive displacement creates zones perpendicular to the elevation gradient. They are inundated almost twice daily (the UK exhibits a semi-diurnal cycle) with seawater and this disturbance mediates an array of physical and chemical stresses, which also varies temporally. Chemical stresses are affected by a multitude of variables (air exposure, precipitation levels, seawater nutrient loading, tidal cycle), which underpin huge surface level fluctuations in pH (7.2-8.3), salinity (2-3‰ and above), temperature (sub-zero to 25°C), pCO₂ and O₂ (1-14 mg L⁻¹) (Baumann et al., 2015).

In Britain, although salt marsh biotopes are highly variable, salt marshes are usually dominated by one of two salt marsh grasses *Spartina maritima* (native) or *Spartina anglica* (invasive hybrid of *S. maritima* and the introduced *S. alterniflora*). These grasses are perennial halophytes that secrete salts of a similar composition to seawater through specialised salt bladders and salt glands that transport salt ions directly out of the plant (McGovern et al., 1979). Plants that use this strategy are collectively known as recretohalophytes (Yuan et al., 2016). It is thought there is only 370 known recretohalophytes, including *Spartina* (Flowers et al., 2015, Yuan et al., 2016). The genus *Spartina* belongs to the family *Poaceae*, which exhibit the highest salt secretion ability of all *Plantae* (rivalled

only by *Tamaricaceae* which also contains coastal dwelling representatives) (Yuan et al., 2016, Zhang et al., 2018b). It is this unique ability that has allowed *Spartina spp.* to colonise and dominate salt marshes. When considering the adverse conditions salt marshes present, it is unexpected that salt marshes are considered some of the most productive ecosystems in the world comparable with intensive agriculture (Hopkinson and Giblin, 2008). This is due to favorable nutrient fluxes mediated by tidal transportation. This level of productivity acts synergistically with the seasonal dieback of the above ground biomass to introduce large amounts of terrestrial lignocellulose into the marine water column. Net aerial primary productivity (NAPP) for UK salt marshes have been reported to be 458-625 g m⁻² year⁻¹ (Boorman and Ashton, 1997, Bouchard et al., 1998), while warmer climates (Argentina, East-coast USA) have reported significantly higher values in excess of 1-2 kg m⁻² year⁻¹ (Trilla et al., 2010, Vera et al., 2009, Dai and Wiegert, 1996). Notably these figures are aerial and exclusive of below ground productivity, which are often multiples of aerial productivity for some species, and may significantly undervalue the ecosystem productivity potential (Chmura, 2013). These values are comparable with the net primary production (NPP) of agricultural crops and highlight the level of productivity in these ecosystems (Li et al., 2014).

Seasonal dieback occurs during the milder months of autumn (Ellison et al., 1986) or is orchestrated by epidemic fungal pathogens (notably *Phaeosphaeria spartinicola* and *Fusarium spp.*) which produces detrital organic matter as necromass, which enters the remineralisation phase of the carbon cycle (Elmer, 2016, Marsh et al., 2016). It is thought the majority of remineralisation occurs within the salt marsh, with only 0.05% NAPP exported and up to 89% of the native lignocellulosic necromass redistributed to the mid and upper zones; which act as carbon sinks, or deposited on the strandline (Bouchard and Lefeuvre, 2000, Jackson et al., 1986). However the exact fate of this necromass is still debated and conflicting evidence exists, it is highly likely that the majority is degraded on site and insignificant fractions are exported to the surrounding waters with the levels of each process variable both within and between marshes (Gordon et al., 1985, Hopkinson et al., 2012, Montemayor et al., 2015, Chmura, 2013, Hemminga et al., 1993).

The remaining biomass is eventually buried in the sediments where the marsh acts as a carbon sink with organic carbon burial rates in salt marshes estimated to be 57-218 g m⁻² year⁻¹, which is significantly higher than any terrestrial ecosystem (Chmura, 2013, Hopkinson et al., 2012). The burial process is strongly correlated and dependent upon key factors such as sediment characteristics (with finer silt and clay sediments often containing higher organic carbon), and inundation frequency and intensity (Zhang et al., 2018a, Kelleway et al., 2016). High organic carbon loadings within salt marsh sediments of 0.73 – 16.2 kg C m⁻² have been reported and interestingly these loadings appear to decrease with depth (Drake et al., 2015, Bai et al., 2016, Zhang et al., 2018a). Not only do the carbon

loadings decrease with depth, but the composition of the organic carbon also likely changes with depth, with refractory lignin derivatives becoming enriched and polysaccharides being lost (Fogel et al., 1989, Benner et al., 1987). This has also been observed in mangrove sediments (Marchand et al., 2005). It should be noted that not all the carbon in this environment is derived directly from lignocellulose, much is derived from algal sources or bacterial exopolysaccharides.

The reduction in carbon loading and profile discontinuity with depth suggests organic matter in the form of lignocellulose is at least partially degraded *in situ*, likely by microbial consortia. The sediment surface is the first interface between pristine lignocellulosic detritus and microbial degraders. Various studies utilising litter bags have ascertained degradation rates of native salt marsh plant biomass (namely *Spartina spp.*) *in situ* at surface levels (Bouchard and Lefeuvre, 2000, Negrin et al., 2012, Jackson et al., 1986, White et al., 1978, Valiela et al., 1984, Bouchard et al., 1998). Decomposition rates vary within and between sites and are influenced by temperature, frequency of inundation and submersion times (Bouchard et al., 2003, Wilson et al., 1986, Simoes et al., 2011, Menendez and Sanmarti, 2007).

Historically and currently salt marshes are in a state of decline. Impending encounters with rising sea level reveal a disconcertingly delicate position for salt marshes where vertical accretion rates, which have been estimated to be in the order of $1.6 - 2.4 \text{ mm } \text{y}^{-1}$, are below projected rates of sea level rise (Wright et al., 2017). Understanding carbon cycling in salt marshes with respect to organic matter, of which a major constituent is lignocellulose is of great ecological importance to ensure the security of these ecosystems (Chmura, 2013, Morris et al., 2016).

This lignocellulose-loaded environment is subjected huge fluctuations in chemical stresses, to which the native microbial consortia is adapted to or tolerant of. The most intriguing of these chemical fluctuations is salinity, which imposes ionic stress on microbes and enzymes rarely experienced in other systems. The tolerance to salt (halotolerance) is of major biotechnological interest, with halotolerant enzymes and organisms sought and currently used for a variety of applications (Yin et al., 2015a). Additionally, salt marshes are predominantly alkaline where lignocellulolytic enzymes of terrestrial origin favour acidic conditions. The biotechnological potential of these ecosystems has remained an enigma largely due to under exploration, however it is apparent that salt marshes are a model ecosystem for bioprospecting halotolerant, alkaliphilic, lignocellulolytic enzymes. Beyond lignocellulose degradation, there are likely vast biotechnological resources given the magnitude of complexity in this environment.

Despite the perceived importance in this field of study, little has been done understand the molecular mechanisms of carbon deconstruction in this environment. The microbial community that

drives this process and is also poorly understood. To interpose this knowledge in this understudied and underexplored ecosystem, an *in situ* decomposition experiment (ISDE) was designed to probe the carbon deconstruction process and provide both ecologically and biotechnologically relevant insight.

In this Chapter, carbon cycling knowledge was assimilated to construct the ISDE. The aim of which was to assess the process of carbon deconstruction of a native salt marsh grass *in situ* using a multi-faceted 'omics approach. *Spartina anglica* biomass was placed in Welwick salt marsh (North bank of the Humber estuary, UK) and allowed to degrade on site for 16 weeks. Nucleic acids, extracellular proteins and biomass samples were harvested periodically. The resulting data was then explored to identify the key organisms and enzymes that are involved in lignocellulose deconstruction in this salt marsh.

Salt marshes are repositories of buried carbon which exists predominantly as lignocellulosic material. Therefore, the hypothesis that the microbiome had adapted to utilize this substrate as a carbon source was developed. The aim of this study was to assess the native salt marsh lignocellulolytic profile and identify the key functional taxonomic groups orchestrating the process of lignocellulose deconstruction using an integrated 'omics approach during a time series of decomposition on *Spartina anglica* biomass in a natural salt marsh. A large proportion of lignocellulose acting enzymes are secreted. However, salt marshes are inherently dynamic, driven in particular, by tidal inundation and it was therefore hypothesized that many of the secreted enzymes are tightly bound to the lignocellulosic substrate and would be overlooked by conventional approaches. To address this, a biotin labelling approach was undertaken, using biotin derivative that is impermeable to cell membranes to label extracellular proteins at the biomass interface, that could not otherwise be captured (Alessi et al., 2017).

3.2 Materials and methods

3.2.1 Nomenclature rationalisation: (meta)secretome or (meta)exoproteome

Nomenclature regarding organisms' peripheral protein profile is broadly categorized as a (meta)secretome or (meta)exoproteome. The former includes only proteins that are actively secreted by the cell(s) into the extracellular matrix whereas exoproteome encompasses both proteins secreted into the extracellular matrix and outer membrane proteins. In the context of this study, the target data was actively secreted lignocellulose degrading proteins which are often bound

to the biomass, however; the biotin labelling technique employed to capture these proteins will have invariably captured outer membrane and transmembrane proteins. Therefore, the outer membrane proteins captured in these datasets are largely sampling artefacts and were not the subject of analyses. Secreted lignocellulolytic proteins were the focus of this study, therefore, despite the outer membrane artefacts, the extracellular proteomic datasets generated in this study are referred to a (meta)secretomes.

3.2.2 Field experiment

3.2.2.1 Field experiment synopsis

A field experiment was designed to study the process of native biomass decomposition in an established UK salt marsh (Figure 3.1). Above ground *Spartina anglica* biomass was collected during winter dieback in February-March 2015 on an adjacent intertidal mud flat (Cherry Cobb sands, Humber estuary, Hull, UK). The biomass was washed free of sediment, dried and size reduced. Nylon 66 monofilament woven bags (18.5cmx18.5cm), 'degradation bags' of aperture size 200µm were filled with 50g biomass and sealed with 100% polyester thread.

The degradation bags were placed in a 3 x 3 conformation in five stainless steel cages (711.2mm x 711.2mm x 63.5mm) with 25cm legs. These cages were interspersed by 75m along a 300m parallel elevation transect, defined by plant zonation patterns (dominance of *Spartina anglica*, *Puccinellia maritima* and *Salicornia europaea*) in Welwick salt marsh (Humber estuary, Hull, UK). All bags were secured with cable ties and the under canopy was removed to expose the sediment prior to deployment to facilitate biomass-sediment crosstalk. The cages were mapped to position with global position system (GPS) coordinate android application (Woozilli Inc.) to within 4m accuracy.

The biomass decayed for 16 weeks from 16/7/15 to 6/11/15 and samples were extracted weekly for the first six weeks, thereafter at week eight, ten and 16. Bag selection was randomised *a priori*, and sampling began at the point of low tide and was completed within two hours. During sampling, around 2.5g of (wet weight) biomass from each selected degradation bag was placed in a pre-weighed tube containing 12.5mL RNeasy lysis buffer, the remaining biomass and bag were placed in a sealed container and kept at 4°C or below during transport and sample processing began within four hours of harvest.

Processing included the removal of 10g (wet weight) biomass for nucleic acid extraction. Four aliquots (technical replicates) of 7.5g (wet weight) biomass was removed for protein extraction

Time series in situ decomposition of lignocellulose in a salt marsh

(inclusive of one negative control). The remaining biomass was weighed, washed free of sediment (under constant flow and agitation), frozen in liquid N₂, freeze-dried and weighed. The wet weight and dry weight allowed for a partial conversion factor to be applied to the extraction fractions and elucidation of total harvested mass.

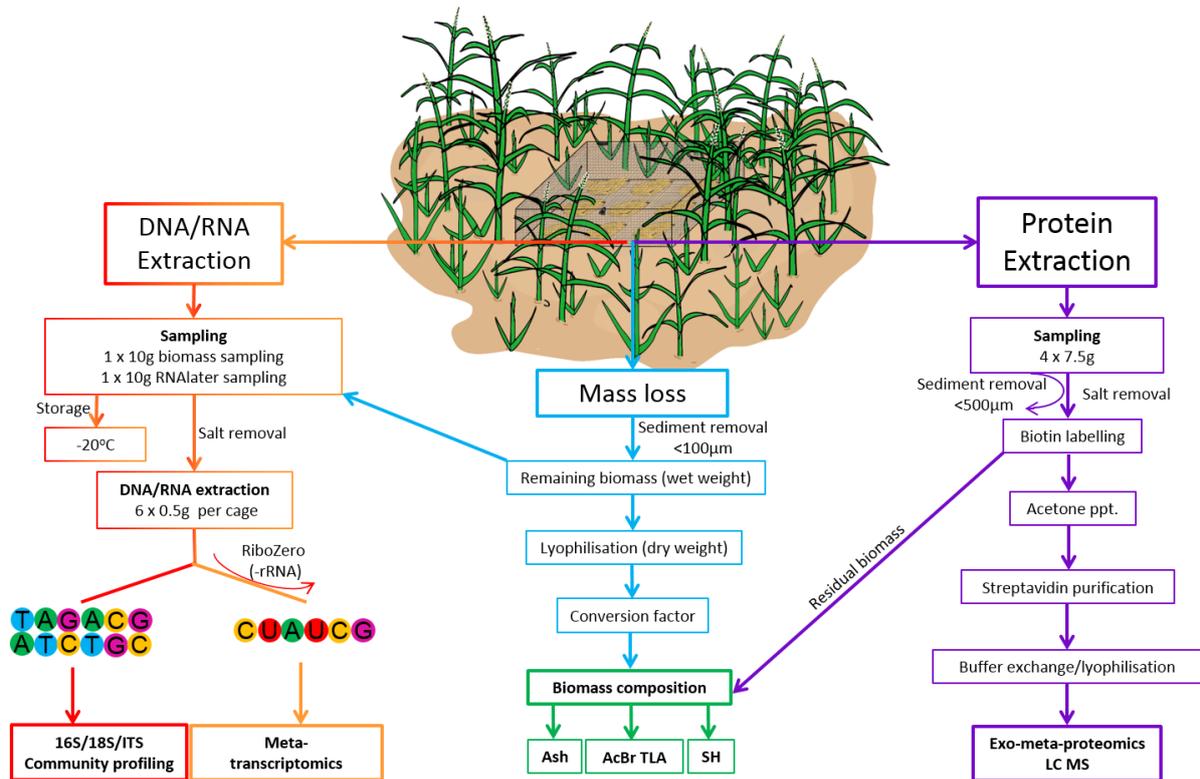


Figure 3.1 Schematic overview of the *in situ* degradation experiment. 16S/18S/ITS; markers for taxonomic identification, AcBr TLA; Acetyl bromide total lignin assay, LC-MS; tandem liquid chromatography mass spectrometry, ppt; precipitation, SH; sequential hydrolysis.

3.2.2.2 Biomass collection and preparation

Cherry Cobb sands (53.679599, -0.181000; Humber estuary, Hull, UK) is an intertidal mudflat spanning around 9km in length dominated by *Spartina anglica* (Figure 3.2). *Spartina anglica* was identified by presence of dense 2-3mm long ciliate legules between the stem and leaf blade. Above ground *S. anglica* biomass from 1-5cm above the sediment was collected several times between the winter dieback in February and March 2015. The harvested plants were cut into manageable pieces (20-50cm in length) and washed free of sediment with freshwater under constant flow. The fragments were dried in a fan assisted oven at 65°C for 48 hours to replicate strandline dehydration. The resulting dried biomass was further fractionated size with a Retsch Cutting Mill SM 300 installed

Time series in situ decomposition of lignocellulose in a salt marsh

with a 7mm bottom sieve at 2300rpm; the biomass was further fractioned with 1.12mm, 500 μ m, 250 μ m aperture sieves). Two of these fractions were pooled and the final biomass fraction consisted of 80% of the >1.12mm (large) fraction and 20% of the <1.12mm ->500 μ m (small) fraction, to replicate natural detrital fragmentation.

Nylon 66 monofilament woven bags (18.5cmx18.5cm), 'degradation bags' of aperture size 200 μ m were used to exclude larger invertebrates and focus primarily on microbial degradation whilst maintaining gaseous exchange. These bags were weighed (W_{bag}) and then filled with 50g biomass and weighed again (W_{filled}), they were then sealed with 100% polyester thread and weighed again (W_{thread}). This allowed for the exact mass of lignocellulosic material to be calculated for each bag.

3.2.2.3 Deployment

The bags were placed in a 3 x 3 conformation in five stainless steel cages (711.2mm x 711.2mm x 63.5mm) with 25cm legs, were interspersed by 75m along a 300m parallel elevation transect, defined by plant zonation patterns (dominance of *Spartina anglica*, *Puccinellia maritima* and *Salicornia europaea*) in Welwick salt marsh (53.650715, 0.023144; Humber estuary, Hull, UK) (Figure 3.2). There was roughly 13km between where the biomass was collected in Cherry Cobb sands and where the transect was placed. All bags were secured with cable ties to the underside of the cage mesh.

Time series in situ decomposition of lignocellulose in a salt marsh

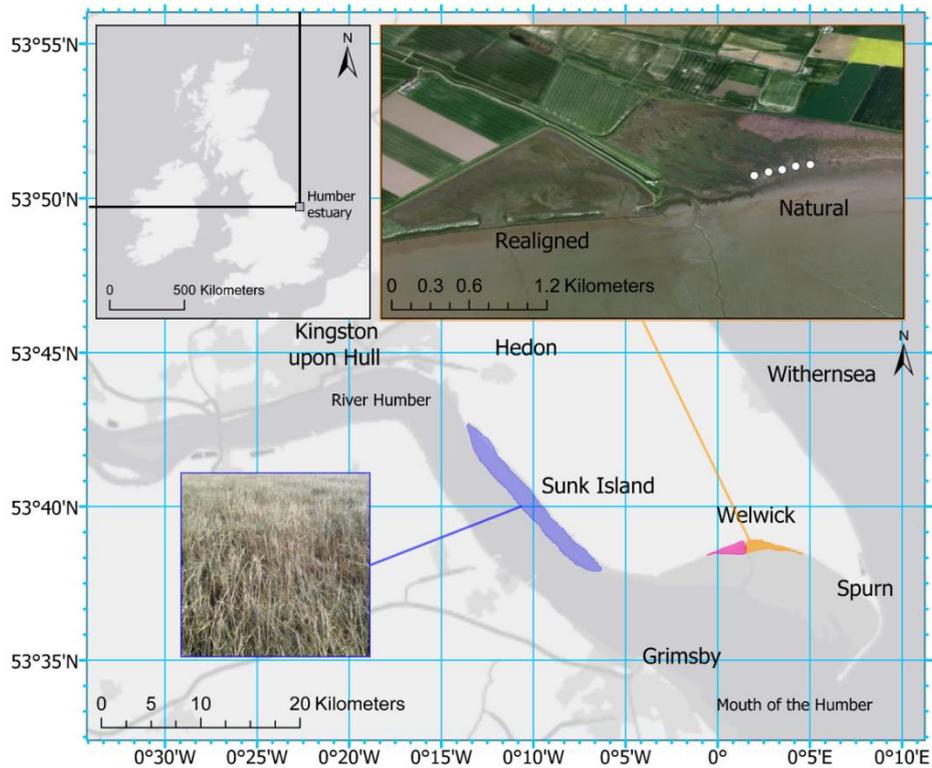


Figure 3.2 Experiment location. A) The Humber estuary (Hull), in relation to the UK. B) The North bank of the Humber estuary, Welwick salt marsh is highlighted in orange and the cage locations along the transect in white (C), Cherry Cobb Sands is highlighted in blue with the location of the collected *Spartina anglica* (D). Imagery sourced from Esri (2018, colour) and Ordnance Survey (2016, grey) and created using ArcGIS Pro (version 2.2).

Deployment was conducted during the point of low tide (12:04pm, 1.2m above chart datum (ACD)) on the 16/07/2015. The faunal under canopy was removed to expose the sediment prior to deployment to facilitate biomass-sediment crosstalk. The cages were mapped to position with GPS coordinate android application (Woozilli Inc.) to within 4m accuracy (Table 3.1). The cages were also fitted with a water trap to ascertain water salinity and pH throughout the time course. All cages were orientated with the water traps facing shoreward. The cage legs were hammered into the sediment until the flexible mesh was in contact with the sediment to ensure stability.

Time series in situ decomposition of lignocellulose in a salt marsh

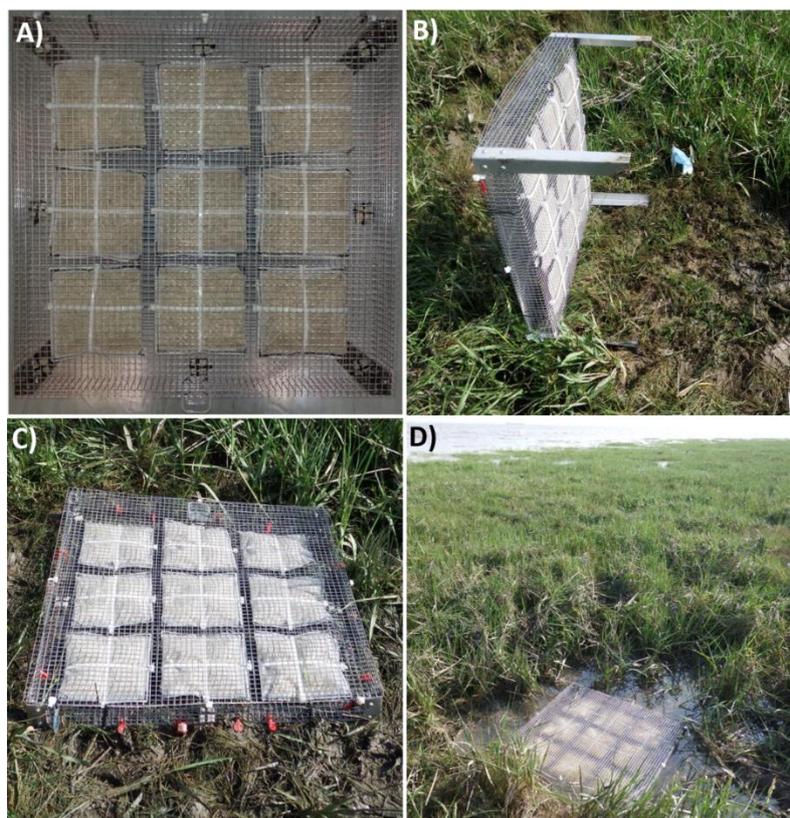


Figure 3.3 Deployment of the *in situ* degradation experiment. A) Cage layout. B) Under canopy removal. C) Cage deployment and fixation to the sediment. D) First tidal inundation on the 16/07/2015.

Table 3.1 Transect position of the five biological replicates in Welwick salt marsh.

| Cage | OS GR | Coordinates | Latitude/Longitude | Elevation (m) |
|------|----------------|------------------------|-----------------------------|---------------|
| A | TA 33850 18626 | 53.647022, 0.023465794 | 53°38'49.28"N/ 0° 1'24.48"E | 2 |
| B | TA 33915 18656 | 53.647271,0.024455525 | 53°38'50.18"N/ 0° 1'28.04"E | 2.2 |
| C | TA 33977 18676 | 53.647439, 0.025407313 | 53°38'50.78"N/ 0° 1'31.47"E | 2.7 |
| D | TA 34038 18704 | 53.647678,0.026343800 | 53°38'51.64"N/ 0° 1'34.84" | 3 |
| E | TA 34105 18719 | 53.647792, 0.027360971 | 53°38'52.05"N/ 0° 1'38.50"E | 3 |

† Coordinate accuracy $\pm 4\text{m}$

‡ Coordinates mapped with GPS Coordinates android application (Woozilli Inc.)

3.2.2.4 Sampling harvesting

Random sampling was conducted using *a priori* selected bags using a custom Python script for randomisation (Python 2.7). The tidal regime was estimated from the Sunk Dredged Channel station and missing data was kindly provided via personal communication (UK Hydrographic Office, <http://www.ukho.gov.uk/Easytide/easytide/SelectPort.aspx>). Sampling began at the point of low tide for each timepoint and was finished within two hours (Table 3.2). A single bag from each of the five cages was harvested weekly during the first six weeks, then at week eight, ten and 16. The bags

Time series in situ decomposition of lignocellulose in a salt marsh

were opened in the field and approximately 2.5g of wet biomass was placed into a 15mL pre-weighed Sarstedt tube containing 12.5mL RNAlater® (Thermo). The remaining biomass (including the bag) and tube containing the RNAlater and biomass were then placed into a pre-cooled (4°C or below) 1L container, sealed (air-tight) and placed into a thermally insulated field bag (chilled with ice-bricks that had been pre-cooled to -80°C).

The surrounding top 5cm of sediment was randomly sampled from each cage and placed into a 50mL Sarstedt tube until full. The water traps were also emptied into 15mL Sarstedt tubes. During deployment (day 0) only coastal water was sampled. The sediment and water samples were kept in the thermally insulated bags. The harvested samples were transported from Welwick salt marsh in Hull to the University of York within two hours where sample processing began immediately, therefore, the longest possible time between harvest to processing was four hours.

Time series in situ decomposition of lignocellulose in a salt marsh

Table 3.2 Sampling and tidal regime. The bold times indicate the low tide during which sampling began. Tidal regime was measured from the Sunk Dredged channel station.

| Timepoint | Date | Tide | Time (24hr) | ACD(m) |
|------------------|------------------------|-------------|--------------------|---------------|
| 0 | 16/07/2015 Thursday | HW | 05:39 | 6.9 |
| | | LW | 12:04 | 1.2 |
| | | HW | 18:11 | 6.8 |
| 1 | 23/07/2015 Thursday | LW | 03:42 | 2.3 |
| | | HW | 09:42 | 6.3 |
| | | LW | 16:19 | 2.1 |
| | | HW | 22:22 | 5.9 |
| 2 | 30/07/2015 Thursday | HW | 04:24 | 6.8 |
| | | LW | 10:52 | 1.6 |
| | | HW | 16:55 | 6.7 |
| | | LW | 23:12 | 1.5 |
| 3 | 06/08/2015 Thursday | LW | 03:42 | 1.5 |
| | | HW | 09:37 | 7 |
| | | LW | 16:17 | 1.3 |
| 4 | 13/08/2015 Thursday | HW | 04:45 | 6.7 |
| | | LW | 11:03 | 1.5 |
| | | HW | 17:21 | 6.6 |
| | | LW | 23:15 | 1.8 |
| 5 | 20/08/2015 Thursday | LW | 02:35 | 1.9 |
| | | HW | 08:30 | 6.7 |
| | | LW | 15:04 | 1.8 |
| | | HW | 20:57 | 6.4 |
| 6 | 27/08/2015 Thursday | HW | 03:13 | 6.4 |
| | | LW | 09:33 | 2 |
| | | HW | 15:51 | 6.5 |
| | | LW | 22:01 | 1.9 |
| 8 | 10/09/2015 Thursday | HW | 03:40 | 6.3 |
| | | LW | 09:55 | 1.9 |
| | | HW | 16:25 | 6.3 |
| | | LW | 22:11 | 2.1 |
| 10 | 24/09/2015 Thursday | HW | 01:43 | 5.9 |
| | | LW | 07:49 | 2.4 |
| | | HW | 14:34 | 6.1 |
| | | LW | 20:32 | 2.5 |
| 16 | 05/11/2015 Thursday | HW | 00:01 | 5.8 |
| | | LW | 06:40 | 2.5 |
| | | HW | 12:57 | 5.7 |
| | | LW | 19:01 | 3.1 |

Time series in situ decomposition of lignocellulose in a salt marsh

3.2.2.5 *Sample processing*

Sample processing began immediately upon return to the laboratory. Initially all containers were weighed and recorded (large containers and RNAlater tubes). The biomass was homogenised and for each of the five containers split into a single 10g wet weight aliquot for nucleic acid extraction (section 3.2.1.5.1) and four 7.5g wet weight aliquots for protein extraction (section 3.2.1.5.2). The remaining biomass and bags were used to assess the mass balance (section 3.2.1.5.3). The nucleic acid extraction was initiated first, and one hour into the protocol the protein extracted was initiated, both extractions were conducted in parallel.

3.2.2.5.1 Simultaneous DNA and RNA extraction

All H₂O in this extraction was diethyl pyrocarbonate (DEPC) treated for 2 hours at 37°C and autoclaved at 121°C for 15 min. 10g of biomass was aliquoted into 50mL Sarstedt tubes and ice-cold (4°C) 1x PBS pH 8.15 was added (to 50mL), and the tubes were centrifuged for 20 min at 4600rpm in Haraeus Multifuge3 S-R with a swing-out rotor (Sorvall). The supernatant was discarded, and another wash step was conducted. 0.5g of biomass was added to screw cap tubes (2mL) containing 0.5g 0.5mM glass beads (Sigma G9268) and 0.5g 0.1mM glass beads (Sigma G8893). 0.5mL cetyl trimethylammonium bromide (CTAB) buffer containing 10% CTAB (m/v) in 0.7M NaCl, 240mM potassium phosphate pH8 and 0.1% β-mercaptoethanol was added to the tubes. 0.4mL phenol/chloroform/isoamyl alcohol (25:24:1) pH 8. The samples were homogenised in a TissueLyser II (Qiagen) for 2 x 2.5 min at 30/s. The tubes were then centrifuged in a pre-cooled (4°C) accuSpin Micro17R (Fisher Scientific) at 13,300 rpm for 15 min. The aqueous phase was transferred to a new tube and an equal volume of chloroform:isoamyl alcohol (24:1) was added and centrifuged as previously. The aqueous phase was transferred to a new tube and precipitated for 16 hours at 4°C with two volumes of PEG precipitation solution containing 20% (w/v) PEG8000 (Sigma) in 1.6M NaCl. The nucleic acid pellet was collected by centrifugation as above for 30 min at 4°C. The pellet was washed twice in 75% ethanol. The quality of the DNA and RNA mixture was assessed by agarose gel electrophoresis, quantified using a NanoDrop™ (Thermo) spectrophotometer and stored at -80°C for downstream processing in Chapters 4, 5 and 6.

Time series in situ decomposition of lignocellulose in a salt marsh

3.2.2.5.2 Protein processing

Two washes were performed on 7.5g biomass with 40mL ice cold 0.5x PBS pH 8.15, centrifuged at 4500rpm for 20 min and the supernatant was discarded using a Haraeus Multifuge3 S-R with a swing-out rotor (Sorvall). The extracellular and transmembrane proteins were labelled in triplicate and 2.5g biomass aliquots for each of the biological replicates were resuspended in 10mM EZ-link-Sulfo-NHS-SS-biotin (Thermo Scientific #21331) in 0.5x PBS and incubated at 4°C for 1 hour. The biomass was centrifuged for 10 min at 4500rpm as above, the supernatant was discarded and the biontylation reaction was quenched with 25mL 50mM Tris-HCl pH8 and incubated for 30 min at 4°C. Excess Tris-HCl and residual biotin was removed with two washes with 20mL ice cold 0.5X PBS pH8 with centrifugation steps for 5 min at 4500rpm.

Proteins were extracted from the biomass with 10mL 2% (w/v) SDS pre-heated to 60°C and incubated for 1 hour. The supernatant was extracted, and the proteins were precipitated with five volumes of pre-chilled (-20°C) 100% acetone and incubated at -20°C for 16 hours. The residual biomass at this point was utilised in section 3.2.1.5.3. Precipitated proteins were pelleted by centrifuging at 4500rpm for 20 min and the residual acetone was discarded. The pellets air dried and resuspended in 1mL 0.1% SDS in 1x PBS, filtered through 0.22µm syringe driven filter units and loaded onto 1mL HiTrap Streptavidin HP columns (GE Healthcare #17-5112-01) and incubated for 1 hour at 4°C. The proteins were eluted with 1mL 50mM dithiothreitol (DTT) in 1x PBS, the column was incubated for a further 1 hour and eluted again, this was performed three times and the first two 1mL fractions were pooled.

The protein fractions were desalted, and buffer exchanged into H₂O using 5mL Zeba™ Spin 7k MWCO columns (Thermo 89882) according to the manufacturer's protocol. The buffer exchanged protein was frozen in liquid nitrogen, lyophilised using a Heto PowerDry LL3000 Freeze Dryer (Thermo) and resuspended in 210µL H₂O. The proteins were stored for LC-MS/MS analysis by solubilising in NuPAGE LDS sample buffer (Life Technologies) and incubating at 70°C for 10 mins prior to a short (6 min) run into a 7 cm NuPAGE Novex 10% Bis-Tris Gel (Life Technologies) at 200 V. The gels were stained with SafeBLUE protein stain (NBS biologicals) for 1 hour before de-staining with H₂O for 1 hour. The stained gels were sliced into 1mm² fragments and stored at -20°C prior to LC-MS/MS analysis (Chapter 4).

Time series in situ decomposition of lignocellulose in a salt marsh

3.2.2.5.3 Biomass preparation

The residual biomass from 3.2.1.5.2 was utilised in this section. The biomass was washed free of SDS with four washes of 40mL ethanol:H₂O (1:1) followed by two washes of 40mL H₂O. The biomass was flash frozen in liquid N₂, lyophilised and stored at room temperature.

3.2.2.5.4 Mass balance

Mass loss was obtained using a partial conversion factor (equation 3.1). For each biological replicate, after the nucleic acid extraction fraction (10g; B_{NA}) and protein extraction fraction (4x7.5g; B_P) had been removed, the remaining wet biomass and bag were weighed (B_w). The biomass was then washed free of sediment through a 100µM mesh under constant agitation and lyophilised in a Heto PowerDry LL3000 Freeze Dryer (Thermo). The dry biomass was weighed (B_D) and a partial conversion factor was calculated and applied to the sum of all other fractions (including the RNAlater field sample fraction; B_{RL}). This was applied to every biological replicate at every time point.

Equation 3.1 Mass balance partial conversion factor. Where B_D is the dry weight biomass remaining, B_w is the wet weight biomass remaining, and E is the sum of all fractions utilised for extraction as wet weight (B_{NA}, B_P, B_w and B_{RL}; see text).

$$\text{Biomass remaining} = \left(\frac{B_D}{B_w} \right) \sum E$$

The partial conversion factor was also applied to each of the nucleic acid and protein fractions to determine the dry weight biomass used in the extraction. This was necessary as despite normalising sampling to the time at which low tide occurs, weekly changes in tidal height effects the exposure time between inundation and sampling, resulting in wet weight variation between weeks. This was particularly notable for week one where the hygroscopic biomass was not fully saturated with water.

3.2.2.5.5 Salinity and pH measurements

The tubes containing the collected sediment and water were centrifuged at 4°C at 4600 rpm for 30 min in a Haraeus Multifuge3 S-R with a swing-out rotor (Sorvall). The supernatant was collected, warmed to ambient temperature and the pH was measured. Salinity was assessed with a temperature compensated H₂Ocean Refractometer (D-D The Aquarium Solution Ltd.).

3.2.3 Enzyme assays

Enzyme assays were conducted to ascertain (in consideration with mass loss) which time points to take forward for sequencing. The proteins that were pooled in equal mass ratios (those placed into gel storage for LC-MS/MS analysis) were utilised. All assays were conducted in triplicate. Cellulase, hemicellulose and peroxidase activity were assessed. For cellulase activity, reducing sugar assays (4-hydroxybenzoic acid hydrazide; PAHBAH) for carboxymethyl cellulose (CMC; 0.1%) and phosphoric acid swollen cellulose (PASC; 0.1%) were conducted. Peroxidase activity was assessed with 2,2'-azino-bis (ABTS, 1mM). Hemicellulase activity was assessed with beechwood xylan (1%; Megazyme) and chromogenic substrates (at 0.1% (w/v) concentration); AZCL xylan (Megazyme), AZCL galactomannan (Megazyme) and AZCL arabinoxylan (Megazyme). Activity was also assessed with non-pretreated *Spartina anglica* biomass (1%).

The assays were conducted in four buffers designed to mimic various levels of environmentally relevant salt concentrations (20mM Tris-HCl pH8, 1.75% NaCl in 20mM Tris-HCl pH8, 3.5% NaCl in 20mM Tris-HCl pH 8 and 1x artificial seawater (SeaChem) pH 8), in a final volume of 200µL and incubated for 16 hours. In depth analysis of xylanase activity was conducted in triplicate, with 1% beechwood xylan, 3µg crude metasecretome protein in a final reaction volume of 200µL for 9 hours and measured using the PAHBAH method. The buffers described above were applied.

3.2.4 Biomass composition analysis

Biomass composition analysis was conducted to determine and quantify the composition of the important lignocellulose fractions; crystalline cellulose, hemicellulose (matrix polysaccharides) and lignin. The matrix polysaccharides were analysed first followed by the crystalline cellulose (from the same biomass fraction). Lignin was measured using an independent biomass aliquot by assessing the acetyl bromide soluble fraction. Ash (non-carbon) content was also determined. For each of the five biological replicates for each of the nine time points including starting day 0 biomass (for a total of ten), analysis was conducted in quintuplicate for a total of 225 reactions per analysis fraction.

Time series in situ decomposition of lignocellulose in a salt marsh

3.2.4.1 *Analysis of matrix polysaccharides*

Biomass samples were aliquoted into five 5mg technical replicates for each biological replicate for each time point. The biomass was hydrolyzed with 500 μ L 2M trifluoroacetic acid (TFA) under an argon atmosphere during a four hour incubation at 100°C. The samples were mixed every hour during the incubation. Upon completion the samples were cooled and dehydrated in a Savant SpeedVac concentrator SPD131DDA (Thermo) used in conjunction with a Savant refrigerated vapour trap RVT4104 (Thermo). The samples were washed (to remove residual TFA) twice with 500 μ L propan 2-ol, mixed and dried again as above. The sugars were solubilised in 200 μ L H₂O and the supernatant was extracted, filtered through 0.45 μ m PTFE filters and placed into HPLC vials for analysis using an ICS-3000 Ion Chromatography System with AS loading units (Dionex, Thermo). The residual pellet was used to assess the crystalline cellulose content (section 3.2.2.2).

Samples were quantified against a standard curve containing various concentrations (100 μ M, 250 μ M and 350 μ M) of each of the expected monosaccharides; arabinose, fucose, galactose, galacturonic acid, glucose, glucuronic acid, mannose, rhamnose, and xylose. As 5 μ L of hydrolysate (and standard) was injected onto the column this facilitated the absolute quantification of each monosaccharide present in the matrix polysaccharide fraction. Peak area integration and quantification was performed using the Chromeleon software package (version 6.80 SR16 Build 5387, Thermo).

3.2.4.2 *Analysis of crystalline cellulose*

The pellet generated in section 3.2.2.1 was cleansed of impurities using 1mL Updegraff reagent (acetic acid: nitric acid: water, 8:2:1 v/v) at 100°C for 30 min to generate an insoluble cellulose pellet (Updegraff, 1969). The samples were centrifuged at 10,000 rpm for 15 min using a Centrifuge 5424 (Eppendorf), the supernatant was discarded. A single wash was performed with 1.5mL of H₂O and three times with 1.5mL acetone and centrifuged as above. The cellulose pellet was air dried at room temperature after the final wash. Saeman hydrolysis was performed with 90 μ L of 72% H₂SO₄ and incubated at room temperature for four hours with agitation (Saeman, 1945). The samples were diluted to 3.2% H₂SO₄ with 1890 μ L H₂O and incubated at 120°C for four hours. The samples were centrifuged at 10,000 rpm for 5 min to remove residual impurities.

The glucose content of the sample was assayed using the anthrone spectrophotometric assay and quantified against a standard curve of serially diluted glucose. Anthrone reagent (10mM anthrone in 72% H₂SO₄) was added to diluted hydrolysate (anthrone reagent, H₂O, sample, 20:9:1, v/v) and

Time series in situ decomposition of lignocellulose in a salt marsh

incubated at 80°C for 30 min. Upon cooling, a 200µL aliquot was transferred to an optical microplate and measured at 620nm using a Tecan Sunrise™ microplate spectrophotometer.

3.2.4.3 Analysis of acetyl bromide soluble lignin

Biomass samples were aliquoted into five 4mg technical replicates for each biological replicate for each time point and the lignin content was measuring using the acetyl bromide assay (Moreira-Vilar et al., 2014). Lignin was solubilised by the addition of 250µL acetyl bromide solution (25% acetyl bromide, 75% glacial acetic acid, v/v) to the biomass and incubated at 50°C for three hours and periodically vortexed. Upon cooling the liquid was transferred to a 5mL volumetric flask and 1mL 2M NaOH and 175µL hydroxylamine HCl was added. The mixture was vortexed and made up to 5mL with glacial acetic acid and the sample was taken for measuring immediately. Samples diluted 1:10 with glacial acetic acid and measured at 280nm in quartz cuvette using a Cary 50 UV-visible spectrophotometer (Agilent technologies). The percentage acetyl bromide soluble lignin was determined using equation 3.2, using a previously cited extinction coefficient of 17.75 for grasses (Foster et al., 2010).

Equation 3.2 Determination of the proportion of acetyl bromide soluble lignin (ABSL). Absorbance is measured at 280nm. The coefficient for grasses was used at 17.75, the pathlength used was 1, the total volume was 5mL, the aliquot was 0.1mL and the dilution factor was 10 (1:10). Biomass weight was individually variable (0.005g).

$$ABSL(\%) = \frac{Absorbance}{Coefficient \times Pathlength} \times \frac{Total\ volume\ (mL) \times Aliquot(mL) \times Dilution\ factor}{Biomass\ weight\ (g)}$$

3.2.4.4 Analysis of total ash

Around 1g of biomass was placed in a pre-weighed crucible (W_C) and weighed (W_{C+B}). The biomass was incubated in a Carbolite AAF 1100 Type 301 furnace (Barloworld Scientific) for 24 hours at 600°C. The cooled samples were weighed (W_A) and the non-carbon ash fraction was determined.

Time series in situ decomposition of lignocellulose in a salt marsh

3.2.4.5 Absolute quantification of lignocellulose fractions

The fractional composition analysis conducted (sections 3.2.2.1-3.2.2.4) was used with the mass remaining data (section 3.2.1.5.4) to extrapolate absolute quantification of biomass composition during the *in situ* degradation experiment.

3.3 Results

3.3.1 Field observations

Welwick salt marsh, as expected from the coastal proximity and distance from the estuary mouth, was predominantly marine (Figure 3.4). Salinity in the sediments and water correlated closely at 26-40ppt. The pH in the sediments were borderline neutral at pH 7.1 where the water pH was consistently alkaline at pH 7.9-8.7. Oceanic seawater is often cited at 35ppt and pH 8.15. Samples were not collected during weeks 10 or 16.

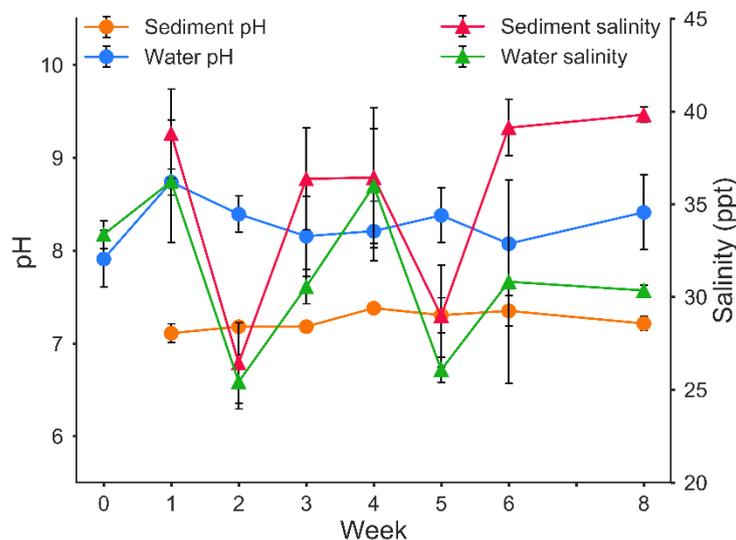


Figure 3.4 Measured pH and salinity throughout the time course experiment.

The experiment was deployed during the middle of a neap tide, and the experiment ran for four spring and four neap tides (Figure 3.5). There appeared to be no correlation between pH and salinity with respect to tidal regime. Cage elevation varied between 2m and 3m and were therefore inundated during every tide throughout the experiment.

Time series in situ decomposition of lignocellulose in a salt marsh

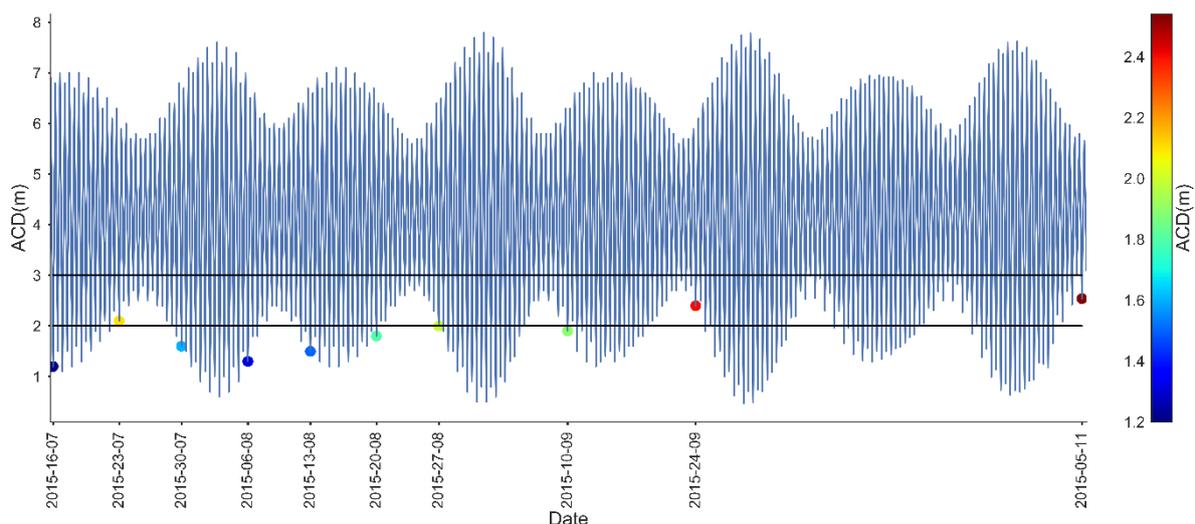


Figure 3.5 Tidal and sampling regime. Data is displayed as meters above chart datum (ACD). The black lines indicate the approximate elevation range of the cages. Plots indicate the dates during which sampling occurred and the tidal level at that point.

3.3.2 Nucleic acid extractions

Genomic DNA and total RNA was successfully extracted from all nine time points excluding the day 0 outgroup where only genomic DNA was recovered. The rRNA profile was suggestive of prokaryotic dominance, with 18S and 28S rRNA (which are comparably larger than their 16S and 23S prokaryotic equivalents) bands only faintly observable in a few time points and only in select replicates. The quality of the extractions was assessed with agarose gel electrophoresis and the strength of rRNA bands were used to determine qualitative quality (Figure 3.6). Total nucleic acids were stored at -80°C until required for downstream processing.

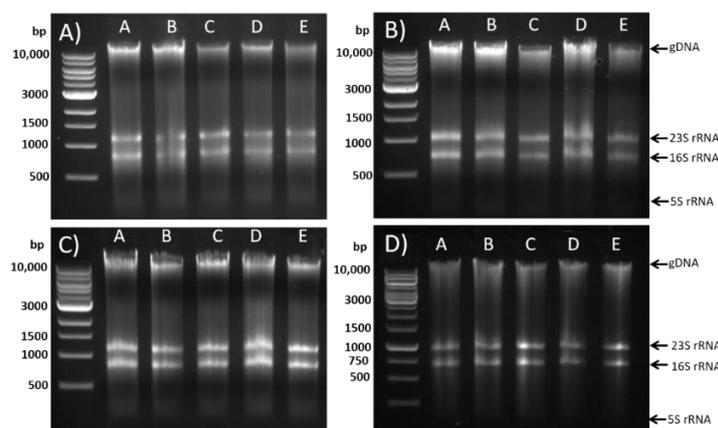


Figure 3.6 Genomic DNA and total RNA extractions. A) Week one. B) Week three. C) Week five. D) Week ten.

Time series in situ decomposition of lignocellulose in a salt marsh

3.3.3 Mass balance

Relative to the starting biomass, during the 16-week time course a total of $69.55 \pm 1.18\%$ mass had been lost (Figure 3.7A). The greatest loss in mass occurred between day 0 and week one (28.26%), however the following week exhibited the least temporal loss in mass (2.53%). The second largest loss in mass was observed between week two and three (10.7%) before the rate of mass loss plateaued to 3-5% to week 10, before declining to an average weekly loss of 1.28% between weeks 10 and 16. Relative mass loss fit a first order decay model ($R^2=0.99$) Mass loss as a function of mass remaining exhibited a non linear profile with a gradual decline in rate over time, from $4.05 \pm 0.22\%$ to $0.17 \pm 0.03\%$ mass loss per day in weeks one and 16 respectively (Figure 3.7B).

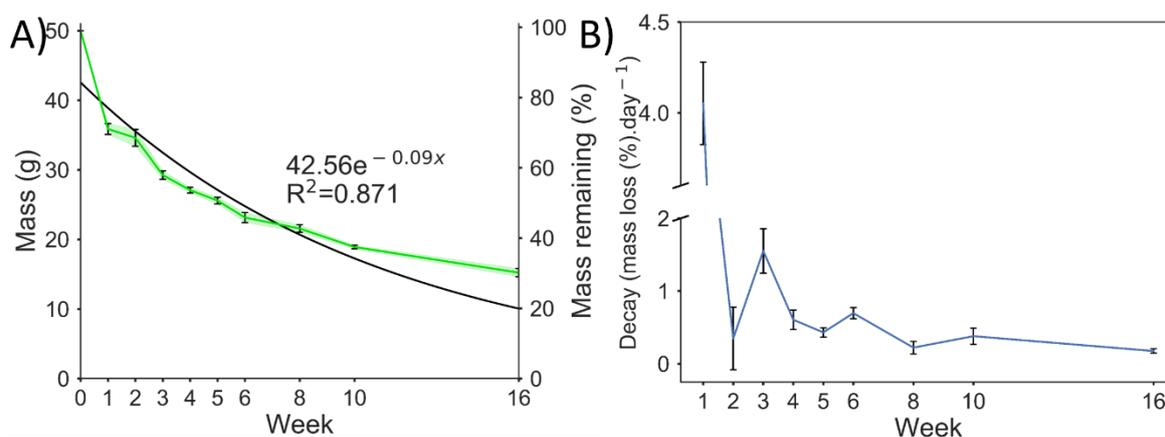


Figure 3.7 Temporal mass loss and decomposition rates. A) Relative and absolute mass loss. B) Decomposition rate as a function of remaining biomass.

3.3.4 Protein extraction

Extracellular protein extractions were successfully performed on biomass from all nine time points (excluding day 0). Extraction success was assessed using SDS-PAGE, upon staining a visible smear was indicative of a sufficient quantity of metasecretome protein for downstream processing (Figure 3.8A). The metasecretome extraction was quantified using the Bradford method (Table 3.3). Biological replicates were pooled in equal mass ratios for each time point, and aliquots of 25 μ g, 50 μ g and 100 μ g were short-run with SDS-PAGE. The resulting gels were stained (Figure 3.8B) and excised into 1mm² fragments for storage at -20°C. Individual biological replicates were run into gel and stored but not used in this study. Extracellular biotinylated protein represented $9.06 \pm 0.65\%$ of the

Time series in situ decomposition of lignocellulose in a salt marsh

total protein from week five to ten (this was not assessed in earlier time points) and was unexpectedly consistent between time points and biological replicates.

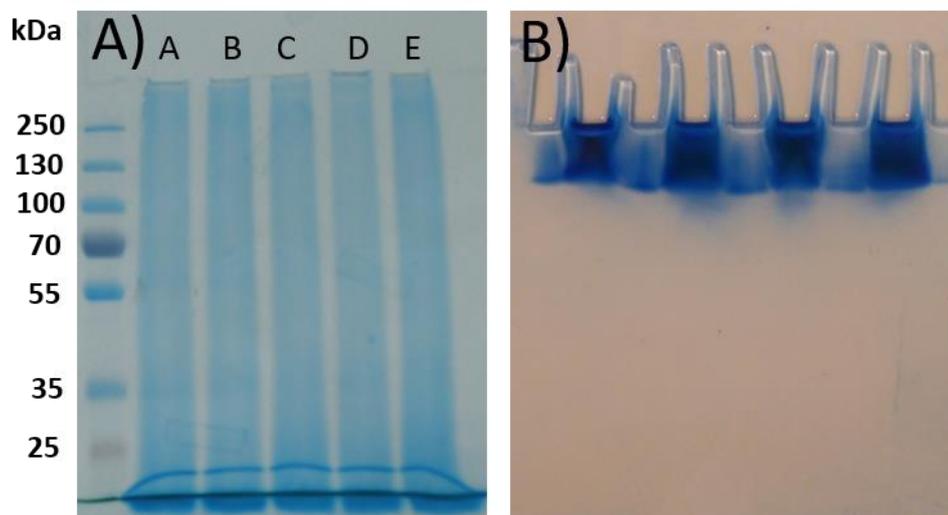


Figure 3.8 SDS-PAGE of metasecretome extraction and storage. A) Week three metasecretome for each biological replicate (cages A-E) post lyophilisation and buffer exchange. B) Four 100µg technical replicates of pooled week three biological replicates short-run SDS-PAGE.

Table 3.3. Quantitative metasecretome recovery by biological replicate per time point. All values expressed are µg of protein.

| Week | Biological replicate | | | | | Mean | Error | Total |
|-----------|----------------------|--------|--------|--------|--------|--------|-------|---------|
| | A | B | C | D | E | | | |
| 1 | 274.99 | 335.10 | 321.71 | 329.59 | 315.28 | 315.33 | 10.64 | 1576.66 |
| 2 | 178.93 | 270.28 | 172.37 | 208.07 | 330.79 | 232.09 | 30.14 | 1160.43 |
| 3 | 385.58 | 323.33 | 373.13 | 347.33 | 408.23 | 367.52 | 14.79 | 1837.60 |
| 4 | 328.78 | 227.83 | 285.58 | 292.63 | 437.68 | 314.50 | 34.79 | 1572.50 |
| 5 | 310.17 | 346.79 | 345.48 | 293.11 | 384.06 | 335.92 | 15.85 | 1679.61 |
| 6 | 183.25 | 234.05 | 300.59 | 215.15 | 295.08 | 245.62 | 22.83 | 1228.11 |
| 8 | 172.41 | 317.39 | 415.82 | 215.55 | 334.93 | 291.22 | 43.57 | 1456.10 |
| 10 | 401.88 | 307.55 | 428.36 | 319.55 | 432.48 | 377.96 | 26.88 | 1889.82 |

The dried *Spartina biomass* was hygroscopic and determined to absorb up to 350% dry weight in water. If the biomass was not saturated within the first 7 days this may have affected the relative dry weight biomass used to extract the metasecretome. The partial conversion factor was applied to this data and week one and two were significantly larger ($p < 0.05$ assessed by a one-way ANOVA) from all other time points (week two was not significantly different from week four) (Figure 3.9).

Time series in situ decomposition of lignocellulose in a salt marsh

Suggesting saturation occurred during week two to three. This may have had an effect on the amount of biotinylated protein extracted, as greater dry weight mass can harbor more organisms and bind more enzymes

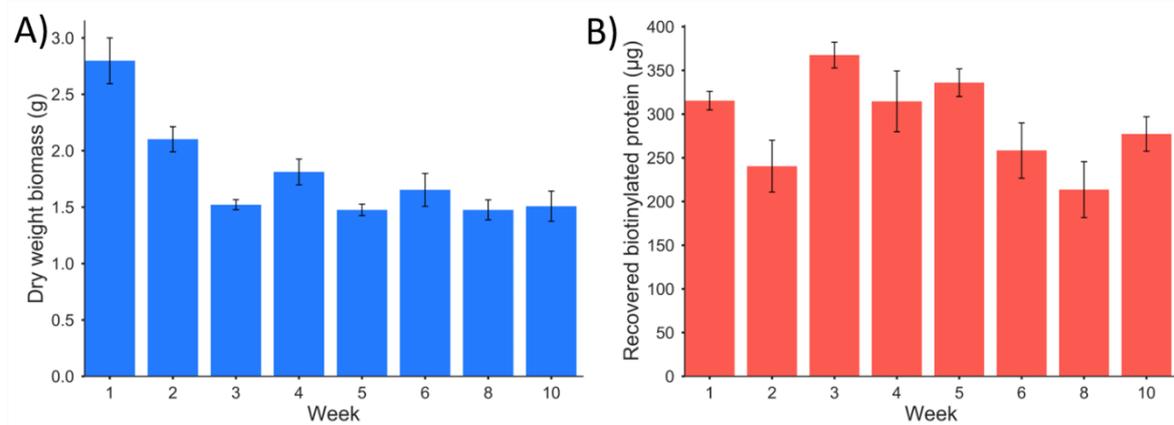


Figure 3.9 Dry weight biomass and recovered protein.

To explore if the effect of dry weight biomass as a function of the amount of protein recovered, regression analysis was conducted which revealed no correlation ($R^2 = 0.012$), however as a function of time distinct clusters were observable (Figure 3.10). Weeks one and two were distinct outliers (Figure 3.10B), however all other time points (weeks three to six) and weeks eight and ten were not distinctly different. These results suggest that the median weeks (belonging to weeks three, four, five and six) were the most productive (highest metasecretome yield per dry weight biomass), with an observable decline in weeks eight and ten. To assess this; protein yields were normalised to per unit dry weight and analysed with regression.

Time series in situ decomposition of lignocellulose in a salt marsh

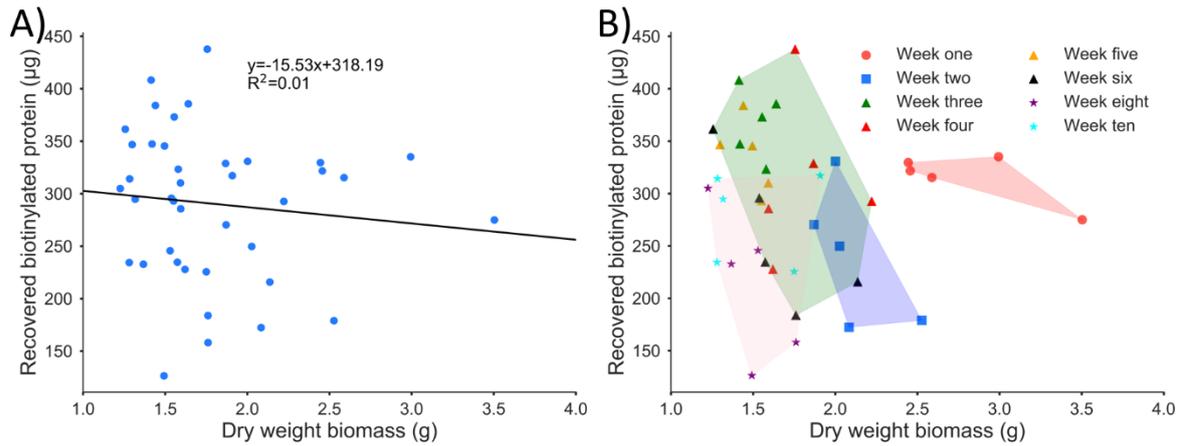


Figure 3.10 Analysis of protein recovered as a function of dry weight biomass. A) Regression analysis. B) Temporal cloud analysis.

The normalised yields revealed a weak correlation ($R^2=0.10$) indicative that water content of the biomass would not significantly effect protein yields (Figure 3.11). This also revealed the least productive weeks to be week one and two.

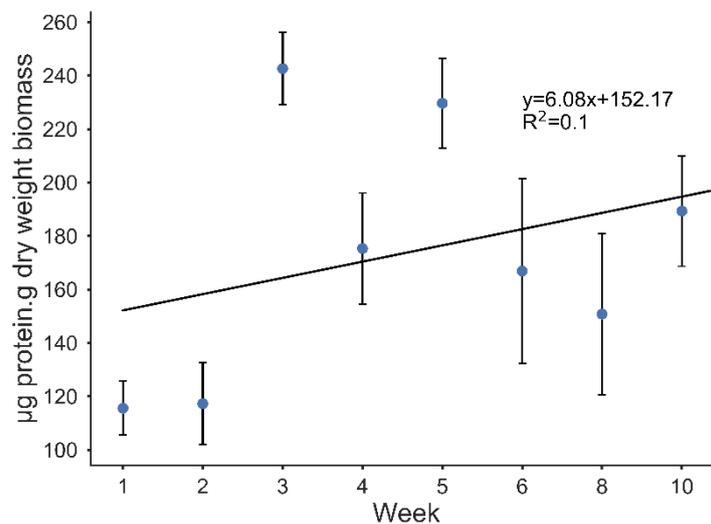


Figure 3.11 Regression analysis of normalised protein yields per dry weight unit biomass.

3.3.4.1 Metasecretome activity

Enzyme assays were conducted to ascertain (in consideration with mass loss) which time points to take forward for sequencing. The proteins that were pooled in equal mass ratios (those placed into gel storage for LC-MS/MS analysis) were utilised and pooled further to generate an equal-mass pool

Time series in situ decomposition of lignocellulose in a salt marsh

for all time points. Significant activity was measured against all substrates tested (CMC, PASC, beechwood xylan, AZCL-xylan, and AZCL-arabinoxylan) except for AZCL-galactomannan and ABTS. An anomalous reading in xylan beechwood experiment yielded a significance value of $p=0.0647$, exclusion of which yielded $p=0.033$ and was taken as significant (Figure 3.12). The high metasecretome loadings ($0.5\text{-}3\mu\text{g}$) required to observe activity restricted further analysis to one substrate, for which the most active substrate (beechwood xylan) was selected, which required $3\mu\text{g}$ crude protein for observable activity in reasonable timeframes.

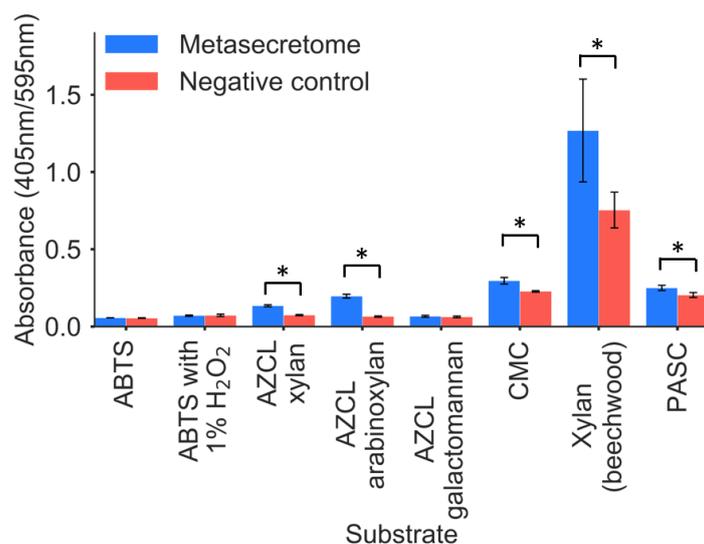


Figure 3.12 Lignocellulolytic activity of the pooled metasecretome across all time points. Asterisks indicate significance at a confidence interval of $p < 0.05$ assessed by a one-way ANOVA.

Analysis of metasecretome xylanase activity for the time course revealed an increase in activity over time (Figure 3.13). Increased activity was observed in saline buffers compared to the non-saline counterparts. Activity was lowest in weeks one and two and began to increase to week six where it largely plateaued to week ten (excluding a slight decline in week eight). This suggests an increase in activity within the *in situ* biomass.

Time series in situ decomposition of lignocellulose in a salt marsh

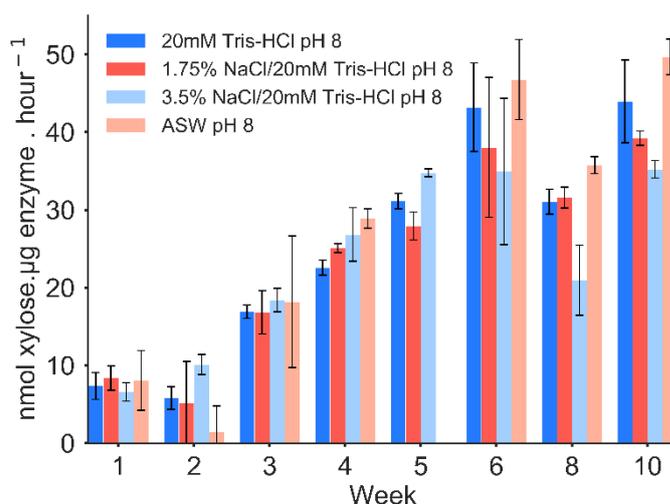


Figure 3.13 Metasecretome xylanase activity by time point. ASW; artificial seawater.

The rates of activity observed here were normalised against the dry weight biomass used to extract the protein (as both were quantified) and extrapolated to total *in situ* activity by assessing dry weight remaining. Extrapolating activity to dry weight biomass remaining in the bags indicates relative activity rates (Figure 3.14A). Significant activity was observed increasing from 7.23 to 92.18 nmol xylose reducing sugar equivalents per hour per gram of dry weight biomass (1.08 and 13.8mg of xylose respectively) in artificial seawater (ASW). Extrapolating this rate further to dry weight remaining indicates the theoretical maximum potential xylanase activity of a specific *in situ* degrading biomass bag at a certain time point (Figure 3.14B). This translates to total activities in the *in situ* degrading biomass bags of 259.5 to 1743.49 nmol xylose reducing sugar equivalents per hour (week one and week ten respectively). However, these experiments were performed on pure polysaccharide substrates and lignocellulosic substrates are significantly more complex, conferring recalcitrance which would display significantly lower rates of product evolution.

Time series in situ decomposition of lignocellulose in a salt marsh

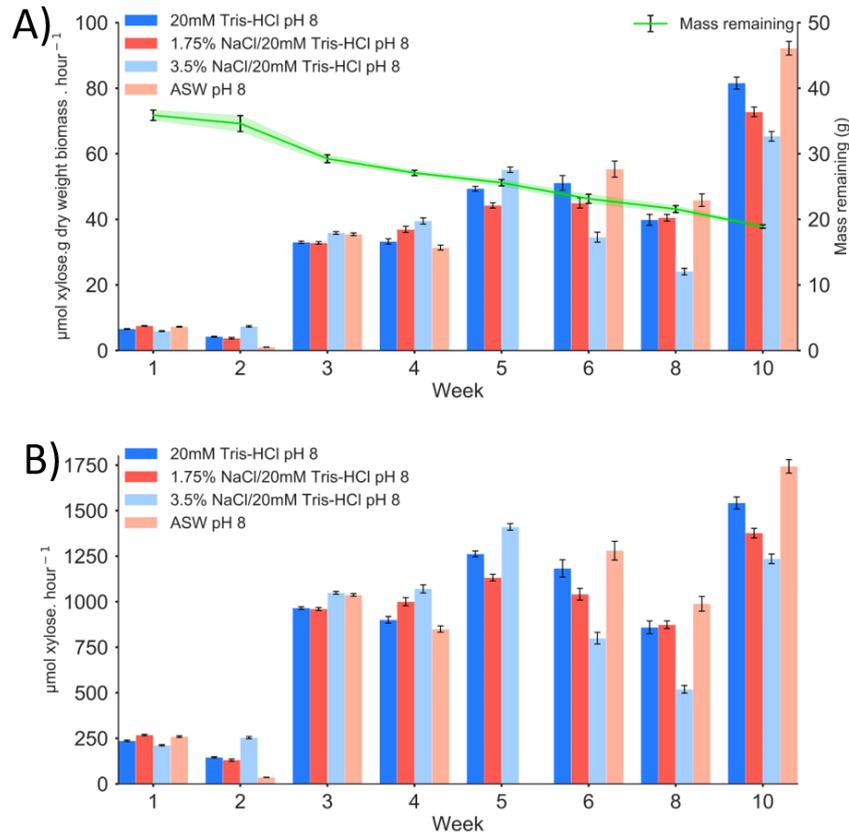


Figure 3.14 Xylanase activity of the metasecretome. A) Metasecretome xylanase activity normalised to per unit dry weight biomass. B) Maximum potential metasecretome xylanase activity as a function of dry weight biomass remaining per time point.

In all of the presented data, the temporal activity increases, suggesting that enzymatic activity is increasing in concentration per unit biomass, when considering the decreasing trend in remaining biomass and increased enzyme activity. This is confirmed with regression analysis ($R^2=0.647$) (Figure 3.15).

Time series in situ decomposition of lignocellulose in a salt marsh

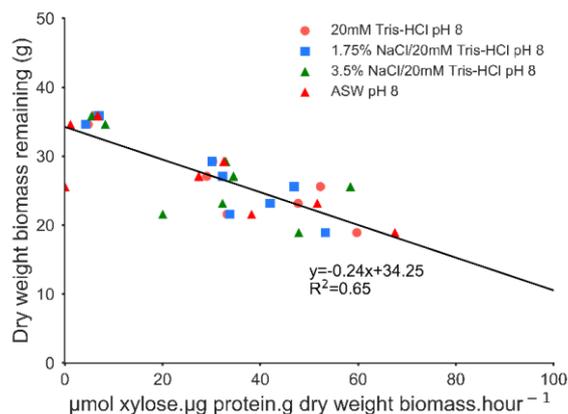


Figure 3.15 Regression analysis of xylanase activity concentration.

Residual metasecretome protein was used to assess hemicellulose activity with AZCL-arabinoxylan and AZCL-xylan. Protein quantity restricted these analyses to relative expression and, therefore, could not be extrapolated to biomass remaining. Whilst the profiles displayed differ, the increase over time trend persisted (Figure 3.16).

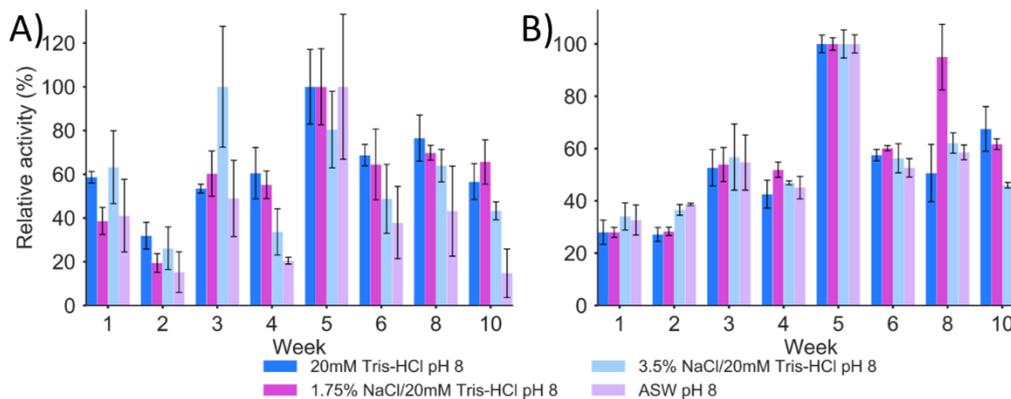


Figure 3.16 Relative metasecretome activity rates for AZCL substrates. A) AZCL-xylan. B) AZCL-arabinoxylan.

The results were generated to indicate which time points were most active, in order to maximise the probability of identifying these enzymes in downstream analyses. As each profile loosely suggested an increase in activity over time, a temporal distribution of time points was selected for further investigation. This included week one, three, five and ten.

Time series in situ decomposition of lignocellulose in a salt marsh

3.3.5 Biomass composition analysis

Biomass composition analysis was conducted to elucidate the temporal composition profile of the *in situ* biomass. This analysis determined the composition of the major biomass fractions; crystalline cellulose, matrix polysaccharides, lignin and the non-carbon fraction determined by ash.

3.3.5.1 Absolute composition values

3.3.5.1.1 Matrix polysaccharide composition

The hemicellulose fraction increased in relative mass over the time course (Figure 3.17). Total matrix polysaccharides in the day 0 biomass were 19.2% increasing to a temporal maximum of 27.3% in week 5. This would suggest other fractions are preferentially degraded. The largest monosaccharide component was expectedly xylose (xylan is predominantly composed of xylose and is the largest hemicellulose polysaccharide in grasses) followed by galactose and glucose, with notable contributions from rhamnose, arabinose and galacturonic acid. Mannose, glucuronic acid and fucose were not highly abundant. There were significantly different neighbouring time points ($p < 0.05$ assessed by a one-way ANOVA), and a profile suggesting initial lack of degradation, followed by an increased rate, suggesting degradation proceeded non-linearly (Figure 3.18).

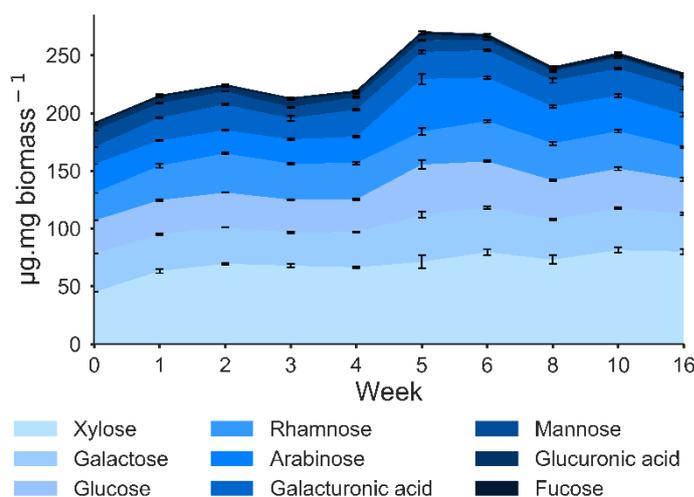


Figure 3.17 Temporal composition of matrix polysaccharide constituent sugars.

Time series in situ decomposition of lignocellulose in a salt marsh

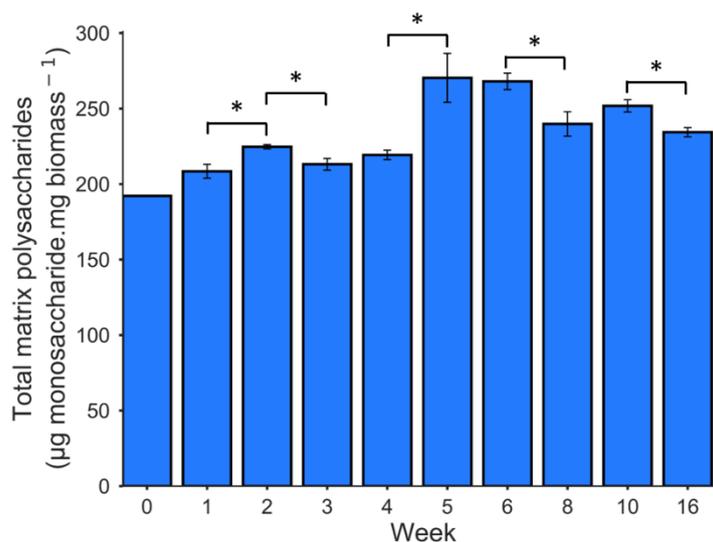


Figure 3.18 Total matrix polysaccharide temporal profile. Asterisks indicate significance at a confidence interval of $p < 0.05$ assessed by a one-way ANOVA.

Interestingly, the relative proportions of the constituent sugars remained largely unchanged, suggesting parallel degradation of all the complex hemicellulose polysaccharides. This would suggest a plethora of enzymatic activators exist within the salt marsh environment to catalyze the deconstruction of the variety of monosaccharide units and bonds in the hemicellulose matrix.

3.3.5.1.2 Crystalline cellulose

Crystalline cellulose displayed an interesting temporal profile with no change within the first two weeks, followed by an increase to week 5 which declines thereafter (Figure 3.19). The pristine day 0 biomass consisted of 15.5% crystalline cellulose, with temporal minima and maxima (13.1% and 20.9% respectively) at weeks five and eight respectively. The initial stability would suggest no preferential cellulose degradation was occurring whilst the following increase also suggests that other fractions were degraded during this time point, however the matrix polysaccharides exhibited a similar temporal profile.

There are no statistically significantly different neighboring data points for the temporal crystalline cellulose profile. This would suggest gradual linear degradation of this fraction with no rates of increase, likely due to limited accessibility by the matrix polysaccharides.

Time series in situ decomposition of lignocellulose in a salt marsh

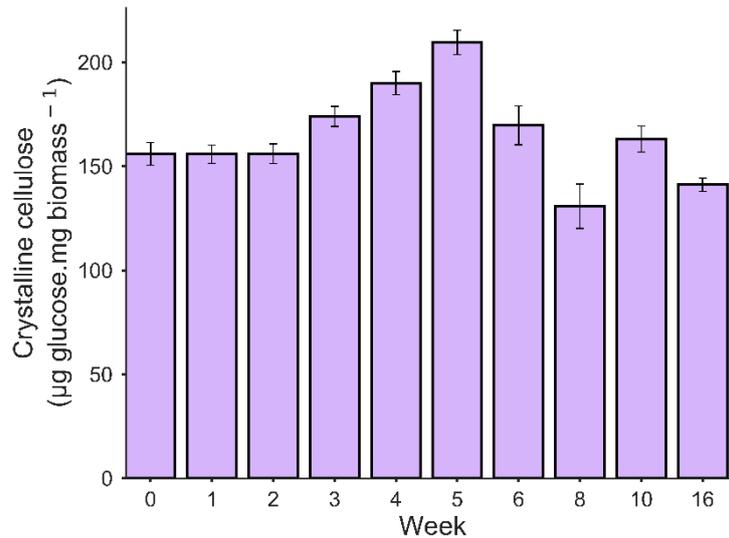


Figure 3.19 Temporal *in situ* biomass crystalline cellulose profile. There were no statistically significant neighbouring data points (confidence interval of $p < 0.05$ assessed by a one-way ANOVA).

3.3.5.1.3 Acetyl bromide soluble lignin

The lignin fraction measured as the acetyl bromide soluble fraction exhibits an increase in abundance per unit weight over time (Figure 3.20). Despite a decline in the first two weeks from 26.97% in the day 0 outgroup to 20.5% in week two, which is the lowest in the time course. The profile gradually increases to the later time points, suggesting accumulation of this fraction. Week eight and ten are statistically different neighbouring time points.

The initial removal of lignin in the first two weeks is offset by an increase in matrix polysaccharides. The matrix polysaccharides may have increased in abundance as lignin restricts enzyme accessibility and is therefore the first to be degraded. In week two lignin is observed at its least abundant, during this week, hemicellulose begins to degrade as lignin deconstruction may have exposed matrix polysaccharides. During this time cellulose has remained completely unchanged, likely still sheltered by matrix polysaccharides.

Time series in situ decomposition of lignocellulose in a salt marsh

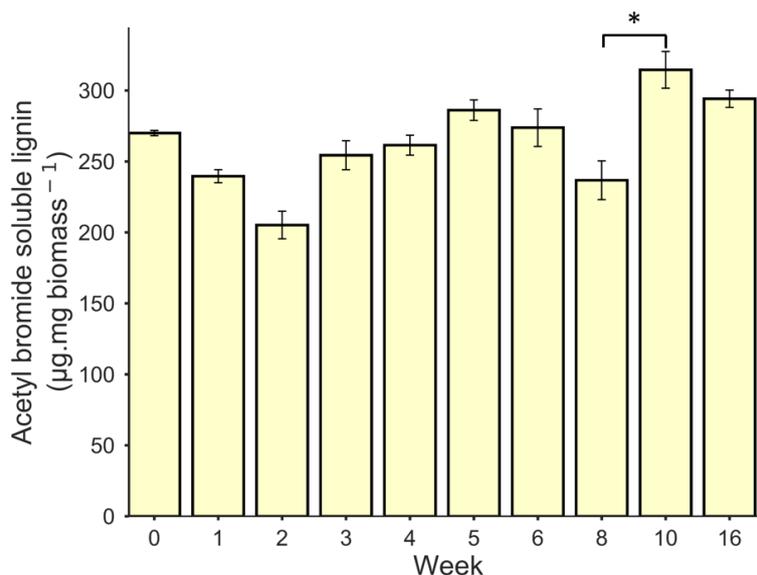


Figure 3.20 Temporal *in situ* biomass lignin profile. Lignin measured as the acetyl bromide soluble fraction. Asterisks indicate significance at a confidence interval of $p < 0.05$ assessed by a one-way ANOVA.

3.3.5.1.4 Ash

The ash profile displayed huge fluctuations, multiples of the original day 0 outgroup biomass. This suggested external factors were influencing the ash constituent in the biomass (Figure 3.21). Ash measured the non-carbon (non-combustible) fraction. The salt marsh is a moderately disturbed environment rich in iron, sulfur and silica, it is, therefore, likely that this material was washed into the biomass containing bags. The *in situ* bags were 200µM, and the biomass was washed through a 100µM filter during processing. Degraded biomass becomes increasingly fibrous and mesh sizes larger than 100µM under constant flow may have resulted in unnecessary biomass loss, resulting in inaccurate measurements. Therefore, particle sizes between 100µM-200µM would have been retained. Due to the entry of exogenous non-carbon material, this profile is unlikely to be a true representation of the biomass ash profile.

Time series in situ decomposition of lignocellulose in a salt marsh

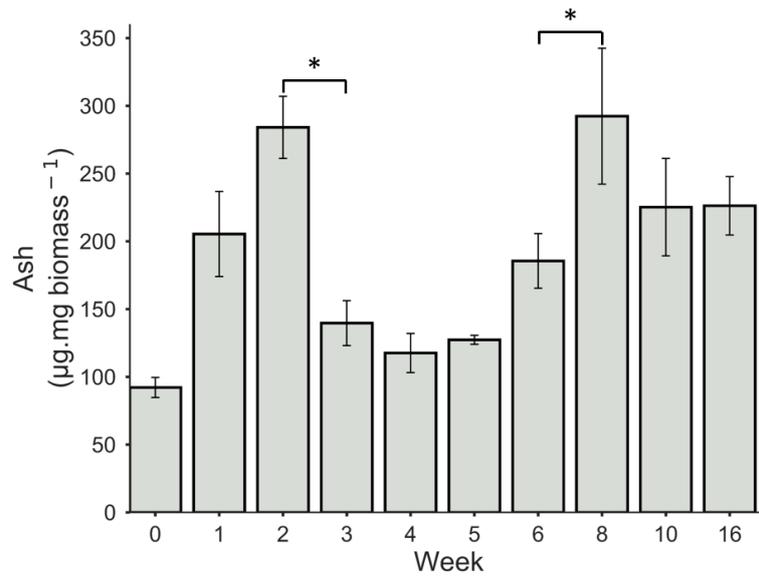


Figure 3.21 Temporal *in situ* biomass ash profile. Asterisks indicate significance at a confidence interval of $p < 0.05$ assessed by a one-way ANOVA.

In consideration of potential detrital entry due to mesh size discrepancy, the change in ash content was explored in relation to tidal height. Tidal energy (and turbidity) correlates with tidal height, therefore larger (spring) tides can resuspend and deposit greater amounts of sediment (Wang, 2002, Byun and Wang, 2005). To affirm this regression analysis was conducted (Figure 3.22). A positive correlation between inter week ash deviation and tidal height was found ($R^2=0.63$). Therefore, it is concluded that the proportions of ash measured in this study are not inherent to the *in situ* biomass, but also contain externally deposited or internally eroded material.

Time series in situ decomposition of lignocellulose in a salt marsh

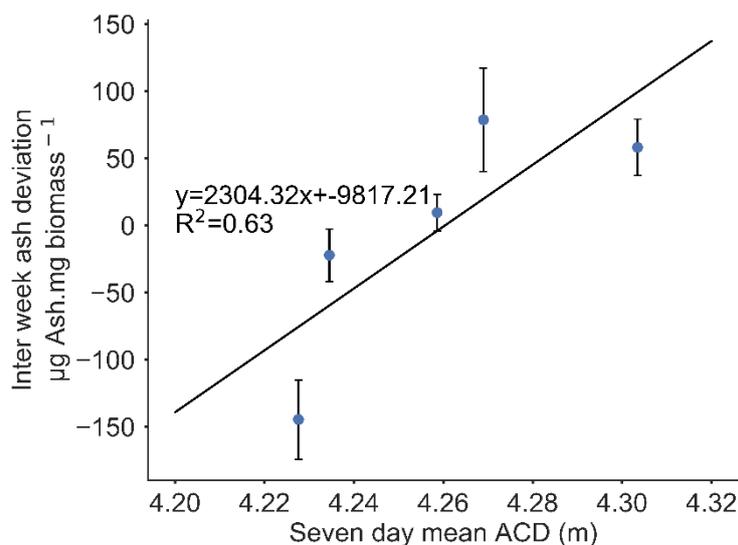


Figure 3.22 Regression analysis of inter week ash deviation and mean tide height. ACD; above chart datum.

3.3.5.1.5 Integrated biomass composition

A holistic view of all the fractions is indicative of a rate limiting hierarchy of fractions, where lignin is rate limiting for hemicellulose degradation, and hemicellulose is rate limiting for cellulose degradation (Figure 3.23-24). Whilst all fractions are degraded at every time point, the changes in relative rate of degradation or accumulation (inter week deviation) displays three phases of degradation during which complex interplay is observable; a likely prerequisite of equilibrium between fractions (Figure 3.24). In the early weeks of the experiment (week one to three; Figure 3.24, phase I), the proportion of cellulose remains unchanged whilst lignin is degraded at an increasing rate. Hemicellulose increases but the rate of change remains constant, suggesting minimal degradation. This would suggest lignin is limiting degradation in the early time points. These results are concordant with the hemicellulose assays in section 3.3.4 where hemicellulase activity is largely muted in the first two weeks for three hemicellulose substrates.

During the mid-stage of degradation (weeks three to five), cellulose begins to accumulate at an increasing rate suggesting reduced degradation (Figure 3.24, phase II). This is likely as the fractionation technique used to size reduce biomass exposed cellulose microfibrils that were degraded and exhausted within the first two weeks. During this phase, hemicellulose begins to be selectively degraded as the rate of hemicellulose degradation begins to increase whilst simultaneously lignin is increasingly disregarded and begins to accumulate. Only during week five (phase III) the rate of cellulose degradation begin to increase.

Time series in situ decomposition of lignocellulose in a salt marsh

The enzyme assays (section 3.3.4) revealed week five to be the most active hemicellulolytic time point, where derivative mass loss revealed the greatest accumulation in this time point. It may be week five was an inflection point to increased hemicellulolytic activity. This may have resulted in the consistent decline in hemicellulose observed afterwards.

During phase I, lignin degradation likely exposed matrix polysaccharides which were selectively degraded during this phase, with cellulose remaining centrally barricaded and not easily accessible to enzymes. During the later time points (week 5 to 16) all fractions appear to be lost (week 10 appears to be an anomaly), with preferential degradation of crystalline cellulose suggesting cellulose is sufficiently exposed (Figure 3.24, phase III). These time points are the only time points in which all fractions are lost and excluding cellulose, appear to be lost at equal rates. This suggests degradative equilibrium has been achieved and complex interplay between the rate-limitation hierarchy is occurring to progress degradation optimally.

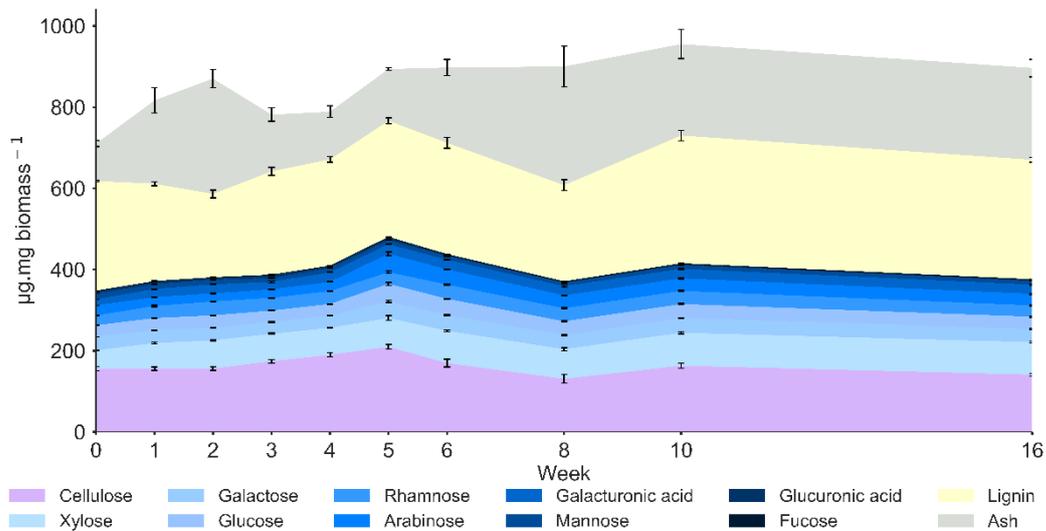


Figure 3.23 Lignocellulose composition of *in situ* biomass.

Time series in situ decomposition of lignocellulose in a salt marsh

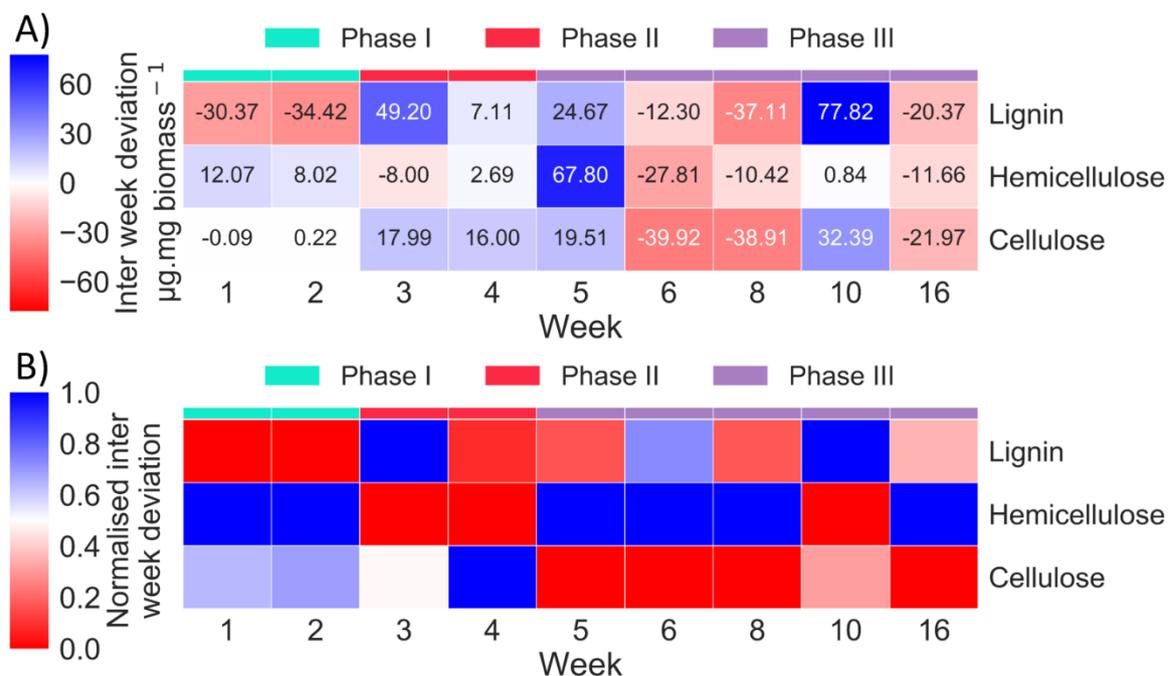


Figure 3.24 Rate of change of the lignocellulose fractions. Red indicates decrease, blue indicates increase, relative to the previous week. A) Absolute rate of change. B) Normalised rate of change by time point. Phase I: accelerated lignin degradation, Phase II: accelerated hemicellulose degradation, Phase III: Accelerated cellulose degradation and degradative equilibrium reached.

3.3.5.2 Extrapolated composition values

Extrapolating absolute mass composition values to mass remaining *in situ* elucidates a complete profile of temporal change (Figure 3.25). All mass loss was closed (relative to the total) to greater than 70%, with the later time points closed to 90% or above. This would suggest that amorphous cellulose and other fractions represent a significant fraction of *Spartina anglica* biomass which may be entirely degraded in the later time points (or supplanted with ash).

Time series in situ decomposition of lignocellulose in a salt marsh

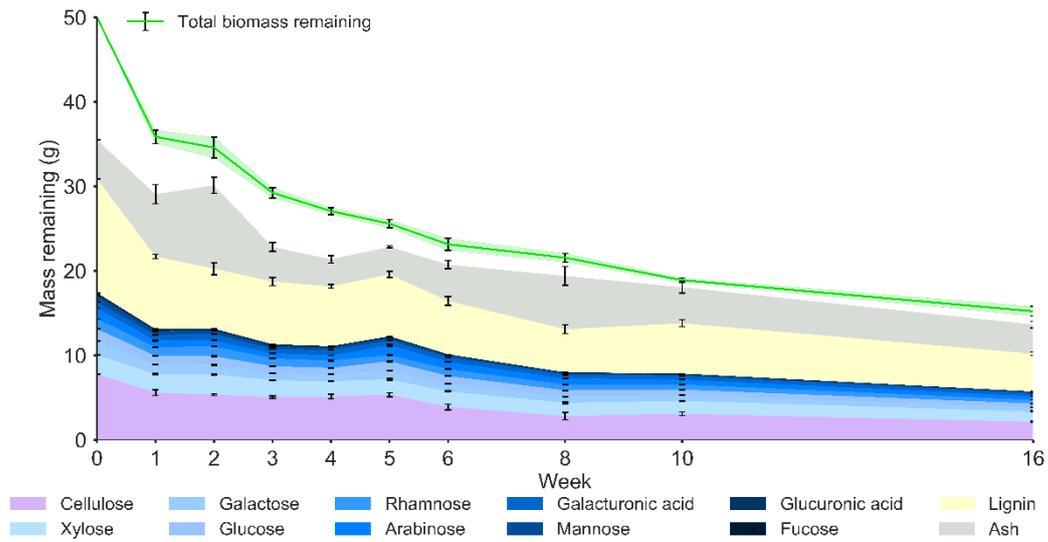


Figure 3.25 Composition of *in situ* remaining biomass.

Normalisation of all values reveals the extent of tidal-driven sediment deposition and erosion (Figure 3.26). Therefore, ash was excluded from further analysis (Figure 3.27). The ash excluded profile reveals largely linear degradation of all fractions, despite subtle changes in the rates of degradation explored in section 3.3.5.1.5. This would suggest a diversity of lignocellulolytic enzymes targeting all fractions should be present at all time points.

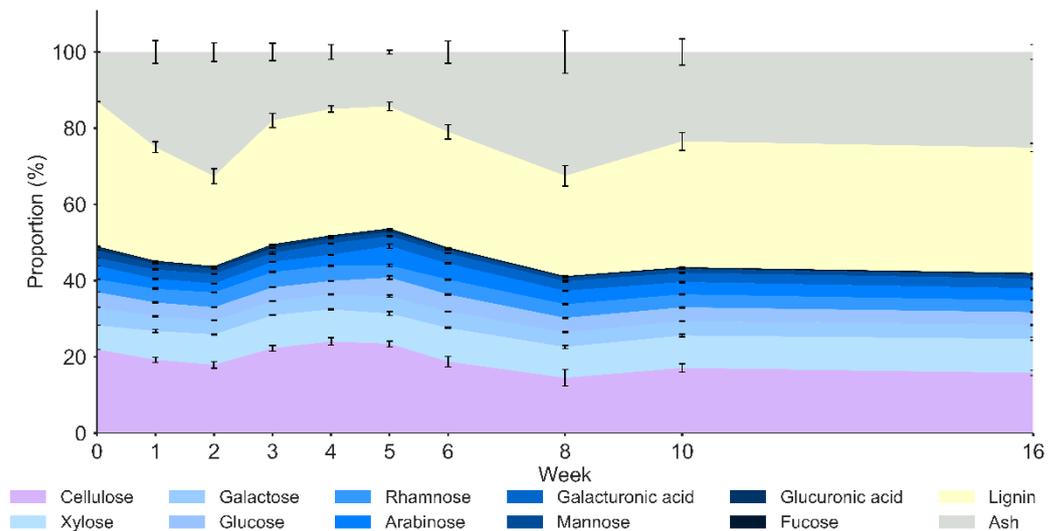


Figure 3.26 Normalised composition of remaining biomass inclusive of ash.

Time series in situ decomposition of lignocellulose in a salt marsh

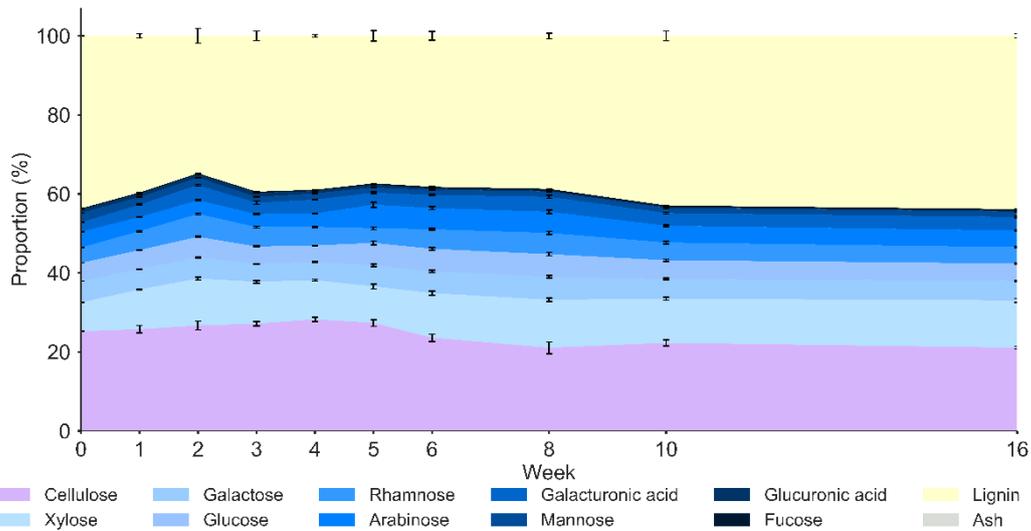


Figure 3.27 Normalised composition of remaining biomass excluding ash.

3.4 Discussion

The Humber estuary is one of the largest estuaries in England and Welwick salt marsh is the largest contiguous well-established marsh on the north bank of the Humber estuary. Historically, Welwick has been fractionated by extensive reclamation since the 1770s for agricultural and woodland purposes, with an estimated minimum of 0.3km² of natural marsh remaining (Armstrong et al., 1985). Welwick is now an actively accreting, expanding salt marsh which is a rare occurrence in the current state of global salt marshes (Andrews et al., 2008). Salt marshes accrete when the rates of material retention exceed that of erosion and is imperative to salt marsh persistence against rising sea level (Chmura, 2013). More recently, during this study, the banks surrounding the reclaimed areas (adjacent to Welwick) were breached to allow salt marsh realignment, further facilitating salt marsh accretion.

One major source of accretion facilitating material is the primary production and retention of lignocellulosic biomass (Chmura, 2013). Therefore, accreting salt marshes are likely to have healthy lignocellulolytic microbiomes capable of remineralising this sugar rich feedstock. In addition to this sediments at Welwick (and surrounding areas) exhibit invariably high carbon concentrations of 2-4% (by weight) from the lower to upper marsh, which decreases gradually with depth, suggestive of decomposition and remineralisation (Andrews et al., 2000, Lamb et al., 2007, Andrews et al., 2008).

As Welwick lies on the outer estuary it was unlikely to be significantly affected by the estuarine freshwater input. This was reflected in the measurements during the experiment, where salinity was

comparable with seawater and stable between 26-40ppt. Therefore, Welwick salt marsh can be categorised as a marine salt marsh. This is key to the identification of halotolerant enzymes.

The nonlinear rate and magnitude of mass loss (69% during the 16-week season) in this study fits a first order decay model which has been alluded in previous necromass-bag approaches (Curco et al., 2002, Negrin et al., 2012, Simoes et al., 2011). These results also suggest the majority of biomass degradation occurs within the first year of entry into the system. Previously three phases of decay in salt marshes have been described which include a leaching phase (leaching of soluble biomass components), followed by a decomposition period (characterised by a slower decomposition) and a final refractory phase (where decomposition rates are greatly diminished) (Valiela et al., 1985). The leaching phase reportedly lasts two to four weeks. In this study, most of the leaching occurred within the first week, putatively due to the extent of biomass size reduction (increase surface area for leaching) and immobilisation on the lower marsh (greater inundation frequency). The greatest loss in mass occurred between day 0 and week one ($28.26 \pm 1.58\%$), followed by the lowest loss in mass between week one and two ($2.52 \pm 2.44\%$) which is where the decomposer phase likely begins.

The initial rate of decomposition in the decomposer phase likely resulted from the established microbiome satisfying their nutritional requirements from the leaching compounds. It is probable the lignocellulolytic enzyme production coincided with the relative exhaustion of leaching compounds. The enzyme assays revealed minimal activity in weeks one and two, concordant with the sudden change in rate of loss. This phenomenon has been reported to an even greater extent where an increase in biomass has been reported directly after a large decline (Buth and Dewolf, 1985). This is ostensibly a consequence of many factors including microbial establishment (mass increase) and entry of exogenous material (mass increase) and an absence of degradative enzymes (lack of mass loss). An increase in this study was not observed, likely due to the washing process not utilised in other studies.

The results published by Valiela et al., (1985) results suggest the refractory period is characterised by and confined to decomposition rates below 0.4% weight loss per day. The refractory period in this study may have been experienced in weeks eight, ten and 16 observed to be 0.22 ± 0.087 , 0.379 ± 0.11 and 0.176 ± 0.03 mass loss per day respectively. Therefore, this study has captured a large proportion of the decomposition process.

Studies have assessed the effect of temperature and inundation on decay rates, with large differences occurring between seasons; however, this experiment was concluded within one season and as such the effect of temperature variation should not have played a significant role (Kirwan and Mudd, 2012). Biotic factors were the focus of this study and rates of decay as a function of abiotic

factors were not explored, instead, the molecular mechanism of biomass decay were assessed through changes in biomass composition.

Previous studies have anatomically fractionated plants to assess degradation and expectedly the recalcitrant, lignified parts (stem, sheath) are more resistant (Liao et al., 2008). Only one study has assessed the biomass composition changes during decomposition (Valiela et al., 1984). This investigation explored the relative composition of *Spartina alterniflora* for 24 months, beginning in winter. The study by Valiela et al., (1984) is not directly comparable to this study due to the location, start date and different species of biomass, which significantly affects decomposition (Simoes et al., 2011); however, there is an undeniable synchrony between the profiles. An initial increase in cellulose and hemicellulose, sequentially followed by lignin degradation, then hemicellulose degradation. The hemicellulose degradation then coincides with cellulose degradation whilst lignin increases. This fits the rate limitation hierarchy asserted in this Chapter.

In all time points in this study, all biomass fractions are deconstructed. However, the rates of deconstruction assessed from the biomass composition indicate three unique phases that follow a rate limiting hierarchy exist that prohibit optimal degradation. All of these phases occurred within the decomposer phase discussed above. To date, microbial decomposition as a function of the complex interplay of biomass composition deviation has not been explored.

In phase I, during decomposition which spanned weeks one and two, the rate of lignin deconstruction increases, whilst the rates of hemicellulose and cellulose deconstruction decreases. Lignin is known to be the most recalcitrant cell wall polymer to deconstruct (Li et al., 2016a). It also structurally fortifies the inner matrix polysaccharides (and cellulose), limiting accessibility for deconstructive enzymes. For lignocellulose degradation to progress, initially (phase I) lignin must be removed, to expose the matrix polysaccharides to enzymatic digestion.

Phase II occurred from week three to four, and is characterised by an increased rate of hemicellulose deconstruction, which is now likely more exposed than during phase I. As with lignin which shields the matrix polysaccharides, the matrix polysaccharides provide a physical barrier to cellulose. Progression of the biomass deconstruction reaction requires matrix polysaccharides to be removed. During this phase, the rate of cellulose hydrolysis decreases, resulting in relative accumulation. This phase begins (week three) with a reduction in the rate of lignin deconstruction, and by week four the rate is significantly increased. The sudden increase in hemicellulose hydrolysis in week three likely offsets production of ligninolytic enzymes, which are required during week four to maintain equilibrate degradation of the two fractions.

Time series in situ decomposition of lignocellulose in a salt marsh

Only in the final phase (Phase III) does the rate of cellulose hydrolysis increase significantly, which is now exposed post matrix polysaccharide removal in phase II. Cellulose has the greatest rate of deconstruction in this phase for all but one time point (week 10). This would suggest complex interplay between the rate limiting fractions as the relative rate of degradation of all fractions approaches equilibrium, allowing the reactions to process optimally.

The enzyme assays are concordant with the biomass analysis. On all hemicellulose substrates, weeks one and two displayed significantly lower hemicellulose activity, which coincides with the rate of loss decline. The greatest hemicellulolytic activity is observed in week five, biomass deviation suggests the rate of change for hemicellulose deconstruction is greatest in week four (compared to previous weeks), with week five eliciting the first increase in the rate of cellulose deconstruction. This suggests week five is an inflection point during which hemicellulose is no longer rate limiting and after this point, the composition of hemicellulose decreases in relative abundance.

To identify the halotolerant enzymes involved in the process of lignocellulose deconstruction in salt marsh environments, a temporal profile of time points were selected for further processing. Weeks one, three, five and ten were selected for DNA, RNA and protein sequencing. This information was used in conjunction to identify the key organisms and enzymes which are discussed in further detail in Chapters 4 and 5. Selected identified enzymes are recombinantly expressed in Chapter 6 and assessed for halotolerant and biotechnological potential.

Chapter 4 Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

4.1 Introduction

A large proportion of lignocellulolytic enzymes are extracellular and often strongly bind to the lignocellulosic biomass (Yu et al., 2013, Alessi et al., 2017). In salt marshes, this phenomena may be magnified as retaining proximity to the secreting organism and biomass upon tidal inundation (and potential loss in the water column) would be energetically favourable. The EZ-link-sulfo-NHS-SS-biotin labelling technique in this study developed by Alessi et al., (2017) is well suited for the identification of these biomass bound enzymes in a salt marsh environment, which would otherwise not be observable. This technique enables labelling of proteins at the biomass interface and facilitates the enrichment of extracellular (biomass-bound and unbound) and transmembrane proteins from the total protein fraction (the tag is impermeable to cell membranes) through a streptavidin affinity column. The resultant protein fraction represents the exo-metaproteome (metasecretome).

Recent advancements in the development of high throughput DNA sequencing technology has facilitated growth in its use driven by cost effective, high quality data acquisition. However, the datasets generated are often complex and their value is often restricted by the availability of well curated archive datasets. This is not an obstacle for well characterised model organisms; however, metagenomic and metatranscriptomic sequence datasets contain orders of magnitude greater complexity than single organism sequencing, for which current archive databases are inadequate for feasible valuable data extraction. This is a significant issue for unexplored environments such as salt marshes where little genetic data has been published.

In this study Illumina sequencing generated large metatranscriptomic databases. The complexity of salt marsh microbiomes, the lack of archived reference datasets and the data size and depth achieved with Illumina sequencing required *de novo* assembly of raw reads into contiguous sequences (contigs) (Grabherr et al., 2011). Proteomic techniques can be used in conjunction with assembled databases to aid in filtering redundant data as the data acquired is generated from putatively functional proteins, alleviating the limitations arising from loose correlations between transcription levels and expression levels (Vogel and Marcotte, 2012).

The proteomic approach used in this study involved liquid chromatography tandem mass spectrometry (LC-MS/MS). An isolated heterogeneous protein fraction is digested to smaller

peptides which are size fractionated by liquid chromatography before mass identification by mass-spectrometry. The unique spectra for each peptide can then be cross-referenced with the metatranscriptomic database and significant matches can be used to discern its originating nucleotide sequence. An enriched database of peptide matching sequences can then be constructed (metasecretome), and non-matching sequences discarded. This metasecretome database provides a non-redundant foundation upon which lignocellulose degrading enzymes can be identified.

This approach facilitates the discovery of novel enzymes by reducing database redundancy by creating a profile of captured proteins consisting of information acquired at the protein level only; this can be semi-quantitatively compared in relative terms. Additionally, this centralises focus on proteins that would be otherwise overlooked if archive databases are inadequate. Multi-faceted 'omics approaches are becoming well practised in investigating enzyme associations and functionality in lignocellulose deconstruction (Zhu et al., 2016, Jimenez et al., 2016, Alessi et al., 2017, Alessi et al., 2018).

There are a variety of tools available for the identification of putatively lignocellulose-associated proteins. These tools rely upon homology modelling (Hidden Markov Models) which use a probabilistic score approach in recognising regular expressions within consensus sequences logos, which are determined by the database curator. Therein lies the caveat of these tools, whereby they are only as powerful as the information contained within the reference (sequence logos). These tools can be applied to the sequences contained within a metasecretome database to identify signal peptides (SignalP v4.1) and regions homologous to known carbohydrate active domains (dbCAN, pFAM, NCBI NR_Prot) (Petersen et al., 2011, Yin et al., 2012, Finn et al., 2016). The resulting information can be parsed to identify the secreted lignocellulose associated proteins from transmembrane and contaminating proteins.

The depolymerisation of cellulose for industrial applications has long been known to be not only rate limiting but also prohibitively expensive (Cherry and Fidantsef, 2003, Klein-Marcuschamer et al., 2012). There are a number of approaches to increase the economic viability of enzyme cocktails and currently the most popular approach is the top down approach. This approach improves current cocktails by supplementation (post production or by engineering *Trichoderma reesei* strains) of missing synergistic activities (such as AA9 family enzymes, β -glucosidase and xylanase to Cellic[®] CTec1 (Novozymes) to formulate CTec3) (Sun et al., 2015a, Druzhinina and Kubicek, 2017, Takashima et al., 1999). The inability to modify the starting cocktail (beyond culture conditions) represents the major disadvantage of a top down approach, which in the case of *T. reesei* is surprisingly not optimal (Billard et al., 2012).

The alternative bottom up approach is being increasingly reported where synthetic cocktails are constructed and optimised on an individual enzyme and substrate basis, facilitating total customisation (Karnaouri et al., 2016, Billard et al., 2012). These cocktails require well characterised, highly active enzymes that are tolerant of inhibitors to be competitive. A bottom up approach also provides an opportunity to exchange the media from freshwater to seawater, the world's most abundant and inexpensive resource, as the solvent of choice in biorefining processes. Seawater is a complex chemical substance containing approximately 3.5% salt (84% sodium chloride, 14% potassium chloride, and over 82 trace salts), and complicated further by the presence of diverse chemical species (Elderfield et al., 2006, Turner et al., 1981, Byrne et al., 1988). Current cocktails are impaired in this medium (Grande and de Maria, 2012). Therefore, novel halotolerant enzymes must be identified, and synthetic cocktails constructed to apply untreated seawater as a solvent in biorefining.

The techniques described above were used in conjunction to simultaneously examine the ecology of the lignocellulose associated exo-metasecretome and identify putative halotolerant lignocellulolytic enzymes with biotechnological potential. These techniques were applied during an *in situ* degradation experiment of *Spartina anglica* biomass in Welwick salt marsh and 'omics approaches were undertaken during weeks one, three, five and ten of a 16-week time course (described in Chapter 3). Data redundancy was managed and CAZy enzymes were identified using a novel pipeline. Protein profiles were assimilated from the generated databases and analysed at the ecosystem, enzyme family and protein levels. The most likely taxonomic origin of these identified proteins was compiled to further investigate the functional ecology of the system.

4.2 Methods

4.2.1 Transcriptomic analysis

4.2.1.1 DNA depletion

The extracted nucleic acid mixture containing genomic DNA and total RNA obtained in Chapter 3 was utilised in this section for the depletion of DNA and subsequent enrichment and isolation of mRNA. Biological replicates (cages A-E) were pooled in equal ratios for each time point (week one, three, five and ten) and DNA contamination was removed using a DNase Max kit (MoBio UC-15200-50) according to the manufacturer's protocol.

4.2.1.2 *mRNA enrichment*

These experiments were conducted in triplicate for each time point (week one, three, five and ten), from the pooled DNA depleted samples in section 4.2.1.1. The samples were cleaned using Clean & Concentrator™ columns (ZymoResearch R1015) according to the retention of >17nt protocol. The quality of the resulting total RNA was assessed with a Bioanalyser 2100 (Agilent) and samples with an RNA Integrity Number (RIN) of greater than 7.5 were taken further. Messenger RNA (mRNA) was enriched by depleting ribosomal RNA (rRNA) using Ribo-Zero™ Magnetic Epidemiology rRNA removal kit (RZE1224/MRZ11124C; Illumina) according to the manufacturer's protocol. The enriched mRNA was cleaned using Clean and Concentrator™ columns as above according to the retention of >200nt protocol. Technical replicates were assessed using the Bioanalyser and pooled if suitable quality was obtained.

4.2.1.3 *RNA sequencing*

A 25µL volume (at a minimum concentration of 2 ng µl⁻¹) of rRNA-depleted RNA (section 4.2.1.2) was provided to the Next Generation Sequencing Facility (NGS) at the University of Leeds. RNA-seq libraries were prepared using a NEBnext RNA Ultra Library preparation kit with NEBnext single 6bp indexing primers (New England BioLabs) according to the manufacturer's protocol. Quality checking was conducted using an Agilent Bioanalyser (HS DNA chip) and quantified using the Qubit high sensitivity DNA kit (Thermo) and pooled in equimolar ratios. The pooled RNA-seq library was spiked with 1% PhiX and sequenced on a single lane of an Illumina HiSeq 3000 2 x 150 base pair chip at the University of Leeds NGS.

4.2.1.4 *Transcriptome assembly (de novo)*

The transcriptome assembly in this study was performed by Dr Yi Li (Centre for Novel Agricultural Products, University of York). The raw reads of the sequenced libraries were searched against Silva_115 database to identify ribosomal RNA (rRNA) genes using the Bowtie2 software package (Quast et al., 2013, Langmead and Salzberg, 2012). rRNA reads, orphan reads in the paired reads and poor-quality sequences were removed with the ngsShoRT software (Chen et al., 2014). Duplicated reads in each library were removed and the remaining sequences were assembled de novo with the

Trinity software package (Grabherr et al., 2011). Read counts and gene expression abundance (estimated) were obtained with the utility programs within the Trinity software package. The databases for week one and five were obtained with an archive version of Trinity (trinityrnaseq_r20140413p1), the databases for week three and ten were obtained using Trinity version 2.2.0.

4.2.1.5 Filtering and processing

The de novo assembled transcriptome databases (metatranscriptomes) were subject to various filtering techniques to reduce dataset size, remove background noise and remove redundancy. There is currently no consensus approach to preparing metatranscriptomic datasets. A common filtering technique for assembled transcriptomes originating from single organisms is filtering by fragments per kilobase of transcript per million mapped reads (FPKM) as a proxy of relative contig abundance (Hart et al., 2013). Where density against FPKM elucidates an abundance profile which can be normalised, to which a gaussian distribution can be fitted to; contigs that fall within the gaussian distribution can be considered highly expressed.

FPKM values for every contig in each database was extracted and assimilated relative to FPKM value density. The zFPKM values were calculated as in equation 4.1 (Hart et al., 2013). The best fit gaussian distribution was calculated and fit to this data and used to determine the best zFPKM threshold for data filtering. It was found that the metatranscriptomic databases in this study did not conform to a typical binomial distribution (with a larger proportion of less abundance contigs, the opposite of which is true for single organism transcriptomics, likely arising from the complexity within these samples), causing the binomial distribution to fit to the smallest, less abundant contigs, which would filter out a large proportion of valuable data.

Equation 4.1 zFPKM. Gaussian distribution parameters are given by μ and σ .

$$zFPKM = \frac{\log_2(FPKM) - \mu}{\sigma}$$

An alternate method of filtering is to filter smaller sequences (and by proxy of their small size filters low FPKM value contigs). A contig size of greater than or equal to 500bp was selected for filtering. The transcriptome databases were pooled into a single database containing 29,938,8668 contigs and filtered as above to 1,725,357 sequences. The contigs in this filtered database were

separated into all open reading frames (ORFs) using Emboss GETORF (<http://www.bioinformatics.nl/cgi-bin/emboss/getorf>) that are greater than or equal to 300bp and contain alternative initiation codons (including methionine and leucine). Using this methodology an ORF database containing 2,726,694 sequences was generated. This ORF database was dereplicated by sequences to generate 2,400,360 unique ORFs using a custom python script. Finally, the databases were concatenated into a single database as merged databases yield more accurate results in proteomic searches (Tanca et al., 2016).

4.2.2 Proteomic analysis

4.2.2.1 *Liquid chromatography tandem mass spectrometry (LS_MS/MS)*

LC-MS/MS analysis was undertaken to identify proteins present with the collected fractions from their constituent peptides. The proteomic experiments in this section (4.2.2.1) were conducted by Dr Adam Dowle (Technology Facility, University of York) after (Dowle et al., 2016).

Tryptic digestion was performed for in-gel proteins post reduction with DTE and *S*-carbamidomethylation with iodoacetamide. Gel pieces for weeks one, three, five and ten obtained in Chapter 3 were washed twice with 50% (v:v) aqueous acetonitrile containing 25 mM ammonium bicarbonate. A final wash was conducted with acetonitrile which was then dried in a vacuum concentrator for 20 min. Modified porcine trypsin (Promega) was dissolved in 50 mM acetic acid and diluted with 25 mM ammonium bicarbonate to generate a trypsin solution of 0.02 µg/µL. 25 µL of trypsin solution was added to the gel pieces and incubated for 10 min before adding 25 mM ammonium bicarbonate (until the gel pieces were submerged). The gel pieces were incubated for 16 hours at 37°C. Three washes were performed with 50% (v:v) aqueous acetonitrile containing 0.1% TFA (v:v), dried in a vacuum concentrator and reconstituted in aqueous 0.1% trifluoroacetic acid (v:v).

The acquisition of peptide spectra was achieved using a nanoLC system interfaced with a maXis HD LC-MS/MS system (Bruker Daltonics) and a CaptiveSpray ionisation source (Bruker Daltonics). Positive ESI-MS and MS/MS spectra were acquired using AutoMSMS mode. Instrument control, data acquisition and processing were performed using Compass 1.7 software (microTOF control, Hystar and DataAnalysis, Bruker Daltonics). Instrument settings were: dry gas; 3 L/min, ion acquisition range; m/z 150-2,000, MS/MS spectra rate; 5 Hz at 2,500 cts to 20 Hz at 250,000 cts, quadrupole low mass; 300 m/z , cycle time; 1 s, ion spray voltage; 1,450 V, collision RF; 1,400 Vpp, transfer time; 120

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

ms, MS spectra rate; 5 Hz and dry gas temperature; 150°C. Isolation and collision energy width settings were calculated automatically using the AutoMSMS fragmentation table; absolute threshold 200 counts, preferred charge states: 2–4, singly charged ions excluded. A single MS/MS spectrum was acquired for each precursor and former target ions were excluded for 0.8 min unless the precursor intensity increased fourfold.

4.2.2.2 Peptide to protein identification pipeline

Peptide spectra generated with LC-MS/MS were cross referenced with the nucleic acid sequences contained within the pooled ORF database using Mascot version 2.5.1 (Matrix Science Ltd.), through the ProteinScape interface version 2.1 (Perkins et al., 1999). Mascot uses a probability bases scoring algorithm to assess peptide significance which is used predominantly to mitigate false positive results; the peptides are small (5-9 amino acids in length) and as such are likely to match to proteins they do not originate from (Perkins et al., 1999). Mascot searching uses a database of theoretically redundant nucleotide sequences to search against the peptide spectra, and as these sequences are redundant, no peptide matches are expected. Therefore this is thought to give an accurate representation of the number of false positives found in the target (ORF) database which is recorded as the false discovery rate (FDR) (Elias et al., 2005). If a peptide spectra scores more highly in the decoy database search than the target database search, it is recorded as a false positive.

There are issues with FDR and erroneous false positives using decoy database as the accuracy is effected by the size of the spectra database and the size of the target database (intrinsically linked to the size of the decoy database) (Tanca et al., 2016). When the target database is large, the theoretically redundant decoy database contains significantly more information, resulting in the identification of more false positives. In the context of this study where the target secretome (non-redundant fraction) is proportionally insignificant compared to the redundant fraction. This is particularly inconvenient for low abundant proteins and low abundant transcripts. This is inconsequential for single organism target databases (e.g. 8,000 and 26,00 sequences for *E. coli* and *H. sapiens* respectively), but for databases exceeding ~80,000 sequences these principles begin to fail (Noble, 2015, Tanca et al., 2016, Shanmugam and Nesvizhskii, 2015). In the case of this study, the incredible size and complexity of the metatranscriptome and its originating environment magnifies this detrimental effect.

To overcome these limitations redundant sequences were removed by leveraging non-redundant sequences in an initial round of sequencing against subsets of sequences with high search stringencies (Figure 4.0). The ORF database was fragmented into 21 blocks of ~115,000 random sequences. Each of these sequence blocks were searched against each of the 12 peptide spectra databases (252 searches in total) with Mascot. The search criteria both searches were +2/+3/+4 peptide charge, peptide tolerance ± 10 ppm, modifications; carbamidomethyl and oxidation. The initial round of searching was performed with a significance threshold of $p=0.05$. The peptide matching ORFs (PMOs) from each of these outputs were extracted and concatenated into a single 'true hit' secondary database. This secondary database was finally searched against the peptide spectra databases with a significance threshold of $p=0.1$ and the result of these searches used to construct a metasecretome database.

Analysis was primarily conducted using a quantitative exponentially modified protein abundance index (emPAI, equation 4.2) which estimates the absolute abundance of proteins from LC-MS/MS data (Ishihama et al., 2005). emPAI values for each protein were then normalised against the emPAI sum for each peptide spectra database to generate the approximate molar percentage.

Equation 4.2 Exponentially modified protein abundance index (emPAI). Where $N_{observed}$ is the number of experimentally observed peptides in the protein and $N_{observable}$ is the maximum calculated number of observable peptides.

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

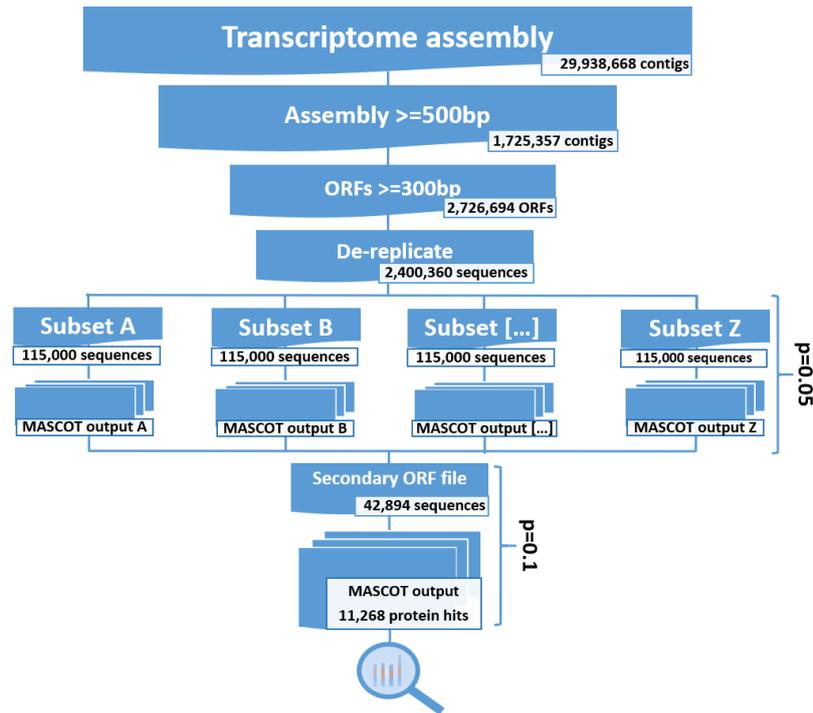


Figure 4.1 Protein identification pipeline. The primary database is fragmented into 21 subsets of 115,000 sequences and searched against peptide spectra Mascot searched independently with a significance threshold of $p=0.05$, the results of these are used to construct a secondary database with reduced redundancy. A second search is performed with the secondary database with a significance threshold of $p=0.1$.

4.2.3 Bioinformatic techniques in the analysis of the metasecretome

Multiple databases and methods were used to identify and extract information from the sequences contained within the metasecretome (Figure 4.2). Carbohydrate active enzymes (CAZy) within the metasecretome databases were identified by searching against the dbCAN database using HHMER3 (Yin et al., 2012). Carbohydrate binding domain activities were also assimilated from the dbCAN database. Additional annotations were ascribed with the Pfam database using HMMER3, identifications from the dbCAN database were valued more highly than Pfam and used as the strongest reference (Finn et al., 2016).

The metasecretome database was also searched against the NCBI non-redundant protein database (NR_prot; 62 files downloaded in February 2017) using BLAST+ (BlastP) version 2.2.31 with an expect value (e value) threshold of $1e^{-5}$ (Camacho et al., 2009). The resulting best-hit (highest expect value) was obtained for each protein in the proteome and an NCBI Accession number database was compiled. The most likely taxonomic origin of these proteins were extracted from their respective Accession numbers using tools within the Environment for Tree Exploration (ETE)

version 3 toolkit (Huerta-Cepas et al., 2016). Accession numbers were used to extract taxonomic identifiers (TaxID). The TaxID was used to extract a full taxonomic lineage. All subsequent analysis was conducted using custom programs written in python versions 2.7, 3.4, 3.5 and 3.6.

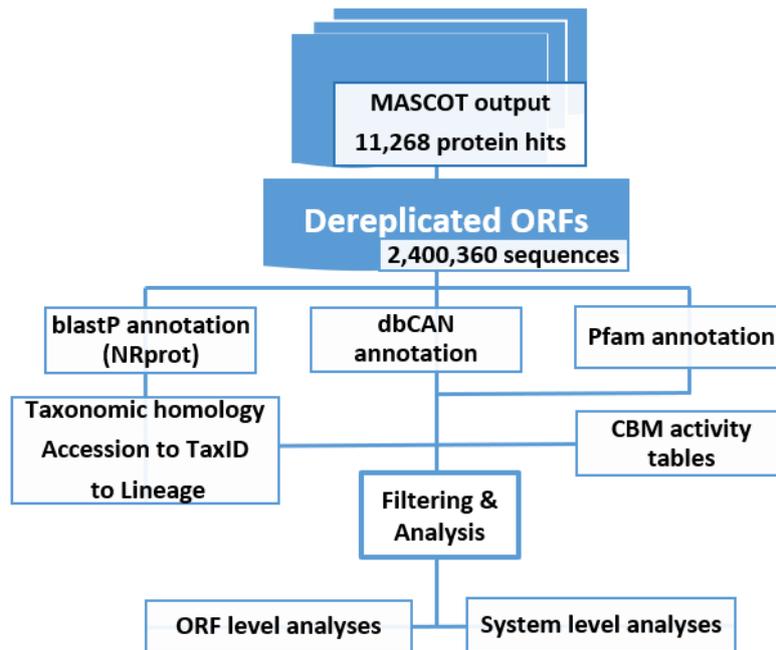


Figure 4.2 Analysis pipeline for the annotation and analysis of the PMO database. PMO; peptide matching open reading frame.

4.3 Results

4.3.1 Transcriptomic filtering and processing

4.3.1.1 *zFPKM and size exclusion filtering*

Illumina sequencing generated a total of 383,122,461 raw reads across the four time points (weeks one, three, five and ten). These reads were filtered for ribosomal RNA contamination (rRNA) and dereplicated before assembly with Trinity. The assembly generated 29,938,668 contiguous sequences (contigs) across the four time points (Table 4.1).

Table 4.1 Sequence reads throughout RNA sequence processing and assembly.

| | Week 1 | Week 3 | Week 5 | Week 10 | Total |
|---|------------------|------------------|------------------|------------------|-------------------|
| Raw reads | 82,966,972 | 99,319,322 | 95,318,915 | 105,517,252 | 383,122,461 |
| Percentage reads | | | | | |
| rRNA | 10.71% | 1.58% | 2.74% | 2.04% | - |
| Paired end reads (- rRNA, -duplicates) | 68,307,441 | 91,948,286 | 85,685,871 | 93,709,355 | 339,650,953 |
| Paired end reads (- post QC) | 68,302,150 | 91,926,491 | 85,676,818 | 93,700,742 | 339,606,201 |
| Contigs (Trinity) | 4,100,865 | 9,720,697 | 7,501,904 | 8,615,202 | 29,938,668 |
| Genes (Trinity) | 3,869,004 | 9,132,509 | 7,071,997 | 8,153,028 | 28,226,538 |
| Median contig length | 187 | 195 | 185 | 201 | - |
| Average contig length | 268.28 | 247.46 | 255.67 | 249.49 | - |

4.3.1.2 zFPKM and size exclusion filtering

The assembled metatranscriptomes were subjected to two different filtering methodologies. There is no accepted method for preparing metatranscriptomic databases. Filtering with zFPKM is often used for single organism transcriptome filtering and was applied to the metatranscriptomic databases generated in this study (Figure 4.3A-C). The metatranscriptome profile exhibited an inverse density profile to single organism metatranscriptomics, whereby larger, more abundant (high FPKM) values are more densely populated within the profile. The metatranscriptomic profile in this study was populated largely by smaller, less abundant contigs. This is likely a result of the complex nature of this environment which exhibits diverse biological functions acting competitively and synergistically. The use of zFPKM to filter this data would have removed excessive putatively valuable data and was deemed unsuitable (particularly for week one and week five which exhibited an interesting bimodal distribution (Figure 4.3C)

Size exclusion filtering of sequences less than 500bp in length was undertaken to remove smaller contigs and by extension sequences with low FPKM values (Figure 4.3D-E). However, not all sequences had low FPKM values (a small but highly abundant sequence may display high FPKM values (Figure 4.3F). However, the number of these sequences were insignificant therefore this method was deemed suitable and the resultant databases contained 1,725,354 contigs; a significant reduction from 29,938,868 contigs.

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

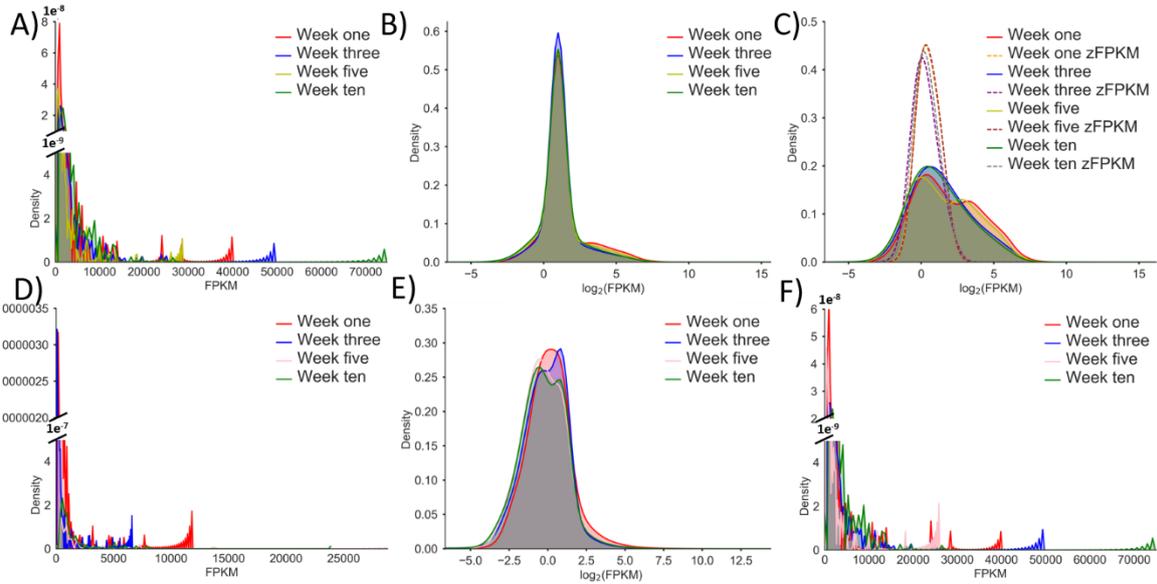


Figure 4.3 zFPKM and size exclusion filtering of metatranscriptomic databases. A) FPKM density distribution for unfiltered databases. B) \log_2 normalised FPKM density distributions, 0 FPKM values were removed. C) zFPKM distributions (overlay) against \log_2 normalised FPKM density distributions, 0 FPKM values were removed. D) FPKM density distribution for size excluded (≤ 500 bp) databases. E) \log_2 normalised FPKM density distribution for size excluded databases. F) FPKM density distribution for the size excluded sequences in D and E.

4.3.2 LC-MS/MS acquired data

Protein from each time point was analysed in triplicate for a total of 12 replicates. Peptide sequences were identified from the proteins recovered from the *in situ* decomposition experiment by Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS), unique spectra (number of peptides) acquired per replicate were consistent irrespective of sample loading (Table 4.2). These databases were searched against the fragmented sequence blocks from the ORF database. The results of this were used to construct a peptide matching ORF database containing 42,894 unique sequences and searched again.

Table 4.2 Protein sample origin, loading and spectra acquired from LC-MS/MS analysis.

| Database identifier | Protein gel loading (μg) | Week | Replicate | Spectra acquired |
|----------------------------------|---------------------------------------|------|-----------|------------------|
| W1 - C201 | 50 | 1 | a | 32,877 |
| W1_3 - C429 | 100 | 1 | b | 39,998 |
| W1_4 - C429 | 100 | 1 | c | 39,735 |
| W3 - C201 | 50 | 3 | a | 32,159 |
| W3_1 - C429 | 100 | 3 | b | 39,509 |
| W3_2 - C429 | 100 | 3 | c | 40,065 |
| W5 - C201 | 50 | 5 | a | 31,713 |
| W5_100 μg _C429 | 100 | 5 | b | 39,369 |
| W5_25 μg - C429 | 25 | 5 | c | 39,642 |
| W10 - C201 | 50 | 10 | a | 31,456 |
| W10_100 μg _i - C429 | 100 | 10 | b | 39,793 |
| W10_100 μg _ii - C429 | 100 | 10 | c | 39,291 |

4.3.3 Protein profile

The final Mascot search with the secondary databases identified 11,268 unique peptide matching ORFs (17,841 total). Of these, 410 were found to be carbohydrate-active enzyme (CAZy) containing within 307 ORFs (2.7%). This ranged between 0.717% and 0.986% of the total proteome by molar percentage (Figure 4.4). The remaining ORFs were not analysed. Glycoside hydrolases were the most abundant CAZy class in all proteomes, representing 86.8% of all CAZy enzymes in week one, decreasing to 74.6% in week five. Carbohydrate esterases increased over the time course from 8.5% in week one, to 14.5% in week three where they then stabilise at 17.4% and 17.1% for week five and ten respectively. Auxiliary activities were not abundant throughout time course ranging between 1.6% to 5.9% (week one and five respectively). Polysaccharide lyases are only observed in weeks three and ten (3.1% and 2.1% respectively). S-layer homology domains commonly associated with membrane anchoring and cellulosome complexes are only abundant in week five and ten (2.0% and 0.59% respectively).

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

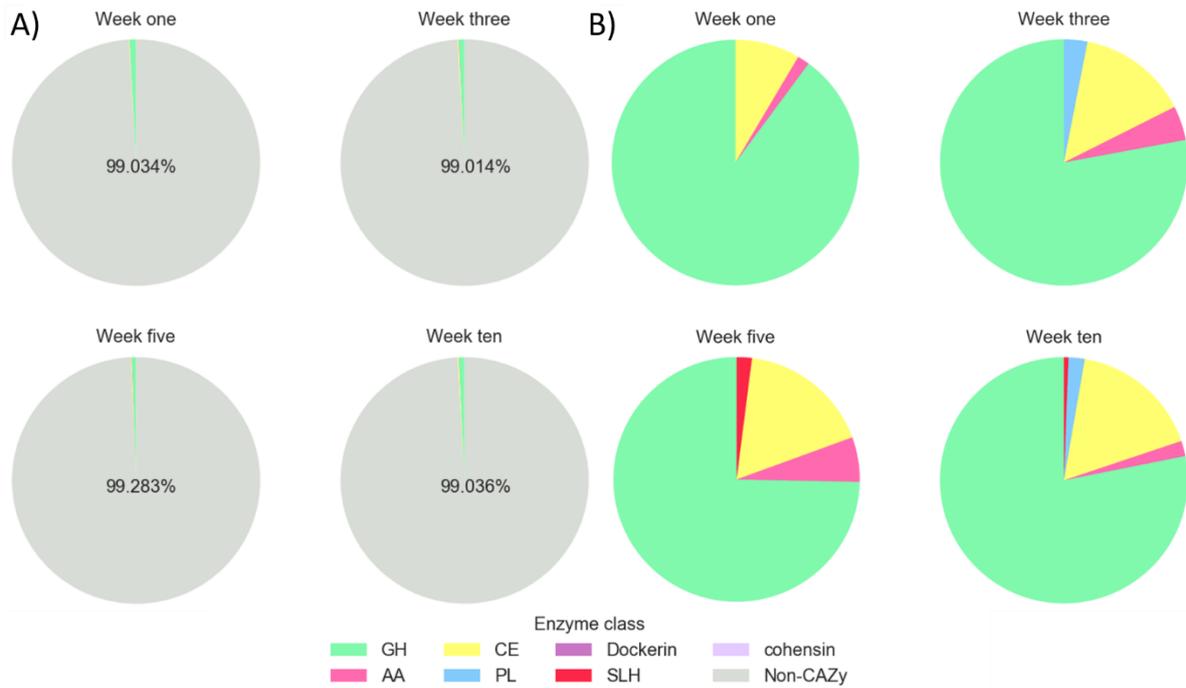


Figure 4.4 Protein and CAZy distribution. A) Protein distribution by molar percentage for the whole proteome including CAZy and Non-CAZy proteins by time point. B) Distribution of CAZy classes by molar percentages by time point. *GH*: Glycoside hydrolase, *CE*: Carbohydrate esterase, *AA*: Auxiliary activity, *PL*: Polysaccharide lyase. *SLH*: S-layer homology.

4.3.3.1 Carbohydrate active enzyme profile

Of the 410 CAZy domains across 307 ORFs, 78 domains within 76 ORFs were below the expected value (e-value) threshold of $1e^{-10}$. Where a domain contained both a domain of high confidence (less than $1e^{-10}$) and low confidence (greater than $1e^{-10}$), the low confidence domain was supplanted back onto its originating ORF. There were 21 supplanted domains, resulting in 55 ORFs being filtered in total (252 remain). The database contained 49 glycosyl transferases (GTs) which were filtered, 203 ORFs remain which were used for all subsequent analysis. Hierarchical Euclidean clustering of the enzyme classes revealed three clusters of temporally highly abundant classes which contain a diverse suite of activities, but are predominantly cellulose targeting (Figure 4.5).

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

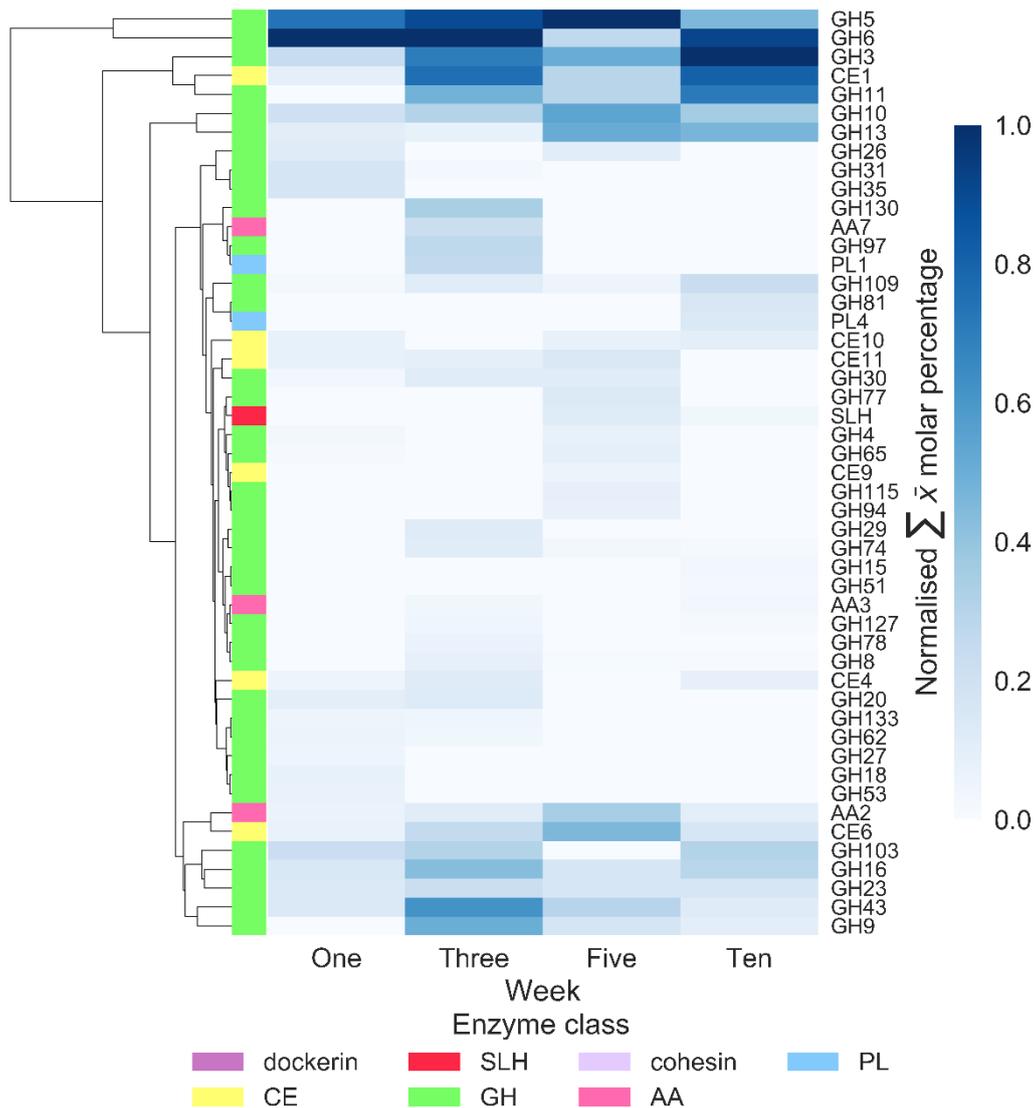


Figure 4.5 Hierarchical clustering of the temporal enzyme class profile (Euclidean). Standard scale normalisation of the sum of the mean molar percentage for each CAZy class was applied. *GH*: Glycoside hydrolase, *CE*: Carbohydrate esterase, *AA*: Auxiliary activity, *PL*: Polysaccharide lyase. *SLH*: S-layer homology.

Throughout the experiment, the number of CAZy classes identified remained consistent with 24 families at every time point excluding week three which contained 30. The most abundant enzyme class were glycoside hydrolases with 37 families identified (Figure 4.5). The most abundant of which were GH6 homologs which contain β -1,4-endoglucanases and cellobiohydrolases (Figure 4.6). This group also contained a_c570829_g1_i1_1, which was the most abundant ORF in both week one and ten whilst being absent from the intermediate weeks, with a mean molar percentage of 0.315% and 0.165% respectively (Figure 4.7). This group was also bolstered by the 7th most abundant ORF overall a_c489891_g1_i1_1. The second and third most abundant classes were GH5 and GH3 respectively,

both of which are diverse classes with a plethora of potential activities. The fourth most abundant group was the carbohydrate esterase 1 (CE1) family which represent acetyl xylan esterases and lignin-associated esterases which contained three highly abundant ORFs a_c412941_g1_i1_1, b_TRINITY_DN851751_c0_g1_i1_1 and c_c997747_g1_i1_1.

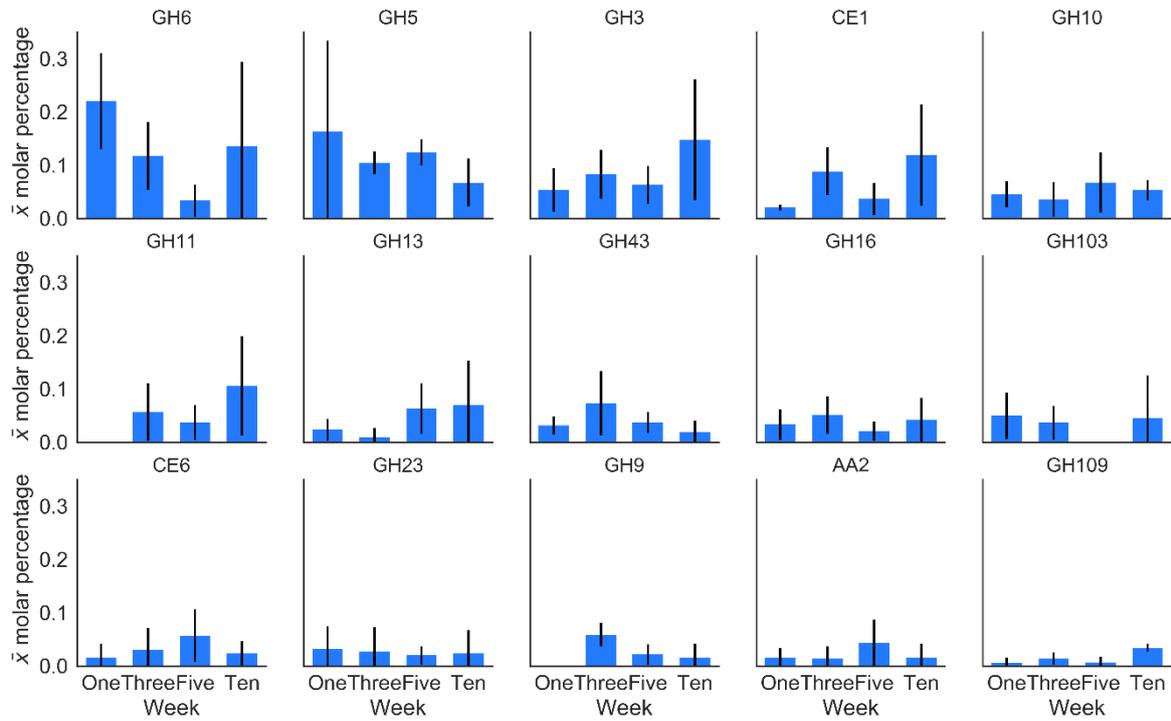


Figure 4.6 Temporal profile of the 15 most abundant enzyme classes. Mean molar percentage is displayed in descending order from left to right. Error bars are displayed as standard deviation. *GH*: Glycoside hydrolase, *CE*: Carbohydrate esterase.

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

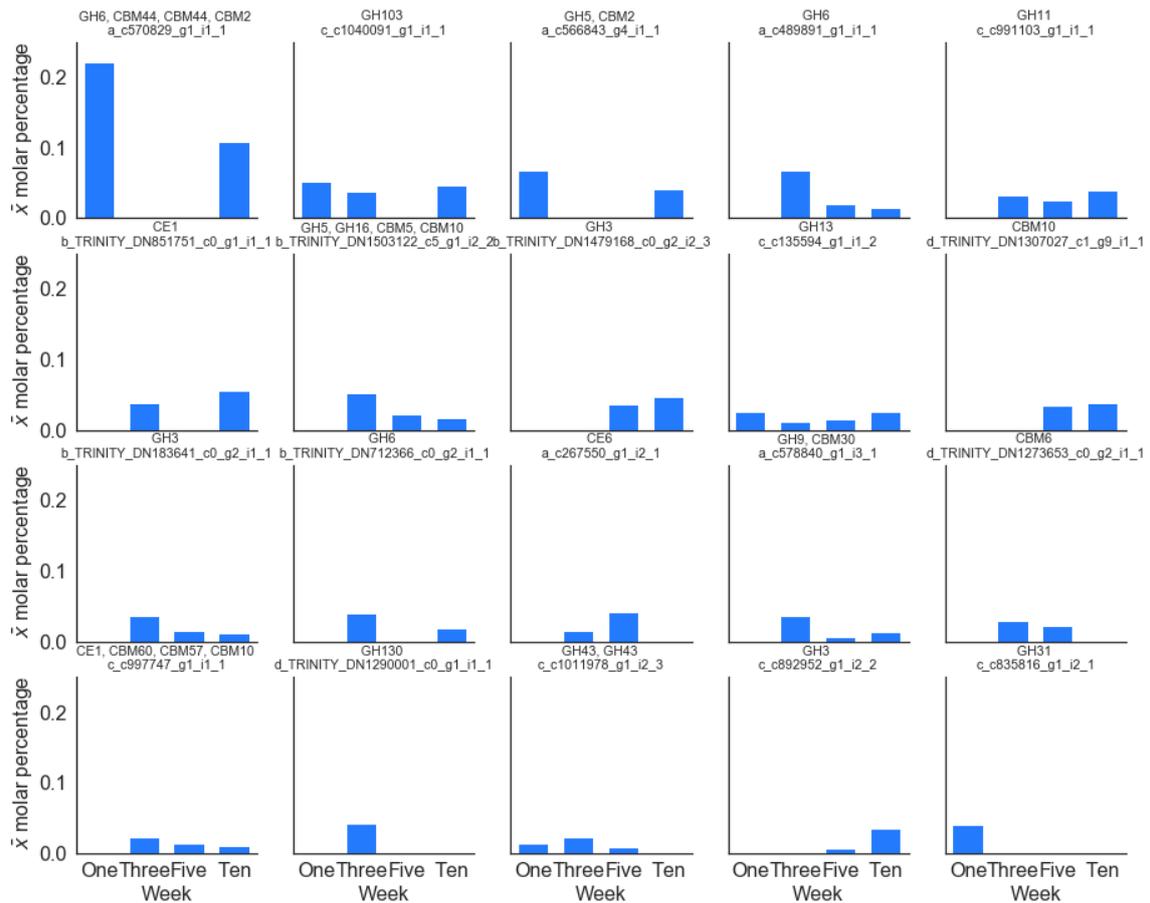


Figure 4.7 Temporal profile of the 20 most abundant ORFs. All contained CAZy domains are displayed, descending order by mean molar percentage from left to right. *GH*: Glycoside hydrolase, *CE*: Carbohydrate esterase, *AA*: Auxiliary activity, *CBM*: Carbohydrate binding domain.

4.3.3.1.1 Carbohydrate esterases

Six carbohydrate esterase (CE) families passed filtering with CE1 being the most abundant in every time point (Figure 4.8). Proteins homologous to CE10 were present in every week with the exception of week three. Interestingly both CE1 and CE10 are putatively involved in lignin degradation. The less abundant families with more sporadic temporal profiles (CE11, CE4, CE6, CE9) are all ostensibly involved in deacetylation of various substrates (acetyl xylan, N-acetylglucosamine and peptidoglycan).

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

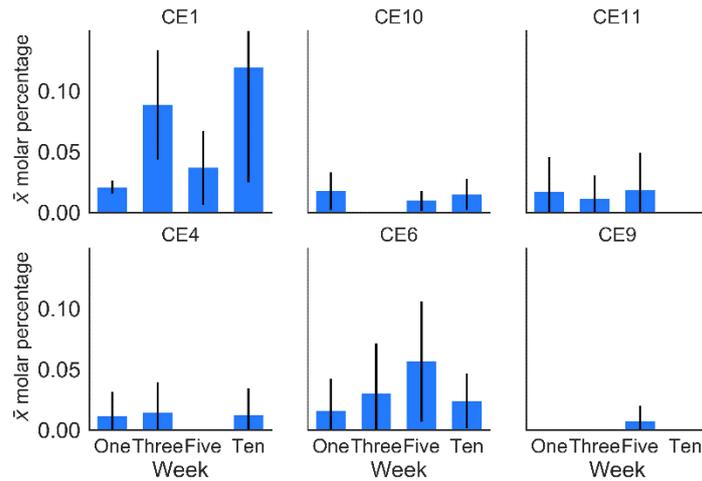


Figure 4.8 The temporal profile of the observed carbohydrate esterase (CE) families. Displayed by the sum of mean molar percentage in descending order from left to right.

4.3.3.1.2 Auxiliary activities

The most abundant auxiliary family present throughout the temporal profile was AA2, a family of peroxidases involved in lignin modification (Figure 4.9). Other families in this class were observed infrequently, which likely originates from the innate anoxic environment.

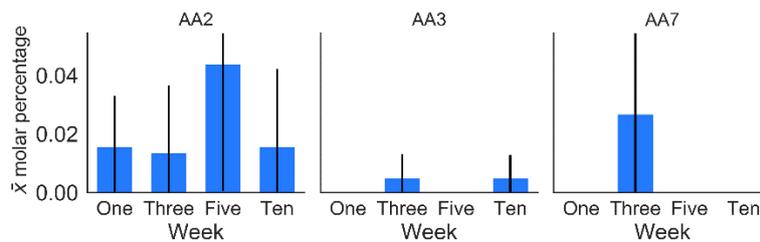


Figure 4.9 Temporal profile of the observed auxiliary families (AA). Displayed by the sum of mean molar percentage in descending order from left to right.

4.3.3.1.3 Polysaccharide lyases

Polysaccharide lyases (PL) were present throughout the experiment but no dominant families were observed (Figure 4.10). PL1 and PL4 appear to be involved in pectin deconstruction; however, the relative abundance of these families and their profile does not correspond to the proportion of rhamnose and galacturonic acid within the composition analysis. This suggests PL are not favoured and the required activities are likely to be found within the glycoside hydrolase profile.

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

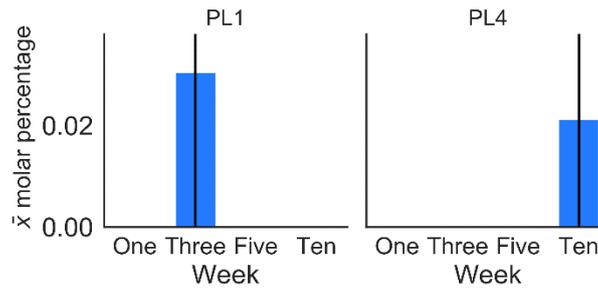


Figure 4.10 Temporal profile of the observed polysaccharide lyase (PL) families. Displayed by the sum of mean molar percentage in descending order from left to right.

4.3.3.1.4 Cellulosome associated domains

While cohesin and dockerin domains were identified, none passed filtering thresholds, suggesting these sequences could be highly divergent from known homologs. S-layer homology (SLH) domains are involved in anchoring protein complexes on cell membranes to form cellulosomes indicative of anaerobism. SLH domains were present in the later time points only becoming abundant during week five through to ten where the experiment was visibly anaerobic (Figure 4.11).

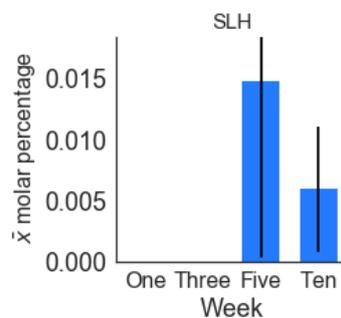


Figure 4.11 Temporal profile of the observed cellulosome associated domains. Displayed by the sum of mean molar percentage in descending order from left to right. *SLH*: S-layer homology.

4.3.3.2 Carbohydrate binding domain profile

Carbohydrate binding domains (CBMs) were less diverse than the enzymatically active CAZy profile with 22 families observed. Two notable clusters (Euclidean) with abundant classes are apparent in this dataset (Figure 4.12). A cluster displaying a similar temporal profile containing CBM2 and CBM44; both of which are predominantly associated with cellulose and hemicellulose binding and are dominant in weeks one and ten. A larger, secondary cluster is apparent containing CBM10,

CBM5, CBM60 which are present in week one but begin to dominate in week three through to ten, these classes are associated with cellulose, hemicellulose and chitin.

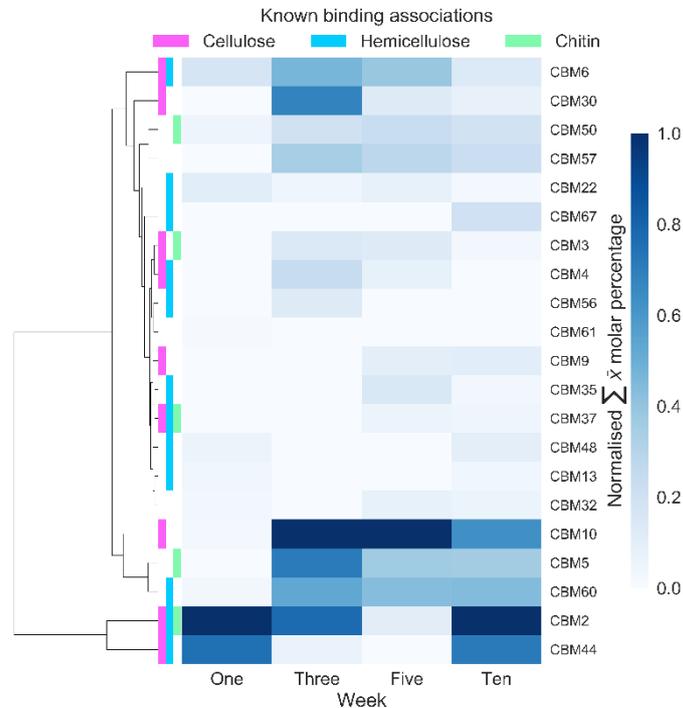


Figure 4.12 Hierarchical clustering of the temporal carbohydrate binding domain profile. Euclidean clustering applied. Standard scale normalisation of the sum of the mean molar percentage for each CBM family has been applied.

4.3.3.2.1 Catalytic associations to carbohydrate binding domains

CBM2 was the most abundant family followed by CBM4 and CBM44, the binding associations suggest proteins accompanying these were targeting exposed cellulose and hemicellulose polymers during the initial phases of the experiment for optimal energetic return (Figures 4.12-14). The onset of a more diverse suite of CBMs in the secondary cluster suggest exhaustion of the favoured exposed polymers and the requirement for a more dynamic set of affinities.

Catalytic associations to CBMs reveal potential activities when analysed in conjunction, which is useful in selecting expression targets. This information is also useful in supplementing knowledge gaps, such as CBM57, which currently has no known function, appearing to be associated to GH10, GH11 and CE11 families, which would convey a link to hemicellulose binding. CBM families 32, 35, 51, 61 and 67 were present in the time course, although no associations to catalytic domains were

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

found. This may be due to a freestanding cellulosome complexation or having associations with currently unknown catalytic domains with potential novel activities.

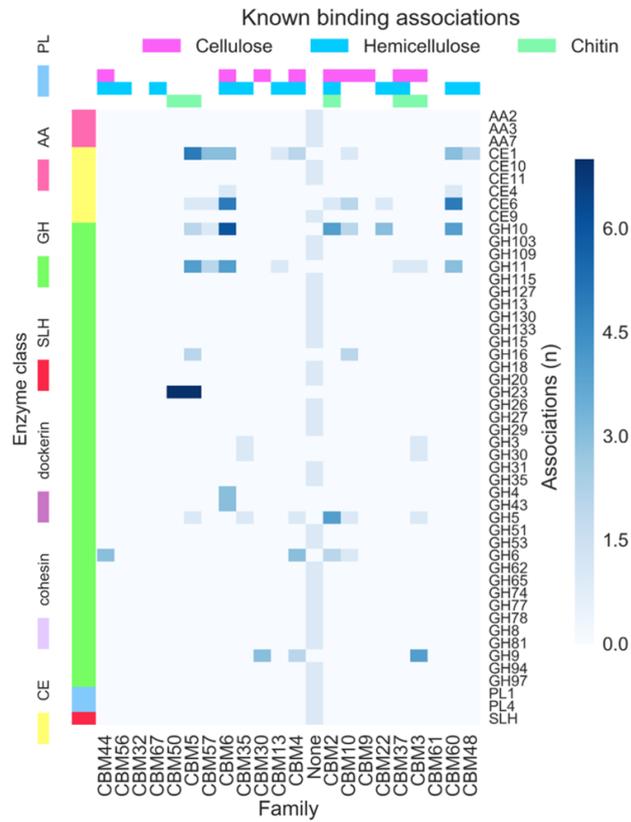


Figure 4.13 Association of observed carbohydrate binding domains (CBM) to catalytic domains. *GH*: Glycoside hydrolase, *CE*: Carbohydrate esterase, *AA*: Auxiliary activity, *PL*: Polysaccharide lyase. *SLH*: S-layer homology.

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

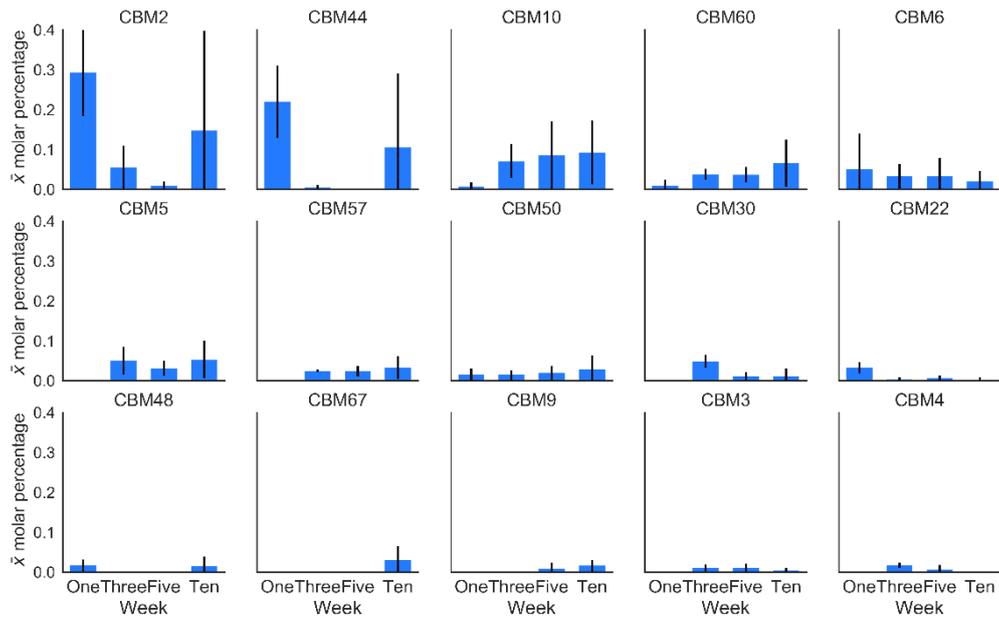


Figure 4.14 Temporal profile of the 15 most abundant CBM families. Displayed by the mean molar percentage in descending order from left to right. CBM; carbohydrate binding domain.

4.3.4 Phylogenetic and functional classification of the metasecretome

A holistic view of the metasecretome places prokaryotes (although only a minor contribution by archaea was observed) as the dominant protein secreting organisms in this salt marsh which were consistently responsible for 66.5-79.45% of the metasecretome and enriched over time (Figure 4.15). Archaea were expectedly inactive presenting only 0.04-0.2% of the total metasecretome. Negligible contributions were also made by Viruses. Eukaryotes were also responsible for 7.7-18.7%, decreasing temporally suggesting they may not have been active in carbon deconstruction. Of these eukaryotic proteins only 0.28-1.46% could be attributed to fungi, none of which were identified as lignocellulose associated, which was unexpected given the role ascribed to fungi as detrital degraders in most ecosystems. Almost all carbohydrate active proteins identified in this study were homologous to bacterial proteins, with exceptions of *Annelida* (known to salt marshes) and *Chlorophyta* (zooplankton).

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

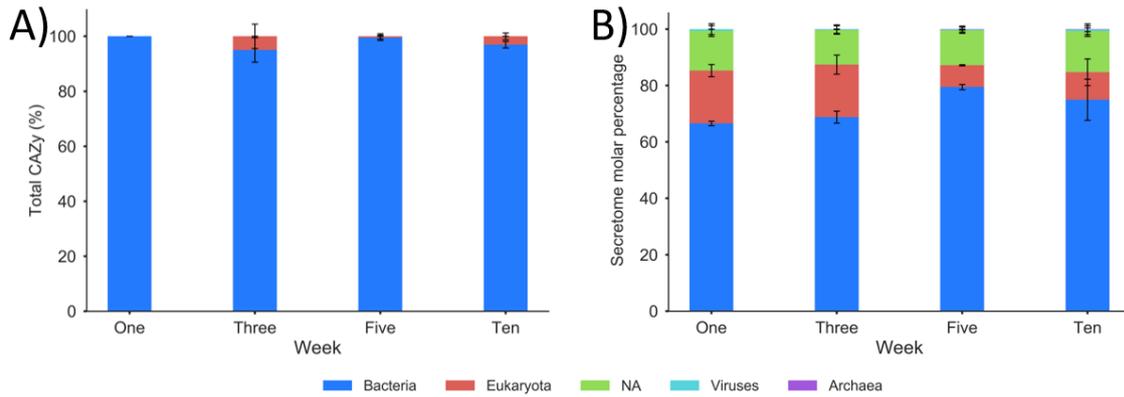


Figure 4.15 Metasecretome distribution at the superkingdom level. A) Distribution of the CAZy fraction only. B) Distribution of the entire metasecretome.

At phylum level resolution, the most likely taxonomic homology (BLAST best-hit) suggests *Proteobacteria* and *Bacteroidetes* were the dominant CAZy producing bacteria followed by *Verrucomicrobiae* (Figure 4.16). *Proteobacteria* exhibited a decline in enzyme production in week 5, suggesting the representatives in this phylum responsible for carbon cycling are not as tolerant as the *Bacteroidetes* to anoxic stress during this visibly anaerobic time point.

At lower taxonomic resolutions, the classes responsible for the majority of CAZy production are *Gammaproteobacteria* and *Deltaproteobacteria* (belonging to *Proteobacteria*) and *Bacteroidia*, *Cytophagia* and *Flavobacteriia* (belonging to *Bacteroidetes*) (Figure 4.17). Analysis of within phyla CAZy profiles elucidated intrinsically distinct profiles with greater degrees of concurrence than distinction, with the distinct fraction constituted by less abundant CAZy classes. (Figure 4.18). All of the top five most abundant CAZy families and CBMs were present concurrently among *Proteobacteria* and *Bacteroidetes*, with the exception of CBM44 (produced exclusively by the former).

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

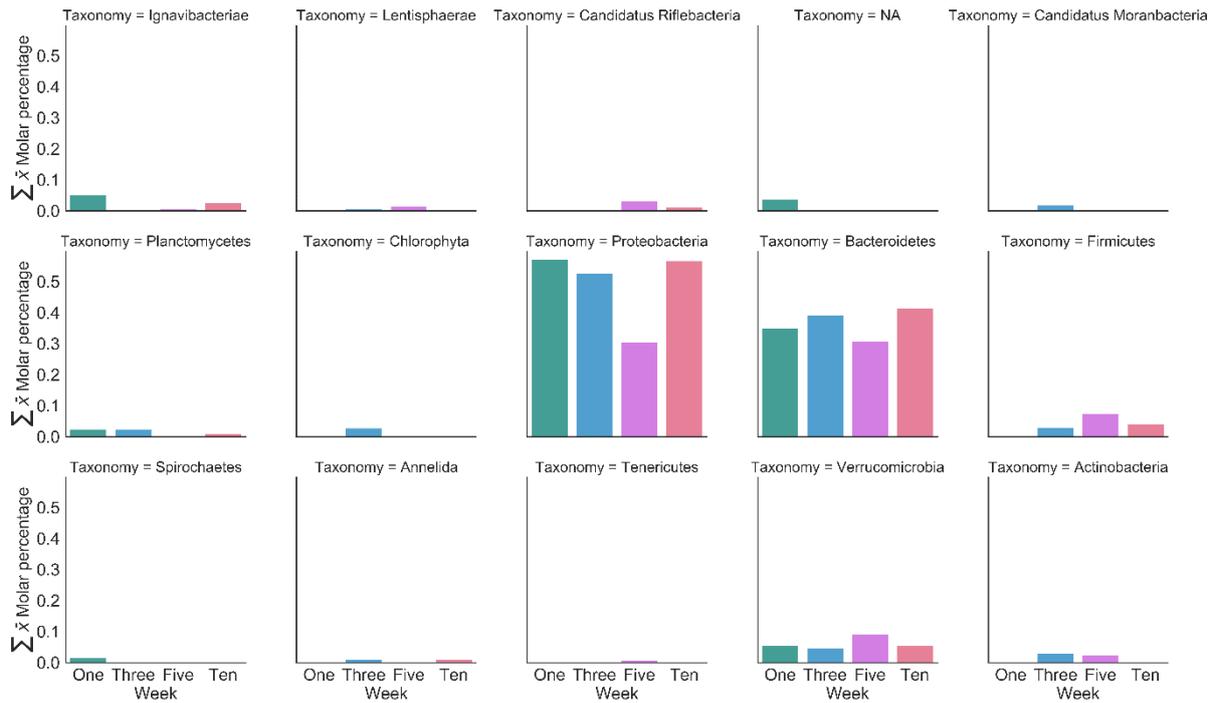


Figure 4.16 Most likely taxonomic origin of carbohydrate active enzymes at phylum level. No taxonomic filtering has been undertaken.

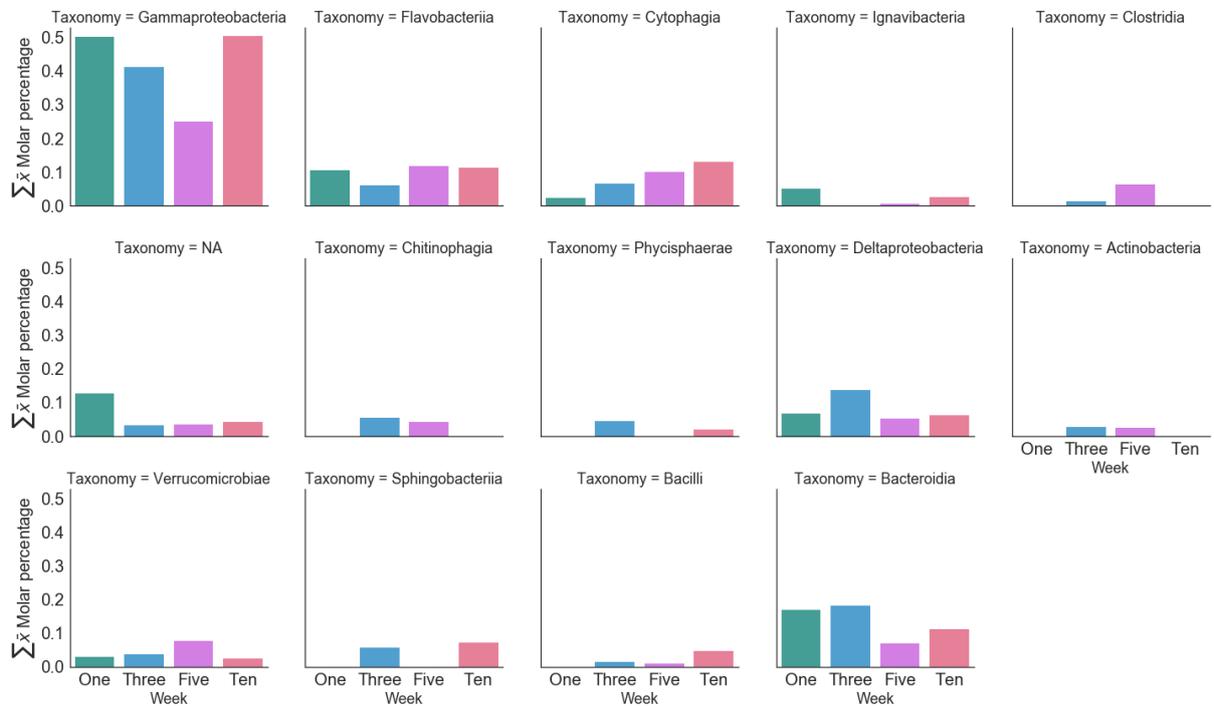


Figure 4.17 Most likely taxonomic origin at class level resolution of carbohydrate active enzymes (CAZy). Classes responsible for less than 1% of the total CAZy have been excluded.

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

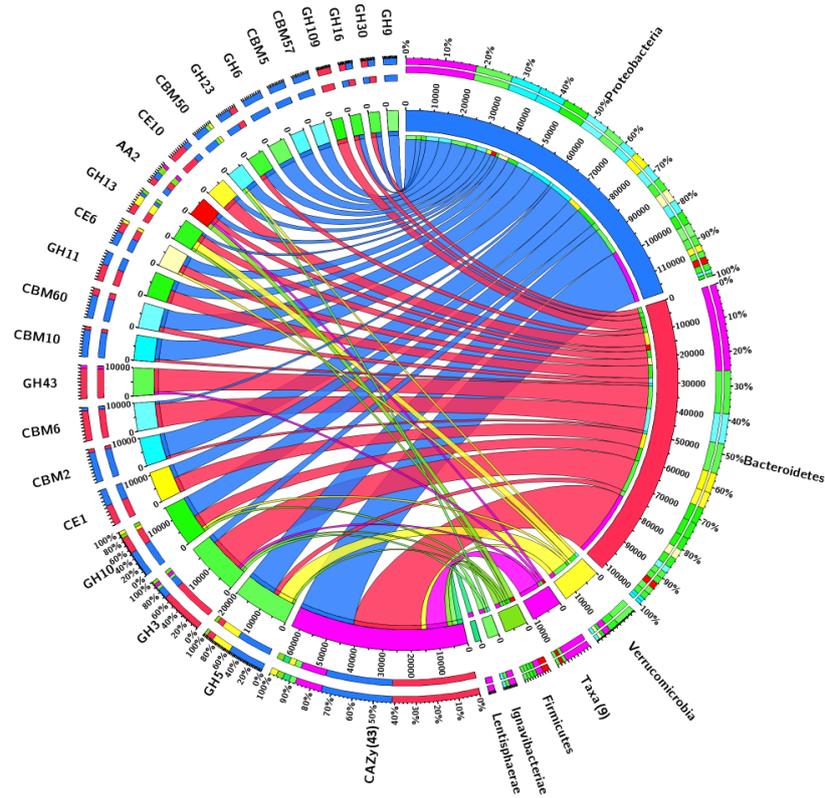


Figure 4.18 Phylogenetic distribution of CAZy ORFs at phylum resolution for all time points. Absolute quantification is 1×10^3 (1000 is equivalent to one ORF). Only phyla producing greater than 1% of the total CAZy are displayed. Only CAZy greater than 1% total CAZy are shown, filtered CAZy are grouped into 'CAZy'. Chord diagram created using the Circos software package (Krzywinski et al., 2009).

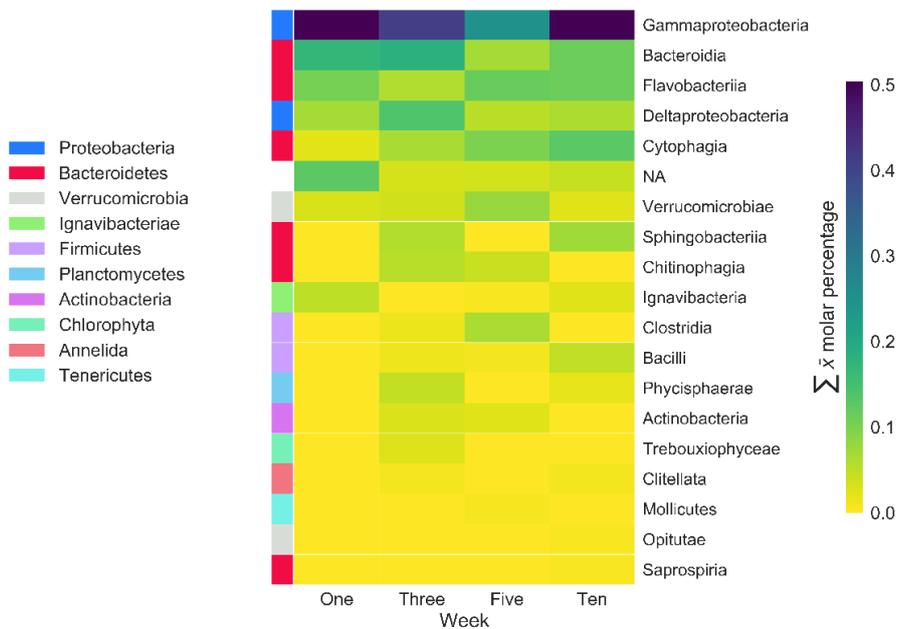


Figure 4.19 Phylogenetic distribution of CAZy ORFs at class level resolution. Respective phyla are indicated in the left column. Descending abundance by total temporal sum is displayed.

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

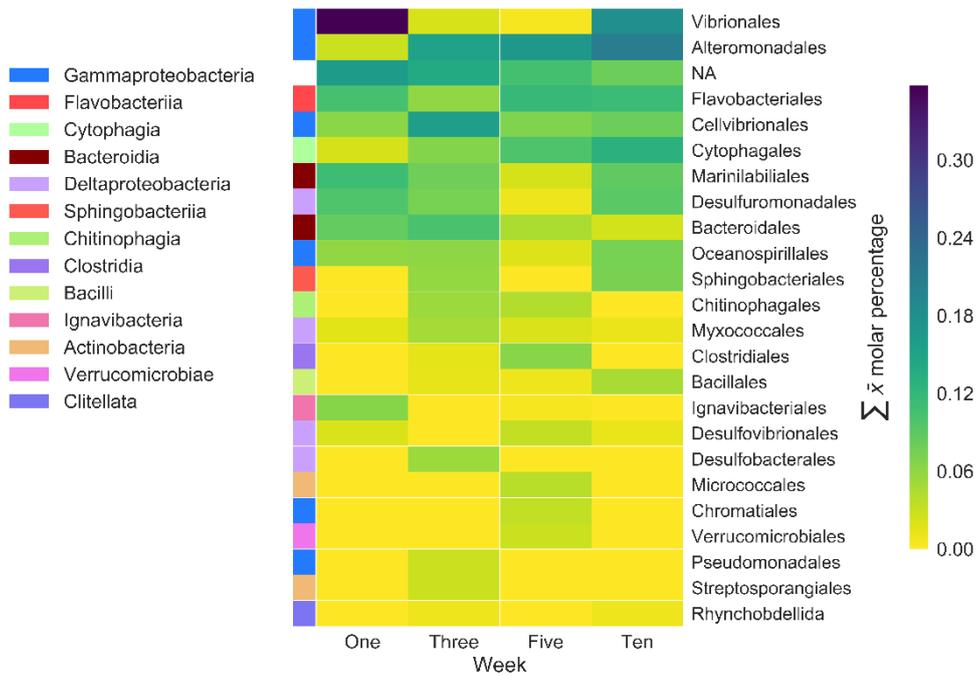


Figure 4.20 Phylogenetic distribution of CAZy ORFs at class-order level resolution. Descending abundance by total temporal sum is displayed. Orders responsible for producing less than 1% (total temporal sum) are filtered.

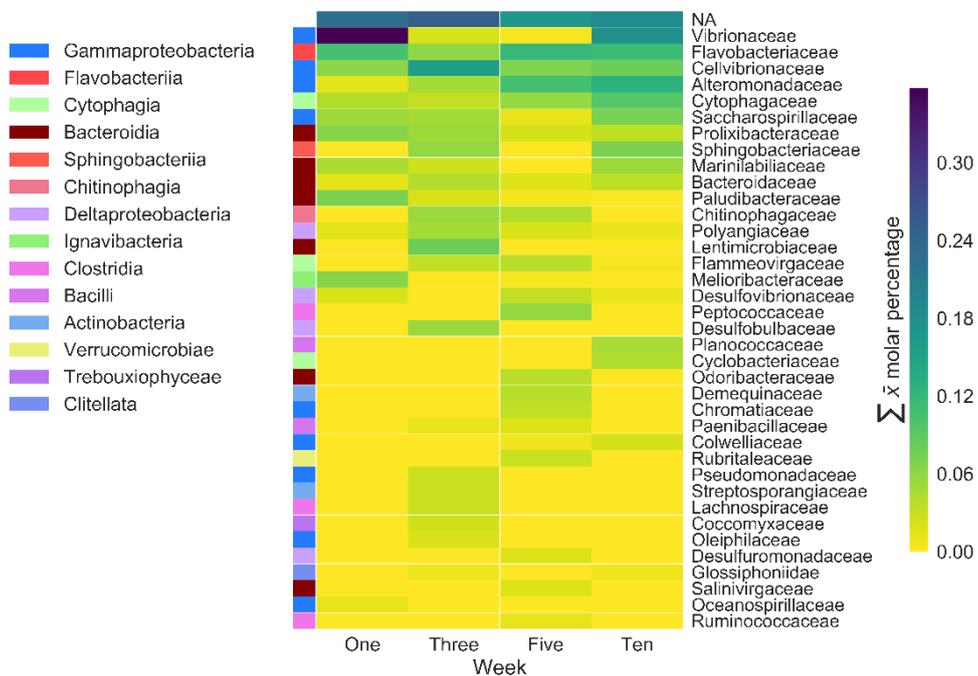


Figure 4.21 Phylogenetic distribution of CAZy ORFs at class-family level resolution. Respective classes are indicated in the left column. Descending abundance by total temporal sum is displayed. Families responsible for producing less than 1% (total temporal sum) are filtered.

Temporal analysis of taxonomically categorised CAZy and CBM profiles further demonstrates the existence of a core suite of enzymes and the dominance of *Proteobacteria* and *Bacteroidetes* (Figure 4.18). These profiles exhibit an abundance of noise in the form of scarce families present in few time points, likely due to poor depth in the proteomic databases. *Gammaproteobacteria* and *Bacteroidetes* had similar CAZy profiles with succinct differences in the less abundant fraction (Figure 4.18).

Due to the imperfect method of attributing a most likely taxonomic origin, greater taxonomic depth confers greater inaccuracy. In this study the lowest taxonomic resolution (for the most likely origin) is the class level which reveals an interesting temporal progression. Whilst *Gammaproteobacteria* remain dominant throughout the entire time series, *Bacteroidia*, *Cytophagia* and *Flavobacteriia* encroach significantly into this position progressively throughout the later time points (Figure 4.19-24). At family level resolution, the dominant CAZy producing taxa include (in descending order of abundance); *Vibrionaceae*, *Flavobacteriaceae*, *Cellvibrionaceae*, *Cytophagaceae*, *Saccharospirillaceae*, *Prolixibacteraceae*, *Sphingobacteriaceae*, *Marinilabilaceae* and *Bacteroidaceae* (Figure 4.21).

Gammaproteobacteria and *Bacteroidetes* displayed similar levels of CAZy diversity throughout the time series increasing over time while other taxa remained largely static. At the class level the diversity within *Bacteroidetes* were split between *Bacteroidia*, *Flavobacteria* and *Cytophagia*. Interestingly *Bacteroidia* displayed high levels of diversity in week one but gradually became less diverse. The inverse was true for *Cytophagia* whose diversity increased consistently. *Flavobacteria* were largely static. Whilst there were some temporal shifts in dominance between the *Bacteroidetes* classes, their attributed CAZy molar percentage remained largely unchanged.

Gammaproteobacteria exhibited the highest diversity throughout the entire time course, this also translated to productivity with *Gammaproteobacteria* producing more CAZy than any other taxa in all time points. As this level of diversity or abundance was not observed for any other taxa this suggests an important role of carbon cycling within salt marshes (Figure 4.22).

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

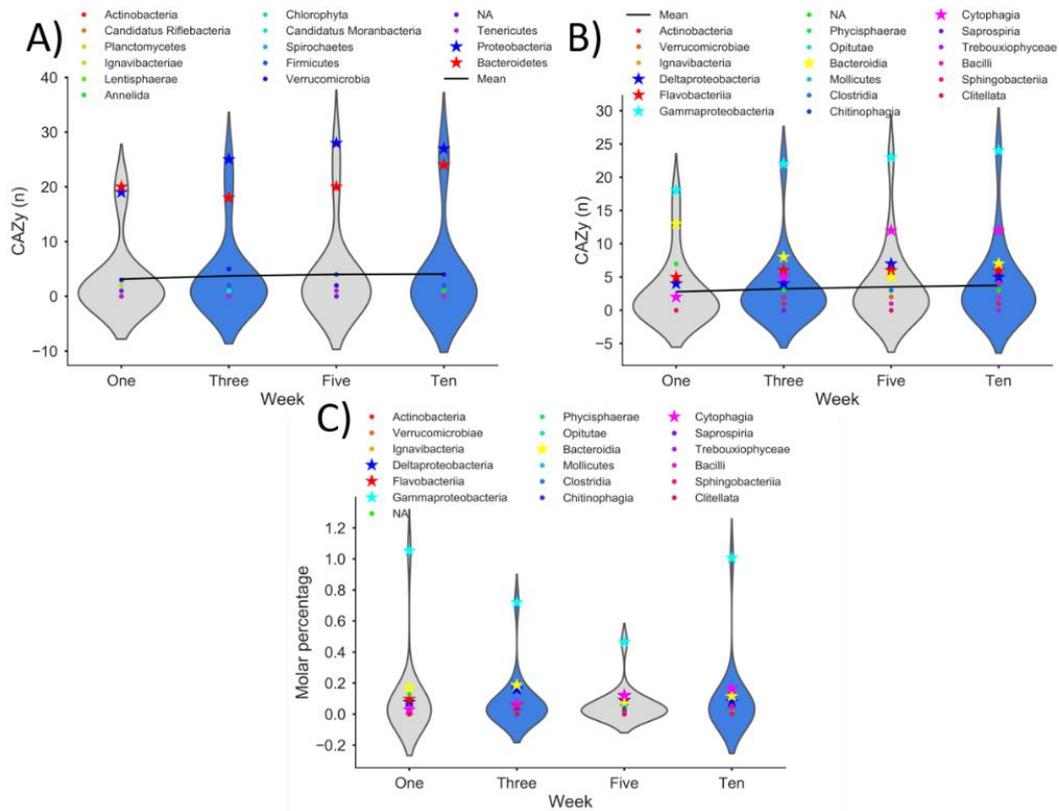


Figure 4.22 Taxonomic CAZy diversity of the metasecretome. A) Violinplot of the diversity of CAZy families within Phyla. B) Violinplot of the diversity of CAZy families within classes. C) Class level origin of the CAZy fraction. Violin widths are proportional to the relative density of independent variables (taxonomic groups).

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

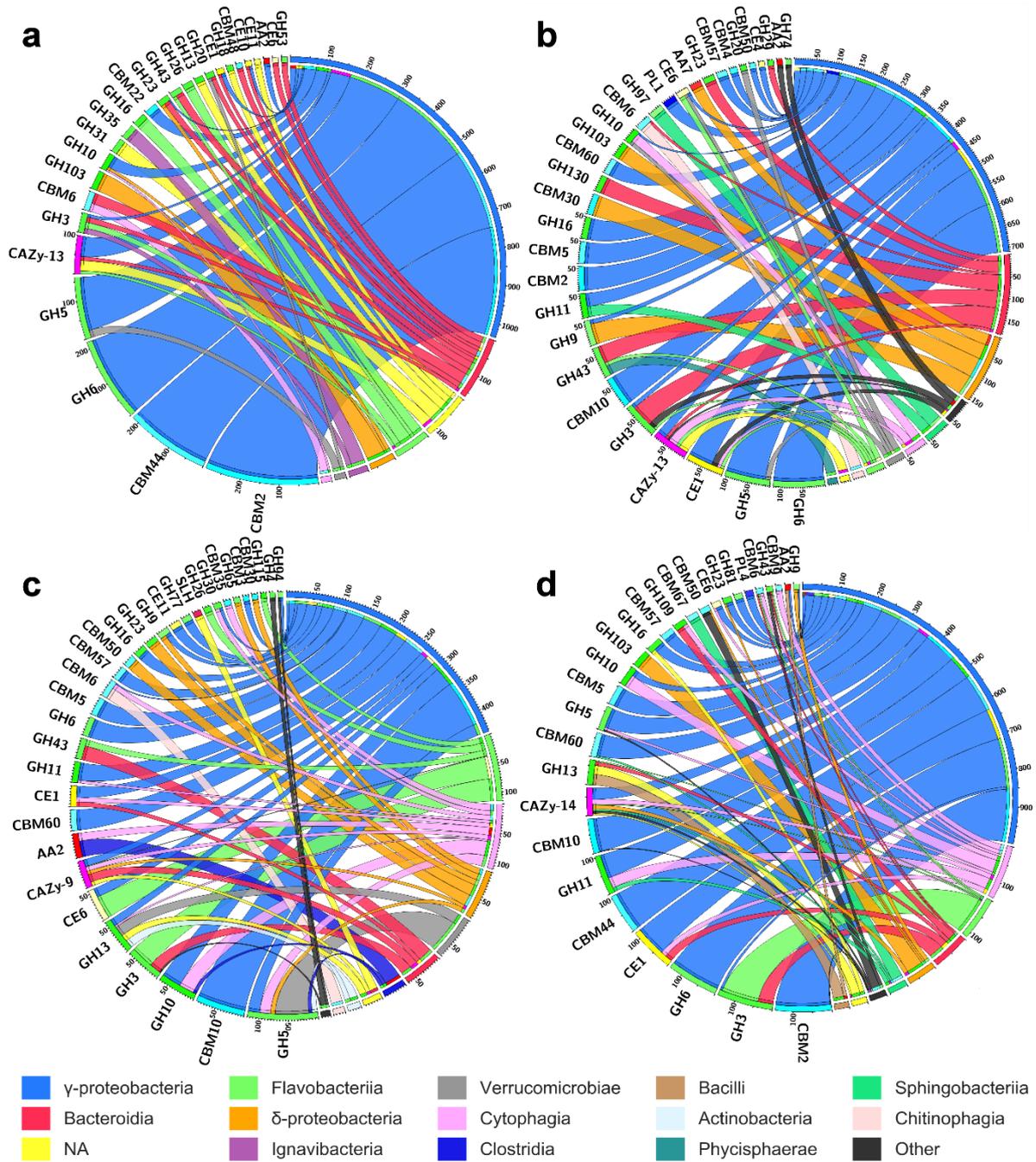


Figure 4.23 Phylogenetic distribution of CAZy classes at class resolution. a) week one. b) week three. c) week five. d) week ten. Absolute values of segments are the sum molar percentage $1e^3$ (1000 = 1%). Ribbon size is relative to the contribution to or of the taxa/CAZy. Segment size is intraspecifically relative only to taxa or only to CAZy and not interspecifically comparable. Only phyla producing greater than 1% of the total CAZy are displayed, filtered taxa are grouped into 'Taxa (n)'. Only CAZy greater than 0.85% total CAZy are shown, filtered CAZy are grouped into 'CAZy (n)'. Chord diagrams were generated using the Circos software package (Krzywinski et al., 2009).

Proteomic analysis of salt marsh derived lignocellulose associated exo-metaproteome

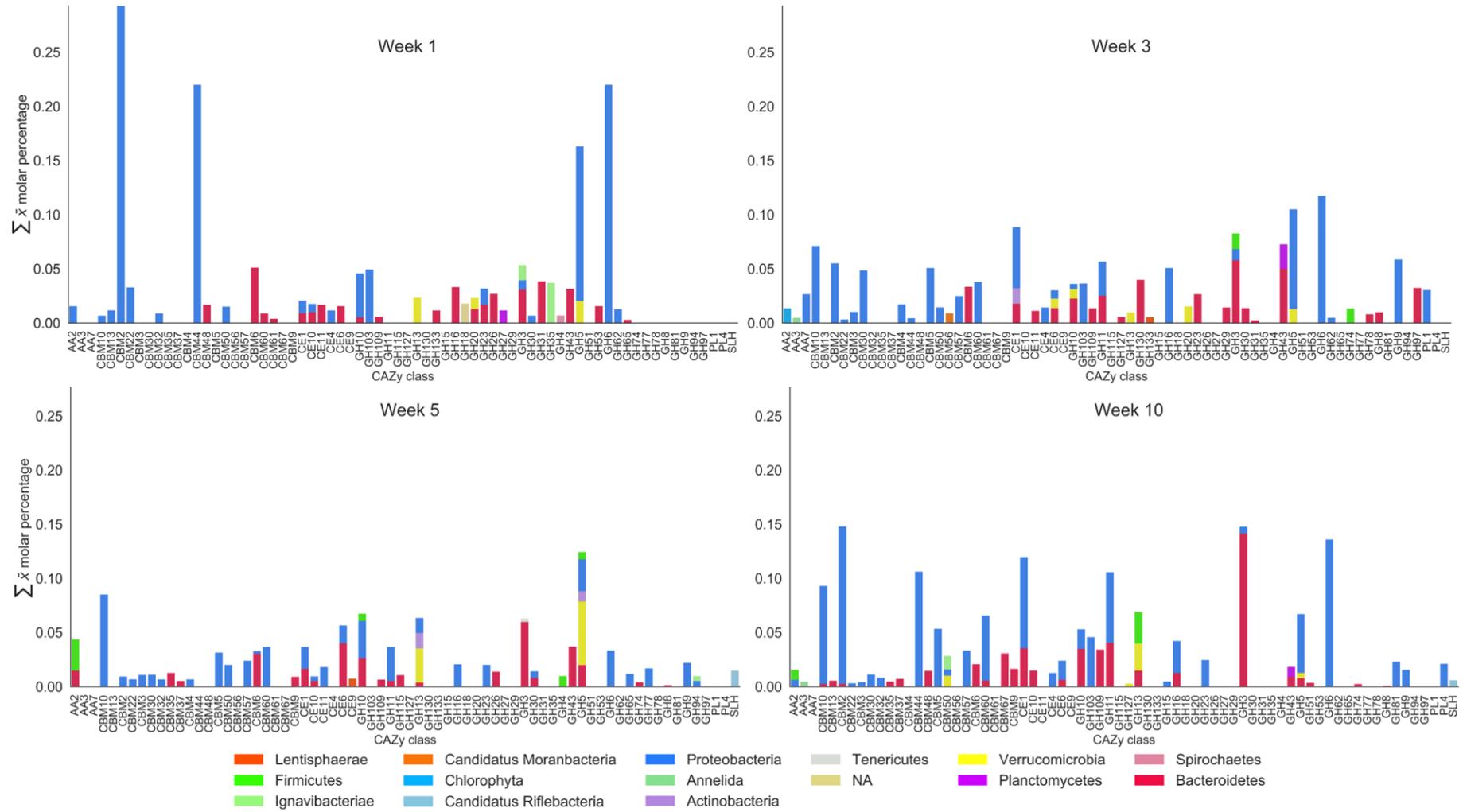


Figure 4.24 Sum of mean molar percentage CAZy profile by most likely taxonomic origin by time point.

4.4 Discussion

The data presented in this Chapter represents the first in depth investigation into natural carbon cycling within a salt marsh environment. The time points chosen during the time series (Chapter 3; weeks one, three, five and ten) provided a temporally interspersed overview of biotic lignocellulose deconstruction. The process of lignocellulose deconstruction in native conditions was examined at the extracellular protein level by assigning putative function to proteins and probing the phylogeny to elucidate function.

The relatively low proportion of the proteome represented by CAZy members reflects the complex nature of salt marsh environments and the diverse communities within. With the non-CAZy proteome representing around 99% (Figure 4.4) of the proteome this suggests many major processes are occurring. The nature of the labelling technique utilised in this study may also underrepresent the CAZy fraction by detecting degraded, partial or inactive proteins which are the result of natural lysis and environmental conditions. Additionally, many halotolerant proteins are tolerant to denaturing agents, and the desorption method in this study may have been too gentle to desorb bound proteins from the biomass (Sinha and Khare, 2014). It has also been shown by Alessi et al., (2017) that the soluble and insoluble fractions are significantly different, the soluble (water column suspended) fraction was not assessed in this study, and as only 29% of ORFs contained CBMs it is possible many proteins are suspended, despite this being energetically incongruous.

It is also true that CAZy homology in this unexplored environment has yet to be established. To date 'omics databases for salt marsh environments are lacking. If these novel sequences have yet to have homologs identified, then the homology-based approach in this study would have overlooked these sequences by omitting them from the analyses. These factors may have attributed to the low representation of CAZy, although no other *in situ* studies have been conducted, *in vitro* enrichment studies (not directly comparable) have observed higher total CAZy representations (5-15%) (Ma et al., 2016, Jimenez et al., 2015b) .

4.4.1 Functional assignment of the metasecretome

The greatest shift in CAZy profile occurs between week one and three. The relative proportions of CAZy classes stabilises from week three to ten after an initial reduction in glycoside hydrolase families (GH), substituted by an increase in carbohydrate esterases (CE) and auxiliary activities (AA).

This observation is concordant with the community succession analysis in Chapter 5. Whilst the dominance of GH families is a common feature of CAZy profiles (Zhu et al., 2016, Wong et al., 2017).

It is apparent that cellulose is the target substrate within the most abundant GH families (GH3, GH5 and GH6). GH6 and GH3 predominantly contain endoglucanases, cellobiohydrolases and β -glucosidases. The GH5 family is diverse and members of this family may be active on the majority of lignocellulose polysaccharides. The identified GH5 members in this study were predominantly associated with cellulose targeting carbohydrate binding modules (CBM10, CBM4, CBM2), but also chitin (CBM5, CBM3, CBM2) and hemicellulose (CBM35, CBM3, CBM2). These three dominant families contain all required catalytic activities to completely hydrolyse cellulose. The focus on cellulose deconstruction likely arises from the favourable energetics of glucose metabolism compared to xylose (lignocelluloses second most abundant monomeric constituent) where net ATP yield is significantly greater for glucose (Liu et al., 2012, Temudo et al., 2009).

Hemicellulose targeting families were also highly abundant, such as GH10, GH11 and GH16 all of which putatively require xylan or associated xylans as a substrate. These are the most abundant hemicellulose polysaccharides in grasses, the depolymerisation of which is required to access the cellulose. The abundance of these groups in later time points relative to week one is likely due to preferential binding to an artificially greater number of exposed cellulose microfibers from mechanical fractionation of the biomass. Interestingly the number of CAZy families across all time points remained similar, suggesting the 'noisy' families were likely transient or below detectable limits.

The low abundance of auxiliary activity (AA) families (with the exception of AA2 in week five) established largely as oxidative enzymes targeting lignin and cellulose coincides with the oxygen deficiency within the environment from which they were sampled; inhibiting their function and ultimately their production. There is increasing evidence that anaerobic communities operate devoid of auxiliary redox activities with many studies reporting their complete absence (Svartstrom et al., 2017, Al-Masaudi et al., 2017, Wong et al., 2017). Two of these anaerobic studies also reported high relative abundances of carbohydrate esterase family 1 (CE1) genes but not of the order observed in this study (highlighting weaknesses in metagenomic approaches) (Wong et al., 2017, Al-Masaudi et al., 2017).

Interestingly the fourth most abundant enzyme family in this study was CE1, which contains esterases with the capability to catalyse the removal of O- or N-acetyl decorations from polysaccharides (Nakamura et al., 2017). The CE1 family was most strongly associated with a diversity of CBMs and the strongest association was with chitin binding CBM5 (weak cellulose

binding in this family has been observed (Hogg et al., 2003)). Currently no chitin degrading CE1 family enzymes have been described and remain targets of interest for expression and characterisation studies. Chitin is a polymer of N-acetylglucosamine and is highly abundant in marine environments, therefore it is fitting that the enzymes identified in the present study act upon this substrate.

Other associations of the CE1 family included CBM families 4, 6, 10, 13 and 60 which are all associated with cellulose and hemicelluloses. Activities within the CE1 family can deacetylate xylan decorations and non-oxidatively disrupt the lignin-carbohydrate complexes (LCC) at the interface between lignin and cell wall polysaccharides via hydrolysis. LCC linkages are thought to consist mainly of aryl ester (furulic linkages from furulic acid to arabinose in grasses) and aryl ether bonds; hydrolysis of which would depolymerise the lignin, exposing the surface of the remaining hemicellulose/cellulose construct (Buanafina, 2009). The disengaged lignin would resuspend and redistribute within the upper sediments of the salt marsh where it would be subject to modification by function of oxidative attack by AA families (e.g. AA2), allowing the microbial community access to high value polysaccharides without the expenditure of fully deconstructing the recalcitrant lignin fraction.

It is also possible that oxygen-independent mechanisms of lignin disruption exist that have yet to be characterised. To address this knowledge gap, the data generated here can be analysed with respect to highly abundant proteins of unknown function. In the present study this direction was not explored but remains an interesting avenue for future research.

It has been reported that with increasing sediment depth carbohydrate content decreases whilst aromatic compounds increase, which can be explained by the gradual burial of freshly fallen material into oxygen depleted sediments where oxidative attack is inhibited by decreasing redox potential with depth (Zhang et al., 2014, Zhang et al., 2016b, Santin et al., 2008, Burdige, 2007). This suggests the fragmented lignin is not fully catabolised in this ecosystem and continues to persist in the sediment contributing to carbon sequestration. It is therefore possible the oxidative activities of AA families have been substituted for enzymes homologous to the CE1 family to fragment rather than deconstruct lignin.

Additionally, lignin is predominantly composed of β -O-4 aryl ether linkages, which can be cleaved by β -etherases (Marinovic et al., 2018). The dbCAN database does not incorporate β -etherase models. Currently only three β -etherases of prokaryotic origin have been described (Picart et al., 2014, Masai et al., 1989, Masai et al., 1991). Sequence logos can be constructed and used as a

template for Hidden Markov Model searching against the databases in this study and remains an avenue for future research.

The novelty is these databases arise from the potential for halotolerance conferring enzymes originating from the salt marsh environment. Beyond this there is scope to probe for novel activities by cross referencing domains of high and low confidence. As above for CE1, is it also associated with CBM57 which has no known binding function, exploring these identified sequences could yield interesting results. CBM57 is most strongly associated with GH11s (endo- β -1,4 and endo- β -1,3-xylanases) and as such xylan (hemicellulose) binding (Figure 4.11). There were numerous ORFs (9.9% of all ORFs) containing single CBM domains (CBM families 9, 32, 56, 61, 67), if the catalytic domain in these sequences has escaped annotation in these databases they may be truly novel (both in terms of sequence and function).

4.4.2 Functional phylogeny of the lignocellulose associated salt marsh microbiome

The phylogenetic distributions of CAZy classes across all time points suggest *Proteobacteria* are both first responders to lignocellulosic biomass in salt marshes and remain temporally dominant extracellular CAZy producers. Organisms belonging to *Bacteroidetes* were the second most dominant CAZy producing taxa. These observations are concordant with salt marsh microbial profiles described previously and the profiles observed in this study; however, no other similar functional phylogenetic study has been undertaken (Bowen et al., 2012, Lasher et al., 2009, Engel et al., 2017). These phyla produced a tremendous diversity of CAZy that was not observed in any other taxa in this study. Other proteomic studies have observed a dominance of *Proteobacteria* and *Bacteroidetes* (Schneider et al., 2012, Alessi et al., 2017, Jimenez et al., 2015b). Metagenomic studies (whilst not directly comparable to proteomic studies) have also implicated *Proteobacteria* and *Bacteroidetes* in lignocellulose deconstruction (Zhou et al., 2017, Jimenez et al., 2015a). Enrichment studies of salt marsh sediment on lignocellulose have revealed *Gammaproteobacteria* and *Bacteroidetes* to be dominant, which is synonymous with the *in situ* observations in this study (Cortes-Tolalpa et al., 2018).

The capability of representatives of *Proteobacteria* to degrade plant cell wall polysaccharides in the marine environment is known, however their contribution to carbon cycling has never been established or studied (Weiner et al., 2008). Despite *Proteobacteria* being copiotrophic, which explains their high relative abundance in detritus rich salt marsh sediments, the major contribution

of *Proteobacteria* to lignocellulose degradation in this study is unexpected (Zhang et al., 2016a). *Gammaproteobacteria* in this study were the most abundant CAZy producing class. Interestingly enrichment studies on lignin in seawater resulted in a microbiome dominated by *Gammaproteobacteria* and absence of fungi (Woo and Hazen, 2018).

Bacteroidetes have been implicated as key in the marine carbon cycle with putative adhesive capabilities (which likely aided their establishment within the biomass) and contain the largest relative genomic arsenal of polymer targeting enzymes of all bacterial phyla (Bauer et al., 2006, Huang et al., 2017). Despite this their precise molecular mechanisms of action have not been captured until now. Sedimentary dwelling bacteria (particularly *Bacteroidetes*) originating from salt marsh environments are poorly characterised in terms of their ecosystem function, despite being implicated repeatedly in carbon cycling.

The fate of lignocellulose within salt marshes with stable isotope probing has shown that most decomposing carbon is utilised by bacteria (predominantly *Gammaproteobacteria* and *Bacteroidetes*, however this provides no insight as to their role in degradation (Darjany et al., 2014). The data in this study provides evidence that not only are bacteria the major benefactors of lignocellulosic breakdown products, but also orchestrate its deconstruction. Terrestrially fungi are thought to drive lignocellulose deconstruction, and proteomic evidence suggests they are the dominant producers of lignocellulolytic enzymes (de Boer et al., 2005, Schneider et al., 2012). The results obtained by Darjany et al., (2014) in a Californian salt marsh and this study (UK salt marsh) emphasise the importance of *Gammaproteobacteria* and *Bacteroidetes* to carbon cycling in independent salt marshes. Priming studies (the effect of temporary lignocellulose enrichment on the degradation of previously sequestered carbon) have also implicated *Gammaproteobacteria* and *Bacteroidetes* in primed marine sediments (Gontikaki et al., 2015).

Salt marshes contain legacy microbiomes of terrestrial and marine origin and whilst marine environments are thought to be unfavourable for fungi, salt marshes are sheltered and less extreme than open oceans (Woo and Hazen, 2018). This challenges current theory that cross-kingdom interactions modulate carbon cycling processes with bacteria either facilitating fungi or interacting commensally in litter microbiomes (Purahong et al., 2016, Urbanova et al., 2015). The lack of fungi-originating CAZy in this data would suggest the inverse is true and fungi are commensal in salt marsh environments (if they benefit from soluble degradation products), however, the extent of their interactions has yet to be fully elucidated.

Fungi from mangrove ecosystems have been shown to be capable of thriving in saline environments and produce lignocellulolytic enzymes (Arfi et al., 2013). In this study fungi are only

responsible for 0.28% (week five) to 1.46% (week one) molar percentage of the metasecretome, none of which are identified as CAZy. An enrichment of salt marsh sediment also reported an absence of fungi, which the author attributed to nitrogen depletion (Cortes-Tolalpa et al., 2018).

The majority of studies have focussed on representatives of the kingdom fungi, as sources of lignocellulolytic enzymes for biotechnological applications, likely encouraged by the fact industrially derived cocktails are developed from *Trichoderma reesei* (filtered culture supernatant) which is capable of producing up to 100g/L cellulases (Cherry and Fidantsef, 2003, Brijwani et al., 2010). In this study fungal-derived lignocellulose degrading proteins were absent. Whilst eukaryotic derived proteins were identified (Annelida; AA2 and Chlorophyta; AA3), the majority of the carbohydrate active (CAZy) fraction of the metasecretome was of prokaryotic origin (100% in week one to 95% in week three). More recent studies are engaging the role of prokaryotes which also exhibit a diverse suite of lignocellulolytic enzymes, with genomes consisting of up to 8% CAZy for representatives of the *Bacteroidetes* phylum (Huang et al., 2017, Cragg et al., 2015). However fungal strains have been optimised to produce lignocellulose degrading enzymes and the use of prokaryotic equivalents (such as *Saccharophagus degradans 2-40(T)*) requires intensive optimisation to be competitive (Jung et al., 2014).

Globally, salt marshes are in a state of decline. Historically salt marshes were lost through anthropogenic action (reclamation of land). Currently the loss of salt marshes in the UK have been mainly attributed to coastal squeeze although other mechanisms have been hypothesised (Hughes and Paramor, 2004). Coastal squeeze is when the rate of sea level rise exceeds the accretion rates in the salt marsh (by deposition of organic matter or sediment). This 'squeezes' the salt marsh against the coast, where it is eventually lost to the sea.

These ecosystems perform valuable services including nursery habitats, pollution and carbon sequestration, erosion barriers and coastal flood defences (Barbier et al., 2011). There is now a large intangible value (\$24.8 trillion (2011)) placed on the global services provided by salt marshes (de Groot et al., 2012, Costanza et al., 2014). Therefore, the knowledge that *Gammaproteobacteria* and *Bacteroidetes* have a significant role in carbon cycling *in situ* may indicate a potential role as indicator taxa in assessing salt marsh susceptibility to coastal squeeze. Additionally, this data shows that salt marshes also contain a plethora of biotechnologically valuable organisms and enzymes that have yet to be exploited or valued. This further bolsters the defensive case for salt marsh protection.

4.4.3 The identification of novel enzymes from the salt marsh metasecretome

To exploit the enzymes identified in this study, CAZy families were selected based upon their abundance at the ecosystem level to replicate the natural cocktail observed in this study. The most abundant groups in the ecosystem level profile were cellulose targeting, and these groups were selected (GH3, GH5 and GH6). The ecosystem level CBM profile was also considered when assessing individual ORFs. Additional auxiliary groups were also targeted including CE1, AA2, GH10, GH11, GH16 and other groups discerned only from the transcriptome not mentioned here (see Chapter 6). Only ORFs containing signal peptides were chosen as targets for further study (however this relies on sequence homology and sequence accuracy).

Given the complex architectures of some ORFs (where multiple CBM and multiple catalytic domains may be present), ORFs with and without CBMs for each family were selected. These subsets of sequences were further disintegrated by creating groups of high (>75%) and low (<50%) confidence based on sequence identity. This created a group of enzymes with a high likelihood of having a known function, and a divergent group of enzymes with potentially novel functions. It was expected all enzymes would have some degree of halotolerance to support their novelty. The result of these further studies is explained in Chapter 6.

Chapter 5 Decomposition associated microbial community profile

5.1 Introduction

Salt marshes exist as transitional ecosystems between terrestrial and marine environments and putatively contain legacy microbiomes of each and an additional salt marsh adapted microbiome. The complex nature of the environment confers fluctuations in abiotic and biotic factors (pH, temperature, salinity, inundation frequency and longevity, pCO₂, dissolved oxygen) at various heterogeneous spatiotemporal scales (Baumann et al., 2015). Therefore, this ecosystem convergence and perpetual inconsistency of deterministic factors is thought to confer great levels of microbial diversity.

Despite being regarded as highly diverse ecosystems, salt marshes are largely unexplored with few studies reporting in depth characterisations of the microbial communities. Salt marsh microbiomes contain bacterial, archaeal and fungal constituents (Bowen et al., 2012, Weber et al., 2011, Seyler et al., 2014). Evidence suggests bacteria are dominant in this ecosystem; abundant at levels that are orders of magnitude greater than archaea, with fungi orders of magnitude less abundant than archaea (Lee et al., 2017, Chaudhary et al., 2018). Microbiome knowledge is important in facilitating potential assignment of function to taxonomic groups; for the few taxonomic groups with known and accepted functions. The relative size of these taxonomic groups allows extrapolation of the relative contribution of these functions to ecosystem processes, which may be compared with experimental data. Consequently, this may provide information about ecosystem functioning and status derived from a biological phylogenetic catalogue.

Cultivation techniques have historically been used to identify microbial constituents; however, the cultivable fraction of microorganisms is thought to represent only 1%, with many microbes putatively obligate symbionts (which likely form a complex network of dependence), isolation and cultivation of which is technically challenging (Buchan et al., 2003). Furthermore, these techniques are dependent upon isolation media, growth conditions and sampling techniques. These studies are empirical as it elucidates taxonomic characteristics, often at species level resolutions. Therefore, these techniques, whilst often successful, incredibly informative and imperative to the advancement of functional taxonomic understanding, only captures a fraction of species present. For this reason, a more appropriate approach is to identify organisms using DNA sequencing.

Decomposition associated microbial community profile

Genomic DNA sequences extracted from a microbiome contains the taxonomic information of the source organisms. Hypervariable regions (often located between conserved regions) confer great sequence heterogeneity, the magnitude of which is considered to positively correlate with evolutionary distance between taxa. Whilst caveats exist, the hypervariable sequences can be assigned to sequences originating from known taxa in curated archive databases by similarity. Regions for amplification and subsequent sequencing include the ribosomal 16S rRNA gene (prokaryotes; bacteria and archaea), ribosomal 18S rRNA gene (eukaryotes; universal) and internal transcribed spacers (ITS; eukaryotes and specifically fungi). The amplicon pool generated from the microbiome genomic DNA can then be clustered by similarity and the taxonomic information for these clusters allocated. The size of the clusters, and the taxonomic assignment facilitates the construction of a taxonomic profile presented in relative terms.

Delineating taxonomic profiles allows the assignment of putative functions, which provides insight into the state of ecosystem processes at the time of sampling. However, as salt marshes are underexplored, many of the taxonomic groups are likely unknown, or have no attributable ecological function. Identifiable organisms within salt marshes are also likely far diverged from their well-studied, terrestrial counterparts (or equally understudied marine counterparts); therefore, ascertaining and ascribing function should be undertaken cautiously.

This task is challenging and is further exacerbated by the vast taxonomic and functional complexity within the salt marsh ecosystem. In depth functional characterisations of taxa are required, that can be cross referenced with sequence data to inform decisions about assigning ecological function to taxa; however, conclusive functional data in this ecosystem is severely limited. In this Chapter, community profiling is undertaken to identify the integral organisms in lignocellulose deconstruction. Bacteria, archaea and fungi are identified from the genomic DNA extracted in Chapter 3 by amplifying the hypervariable regions of the conserved ribosomal 16S, 18S and internal transcribed spacer regions.

Interestingly, fungi and archaea were not identified as CAZy producers in Chapter 4; 99-100% of the carbohydrate active enzyme fraction of the metasecretome identified in Chapter 4 was of bacterial origin; therefore, this kingdom is the primary focus of this Chapter. Notable, yet minor exceptions were enzymes homologous to an *Annelida* and *Chlorophyta* protein. These bacterial enzymes spanned diverse taxonomic lineages but were predominantly of *Proteobacterial* and *Bacteroidetes* origin.

Gammaproteobacteria were identified as the major CAZy producing taxonomic group, particularly the families *Vibrionaceae*, *Cellvibrionaceae* and *Alteromonadaceae*. This class was

followed closely by *Bacteroidetes*, which displayed greater a greater taxonomic diversity of CAZy producing lineages; *Flavobacteriia* (family; *Flavobacteriaceae*), *Cytophagia* (family; *Cytophagaceae*), *Bacteroidia* (families; *Prolixibacteraceae*, *Marinilabiaceae*, *Bacteroidaceae*, *Paludibacteraceae* and *Lentimicrobiaceae*) and *Sphingobacteriia* (family *Sphingobacteriaceae*). *Deltaproteobacteria* were also active producers (families; *Polyangiaceae*, *Desulfovibrionaceae*, *Desulfobulbaceae* and *Desulfomonadaceae*). Minor contributions were also made by the phyla *Firmicutes* (families; *Peptococcaceae*, *Planococcaceae* and *Paenibacillaceae*), *Verrucomicrobiae* (order *Verrucomicrobiae*) and *Ignavibacteriae* (order *Ignavibacteriales*). Many of these groups have been previously implicated as carbon cyclers (see section 1.7); however, direct evidence (such as metasecretome homology described in Chapter 4) for this has yet to be provided.

The aim of this Chapter is to identify and characterise the changes in community structure of the biomass associated microbiome during the time series of *Spartina anglica* decomposition in the natural salt marsh using amplicon sequencing and reference datasets. An additional aim was to cross reference the community profile generated in this Chapter with the proteomic taxonomic profile generated independently in Chapter 4, which, upon cross referencing, facilitates the identification of lignocellulolytic enzyme producing lineages. This will inform the assignment of function and productivity to these lignocellulose-associated taxa, thus identifying the organisms directly involved in salt marsh lignocellulosic carbon cycling. This is necessary as currently no available lignocellulose active enzyme databases exists for salt marshes.

5.2 Materials and methods

5.2.1 DNA Sequencing

5.2.1.1 Genomic DNA preparation and RNA depletion

Total nucleic acids (genomic DNA (gDNA) and total RNA (tRNA)) obtained in Chapter 3 were diluted 1:2 with nuclease free H₂O. Reactions were conducted in triplicate for each of the 46 time points. 100µL of diluted nucleic acids were incubated at room temperature for 5 min with 0.2µL RNase A (Qiagen 1006671, 100mg/mL). tRNA depletion was assessed with agarose gel electrophoresis. The resulting genomic DNA was purified using a Genomic DNA Clean & Concentrator™ Kit (ZymoResearch D4065) according to the manufacturer's protocol. gDNA was quantified using a nanodrop spectrophotometer (Thermo), technical replicates were pooled and diluted to stock concentrations of 100ng/µL with H₂O as above and stored at -20°C.

Decomposition associated microbial community profile

5.2.1.2 Target gene selection and primer selection and development

It was *a priori* determined that sequencing was to be conducted on an Illumina MiSeq with a 2 x 250 base pair cartridge. This limited primer selection to a region of less than 500 base pairs (inclusive of primers and index tags, to ensure high confidence in overlapping regions of forward and reverse reads. Primers targeting bacterial and archaeal ribosomal 16S rRNA genes, eukaryotic ribosomal 18S rRNA genes and fungal internal transcribed spacer units (ITS1 and ITS2) were selected from the literature (Table 5.1).

For fungi, the internal transcribed spacer units (both region 1 and 2) are considered hypervariable universal DNA barcodes for fungi and were selected as markers to assess the fungal community in this study (Schoch et al., 2012). As these regions are significantly variable in size between taxa (700-5000 base pairs) it was unlikely overlapping regions could be established for all sequences. Consequently, only the forward read were to be analysed, and primers were selected solely on taxonomic coverage. In total 8 forward primers and 5 reverse primers were selected for testing (Figure 5.1). As there is little difference between taxonomic identification capabilities of ITS1 and ITS2 (Wang et al., 2015c, Balaalid et al., 2013), ITS2 was selected due to novel degenerate primer availability. ITS4ngs was selected to compliment fITS7, which is degenerate in one position and exhibits enhanced fungal specificity and increased eukaryote exclusion (Tedersoo et al., 2014, Ihrmark et al., 2012).

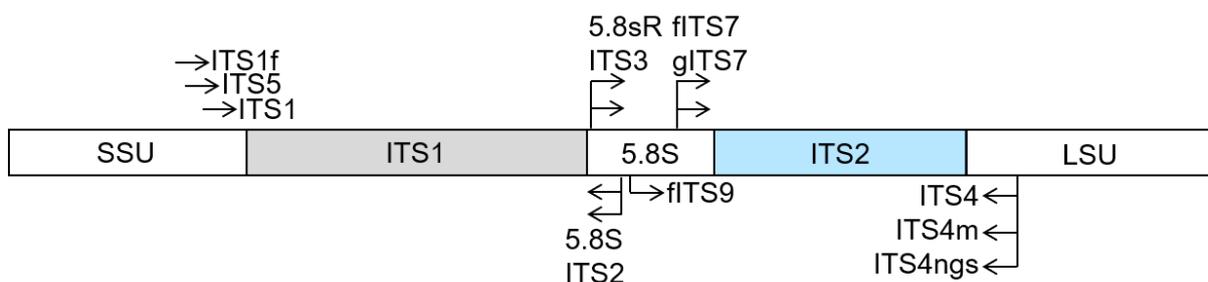


Figure 5.1 General ribosomal domain architecture. Highly cited primers are displayed and arrows indicate their position and direction. SSU: small rRNA subunit encoding gene, ITS1: internal transcribed spacer region 1, 5.8S: 5.8S rRNA gene, ITS2: internal transcribed spacer region 2, LSU: large rRNA subunit encoding gene.

For the universal assessment of eukaryotes, with a focus on fungi (contingent upon the taxonomic coverage of ITS2 as above), the hypervariable regions of the ribosomal 18S rRNA gene was selected

Decomposition associated microbial community profile

as targets for amplification (Figure 5.2). A multitude of primers were assessed for their coverage of their respective taxonomic targets using TestPrime v1.0 (<https://www.arb-silva.de/search/testprime/>) by cross referencing to the SILVA ssu-132 database RefNR ribosomal RNA sequence collection. This software performs an in silico PCR to determine taxonomic coverage.

The hypervariable regions of the ribosomal 18S rRNA gene vary in their degrees of heterogeneity, with regions V9, V2 and V4 (in descending order) being the most diverse and providing the greatest taxonomic resolution (Hadziavdic et al., 2014). In a marine context, V4 has been shown to be superior and was targeted in this study (Stoeck et al., 2010).

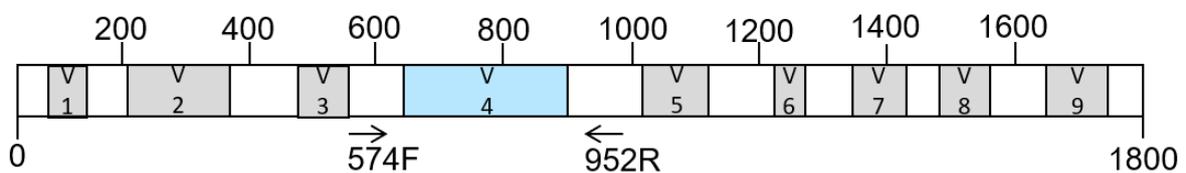


Figure 5.2 Hypervariable regions of the ribosomal 18S rRNA gene. Arrows indicate primer position and direction. Sizes displayed are base pairs and correspond to the *Saccharomyces cerevisiae* ribosomal 18S gene template, adapted after Hadziavdic et al., (2010).

To assess the bacterial and archaeal salt marsh profile, the ribosomal 16S rRNA gene was targeted using the well established V4 region with primers that carry degeneracy to enhance taxonomic coverage by alleviating negative bias (Figure 5.3) (Vasileiadis et al., 2012). The 515f-Y primer contains degeneracy that supports detection of *Crenarchaea* and *Thaumarchaeota* (Parada et al., 2016). Both of which are considered abundant constituents of the archaeal salt marsh profile (Nelson et al., 2009, Seyler et al., 2014). The 806R primer carries degeneracy that detects the SAR-11 clade in marine samples (Aprill et al., 2015).

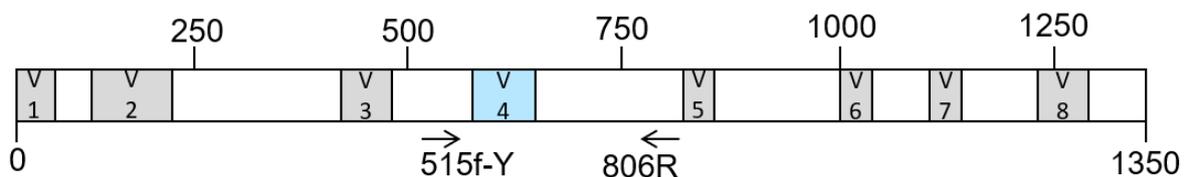


Figure 5.3 Hypervariable regions of the ribosomal 16S rRNA gene. Arrows indicate primer position and direction. Sizes displayed are base pairs and correspond to the *Escherichia coli* ribosomal 16S gene template, adapted after Vasileiadis et al., (2012).

Decomposition associated microbial community profile

Table 5.1 Primer selection for salt marsh community profiling.

| Target | Primer | Sequence (5'-3') | Coverage | Size (bp) | Reference |
|---|---------|----------------------|------------------|-----------|---------------------------|
| 16s rRNA gene V4 hypervariable region: Bacteria and Archaea | | | | | |
| 515f-Y | Forward | GTGYCAGCMGCCGCGGTAA | 85.3% Archaea | 290 | Parada et al., (2016) |
| 806R | Reverse | GGACTACNVGGGTWTCTAAT | 87.7% Bacteria | | Apprill et al., (2015) |
| Internal transcribed spacer region 2 (ITS2): Fungi | | | | | |
| fITS7 | Forward | GTGARTCATCGAATCTTTG | n/a | Variable | Ihrmark et al., (2014) |
| ITS4ngs | Reverse | TCCTSCGCTTATTGATATGC | n/a | | Tedersoo et al., (2014) |
| 18S rRNA gene V4 hypervariable region: Eukaryotes | | | | | |
| 574F | Forward | GCGTAATTCAGCTCCAA | 85.1% Fungi* | 380 | Hadziavdic et al., (2011) |
| 952R | Reverse | TTGGCAAATGCTTTCGC | 74.3% Eukaryota* | | Hadziavdic et al., (2011) |

* Coverage assumes 1 base pair mismatch

Amplicon libraries are low diversity as the same region is amplified. This is problematic for cluster identification on Illumina systems, which ostensibly uses the first 1-12 base pairs or the first read to determine cluster identity and quality. A random degenerate dodecamer sequence NNNHNNWNNN (5'-3') was prepended to the forward primers to reduce sample homogeneity by artificial inflation (Lundberg et al., 2013). The primer pairs used for direct amplification from genomic DNA had four regions; Nextera XT™ library prep adapter overhangs, flexible linker region, degenerate dodecamer sequence and target primer (Table 5.2).

Table 5.2 Primer architecture for direct amplification to indexing. Base colours represent primer regions, Orange; Nextera XT™ adapter sequence, Purple; flexible linker, Green; degenerate random dodecamer, Red; target specific primer sequence.

| Primer | Sequence (5'-3') |
|-----------|--|
| 515F-YMN | TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGANNHNNWNNNHGTGYCAGCMGCCGCGGT AA |
| 806RMN | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGGACTACNVGGGTWTCTAAT |
| fITS7MN | TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGANNHNNWNNNHGTGARTCATCGAATCTTTG |
| ITS4ngsMN | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTCCTSCGCTTATTGATATGC |
| 574FMN | TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGANNHNNWNNNHGCGGTAATTCAGCTCCAA |
| 952RMN | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTGGCAAATGCTTTCGC |

Consistent primer use within an environment allows comparable datasets to be generated; however, as amplicon sequencing for salt marshes has yet to be established, many combinations of primers were assessed. Primers were ranked by coverage (described above), if reproducible amplification was not achieved in all samples they were deemed inadequate. The chosen pairs (Table 5.1-2) were selected as they were the pair that were both most reliable and achieved the most coverage. This was exacerbated as Nextera XT™ library prep (Illumina) adapter overhangs were

Decomposition associated microbial community profile

appended to primers, which interfered with the specificity and functionality of the primers (Table 5.2). The combinations targeting the ribosomal 18S v4 regions that were considered inadequate, despite superior taxonomic coverage are shown in Table 5.3.

Table 5.3 Primer pairs tested that were deemed inadequate for salt marsh samples. **A-** represents TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGANNHNNHNNH (5'-3'), **B-** represents GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGT (5'-3')

| Target | Primer | Sequence (5'-3') | Coverage* | Size (bp) | Notes |
|---|--------|-----------------------------|-----------------|-----------|----------------------------|
| 18S rRNA gene V4 hypervariable region: Eukaryotes | | | | | |
| | 575F | A -CGGTAATCCAGCTCC | 86.1% Fungi | 340 | Poor yield |
| | 915R | B -TCCAAGAATTC ACCTC | 81.1% Eukaryota | | |
| | 566F | A -CAGCAGCCGCGTAATCC | 85.6% Fungi | 349 | Non-specific amplification |
| | 915R | B -TCCAAGAATTCACCTC | 80.7% Eukaryota | | |

* Assumes 1 base pair mismatch

5.2.1.3 Amplification of conserved regions

Polymerase chain reactions (PCR) were conducted using the genomic DNA (gDNA) obtained in 5.2.1.1. The gDNA was re-quantified using a Qubit 4 fluorometer (Invitrogen) with a dsDNA HS Assay Kit (Q32854). PCR conditions were as described in Table 5.4, and thermocycling conditions for each profile are displayed in tables 5.5-7. Amplifications were conducted for each biological replicate (5) at each time point (9) and including a day 0 outgroup, for each amplicon profile (138 in total).

Table 5.4 PCR conditions for amplification of hypervariable regions.

| Reagent | Final concentration | Volume per reaction (μL) |
|---------------------------|---------------------|--------------------------|
| dH2O to 50μL | - | 14.75 |
| Phusion HF Buffer (5x) | 1x | 5 |
| dNTP (10mM) | 200μM | 0.5 |
| pF (10μM) | 0.5μM | 1.25 |
| pR (10μM) | 0.5μM | 1.25 |
| DNA (25ng/μL) | 2ng/μL | 2 |
| Phusion (2000U/mL) #M0530 | 1U/50μL | 0.25 |
| Total | - | 25 |

Decomposition associated microbial community profile

Table 5.5 Thermocycling conditions for ribosomal 16S amplification using 515F-Y and 806R.

| Step | Temperature (°C) | Time | |
|----------------------|------------------|--------|---------------|
| Initial denaturation | 98 | 30s | |
| Denaturation | 98 | 10s | Repeat x28 |
| Annealing | 53 | 30s | |
| Extension | 72 | 15s | |
| Final extension | 72 | 10 min | |
| Hold | 4 | - | |

Table 5.6 Thermocycling conditions for ITS2 amplification using fITS7 and ITS4ngs.

| Step | Temperature (°C) | Time | |
|----------------------|------------------|------|---------------|
| Initial denaturation | 98 | 30s | |
| Denaturation | 98 | 10s | Repeat x34 |
| Annealing | 57 | 30s | |
| Extension | 72 | 20s | |
| Final extension | 72 | 5min | |
| Hold | 4 | - | |

Table 5.7 Thermocycling conditions for ribosomal 18S amplification using 574F and 952R.

| Step | Temperature (°C) | Time | |
|----------------------|------------------|--------|---------------|
| Initial denaturation | 98 | 30s | |
| Denaturation | 98 | 10s | Repeat x28 |
| Annealing | 60 | 30s | |
| Extension | 72 | 15s | |
| Final extension | 72 | 10 min | |
| Hold | 4 | - | |

5.2.1.4 *Illumina sequencing*

Amplicon profiles were generated using the primer pairs and conditions stated in section 5.2.1.3. Amplicon profiles were assessed for purity and quantified using an Agilent 4200 TapeStation with High Sensitivity D1000 ScreenTapes (5067-5584 and 5067-5585). Amplicons were cleaned and purified using AMPure XP beads, at size specific bead concentrations (Beckman Coulter A63881)

Decomposition associated microbial community profile

(Table 5.8; clean I). Amplicon profiles were then indexed using the Nextera XT™ library preparation kit (Illumina FC-131-1001) according to the manufacturer’s protocol with thermocycling conditions presented in Table 5.9. As the number of samples (138) was greater than the number of indexing combinations in this kit (96), the 16S and ITS2 profiles were indexed with the same combinations and separated *a posteriori* by primer sequence. Indexed profiles were cleaned again using AMPure XP beads as above, at different bead concentrations (Table 5.8; clean II). Profiles were assessed for successful indexing by assessing the size change using the Agilent 4200 TapeStation with D1000 ScreenTapes. The amplicon profiles were quantified using a Qubit 4 fluorometer (Invitrogen) with a dsDNA HS Assay Kit (Q32854), from this nanomolar concentrations were calculated. Amplicon profiles were pooled in their respective fraction (16S, 18S, ITS2) in equimolar concentrations, re-quantified using the Qubit 4 fluorometer. The three mixtures were pooled again in equimolar concentrations, re-quantified and finally diluted to 4nM.

Table 5.8 AMPure XP bead concentrations at each purification stage for each amplicon profile.

| Amplicon profile | Clean I (initial PCR) | Clean II (indexing) |
|------------------|-----------------------|---------------------|
| 16S | 0.8x | 0.9x |
| 18S | 0.8x | 0.9x |
| ITS2 | 1.2x | 0.9x |

Table 5.9 Thermocycling conditions for Nextera XT™ indexing.

| Step | Temperature (°C) | Time | |
|----------------------|------------------|------|--------------|
| Initial denaturation | 95 | 3min | |
| Denaturation | 95 | 30s | Repeat x8 |
| Annealing | 55 | 30s | |
| Extension | 72 | 30s | |
| Final extension | 72 | 5min | |
| Hold | 4 | - | |

5µL of 0.2N NaOH was combined with 5µL of the 4nM pooled library. The mixture was vortexed briefly and incubated at room temperature for 5 mins to denature the dsDNA. 990µL of pre-chilled (4°C) Hybridisation Buffer (HT1) was added to the denatured library. The denatured library was further diluted to 4pM (120µL denatured library and 480µL HT1). The PhiX library was created with 2µL 10nM PhiX (PhiX Control Kit v3 FC-110-3001) and 3µL of 10mM Tris pH 8.5, to a final

Decomposition associated microbial community profile

concentration of 4nM PhiX. The 4nM PhiX library was denatured with the addition of 5µL 0.2N NaOH and incubated at room temperature for 5 min to denature the dsDNA. 10µL of denatured PhiX library was diluted with 990µL HT1 to a final concentration of 20pM. The 20pM denatured PhiX library was diluted to 4pM with HT1.

The PhiX library (6µL) was added to 594µL of amplicon library, to a final PhiX spike concentration of 1%. The pooled library was incubated at 96°C for 2 min and incubated on ice for 5min. The samples were then immediately loaded onto a pre-prepared MiSeq 250bp x 2 cartridge (MiSeq Reagent Kit v2 (500 cycles) MS-102-2003, Illumina). This work was performed at the Biorenewables Development Centre (Dunnington, York), with the assistance of Dr. Deborah Rathbone and Susan Heywood.

5.2.2 Bioinformatic analysis

Bioinformatic analysis of the sequence data generated in section 5.2.1.4 allows the construction of an operational taxonomic unit (OTU) table. The hypervariable fragments of sequences are clustered to an identity threshold. A representative sequence of all the sequences in a cluster are generated and assessed against archive reference databases for homology to known taxa (at an identity threshold). The size of the OTUs (sequences contained within the cluster) can then be related to the taxonomic assignments to generate a quantitative taxonomic profile.

5.2.2.1 *Sequence to community profile analysis pipeline*

An analysis pipeline was developed to assess the community profile from the indexed Illumina MiSeq generated databases (Figure 5.4). The forward and reverse reads for the 16S and ITS2 profiles sharing the same index and the forward and reverse reads for the 18S profile were merged using Vsearch version 1.11.1 (Rognes et al., 2016). The 16S/ITS2 and 18S files merged to 91.61% and 93.21% respectively. The 16S/ITS2 files for the forward read were trimmed of the random dodecamer sequence and linkers using Cutadapt version 1.11 (Martin, 2011). The ITS2 sequences from the trimmed forward read files were extracted from the 16S and ITS2 pool using QIIME (version 1.9.1; split_library.py) with a phred quality score of 20 (Caporaso et al., 2010). The merged 16S/ITS2 pool and merged 18S files were trimmed (random dodecamer and linkers) and quality checked with an e value of 20 using Cutadapt (version 1.11.1). The merged, trimmed and quality checked 16S/ITS2

Decomposition associated microbial community profile

pool and 18S files were demultiplexed using Qiime (version 1.9.1; `split_library.py`), the ITS2 sequences in the 16S/ITS2 pools were discarded.

The primer sequences were trimmed from the 16S and 18S profiles using Cutadapt (version 1.11.). Primer sequences, 5.8S and 28S regions were removed from the ITS2 amplicon profile using ITSx and the fungi only pool was extracted for further processing, the non-fungi pool of sequences were discarded (Bengtsson-Palme et al., 2013). Biological replicates for each profile (16S, 18S, ITS2) were concatenated into single files, the headers were formatted to Usearch compatible formats ("`barcodelabel=sample_id;sequence_number_integer`"), and globally trimmed to equal lengths (16S; 250bp, 18S;250bp, ITS2;100bp) using Usearch (version 9; `-fastx_truncate`) (Edgar, 2010). All amplicon profiles were dereplicated, purged of singletons, assigned abundance and sorted by size using Usearch (version 7: `-derep_fulllength`) (Edgar, 2010). The profiles were then clustered into OTUs using the UPARSE algorithm (Edgar, 2013), with concurrent de novo chimera detection using Usearch (version 9 `-cluster_otus`) with a 97% identity threshold. OTUs were relabelled from sample IDs to OTU numbers (Usearch; `fasta_number.py`). Representative sequences for each OTU were then mapped to original sequences using Usearch (version 7: `-usearch_global`). Taxonomy was assigned using QIIME (version 1.9; `assign_taxonomy.py`) for 16S, 18S and ITS2 using Greengenes 13.8, SILVA 123 SSU QIIME release and Unite version 7.1 reference databases respectively. The `.uc` output (uclust file format) was converted to text file format (Usearch: `uc2otutab.py`), and further converted to `.biom` and `.tsv` (Biom: `convert`) for custom analysis in Python (version 3.6).

Decomposition associated microbial community profile

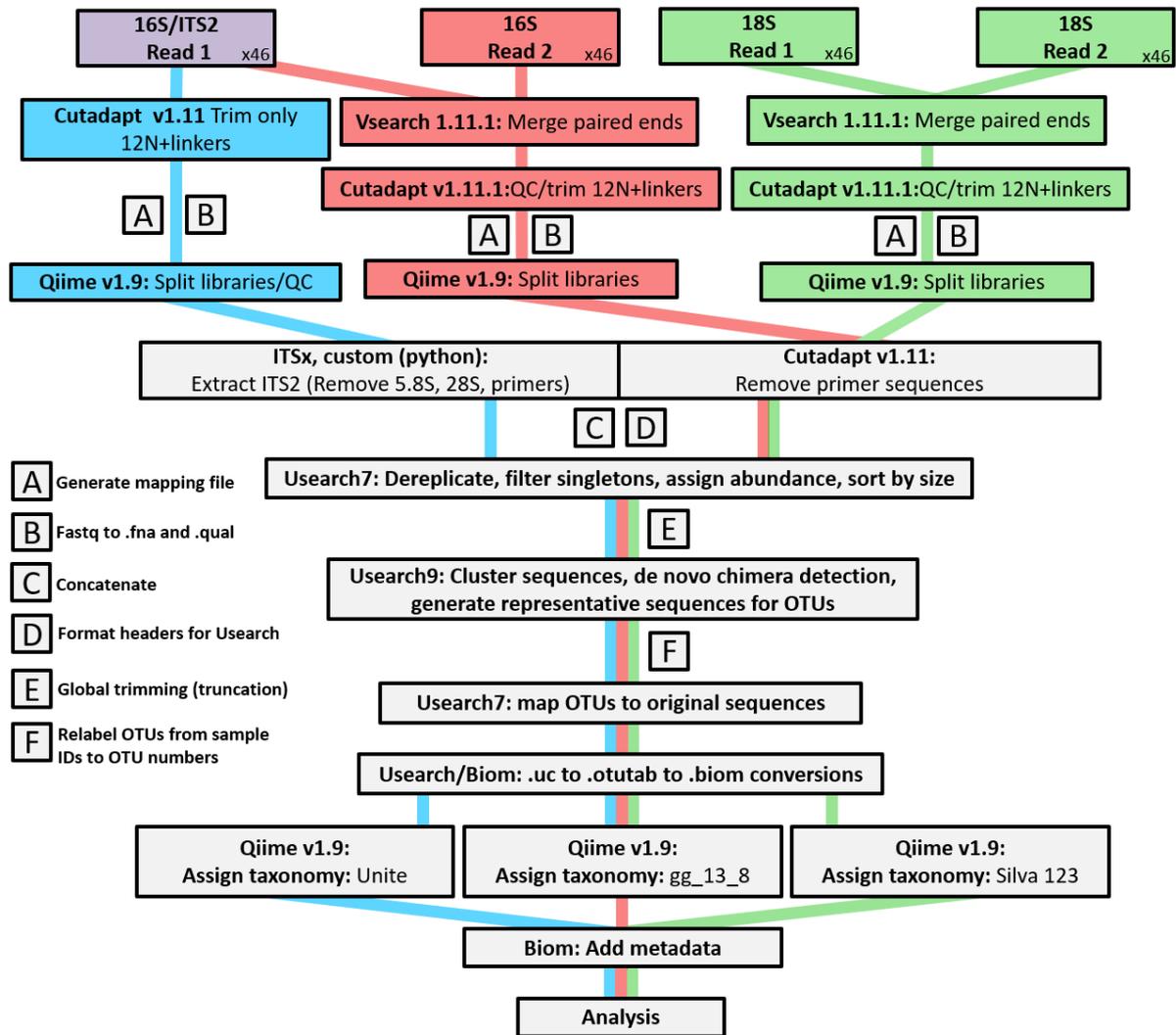


Figure 5.4 Analysis pipeline for amplicon sequence databases.

Decomposition associated microbial community profile

Table 5.10 Commands for the replication of the 16S rRNA amplicon database processing pipeline. Emboldened characters are specific input or output files. Only steps utilising third party software is shown, custom steps are not shown. A1-A12 represent input or output files.

| Step | Function | Command |
|----------|------------------|---|
| 1 | Merge | vsearch --fastq_mergepairs X1_001.fastq.gz --reverse X2_001.fastq.gz --fastqout X_16Smerged |
| 2A | Trim adapter | cutadapt -o nexterearemoved/ X -a CTGTCTCTTATACACATCTGACGCTGCCGACGA X |
| 2B | Trim 13N | cutadapt --cut 13 -o X X |
| 3 | Fastq split | convert_fastaqual_fastq.py -c fastq_to_fastaqual -f X |
| 4 | Demultiplex | split_libraries.py -f X --max_ambig 0 -r -k -B -H 10 -M 2 -e 2 -b 7 -o X -m MAP |
| 5 | Trim primer | cutadapt --cut 19 --cut -20 -o X X |
| 6 | Concatenate | cat *.fna > catfile.fna |
| 7 | Format header | Format headers to >barcode label=sample_id;sequence_number_integer - custom |
| 8 | Global trim | usearch_v9 -fastx_truncate A1 -truncLen 250 -fastaout A2 |
| 9 | Dereplicate | usearch_v7 --derep_fulllength A2 --output A3 --log=log --sizeout --minuniquesize 2 |
| 10 | Sort by size | Usearch_v7 -sortbysize A3 -output A4 -minsize2 |
| 11 | Cluster | usearch_v9 -cluster_otus A4_cat16S.fna -otus A5 -minsize 2 |
| 12 | Relabel | fasta_number.py A5 OTU_> A6 |
| 13 | Map OTUs | Usearch_v7 -usearch_global A1 -db A6 -strand plus -id 0.95 -uc A7 |
| 14 | Assign taxonomy | Assign_taxonomy.py -i A6 -o A8 --similarity 0.9 -r ref_set.fna -t ref_set.txt |
| 15 | OTU table (.txt) | uc2otutab.py A7 > A9.txt |
| 16 | .biom | Biom convert -l A9 -o A10 --table-type="OTU table" --to-hdf5 |
| 17 | .tsv | Biom convert -i A10 -o A11 --to-tsv |
| Optional | | |
| 18 | Metadata | Biom add-metadata -i A10 -o A12 --observation-metadata-fp otus_tax_assignments.txt --observation-header OTUID,taxonomy,confidence --sc-separated taxonomy --float-fields confidence |

5.3 Results

5.3.1 Database processing

One hundred and thirty eight amplicon profiles (46 for each of the 16S, ITS2 and 18S target regions) were successfully generated (Figure 5.5). A total of 8,768,150 reads were achieved across the 138 loaded samples, despite sub-optimal cluster density of 430 (optimal cluster densities for V2 flowcells are 900-1000) suggesting underloading of the flowcell (4pM was loaded, 8pM would have been more appropriate). The low density appeared to be offset by artificially inflated sequence

Decomposition associated microbial community profile

heterogeneity conferred by the 12N random dodecamer. Amplicon libraries were pooled in equimolar quantities successfully, with the average database containing (post library split by primer sequence and separation of the 16S/ITS2 pool) 54,929, 50,844 and 52,853 for 16S, ITS2 and 18S, respectively (Figure 5.6). Within the ITS2 databases a mean proportion of 73.80% sequences were of fungal origin (due to primer specificity of fITS7) which reduced the mean database size to 28,972 sequences (Figure 5.6).

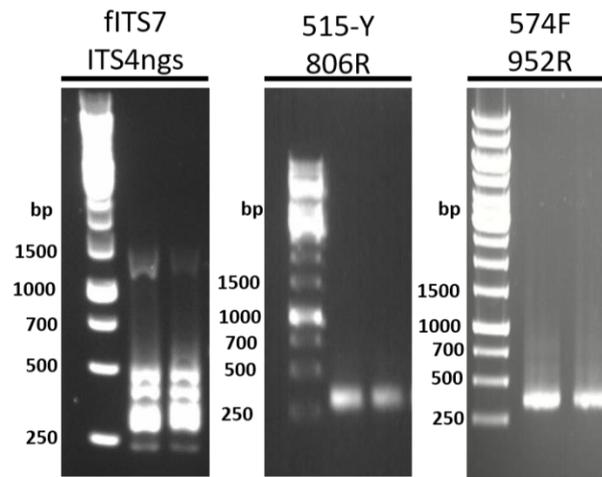


Figure 5.5 A selection of amplicon profiles for teach target region. ITS2 (fITS7 and ITS4ngs), 16S (fITS7 and ITS4ngs) and 18S 574F and 952R).

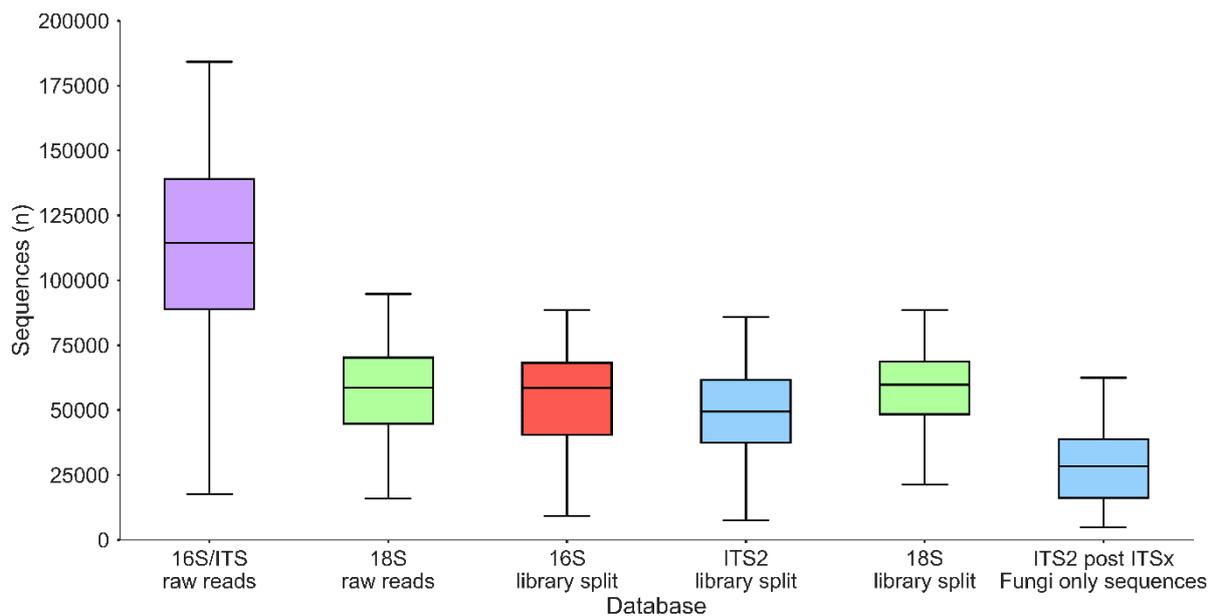


Figure 5.6 Read distribution of the amplicon profile databases, pre- and post-assignment.

Decomposition associated microbial community profile

The 16S/ITS2 profiles and 18S profiles merged 91.61% and 93.21% successfully, respectively. The degree of merging in the 16S/ITS2 profile was likely due to the small range of ITS2 amplicon sizes generated (300-500bp; Figure 5.5). However, as some sequences appear larger than 500 the forward read (read one) was taken forward unmerged to prevent bias against taxa with longer ITS2 reads. The UPARSE clustering algorithm requires global amplicon lengths; a length of 100bp was selected for ITS2 amplicons, and 250bp for 16S and 18S profiles (Table 5.11).

Table 5.11 Statistics for the amplicon database processing pipeline.

| Stage | | | | |
|------------------|-----------------|-------------------------|-----------------------|--|
| Truncation | | | | |
| Profile | Trimmed to (bp) | Sequences discarded (%) | Sequences remaining | |
| 16S | 250 | 0.4 | 2,516,510 | |
| 18S | 250 | 0.1 | 2,427,394 | |
| ITS2 | 100 | 0.3 | 1,329,125 | |
| Clustering (97%) | | | | |
| Profile | Total OTUs (n) | Chimeric OTUs (n) | Non-chimeric OTUs (n) | |
| 16S | 26797 | 21675 | 5122 | |
| 18S | 5221 | 3251 | 1970 | |
| ITS2 | 1406 | 225 | 1181 | |
| Mapping (95%) | | | | |
| Profile | OTUs mapped (%) | | | |
| 16S | 74.1 | | | |
| 18S | 97.6 | | | |
| ITS2 | 99.0 | | | |
| Assignment | | | | |
| Profile | Assigned | Unassigned | | |
| 16S | 4403 | 719 | | |
| 18S | 1156 | 814 | | |
| ITS2 | 569 | 612 | | |

5.3.2 Prokaryotic profiles

5.3.2.1 *Archaea*

The prokaryotic profile elucidated from the 16S analysis revealed archaea to be orders of magnitude less abundant than bacteria. archaea were most abundant in week 16 at 0.44% with an average temporal abundance of 0.12% (Figure 5.7).

Decomposition associated microbial community profile

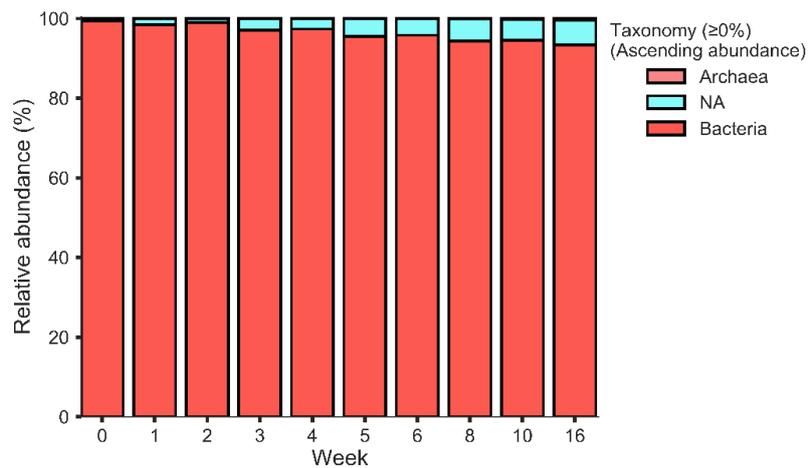


Figure 5.7 Prokaryotic profile from 16S rRNA sequence homology at Kingdom resolution. Archaea are not visible due to their low abundance.

Within the archaeal community, *Crenarchaeota* and *Parvarchaeota* were the most abundant phyla, with the former dominant in the day 0 outgroup, before encroachment of *Parvarchaeota* which claims 93% of the profile in later timepoints (Figure 5.8A). This phylum consisted of a single class *Parvarchaeo*, that comprised of *YLA114* and *WCHD3-30* orders (Figure 5.8B). The *Crenarchaeota* were constituted by *Cenarchaeales*. Further taxonomic resolution (family and below) could not be reliably resolved for these taxa. There were no carbohydrate active enzymes originating from archaea in Chapter 4 and, therefore, the Archaeal sp. profile was not further explored.

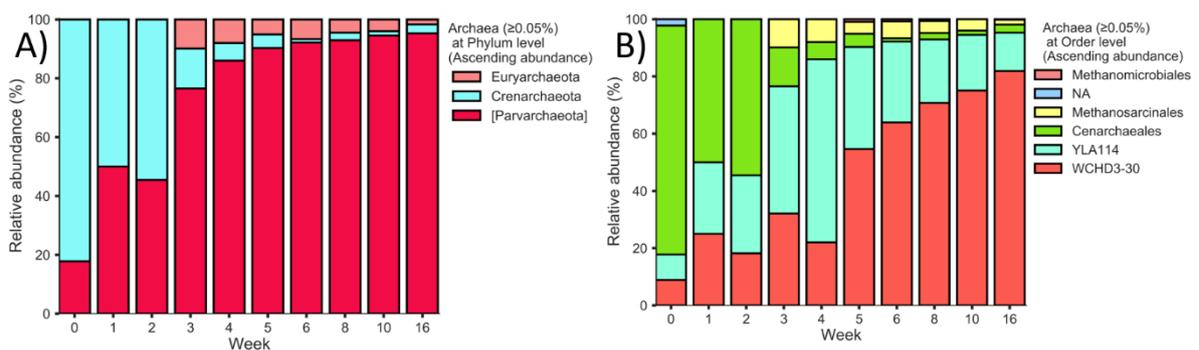


Figure 5.8 Archaeal profiles elucidated from 16S rRNA sequence homology. A) Phylum level resolution, B) Order level resolution.

Decomposition associated microbial community profile

5.3.2.2. Bacteria

OTU richness was lowest in week one ($n=1020 \pm 65$), increasing up to week six ($n=2126 \pm 84$) where richness was comparable to the richness observed in the day 0 sediment outgroup ($n=2277$), before plateauing and subsequently declining (Figure 5.9). The bacterial fraction was consistently dominated by *Proteobacteria* and *Bacteroidetes* at a mean sum total of $80.51 \pm 1.34\%$ relative abundance across all time points. The minor phyla were constituted by *Firmicutes*, *Verrucomicrobia*, *Actinobacteria*, *Spirochaetes* and *Chloroflexi*. The major *Proteobacterial* representatives at the class level consisted of (in descending order of abundance) *Gamma*proteobacteria, *Delta*proteobacteria and *Alpha*proteobacteria (Figure 5.10). Abundant class level representatives of *Bacteroidetes* consisted of *Bacteroidia* and *Flavobacteria* which contain the most abundant orders (*Bacteroidales* and *Flavobacteriales* respectively) (Figure 5.11).

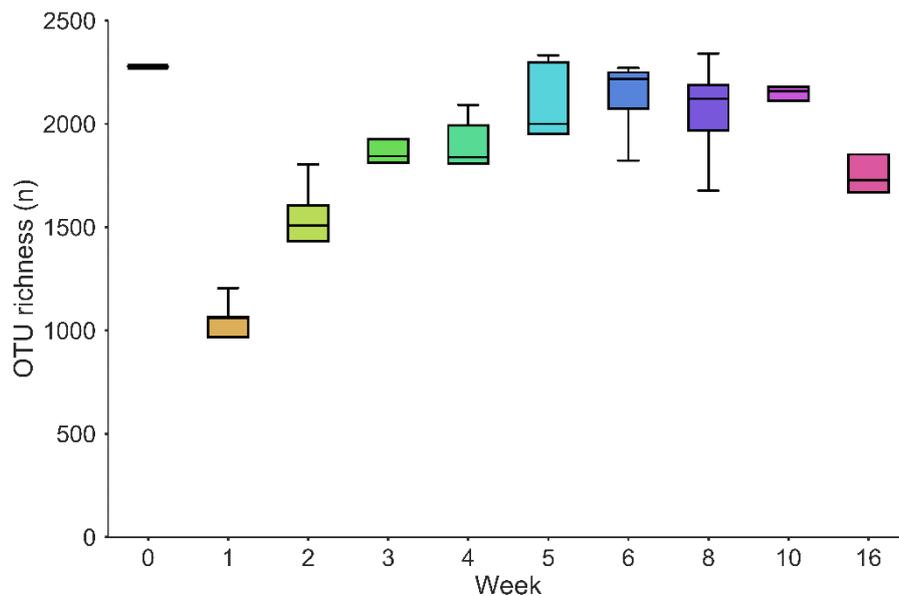


Figure 5.9 Temporal OTU richness for bacterial and archaea lineages.

Proteobacteria consistently declined in abundance and the most pronounced change is exhibited by *Bacteroidetes* who encroach into this space. *Gamma*proteobacteria were responsible for the marked decline of *Proteobacteria*, most of this decline is offset by a concurrent increase in *Delta*proteobacterial abundance and comparative stability of *Alpha*proteobacteria and *Epsilon*proteobacteria. The majority of *Bacteroidetes* expansion is due to *Bacteroidia*.

Decomposition associated microbial community profile

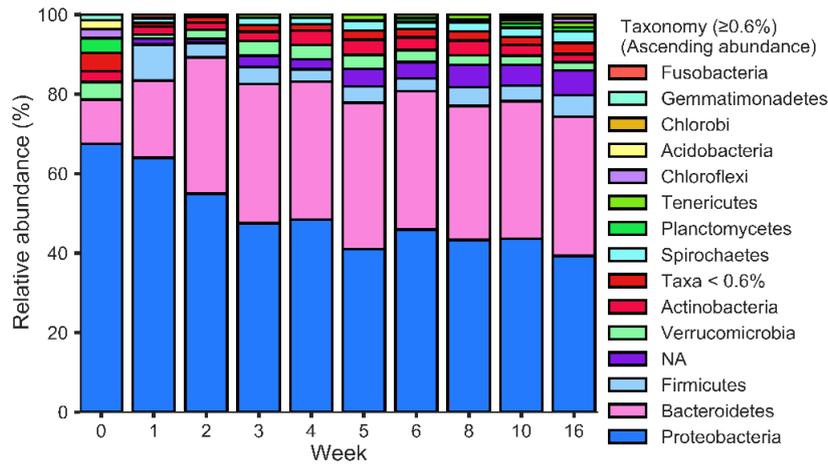


Figure 5.10 Phylum level bacterial profile elucidated from 16S rRNA sequence homology.

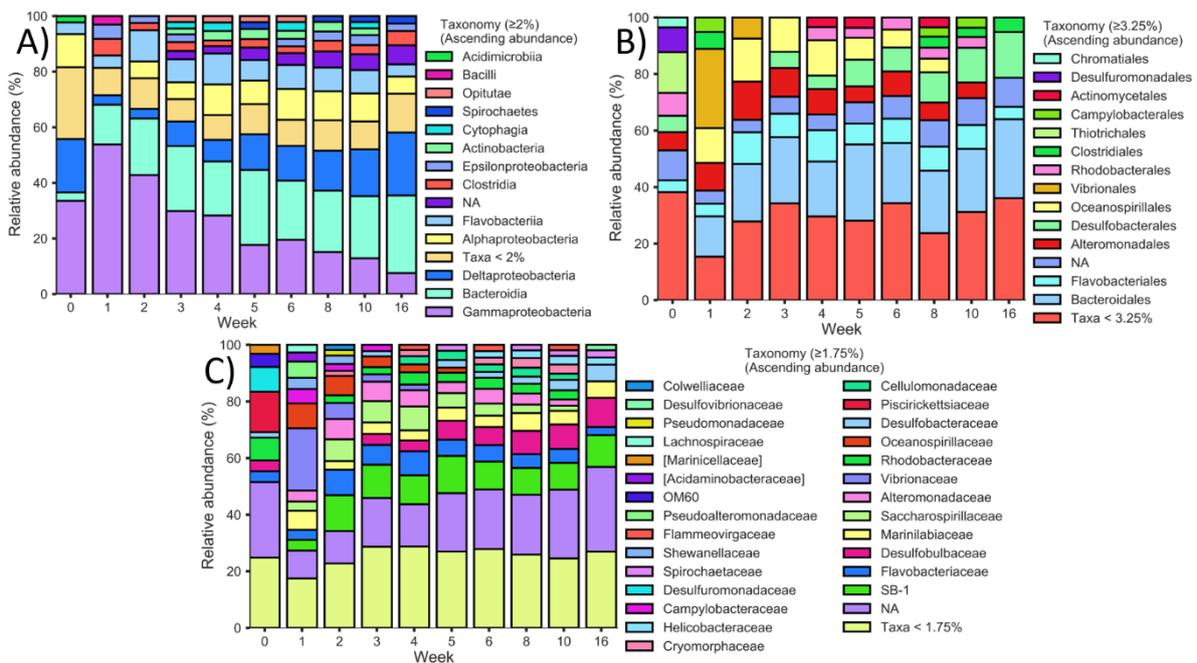


Figure 5.11 Bacterial profiles elucidated from 16S rRNA sequence homology. A) Class level resolution B) Order level resolution C) Family level resolution.

5.3.2.2.1 Gammaproteobacteria

Within *Gammaproteobacteria* the *Alteromonadales* and *Oceanospirillales* were temporally stable and highly abundant (Figure 5.12A); these orders were the third and fifth most abundant order in this study (Figure 5.11B). At the family level, *Alteromonadales* were characterised by *Alteromonadaceae* $53.46 \pm 6.59\%$, with a minor contribution by *OM60* which were dominant in the sedimentary day 0 outgroup (77.14% of *Alteromonadales*) and became more abundant in later

Decomposition associated microbial community profile

timepoints (10.9-28.26% in weeks six to 16 respectively). *Oceanospirillales* were dominated by two families *Oceanospirillaceae*, which decreased in abundance temporally from 71.25% to 16.58% from week one to 16, and *Saccharospirillaceae* which increased in abundance from 6.54% to 70% in from the day 0 outgroup to week eight, before declining to 54% in week 16. The sedimentary day 0 outgroup was dominated by the order *Thiotrichales*; however, they were not abundant on biomass samples. *Vibrionales* were also transiently highly abundant during week one, declining in abundance thereafter and consisted predominantly of *Vibrionaceae*.

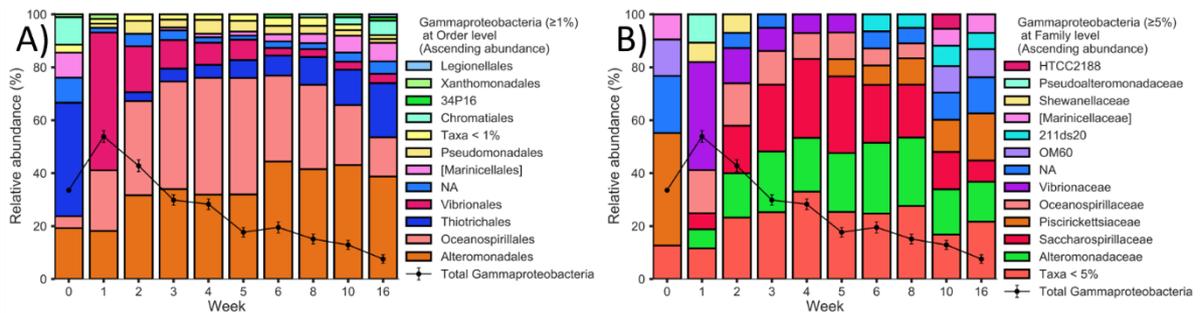


Figure 5.12 Normalised *Gammaproteobacterial* fraction. A) Order level resolution, B) Family level resolution. The lineplot represents the visualised fractions (bars) total proportion of the entire microbiome.

Filtering and cross referencing these profiles by CAZy producing taxa identified in Chapter 4 reveals that despite the decline in relative abundance of the *Gammaproteobacterial* community, the representatives of these groups are enriched for carbohydrate associated groups (Figure 5.13). At the family level *Gammaproteobacteria* became significantly enriched for CAZy producing clades, increasing from 6.01% in the day 0 sediment outgroup, to $76.0 \pm 0.71\%$ during weeks one to five, before receding (Figure 5.13B).

These results are concordant with the proteomic results in Chapter 4. *Vibrionales* are the dominant carbohydrate active enzyme (CAZy) producing order in week one followed by a period of inactivity, which is synchronous to their abundance profile (this also translates to family level resolution for *Vibrionaceae*). *Alteromonadales* exhibit a positive temporal increase in abundance and are the second largest CAZy producing order (which also translated to family level to *Alteromonadaceae*). At the order level, *Oceanospirillales* are dominant CAZy producers which increase in relative abundance, the dominant family within this taxon is the *Saccharospirillaceae* which also become enriched and are the sixth greatest CAZy producing family.

Decomposition associated microbial community profile

The order *Cellvibrionales* is the fourth most active CAZy producing clade, containing the *Cellvibrionaceae* family (the third most active family, Chapter 4). These clades have been observed in saline environments and have only recently been described (Spring et al., 2015). As a direct consequence of their novelty they are not represented in the taxonomic database (GreenGenes 13.8) and are likely not assigned (NA) or assigned to closely related taxa.

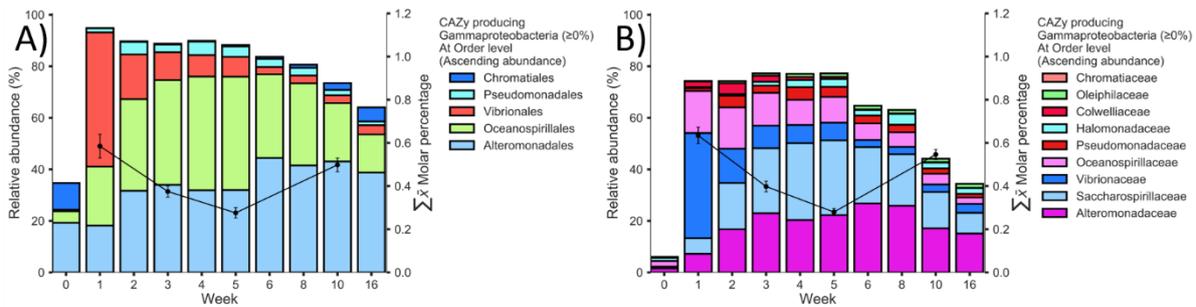


Figure 5.13 CAZy producing *Gammaproteobacteria* normalised within the *Gammaproteobacterial* profile. A) Order level resolution, B) Family level resolution. Non-CAZy associated taxa have been filtered, relative abundances are relative to total *Gammaproteobacterial* representation. Plots display the sum of the mean molar percentage originating from visualised taxa (bars).

5.3.2.2.2 Bacteroidetes

Bacteroidetes were the second most abundant phylum at microbiome level, exhibiting persistent stability after initial encroachment during weeks one ($11.0 \pm 1.8\%$) and two ($19.4 \pm 2.1\%$), occupying the later time points at $34.9 \pm 0.3\%$ (Figure 5.10, 5.14A-B). Within the *Bacteroidetes* phylum the *Bacteroidia* and *Flavobacteriia* families were dominant. *Bacteroidia* was temporally stable and more abundant in the decomposing biomass ($66.0 \pm 2.4\%$ for week one to 16) than the sediment outgroup (27.2%). The relative presence of *Flavobacteria* as a proportion of the *Bacteroidetes* declined from 37.3% in the outgroup sample to $24.1 \pm 2.0\%$ during weeks one to 16. (Figure 5.14).

The *Bacteroidales* exhibited the greatest increase in abundance at the microbiome level, increasing from 3.04% in the day 0 outgroup, to $23.43 \pm 1.93\%$ in week three, before stabilising at an average of $23.36 \pm 1.16\%$ between weeks three and 16. At the family level, the *Bacteroidales* were dominated by the interestingly unclassified *SB-1* clade and *Marinilabiaceae*, both of which were not present in the sediment outgroup samples but were highly abundant in the lignocellulose associated microbiome (Figure 5.11C). *SB-1* increased during week one to 3.8% , before stabilising at $9.9 \pm 0.9\%$ throughout the time course. *Marinillabiaceae* were less abundant than the *SB-1* clade, present from week one through to 16 at a mean abundance of $4.68 \pm 0.43\%$ at the microbiome level. The

Decomposition associated microbial community profile

dominant *Flavobacteriales* clades were *Flavobacteriaceae* and *Cryomorphaceae*. *Flavobacteriaceae* were abundant in the sediment outgroup (3.75%), increasing only to a mean of $5.65 \pm 0.65\%$ at the microbiome level.

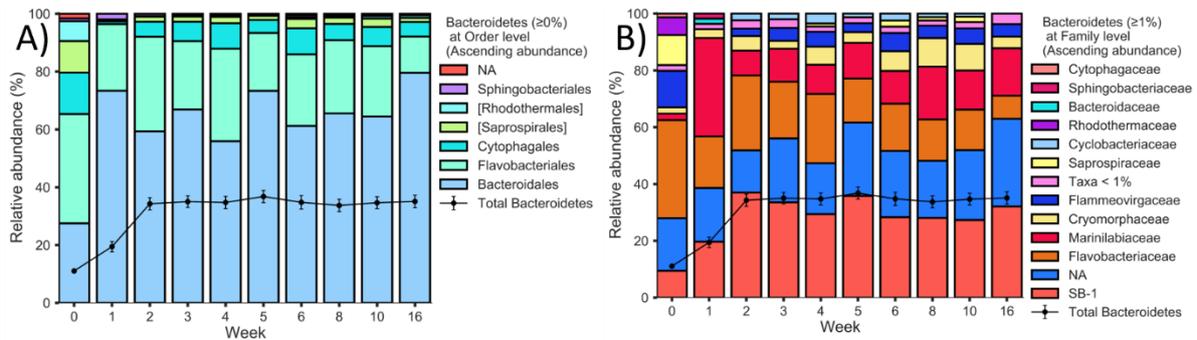


Figure 5.14 Normalised *Bacteroidetes* fraction. A) Order level resolution, B) Family level resolution. The lineplot represents the visualised fractions (bars) total proportion of the entire microbiome.

The *Bacteroidetes* profile was cross referenced to the CAZy producing *Bacteroidetes* taxa identified in Chapter 4 (Figure 5.15). This profile elucidated that *Flavobacteriaceae* were the most abundant CAZy producing family within *Bacteroidetes* and the second most active producer in the microbiome (Chapter 4), followed by *Marinilabiaceae* and *Flammeovirgaceae*, both of which were abundant within the microbiome. The CAZy associated fraction was not significantly enriched at order or family level, but these groups exhibited temporal stability both within *Bacteroidetes* and at the microbiome level.

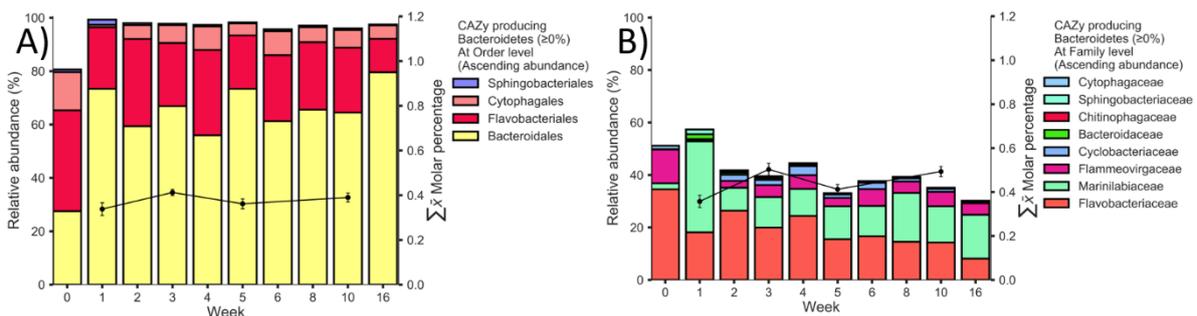


Figure 5.15 CAZy producing *Bacteroidetes* normalised within the *Bacteroidetes* profile. A) Order level resolution, B) Family level resolution. Non-CAZy associated taxa have been filtered, relative abundances are relative to total *Bacteroidetes* representation. Plots display the sum of the mean molar percentage originating from visualised taxa (bars).

Decomposition associated microbial community profile

5.3.2.2.3 Deltaproteobacteria

In contrast to *Gammaproteobacteria*, *Deltaproteobacteria* displayed a temporal increase in abundance at the microbiome level, after an initial recession from the high levels of abundance observed in the day 0 outgroup (19.17%), to $3.4 \pm 1.19\%$ in week one to $22.64 \pm 2.01\%$ in week 16 (Figure 5.16A). Major constituents of this class were *Desulfobacterales*, *Desulfuromonadales*, *Desulfovibrionales* and at later time points *Myxococcales* (order level). *Desulfobacterales* become dominant, occupying $72.94 \pm 0.94\%$ (within *Deltaproteobacteria*) in the final three weeks. *Desulfuromonadales* were the most abundant order in the sediment outgroup (45.9%), however in the lignocellulose associated microbiome they were far less abundant $9.05 \pm 1.54\%$. Both orders were highly abundant at the microbiome level (Figure 5.11B).

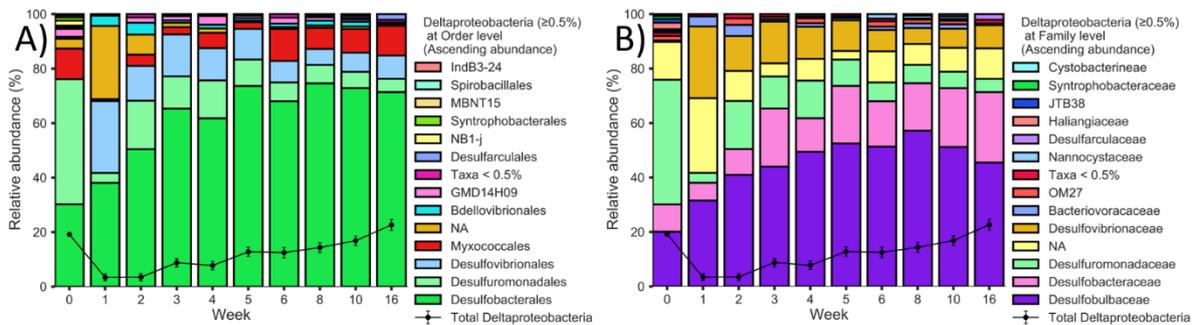


Figure 5.16 Normalised *Deltaproteobacteria* fraction. A) Order level resolution, B) Family level resolution. The lineplot represents the visualised fractions (bars) total proportion of the entire microbiome.

Within the *Desulfobacterales* and *Desulfuromonadales* orders, the families *Desulfobulbaceae*, *Desulfomonadaceae*, *Desulfobacteraceae* and *Desulfovibrionaceae* were dominant (Figure 5.16B). Interestingly all of these families with the exception of *Desulfobacteraceae* were CAZy producers (Figure 5.17). The abundant families within *Deltaproteobacteria* were temporally consistent, and accounting for the increase at the microbiome level suggests this class was enriched for CAZy producing organisms (Figure 5.11C).

Decomposition associated microbial community profile

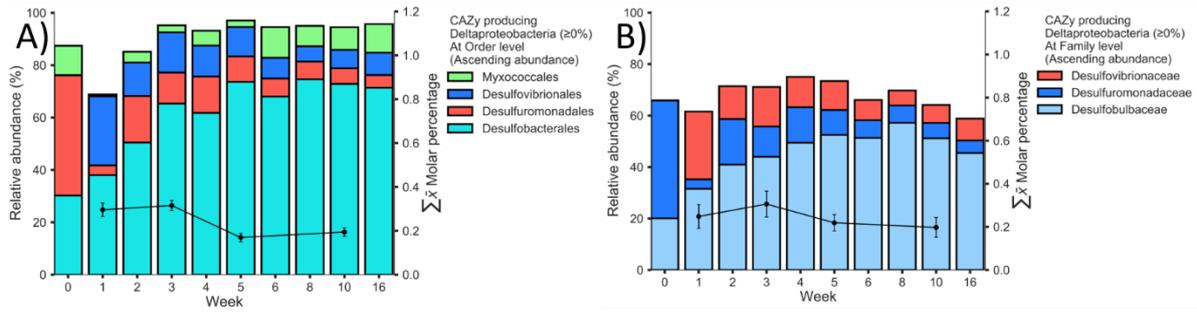


Figure 5.17 CAZY producing *Deltaproteobacteria* normalised within the *Bacteroidetes* profile. A) Order level resolution, B) Family level resolution. Non-CAZY associated taxa have been filtered, relative abundances are relative to total *Deltaproteobacteria* representation. Plots display the sum of the mean molar percentage originating from visualised taxa (bars).

5.3.2.2.4 Firmicutes

Firmicutes was not abundant in sediment samples (0.41%) but increased in abundance within the decomposing biomass ($4.59\% \pm 0.61$) and exhibited temporal stability (Figure 5.18). This phylum was dominated by the order *Clostridiales* with *Bacillales* far less abundant. The dominant families were *Lachnospiraceae*, *Christensenellaceae*, *Acidaminobacteraceae* and *Planococcaceae*. *Lachnospiraceae* (*Clostridiales*) and *Planococcaceae* (*Bacilli*) were highly active CAZY secretors. Less abundant CAZY producers included *Peptococcaceae* (*Clostridia*) and *Paenibacillaceae* (*Bacilli*).

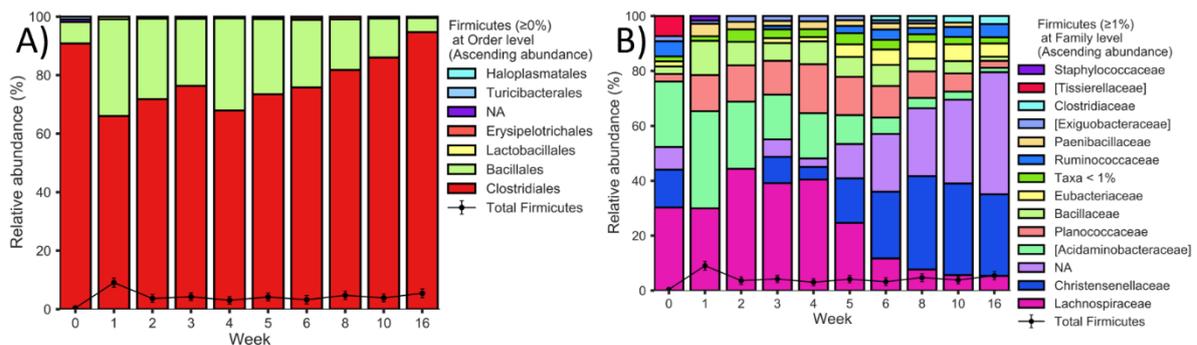


Figure 5.18 Normalised *Deltaproteobacteria* fraction. A) Order level resolution, B) Family level resolution. The lineplot represents the visualised fractions (bars) proportion to the entire microbiome.

5.3.2.2.5 Verrucomicrobia

Verrucomicrobia were also present at low abundances throughout the time course Figure (5.19). *Verrucomicrobiales* appears to be the lineage responsible for the recession, where biomass

Decomposition associated microbial community profile

associated dominant orders became the *R76-B128* clade, *Pelagicoccales*, *Puniceicoccales* and *Verrucomicrobiales*. At the family level *Pelagiococcaceae* and *Puniceicoccaceae* were abundant, however taxonomic assignment was particularly scarce within this phylum. In terms of the CAZy producing lineages, only *Opitutaceae* were identified, whereas proteomic homology to *Rubritaleaceae* was observed.

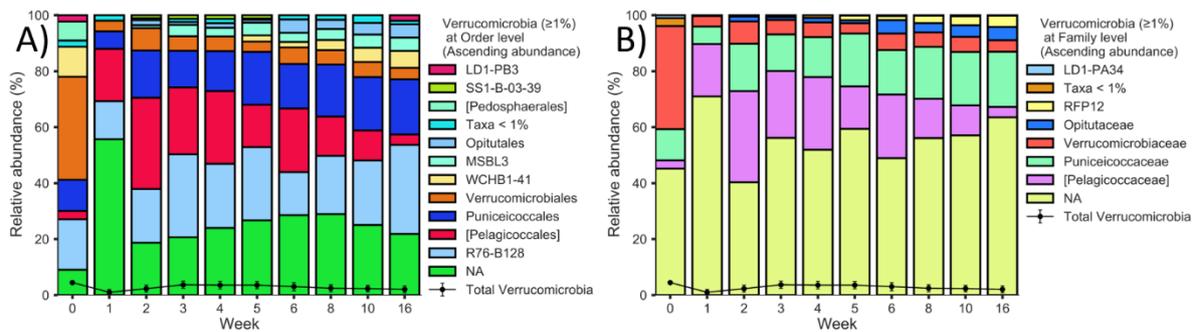


Figure 5.19 Normalised *Verrucomicria* fraction. A) Order level resolution, B) Family level resolution. The lineplot represents the visualised fractions (bars) proportion to the entire microbiome.

5.3.2.2.6 Lignocellulolytic producing taxa at the microbiome level.

To elucidate the temporal profile of lignocellulolytic taxa, the CAZy producing taxa identified in Chapter 4 was cross referenced at the microbiome level and non-CAZy associated taxa were filtered. In contrast to the methodology in Chapter 4, no filtering was conducted on the proteome database and as such the total CAZy fraction is enlarged (Figure 5.20).

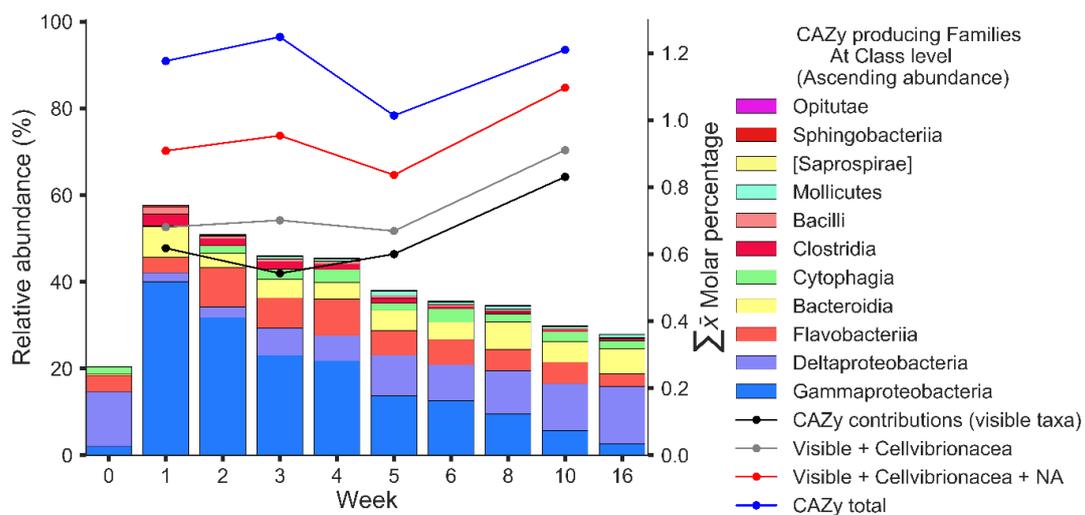


Figure 5.20 Unfiltered CAZy producing families at the microbiome level. Class level resolution.

Decomposition associated microbial community profile

At low taxonomic level resolutions (family), the profile became enriched for CAZy associated taxa, particularly when compared to the sediment outgroup, where only 20.51% of the sedimentary microbiome were CAZy producers, this increased to 57.64% within the first week. The CAZy producing fraction then sequentially temporally receded to 27.97% in week 16, suggesting lower (genus) level taxa may be further enriched, or the degree of carbon cycling (and associated taxa) become reduced (Figure 5.21). This emphasises the need for species level resolution in this type of study.

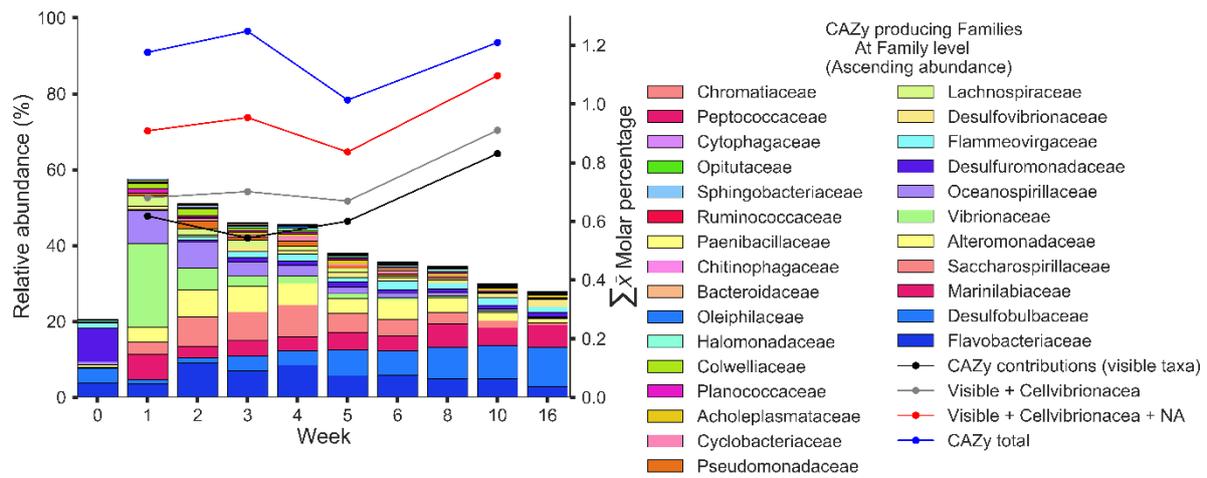


Figure 5.21 Unfiltered CAZy producing taxa at the microbiome level at family level taxonomic resolution. Plots display the sum of the mean molar percentage originating from visualised CAZy producing taxa (bars).

The CAZy producing taxa at the family level that were identified from the amplicon sequencing databases accounted for 74.3, 66.3, 76.6 and 82.9 molar percentage of the CAZy proteomic fraction for weeks one, three, five and ten respectively. The majority of the CAZy enzymes not accounted for in the microbiome were a result of unknown taxonomic homology (lack of taxonomic assignment). The not assigned (NA) fraction accounted for 21.37%, 23.8%, 21.16% and 17.15% CAZy proteome molar percentage for weeks one, three, five and ten respectively. This is expected given the under exploration within salt marsh environments at genomic levels.

The abundance profile appears to correlate with production for the highly abundant taxa, but outliers are apparent where taxonomic groups produce disproportionately large quantities of CAZy enzymes relative to their abundance (Figure 5.22-23). This is particularly apparent for *Cytophagaceae* and *Bacteroidaceae*, whereas *Flavobacteriaceae* were the most active CAZy

Decomposition associated microbial community profile

producing family, and were also the most abundant throughout the time course. It is clear activity does not correlate with abundance.

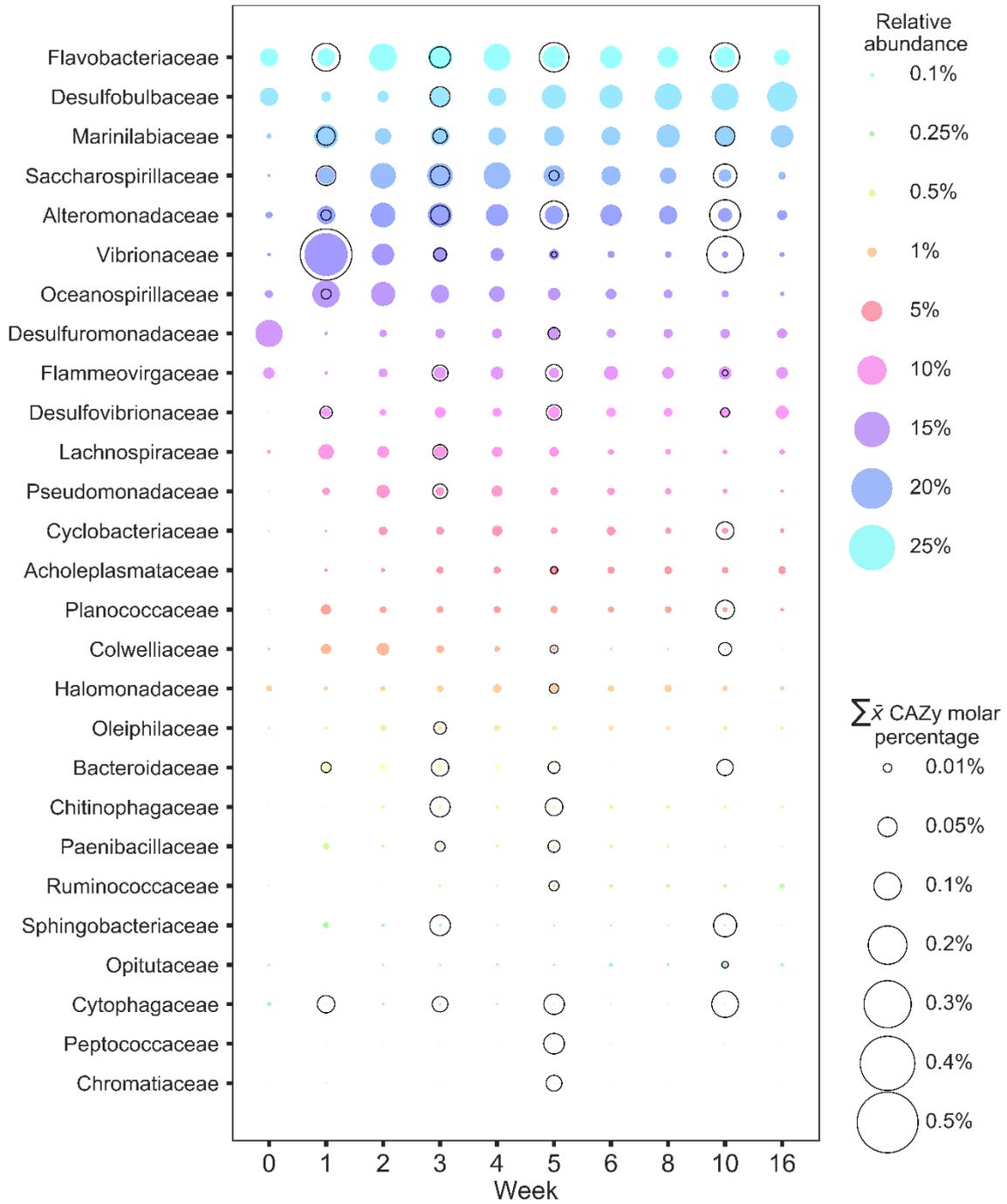


Figure 5.22 Family level resolution of CAZy producers and their respective CAZy contributions. Descending abundance by temporal sum.

Decomposition associated microbial community profile

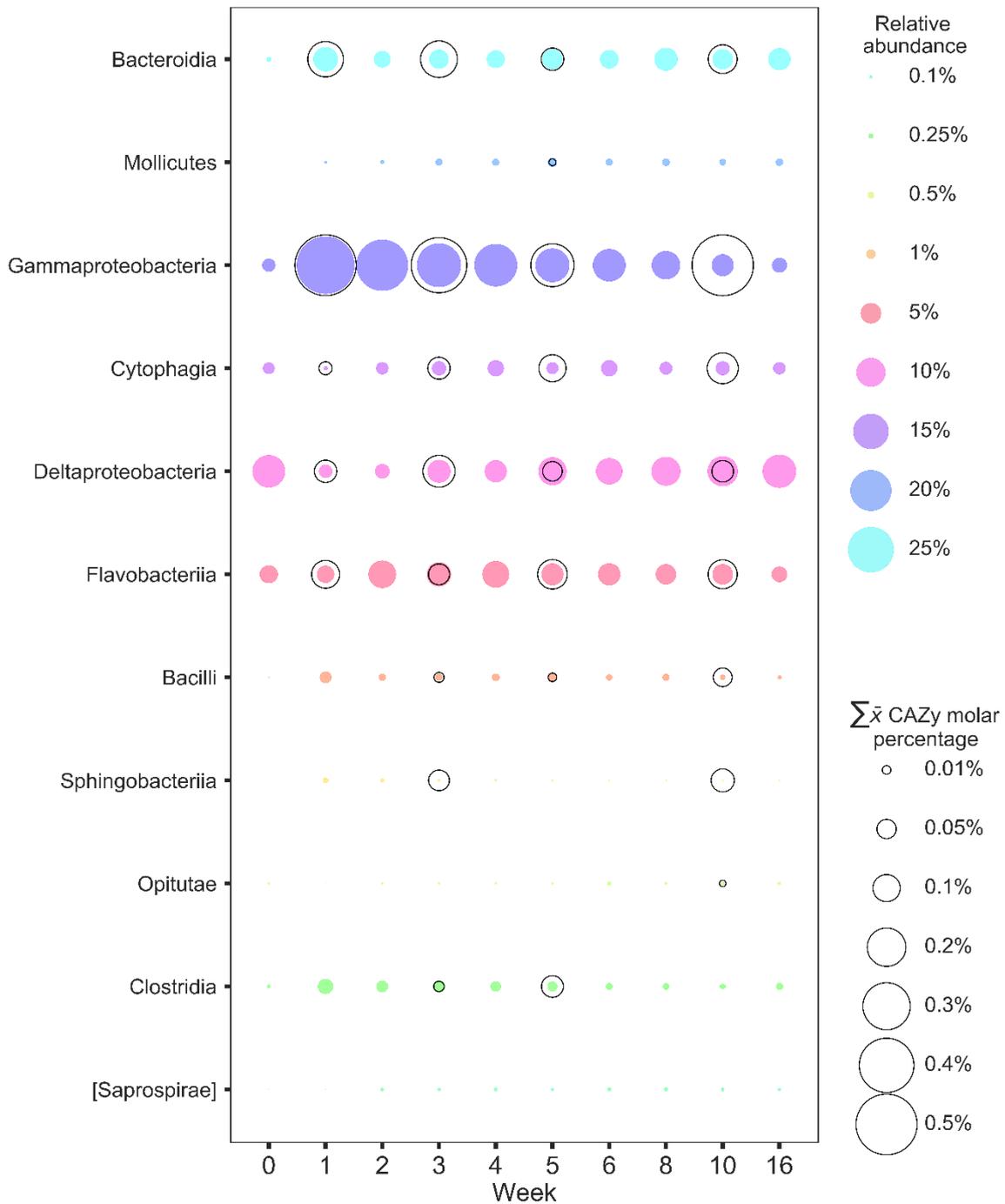


Figure 5.23 CAZY producing families. Displayed at class level taxonomic resolution and their respective CAZY contributions at class level resolution.

5.3.2.2.6.1 Productivity

As CAZY production and abundance was not correlated, a novel method to assess taxonomic activity was undertaken. To explore taxonomic productivity as a function of their abundance a novel productivity index (PI) was devised to assess their relative contribution (equation 5.1). This index

expresses productivity as production activity per unit abundance. Whilst it is true highly abundant taxa, with moderate CAZy expression will feature heavily in lignocellulose construction, the PI highlights those with a relative disproportionate focus which may transfer to a functional ecological role (this can also be observed in Figures 5.22-23). The PI infers functional propensity and does not infer lignocellulolytic proficiency.

Equation 5.1 Taxonomic productivity index (PI).

$$\log_{10} \left(\frac{\Sigma \bar{x} \text{ molar percent}}{\text{Relative abundance (\%)}} \right)$$

A positive productivity index reveals disproportionate CAZy production relative to microbiome abundance. Analysis using the productivity index revealed *Cytophagaceae* (*Cytophagia*), *Chromatiaceae* (*Gammaproteobacteria*), *Bacteroidaceae* (*Bacteroidia*), *Peptococcaceae* (*Clostridia*), *Sphingobacteriaceae* (*Sphingobacteria*), *Planococcaceae* (*Bacilli*) and *Paenicbacillaceae* (*Bacilli*) to be disproportionately more active compared to other members of the cohort (Figure 5.24). This suggests these clades may be more involved or focussed in lignocellulose deconstruction. In particular, *Cytophagaceae*, whose mean abundance reduces from the day 0 outgroup from 0.14% to $0.02 \pm 0.008\%$ concurrently being responsible for the production of 0.02, 0.032, 0.057 and 0.095 molar percentage of CAZy enzymes for weeks one, three, five and ten respectively.

Decomposition associated microbial community profile

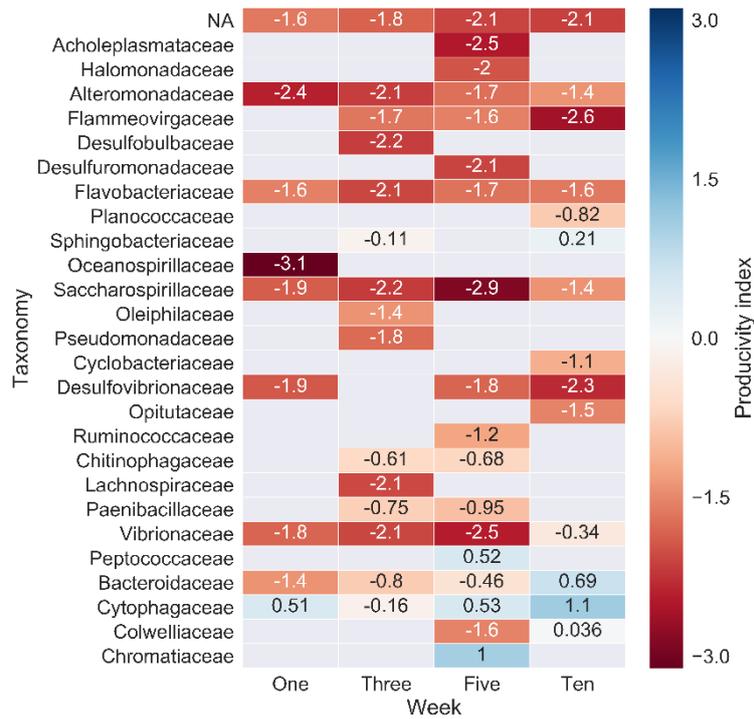


Figure 5.24 Productivity index for CAZY producing taxa at family level resolution.

5.3.3 Fungal profile

Only 48.2% (569 of 1181) of the fungal OTUs were assigned to taxonomic groups, which highlights the magnitude of under exploration in this environment (Figure 5.25A). This creates difficulty in interpreting the fungal profile. Within the characterised fraction, at the phylum level, *Ascomycota* were consistently dominant, with minor contributions in early time points by *Basidiomycota*. The day 0 outgroup also contained *Rozellomycota* and *Zygomycota* (Figure 5.25B).

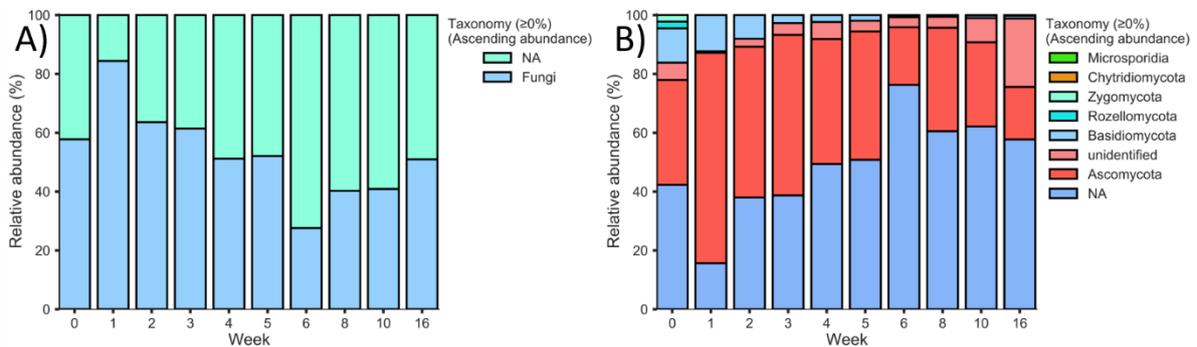


Figure 5.25 Fungi profile at kingdom and phylum levels. A) kingdom level, B) Phylum level.

Decomposition associated microbial community profile

The unassigned fraction decreases in abundance from the day 0 outgroup to biomass, before sequential encroachment in later timepoints. This may suggest uncharacterised marine fungi are involved in lignocellulose decomposition, but as representative sequences are not archived, homology may not have been identified at the proteome or community level.

At the class level, *Saccharomycetes* (which contain the single order *Saccharomycetales*) were the most abundant followed by *Dothideomycetes* (predominantly *Pleosporales* and *Capnodiales*) (Figure 5.26). In the day 0 outgroup, minor contributions were made by *Sordariomycetes* (*Hypocreales*) and *Tremellomycetes* (*Tremellales*). Due to the short length of ITS2 amplicons, and the poor taxonomic coverage in archive databases, family level resolution was non-informative as unassigned OTUs were too abundant (60-90%; Figure 5.27).

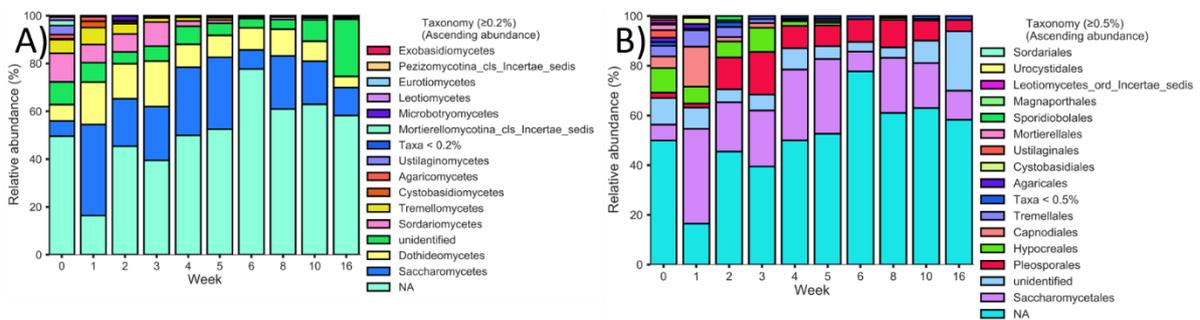


Figure 5.26 Fungi profile at class and order levels. A) Class level, B) Order level.

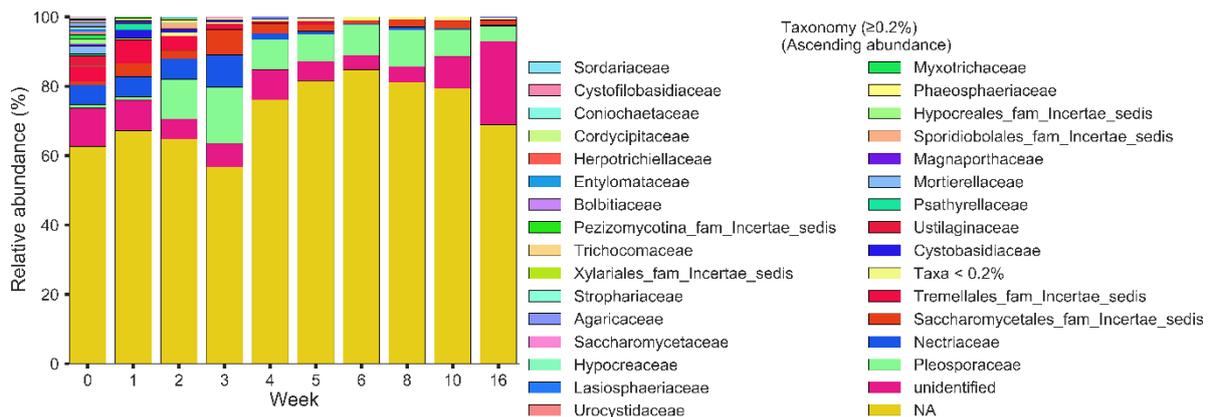


Figure 5.27 Fungi profile at the microbiome level at family level taxonomic resolution.

5.4 Discussion

5.4.1 Bacteria

These results, considered in conjunction with the most likely taxonomic origin of the enzymes identified in Chapter 4, have identified at family level resolution carbohydrate active salt marsh taxa. Complex interplay between taxonomic groups is observed, and enrichment of CAZy producing taxa is apparent. The CAZy associated clades were enriched up to 280.98% in the decomposing *Spartina anglica* biomass compared to levels of abundance in the detrital rich sediments. These results implicate bacterial phyla *Proteobacteria* and *Bacteroidetes* as the dominant producers of lignocellulolytic enzymes, responsible for 47.02% and 34.92% (81.94%) of the observed molar percentage. This is concurrent with the abundance of these two phyla, which is maintained across all decomposition timepoints at $80.51 \pm 1.34\%$.

Within these phyla, varying degrees of enrichment in CAZy producing taxa occur, suggesting these clades are active decomposers. At the class level, *Gammaproteobacteria* and *Deltaproteobacteria* (*Proteobacteria*) were responsible for 39.03% and 7.48% of the total CAZy molar percentage, whilst *Bacteroidia*, *Flavobacteria*, *Cytophagia* and *Sphingobacteria* were responsible for 12.45%, 9.25%, 7.45% and 3.03% of the total CAZy molar percentage respectively. Collectively, these clades were responsible for the production of the majority of CAZy enzymes.

Despite unparalleled levels of productivity within *Gammaproteobacteria*, the decline in abundance is not correlated with their productivity which also decreases (excluding the anomalous timepoint in week 10). The reduction in abundance is offset by an enrichment in CAZy producing clades, particularly *Alteromonadaceae*, *Saccharospirillaceae*, *Oceanospirillaceae* and *Vibrionaceae*. *Alteromonadaceae* and *Saccharospirillaceae* are not abundant in sediments but become major components of the *Gammaproteobacterial* profile.

The rapid colonisation of *Vibrionaceae* and subsequent disappearance; accounting for 22.0%, 5.7% and 2.61% of the total microbiome in weeks one, two and three respectively (this family was undetectable thereafter), is indicative of opportunistic colonisation. Peak abundance during week one precludes a stable decline at the microbiome level and categorises this clade as rapid colonisers and opportunistic oligotrophs. *Vibrionaceae* was predominantly comprised of two genera, *Vibrio* and *Photobacterium*, marine members of both genera are involved in nitrogen fixation (Criminger et al., 2007, Yu et al., 2018), and as evidenced by the high levels of CAZy enzymes attributable to this taxa, they may also dual function as detrital cyclers conferring high ecological importance to this taxa. The

rapid colonisation of this family may be a result of the high carbon to nitrogen ratio in biomass samples, exacerbated by nitrogen limitation in coastal ecosystems, further evidencing a dual function and explaining the observable rapid colonisation. The subsequent recession of *Vibrionaceae* is likely deselection through competition, visualised as enrichment in *Alteromonadaceae*, *Saccharospirillaceae* and *Oceanospirillaceae* and *Bacteroidetes* and *Deltaproteobacterial* groups. The magnitude of enrichment (in respect to the *Gammaproteobacterial* decline) was such that these families were the dominant CAZy producers up to week 5.

Alteromonadales are considered oligotrophic, capable utilising a diversity of compounds, this ability may have facilitated their competitive displacement of the early colonising *Vibrionaceae* (Cho and Giovannoni, 2004, Sun et al., 2015b, Cleary et al., 2017). While genomic level resolution was not fully informative, identifiable abundant genera included *Glaciecola*, *Agarivorans* and *Marinobacter*. Members of these genera are ostensibly aerobic and have been observed as lignocellulolytic (*Glaciecola*), agar hydrolyzing (*Agarivorans*) and capable of hydrocarbon deconstruction (*Marinobacter*) (Klippel et al., 2011, Du et al., 2011, Kostka et al., 2011).

Oceanospirillaceae consisted of two dominant aerobic genera; *Amphiritea* and *Marinomonas*, the former is associated with sponges and exhibits cellulase activity; their abundance profile in this study may implicate them as saprotrophs (Jang et al., 2015). Potential lignocellulose acting enzymes have been identified from *Marinomonas* including putative membrane bound oxidative enzymes (with similarity to laccases and tyrosinases) (Ding et al., 2017, Jimenez-Juarez et al., 2005, Sanchez-Amat et al., 2010).

Saccharospirillaceae could not be resolved to genus level on the *Spartina anglica* substrate, however the day 0 outgroup contained the genus *Reinekea*. *Reinekea* have been observed in abundance on wood in marine conditions and in marine sediments as nitrogen reducers (hence their abundance in sediment, compared to the lignocellulosic substrate) (Kalenitchenko et al., 2015, Romanenko et al., 2004). The CAZy productivity of *Saccharospirillaceae* observed in this study designates them as saprotrophs. Despite this family being extremely poorly characterised (containing only three known species); recently, species within this family have exhibited lignocellulolytic activity, which, when considered in the context of their enrichment and high CAZy productivity, may implicate them as important salt marsh saprotrophs (Fidalgo et al., 2017, Labrenz et al., 2003). In contrast to the other *Gammaproteobacterial* families, *Saccharospirillaceae* range from facultative anaerobes to obligate aerobes (Shahinpei et al., 2014, Fidalgo et al., 2017).

Decomposition associated microbial community profile

Beyond week 5 and during the later timepoints, *Gammaproteobacteria* is supplanted at the microbiome level declining from in week one to week 16. This may be due to competition by *Bacteroidetes* (who exhibited little temporal variation in abundance) particularly the highly abundant CAZy producing families *Flavobacteriaceae* (*Flavobacteriia*), *Marinilabiaceae* (*Bacteroidales*) and *Flammeovirgaceae* (*Flavobacteria*) which increase in abundance during the first two weeks before stabilising. *Gammaproteobacteria* were supplanted in later timepoints (where *Bacteroidetes* abundance was stable) by *Deltaproteobacteria* which consistently increased in abundance throughout the time course and was composed of *Desulfobacterales* and *Desulfuromonadales* orders and the CAZy producing families *Desulfobulbaceae*, *Desulfomonadaceae* and *Desulfovibrionaceae* were dominant. This phylum was dominated by the single family (>50%) *Desulfobulbaceae*. *Desulphobacterales* are an anaerobic lineage with previous connections to both sulfur cycling and carbon cycling, the latter of which is evidenced in this study (Cleary et al., 2017). Interestingly within salt marshes the relative abundance of *Desulfobacterales* increases with depth, suggesting this family may be important for late stage carbon cycling in deeper sediments (Cleary et al., 2017).

As the experiment was conducted at the surface sediments; the increased relative abundance of the anaerobic *Deltaproteobacteria*, the initial increase and stabilisation of *Bacteroidetes* (which consist of the aerobic and anaerobic lineages) and the consistent decline in the aerobic *Gammaproteobacteria* suggests the community was becoming increasingly oxygen depleted throughout the experiment. This is further evidenced by the increase in the respiratory plastic (facultatively anaerobic) *Saccharospirillaceae* (*Gammaproteobacteria*). This likely occurred as a function of microbiome growth increasing oxygen requirements that is limited by the small diffusible surface area, further limited by tidal inundation cycles in a predominantly anaerobic salt marsh.

Genus level identification for the CAZy producing *Bacteroidetes* families could not be performed. However, *Flavobacteriia* are known to deconstruct plant material which is concurrent with their level of lignocellulolytic enzymes in this study (Lydell et al., 2004, Cleary et al., 2017). *Flavobacteriia* contain both aerobic and anaerobic representatives; however, the dominant CAZy producing family is predominantly aerobic (however salt marsh lineages are largely uncharacterised and may be divergent) (McBride, 2014). *Marinilabiaceae* were also pivotal during lignocellulose deconstruction and are obligate anaerobes (Rosenberg, 2014). Interestingly, both *Flavobacteriaceae* and *Marinilabiaceae*; identified in this study to be highly productive carbon cyclers, were also enriched in deep sea sediment subject to phytodetrital supplementation (Hoffmann et al., 2017).

At later time points, the previously stable CAZy producing taxa within *Gammaproteobacteria* declines, however the molar percentage of CAZy enzymes with *Gammaproteobacterial* homology

peaks. This may also have been due to an enrichment in the (*Cellvibrionales*) *Cellvibrionaceae* family (which were third most active clade at the microbiome level, Chapter 4), which were not observed due lack of classification in the archive databases in this study (Spring et al., 2015). This family may have outcompeted the other taxa (within *Gammaproteobacteria*).

The productivity index revealed *Firmicutes* families to be highly active, despite low apparent abundances; particularly the families *Peptococcaceae* (*Clostridia*), *Planococcaceae* (*Bacilli*) and *Paenicbacillaceae* (*Bacilli*). This is insightful as *Firmicutes* was not abundant in sediment samples (0.41%), becoming slightly enriched on the decomposing biomass ($4.59\% \pm 0.61$). This finding may implicate the function of these families as active detritus decomposers. Representatives of all families have hydrocarbon degrading representatives, which is thought to be a good predictor of lignocellulolytic (specifically ligninolytic) activity (Beazley et al., 2012). *Peptococcaceae* are thought to be responsible for ring cleavage events in vanillate (a lignin derived aromatic) (Kato et al., 2015). These capabilities are in accordance with *Firmicutes* preference for deeper sediments where they are observed to be significantly more abundant (deeper sediments are rich in recalcitrant aromatics), which could also explain their low abundance during the surface level experiment described in this Chapter (Petro et al., 2017).

Clostridia (*Peptococcaceae*) were more abundant than *Bacilli*, the latter are aerobic or facultive anaerobes where *Clostridia* are often obligate anaerobes and saprotrophs (Sharmin et al., 2013, Riley, 2012). The abundance of *Clostridia* within *Firmicutes*, and the minimal abundance of *Firmicutes* in the surface level outgroup sediment reinforces the oxygen depleted conditions within the substrate. The propensity for these lineages to degrade aromatics whilst respiring anaerobically and their apparent enrichment on lignocellulosic substrate *in situ* make these families an attractive target to elucidate anaerobic lignin deconstruction.

Many *Bacteroidetes* families exhibited high productivity indexes, including *Cytophagaceae*, *Bacteroidaceae* and *Sphingobacteriaceae*. At the microbiome and family level, CAZy producing members of *Bacteroidetes* comprised only 5.6% of the total abundance in contrast to the *Proteobacterial* cohort which comprised 14.66% of the total. The low abundance of *Bacteroidetes* in consideration with their high productivity indexes would suggest this cohort to be more specialised and targeted toward detrital cycling. These results also suggest, a maximum of 20.5% of the total microbiome in the upper salt marsh sediments at Welwick can function as carbon cyclers. This aligns well with the rich organic content of Welwick sediments (2-4%) (Andrews et al., 2008).

Decomposition associated microbial community profile

Verrucomicrobia were disparately abundant relative to their level of CAZy productivity. In this study, only *Optuaceae* were identified, whereas proteomic homology to *Rubritaleaceae* was observed, which were active CAZy producers. Representatives of these families may be highly divergent to known representatives and require further investigation. This is one source of misalignment of community to productivity.

Accounting for the highly productive but taxonomically unidentified *Cellvibrionaceae* (*Gammaproteobacteria*) family, $63.82 \pm 4.37\%$ of CAZy molar percentage was assigned to identified taxa at family level resolution, whilst proteins with no known homology accounted for $17.88 \pm 1.15\%$. The disparity likely arises from unclassified taxa in archive databases being omitted in this analysis, or a result of differential classification nomenclature between databases. Additionally, the most likely taxonomic origin of these proteins may have been mis-assigned, which is likely given the level of unknowns in salt marsh ecosystems. Uncharacterised organisms will be assigned to the most homologous taxa which may be evolutionarily divergent its native salt marsh counterpart. This highlights the inadequacy of archive databases for this under explored environment and reveals the magnitude of divergence of organisms within this environment. This could also be a function of divergence at the protein level, accentuating the biotechnological potential of this environment for highly divergent, novel enzymes. An interesting observation in this study was the enrichment on biomass of the largest *Bacteroidetes* group, the *SB-1* family, containing unclassified *Bacteroidetes*, suggesting unknown lineages may function as active saprobes. However, this level of alignment within the microbiome in this study was sufficient to elucidate and conclude functional taxonomic groups attributable to carbon cycling within the studied salt marsh.

To conclude, while *Proteobacteria* and *Bacteroidetes* dominate carbohydrate deconstruction in surface sediments of salt marsh ecosystems, minor contributions were made by *Veccuromicrobia* and *Firmicutes*. Native *Verrucomicrobia* appear highly divergent to current archived information and *Firmicutes* likely feature as carbon cyclers at great sediment depths in the aromatic rich sediments. The CAZy producing *Proteobacterial* representatives described here appear to be early colonisers of lignocellulosic substrates, that undergoes taxonomic restructuring to favour anaerobic heterotrophic lineages (likely driven by spatiotemporal oxygen depletion), whilst concurrently competitively deselected by putatively respiratory plastic *Bacteroidetes* clades. Given the propensity of lignocellulosic biomass in salt marshes to form dense, static aggregates, the *in situ* methodology utilised in this experiment is likely a true (small scale) representation of the microbial dynamics within. The abundance of CAZy producing taxa within the sediments, the higher activity and abundance (in earlier timepoints) of *Proteobacteria*, suggest the families described here are the

predominant surface level carbon cyclers. The respiratory plasticity of the families identified within *Bacteroidetes* suggests they are highly active at surface levels, but likely dominate carbon cycling in the oxygen depleted cores of biomass aggregates, and in shallow sub surface sediments. The families associated with carbon cycling in this study, identified from a healthy, accreting salt marsh with sediments invariably rich in organic carbon should assist in status assessments to future microbiome and microbiological studies.

5.4.2 Fungi

The lack of enzyme homology to known fungal proteins suggest fungi are not dominant decomposers in this environment. However, the magnitude of unassigned operational taxonomic units (OTUs) may also indicate that representatives of potential lignocellulose associated taxa are completely uncharacterised and were not detectable using the methodologies of this study. The known fungi within the biomass are highly abundant and include the classes (family); *Saccharomycetes* (*Saccharomycetales*) and *Dothideomycetes* (*Pleosporales*) both of which appeared enriched within the decomposing biomass. *Saccharomycetales* are unicellular yeasts well characterised as both plant pathogens and saprobies (Suh et al., 2006).

The order *Pleosporales* consisted primarily of the family *Pleosporaceae* ($95.15 \pm 0.92\%$) and *Phaeosphaericeae*, which constituted $0.97 \pm 0.21\%$ of the fungal fraction of the microbiome. *Pleosporaceae* were the most abundant family on the decomposing biomass ($9.47 \pm 1.22\%$ from week two through to 16) and are considered saprobic, which would explain their enrichment (Kruys et al., 2006, Zhang et al., 2009). However, the lack of lignocellulolytic enzymes originating from these families would dispute this function. This family has been strongly associated with sugarcane leaves indicating they may function predominantly as plant pathogens (de Souza et al., 2016). Marine originating isolates of this family have exhibited lignocellulolytic activity (Hong et al., 2015).

Phaeosphaericeae are plant pathogens notorious for assisting the autumn dieback of *Spartina spp* (Buchan et al., 2003). It is interesting to note that signs of fungal pathogenic infections on *Spartina anglica* during collection were visible, as is common during winter dieback (Buchan et al., 2003). *Phaeosphaeria* have been observed to be dominant in salt marsh sediments but inferred to be saprotrophs (Dini-Andreote et al., 2016b).

The *Sordariomycetes* family *Nectriaceae* (order *Hypocreale*) were highly abundant in the day 0 sediment (5.64%) and were abundant in the biomass samples up to week three (9.83%) before declining and not recovering ($0.69 \pm 0.23\%$). This family is poorly characterised due to lack of

genomic resources and 79.74% of the assigned abundance in this family could not be attributed to a taxonomic group in the sediment outgroup (Lombard et al., 2015). *Fusarium* was the dominant genus in this family, representing $80.64 \pm 3.0\%$ of all *Nectriaceae* during week one to six, before receding to encroachment by the genus *Gibberella* and unidentified taxa. While species level resolution was not resolved, in the context of salt marshes, both genera have been identified as plant pathogens capable of infecting *Spartina spp.* within salt marshes (Elmer and Marra, 2011).

Tremellomyces (mainly the genus *Cryptococcus*) followed a similar pattern to *Nectriaceae* of the class *Sordariomycetes*. They were transient and did not differ in abundance from the sediment outgroup before becoming an insignificant component of the decomposing *Spartina anglica* microbiome. *Cryptococcus* are thought to be saprotrophs (Dini-Andreote et al., 2016b), as such their disappearance from the lignocellulosic substrate is counter intuitive. One possible reason is that they are present as residual dormant spores on the *Spartina anglica* biomass collected during the early phase decay. Pathogenic *Ascomycetes* on *Spartina spp.* are known to intensively sporulate during standing decay (when the biomass for this study was collected) which supports this theory (Newell, 2001).

The proportion of unassigned OTUs in this study predicates the level of under exploration in this environment and affirms the potential for biotechnological novelty. In the current study, the role of fungi from the identified fraction appear to be plant pathogens. The putative saprotrophic clades appear enriched within the biomass; however, no biomass-only analysis was performed. The observed 'enrichment' of saprotrophic clades is likely a consequence of residual spores from the collected standing biomass as rapid recession of these clades follows (the spores likely decomposed). The lack of homology to fungal enzymes, despite the abundance of putative saprotrophs (likely as a function of the comparative scarcity of fungi to bacteria (orders of magnitude less) conferring levels of proteins below detectable limits in this study), infers fungi are not major contributors to the detrital cycle in upper salt marsh sediments.

These result in this study suggest sediments contain a biological reservoir of plant pathogens (likely in spore form), that are either mischaracterised as saprotrophs or inactive (as dormant spores) and known to act upon salt marsh flora, resulting in the introduction of carbon into the environment. Therefore, fungi may play a greater indirect role in salt marsh carbon cycling, through the introduction of lignocellulosic biomass into the environment, but the evidence presented in this study suggests they are negligibly active during the subsequent marine *in situ* decomposition (standing biomass is not fully submerged and can not be considered truly marine). However, there may be a very minor (undetected) role of yeasts (e.g. *Saccharomycetales*) in this process, yet,

proteomic evidence was absent. This challenges current theory that fungi are important decomposers in salt marsh ecosystems (Torzilli et al., 2006, Dini-Andreote et al., 2016b, Walker and Campbell, 2010), but is supported by recent observations *in vitro* studied reporting competitive deselection of marine fungi by bacteria (Cortes-Tolalpa et al., 2018).

The function of fungi in salt marsh sediments, given the incredible levels of diversity described here and elsewhere (Dini-Andreote et al., 2016b), in conjunction with the magnitude of unclassified fungal lineages remains to elusive. Fungi may contribute to salt marsh carbon cycling through pathogenic action; however, *Spartina anglica* is perennial and the biomass would be introduced irrespective of fungal infection. The level of unclassified and mischaracterised fungi (at genomic and isolate levels) in this ecosystem is the main barrier for further elucidation of their function in this process. A further area of research would be to determine the factors inhibiting fungal contributions to carbon cycling in marine ecosystems when they appear to dominate the terrestrial carbon cycling landscape.

Chapter 6 Expression and characterisation of putatively lignocellulolytic enzymes

6.1 Introduction

Proteins ostensibly involved in lignocellulose deconstruction as outlined in Chapter 4 during an *in situ* degradation experiment of biomass from the salt marsh cordgrass; *Spartina anglica*, in Welwick salt marsh were identified using bioinformatic and proteomic techniques. This information was broadly categorised into system level and more precisely protein level analyses which were used in conjunction to determine a list of protein targets to take forward for further analysis; including protein expression in a heterologous host, purification and characterisation.

It was inferred that abundance at the system level translates to relative importance *in vivo* which could be replicated within a synthetic enzymatic cocktail for biomass processing. The system level profile elucidated two major and one minor Euclidean clusters containing temporally dominant glycoside hydrolase families GH3, GH5, GH6, GH9, GH10, GH11 and carbohydrate esterase family CE1 (section 4.3.2.1). Cross referencing these families with carbohydrate binding domains (CBMs) associated within their domain architecture revealed relationships with predominantly cellulose and hemicellulose targeting CBMs. Sequences associated with CBMs with an affinity for cellulose binding were ascribed importance. Other enzyme classes were also included due to novelty despite their relatively low molar percentage or including auxiliary activity (AA) families AA2 and AA9.

From these select families, all sequences annotated above the e value threshold from the dbCAN database of $1e^{-10}$ where whole open reading frames (ORFs) were present, they were inspected for the presence of signal peptides; sequences without an identifiable signal peptide were discarded. Most likely taxonomic origin for proteins was used to discard proteins with potential eukaryotic origin, due to their low abundance at the system level likely translating to negligible contributions to biomass deconstruction and the analogous difficulties associated with eukaryotic protein expression. One exception to this included one ostensible dbCAN and pFAM annotated AA9 and GH61 respectively, which was taken forward due to its low maximum percentage identify (36% to NCBI NR_prot) and qualifying confidence scores spanning a significant length of the sequence.

For groups with underrepresented full length ORFs or no observable signal peptide, supplementary sequences were identified from the transcriptomic database that was not identified in the proteomic profiles. This was necessary for GH3, GH9 and AA9 families where one, two and no sequences passed selection criteria respectively. Where applicable within the remaining sequences

subsets were created for each group, determined by Hidden Markov dbCAN and pFAM confidence scores and the maximum identity to the NCBI NR_prot database with the aim of offsetting inherent risk associated with novel, highly divergent sequences with more homologous sequences with predictable activities. Additionally, a selection of single catalytic domain ORFs and ORFs with catalytic domains in conjunction with carbohydrate binding domains were also selected. A total of 42 targets were shortlisted for amplification including ten GH3s, 11 GH5s, six GH6s, six GH9s, one GH11, one GH16, three CE1s, one CE6, two AA2s and one AA9.

To characterise putative enzyme targets from 'omic databases where sequences predominantly originate from unknown, uncultivated organisms, they must be recombinantly expressed in a heterologous host and isolated. As the target shortlist consists of sequences with most likely prokaryotic origin (excluding the AA9), an *Escherichia coli* expression system was deemed suitable. Interestingly, many of the selected sequences contained multiple predicted disulfide bonds and all sequences contained greater than 10% rare codons which are considered problematic in *E. coli* expression systems (Rosano and Ceccarelli, 2014). To alleviate these potentially confounding factors, a strain of *E. coli* containing both the pRARE2 plasmid (encoding seven tRNAs for rare codon transcription) and double silenced (thioredoxin reductase and glutaredoxin) *trxB* and *gor* which act to create and maintain an oxidising cytoplasm, where it is usually reducing and consequently inhibits disulfide bond formation (de Marco, 2009). This strain was used with a cytoplasm targeting expression vector with a highly specific N-terminal tag, in a range of physiological and induction conditions to control and optimise enzyme production and achieve isolation of a high purity protein for further experimentation (Kimple et al., 2013).

Two enzymes were selected for in depth characterisation delineated by the ORF descriptors a_c570829_g1_i1_1 and c_c892255_g1_i1_1 (Figure 6.1). The former exhibits a novel domain architecture containing predicted (with high confidence across databases) GH6, CBM2 and CBM44 domains all of which belong to the most abundant enzyme category at the system level. This protein was also the most abundant CAZy in two time points (week one and ten). The latter is categorised by a single GH5 domain with loose homology to current databases with a maximum identity in the NCBI NR_prot database of 45%.

Expression and characterisation of putatively lignocellulolytic enzymes

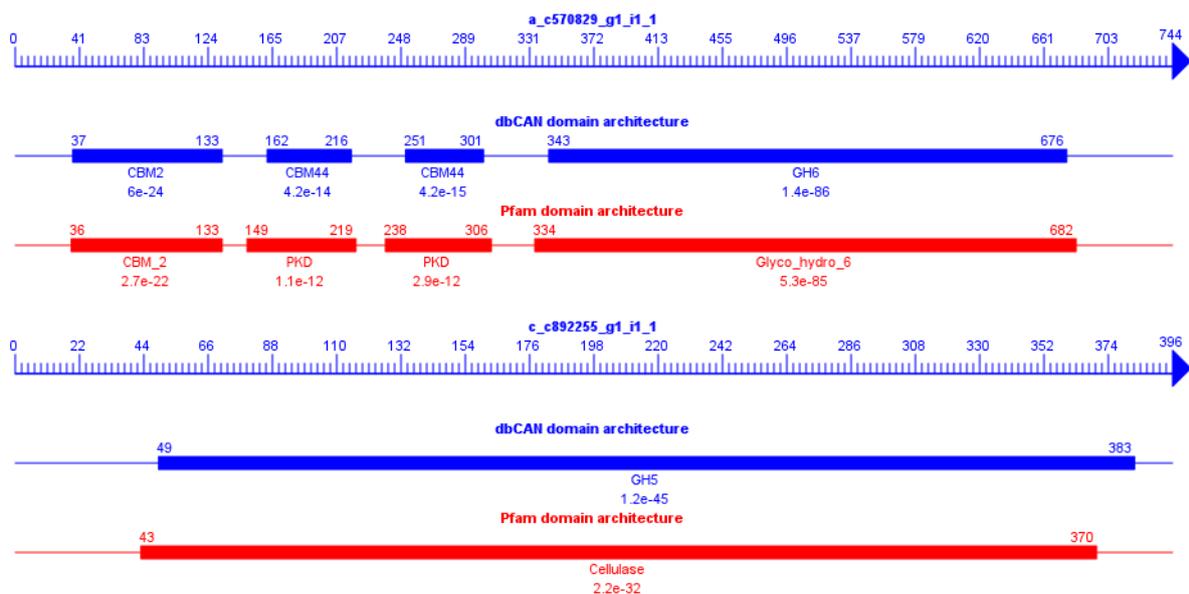


Figure 6.1 Domain architectures of *a_c570829_g1_i1_1* and *c_c892255_g1_i1_1*. CAZy annotated domains are represented in blue, pFAM annotations are represented in red, Hidden Markov confidence values are displayed below their respective domain.

6.2 Methods

6.2.1 cDNA synthesis

Messenger RNA (mRNA) obtained in Chapter 4 (section 4.2.1.1-2) was reverse transcribed to complementary DNA (cDNA) using a SuperScript™ III One-Step RT-PCR System kit (Invitrogen) according to the manufacturer's protocol.

6.2.2 genomic DNA preparation

Genomic DNA (gDNA) was used as a template for amplification of target sequences. The gDNA used in these experiments was obtained in Chapter 5 (section 5.2.1.1).

6.2.3 Plasmid preparation

The pET-52b(+) vector (Novagen) was linearised with PCR (Table 6.1-6.3), the resultant linearised amplicon was deficient of the multiple cloning site, thrombin recognition sequences and histidine

repeats. The remaining sequence contained the start codon, (N-terminal) streptavidin II tag (WSHPQFEK), HRV 3C recognition sequence (for tag cleavage) and a native stop codon. Due to the sensitivity of the InFusion cloning method to background vector, incomplete and non-linearised vector (which originated from *dam+ / dcm+* Stellar™ cells) were removed by treating with 1 unit DPN1 (NEB), compatible with Phusion HF buffer for 90 min at 37°C. The treated amplicon was run on an agarose gel and the band conforming to 5128bp was excised and purified using NucleoSpin® Gel and PCR clean up kit (ClonTech) according to the manufacturer's protocol with two deviations; the loading NTI buffer was reloaded onto the column once and the elution buffer was heated to 70°C and incubated in the column for 5 min to maximise recovery.

Table 6.1 Primers utilised for plasmid linearisation and confirmation of transformation.

| Target | Primers (5'-3') | Size (bp) |
|--------------------------|----------------------------------|-----------|
| pET-52b(+) linearisation | F <u>GGGTCCTGAAAGAGGACTTCAAG</u> | 5128 |
| | R TAATTAACCTAGGCTGCTGCCACC | |
| T7 insert confirmation | F <u>TAATACGACTCACTATAGGG</u> | Variable |
| | R GCTAGTTATTGCTCAGCGG | |

Target sequence forward primers were designed to eliminate the native signal peptide as a cytoplasmic expression system was chosen. The primer was designed to begin at the end of the predicted signal peptide (see section 6.2.10.9). Target sequences were amplified with primers containing suffixed overhangs complimentary to the 5' (forward primer) and 3' (reverse primer) ends of the linearised vector compatible with InFusion cloning. The overhangs were designed to insert the fragment of interest after the HRV 3C site and stop codon (position 209 and 110 on the circular plasmid). This primer design facilitated a single amplification to vector insertion methodology.

Expression and characterisation of putatively lignocellulolytic enzymes

Table 6.2 Primers for amplification of selected proteins. :OF represents the InFusion forward overhang sequence (5'-3') CTCTTTCAGGGACCC, :OR represents the InFusion reverse overhang sequence (5'-3') CAGCCTAGGTTAATTA.

| ORF | Index | Origin | Primer (5'-3') | Size (bp) |
|---------------------------------------|-------|--------|--|-----------|
| a_c570829_g1_i1_1 (GH6) | 1 | P | F GCATCTTGTTTCGATTTCATTGA:OF R TTCCGCAGGAGGACCTG:OR | 2158 |
| c_c987261_g2_i1_1 (GH5) | 2 | P | F GAAACTAAGGAAAACAAAGAGAAAACG:OF R TTGAGTTATTATTTTAATAATTGCAGTATCT:OR | 1132 |
| a_c577269_g3_i1_1 (GH5) | 4 | P | F ACTTTAGGTTTCGCAACTCAGC:OF R TTTTCCAGTTAGGATTGCTCTGA:OR | 1735 |
| a_c566843_g4_i1_1 (GH5) | 5 | P | F GAATGTACTIONTATGAAGTGCAAAGTGA:OF R TTTGACTACAGCTGGGTTTGC:OR | 1108 |
| a_c580607_g3_i1_2 (GH5) | 6 | P | F GTCGACGATGCTCGTCAAG:OF R CAGATAACAACCCGCATTATTAC:OR | 1780 |
| c_c991288_g2_i1_1 (GH5) | 7 | P | F AATAAAAACAAAACAGAACAAAAGGA:OF R TTTTCTTACAAGTATATGTATTAGTGTCTGATC:OR | 2215 |
| c_c892255_g1_i1_1 (GH5) | 8 | P | F GGAAATGGAAAAGAAGAACTAAA:OF R TCTTGTTAGAATATTTATAATAGCGGTATCG:OR | 1126 |
| c_c1007998_g3_i1_1 (GH3) | 9 | P | F CAAAATTATCCTTATTTAAATCCAAACC:OF R TAGAGAACTTAAACTAATCTTTATTGATTTCA:OR | 1153 |
| a_c582689_g1_i1_4 (CE1) | 10 | P | F GAAAATGCGATTATCAATGGC:OF R TGGCAGATCCGATGTTCTC:OR | 2572 |
| c_c1011459_g1_i1_2 (CE1) | 11 | P | F CAAGAAACTCACGTCGAACT:OF R GAAGCTATTAAAGAAATCCAAGCA:OR | 1351 |
| c_c1011459_g1_i2_2 (CE1) | 12 | P | F CAAACTGACATCGAACAAAA:OF R GTTAGCCAGATTTTTGAAGCCA:OR | 2671 |
| d_TRINITY_DN1293433_c62_g2_i4_3 (CE6) | 13 | P | F GCGCCAGACCCCACT:OF R TCGACCGAGATAATTCATCAG:OR | 2674 |
| a_c580900_g1_i2_1 (GH3) | 14 | P2 | F AGTGAAGGTGTACAAGGACCC:OF R TTCAACTATTACTGTAAATACTTGTGAGTATG:OR | 2770 |
| c_c991288_g2_i3_1 (GH5) | 15 | P2 | F ACTGCAAACATCGAAACATTG:OF R AATAAAAACAAAACAGAACAAAAGGA:OR | 3751 |
| a_c561927_g1_i2_1 (GH5) | 16 | P2 | F AGTTGTGATTATCAAGTTGGAAACG:OF R TTGCCATAGTCTTTCAAGCAA:OR | 1138 |
| b_TRINITY_DN1505951_c8_g1_i2_1 (GH5) | 17 | P2 | F GGCACCGCACAACCAG:OF R TTCAGCGCTACACGTTGGT:OR | 1723 |
| a_c571186_g1_i2_1 (GH5) | 18 | P2 | F GGAACCGCACAACCAGG:OF R TTCAGCGCTACACGTTGGT:OR | 2095 |
| a_c578840_g1_i6_1 (GH9) | 20 | P2 | F GAACCCACATTAATCTTTAACGG:OF R AGATTGCTTACGCTTACGTAAAAA:OR | 2941 |
| b_TRINITY_DN1290194_c0_g2_i1_1 (GH16) | 21 | P2 | F AGTTTGTTTTTGTATGTGTAGGTGG:OF R TTGACTCACCAACATTCTAAATTTTT:OR | 934 |
| a_c579021_g1_i2_1 (AA2) | 22 | P2 | F TCAGAAGAACCAAAATCCAATAA:OF R CGAACGTTTTAGATCAAACGG:OR | 2179 |
| c_c558329_g1_i1_2 (AA9) | 23 | T | F CACGGATCCCTGCAGCTAC:OF R CTCCAAGTCTTCAAAGCGCT:OR | 958 |
| c_c1000477_g2_i2_1 (GH6) | 24 | T | F GCGCTATGCGAATACGTCTG:OF R TGGCGCTGGTGGACC:OR | 2155 |
| a_c537502_g2_i1_2 (GH6) | 25 | T | F GCATGTGAATATAGTTTGACAAGCA:OF R TTCTGGTTTACCTGGTGACG:OR | 1894 |

Expression and characterisation of putatively lignocellulolytic enzymes

Table 6.1 continued.

| ORF | Index | Origin | Primer (5'-3') | Size (bp) |
|---------------------------------------|-------|--------|--|-----------|
| a_c579709_g1_i2_1 (GH6) | 26 | T | F TTGGTGTCTCAGCGTGTAAATGG:OF R GTCTTTATCGTAGTCGCTGTCA:OR | 1903 |
| d_TRINITY_DN1312166_c0_g1_i4 (GH3) | 27 | T | F ACCGAAGTCTGGAAGGACAA:OF R TGGCTCCACAACCGTCA:OR | 2569 |
| a_c566242_g1_i1 (GH3) | 28 | T | F GCCAGTTTTTTTGACAACAGAGG:OF R AAGAATCACTGCATCAGCATCT:OR | 1702 |
| b_TRINITY_DN1506521_c17_g3_i3_3 (GH3) | 29 | T | F CAGAGGAAGAAAATTGATCCG:OF R ATTTACTACATTTCAGATAGGCTTCT ATATACTCC:OR | 2311 |
| c_c1013947_g1_i2_4 (GH3) | 30 | T | F GCAAATAAACTTGCCGGAAT:OF R ATCCAAATTCAAATAGAACCTTGC:OR | 2224 |
| c_c1013947_g1_i4 (GH3) | 31 | T | F GCAATTTTGACATTTAAATGGAT:OF R ATCCAAATTCAAATAGAACCTTGC:OR | 2374 |
| c_c988814_g2_i1_2 (GH3) | 32 | T | F TTGTTTGCCAGCAATACC:OF R CGGCATACGGCAGGTAG:OR | 1648 |
| c_c928629_g1_i4_2 (GH3) | 33 | T | F GAAAACCTACCCACTTCTAAAAATGA:OF R GCTCAGTTGATTGCCGG:OR | 1834 |
| a_c554459_g1_i3_1 (GH9) | 34 | T | F GCAGGTAAGTTTGTGCTGAA:OF R AATTATTAATAAACTAATTCTATTTT:OR | 1783 |
| b_TRINITY_DN1460637_c1_g1_i2_1 (GH9) | 35 | T | F AACCCACAAAGTTATACCCTTGA:OF R TTTTATCAACAAGGTCATGTAGCTC:OR | 2173 |
| b_TRINITY_DN1476303_c2_g1_i3_1 (GH9) | 36 | T | F ACCACCGAAAATAGTAGCTCTGC:OF R GTCGGCGTTTTGTACGG:OR | 1846 |
| c_c1010610_g1_i2_3 (GH9) | 37 | T | F ATTTTTAGCCTGAATACTATTGGTTATTT:OF R CGGATTTACAAAACCGGC:OR | 1597 |
| b_TRINITY_DN1488318_c3_g3_i5_1 (GH11) | 38 | P2 | F CAGACAATAAATGGTAGTAGCACAGG:OF R CCATTGTATCTCACAGTCTTGACTT:OR | 1816 |
| c_c993151_g1_i1 (GH3*) | AS4 | T | F CAACCTTCTAATCAAGCTAAACAAA:OF R ATTTTGGATAATTCTGAACGTATCG:OR | 2449 |
| a_c561927_g1_i1_1 (GH5) | AS7 | P | F AGTTGTGATTATCAAGTTGGAAACG:OF R TTGCCATAGTCTTTCAAGCAAGT:OR | 1735 |
| c_c1005042_g2_i1_1 (GH6) | AS12 | P2 | F GCGACATGCGAGTATTTTATTAGC:OF R ATTAGCAGGAGGTCTCTGCTGG:OR | 2155 |
| c_c1008383_g1_i3_2 (GH6) | AS15 | T | F GCTTGTAGCCCAGATGAAAC:OF R TAGTGCAGGGTAAGCGTTG:OR | 1666 |
| c_c946160_g2_i1_4 (GH9*) | AS19 | P | F TTTGATGATTCAAACAATGGC:OF R TTCAGAACCATATTCCAATACTGAAG:OR | 3367 |
| a_c579021_g1_i2_1 (AA2) | AS20 | P2 | F TCAGAAGAACCAAATCCAATAAG:OF R CGAACGTTTTAGATCAAACCGG:OR | 2179 |
| c_c558329_g1_i1_2 (AA9) | AS23 | T | F CACGGATCCCTGCAGCTAC:OF R CTCCAAGTCTTCAAAGCGCTC:OR | 958 |

Table 6.3 Nested primers for amplification of selected sequences.

| ORF | ID | Origin | | Primer (5'-3') | Size (bp) |
|---|----|--------|--------|--|-----------|
| d_TRINITY_DN1293433 _c62_g2_i4_3 (CE6) | 13 | P | F R | AAACTGCCACCGAAAAATTG CATTATGCTCCCACCGTTT | 3002 |
| c_c1000477_g2_i2_1 (GH6) | 24 | T | F R | CTCTCCGCATTAGCGTTAGG TTCCGATCTCTCTGCTTGGT | 2552 |
| c_c928629_g1_i4_2 (GH3) | 33 | T | F R | TCGACAGTTTTGCTGTTTCG CGGCCTTAGTTTCGAGGTT | 2427 |

Targets were amplified from genomic DNA with the exception of the AA9 of eukaryotic origin which was amplified from complimentary DNA. Upon PCR condition optimisation when visualisation of single bands of interest were obtained they were directly cleaned else they were excised from an agarose gel and cleaned (see section 2.2 - 13% NTI buffer protocol). These fragments were inserted into pET-52b(+) following the manufacturer's protocol (InFusion, ClonTech). The resulting vector was then transformed into Stellar™ ultra-competent cells (ClonTech) using a heat shock methodology with a vector to insert ratio of 1:X where X is 0.33 multiplied by the length of the insert in kb (a 1kb insert would have a vector to insert ratio of 1:1). The cells were spread onto LB agar plates containing 50µg/mL streptomycin and grown for 16 hours at 37°C. Colonies were inspected for successful insertions of the fragment of interest by utilising T7 primers targeting upstream and downstream of the region of interest in the pET-52b(+) vector (Table 6.1). Sequences were inspected predominantly for frameshifts and those with no significant amino acid substitutions were taken forward.

6.2.4 Cell preparation

Due to the volume of cloning reactions to be undertaken, Rosetta-gami™ 2 (DE3) competent cells (Novagen) were grown from stocks, treated to induce chemical competence and aliquoted for later use. Stocks were streaked onto an LB agar plates containing 34µg/mL chloramphenicol, 12.5µg/mL tetracycline and 50µg/mL streptomycin and grown at 37°C for 16 hours. A single colony was selected to inoculate a 5mL LB starter culture with the appropriate antibiotics and grown for 16 hours at 37°C. 1mL of starter culture was used to inoculate 100mL (1%) LB containing appropriate antibiotics, which was grown at 37°C until the optical density reached 0.5, at which time the cells were removed and placed on ice to cool. The culture was centrifuged at 4600 rpm for 5 min. The cell pellets were resuspended in a solution containing 50mM MgCl₂ and 20mM CaCl₂ and incubated at 4°C for 30 min. The cells were centrifuged as before and the supernatant was discarded. The cell

pellet was resuspended in a 20% glycerol solution containing 80mM CaCl₂. The cells were aliquoted into 100µL aliquots, snap frozen in liquid nitrogen and stored at -80°C. The transformation efficiency for these cells was determined to be 3.125x10⁵ with a pUC19 control plasmid, comparatively the guaranteed transformation efficiency of the stock was reportedly 2x10⁶. These aliquots were used for all transformations.

6.2.5 Bacterial transformation

Plasmids were extracted from transformed Stellar™ cells, quantified with the nanodrop and 2.5ng of plasmid DNA was transformed into 50µL Rosetta-gami™ 2 (DE3) competent cells (Novagen) prepared as in section 6.2.4. Successful transformants were identified using T7 primers (Table 6.1), successful colonies were grown for 16 hours in 15mL LB media containing 50µg/mL streptomycin in a 50mL falcon tube. Glycerol stocks were made to a final concentration of 20% glycerol, snap frozen in liquid nitrogen and stored at -80°C, these stocks were used for all expression screening and purification experiments.

6.2.6 Expression screening

Expression screening was undertaken to identify and optimise growth and induction conditions for protein production from transformed Rosetta-gami™ 2 cells. A starter culture in 20mL LB was inoculated from pre-prepared glycerol stocks and allowed to grow at 37°C for 16 hours at 175rpm. 1mL of starter culture was then used to inoculate a 100mL culture which was grown at 37°C at 175rpm until the optical density at 600nm reached 0.6, after which 5mL aliquots of culture were decanted into 16 50mL falcon tubes. Each falcon tube was subject to a different treatment with four temperature variables (16°C, 20°C, 30°C and 37°C) and induced with four Isopropyl β-D-1-thiogalactopyranoside (IPTG) concentrations (0mM, 0.1mM, 0.5mM, 1mM) for a total of 16 separate conditions. The cultures grown at 37°C and 30°C were incubated for four hours post induction prior to harvesting, whereas the cultures grown at 20°C and 16°C were incubated for 16-20 hours post induction. Empty vector controls were included in all batches of experiments and subject to every treatment. The entire 5mL culture was harvested by centrifugation at 4600 rpm for 10 min and the cell pellet was lysed with 250µL freshly prepared BugBuster™ (Novagen) solution containing 1x BugBuster™ in 1x PBS, 5mM MgCl₂, 0.1mM AEBSF, 10µg/mL DNase I (NEB) and transferred to a 1.5mL tube. The lysis solution was incubated at room temperature on a shaking platform at 1000

rpm for 25 min before being centrifuged in a benchtop centrifuge at 16,000g for 20 min at 4°C and the supernatant and insoluble protein fraction were separated for analysis. Protein profiles were visualised using SDS-PAGE and treatment success determined by band purity and intensity, proteins with conditions determined to have been successful were concluded as successful via chemiluminescent immunoblotting (section 6.2.9).

6.2.7 Protein purification

Large scale expression and purification was conducted in 500mL volumes LB media in 2 L baffled volumetric flasks inoculated with 5mL (1%) starter culture. The cultures were grown at 37°C at 175rpm until the optical density at 600nm in a disposable 1.6mL cuvette reached 0.6. An optimal concentration of IPTG was added and cultures were incubated at the optimal temperature (usually 0.1mM and 16°C for 16 hours) for 16-20 hours. Cells were harvested by centrifugation at 4600 rpm for 20 min in a Heraeus Multifuge3 S-R with a swing-out rotor (Sorvall). The cell pellet was resuspended in lysis buffer containing 280mM NaCl, 20mM sodium phosphate, 0.1mM 4-benzenesulfonyl fluoride hydrochloride (AEBSF; protease inhibitor) pH 7.4 and sonicated on ice intermittently for 3 sec (with 7 sec breaks) for a total of four min pulse time using an S-4000 ultrasonic liquid processor (Misonix, Inc). The cell lysate was partially purified of insoluble cell debris by high speed centrifugation in a Sorvall Evolution RC (Thermo) equipped with an SS-34 angled rotor at 16,000rpm for 30 min at 4°C. The resulting supernatant was filtered through a 0.22µm syringe driven filter unit prior to loading on a buffer equilibrated 5mL StrepTrap™ HP column (GE Healthcare) at a rate of 1mL per min on an ÄKTA fast protein liquid chromatography system (FPLC, GE Healthcare). The column was washed with three to 20 column volumes of lysis buffer. The bound protein was then displaced from the column and eluted using 2.5mM d-desthiobiotin (Sigma-Aldrich) in lysis buffer at a rate of 1mL per min. Protein containing wells were identified using measured UV absorbance, high intensity wells were combined and quantified using the Bradford methodology (section 2.3).

6.2.8 Thermal shift assay

Protein samples were mixed with SYPRO™ orange protein dye to a final dye concentration of 1-5x (usually 5x) and assayed on a Stratagene Mx300 rtPCR/Thermoflour thermal cycler for 70 cycles

between 25°C and 95°C at increments of 1°C per 30 sec. Data was acquired using the MxPro software package (Stratagene). The fluorescence dataset was normalised between 0 and 1 and the protein melting point (T_m) was estimated by fitting a sigmoid curve between the minimum (before denaturation) and maximum value on the fluorescence trace and selecting the median value of the curve. Samples were run in triplicate and representative traces were analysed. Thermal shift assay results were used to determine an initial temperature range to test for pH and temperature optima.

6.2.9 Chemiluminescent immunoblotting

SDS-PAGE gels of soluble and insoluble protein fractions (Mini-PROTEAN® TGX™ 4-12%) were run for 45 min at 200V and incubated in 20% ethanol for 5 min. Transfer electrophoresis was then conducted on an iBlot™ 2 Dry Blotting System (Invitrogen), with compatible nitrocellulose transfer stacks, filter paper and absorbent pads (Thermo IB23002). Electrophoresis was conducted for seven min at 20V and 1.3A. The protein containing nitrocellulose membrane was then washed twice for 5 min with 1x PBS. Blocking was conducted by incubation for 1 hour with 1x PBS with 5% (w/v) skimmed milk (Sigma). Three further 5 min washes were conducted with PBS-Tween-20 buffer (1x PBS with 0.05% Tween 20 (w/v)). The membrane was finally incubated with 10mL PBS-Tween-20 containing 2µL (Strep.Tag® II Antibody (peroxidase) HRP Conjugate (Novagen)) for 2 hours. Fluorescence was detected using SuperSignal West Pico Fast Western Blot Kit according the manufacturer's protocol and imaged with Syngene PXi imaging station (Thermo).

6.2.10 Assays

Purified proteins were characterised to determine their function in lignocellulose deconstruction. A broad range of substrates were tested depending on the putative enzyme family to determine substrate affinity. Temperature and pH optima were also assessed together with thermostability assays. Salt tolerance and activity in seawater where applicable were also measured. Absolute rates of activity were measured during the linear phase to provide maximum activity rates for comparison to other enzymes.

6.2.10.1 *Reducing sugar assay*

Reducing sugars were detected using 4-hydroxybenzoic acid hydrazide (PAHBAH) (Lever, 1977). Sugar hydrolysate was incubated with a solution containing 50mM PAHBAH, 10mM bismuth nitrate, 10mM sodium, potassium tartrate in 0.5M NaOH at 72°C for 10 min. Absorbance changes were detected at 405nm using a Tecan Sunrise™ microplate spectrophotometer. A standard curve was obtained using serial dilutions of glucose.

6.2.10.2 *Substrate affinities*

A range of complex polysaccharide, oligosaccharide and respective derivatives were used to determine substrate specificity of target enzymes. The polysaccharides utilised and their final concentration in reaction mixtures included; Avicel™ (0.5%), phosphoric acid swollen cellulose (PASC; 1%), carboxy-methyl cellulose (sodium salt; 0.5%), xylan (beechwood 1-4-β-xylose; 0.5%), 1-4-β-glucan (barley; 0.1%), lichenan (1,3:1,4-β-glucan; 0.1%), pachyman (1,3-β-glucan), arabinoxylan (0.1%), arabinogalactan (0.1%), xyloglucan (0.1%), galactan (Lupin; 0.1%), glucomannan (0.1%), galactomannan (0.1%), mannan (1-4-β-mannan; 0.1%).

Monosaccharides and oligosaccharides (glucose, cellobiose, cellotriose, cellotetraose, cellopentaose and cellohexaose) were utilised at a final concentration of 0.1mM or 0.5mM and activity was determined using the PAHBAH method. The 4-nitrophenyl-β-D-R (Carbosynth Limited) fluorometric derivative of these substrates (pNP-β-D-glucopyranoside, pNP-β-D-cellobioside, pNP-β-D-cellotrioside, pNP-β-D-cellotetraoside, pNP-β-D-cellopentaoside) were utilised in a final concentration of 0.5mM. Reactions were terminated with 1M sodium carbonate (Na₂CO₃) to a final concentration of 0.5M. Activity was assessed against 4-nitrophenol standards and measured in a TECAN, Sunrise microplate reader at 405nm. All assays were performed in an Incubator shaker (MaxQ 8000, Thermo) at 250rpm or a thermal cycler (Pelteir PTC 200, MJ Research or Tetrad Z, Pelteir DNA Engine, BioRad) without shaking.

6.2.10.3 *Bivariate pH, temperature optima*

As pH and temperature are not independent variables a two-dimensional (bivariate) pH and temperature experiment was conducted. Relative activity profiles were obtained using a 100mM

citrate-phosphate (CPB) buffer system in a pH range of 4-8 (using various volumes of 0.1M citric acid and 0.2M disodium hydrogen phosphate). This buffer system was chosen as it encompasses both intrinsic salt marsh pH ranges and common glycoside hydrolase optima, additionally this buffer systems buffering capacity is not significantly altered by temperature changes or salt concentration (Herlet et al., 2017). Reactions were carried out in 50µL or 100µL volumes in 96 well plates or split 96 well plates in a Tetrad Z, Pelteir DNA Engine (BioRad) gradient cycle or multiple thermal cyclers (Pelteir PTC 200, MJ Research) for incubation at predetermined temperatures. For the characterisation of a_c570829_g1_i1_1 (GH6), the reaction mixture contained 1µg purified enzyme, 0.2% Avicel™ for 24 hours in a total reaction volume of 100µL for 24 hours in multiple thermal cyclers, activity was measured using the PAHBAH method 6.2.9.2. For the characterisation of c_c892255_g1_i1_1 (GH5), the reaction mixture contained 0.05µg of purified enzyme and 0.5mM pNP-G2 for 1 hour in a single gradient thermal cyler, activity was measured spectrophotometrically at 405nm.

6.2.10.4 Salinity tolerance

Tolerance to salt was measured both using artificial seawater (ASW) and sodium chloride concentrations equivalent to the ionic strength of estuarine waters, seawater and above. Reaction mixtures were as in section 6.2.9.2 but were conducted at the enzymes respective pH and temperature optima; a_c570829_g1_i1_1 (GH6) at 43.8°C and 50mM sodium acetate pH 5, c_c892255_g1_i1_1 (GH5) at 45°C and 50mM sodium phosphate pH 7.6.

6.2.10.5 Thermostability

Thermostability assays were conducted by heat treatment of a_c570829_g1_i1_1 (GH6) at 100°C in various concentrations of sodium chloride. The enzyme was allowed to renature by incubating at 25°C for 1 hour post heat treatment prior to being assayed using the PAHBAH method.

6.2.10.6 *Rate of activity*

The absolute rate of activity of a_c570829_g1_i1_1 (GH6) was determined using the reaction mixture as in section 6.2.9.3 and determined using the PAHBAH method. Independent samples were measured at 1, 2, 3 and four hours.

6.2.10.7 *Processivity*

Cellulose and cellulose derivatives (Avicel (0.2%), phosphoric acid swollen cellulose (0.2%), carboxy-methyl cellulose (0.05%)) were incubated with 1µg of purified a_c570829_g1_i1_1 (GH6) for 1 hour at optimum conditions. Upon completion the reaction was centrifuged at 4600 rpm for 1 min. The supernatant was extracted and 4 volumes of 100% ethanol pre-chilled to 4°C was added to a final concentration of 80% to precipitate proteins and impurities. The hydrolysate was dehydrated in a Savant SpeedVac concentrator SPD131DDA (Thermo) used in conjunction with a Savant refrigerated vapor trap RVT4104 (Thermo). Samples were resuspended in the original volume taken and analysed with ion exchange chromatography using an ICS-3000 Ion Chromatography System with AS loading units (Dionex, Thermo). Peak areas were integrated and quantified against known standards using the Chromeleon software package (version 6.80 SR16 Build 5387, Thermo). Relative product profiles were assessed and processivity was determined (equation 6.1)

Equation 6.1 Processivity. Processivity. G; proportion of glucose, G2; proportion of cellobiose, G3; proportion of cellotriose.

$$\frac{G2}{G1 + G3}$$

6.2.10.8 *Product profile*

Glucosaccharides (cellobiose, cellotriose, cellotetraose, cellopentaose and cellohexaose) of final concentration of 0.1mM were incubated with 1µg a_c570829_g1_i1_1 (GH6) for 4 hours at optimum conditions. Hydrolysate was analysed as in section 6.2.10.7. For c_c892255_g1_i1_1 (GH5), 0.5mM glucosaccharides were incubated with 1µg purified enzyme for 16 hours.

6.2.10.9 *Signal peptide prediction*

Regions depicting signal peptides were identified using the predictive software SignalP (version 4.0) (Petersen et al., 2011).

6.3 Results

6.3.1 Target amplification

A plethora of PCR conditions were tested for all 42 selected target sequences, successful amplification was achieved for 38 of these (Table 6.4). Of the four that failed to amplify, despite testing multiple conditions, nested primers and touch-down PCRs, one, *c_c991288_g2_i3_1* was a predicted isoform of *c_c991288_g2_i1_1* (GH5) which was successful and as such was a likely an assembly artefact. Other targets that failed to amplify included *c_c1011459_g1_i2_2* (CE1), *a_c579709_g1_i2_1* (GH6) and *a_c554459_g1_i3_1* (GH9). All targets originating from proteomic identification amplified successfully.

6.3.2 Target cloning

Four enzyme families (GH3, GH5, GH6, GH9) were focussed upon for cloning and a miscellaneous group containing auxiliary and less abundant enzymes. The aim was to clone five representatives per family. In total, 25 target amplicons were successfully inserted into pET-52b(+) and transformed into Stellar™ competent cells; GH3 (5), GH5 (5), GH6 (4), GH9 (5) and a miscellaneous group containing AA2, AA9, CE1, GH11 (2) and a GH16 (Table 6.5). Confirmation of successful transformation was performed with colony PCR (Figure 6.2). Putative successes were scaled up and the was plasmid extracted and sequenced for confirmation of transformation.

Expression and characterisation of putatively lignocellulolytic enzymes

Table 6.4 PCR conditions for the successful amplification of ORF targets. P/HF; Phusion polymerase, Q5/Q5; Q5 polymerase, GE; Gel extraction. P/HF.

| Index | Enzyme/ Buffer | Primer (μ M) | dNTPs (μ M) | Additive | T _m ($^{\circ}$ C) | Extension (s) | Purification |
|--------|-------------------|----------------------|---------------------|-----------|-----------------------------------|------------------|--------------|
| 1 | P/HF | 0.5 | 200 | None | 60 | 20 | Direct |
| 2 | P/HF | 0.25 | 800 | None | 60 | 20 | Direct |
| 4 | P/HF | 0.25 | 800 | None | 60 | 20 | GE |
| 5 | P/HF | 0.25 | 800 | None | 60 | 20 | Direct |
| 6 | P/HF | 0.5 | 200 | None | 60 | 20 | Direct |
| 7 | P/HF | 0.5 | 200 | None | 60 | 20 | Direct |
| 8 | P/HF | 0.5 | 200 | None | 60 | 20 | Direct |
| 9 | P/HF | 0.5 | 200 | None | 60 | 20 | Direct |
| 10 | P/HF | 0.5 | 200 | None | 60 | 20 | Direct |
| 11 | P/HF | 0.5 | 200 | None | 60 | 20 | Direct |
| 12 | - | - | - | - | - | - | - |
| 13 | P/HF | 0.25 | 800 | None | 60 | 45 | N/A |
| Nested | P/HF | 0.25 | 800 | None | 60 | 30 | GE |
| 14 | P/HF | 0.5 | 200 | None | 58 | 60 | Direct |
| 15 | - | - | - | - | - | - | - |
| 16 | P/HF | 0.5 | 200 | None | 60 | 20 | Direct |
| 17 | P/HF | 0.5 | 200 | None | 60 | 20 | Direct |
| 18 | P/HF | 0.5 | 200 | None | 60 | 20 | Direct |
| 20 | P/HF | 0.5 | 200 | None | 60 | 20 | Direct |
| 21 | P/HF | 0.5 | 200 | None | 60 | 20 | Direct |
| 22 | P/HF | 0.5 | 200 | None | 60 | 20 | Direct |
| 23 | P/HF | 0.25 | 800 | DMSO (3%) | 60 | 20 | Direct |
| 24 | P/HF | 0.25 | 800 | None | 65 | 45 | N/A |
| Nested | P/HF | 0.25 | 800 | None | 60 | 30 | GE |
| 25 | P/HF | 0.25 | 800 | None | 60 | 45 | GE |
| 26 | - | - | - | - | - | - | - |
| 27 | P/HF | 0.25 | 800 | None | 60 | 25 | Direct |
| 28 | P/HF | 0.25 | 800 | None | 60 | 25 | Direct |
| 29 | P/HF | 0.5 | 200 | None | 60 | 25 | Direct |
| 30 | P/HF | 0.25 | 800 | None | 60 | 25 | Direct |
| 31 | P/HF | 0.5 | 200 | None | 60 | 25 | Direct |
| 32 | P/HF | 0.25 | 800 | None | 60 | 25 | GE |
| 33 | P/HF | 0.25 | 800 | None | 60 | 45 | N/A |
| Nested | P/HF | 0.25 | 800 | None | 65 | 30 | GE |
| 34 | - | - | - | - | - | - | - |
| 35 | P/HF | 0.5 | 200 | None | 60 | 25 | Direct |
| 36 | P/HF | 0.25 | 800 | None | 60 | 25 | GE |
| 37 | P/HF | 0.25 | 800 | None | 60 | 25 | GE |
| 38 | P/HF | 0.5 | 200 | None | 60 | 25 | Direct |
| AS4 | P/HF | 0.5 | 200 | DMSO (3%) | 60 | 30 | Direct |
| AS7 | P/HF | 0.5 | 200 | DMSO (3%) | 60 | 20 | Direct |
| AS12 | P/HF | 0.5 | 200 | DMSO (3%) | 60 | 20 | GE |
| AS15 | P/HF | 0.5 | 200 | DMSO (3%) | 60 | 20 | Direct |
| PCR | | | | | | | |
| AS19 | Q5/Q5 | 0.5 | 200 | enhancer | 60 | 45 | Direct |
| AS20 | P/HF | 0.5 | 200 | None | 60 | 20 | Direct |
| AS23 | P/HF | 0.5 | 200 | DMSO (3%) | 60 | 20 | Direct |

Table 6.5 An overview of successfully cloned targets into pET-52(b)+ in Stellar™ cells.

| Transformation target | CBM containing | Single domain | Total | Maximum available |
|-----------------------|----------------|---------------|-------|-------------------|
| GH3 | - | 5 | 5 | 12 |
| GH5 | 1 | 4 | 5 | 10 |
| GH6 | 3 | 1 | 4 | 5 |
| GH9 | 3 | 2 | 5 | 6 |
| AA2 | - | 1 | 1 | 2 |
| AA9 | - | 1 | 1 | 1 |
| CE1 | 2 | - | 2 | 2 |
| CE6 | - | - | 0 | 2 |
| GH11 | 1 | - | 1 | 1 |
| GH16 | - | 1 | 1 | 1 |
| Sum | 10 | 15 | 25 | 42 |

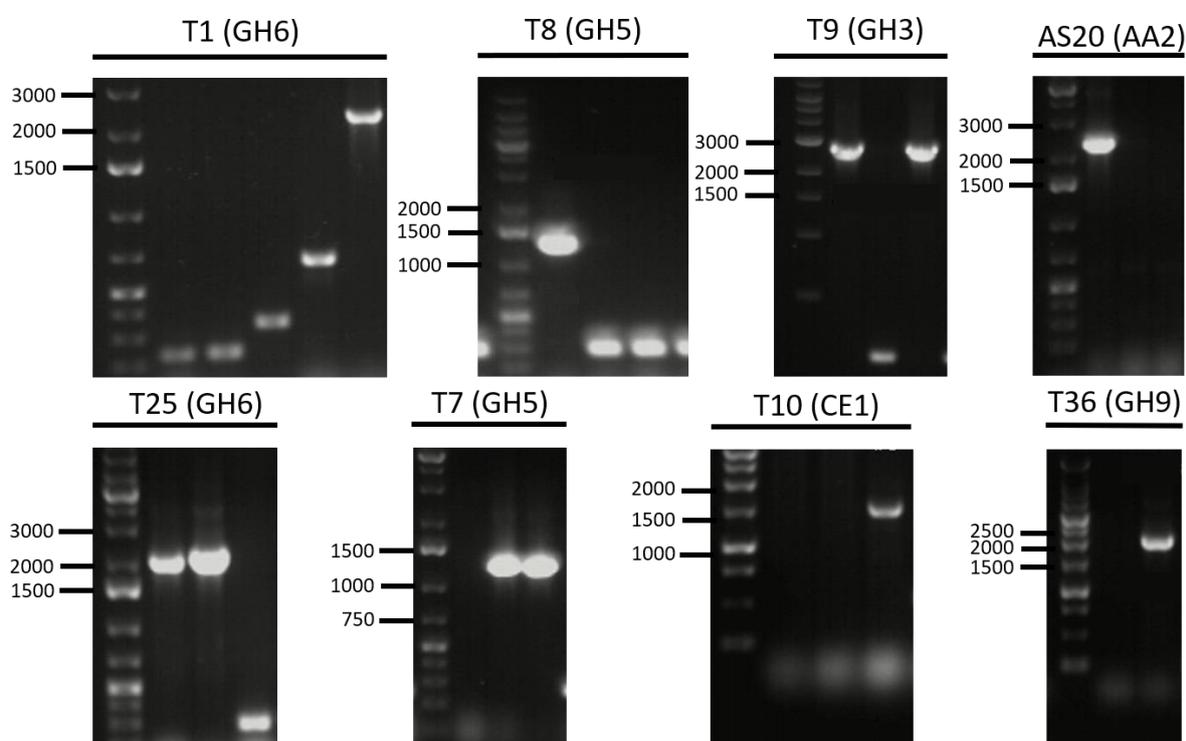


Figure 6.2 T7 colony PCRs. Transformed Stellar™ cells with the putative gene of interest in pET-52b(+). A small subset of transformants and targets are displayed.

6.3.3 Expression screening

Plasmids were harvested from Stellar™ cells and transformed into the chosen expression strain (Rosetta-gami™ 2) and screened for optimal growth conditions (Figure 6.3). Of the 15 targets screened in the 16 conditions, nine expressed soluble protein (Table 6.6).

Expression and characterisation of putatively lignocellulolytic enzymes

Table 6.6 Optimal expression conditions of screened protein targets. n/d; not determined, PF; partially fragmented, S; soluble fraction, I; insoluble fraction.

| Family | Index | CBM | Disulfide bonds | Rare codons (%) | Mol. weight | pI | SDS-PAGE | | Western blot | | Optimal condition |
|--------|-------|-----|-----------------|-----------------|-------------|-----|----------|-----|--------------|-----|-------------------|
| | | | | | | | S | I | S | I | |
| GH6 | | | | | | | | | | | |
| | 1 | Yes | 3 | 12 | 76108 | 4.3 | Yes | n/d | Yes | Yes | 16°C, 0.1mM |
| | AS12 | Yes | 3 | 12 | 73879 | 4.1 | n/d | n/d | Yes | Yes | 16°C, 0.1mM |
| | 25 | Yes | 3 | 13 | 65939 | 4.5 | n/d | n/d | Yes | Yes | 16°C, 0.1mM |
| | AS15 | No | 0 | 11 | 58487 | 4.2 | Yes | n/d | No | No | - |
| GH5 | | | | | | | | | | | |
| | 6 | Yes | 8 | 15 | 78262 | 4.4 | Yes | Yes | PF* | - | 20°C, 0.1mM |
| | 7 | No | 2 | 18 | 43373 | 5.2 | n/d | n/d | No | - | - |
| | 8 | No | 2 | 19 | 43918 | 4.1 | Yes | Yes | Yes | - | 30°C, 0.1mM |
| | AS7 | Yes | 4 | 11 | 62807 | 5.7 | n/d | n/d | No | - | - |
| GH3 | | | | | | | | | | | |
| | 9 | No | 7 | 21 | 95124 | 5.6 | Yes | Yes | Yes | - | 16°C, 0.1mM |
| | T14 | No | 1 | 23 | 134416 | 4.1 | Yes | Yes | Yes | - | 16°C, 0.1mM |
| GH9 | | | | | | | | | | | |
| | AS19 | Yes | 2 | 18 | 126165 | 4.9 | No | Yes | PF* | Yes | 16°C, 500µm |
| | T20 | Yes | 6 | 14 | 105428 | 4.4 | No | No | No | - | - |
| | T35 | No | 0 | 19 | 80079 | 4.9 | Yes | Yes | Yes | - | 20°C, 0.1mM |
| | T36 | Yes | 1 | 13 | 64368 | 4.6 | No | No | No | - | - |
| | T37 | No | 0 | 18 | 58718 | 4.8 | Yes | Yes | Yes | - | 20°C, 0.1mM |

Expression and characterisation of putatively lignocellulolytic enzymes

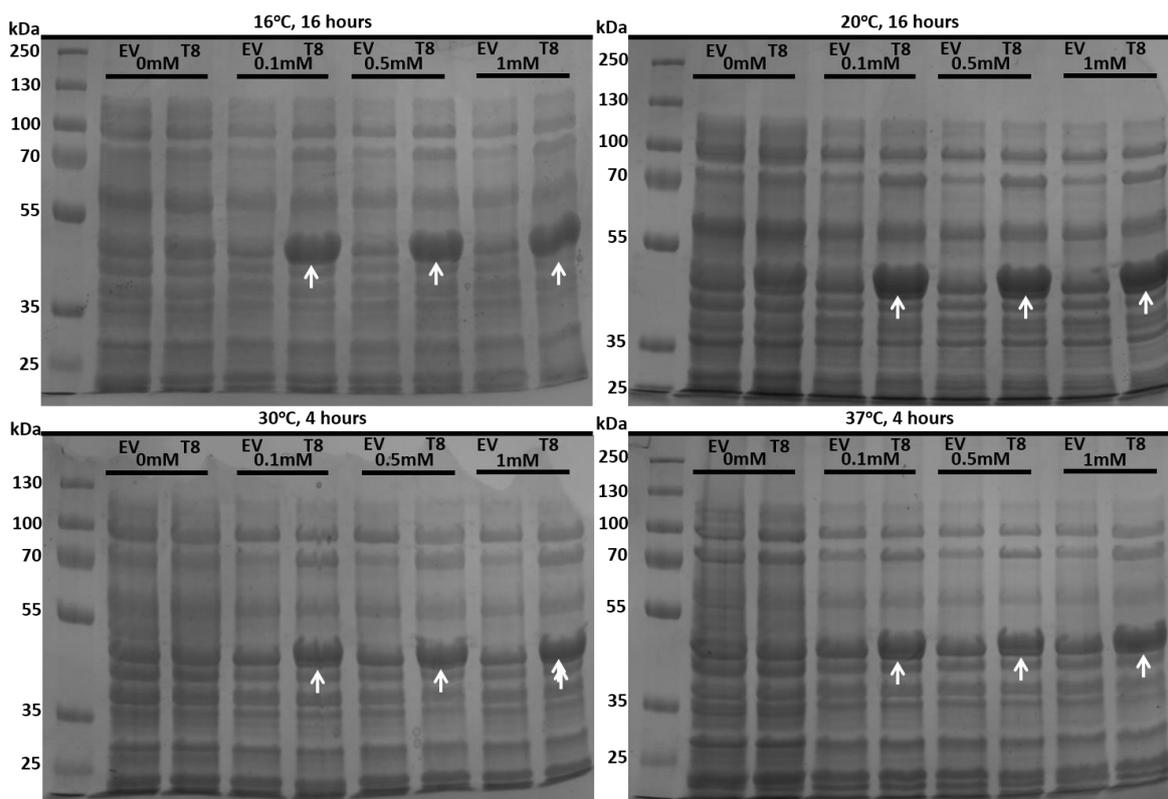


Figure 6.3 Representative SDS-PAGE of the screening process. Soluble proteins harvested from Rosetta-gami™ 2 containing a pET-52b(+) plasmid inserted with T8 (a putative GH5; c_c892255_g1_i1_1) induced at various concentrations of IPTG (0mM, 0.1mM, 0.5mM, 1mM) and incubated at selected temperatures (16°C, 20°C, 30°C or 37°C) for four hours or 16 hours. EV; empty pET-52b(+) vector control.

The presence of a protein of interest was confirmed with a chemiluminescent immunoblot of a selection of conditions where expression was greatest, could not be distinguished or was presumed absent (Figure 6.4). Relative fluorescence intensity between conditions was used to determine the optimal condition. Interestingly a_c580607_g3_i1_2 (GH5) was fragmented at all conditions except the non-induced condition where significant leaky expression was observed producing putatively intact soluble protein. Leaky expression arises after carbon source exhaustion inducing cyclic AMP which stimulates transcription by the T7 RNA polymerase (encoded in the λ DE3 cassette under *lac*-UV5 control) which is produced at basal levels despite the *lac*-UV5 repression in the absence of a suitable induction agent (Rosano and Ceccarelli, 2014, Novy, 2001). Leaky expression is often considered problematic; however, in this instance it proved advantageous, likely by reducing the rate of protein production.

Expression and characterisation of putatively lignocellulolytic enzymes

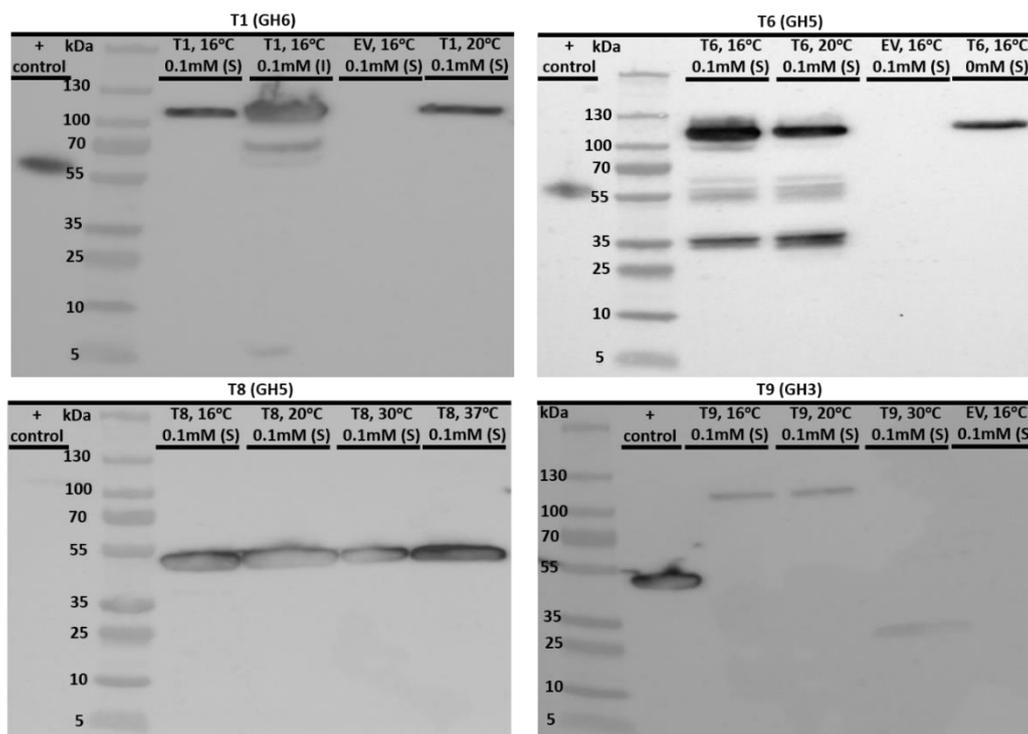


Figure 6.4 Fluorescent immunoblots of a random subset of protein targets. A subset of condition combinations is displayed which were induced at different IPTG concentrations and incubated at different temperatures. (S); soluble protein fraction, (I); insoluble protein fractions.

Two targets were selected for further analyses; a putative GH6 (a_c570829_g1_i1_1) and a putative GH5 (a_c580607_g3_i1_2). The former was selected as it is the most abundant ORF in two time points, it also belonged to the most abundant family (GH6) and contained the two most abundant CBMs (CBM2 and CBM44) within its architecture. The latter was selected as despite its architecture indicating a single highly confident ($1.2e^{-45}$) GH5 domain, nucleotide and protein sequence homology in the NCBI databases yielded a maximum sequence identity of 45%, suggesting significant sequence divergence and possible novelty.

6.3.4 Purification and characterisation of a putative GH6 (a_c570829_g1_i1_1)

6.3.4.1 Purification

Large scale expression and purification of T1 (a_c570829_g1_i1_1; appendix) was achieved from two to four 2 L flasks containing 500mL LB at the determined optimal conditions (Table 6.6). The resultant protein was isolated with affinity tag purification (Figure 6.5). Fractions resulting from elution were analysed with SDS-PAGE, the concentrated fractions were pooled and quantified with

the Bradford method (section 2.3) and subjected to a thermal shift assay to determine the melting point, which was determined to be around 40°C (Figure 6.6).

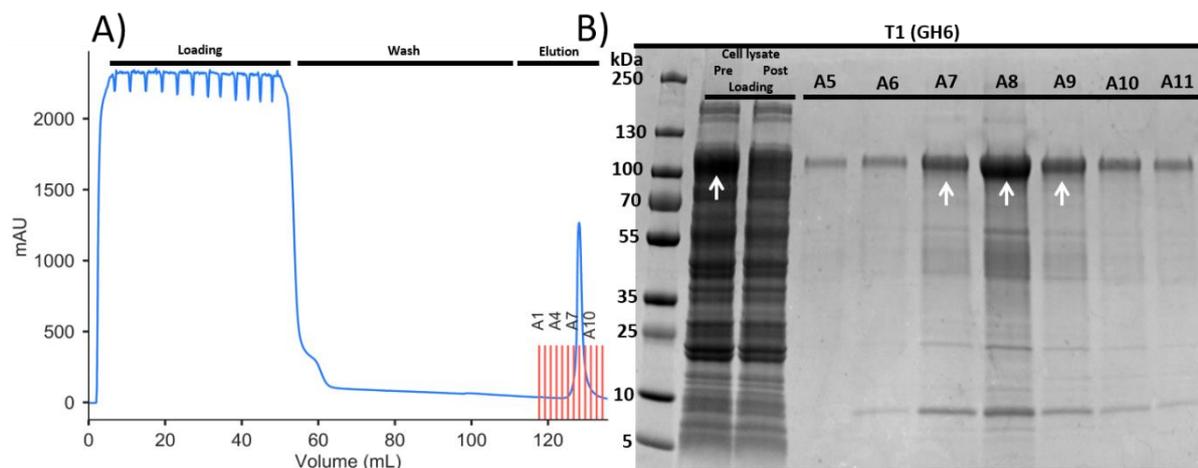


Figure 6.5 Purification of a putative GH6. A) Purification of a_c570829_g1_i1_1 (T1; GH6) from sonicated cell lysate on a StrepTrap™ HP column, resulting elution fractions were analysed by SDS-PAGE (B).

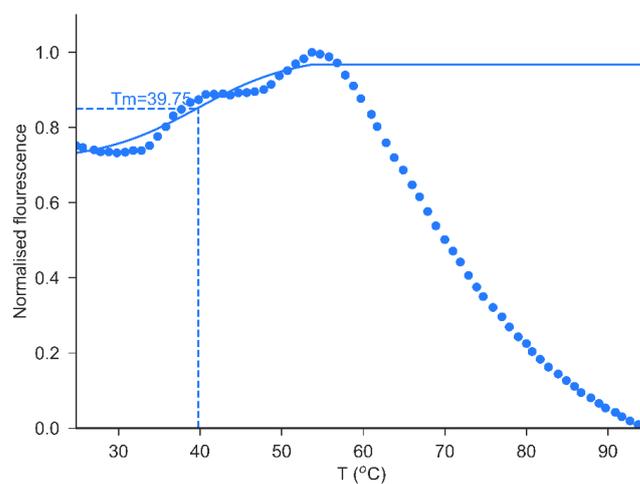


Figure 6.6 A representative thermal shift profile of a_c570829_g1_i1_1 (GH6). A sigmoidal curve between the maximum and minimum before the maximum is shown, the median value of this fit was determined as the melting point (Tm).

6.3.4.2 Substrate specificity

The substrate T1 (a_c570829_g1_i1_1) was active on was determined by incubating the enzyme with multiple polysaccharides, oligosaccharides and derivative oligosaccharides, activity was

measured fluorometrically or by the release of reducing sugars. There was no detectable activity for any pNP substrates tested (pNP-G1 to pNP-G5) after incubation for 24 hours; however, polysaccharide assays revealed significant (p value greater than 0.05 assessed by a one-way ANOVA) sugar release and an unusual affinity for crystalline cellulose (Avicel™) with a weaker affinity for less crystalline cellulose in the form of phosphoric acid swollen cellulose (PASC) and the soluble cellulose derivative carboxy-methyl cellulose (CMC) (Figure 6.7). No significant activity was detected on xylan, lichenan, glucomannan and mannan. These results suggest an affinity for 1-4- β glycosidic bonds and the lack of activity on pNP-R substrates (putatively substrates for cellobiohydrolases and β -glucosidase) suggest endo-acting activity.

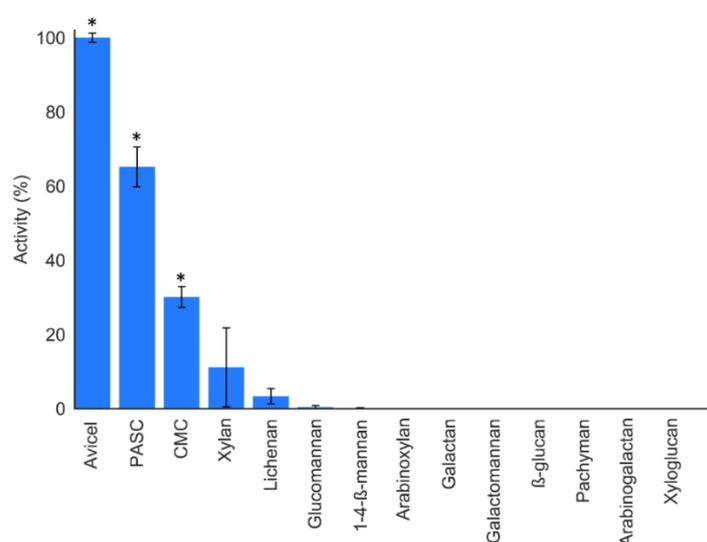


Figure 6.7 Substrate affinities of a_c570829_g1_i1_1 (GH6). 3 μ g of enzyme was incubated with a substrate at optimal conditions at 250rpm for 24 hours in 50mM sodium acetate. Error bars represent standard error of five technical replicates. Asterisks indicate significance at a confidence interval of $p < 0.05$ assessed by a one-way ANOVA performed on non-normalised, raw data between treated (enzyme) and untreated (no enzyme) samples.

6.3.4.3 Bivariate pH, temperature optima

The optimal pH and temperature condition for a_c570829_g1_i1_1 (GH6) was 45°C and pH 5 in 600mM NaCl. Interestingly both the temperature and pH had a broad range within which the enzyme remained highly active with 60% activity down to 25°C and up to pH 7.5 (Figure 6.8). This coincides with the likely environmental maxima where salt marshes rarely exceed 30°C and the pH averages 7.6 (Baumann et al., 2015).

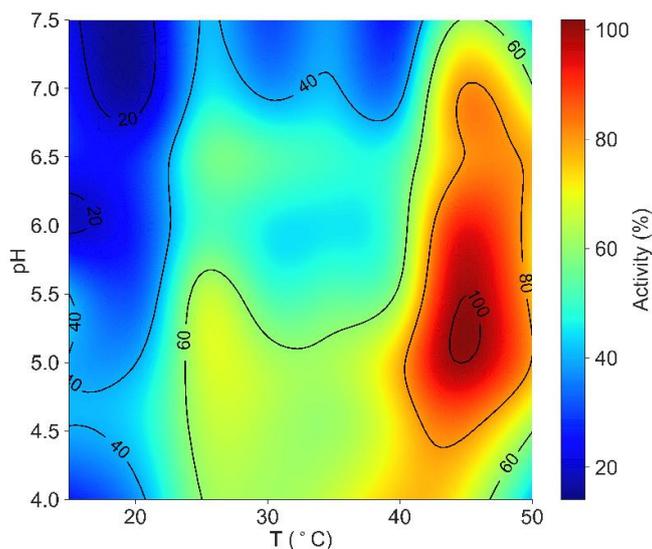


Figure 6.8 Bivariate pH and temperature optima of a_c570829_g1_i1_1 (GH6). 1 μ g of purified enzyme was incubated in 2% Avicel™ for 24 hours.

6.3.4.4 Salinity tolerance

Tolerance to saline media was assessed with various sodium chloride concentrations as a proxy for increasing ionic strength and independently against artificial seawater (ASW). Activity was unaffected by salt concentrations up to 1M before declining rapidly to 4M NaCl, where the rate of decline plateaus to 6M NaCl, where 25.5% activity is measured (Figure 6.9). In artificial seawater, the enzyme is tolerant of 0.5x and 1x concentrated ASW (environmentally relevant) with no significant difference detected between the two concentrations, but in 2x concentrated ASW a significant decline is observed. It is concluded that T1 is tolerant to environmental seawater concentrations.

Expression and characterisation of putatively lignocellulolytic enzymes

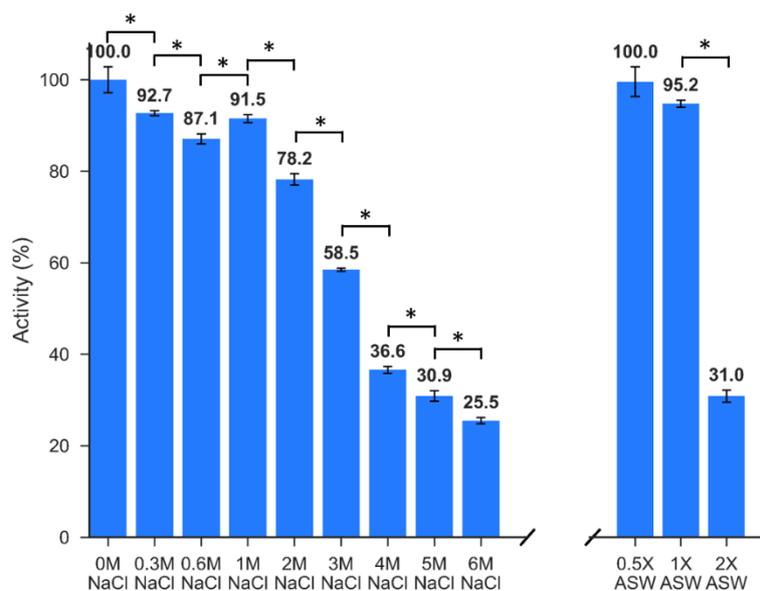


Figure 6.9 Salinity tolerance of a_c570829_g1_i1_1 (GH6). 1 μ g of purified enzyme was incubated with 2% Avicel™ for 16 hours at optimal conditions in 50mM sodium acetate supplemented with sodium chloride pH 5 or artificial seawater (ASW) pH 5. Asterisks indicate significance at a confidence interval of $p < 0.05$ assessed by a one-way ANOVA performed on non-normalised, raw data.

6.3.4.5 Thermostability

Thermostability was assessed by measuring thermal shift profile and relative activity after heat treatment at 100°C for 5 or 15 min in various salt concentrations. The enzyme was allowed a settling period to renature for 1 hour at 25°C. Whilst the thermal shift profile suggests the enzyme should be completely denatured at around 60°C, thermal shift analysis post heat treatment revealed structures were present that were observably denaturing, suggesting some degree of protein restructuring (Figure 6.10). All conditions tested suggested some degree of renaturing, with the most convincing profiles visible for higher salt concentrations; this is particularly true for samples heat treated in sodium chloride and denaturing agents (100mM DTT and 2% SDS) which were supposed to aid denaturation and break disulfide bonds. The lower salt concentration exhibited a profile synonymous with the buffer control suggested no protein refolding had occurred, however at 600mM NaCl the profile was the same as the untreated, suggesting salt may play a protective role against denaturants. Potential refolding was also observed for the non-heat-treated sample diluted to 80% ethanol, dehydrated and resuspended in the presence of 600mM salt.

Expression and characterisation of putatively lignocellulolytic enzymes

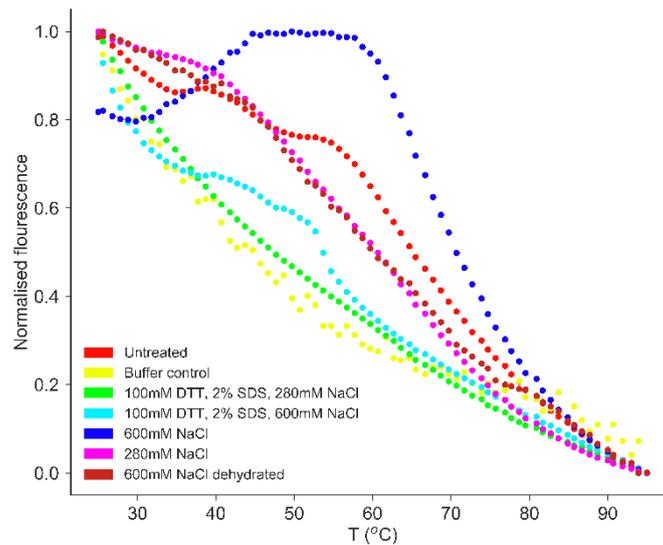


Figure 6.10 Thermal shift profiles of a_c570829_g1_i1_1 (GH6) post heat treatment in different treatments. DTT; dithiothreitol, SDS; sodium dodecyl sulfate.

To assess whether these refolded protein structures were functional, activity assays were conducted. All treatments were significantly different to the untreated control (Figure 6.11). The enzyme remained active up to boiling for 5 and 15 min up to 600mM before a significant decline. Interestingly, in estuarine salt concentrations it retained 65.5% relative activity after boiling for 5 min in 300mM NaCl which was the highest recorded. The only condition where treatment time had a significant effect was 300mM NaCl, suggesting salt concentration rather than treatment longevity was responsible for the loss in activity. Higher salt concentrations (900mM and 1.2M) caused visible protein precipitation (this was also observed at 600mM for 15 min).

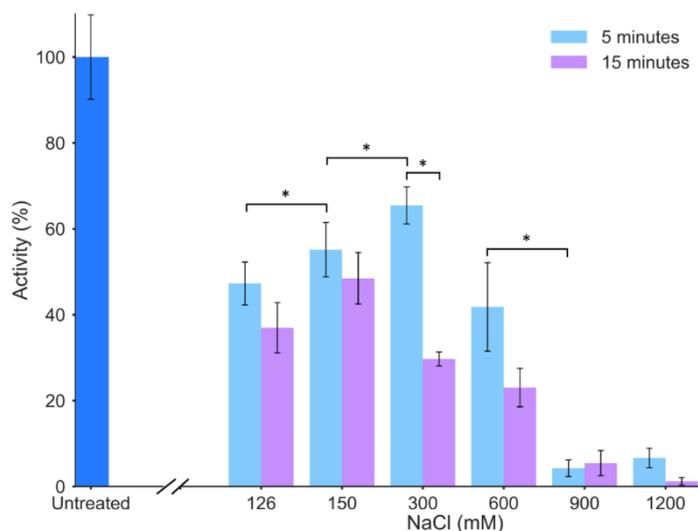


Figure 6.11 Thermostability of a_c570829_g1_i1_1 (GH6) in different salt concentrations. 2.5 μ g of heat treated enzyme was incubated with 0.25% Avicel™ for six hours at optimal conditions. Asterisks indicate significance at a confidence interval of $p < 0.05$ assessed by a one-way ANOVA performed on non-normalised, raw data.

6.3.4.6 Rate of activity

To convey relative activity to facilitate comparisons between similar enzymes, maximum rates of activity were recorded hourly for four hours to ensure measurements were made during the linear phase of degradation (prior to steric inhibition by-products, enzyme degradation or the formation of traffic jams). The maximum rate of activity at optimal conditions at 1 hour was 1.03 μ mol glucose reducing sugar equivalents. $\text{mg}^{-1}.\text{min}^{-1}$, and at 25 $^{\circ}\text{C}$ (likely environmental maxima as observed in the bivariate optima heatmap in section 6.3.4.3) 0.50 μ mol glucose reducing sugar equivalents. $\text{mg}^{-1}.\text{min}^{-1}$ at 2 hours (Figure 6.12). At both temperatures the rate of activity appeared to begin to plateau between hour three and four with a significant decline in the rate of activity, suggesting product inhibition.

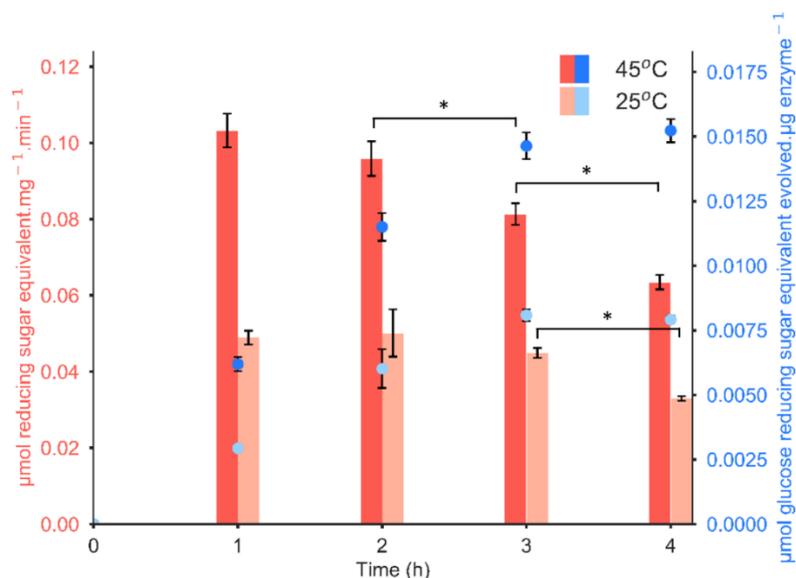


Figure 6.12 Rate of activity of a_c570829_g1_i1_1 (GH6). Reactions were conducted independently. 0.5μg on enzyme was incubated in 2% Avicel™ in 50mM sodium acetate pH 5 at two temperatures (25°C and 45°C) for up to four hours. Total reducing sugars evolved at each timepoint for both temperatures are shown shades of in blue, rates of activity at each temperature are shown in shades of red. Asterisks indicate significance at a confidence interval of $p < 0.05$ assessed by a one-way ANOVA.

6.3.4.7 Product profile

Product profiles for 1-4-β-glucanases can be indicative of their mechanism of action. T1 (a_c570829_g1_i1_1) was incubated with oligosaccharides (from glucose to cellohexasaccharide) and the sugar products evolved were analysed and quantified by ion exchange chromatography (as described in section 6.2.10.7) (Figure 6.13). T1 was unable to hydrolyse cellobiose and showed very weak activity toward cellotriose (negligible amounts of glucose and cellobiose were detected during a cellotriose incubation which were not present in the negative control). The major product from incubation with cellotetraose was cellobiose. Incubation with cellopentaose evolved only cellobiose and cellotriose, and incubation with cellohexasaccharide revealed only cellobiose, cellotriose and cellotetraose (the latter likely due to incomplete hydrolysis or inhibition by the former).

Expression and characterisation of putatively lignocellulolytic enzymes

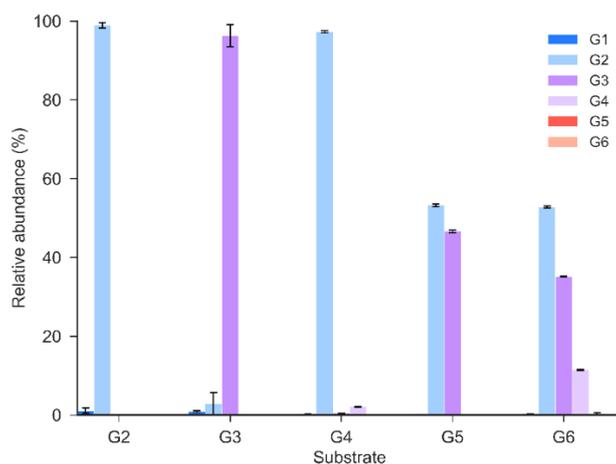


Figure 6.13 Product profile a_c570829_g1_i1_1 after incubation with oligosaccharides. G1; glucose, G2; cellobiose, G3; cellotriose, G4; cellotetraose, G5; cellopentaose, G6; cellohexaose.

Additional incubations were performed on polysaccharides where activity was highest (Avicell™, PASC and CMC). Incubations were performed for 1 hour at the optimal conditions and analysed by ion exchange chromatography (Figure 6.14A). The major product for all polysaccharides was cellobiose, although glucose and cellotriose was produced in all reactions. The product profiles for each polysaccharide was quantified and relative proportions were used to determine enzyme processivity (equation 6.1) on each substrate (Figure 6.14B). These profiles are suggestive of a processive cellobiohydrolase which conflict with the negative results with pNP-R substrates.

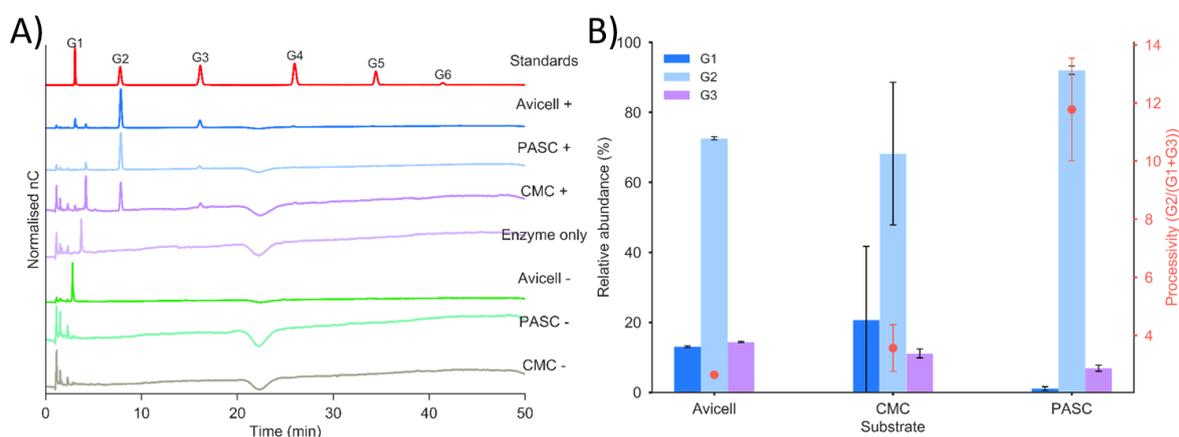


Figure 6.14 Product profile and processivity of a_c570829_g1_i1_1. A) Product profile a_c570829_g1_i1_1 after incubation with polysaccharides. B) Relative abundance of products and processivity PASC; Phosphoric acid swollen cellulose, CMC; Carboxy-methyl cellulose, G1; glucose, G2; cellobiose, G3; cellotriose, G4; cellotetraose, G5; cellopentaose, G6; cellohexaose.

Processivity is used to estimate enzyme efficiency by assessing the proportion of ‘false starts’ (producing cellotriose) and non-target activity (e.g. activity on cellotriose) and was observed to be most processive on PASC (11.77 ± 1.76) than Avicel (2.65 ± 0.06) and CMC (3.56 ± 0.8). This is concordant with the fact PASC is presumably more accessible than crystalline Avicel and is therefore able to initiate and process more effectively.

6.3.5 Purification and characterisation of a putative GH5 (c_c892255_g1_i1_1)

6.3.5.1 Purification

Large scale expression and purification of T8 (c_c892255_g1_i1_1; appendix) was achieved from two to four 2 L flasks containing 500mL LB at the determined optimal conditions (Table 6.6). The resultant protein was isolated with affinity tag purification (Figure 6.15). Fractions resulting from elution were analysed with SDS-PAGE and concentrated fractions pooled and quantified with the Bradford method (see section 2.3) and subjected to a thermal shift assay to determine the melting point, which was determined to be around 53°C (Figure 6.16).

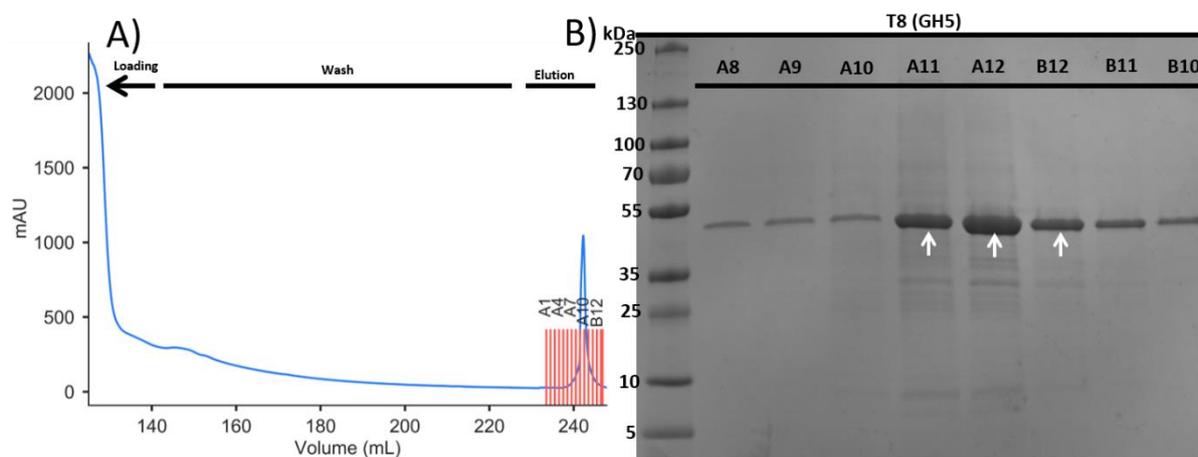


Figure 6.15 Purification of a putative GH5. A) Purification of c_c892255_g1_i1_1 (T8; GH5) from sonicated cell lysate, resulting elution fractions were analysed by SDS-PAGE (B).

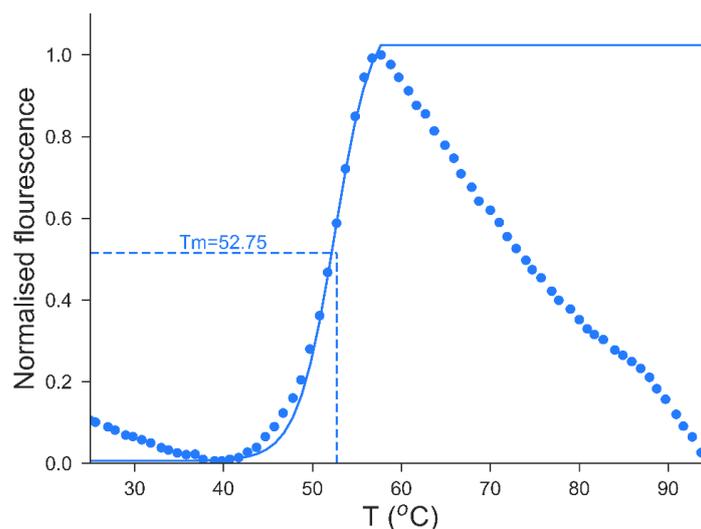


Figure 6.16 A representative thermal shift profile of c_c892255_g1_i1_1 (T8). A sigmoidal curve between the maximum and minimum before the maximum is shown, the median value of this fit was determined as the melting point (T_m).

6.3.5.2 Substrate specificity

Activity was detected using 4-nitrophenyl (pNP-R) oligosaccharides from pNP- β -D-cellobioside to pNP- β -D-cellohexaose, but no significant activity was detected on the cellobiose simulating substrate pNP- β -D-glucopyranoside (pNP-G1) (Figure 6.17). These results would suggest β -glucosidase activity; however, no activity on pNP- β -D-glucopyranoside contradicts this theory. T8 could be a cellobiohydrolase; however, whilst significant activity using pNP-R substrates were observable, no significant activity could be detected using reducing sugar assays (PAHBAH, DNS) on oligosaccharide or polysaccharide substrates. Oligosaccharides as substrates are difficult to measure with reducing sugar assays due to high levels of background noise; however, activity on polysaccharides should have been detectable, particularly when considering the exceptional rates of activity observed on pNP-R substrates. Upon co-incubation with Avicel™ a visible reduction in particle size and amount of substrate was observable (despite a lack of reducing sugars) which encouraged further experimentation (Figure 6.18).

Expression and characterisation of putatively lignocellulolytic enzymes

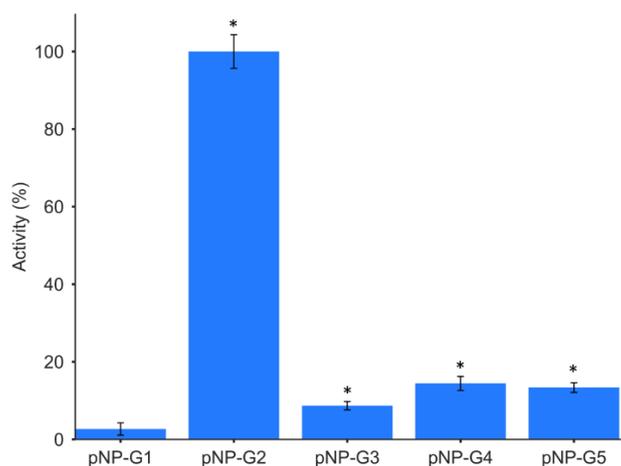


Figure 6.17 Activity of c_c892255_g1_i1_1 on pNP-R substrates. 0.25 μ g on purified enzyme was incubated in 0.5mM substrate for two hours in 20mM sodium phosphate, 280mM sodium chloride, pH 7.6. The reaction was not conducted to saturation. Asterisks indicate significance at a confidence interval of $p < 0.05$ assessed by a one-way ANOVA.

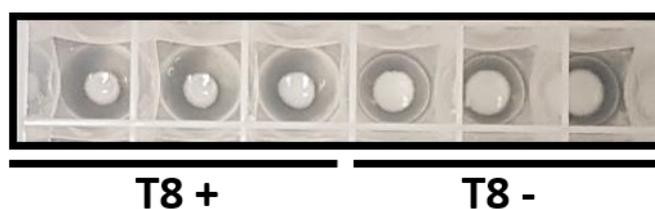


Figure 6.18 Visible reduction in Avicel™ co-incubated with c_c892255_g1_i1_1.

Ion exchange chromatography revealed c_c892255_g1_i1_1 had no activity towards glucose or cellobiose (synonymous findings to the pNP-R substrates), but was capable of hydrolysing cellotriose, cellotetraose, cellopentaose and cellohexaose (Figure 6.19). The more abundant oligosaccharide product was n-1 of the initial substrate, suggestive of β -glucosidase activity. The results suggest that this protein belongs to a transglycosylase superfamily which can both hydrolyse and construct glycosidic bonds. This would explain the lack of significant activity with reducing sugar assays, as the hydrolysis products may be being used as building blocks for polymerisation. The product profile contains both products with a lesser degree of polymerisation (indicative of hydrolysis products) and larger degrees of polymerisation (indicative of trans glycosylation) than the original substrate, that were not comparable to the standard oligosaccharide cocktail. If the hydrolysis or transglycosylation products were specifically 1,4- β -glycosidic bonds, the profile should mirror the standards profile suggesting a degree of branching may be occurring (1,3- or 1,6- β -linkages).

Expression and characterisation of putatively lignocellulolytic enzymes

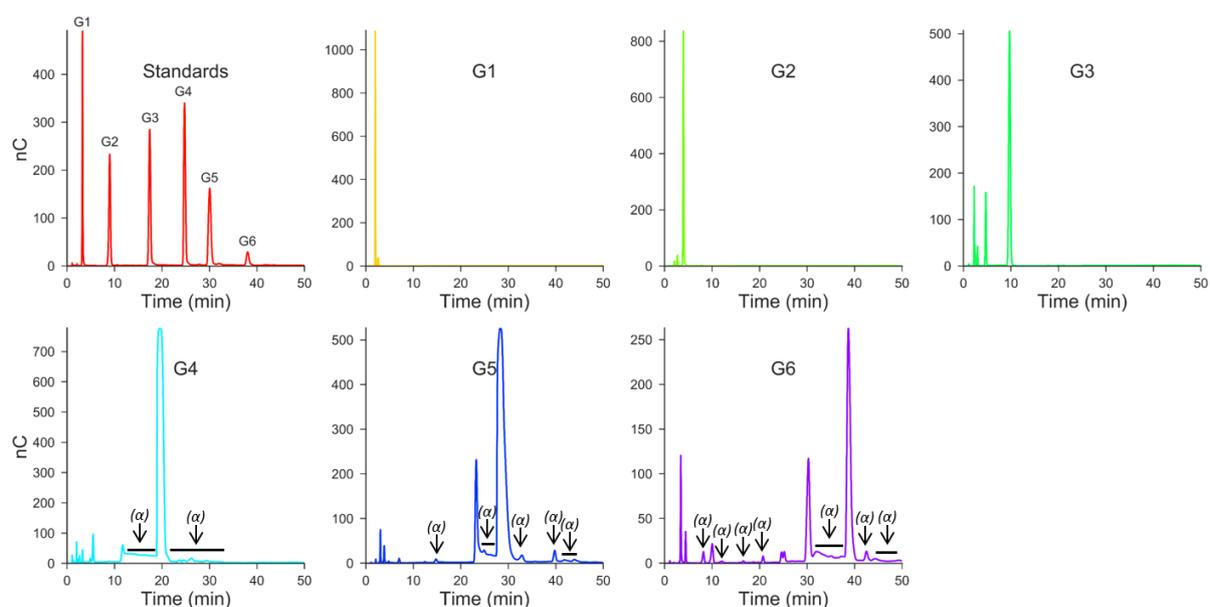


Figure 6.19 Product profile of c_c892255_g1_i1_1 after incubation with oligosaccharides. 1 μ g of purified enzyme was incubated in 0.5mM substrate in 50mM sodium phosphate pH 7.6 for four hours, the resultant hydrolysate was with ion exchange chromatography. (α); unidentifiable products.

6.3.5.3 Bivariate pH, temperature optima

Using pNP-cellobioside as a substrate, 0.05 μ g T8 (c_c892255_g1_i1_1) was incubated for one hour which elucidated the optimal reaction conditions which were determined to be 43.8 $^{\circ}$ C and pH 7.6 (Figure 6.20). The bivariate activity profile exhibited a tight tolerance range with little deviation around the optimum. With respect to environmental conditions, the pH optima is representative of inherent salt marsh conditions; however, the temperature optima is beyond what the enzyme would experience in nature.

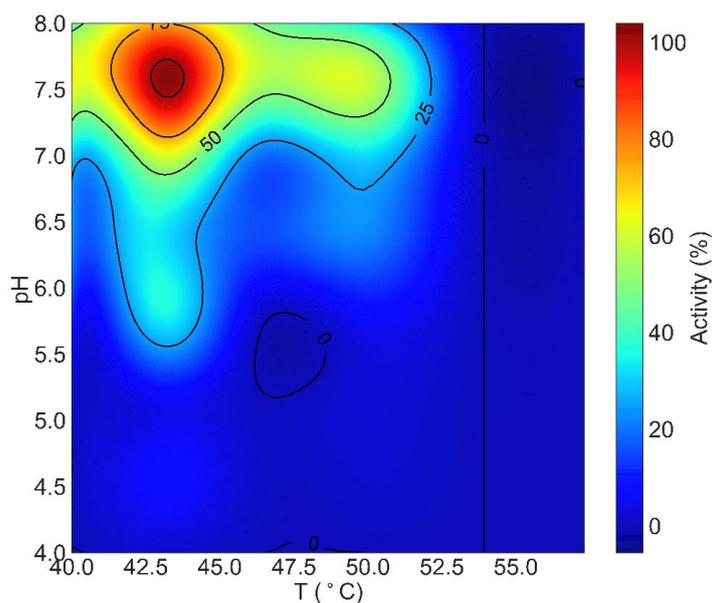


Figure 6.20 Bivariate pH and temperature optima of c_c892255_g1_i1_1 (GH5).

6.3.5.4 Salinity tolerance

An interesting profile for salinity tolerance was observed for T8 (c_c892255_g1_i1_1), where the optimal salinity was between 0.6M and 1M NaCl, representative of artificial seawater (ASW), however the seawater assays displayed a significant decrease in activity between 0.5x ASW (estuarine concentrations) and 1x ASW (marine concentrations) (Figure 6.21). Rate of activity at the optimal salt concentration (1M) was determined to be $2126.6 \mu\text{M pNP} \cdot \mu\text{g}^{-1} \cdot \text{hour}^{-1}$ in 50mM sodium phosphate pH 7.6 and $825 \text{ pNP} \cdot \mu\text{g}^{-1} \cdot \text{hour}^{-1}$ in 0.5x ASW pH 7.6; however, no experiments were conducted to assess if this rate was determined within the linear phase of degradation.

Expression and characterisation of putatively lignocellulolytic enzymes

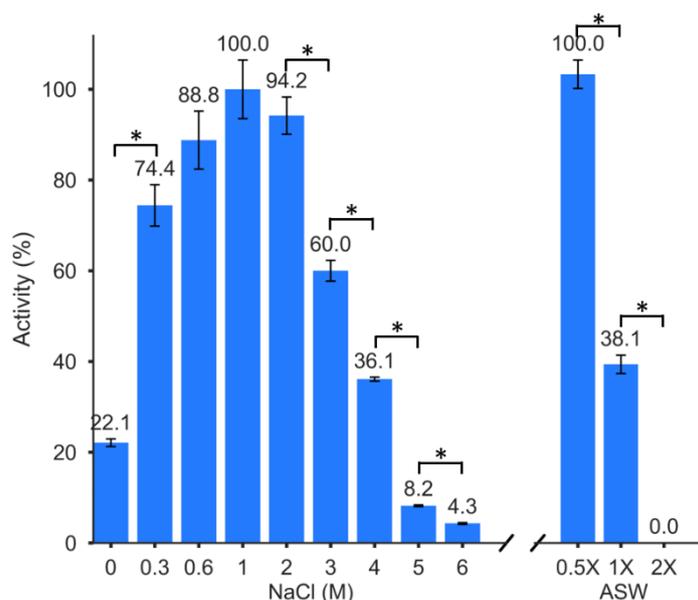


Figure 6.21 Salinity tolerance of c_c892255_g1_i1_1 (GH5). 0.1µg of T8 (c_c892255_g1_i1_1) was incubated with pNP-cellobiose for 1 hour at optimal conditions in 50mM sodium phosphate. Asterisks indicate significance at a confidence interval of $p < 0.05$ assessed by a one-way ANOVA performed on non-normalised, raw data. ASW; artificial seawater.

6.4 Discussion

6.4.1 Cloning approach

The shotgun approach undertaken in this study was undertaken due to the number, diversity and complexity of target sequences requiring a high throughput methodology. The average target sequence contained 2.8 disulfide bonds and 15.8% rare codons which are considered problematic in bacterial expression systems (Rosano and Ceccarelli, 2014). Disulfide bonds are thought to significantly contribute to protein stability, particularly in secreted proteins, which has relevance to the stressful salt marsh environment the proteins originated from (Hatahet et al., 2014). Whilst rare codons are present throughout all prokaryotic lineages, the large proportion observed in these sequences is likely a result of taxonomic divergence between *Bacteroidetes* (*Bacteroides*) and lineages within *Proteobacteria*, as large inter-taxa deviation of codon usage profiles exists down to genus level within *Proteobacteria* (*E. coli* and the majority of target sequences belong to *Gammaproteobacteria*) (Bachvarov et al., 2008).

Despite these potential caveats, a success rate of 60% soluble protein was achieved which is testament to the strategy implemented. Most of the proteins expressed better at mild induction conditions and lower temperatures. Therefore, it is likely that these proteins are constrained at

some point in their production pathway and the rate of production, as a function of induction intensity, inhibits the production of functional protein. This is particularly notable in T6 (a_c580607_g3_i1_2) where basal (leaky) expression produces putatively intact protein and the mildest condition produces fragmented protein.

6.4.2 The GH6 (a_c570829_g1_i1_1)

The activities contained within the GH6 family are categorised as endo-1,4- β -glucanases (endoglucanase; EG) and exo-1,4- β -glucanases (cellobiohydrolases; CBH). These are integral to cellulose deconstruction and in optimised cocktails constituting around 20-40% and 40- 60% respectively, inferring CBH are the most important constituent (Billard et al., 2012, Banerjee et al., 2010, Teeri, 1997). This is also reflected in the system level analysis where GH6 was the most abundant enzyme family.

The product profiles for a_c570829_g1_i1_1 together with the presence of three CBM domains suggest it is a processive cellobiohydrolase (CBH), and CBH within GH6 are thought to act from the non-reducing ends of cellulose which would classify this enzyme as a CBHII (Cantarel et al., 2009). It has a processive index of 2.65 ± 0.06 on crystalline cellulose and 11.77 ± 1.76 on amorphous cellulose. The disparity likely arises from the increased ease in accessibility of cellulose microfibrils (as processive enzymes complex with the chain terminus (Jalak et al., 2012)) decreasing the number of false starts (which produce cellotriose dependent upon the stereochemistry of the enzyme bound chain, cellobiose is produced thereafter) which is inferred from the cellobiose to glucose and cellotriose product ratios (Vuong and Wilson, 2009, Wilson and Kostylev, 2012).

The unusual affinity of a_c570829_g1_i1_1 for crystalline cellulose (more so than amorphous cellulose) likely arises from the carbohydrate binding domain belonging to family 2 (CBM2), which has been shown to target crystalline cellulose, however the CBM44 domains also have an affinity for amorphous cellulose (Gourlay et al., 2012). CBMs are generally thought to increase the activity of the enzymes they are conjugated to, by retaining proximity to the substrate (Reyes-Ortiz et al., 2013). This may be particularly pertinent for this enzyme in a salt marsh environment where tidal submersion could disassociate the enzyme into open waters, to the detriment of the secreting organism.

The refolding capability (reversible renaturation) of a_c570829_g1_i1_1 has been reported to a lesser degree for a *T. reesei* GH7 (Cel7A) treated at 80°C for 30 min (without salt) and allowed to refold at 25°C which retained 67% activity; however, in the presence of 1mM DTT all activity was lost

suggesting disulfide bonds were responsible (Boer et al., 2000). The thermal shift assay displayed refolding (to an unknown degree) of a_c570829_g1_i1_1 after 15 min at 100°C in 100mM DTT and 2% sodium dodecyl sulfate; however, activity after this treatment was not assessed. This is presumably due to a protective role that salt plays against denaturing agents for halo enzymes (Sinha and Khare, 2014). This phenomenon has been observed with other halo enzymes, a β -lactamase retained 75% of its activity after heat treatment at 100°C in 200mM NaCl for 5 min (synonymous conditions were tested in this study) and was observed to completely refold, suggesting this may be a common but unexplored property of halo enzymes (Tokunaga et al., 2004).

Despite the importance of cellobiohydrolases, there are few salt tolerant representatives (Zhang et al., 2011b, Asha and Sakthivel, 2014). This GH6 is tolerant up to 6M NaCl and seawater as would be expected from a halotolerant or halophilic enzyme. The halo tolerance profile would suggest it is tolerant of, rather than adapted to as there is a significant decline in activity between non saline and saline media. Taken together with the low pH optima of 4.8-5.5 which is the common optimum range for terrestrially derived lignocellulolytic enzyme, would suggest the enzyme is of terrestrial origin with adaptations to function within the saline, metal rich environment of the salt marsh.

These adaptations also appear to have consigned considerable plasticity in pH and temperature tolerances within which high activity is retained. This is common feature in halotolerant proteins where adaptations include reduced overall hydrophobicity, greater proportions of acidic surface residues (especially aspartic acid, which constitutes 8.74% of a_c570829_g1_i1_1), fewer cysteine residues and tendencies toward coil structures over helices (Paul et al., 2008, Jaenicke, 1991). These features confer polyextremophilic properties for halo enzymes where a tolerance to one factor often confers tolerance to others, mediated by the presence of salt (Sinha and Khare, 2014, Amoozegar et al., 2017). Similar plasticity in environmental tolerance been observed for an endo-acting halophilic GH6 isolated from *Thermobifida haloterans* YIM 90462(T) (Yin et al., 2015b).

The observed reaction rate 'slowdown' during hours three and four at both temperatures occurs at such low product concentrations steric inhibition should not occur (although this was not tested), suggesting another mode of action is responsible. This slowdown has been reported previously for enzymes targeting cellulose and is ostensibly a result of transient enzyme inactivation and blocking causing 'traffic jams' (blocking by nonproductive enzyme binding to cellulose residues, transient inactivation during processing), which results in an initial 'burst' phase of activity (Bansal et al., 2009, Praestgaard et al., 2011, Igarashi et al., 2011). The rate of activity for a_c570829_g1_i1_1 (during burst phase) was measured at 1.03 μmol glucose reducing sugar equivalents. $\text{mg}^{-1}.\text{min}^{-1}$ at 45°C with

Avicel™ substrates is greater than but comparable to *Trichoderma reesei* enzymes CBHI and CBHII which exhibit rates of 0.035 and 0.027 μmol glucose reducing sugar equivalents. $\text{mg}^{-1}.\text{min}^{-1}$ respectively (Billard et al., 2012).

6.4.3 The GH5 (c_c892255_g1_i1_1)

The GH5 family contains the broadest scope of activities of any GH family (Cantarel et al., 2009). The selection of this enzyme was a result of high confidence levels (1.2e^{-45} and 2.2e^{-32} for dbCAN and pFAM respectively) suggesting this protein was a cellulase, coupled with low sequence identity in both nucleotide and protein databases indicating prospective novelty. Unlike a_c570829_g1_i1_1 (GH6), the pH and temperature optima profile indicate a protein adapted to a salt marsh environment where the average pH is 7-7.4 (Baumann et al., 2015). The lack of plasticity around the optimum is likely reflective of tight environmental conditions. The halo tolerance profile also presents an obligate halophilic relationship where salt is required to achieve optimal activity, and as such is adapted to, rather than tolerant of these conditions. Interestingly the optimal ionic concentration of sodium chloride was measured between 600mM and 1M which is within the upper bounds of natural limits.

The product profile suggests this enzyme is likely a β -glucosidase which cleave terminal residues from the reducing or non-reducing ends of oligosaccharides; however, the profile also suggests transglycosylase activity. The enzyme has homology to a transglycosylase superfamily which confers the capacity to both synthesise and hydrolyse glycosidic bonds. The potential of transglycosylases to produce a range of industrially relevant functional oligosaccharides is known (Mangas-Sanchez and Adlercreutz, 2015). The product profile is suggestive of a range of target activities (hydrolysis of β -1,4, and transglycosylation of β -1,3 or β -1,6), a β -glucosidase with β -1,3, β -1,4 and β -1,6 activity has been identified (Lafond et al., 2012). Interestingly another broadly active GH5 β -glucosidase of marine origin has exhibited β -1,3 and β -1,6 activity with transglycosylation capability to produce larger cello-oligosaccharides (Wang et al., 2016a, Wang et al., 2018b, Park et al., 2005).

It is accepted that most glycoside hydrolases can transglycosylate as well as hydrolyse and the tendency between these processes can be manipulated by increasing substrate and 'building block' (e.g. glucose or cellobiose) concentrations. In this study, no building blocks were supplied and must have originated from hydrolysis of the oligosaccharide. This raises the question as to whether the activity observed in c_c892255_g1_i1_1 is true to nature where the experimental conditions measured were conducted at an unnatural (to salt marsh environments) substrate concentration,

inferring the enzyme may act solely as a β -glucosidase naturally. If this enzyme does transglycosylate in natural conditions, then the observed action on Avicel™ may be to make the structure more amorphous by branching and accessible to attack by other enzymes. However, this enzyme lacks a CBM and as such is unlikely to target crystalline substrates *in situ*.

Whilst the function of this protein has not been fully elucidated, it may be beneficial to cellulase cocktails. Many important cellulolytic enzymes are inhibited by their products through a steric hinderance mechanism, whereby the rate of activity decline is correlated to the product concentration (Olsen et al., 2016). This inhibition could be alleviated with the use of transglycosylation to transform the products, reducing specificity of the products to the active site; a strategy that has previously been hypothesised (Wang et al., 2017). Crystalline cellulose is notoriously recalcitrant and creating complex oligosaccharides from simple products may appear counter intuitive, but these products should be readily deconstructed by other glycosyl hydrolases present in the cocktail. The removal of product inhibition by this enzyme may improve the overall rates of reaction for current cellulase cocktail mixtures.

It has long been known bacteria are capable of synthesising cellulose (Brown, 1886). If hydrolysis is a legacy activity and the main function of this enzyme is to synthesis polymers, it may have a role in the extracellular synthesis of exopolysaccharides (EPSs). EPSs are produced predominantly by marine bacteria to form a matrix (biofilm) of linear or highly branched, hetero- or homopolysaccharides with predicted functions including desiccation resistance, stabilisation (against wave action), energy reserves, aggregate formation, cryoprotection and nutrient sequestration (Poli et al., 2011, Casillo et al., 2018, Liu et al., 2017a). It is thought EPSs are secreted (imposing an energetic burden on cells) or synthesised by extracellular cellulosome components, however this enzyme had a well-defined signal peptide indicating it was secreted, but no obvious affiliation with an extracellular complex (Rehm, 2010).

Interestingly this enzyme has 42% sequence homology to a putative cellodextrinase (endo-acting cellulases predominantly targeting short chain oligosaccharides) from an uncultured organism (NCBI NR_prot). Cellodextrinases have been observed to bind to crystalline cellulose (Ferreira et al., 1991). Dextrin is a homopolymer consisting of 1,4:1,6- α linkages (Rehm, 2010). To conclude the mechanism of action for this protein and further probe the products, further work must be undertaken to elucidate the scope of activity (β -1,3 and β -1,6) and the building blocks used to transglycosylate (it may utilise alternative sugars for heteropolymer synthesis).

EPSs are established as industrially relevant, most notably as gelling agents (e.g. dextran, xanthan and gellan) but also as medical, nutraceutical and pharmaceutical agents (Nwodo et al., 2012).

Expression and characterisation of putatively lignocellulolytic enzymes

Despite their broad use and potential for use, the underlying mechanism to their synthesis is poorly understood. Currently, most commercial EPSs are extracted from bacterial cultures rather than direct enzymatic synthesis. The potential industrial applications of a transglycosylase with these properties that have not been reported previously is novel and encouraging. Whether this protein is involved in lignocellulose deconstruction, the extracellular synthesis of EPSs or both, remains an area for further investigation.

Chapter 7 Final discussion

The aim of the work presented in this thesis was to identify and characterise novel halotolerant lignocellulolytic enzymes with biotechnologically attractive attributes whilst synchronously exploring the functional ecology of carbon cycling within the salt marsh ecosystem. It was hypothesised that as the studied salt marsh was predominantly marine, the saline conditions would confer selective pressures for halotolerant enzymes not found in terrestrial ecosystems. Halotolerance would facilitate the replacement of valuable freshwater for biorefining applications with abundant, accessible and inexpensive seawater, alleviating water security concerns. Highly active enzymes were sought to aid in reducing enzyme loadings improving biorefining economics.

An *in situ* decomposition experiment was conducted to mimic natural lignocellulose decomposition processes in a salt marsh environment for sixteen weeks (Chapter 3). Mass loss and biomass composition analysis revealed a hierarchy of limitation whereby physical complexity impedes accessibility. Rate limiting factors (lignin and matrix polysaccharides) alternate and oscillate for up to sixteen weeks before degradative equilibrium is observed for all fractions. During this time, all lignocellulose fractions analysed were degraded, suggesting a diverse suite of lignocellulolytic enzymes were present at all time points. Quantitative enzyme assays revealed an increase in activity over time, except for the first two weeks which exhibited significantly lower activity. A temporally interspersed strategy was implemented to capture the molecular process of decomposition through RNA and extracellular protein sequencing.

The carbohydrate active enzyme (CAZy) profile (Chapter 4) elucidated by matching the RNA sequencing data and peptide spectra from the extracellular proteins revealed a suite of consistently abundant enzyme families. These dominant classes included predominantly cellulase targeting families GH5, GH6, GH3, GH9 followed by less abundant hemicellulase targeting families GH11, GH10, GH16 and GH43. The carbohydrate binding domains (CBMs) explored in this study reveal potential avenues for novelty (CBM53s are associated to hemicellulases). The dominant CBMs were predominantly cellulose associated (CBM2, CBM44), which reinforce an energetically favorable carbon utilisation strategy (where cellulosic-derived glucose is valued highly despite inherent inaccessibility, as observed in the biomass composition analysis).

Many enzymes in this study belong to diverse families that are still poorly characterised which inhibits assignment of enzymatic function based on sequence. Therefore, the absolute enzymatic activity acting upon each lignocellulosic fraction could not be determined until current databases are mature. This is a major limitation in the molecular exploration of novel ecosystems.

Final discussion

The scarcity of lignin modifying enzymes, such as the oxidative AA2 family, was surprising but CE1 family enzyme abundance implicates this family as the dominant lignin associated family in this environment. Enzymes within the CE1 family can operate independent of oxygen by disconnecting lignin polymers at the lignin-carbohydrate complex interface, increasing accessibility to core polysaccharides, negating the requirement for further modification in the O₂ limited salt marsh sediments. This allusion has biotechnological potential as an enzymatic pretreatment (removal of pristine lignin without the need for acid, alkali or ionic pre-treatments). Additionally, as the CE1 family is observably abundant in the natural salt marsh lignocellulolytic cocktail, but not in commercially available cocktails; representation in commercial cocktails may be beneficial.

Phylogenetic homology of the metasecretome revealed the dominant CAZy producing taxa to belong to *Proteobacteria* (*Gammaproteobacteria* and *Deltaproteobacteria*) and *Bacteroidetes* (*Bacteroidia*, *Flavobacteriia* and *Cytophagia*; Chapter 4). Both exhibited similar CAZy family profiles, suggesting a core suite of enzymes had been inter-taxonomically selected for in this environment. The key families regulating lignocellulose disassembly were (in descending order) *Flavobacteriaceae*, *Alteromonadaceae*, *Cytophagaceae*, *Saccharospirillaceae*, *Marinilabiaceae*, *Flammeovirgaceae*, *Bacteroidaceae* and whilst transient *Vibrionaceae*. Families with the highest productivity indexes included *Cytophagaceae*, *Bacteroidaceae* and *Sphingobacteriaceae*.

CAZy producing taxonomic groups as identified from protein homology were notably enriched (281% enrichment) in the composting biomass sampling to the sediment outgroup (Chapter 5). Temporal shifts in abundance displayed a decline in *Gammaproteobacterial* taxa, encroached upon initially by *Bacteroidetes* and thereafter by *Deltaproteobacterial* enrichment. *Gammaproteobacterial* families were not as abundant as *Bacteroidetes* in sediment samples, however they exhibit early and rapid opportunistic colonisation patterns. Additionally, *Flavobacteriaceae* were the most abundant CAZy producing taxa in the day 0 outgroup and were the most productive throughout the time course (*Cytophagia* also exhibited a similar pattern). The most abundant CAZy producers in the sediment samples were *Deltaproteobacteria*, however their productivity is significantly less than *Gammaproteobacteria* and *Bacteroidetes*. For these reasons, *Bacteroidetes* (specifically *Flavobacteriaceae*, *Cytophagaceae*, *Sphingobacteriaceae*) likely function as the dominant carbon cyclers in salt marsh sediment, where *Gammaproteobacteria* (specifically *Alteromonadales*, *Cellvibrionaceae*, *Saccharospirillaceae* and *Vibrionaceae*) are likely opportunistic carbon cyclers, heavily involved in the initial above ground decomposition processes.

To demonstrate halotolerance of enzymes sourced from salt marsh ecosystems, the analysis was assimilated to select protein targets based on their perceived importance at the ecosystem level

with respective consideration to their abundance, potential function and taxonomic origin (Chapter 6). The most abundant enzyme was a putative GH6 (most abundant enzyme family at ecosystem level) containing two CBM domains (CBM2 and CBM4 belonging to the two most abundant CBM families at the ecosystem level) with putative *Gammaproteobacterial* origin. This enzyme was characterised as a potential non-reducing end acting exo-1,4- β -glucanases (cellobiohydrolase; CBHII) with halotolerance up to saturated NaCl solutions and no significant decline in activity in naturally occurring seawater concentrations. An additional enzyme consisting of a single GH5 domain of low confidence and low identity was only partially characterised due to time constraints. Evidence suggests this enzyme is a β -1,4-glucosidase with transglycosylase activity, with potential environmental functions in biofilm synthesis. This enzyme was also tolerant in NaCl and naturally occurring seawater concentrations (with an optimum in an estuarine range).

The two enzymes characterised in this study provide evidence that salt marshes are a good environment for bioprospecting halotolerant, alkaliphilic, lignocellulolytic enzymes. The techniques in this study have also alluded to potential novel mechanisms and emphasise the requirement for further research, particularly for taxonomic groups with implicit but unknown functions in carbon cycling (such as the aromatophilic and anaerobic *Chloroflexi* and *Firmicutes*). The poor molecular characterisation of salt marshes and wealth of information generated in this study demonstrate that salt marshes contain a vast reservoir of unexploited biotechnological potential. This reinforces the requirement for salt marsh conservation and management to protect these previously undescribed biotechnological troves and the persuasive level of diversity and novelty in these databases emphasise the potential of molecular exploration in salt marshes for industrially relevant enzymes and organisms.

7.1 Future perspectives

The wealth of information generated in this study is the first in depth account of lignocellulose associated metasecretome and microbiome from a salt marsh and provides a foundation through which halotolerant lignocellulolytic enzyme cocktails can be constructed. The highly active cellobiohydrolase characterised in this study (CHBII), can be used to complement halotolerant enzymes sourced from these databases to replicate the ecosystem cocktail profiled in this study or to reconstruct commercially available cellulase cocktails. This would include a reducing end acting cellobiohydrolase (CBHI), an endo-1,4- β -glucanase (EG) and a β -1,4-glucosidase.

The β -1,4-glucosidase in this study with apparent transglycosylase activity requires further investigation to reveal the mechanism of transglycosylation and whether this function is observable in environmental conditions or a product of unrealistic *in vitro* conditions (inflated substrate or monomer concentrations). Transglycosylase activity, particularly for the construction of complex polysaccharides, if controllable, is desirable in the pharmaceutical sector. Refined complex polysaccharides also have a major role in the food market as gelling agents and viscosity inducers. The levels of activity of this enzyme and the unusually broad range of accepted substrates considered in conjunction with the potential biotechnological and industrial relevance makes it an attractive prospect for further research.

These databases also contain a diverse suite of potentially halotolerant hemicellulases to be exploited which are known to improve lignocellulolytic cocktails. Whilst the evidence in this study suggests low to moderate lignin modification in this environment, it has alluded to the enzymatic cleavage of lignin from the core polysaccharides by the highly abundant CE1 family. While further evidence to support this is required, it creates an avenue through which lignin cleavage could be possible (e.g. with carbohydrate esterase family 1 enzymes) independent of economically and resource intensive pre-treatment. Currently, no effective enzymatic biomass pre-treatments have been developed using carbohydrate esterases and these enzymes are not fully represented in commercial cocktails. Supplementation of this family into cocktails for grass derived feedstocks would be an appropriate continuation of this finding.

In the present study, cellulase targeting enzyme families were more abundant than hemicellulases and lignin targeting groups. However, as sequence-assigned enzyme families are highly diverse, ascribing function to the sequences within the enzyme families is still unachievable and significant overlap exists within families (e.g. GH5). In this study semi-quantitative methods were utilised to ascribe approximate relative abundance to proteins (EmPAI derived molar percentage). Methodologies exist for absolute quantification of peptides (such as amine-reactive labelling systems). Absolute quantification would be invaluable when archive databases mature to the extent where function can be assigned from sequence, consequently constructing activity profiles from sequence data. However, the current level of characterisation in archive databases is likely insufficient to generate superior outcomes to the approximation-based method in the current study, but it remains an outstanding area for improving in future research.

Mechanisms of anaerobic modification of the aromatic polymers of lignin is still an enigma and little evidence of their existence has been reported. However, the findings of this study in consideration of the wider field would implicate *Firmicutes* (particularly Clostridia) and *Chloroflexi* as

interesting taxonomic targets to elucidate these mechanisms. This is a particularly valuable insight given the recalcitrant nature of lignin and laborious processing attributable to overcoming this feature. Future studies should attempt to characterise carbon cycling at depth, in oxygen depleted sediments, which may reveal oxygen independent mechanisms of aromatic modification.

The functional assignments of CAZy productivity to taxonomic groups in this study could be used as a predictive template in targeting organisms from these groups for in depth characterisation of their respective CAZy armament. It is likely groups that were both highly abundant and highly active (*Flavobacteriaceae*), utilise different mechanisms and life strategies to the lowly abundant yet highly productive (high productivity index) families (*Cytophagaceae*, *Bacteroidaceae* and *Sphingobacteriaceae*). Elucidating the reasons for the disparity may be insightful, particularly if differences in enzymatic activities are observed.

The observation that *Flavobacteriaceae* are highly productive carbon cyclers both within the sediments and detrital composts (among other highly productive families) could implicate them as potential bioindicator species. Additionally, the families with high productivity indexes (and low abundances) may be suited as bioindicator species, as their low abundance may make them more susceptible to unfavorable environmental conditions. Detrital cycling and retention is critical to salt marsh accretion and ecosystem health. The current study was conducted in Welwick salt marsh which is long established, healthy and accreting. Comparisons to unstable salt marshes (eroding and receding in area) would provide a basis upon which bioindicator species could be assigned.

The relatively small CAZy producing consortia identified in this study could also provide a template for assessing the status of realigned salt marshes. Microbiome assessment of nascent salt marshes that have been freshly realigned could reveal the propensity of that microbiome to degrade detritus. A highly active microbiome (a microbiome abundant in key taxa identified in Chapter 4 and 5) will assist in salt marsh accretion, stabilisation and contribute to the long-term success of the realignment. This could be an invaluable tool given the current focus on salt marsh rehabilitation and reclamation.

The observation that prokaryotes are the dominant CAZy producing domain designates them as targets for further exploration. The focal groups for lignocellulolytic exploration have been predominantly fungi, however as evidenced in this study and many *in vitro* microbiomes, bacteria are equally capable of lignocellulose deconstruction. This is the first instance of carbon cycling in an aerobic natural ecosystem that is not fungal dominated. The deterministic factors for this should be further elucidated (unfavorable environmental conditions for fungi or a highly specialised prokaryotic consortia).

Final discussion

In this study the deconstruction of newly introduced lignocellulose is studied. This process predominantly occurs at the surface to shallow sub surface level. The observed taxonomic restructuring in favor of anaerobic lineages suggests oxygen limitation may result in heterogeneous stratification of the decomposition process. Only the top few millimeters of sediment is reliably aerated, therefore, it is highly likely that the surface level functional consortia identified in this study is significantly different to the consortia responsible for cycling in the aromatic rich, deep, anoxic sub surface sediments. This stratification has been observed at the community level (Petro et al., 2017) but not at a functional level. This consortia may employ novel deconstructive strategies that have yet to be discovered and provide an avenue for future research.

During the course of this study, the utility of the approach has been demonstrated in a salt marsh environment; however, it is readily transferable to characterize lignocellulose decomposition in other ecosystems. The novelty lies in the enrichment of putatively substrate bound enzymes that would be overlooked with conventional approaches and its ability to identify enzymes involved in lignocellulose decomposition and delineate their most likely taxonomic origin. This technique can be applied to any environment with readily cycled organic carbon or detrital stocks, such as soils, sands and leaf litters which may be insightful for identifying functional members of the communities. This approach yielded insightful results despite the poor level of characterization in salt marshes. Therefore, microbial cycling of any material can be achieved with this approach following the application of a native substrate (which can be achieved for example with mesh bags as utilized in this study) in any terrestrial, aquatic or marine environments. The approach is not limited to lignocellulose, it can be applied to any substrate and may be particularly useful for other polymeric or complex substrates, such as plastics, that may have bound enzymes.

The strength of this approach, in contrast to culture dependent methods, is that it can capture the activity of an entire microbiome at the point in time where they are conducting a process. Culture dependent methodologies attempt to identify an organism and assess its potential activities *in vitro*. While these efforts provide insight into potential functions or an involvement in processes, *in vitro* conditions are divorce from environmental attributes and so the observed activities are often artificial or not environmentally relevant. These observations are then often generalized for closely related taxa, which magnifies errors carried through as a result of environmental dissimilarity to *in vitro* conditions. Too often, the relationship between ecosystem processes and microbiome frameworks are implicit or assumed rather than experimentally validated, as they are informed from environmentally divorced, culture dependent methodologies. The integrated approach in this study provides experimental validation for the involvement of these taxonomic lineages in these processes and the degree to which they are involved.

Final discussion

While this approach appears strong for capturing processes as they occur, it is reliant on databases which have largely been generated through culture dependent methods. While not all organisms are cultivable, archive strains are integral to the generation of primary data, particularly given the reliance of metagenomics on primary genomes as reference datasets. Furthermore, culture dependent methods are crucial in understanding organism physiology, as physical manipulation of an organism allows further experimental resolution than any 'omic technique currently available and biotechnological exploitation of an organism requires cultivation. Therefore, these methods should be considered co-dependent as they are complimentary such that one informs the other in a cyclic manner. For example, despite microbes native to salt marshes being poorly characterized, the level of characterization was sufficient to yield insightful results and synonymously reveal that the native salt marsh inhabitants are likely divergent from current archive strains and should be targeted in future culture studies.

There are three major outcomes of this study. The first is the experimentally validated identification of functional taxonomic groups involved in lignocellulose deconstruction in a salt marsh ecosystem. The second is the generation of proteomic databases with enzymes that have been biochemically characterized as halotolerant and have biotechnological potential for use in industry to address current concerns in the bioprocessing of lignocellulosic feedstocks. In addition, there is residual novelty within these databases as thousands of proteins remain to be researched. The approach undertaken in this study is fundamentally the consolidation and linkage of archive datasets to construct an interpretable image of a biological process. The final outcome of this study is the validation of this approach, the concept of which was proven in this study by characterizing the complex biological process that is lignocellulose decomposition in a natural salt marsh.

Appendix

Appendix

T1 (a_c570829_g1_i1_1)

| | | |
|-----|---|-----|
| 0 | MNF RRL EMR KGK KQI RLL PLV FAI GAA ISS IGA QAA SCS YSL TNE WDK | 48 |
| 49 | GFQ GVV TIT NDE NAD IND WQI GMK FPA GIA VTQ SWS GKL SGS NPY YIK | 96 |
| 97 | NTE WNG TIK TGQ SLA VNF MGT RNN LSA ANV ELM GST CSG SDS PPS NTL | 144 |
| 145 | PDA VAD ATP VSG IAP LEV SFS GTR SSD EDG DAL TYI WDF GDG ETA TGK | 192 |
| 193 | DVL HTY TEP GQY TAK LTV DDG TGV DTA TAV ISV TTE SGN TPP VAA MTA | 240 |
| 241 | TPA SGQ APL LVN FDA SAS TDA DND VLT YIW DFG DGA TGS GVS TSH TYA | 288 |
| 289 | DAG KYT VSL IVT DGI DTN KII KTI NVT DDD VPP TTG RVD NPF RDA KWY | 336 |
| 337 | VNP EWS AKA AAE PGG SAI SDQ NTA VWL DRI GAI AGT EDK MGL RGH LDE | 384 |
| 385 | ALE QGA NLF TVV VYD LPN RDC KAV ASN GEL LIS ENG IER YRE EFI DPI | 432 |
| 433 | AEI LAD PKY QSI RIA AVI ELD SLP NLV TNL DIP KCA EAN GPG GYR EGI | 480 |
| 481 | TYA LNK FSP IKN VYS YID AAH SGW LGW DSN FGP AVT LIS NVI KGT DAE | 528 |
| 529 | WDS VAG FVT NTA NYS PLV ETY LPN PDK NVG GGP IRS ADF YEW NPN FDE | 576 |
| 577 | KGY AIE FRK RMI AKG APT TIG MLI DTG RNG WGG PER PTQ VSS SSD KNT | 624 |
| 625 | YVD ESR IDR RYH RGN WCN QPS GIG YKP WAD PYN GVD AFV WVK PPG ESD | 672 |
| 673 | GMS EPD FQP DPD DPA KQY DPM CDP DKM SNA AAY PTG AMD EAP HAG RWY | 720 |
| 721 | SEG FQY LVE NAY PPV SEP AGP PAE | 744 |

T8 (c_c892255_g1_i1_1)

| | | |
|-----|---|-----|
| 0 | MRK LYS IIV AIF ITS AFF SCN SGN GKE ETK SEK SET SDM YYT PNP SGF | 48 |
| 49 | VIS KGV NLS HWL SQS FGW SPK DTY ITE DDI QAI KDF GYD HVR LPV DED | 96 |
| 97 | ELW YED GTF KDT VLN YVK SCI DWC IKR DLR VVF DLH ILR SHH FNA RNG | 144 |
| 145 | EGK ITL WSD TSA QSN FLN LWS SLS DEL KDY PNS MLA YEP MNE PVA PEH | 192 |
| 193 | DMW NDL LIR MVN SIR EKE PER VII FGS NRW QKP YTY PYL KVP END KNI | 240 |
| 241 | ILS FHS YHP YFL THY NAY WSV AKD YKG EVA YPG VTI DEE NFQ RYV DTT | 288 |
| 289 | DSA LMA RIH EEK ALF HHD RDE LVK IFQ PAI DKA VKM KLQ LYC NEF GCL | 336 |
| 337 | PSA PLE MRK QYY SDI KSI FEV YNI AFA NWD YKG DFG IVG WDR ENY KNL | 384 |
| 385 | EPD TAI INI LTR | 396 |

Abbreviations

Abbreviations

| | |
|-----------------|---|
| [Me]GlcA | methylated glucuronic acid |
| 5C | pentose sugar |
| 6C | hexose sugar |
| AA | auxiliary Activities |
| ABTS | 2,2'-azino-bis |
| ACD | above chart datum |
| AESBF | 4-benzenesulfonyl fluoride hydrochloride |
| ANOVA | analysis of variance |
| APS | ammonium persulfate |
| Araf | ferulic acid cross linked arabinosyl side chain |
| ASW | artificial seawater |
| BLAST | basic local alignment search tool |
| CAZy | carbohydrate active enzyme |
| CBH | cellobiohydrolase |
| CBM | carbohydrate binding module |
| cDNA | complimentary DNA |
| CE | carbohydrate Esterases |
| CMC | carboxymethyl cellulose |
| CTAB | cetyl trimethylammonium bromide |
| dbCAN | database for automated carbohydrate-active enzyme annotation |
| DEPC | diethyl pyrocarbonate |
| DNS | 3,5-dinitrosalicylic acid |
| DP | degrees of polymerisation |
| DTT | dithiothreitol |
| emPAI | exponentially modified protein abundance |
| EV | empty vector |
| FPKM | fragments per kilobase of transcript per million mapped reads |
| G | guaiacyl |
| GAX | glucoarabinoxylan |
| gDNA | genomic DNA |
| GH | glycoside Hydrolases |
| GT | glycosyltransferases |
| H | p-hydroxyphenyl |
| HMF | 5-hydroxymethylfurfural |
| HMM | hidden markov model |
| HPLC | high performance liquid chromatography |
| IPTG | isopropyl β -D-1-thiogalactopyranoside |
| ISDE | <i>in situ</i> decomposition experiment |
| ITS | internal transcribed spacer |
| LB | luria broth |
| LC-MS/MS | liquid chromatography tandem mass spectrometry |
| LCC | lignin-carbohydrate complex |

Abbreviations

| | |
|------------------------|---|
| LP | lignin peroxidase |
| LPMO | lytic polysaccharide monooxygenase |
| MBO | million barrels of oil |
| MnP | manganese peroxidase |
| mRNA | messenger RNA |
| na | not applicable |
| NA | not assigned |
| NAPP | net aerial primary production |
| NCBI | national center for biotechnology information |
| NPP | net primary production |
| NR_prot | nCBI non-redundant protein database |
| ORF | open reading frame |
| OTU | operational taxonomic unit |
| P3HB | poly-3-hydroxybutyrate |
| PAHBAH | 4-hydroxybenzoic acid hydrazide |
| PASC | phosphoric acid swollen cellulose |
| PBS | phosphate buffered saline |
| pCO₂ | partial pressure of CO ₂ |
| PCR | polymerase chain reaction |
| PDB | protein data bank |
| pFAM | protein family database |
| PHA | polyhydroxyalkanoates |
| PI | productivity index |
| PL | polysaccharide Lyases |
| PMO | peptide matching open reading frame |
| pNP | para-nitrophenyl |
| rRNA | ribosomal RNA |
| S | syringyl |
| SDS | sodium dodecyl sulfate |
| SDS-PAGE | sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| TaxID | taxonomic identified |
| TEMED | tetramethylethylenediamine |
| TFA | trifluoroacetic acid |
| tRNA | total RNA |
| v/v | volume to volume ratio |
| Vp | versatile peroxidase |
| w/v | weight to volume ratio |

References

- ABDELAZIZ, O. Y., BRINK, D. P., PROTHMANN, J., RAVI, K., SUN, M. Z., GARCIA-HIDALGO, J., SANDAHL, M., HULTEBERG, C. P., TURNER, C., LIDEN, G. & GORWA-GRAUSLUND, M. F. 2016. Biological valorization of low molecular weight lignin. *Biotechnology Advances*, 34, 1318-1346.
- AGBOR, V. B., CICEK, N., SPARLING, R., BERLIN, A. & LEVIN, D. B. 2011. Biomass pretreatment: Fundamentals toward application. *Biotechnology Advances*, 29, 675-685.
- AIROLDI, L. & BECK, M. W. 2007. Loss, status and trends for coastal marine habitats of Europe. In: GIBSON, R. N., ATKINSON, R. J. A. & GORDON, J. D. M. (eds.) *Oceanography and Marine Biology*, Vol 45. Boca Raton: Crc Press-Taylor & Francis Group.
- AKINTOLA, A. I., OYEDEJI, O., BAKARE, M. K. & ADEWALE, I. O. 2017. Purification and characterization of thermostable cellulase from *Enterobacter cloacae* IP8 isolated from decayed plant leaf litter. *Biocatalysis and Biotransformation*, 35, 379-387.
- AL-MASAUDI, S., EL KAOUTARI, A., DRULA, E., AL-MEHDAR, H., REDWAN, E. M., LOMBARD, V. & HENRISSAT, B. 2017. A Metagenomics Investigation of Carbohydrate-Active Enzymes along the Gastrointestinal Tract of Saudi Sheep. *Frontiers in Microbiology*, 8, 9.
- ALESSI, A. M., BIRD, S. M., BENNETT, J. P., OATES, N. C., LI, Y., DOWLE, A. A., POLIKARPOV, I., YOUNG, J. P. W., MCQUEEN-MASON, S. J. & BRUCE, N. C. 2017. Revealing the insoluble metasecretome of lignocellulose-degrading microbial communities. *Scientific Reports*, 7.
- ALESSI, A. M., BIRD, S. M., OATES, N. C., LI, Y., DOWLE, A. A., NOVOTNY, E. H., DEAZEVEDO, E. R., BENNETT, J. P., POLIKARPOV, I., YOUNG, J. P. W., MCQUEEN-MASON, S. J. & BRUCE, N. C. 2018. Defining functional diversity for lignocellulose degradation in a microbial community using multi-omics studies. *Biotechnology for Biofuels*, 11, 16.
- ALLDRED, M., LIBERTI, A. & BAINES, S. B. 2017. Impact of salinity and nutrients on salt marsh stability. *Ecosphere*, 8, 10.
- ALVIRA, P., NEGRO, M. J. & BALLESTEROS, M. 2011. Effect of endoxylanase and alpha-L-arabinofuranosidase supplementation on the enzymatic hydrolysis of steam exploded wheat straw. *Bioresource Technology*, 102, 4552-4558.
- ALVIRA, P., TOMAS-PEJO, E., BALLESTEROS, M. & NEGRO, M. J. 2010. Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresource Technology*, 101, 4851-4861.
- AMOOZEGAR, M. A., SIROOSI, M., ATASHGAHI, S., SMIDT, H. & VENTOSA, A. 2017. Systematics of haloarchaea and biotechnological potential of their hydrolytic enzymes. *Microbiology-Sgm*, 163, 623-645.
- ANDREI, A. S., BARICZ, A., ROBESON, M. S., PAUSAN, M. R., TAMAS, T., CHIRIAC, C., SZEKERES, E., BARBU-TUDORAN, L., LEVEL, E. A., COMAN, C., PODAR, M. & BANCIU, H. L. 2017. Hypersaline sapropels act as hotspots for microbial dark matter. *Scientific Reports*, 7, 8.
- ANDREWS, J. E., SAMWAYS, G., DENNIS, P. F. & MAHER, B. A. 2000. Origin, abundance and storage of organic carbon and sulphur in the Holocene Humber Estuary: emphasizing human impact on storage changes. *Geological Society, London, Special Publications*, 166, 145.
- ANDREWS, J. E., SAMWAYS, G. & SHIMMIELD, G. B. 2008. Historical storage budgets of organic carbon, nutrient and contaminant elements in saltmarsh sediments: Biogeochemical context for managed realignment, Humber Estuary, UK. *Science of the Total Environment*, 405, 1-13.
- ANGERMEYER, A., CROSBY, S. C. & HUBER, J. A. 2016. Decoupled distance-decay patterns between *dsrA* and 16S rRNA genes among salt marsh sulfate-reducing bacteria. *Environmental Microbiology*, 18, 75-86.
- ANGERMEYER, A., CROSBY, S. C. & HUBER, J. A. 2018. Salt marsh sediment bacterial communities maintain original population structure after transplantation across a latitudinal gradient. *Peerj*, 6, 21.

References

- ANWAR, Z., GULFRAZ, M. & IRSHAD, M. J. J. O. R. R. A. A. S. 2014. Agro-industrial lignocellulosic biomass a key to unlock the future bio-energy: a brief review. *7*, 163-173.
- APPRILL, A., MCNALLY, S., PARSONS, R. & WEBER, L. 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology*, *75*, 129-137.
- ARAI, H., HOSEN, Y., HONG, V. N. P., THI, N. T., HUU, C. N. & INUBUSHI, K. 2015. Greenhouse gas emissions from rice straw burning and straw-mushroom cultivation in a triple rice cropping system in the Mekong Delta. *Soil Science and Plant Nutrition*, *61*, 719-735.
- ARFI, Y., CHEVRET, D., HENRISSAT, B., BERRIN, J. G., LEVASSEUR, A. & RECORD, E. 2013. Characterization of salt-adapted secreted lignocellulolytic enzymes from the mangrove fungus *Pestalotiopsis* sp. *Nature Communications*, *4*, 9.
- ARMSTRONG, W., WRIGHT, E. J., LYTHER, S. & GAYNARD, T. J. 1985. Plant zonation and the effects of the spring-neap tidal cycle on soil aeration in a humber salt-marsh. *Journal of Ecology*, *73*, 323-339.
- ASHA, B. M. & SAKTHIVEL, N. 2014. Production, purification and characterization of a new cellulase from *Bacillus subtilis* that exhibit halophilic, alkalophilic and solvent-tolerant properties. *Annals of Microbiology*, *64*, 1839-1848.
- ATKINSON, M. & BINGMAN, C. 1997. Elemental composition of commercial seasalts. *Journal of Aquaculture and Aquatic Sciences*, *8*, 39-43.
- BACHVAROV, B., KIRILOV, K. & IVANOV, I. 2008. Codon usage in prokaryotes. *Biotechnology & Biotechnological Equipment*, *22*, 669-682.
- BADHAN, A., WANG, Y. X., GRUNINGER, R., PATTON, D., POWLOWSKI, J., TSANG, A. & MCALLISTER, T. 2014. Formulation of enzyme blends to maximize the hydrolysis of alkaline peroxide pretreated alfalfa hay and barley straw by rumen enzymes and commercial cellulases. *Bmc Biotechnology*, *14*, 14.
- BAI, J. H., ZHANG, G. L., ZHAO, Q. Q., LU, Q. Q., JIA, J., CUI, B. S. & LIU, X. H. 2016. Depth-distribution patterns and control of soil organic carbon in coastal salt marshes with different plant covers. *Scientific Reports*, *6*, 12.
- BAIOCCO, P., BARRECA, A. M., FABBRINI, M., GALLI, C. & GENTILI, P. 2003. Promoting laccase activity towards non-phenolic substrates: a mechanistic investigation with some laccase-mediator systems. *Organic & Biomolecular Chemistry*, *1*, 191-197.
- BANERJEE, G., CAR, S., SCOTT-CRAIG, J. S., BORRUSCH, M. S. & WALTON, J. D. 2010. Rapid optimization of enzyme mixtures for deconstruction of diverse pretreatment/biomass feedstock combinations. *Biotechnology for Biofuels*, *3*, 15.
- BANSAL, P., HALL, M., REALFF, M. J., LEE, J. H. & BOMMARIUS, A. S. 2009. Modeling cellulase kinetics on lignocellulosic substrates. *Biotechnology Advances*, *27*, 833-848.
- BARBIER, E. B., HACKER, S. D., KENNEDY, C., KOCH, E. W., STIER, A. C. & SILLIMAN, B. R. 2011. The value of estuarine and coastal ecosystem services. *Ecological Monographs*, *81*, 169-193.
- BAUER, M., KUBE, M., TEELING, H., RICHTER, M., LOMBARDOT, T., ALLERS, E., WURDEMANN, C. A., QUAIST, C., KUHL, H., KNAUST, F., WOEBKEN, D., BISCHOF, K., MUSSMANN, M., CHOUDHURI, J. V., MEYER, F., REINHARDT, R., AMANN, R. I. & GLOCKNER, F. O. 2006. Whole genome analysis of the marine Bacteroidetes 'Gramella forsetii' reveals adaptations to degradation of polymeric organic matter. *Environmental Microbiology*, *8*, 2201-2213.
- BAUMANN, H., WALLACE, R. B., TAGLIAFERRI, T. & GOBLER, C. J. 2015. Large Natural pH, CO₂ and O₂ Fluctuations in a Temperate Tidal Salt Marsh on Diel, Seasonal, and Interannual Time Scales. *Estuaries and Coasts*, *38*, 220-231.
- BEAZLEY, M. J., MARTINEZ, R. J., RAJAN, S., POWELL, J., PICENO, Y. M., TOM, L. M., ANDERSEN, G. L., HAZEN, T. C., VAN NOSTRAND, J. D., ZHOU, J. Z., MORTAZAVI, B. & SOBECKY, P. A. 2012. Microbial Community Analysis of a Coastal Salt Marsh Affected by the Deepwater Horizon Oil Spill. *Plos One*, *7*, 13.

References

- BENGTSSON-PALME, J., RYBERG, M., HARTMANN, M., BRANCO, S., WANG, Z., GODHE, A., DE WIT, P., SANCHEZ-GARCIA, M., EBERSBERGER, I., DE SOUSA, F., AMEND, A. S., JUMPPONEN, A., UNTERSEHER, M., KRISTIANSSON, E., ABARENKOV, K., BERTRAND, Y. J. K., SANLI, K., ERIKSSON, K. M., VIK, U., VELDRE, V. & NILSSON, R. H. 2013. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology and Evolution*, 4, 914-919.
- BENNER, R., FOGEL, M. L., SPRAGUE, E. K. & HODSON, R. E. 1987. Depletion of c-13 in lignin and its implications for stable carbon isotope studies. *Nature*, 329, 708-710.
- BENNER, R., MACCUBBIN, A. E. & HODSON, R. E. 1984a. Anaerobic biodegradation of the lignin and polysaccharide components of lignocellulose and synthetic lignin by sediment microflora. *Applied and Environmental Microbiology*, 47, 998-1004.
- BENNER, R., NEWELL, S. Y., MACCUBBIN, A. E. & HODSON, R. E. 1984b. Relative contributions of bacteria and fungi to rates of degradation of lignocellulosic detritus in salt-marsh sediments. *Applied and Environmental Microbiology*, 48, 36-40.
- BERTINI, L., BREGLIA, R., LAMBRUGHI, M., FANTUCCI, P., DE GIOIA, L., BORSARI, M., SOLA, M., BORTOLOTTI, C. A. & BRUSCHI, M. 2018. Catalytic Mechanism of Fungal Lytic Polysaccharide Monooxygenases Investigated by First-Principles Calculations. *Inorganic Chemistry*, 57, 86-97.
- BIANCHI, T. S. 2011. The role of terrestrially derived organic carbon in the coastal ocean: A changing paradigm and the priming effect. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 19473-19481.
- BILAL, M., ASGHER, M., IQBAL, H. M. N., HU, H. B. & ZHANG, X. H. 2017. Biotransformation of lignocellulosic materials into value-added products A review. *International Journal of Biological Macromolecules*, 98, 447-458.
- BILLARD, H., FARAJ, A., LOPES FERREIRA, N., MENIR, S. & HEISS-BLANQUET, S. 2012. Optimization of a synthetic mixture composed of major *Trichoderma reesei* enzymes for the hydrolysis of steam-exploded wheat straw. *Biotechnology for Biofuels*, 5, 13.
- BLAALID, R., KUMAR, S., NILSSON, R. H., ABARENKOV, K., KIRK, P. M. & KAUSERUD, H. 2013. ITS1 versus ITS2 as DNA metabarcodes for fungi. *Molecular Ecology Resources*, 13, 218-224.
- BLANCO-CANQUI, H. & LAL, R. 2009. Crop Residue Removal Impacts on Soil Productivity and Environmental Quality. *Critical Reviews in Plant Sciences*, 28, 139-163.
- BOER, H., TEERI, T. T. & KOIVULA, A. 2000. Characterization of *Trichoderma reesei* cellobiohydrolase Cel7A secreted from *Pichia pastoris* using two different promoters. *Biotechnology and Bioengineering*, 69, 486-494.
- BOERJAN, W., RALPH, J. & BAUCHER, M. 2003. Lignin biosynthesis. *Annual Review of Plant Biology*, 54, 519-546.
- BOLHUIS, H. & STAL, L. J. 2011. Analysis of bacterial and archaeal diversity in coastal microbial mats using massive parallel 16S rRNA gene tag sequencing. *Isme Journal*, 5, 1701-1712.
- BONUGLI-SANTOS, R. C., VASCONCELOS, M. R. D., PASSARINI, M. R. Z., VIEIRA, G. A. L., LOPES, V. C. P., MAINARDI, P. H., DOS SANTOS, J. A., DUARTE, L. D., OTERO, I. V. R., YOSHIDA, A. M. D., FEITOSA, V. A., PESSOA, A. & SETTE, L. D. 2015. Marine-derived fungi: diversity of enzymes and biotechnological applications. *Frontiers in Microbiology*, 6.
- BOORMAN, L. A. & ASHTON, C. 1997. The productivity of salt marsh vegetation at Tollesbury, Essex, and Stiffkey, Norfolk, England. *Mangroves and Salt Marshes*, 1, 113-126.
- BORASTON, A. B., BOLAM, D. N., GILBERT, H. J. & DAVIES, G. J. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemical Journal*, 382, 769-781.
- BOUCHARD, V., CREACH, V., LEFEUVRE, J. C., BERTRU, G. & MARIOTTI, A. 1998. Fate of plant detritus in a European salt marsh dominated by *Atriplex portulacoides* (L.) Aellen. *Hydrobiologia*, 374, 75-87.

References

- BOUCHARD, V., GILLON, D., JOFFRE, R. & LEFEUVRE, J. C. 2003. Actual litter decomposition rates in salt marshes measured using near-infrared reflectance spectroscopy. *Journal of Experimental Marine Biology and Ecology*, 290, 149-163.
- BOUCHARD, V. & LEFEUVRE, J. C. 2000. Primary production and macro-detritus dynamics in a European salt marsh: carbon and nitrogen budgets. *Aquatic Botany*, 67, 23-42.
- BOWEN, J. L., CRUMP, B. C., DEEGAN, L. A. & HOBBI, J. E. 2009. Salt marsh sediment bacteria: their distribution and response to external nutrient inputs. *Isme Journal*, 3, 924-934.
- BOWEN, J. L., MORRISON, H. G., HOBBI, J. E. & SOGIN, M. L. 2012. Salt marsh sediment diversity: a test of the variability of the rare biosphere among environmental replicates. *Isme Journal*, 6, 2014-2023.
- BRIJWANI, K., OBEROI, H. S. & VADLANI, P. V. 2010. Production of a cellulolytic enzyme system in mixed-culture solid-state fermentation of soybean hulls supplemented with wheat bran. *Process Biochemistry*, 45, 120-128.
- BROWN, A. J. 1886. XLIII.—On an acetic ferment which forms cellulose. *Journal of the Chemical Society, Transactions*, 49, 432-439.
- BUANAFINA, M. M. D. 2009. Feruloylation in Grasses: Current and Future Perspectives. *Molecular Plant*, 2, 861-872.
- BUCHAN, A., NEWELL, S. Y., BUTLER, M., BIERS, E. J., HOLLIBAUGH, J. T. & MORAN, M. A. 2003. Dynamics of bacterial and fungal communities on decaying salt marsh grass. *Applied and Environmental Microbiology*, 69, 6676-6687.
- BURDIGE, D. J. 2005. Burial of terrestrial organic matter in marine sediments: A re-assessment. *Global Biogeochemical Cycles*, 19, 7.
- BURDIGE, D. J. 2007. Preservation of organic matter in marine sediments: Controls, mechanisms, and an imbalance in sediment organic carbon budgets? *Chemical Reviews*, 107, 467-485.
- BUSSE-WICHER, M., GOMES, T. C. F., TRYFONA, T., NIKOLOVSKI, N., STOTT, K., GRANTHAM, N. J., BOLAM, D. N., SKAF, M. S. & DUPREE, P. 2014. The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of *Arabidopsis thaliana*. *Plant Journal*, 79, 492-506.
- BUTH, G. J. C. & DEWOLF, L. 1985. Decomposition of *spartina-anglica*, *elytrigia-pungens* and *halimione-portulacoides* in a dutch salt-marsh in association with faunal and habitat influences. *Vegetatio*, 62, 337-355.
- BYRNE, R. H., KUMP, L. R. & CANTRELL, K. J. 1988. THE INFLUENCE OF TEMPERATURE AND PH ON TRACE-METAL SPECIATION IN SEAWATER. *Marine Chemistry*, 25, 163-181.
- BYUN, D. S. & WANG, X. H. 2005. The effect of sediment stratification on tidal dynamics and sediment transport patterns. *Journal of Geophysical Research-Oceans*, 110, 16.
- CAI, Z. J., HOU, C. W. & YANG, G. 2012. Characteristics and bending performance of electroactive polymer blend made with cellulose and poly(3-hydroxybutyrate). *Carbohydrate Polymers*, 87, 650-657.
- CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. & MADDEN, T. L. 2009. BLAST plus : architecture and applications. *Bmc Bioinformatics*, 10, 9.
- CAMPBELL, B. J. & KIRCHMAN, D. L. 2013. Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. *Isme Journal*, 7, 210-220.
- CANTAREL, B. L., COUTINHO, P. M., RANCUREL, C., BERNARD, T., LOMBARD, V. & HENRISSAT, B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, 37, D233-D238.
- CAPORASO, J. G., KUCZYNSKI, J., STOMBAUGH, J., BITTINGER, K., BUSHMAN, F. D., COSTELLO, E. K., FIERER, N., PENA, A. G., GOODRICH, J. K., GORDON, J. I., HUTTLEY, G. A., KELLEY, S. T., KNIGHTS, D., KOENIG, J. E., LEY, R. E., LOZUPONE, C. A., MCDONALD, D., MUEGGE, B. D., PIRRUNG, M., REEDER, J., SEVINSKY, J. R., TUMBAUGH, P. J., WALTERS, W. A., WIDMANN, J., YATSUNENKO, T., ZANEVELD, J. & KNIGHT, R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7, 335-336.

References

- CASILLO, A., LANZETTA, R., PARRILLI, M. & CORSARO, M. M. 2018. Exopolysaccharides from Marine and Marine Extremophilic Bacteria: Structures, Properties, Ecological Roles and Applications. *Marine Drugs*, 16, 34.
- CERQUEIRA, D. A., RODRIGUES, G. & MEIRELES, C. D. 2007. Optimization of sugarcane bagasse cellulose acetylation. *Carbohydrate Polymers*, 69, 579-582.
- CHAUDHARY, D. R., KIM, J. & KANG, H. 2018. Influences of Different Halophyte Vegetation on Soil Microbial Community at Temperate Salt Marsh. *Microbial Ecology*, 75, 729-738.
- CHEN, C., KHALEEL, S. S., HUANG, H. & WU, C. H. 2014. Software for pre-processing Illumina next-generation sequencing short read sequences. *Source code for biology and medicine*, 9, 8.
- CHEN, H. Y., LIU, J. B., CHANG, X., CHEN, D. M., XUE, Y., LIU, P., LIN, H. L. & HAN, S. 2017a. A review on the pretreatment of lignocellulose for high-value chemicals. *Fuel Processing Technology*, 160, 196-206.
- CHEN, L., GAO, J., ZHU, Q., WANG, Y. & YANG, Y. 2017b. Accumulation and output of heavy metals in *Spartina alterniflora* in a salt marsh. *Pedosphere*, In press.
- CHEN, X. W., KUHN, E., JENNINGS, E. W., NELSON, R., TAO, L., ZHANG, M. & TUCKER, M. P. 2016. DMR (deacetylation and mechanical refining) processing of corn stover achieves high monomeric sugar concentrations (230 g L⁻¹) during enzymatic hydrolysis and high ethanol concentrations (> 10% v/v) during fermentation without hydrolysate purification or concentration. *Energy & Environmental Science*, 9, 1237-1245.
- CHERRY, J. R. & FIDANTSEF, A. L. 2003. Directed evolution of industrial enzymes: an update. *Current Opinion in Biotechnology*, 14, 438-443.
- CHERUBIN, M. R., OLIVEIRA, D. M. D., FEIGL, B. J., PIMENTEL, L. G., LISBOA, I. P., GMACH, M. R., VARANDA, L. L., MORAIS, M. C., SATIRO, L. S., POPIN, G. V., DE PAIVA, S. R., DOS SANTOS, A. K. B., DE VASCONCELOS, A. L. S., DE MELO, P. L. A., CERRI, C. E. P. & CERRI, C. C. 2018. Crop residue harvest for bioenergy production and its implications on soil functioning and plant growth: A review. *Scientia Agricola*, 75, 255-272.
- CHMURA, G. L. 2013. What do we need to assess the sustainability of the tidal salt marsh carbon sink? *Ocean & Coastal Management*, 83, 25-31.
- CHO, J. C. & GIOVANNONI, S. J. 2004. Cultivation and growth characteristics of a diverse group of oligotrophic marine Gammaproteobacteria. *Applied and Environmental Microbiology*, 70, 432-440.
- CHRISTOPHER, L. P., YAO, B. & JI, Y. 2014. Lignin biodegradation with laccase-mediator systems. *Frontiers in Energy Research*, 2, 12.
- CHURCH, J. A., CLARK, P. U., CAZENAVE, A., GREGORY, J. M., JEVREJEVA, S., LEVERMANN, A., MERRIFIELD, M. A., MILNE, G. A., NEREM, R. S., NUNN, P. D., PAYNE, A. J., PFEFFER, W. T., STAMMER, D. & UNNIKISHNAN, A. S. 2013. Sea-Level Rise by 2100. *Science*, 342, 1445-1445.
- CLEARY, D. F. R., COELHO, F., OLIVEIRA, V., GOMES, N. C. M. & POLONIA, A. R. M. 2017. Sediment depth and habitat as predictors of the diversity and composition of sediment bacterial communities in an inter-tidal estuarine environment. *Marine Ecology-an Evolutionary Perspective*, 38, 15.
- COCINERO, E. J., GAMBLIN, D. P., DAVIS, B. G. & SIMONS, J. P. 2009. The Building Blocks of Cellulose: The Intrinsic Conformational Structures of Cellobiose, Its Epimer, Lactose, and Their Singly Hydrated Complexes. *Journal of the American Chemical Society*, 131, 11117-11123.
- COCK, P. J. A., ANTAO, T., CHANG, J. T., CHAPMAN, B. A., COX, C. J., DALKE, A., FRIEDBERG, I., HAMELRYCK, T., KAUFF, F., WILCZYNSKI, B. & DE HOON, M. J. L. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25, 1422-1423.
- COLEMAN, M. L., HEDRICK, D. B., LOVLEY, D. R., WHITE, D. C. & PYE, K. 1993. REDUCTION OF FE(III) IN SEDIMENTS BY SULFATE-REDUCING BACTERIA. *Nature*, 361, 436-438.

References

- COLUSSI, F., SORESENSEN, T. H., ALASEPP, K., KARI, J., CRUYS-BAGGER, N., WINDAHL, M. S., OLSEN, J. P., BORCH, K. & WESTH, P. 2015. Probing Substrate Interactions in the Active Tunnel of a Catalytically Deficient Cellobiohydrolase (Cel7). *Journal of Biological Chemistry*, 290, 2444-2454.
- COPPOLA, F., SIMONCINI, E. & PULSELLI, R. M. 2009. Bioethanol potentials from marine residual biomass: an Energy evaluation. In: BREBBIA, C. A. & TIEZZI, E. (eds.) *Ecosystems and Sustainable Development VII*.
- CORTES-TOLALPA, L., NORDER, J., VAN ELSAS, J. D. & SALLES, J. F. 2018. Halotolerant microbial consortia able to degrade highly recalcitrant plant biomass substrate. *Applied Microbiology and Biotechnology*, 102, 2913-2927.
- COSGROVE, D. J. 2005. Growth of the plant cell wall. *Nature Reviews Molecular Cell Biology*, 6, 850-861.
- COSGROVE, D. J. & JARVIS, M. C. 2012. Comparative structure and biomechanics of plant primary and secondary cell walls. *Frontiers in Plant Science*, 3, 6.
- COSTANZA, R., DE GROOT, R., SUTTON, P., VAN DER PLOEG, S., ANDERSON, S. J., KUBISZEWSKI, I., FARBER, S. & TURNER, R. K. 2014. Changes in the global value of ecosystem services. *Global Environmental Change-Human and Policy Dimensions*, 26, 152-158.
- COURTADE, G., WIMMER, R., ROHR, A. K., PREIMS, M., FELICE, A. K. G., DIMAROGONA, M., VAAJEKOLSTAD, G., SORLIE, M., SANDGREN, M., LUDWIG, R., EIJSINK, V. G. H. & AACHMANN, F. L. 2016. Interactions of a fungal lytic polysaccharide monooxygenase with beta-glucan substrates and cellobiose dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 5922-5927.
- COUTINHO, P. M., DELEURY, E., DAVIES, G. J. & HENRISSAT, B. 2003. An evolving hierarchical family classification for glycosyltransferases. *Journal of Molecular Biology*, 328, 307-317.
- COUTURIER, M., LADEVEZE, S., SULZENBACHER, G., CIANO, L., FANUEL, M., MOREAU, C., VILLARES, A., CATHALA, B., CHASPOUL, F., FRANSEN, K. E., LABOUREL, A., HERPOEL-GIMBERT, I., GRISEL, S., HAON, M., LENFANT, N., ROGNIAUX, H., ROPARTZ, D., DAVIES, G. J., ROSSO, M. N., WALTON, P. H., HENRISSAT, B. & BERRIN, J. G. 2018. Lytic xylan oxidases from wood-decay fungi unlock biomass degradation. *Nature Chemical Biology*, 14, 306-+.
- CRAGG, S. M., BECKHAM, G. T., BRUCE, N. C., BUGG, T. D. H., DISTEL, D. L., DUPREE, P., ETXABE, A. G., GOODELL, B. S., JELLISON, J., MCGEEHAN, J. E., MCQUEEN-MASON, S. J., SCHNORR, K., WALTON, P. H., WATTS, J. E. M. & ZIMMER, M. 2015. Lignocellulose degradation mechanisms across the Tree of Life. *Current Opinion in Chemical Biology*, 29, 108-119.
- CRIMINGER, J. D., HAZEN, T. H., SOBECKY, P. A. & LOVELL, C. R. 2007. Nitrogen fixation by *Vibrio parahaemolyticus* and its implications for a new ecological niche. *Applied and Environmental Microbiology*, 73, 5959-5961.
- CROUCH, L. I., LABOUREL, A., WALTON, P. H., DAVIES, G. J. & GILBERT, H. J. 2016. The Contribution of Non-catalytic Carbohydrate Binding Modules to the Activity of Lytic Polysaccharide Monooxygenases. *Journal of Biological Chemistry*, 291, 7439-7449.
- CUI, J., CHEN, X. P., NIE, M., FANG, S. B., TANG, B. P., QUAN, Z. X., LI, B. & FANG, C. M. 2017. Effects of *Spartina alterniflora* Invasion on the Abundance, Diversity, and Community Structure of Sulfate Reducing Bacteria along a Successional Gradient of Coastal Salt Marshes in China. *Wetlands*, 37, 221-232.
- CURCO, A., IBANEZ, C., DAY, J. W. & PRAT, N. 2002. Net primary production and decomposition of salt marshes of the Ebre delta (Catalonia, Spain). *Estuaries*, 25, 309-324.
- DAI, T. & WIEGERT, R. G. 1996. Ramet population dynamics and net aerial primary productivity of *Spartina alterniflora*. *Ecology*, 77, 276-288.
- DARJANY, L. E., WHITCRAFT, C. R. & DILLON, J. G. 2014. Lignocellulose-responsive bacteria in a southern California salt marsh identified by stable isotope probing. *Frontiers in Microbiology*, 5, 9.

References

- DAVIDE, D. M. & DAVID, T. 2014. Lignocellulosic feedstock in the UK. *In*: NNFCC (ed.). Lignocellulosic Biorefinery Network.
- DAVIS, S. C., DIEGEL, S. W. & BOUNDY, R. G. 2011. Transportation Energy Data Book, 30th Edn. Oak Ridge National Laboratory. URL: <http://info.ornl.gov/sites/publications/files/Pub31202.pdf>.
- DAVIS, S. C., WILLIAMS, S. E. & BOUNDY, R. G. 2018. Transportation Energy Data Book, 36th Edn. Oak Ridge National Laboratory. URL: https://cta.ornl.gov/data/tedbfiles/Edition36_Full_Doc.pdf.
- DE BOER, W., FOLMAN, L. B., SUMMERBELL, R. C. & BODDY, L. 2005. Living in a fungal world: impact of fungi on soil bacterial niche development. *Fems Microbiology Reviews*, 29, 795-811.
- DE GROOT, R., BRANDER, L., VAN DER PLOEG, S., COSTANZA, R., BERNARD, F., BRAAT, L., CHRISTIE, M., CROSSMAN, N., GHERMANDI, A., HEIN, L., HUSSAIN, S., KUMAR, P., MCVITTIE, A., PORTELA, R., RODRIGUEZ, L. C., TEN BRINK, P. & VAN BEUKERINGH, P. 2012. Global estimates of the value of ecosystems and their services in monetary units. *Ecosystem Services*, 1, 50-61.
- DE MARCO, A. 2009. Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microbial Cell Factories*, 8, 18.
- DE OLIVEIRA, D. M., FINGER-TEIXEIRA, A., MOTA, T. R., SALVADOR, V. H., MOREIRA-VILAR, F. C., MOLINARI, H. B. C., MITCHELL, R. A. C., MARCHIOSI, R., FERRARESE, O. & DOS SANTOS, W. D. 2015. Ferulic acid: a key component in grass lignocellulose recalcitrance to hydrolysis. *Plant Biotechnology Journal*, 13, 1224-1232.
- DE SOUZA, R. S. C., OKURA, V. K., ARMANHI, J. S. L., JORRIN, B., LOZANO, N., DA SILVA, M. J., GONZALEZ-GUERRERO, M., DE ARAUJO, L. M., VERZA, N. C., BAGHERI, H. C., IMPERIAL, J. & ARRUDA, P. 2016. Unlocking the bacterial and fungal communities assemblages of sugarcane microbiome. *Scientific Reports*, 6, 15.
- DELGADO-GARCIA, M., VALDIVIA-URDIALES, B., AGUILAR-GONZALEZ, C. N., CONTRERAS-ESQUIVEL, J. C. & RODRIGUEZ-HERRERA, R. 2012. Halophilic hydrolases as a new tool for the biotechnological industries. *Journal of the Science of Food and Agriculture*, 92, 2575-2580.
- DIN, N., GILKES, N. R., TEKANT, B., MILLER, R. C., WARREN, A. J. & KILBURN, D. G. 1991. NON-HYDROLYTIC DISRUPTION OF CELLULOSE FIBERS BY THE BINDING DOMAIN OF A BACTERIAL CELLULASE. *Bio-Technology*, 9, 1096-1099.
- DING, H. T., ZENG, Q., ZHOU, L. L., YU, Y. & CHEN, B. 2017. Biochemical and Structural Insights into a Novel Thermostable beta-1,3-Galactosidase from *Marinomonas* sp BSi20414. *Marine Drugs*, 15, 15.
- DINI-ANDREOTE, F., BROSSI, M. J. D., VAN ELSAS, J. D. & SALLES, J. F. 2016a. Reconstructing the Genetic Potential of the Microbially-Mediated Nitrogen Cycle in a Salt Marsh Ecosystem. *Frontiers in Microbiology*, 7, 13.
- DINI-ANDREOTE, F., PYLRO, V. S., BALDRIAN, P., VAN ELSAS, J. D. & SALLES, J. F. 2016b. Ecological succession reveals potential signatures of marine-terrestrial transition in salt marsh fungal communities. *Isme Journal*, 10, 1984-1997.
- DINI-ANDREOTE, F., SILVA, M. D. C. P. E., TRIADO-MARGARIT, X., CASAMAYOR, E. O., VAN ELSAS, J. D. & SALLES, J. F. 2014. Dynamics of bacterial community succession in a salt marsh chronosequence: evidences for temporal niche partitioning. *Isme Journal*, 8, 1989-2001.
- DOWLE, A. A., WILSON, J. & THOMAS, J. R. 2016. Comparing the Diagnostic Classification Accuracy of iTRAQ, Peak Area, Spectral-Counting, and emPAI Methods for Relative Quantification in Expression Proteomics. *Journal of Proteome Research*, 15, 3550-3562.
- DRAKE, K., HALIFAX, H., ADAMOWICZ, S. C. & CRAFT, C. 2015. Carbon Sequestration in Tidal Salt Marshes of the Northeast United States. *Environmental Management*, 56, 998-1008.
- DRUZHININA, I. S. & KUBICEK, C. P. 2017. Genetic engineering of *Trichoderma reesei* cellulases and their production. *Microbial Biotechnology*, 10, 1485-1499.
- DU, Z. J., LV, G. Q., ROONEY, A. P., MIAO, T. T., XU, Q. Q. & CHEN, G. J. 2011. *Agarivorans gilvus* sp. nov. isolated from seaweed. *International Journal of Systematic and Evolutionary Microbiology*, 61, 493-496.

References

- DUAN, C. J., FENG, Y. L., CAO, Q. L., HUANG, M. Y. & FENG, J. X. 2016. Identification of a novel family of carbohydrate-binding modules with broad ligand specificity. *Scientific Reports*, 6, 8.
- DUTTA, A. & PHILLIPS, S. D. 2009. Thermochemical ethanol via direct gasification and mixed alcohol synthesis of lignocellulosic biomass. National Renewable Energy Lab.(NREL), Golden, CO (United States), NREL/TP-510-45913.
- DYKSMA, S., BISCHOF, K., FUCHS, B. M., HOFFMANN, K., MEIER, D., MEYERDIERKS, A., PJEVAC, P., PROBANDT, D., RICHTER, M., STEPANAUSKAS, R. & MUSSMANN, M. 2016. Ubiquitous Gammaproteobacteria dominate dark carbon fixation in coastal sediments. *Isme Journal*, 10, 1939-1953.
- EDGAR, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, 2460-2461.
- EDGAR, R. C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10, 996-+.
- ELDERFIELD, H., HOLLAND, H. D. & TUREKIAN, K. K. 2006. *The oceans and marine geochemistry, 1st Edition, Vol. 6, pp. 23-47*, Elsevier.
- ELIAS, J. E., HAAS, W., FAHERTY, B. K. & GYGI, S. P. 2005. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nature Methods*, 2, 667-675.
- ELLISON, A. M., BERTNESS, M. D. & MILLER, T. 1986. Seasonal patterns in the belowground biomass of spartina-alterniflora (gramineae) across a tidal gradient. *American Journal of Botany*, 73, 1548-1554.
- ELMER, W. H. 2016. Pathogenic Microfungi Associated with Spartina in Salt Marshes. In: LI, D.-W. (ed.) *Biology of Microfungi*. Cham: Springer International Publishing.
- ELMER, W. H. & MARRA, R. E. 2011. New species of Fusarium associated with dieback of Spartina alterniflora in Atlantic salt marshes. *Mycologia*, 103, 806-819.
- EMERY, N. C., EWANCHUK, P. J. & BERTNESS, M. D. 2001. Competition and salt-marsh plant zonation: Stress tolerators may be dominant competitors. *Ecology*, 82, 2471-2485.
- ENACHE, M. & KAMEKURA, M. 2010. Hydrolytic enzymes of halophilic microorganisms and their economic values. *Romanian Journal of Biochemistry*, 47, 47-59.
- ENGEL, A. S., LIU, C., PATERSON, A. T., ANDERSON, L. C., TURNER, R. E. & OVERTON, E. B. 2017. Salt Marsh Bacterial Communities before and after the Deepwater Horizon Oil Spill. *Applied and Environmental Microbiology*, 83, 22.
- ERDEI, B., BARTA, Z., SIPOS, B., RECZEY, K., GALBE, M. & ZACCHI, G. 2010. Ethanol production from mixtures of wheat straw and wheat meal. *Biotechnology for Biofuels*, 3, 9.
- FANG, C. J., THOMSEN, M. H., BRUDECKI, G. P., CYBULSKA, I., FRANKAER, C. G., BASTIDAS-OYANEDEL, J. R. & SCHMIDT, J. E. 2015a. Seawater as Alternative to Freshwater in Pretreatment of Date Palm Residues for Bioethanol Production in Coastal and/or Arid Areas. *Chemsuschem*, 8, 3823-3831.
- FANG, Z. M., LIU, X. M., CHEN, L. Y., SHEN, Y., ZHANG, X. C., FANG, W., WANG, X. T., BAO, X. M. & XIAO, Y. Z. 2015b. Identification of a laccase Glac15 from Ganoderma lucidum 77002 and its application in bioethanol production. *Biotechnology for Biofuels*, 8, 12.
- FERREIRA, L. M. A., HAZLEWOOD, G. P., BARKER, P. J. & GILBERT, H. J. 1991. The cellodextrinase from pseudomonas-fluorescens subsp cellulosa consists of multiple functional domains. *Biochemical Journal*, 279, 793-799.
- FIDALGO, C., ROCHA, J., PROENCA, D. N., MORAIS, P. V., ALVES, A. & HENRIQUES, I. 2017. Saccharospirillum correaiae sp nov., an endophytic bacterium isolated from the halophyte Halimione portulacoides. *International Journal of Systematic and Evolutionary Microbiology*, 67, 2026-2030.
- FINGERMAN, K. R., TORN, M. S., O'HARE, M. H. & KAMMEN, D. M. 2010. Accounting for the water impacts of ethanol production. *Environmental Research Letters*, 5, 7.

References

- FINN, R. D., COGGILL, P., EBERHARDT, R. Y., EDDY, S. R., MISTRY, J., MITCHELL, A. L., POTTER, S. C., PUNTA, M., QURESHI, M., SANGRADOR-VEGAS, A., SALAZAR, G. A., TATE, J. & BATEMAN, A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, 44, D279-D285.
- FLOWERS, T. J., MUNNS, R. & COLMER, T. D. 2015. Sodium chloride toxicity and the cellular basis of salt tolerance in halophytes. *Annals of Botany*, 115, 419-431.
- FOGEL, M. L., SPRAGUE, E. K., GIZE, A. P. & FREY, R. W. 1989. DIAGENESIS OF ORGANIC-MATTER IN GEORGIA SALT MARSHES. *Estuarine Coastal and Shelf Science*, 28, 211-230.
- FOSTER, C. E., MARTIN, T. M. & PAULY, M. 2010. Comprehensive compositional analysis of plant cell walls (lignocellulosic biomass) part I: lignin. *Journal of visualized experiments: JoVE*, 37, 1745.
- FRANSDEN, K. E. H., SIMMONS, T. J., DUPREE, P., POULSEN, J. C. N., HEMSWORTH, G. R., CIANO, L., JOHNSTON, E. M., TOVBORG, M., JOHANSEN, K. S., VON FREIESLEBEN, P., MARMUSE, L., FORT, S., COTTAZ, S., DRIGUEZ, H., HENRISSAT, B., LENFANT, N., TUNA, F., BALDANSUREN, A., DAVIES, G. J., LO LEGGIO, L. & WALTON, P. H. 2016. The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases. *Nature Chemical Biology*, 12, 298-+.
- FROMMHAGEN, M., MUTTE, S. K., WESTPHAL, A. H., KOETSIER, M. J., HINZ, S. W. A., VISSER, J., VINCKEN, J. P., WEIJERS, D., VAN BERKEL, W. J. H., GRUPPEN, H. & KABEL, M. A. 2017. Boosting LPMO-driven lignocellulose degradation by polyphenol oxidase-activated lignin building blocks. *Biotechnology for Biofuels*, 10, 16.
- GADDE, B., BONNET, S., MENKE, C. & GARIVAIT, S. 2009. Air pollutant emissions from rice straw open field burning in India, Thailand and the Philippines. *Environmental Pollution*, 157, 1554-1558.
- GALDOS, M., CAVALETT, O., SEABRA, J. E. A., NOGUEIRA, L. A. H. & BONOMI, A. 2013. Trends in global warming and human health impacts related to Brazilian sugarcane ethanol production considering black carbon emissions. *Applied Energy*, 104, 576-582.
- GERBENS-LEENES, W., HOEKSTRA, A. Y. & VAN DER MEER, T. H. 2009. The water footprint of bioenergy. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 10219-10223.
- GHRSSST 2010. JPL MUR MEaSURES Project. GHRSSST Level 4 MUR Global Foundation Sea Surface Temperature Analysis. Data accessed [23/08/2018] at <https://podaac.jpl.nasa.gov/dataset/MUR-JPL-L4-GLOB-v4.1>.
- GIACOBBE, S., PEZZELLA, C., LETTERA, V., SANNIA, G. & PISCITELLI, A. 2018. Laccase pretreatment for agrofood wastes valorization. *Bioresource Technology*, 265, 59-65.
- GIBSON, A., GROCHOWSKI, J., MALEK, L. & DEKKER, R. 2013. The effects of Changing Growth Conditions on the Demethylation of a Lignin-Related Compound by *Absidiocyindrospora*, *Fomesfomentarius*, *Ischnodermaresinosum* and *Cylindrocladiumcamelliae*. *Research Journal of Biotechnology Vol*, 8, 12.
- GILBERT, H. J. 2010. The Biochemistry and Structural Biology of Plant Cell Wall Deconstruction. *Plant Physiology*, 153, 444-455.
- GILLE, S. & PAULY, M. 2012. O-acetylation of plant cell wall polysaccharides. *Frontiers in Plant Science*, 3, 7.
- GOMES, D., GAMA, M. & DOMINGUES, L. 2018. Determinants on an efficient cellulase recycling process for the production of bioethanol from recycled paper sludge under high solid loadings. *Biotechnology for Biofuels*, 11, 12.
- GOMEZ, L. D., STEELE-KING, C. G. & MCQUEEN-MASON, S. J. 2008. Sustainable liquid biofuels from biomass: the writing's on the walls. *New Phytologist*, 178, 473-485.
- GONTIKAKI, E., THORNTON, B., CORNULIER, T. & WITTE, U. 2015. Occurrence of Priming in the Degradation of Lignocellulose in Marine Sediments. *Plos One*, 10, 13.
- GORDON, D. C., CRANFORD, P. J. & DESPLANQUE, C. 1985. OBSERVATIONS ON THE ECOLOGICAL IMPORTANCE OF SALT MARSHES IN THE CUMBERLAND BASIN, A MACROTIDAL ESTUARY IN THE BAY OF FUNDY. *Estuarine Coastal and Shelf Science*, 20, 205-227.

References

- GOURLAY, K., ARANTES, V. & SADDLER, J. N. 2012. Use of substructure-specific carbohydrate binding modules to track changes in cellulose accessibility and surface morphology during the amorphogenesis step of enzymatic hydrolysis. *Biotechnology for Biofuels*, 5, 14.
- GRABBER, J. H., HATFIELD, R. D. & RALPH, J. 1998. Diferulate cross-links impede the enzymatic degradation of non-lignified maize walls. *Journal of the Science of Food and Agriculture*, 77, 193-200.
- GRABHERR, M. G., HAAS, B. J., YASSOUR, M., LEVIN, J. Z., THOMPSON, D. A., AMIT, I., ADICONIS, X., FAN, L., RAYCHOWDHURY, R., ZENG, Q. D., CHEN, Z. H., MAUCELLI, E., HACOEN, N., GNIRKE, A., RHIND, N., DI PALMA, F., BIRREN, B. W., NUSBAUM, C., LINDBLAD-TOH, K., FRIEDMAN, N. & REGEV, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29, 644-U130.
- GRANDE, P. M. & DE MARIA, P. D. 2012. Enzymatic hydrolysis of microcrystalline cellulose in concentrated seawater. *Bioresource Technology*, 104, 799-802.
- GUILLEN, D., SANCHEZ, S. & RODRIGUEZ-SANOJA, R. 2010. Carbohydrate-binding domains: multiplicity of biological roles. *Applied Microbiology and Biotechnology*, 85, 1241-1249.
- GUNNE, M., HOPNER, A., HAGEDOORN, P. L. & URLACHER, V. B. 2014. Structural and redox properties of the small laccase Ssl1 from *Streptomyces sviveus*. *Febs Journal*, 281, 4307-4318.
- GUO, B., LI, P. Y., YUE, Y. S., ZHAO, H. L., DONG, S., SONG, X. Y., SUN, C. Y., ZHANG, W. X., CHEN, X. L., ZHANG, X. Y., ZHOU, B. C. & ZHANG, Y. Z. 2013. Gene Cloning, Expression and Characterization of a Novel Xylanase from the Marine Bacterium, *Glaciecola mesophila* KMM241. *Marine Drugs*, 11, 1173-1187.
- GUO, H. L., WANG, X. D. & LEE, D. J. 2018. Proteomic researches for lignocellulose-degrading enzymes: A mini-review. *Bioresource Technology*, 265, 532-541.
- HADZIAVDIC, K., LEKANG, K., LANZEN, A., JONASSEN, I., THOMPSON, E. M. & TROEDSSON, C. 2014. Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers. *Plos One*, 9, 10.
- HALLAC, B. B. & RAGAUSKAS, A. J. 2011. Analyzing cellulose degree of polymerization and its relevancy to cellulosic ethanol. *Biofuels Bioproducts & Biorefining-Biofpr*, 5, 215-225.
- HARRIS, P. V., WELNER, D., MCFARLAND, K. C., RE, E., POULSEN, J. C. N., BROWN, K., SALBO, R., DING, H. S., VLASENKO, E., MERINO, S., XU, F., CHERRY, J., LARSEN, S. & LO LEGGIO, L. 2010. Stimulation of Lignocellulosic Biomass Hydrolysis by Proteins of Glycoside Hydrolase Family 61: Structure and Function of a Large, Enigmatic Family. *Biochemistry*, 49, 3305-3316.
- HART, T., KOMORI, H. K., LAMERE, S., PODSHIVALOVA, K. & SALOMON, D. R. 2013. Finding the active genes in deep RNA-seq gene expression studies. *Bmc Genomics*, 14, 7.
- HATAHET, F., BOYD, D. & BECKWITH, J. 2014. Disulfide bond formation in prokaryotes: History, diversity and design. *Biochimica Et Biophysica Acta-Proteins and Proteomics*, 1844, 1402-1414.
- HAVLIK, P., SCHNEIDER, U. A., SCHMID, E., BOTTCHER, H., FRITZ, S., SKALSKY, R., AOKI, K., DE CARA, S., KINDERMANN, G., KRAXNER, F., LEDUC, S., MCCALLUM, I., MOSNIER, A., SAUER, T. & OBERSTEINER, M. 2011. Global land-use implications of first and second generation biofuel targets. *Energy Policy*, 39, 5690-5702.
- HAYATSU, R., WINANS, R. E., MCBETH, R. L., SCOTT, R. G., MOORE, L. P. & STUDIER, M. H. 1979. LIGNIN-LIKE POLYMERS IN COALS. *Nature*, 278, 41-43.
- HAZEN, T. C., DUBINSKY, E. A., DESANTIS, T. Z., ANDERSEN, G. L., PICENO, Y. M., SINGH, N., JANSSON, J. K., PROBST, A., BORGLIN, S. E., FORTNEY, J. L., STRINGFELLOW, W. T., BILL, M., CONRAD, M. E., TOM, L. M., CHAVARRIA, K. L., ALUSI, T. R., LAMENDELLA, R., JOYNER, D. C., SPIER, C., BAEUM, J., AUER, M., ZEMLA, M. L., CHAKRABORTY, R., SONNENTHAL, E. L., D'HAESELEER, P., HOLMAN, H. Y. N., OSMAN, S., LU, Z. M., VAN NOSTRAND, J. D., DENG, Y., ZHOU, J. Z. & MASON, O. U. 2010. Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria. *Science*, 330, 204-208.

References

- HEDGES, J. I. & KEIL, R. G. 1995. Sedimentary organic-matter preservation - an assessment and speculative synthesis. *Marine Chemistry*, 49, 81-115.
- HEMMINGA, M. A., KLAP, V. A., VANSOELEN, J. & BOON, J. J. 1993. Effect of salt-marsh inundation on estuarine particulate organic-matter characteristics. *Marine Ecology Progress Series*, 99, 153-161.
- HEMSWORTH, G. R., DAVIES, G. J. & WALTON, P. H. 2013. Recent insights into copper-containing lytic polysaccharide mono-oxygenases. *Current Opinion in Structural Biology*, 23, 660-668.
- HENRISSAT, B. 1991. A classification of glycosyl hydrolases based on amino-acid-sequence similarities. *Biochemical Journal*, 280, 309-316.
- HENRISSAT, B. & DAVIES, G. 1997. Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology*, 7, 637-644.
- HERLET, J., KORNBURGER, P., ROESSLER, B., GLANZ, J., SCHWARZ, W. H., LIEBL, W. & ZVERLOV, V. V. 2017. A new method to evaluate temperature vs. pH activity profiles for biotechnological relevant enzymes. *Biotechnology for Biofuels*, 10, 12.
- HIMES-CORNELL, A., PENDLETON, L. & ATIYAH, P. 2018. Valuing ecosystem services from blue forests: A systematic review of the valuation of salt marshes, sea grass beds and mangrove forests. *Ecosystem Services*, 30, 36-48.
- HIRASAWA, K., UCHIMURA, K., KASHIWA, M., GRANT, W. D., ITO, S., KOBAYASHI, T. & HORIKOSHI, K. 2006. Salt-activated endoglucanase of a strain of alkaliphilic *Bacillus agaradhaerens*. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology*, 89, 211-219.
- HOFFMANN, K., HASSENBUCK, C., SALMAN-CARVALHO, V., HOLTAPPELS, M. & BIENHOLD, C. 2017. Response of Bacterial Communities to Different Detritus Compositions in Arctic Deep-Sea Sediments. *Frontiers in Microbiology*, 8, 18.
- HOGG, D., PELL, G., DUPREE, P., GOUBET, F., MARTIN-ORUE, S. M., ARMAND, S. & GILBERT, H. J. 2003. The modular architecture of *Cellvibrio japonicus* mannanases in glycoside hydrolase families 5 and 26 points to differences in their role in mannan degradation. *Biochemical Journal*, 371, 1027-1043.
- HONG, J. H., JANG, S., HEO, Y. M., MIN, M., LEE, H., LEE, Y. M. & KIM, J. J. 2015. Investigation of Marine-Derived Fungal Diversity and Their Exploitable Biological Activities. *Marine Drugs*, 13, 4137-4155.
- HOPKINSON, C. & GIBLIN, A. 2008. *Nitrogen Dynamics of Coastal Salt Marshes. Nitrogen in the Marine Environment (2nd Edition)*, Academic Press, 991-1036.
- HOPKINSON, C. S., CAI, W. J. & HU, X. P. 2012. Carbon sequestration in wetland dominated coastal systems - a global sink of rapidly diminishing magnitude. *Current Opinion in Environmental Sustainability*, 4, 186-194.
- HUANG, L., ZHANG, H., WU, P., ENTWISTLE, S., LI, X., YOHE, T., YI, H., YANG, Z. & YIN, Y. 2017. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Research*, gkx894-gkx894.
- HUBER, D., ORTNER, A., DAXBACHER, A., NYANHONGO, G. S., BAUER, W. & GUEBITZ, G. M. 2016. Influence of Oxygen and Mediators on Laccase-Catalyzed Polymerization of Lignosulfonate. *Acs Sustainable Chemistry & Engineering*, 4, 5303-5310.
- HUERTA-CEPAS, J., SERRA, F. & BORK, P. 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution*, 33, 1635-1638.
- HUGHES, R. G. & PARAMOR, O. A. L. 2004. On the loss of saltmarshes in south-east England and methods for their restoration. *Journal of Applied Ecology*, 41, 440-448.
- HUMBIRD, D., DAVIS, R., TAO, L., KINCHIN, C., HSU, D., ADEN, A., SCHOEN, P., LUKAS, J., OLTHOF, B. & WORLEY, M. 2011. Process design and economics for biochemical conversion of lignocellulosic biomass to ethanol: dilute-acid pretreatment and enzymatic hydrolysis of corn stover. National Renewable Energy Lab.(NREL), Golden, CO (United States), NREL/TP-5100-47764.

References

- HUNG, K. S., LIU, S. M., TZOU, W. S., LIN, F. P., PAN, C. L., FANG, T. Y., SUN, K. H. & TANG, S. J. 2011. Characterization of a novel GH10 thermostable, halophilic xylanase from the marine bacterium *Thermoanaerobacterium saccharolyticum* NTOU1. *Process Biochemistry*, 46, 1257-1263.
- HUNTER, J. D. 2007. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9, 90-95.
- IGARASHI, K., UCHIHASHI, T., KOIVULA, A., WADA, M., KIMURA, S., OKAMOTO, T., PENTTILA, M., ANDO, T. & SAMEJIMA, M. 2011. Traffic Jams Reduce Hydrolytic Efficiency of Cellulase on Cellulose Surface. *Science*, 333, 1279-1282.
- IHRMARK, K., BODEKER, I. T. M., CRUZ-MARTINEZ, K., FRIBERG, H., KUBARTOVA, A., SCHENCK, J., STRID, Y., STENLID, J., BRANDSTROM-DURLING, M., CLEMMENSEN, K. E. & LINDAHL, B. D. 2012. New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *Fems Microbiology Ecology*, 82, 666-677.
- ISHIHAMA, Y., ODA, Y., TABATA, T., SATO, T., NAGASU, T., RAPPSILBER, J. & MANN, M. 2005. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular & Cellular Proteomics*, 4, 1265-1272.
- ISOZAKI, K., SHIMOAKA, T., OSHIRO, S., YAMAGUCHI, A., PINCELLA, F., UENO, R., HASEGAWA, T., WATANABE, T., TAKAYA, H. & NAKAMURA, M. 2018. Robust Surface Plasmon Resonance Chips for Repetitive and Accurate Analysis of Lignin-Peptide Interactions. *Acs Omega*, 3, 7483-7493.
- JACKSON, D., LONG, S. P. & MASON, C. F. 1986. Net primary production, decomposition and export of *spartina-anglica* on a suffolk salt-marsh. *Journal of Ecology*, 74, 647-662.
- JAENICKE, R. 1991. Protein stability and molecular adaptation to extreme conditions. *European Journal of Biochemistry*, 202, 715-728.
- JAISAMUT, K., PAULOVA, L., PATAKOVA, P., KOTUCOVA, S. & RYCHTERA, M. 2016. Effect of sodium sulfite on acid pretreatment of wheat straw with respect to its final conversion to ethanol. *Biomass & Bioenergy*, 95, 1-7.
- JALAK, J., KURASIN, M., TEUGJAS, H. & VALJAMAE, P. 2012. Endo-exo Synergism in Cellulose Hydrolysis Revisited. *Journal of Biological Chemistry*, 287, 28802-28815.
- JANG, H., YANG, S. H., SEO, H. S., LEE, J. H., KIM, S. J. & KWON, K. K. 2015. *Amphritea spongicola* sp nov., isolated from a marine sponge, and emended description of the genus *Amphritea*. *International Journal of Systematic and Evolutionary Microbiology*, 65, 1866-1870.
- JENSEN, M. S., FREDRIKSEN, L., MACKENZIE, A. K., POPE, P. B., LEIROS, I., CHYLENSKI, P., WILLIAMSON, A. K., CHRISTOPEIT, T., OSTBY, H., VAAJE-KOLSTAD, G. & EIJSINK, V. G. H. 2018. Discovery and characterization of a thermostable two-domain GH6 endoglucanase from a compost metagenome. *Plos One*, 13, 22.
- JIMENEZ, D. J., BROSSI, M. J. D., SCHUCKEL, J., KRACUN, S. K., WILLATS, W. G. T. & VAN ELSAS, J. D. 2016. Characterization of three plant biomass-degrading microbial consortia by metagenomics- and metasecretomics-based approaches. *Applied Microbiology and Biotechnology*, 100, 10463-10477.
- JIMENEZ, D. J., CHAVES-MORENO, D. & VAN ELSAS, J. D. 2015a. Unveiling the metabolic potential of two soil-derived microbial consortia selected on wheat straw. *Scientific Reports*, 5, 16.
- JIMENEZ, D. J., MARUTHAMUTHU, M. & VAN ELSAS, J. D. 2015b. Metasecretome analysis of a lignocellulolytic microbial consortium grown on wheat straw, xylan and xylose. *Biotechnology for Biofuels*, 8, 15.
- JIMENEZ-JUAREZ, N., ROMAN-MIRANDA, R., BAEZA, A., SANCHEZ-AMAT, A., VAZQUEZ-DUHALT, R. & VALDERRAMA, B. 2005. Alkali and halide-resistant catalysis by the multipotent oxidase from *Marinomonas mediterranea*. *Journal of Biotechnology*, 117, 73-82.
- JOHNSON, E. 2009. Goodbye to carbon neutral: Getting biomass footprints right. *Environmental Impact Assessment Review*, 29, 165-168.

References

- JONES, E., OLIPHANT, T. & PETERSON, P. 2014. SciPy: open source scientific tools for Python. URL: <http://www.scipy.org/>.
- JONSSON, L. J. & MARTIN, C. 2016. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. *Bioresource Technology*, 199, 103-112.
- JUNG, Y. H., KIM, H. K., SONG, D. S., CHOI, I. G., YANG, T. H., LEE, H. J., SEUNG, D. & KIM, K. H. 2014. Feasibility test of utilizing *Saccharophagus degradans* 2-40(T) as the source of crude enzyme for the saccharification of lignocellulose. *Bioprocess and Biosystems Engineering*, 37, 707-710.
- KADAM, K. L. & MCMILLAN, J. D. 2003. Availability of corn stover as a sustainable feedstock for bioethanol production. *Bioresource Technology*, 88, 17-25.
- KALENITCHENKO, D., FAGERVOLD, S. K., PRUSKI, A. M., VETION, G., YUCEL, M., LE BRIS, N. & GALAND, P. E. 2015. Temporal and spatial constraints on community assembly during microbial colonization of wood in seawater. *Isme Journal*, 9, 2657-2670.
- KARLEN, S. D., FREE, H. C. A., PADMAKSHAN, D., SMITH, B. G., RALPH, J. & HARRIS, P. J. 2018. Commelinid Monocotyledon Lignins Are Acylated by p-Coumarate. *Plant Physiology*, 177, 513-521.
- KARLIN, S., BROCCIERI, L., MRAZEK, J. & KAISER, D. 2006. Distinguishing features of delta-proteobacterial genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 11352-11357.
- KARNAOURI, A., MATSAKAS, L., TOPAKAS, E., ROVA, U. & CHRISTAKOPOULOS, P. 2016. Development of Thermophilic Tailor-Made Enzyme Mixtures for the Bioconversion of Agricultural and Forest Residues. *Frontiers in Microbiology*, 7, 14.
- KASHYAP, R., MONIKA & SUBUDHI, E. 2014. A novel thermoalkaliphilic xylanase from *Gordonia* sp. is salt, solvent and surfactant tolerant. *Journal of Basic Microbiology*, 54, 1342-1349.
- KATO, S., CHINO, K., KAMIMURA, N., MASAI, E., YUMOTO, I. & KAMAGATA, Y. 2015. Methanogenic degradation of lignin-derived monoaromatic compounds by microbial enrichments from rice paddy field soil. *Scientific Reports*, 5, 11.
- KEARNS, P. J., ANGELL, J. H., FEINMAN, S. G. & BOWEN, J. L. 2015. Long-term nutrient addition differentially alters community composition and diversity of genes that control nitrous oxide flux from salt marsh sediments. *Estuarine Coastal and Shelf Science*, 154, 39-47.
- KELLEWAY, J. J., CAVANAUGH, K., ROGERS, K., FELLER, I. C., ENS, E., DOUGHTY, C. & SAINTILAN, N. 2017. Review of the ecosystem service implications of mangrove encroachment into salt marshes. *Global Change Biology*, 23, 3967-3983.
- KELLEWAY, J. J., SAINTILAN, N., MACREADIE, P. I. & RALPH, P. J. 2016. Sedimentary Factors are Key Predictors of Carbon Storage in SE Australian Saltmarshes. *Ecosystems*, 19, 865-880.
- KIM, S. & DALE, B. E. 2004. Global potential bioethanol production from wasted crops and crop residues. *Biomass & Bioenergy*, 26, 361-375.
- KIMPLE, M. E., BRILL, A. L. & PASKER, R. L. 2013. Overview of affinity tags for protein purification. *Current protocols in protein science*, 9.9. 1-9.9. 23.
- KIRWAN, M. L. & MUDD, S. M. 2012. Response of salt-marsh carbon accumulation to climate change. *Nature*, 489, 550-+.
- KLEIN-MARCUSCHAMER, D., OLESKOWICZ-POPIEL, P., SIMMONS, B. A. & BLANCH, H. W. 2012. The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnology and Bioengineering*, 109, 1083-1087.
- KLIPPEL, B., LOCHNER, A., BRUCE, D. C., DAVENPORT, K. W., DETTER, C., GOODWIN, L. A., HAN, J., HAN, S. S., LAND, M. L., MIKHAILOVA, N., NOLAN, M., PENNACCHIO, L., PITLUCK, S., TAPIA, R., WOYKE, T., WIEBUSCH, S., BASNER, A., ABE, F., HORIKOSHI, K., KELLER, M. & ANTRANIKIAN, G. 2011. Complete Genome Sequence of the Marine Cellulose- and Xylan-Degrading Bacterium *Glaciecola* sp Strain 4H-3-7+YE-5. *Journal of Bacteriology*, 193, 4547-4548.

References

- KNOTT, B. C., MOMENI, M. H., CROWLEY, M. F., MACKENZIE, L. F., GOTZ, A. W., SANDGREN, M., WITHERS, S. G., STAHLBERG, J. & BECKHAM, G. T. 2014. The Mechanism of Cellulose Hydrolysis by a Two-Step, Retaining Cellobiohydrolase Elucidated by Structural and Transition Path Sampling Studies. *Journal of the American Chemical Society*, 136, 321-329.
- KOSTKA, J. E., PRAKASH, O., OVERHOLT, W. A., GREEN, S. J., FREYER, G., CANION, A., DELGARDIO, J., NORTON, N., HAZEN, T. C. & HUETTEL, M. 2011. Hydrocarbon-Degrading Bacteria and the Bacterial Community Response in Gulf of Mexico Beach Sands Impacted by the Deepwater Horizon Oil Spill. *Applied and Environmental Microbiology*, 77, 7962-7974.
- KRUYS, A., ERIKSSON, O. E. & WEDIN, M. 2006. Phylogenetic relationships of coprophilous Pleosporales (Dothideomycetes, Ascomycota), and the classification of some bitunicate taxa of unknown position. *Mycological Research*, 110, 527-536.
- KRZYWINSKI, M., SCHEIN, J., BIROL, I., CONNORS, J., GASCOYNE, R., HORSMAN, D., JONES, S. J. & MARRA, M. A. 2009. Circos: An information aesthetic for comparative genomics. *Genome Research*, 19, 1639-1645.
- KUMAR, D. & SINGH, V. 2016. Dry-grind processing using amylase corn and superior yeast to reduce the exogenous enzyme requirements in bioethanol production. *Biotechnology for Biofuels*, 9, 12.
- KUMAR, M., CAMPBELL, L. & TURNER, S. 2016. Secondary cell walls: biosynthesis and manipulation. *Journal of Experimental Botany*, 67, 515-531.
- LABRENZ, M., LAWSON, P. A., TINDALL, B. J., COLLINS, M. D. & HIRSCH, P. 2003. *Saccharosporillum impatiens* gen. nov., sp nov., a novel gamma-Proteobacterium isolated from hypersaline Ekho Lake (East Antarctica). *International Journal of Systematic and Evolutionary Microbiology*, 53, 653-660.
- LAFOND, M., NAVARRO, D., HAON, M., COUTURIER, M. & BERRIN, J. G. 2012. Characterization of a Broad-Specificity beta-Glucanase Acting on beta-(1,3)-, beta-(1,4)-, and beta-(1,6)-Glucans That Defines a New Glycoside Hydrolase Family. *Applied and Environmental Microbiology*, 78, 8540-8546.
- LAMB, A. L., VANE, C. H., WILSON, G. P., REES, J. G. & MOSS-HAYES, V. L. 2007. Assessing delta C-13 and C/N ratios from organic material in archived cores as Holocene sea level and palaeoenvironmental indicators in the Humber Estuary, UK. *Marine Geology*, 244, 109-128.
- LAMBAIS, M. R., OTERO, X. L. & CURY, J. C. 2008. Bacterial communities and biogeochemical transformations of iron and sulfur in a high saltmarsh soil profile. *Soil Biology & Biochemistry*, 40, 2854-2864.
- LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357-U54.
- LASHER, C., DYSZYNSKI, G., EVERETT, K., EDMONDS, J., YE, W. Y., SHELDON, W., WANG, S. Y., JOYE, S. B., MORAN, M. A. & WHITMAN, W. B. 2009. The Diverse Bacterial Community in Intertidal, Anaerobic Sediments at Sapelo Island, Georgia. *Microbial Ecology*, 58, 244-261.
- LBNET & E4TECH 2016. As assessment of the potential for the establishment of lignocellulosic biorefineries in the UK. *Lignocellulosic Biorefinery Network. UC/PROC/415*.
- LEE, S. H., MEGONIGAL, P. J., LANGLEY, A. J. & KANG, H. 2017. Elevated CO₂ and nitrogen addition affect the microbial abundance but not the community structure in salt marsh ecosystem. *Applied Soil Ecology*, 117, 129-136.
- LEE, T. M., FARROW, M. F., ARNOLD, F. H. & MAYO, S. L. 2011. A structural study of Hypocrea jecorina Cel5A. *Protein Science*, 20, 1935-1940.
- LEVASSEUR, A., DRULA, E., LOMBARD, V., COUTINHO, P. M. & HENRISSAT, B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels*, 6, 14.
- LEVER, M. 1977. Carbohydrate determination with 4-hydroxybenzoic acid hydrazide (pahbah) - effect of bismuth on reaction. *Analytical Biochemistry*, 81, 21-27.

References

- LI, H. L., LIU, Z. Y., WANG, C. L., HUANG, S. C. & ZHAO, M. 2015. Secretory expression and characterization of a novel thermo-stable, salt-tolerant endo-1,4-beta-mannanase of *Bacillus subtilis* WD23 by *Pichia pastoris*. *European Food Research and Technology*, 240, 671-677.
- LI, J. B., CAI, S. C., LUO, Y. M. & DONG, X. Z. 2011. Three Feruloyl Esterases in *Cellulosilyticum ruminicola* H1 Act Synergistically To Hydrolyze Esterified Polysaccharides. *Applied and Environmental Microbiology*, 77, 6141-6147.
- LI, M., PU, Y. Q. & RAGAUSKAS, A. J. 2016a. Current Understanding of the Correlation of Lignin Structure with Biomass Recalcitrance. *Frontiers in Chemistry*, 4, 8.
- LI, Q. Q., WANG, F. P., CHEN, Z. W., YIN, X. J. & XIAO, X. 2012. Stratified active archaeal communities in the sediments of Jiulong River estuary, China. *Frontiers in Microbiology*, 3, 14.
- LI, X. & YU, H. Y. 2013. Halostable cellulase with organic solvent tolerance from *Haloarcula* sp LLSG7 and its application in bioethanol fermentation using agricultural wastes. *Journal of Industrial Microbiology & Biotechnology*, 40, 1357-1365.
- LI, Z. B., YANG, J. & LOH, X. J. 2016b. Polyhydroxyalkanoates: opening doors for a sustainable future. *Npg Asia Materials*, 8, 20.
- LI, Z. P., LIU, S. G., TAN, Z. X., BLISS, N. B., YOUNG, C. J., WEST, T. O. & OGLE, S. M. 2014. Comparing cropland net primary production estimates from inventory, a satellite-based model, and a process-based model in the Midwest of the United States. *Ecological Modelling*, 277, 1-12.
- LIAO, C. Z., LUO, Y. Q., FANG, C. M., CHEN, J. K. & LI, B. 2008. Litter pool sizes, decomposition, and nitrogen dynamics in *Spartina alterniflora*-invaded and native coastal marshlands of the Yangtze Estuary. *Oecologia*, 156, 589-600.
- LIU, J., WANG, X. C., PU, H. M., LIU, S., KAN, J. & JIN, C. H. 2017a. Recent advances in endophytic exopolysaccharides: Production, structural characterization, physiological role and biological activity. *Carbohydrate Polymers*, 157, 1113-1124.
- LIU, R. M., LIANG, L. Y., CHEN, K. Q., MA, J. F., JIANG, M., WEI, P. & OUYANG, P. K. 2012. Fermentation of xylose to succinate by enhancement of ATP supply in metabolically engineered *Escherichia coli*. *Applied Microbiology and Biotechnology*, 94, 959-968.
- LIU, X., HOEKMAN, S. K. & BROCH, A. 2017b. Potential water requirements of increased ethanol fuel in the USA. *Energy Sustainability and Society*, 7, 13.
- LIU, X. S., HUANG, Z. Q., ZHANG, X. N., SHAO, Z. Z. & LIU, Z. D. 2014. Cloning, expression and characterization of a novel cold-active and halophilic xylanase from *Zunongwangia profunda*. *Extremophiles*, 18, 441-450.
- LOMBARD, L., VAN DER MERWE, N. A., GROENEWALD, J. Z. & CROUS, P. W. 2015. Generic concepts in Nectriaceae. *Studies in Mycology*, 189-245.
- LOMBARD, V., BERNARD, T., RANCUREL, C., BRUMER, H., COUTINHO, P. M. & HENRISSAT, B. 2010. A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochemical Journal*, 432, 437-444.
- LOMBARD, V., RAMULU, H. G., DRULA, E., COUTINHO, P. M. & HENRISSAT, B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, 42, D490-D495.
- LOWE, K. L., DICHRISTINA, T. J., ROYCHOUDHURY, A. N. & VAN CAPPELLEN, P. 2000. Microbiological and geochemical characterization of microbial Fe(III) reduction in salt marsh sediments. *Geomicrobiology Journal*, 17, 163-176.
- LU, Y., LU, Y. C., HU, H. Q., XIE, F. J., WEI, X. Y. & FAN, X. 2017. Structural Characterization of Lignin and Its Degradation Products with Spectroscopic Methods. *Journal of Spectroscopy*, 15.
- LUNDBERG, D. S., YOURSTONE, S., MIECZKOWSKI, P., JONES, C. D. & DANGL, J. L. 2013. Practical innovations for high-throughput amplicon sequencing. *Nature Methods*, 10, 999-+.
- LYDELL, C., DOWELL, L., SIKAROODI, M., GILLEVET, P. & EMERSON, D. 2004. A population survey of members of the phylum Bacteroidetes isolated from salt marsh sediments along the East Coast of the United States. *Microbial Ecology*, 48, 263-273.

References

- MA, J. S., ZHANG, K. K., LIAO, H. D., HECTOR, S. B., SHI, X. W., LI, J. L., LIU, B., XU, T., TONG, C. Y., LIU, X. M. & ZHU, Y. H. 2016. Genomic and secretomic insight into lignocellulolytic system of an endophytic bacterium *Pantoea ananatis* Sd-1. *Biotechnology for Biofuels*, 9, 15.
- MA, Y. X., TAO, W., LIU, J., LIU, C. F., LI, J. & LIU, J. C. 2018. Response of ammonia-oxidizing betaproteobacteria to short-term fertilization in a salt marsh in China. *Journal of Oceanology and Limnology*, 36, 351-361.
- MADERN, D., EBEL, C. & ZACCAI, G. 2000. Halophilic adaptation of enzymes. *Extremophiles*, 4, 91-98.
- MAI, Z. M., YANG, J., TIAN, X. P., LI, J. & ZHANG, S. 2013. Gene Cloning and Characterization of a Novel Salt-Tolerant and Glucose-Enhanced beta-Glucosidase from a Marine Streptomyces. *Applied Biochemistry and Biotechnology*, 169, 1512-1522.
- MANGAS-SANCHEZ, J. & ADLERCREUTZ, P. 2015. Enzymatic preparation of oligosaccharides by transglycosylation: A comparative study of glucosidases. *Journal of Molecular Catalysis B-Enzymatic*, 122, 51-55.
- MARCHAND, C., DISNAR, J. R., LALLIER-VERG, E. & LOTTIER, N. 2005. Early diagenesis of carbohydrates and lignin in mangrove sediments subject to variable redox conditions (French Guiana). *Geochimica Et Cosmochimica Acta*, 69, 131-142.
- MARINOVIC, M., NOUSIAINEN, P., DILOKPIMOL, A., KONTRO, J., MOORE, R., SIPILA, J., DE VRIES, R. P., MAKELA, M. R. & HILDEN, K. 2018. Selective Cleavage of Lignin beta-O-4 Aryl Ether Bond by beta-Etherase of the White-Rot Fungus *Dichomitus squalens*. *Acs Sustainable Chemistry & Engineering*, 6, 2878-2882.
- MARRIOTT, P. E., GOMEZ, L. D. & MCQUEEN-MASON, S. J. 2016. Unlocking the potential of lignocellulosic biomass through plant science. *New Phytologist*, 209, 1366-1381.
- MARSH, A., BLUM, L. K., CHRISTIAN, R. R., RAMSEY, E. & RANGOONWALA, A. 2016. Response and resilience of *Spartina alterniflora* to sudden dieback. *Journal of Coastal Conservation*, 20, 335-350.
- MARTIN, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17, pp. 10-12.
- MARTIN, M., AHMETOVIC, E. & GROSSMANN, I. E. 2011. Optimization of Water Consumption in Second Generation Bioethanol Plants. *Industrial & Engineering Chemistry Research*, 50, 3705-3721.
- MARTIN, M., BIVER, S., STEELS, S., BARBEYRON, T., JAM, M., PORTETELLE, D., MICHEL, G. & VANDENBOL, M. 2014. Identification and Characterization of a Halotolerant, Cold-Active Marine Endo-beta-1,4-Glucanase by Using Functional Metagenomics of Seaweed-Associated Microbiota. *Applied and Environmental Microbiology*, 80, 4958-4967.
- MARTINEZ, D., BERKA, R. M., HENRISSAT, B., SALOHEIMO, M., ARVAS, M., BAKER, S. E., CHAPMAN, J., CHERTKOV, O., COUTINHO, P. M., CULLEN, D., DANCHIN, E. G. J., GRIGORIEV, I. V., HARRIS, P., JACKSON, M., KUBICEK, C. P., HAN, C. S., HO, I., LARRONDO, L. F., DE LEON, A. L., MAGNUSON, J. K., MERINO, S., MISRA, M., NELSON, B., PUTNAM, N., ROBBERTSE, B., SALAMOV, A. A., SCHMOLL, M., TERRY, A., THAYER, N., WESTERHOLM-PARVINEN, A., SCHOCH, C. L., YAO, J., BARBOTE, R., NELSON, M. A., DETTER, C., BRUCE, D., KUSKE, C. R., XIE, G., RICHARDSON, P., ROKHSAR, D. S., LUCAS, S. M., RUBIN, E. M., DUNN-COLEMAN, N., WARD, M. & BRETTIN, T. S. 2008. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nature Biotechnology*, 26, 553-560.
- MASAI, E., KATAYAMA, Y., KAWAI, S., NISHIKAWA, S., YAMASAKI, M. & MOROHOSHI, N. 1991. Cloning and sequencing of the gene for a pseudomonas-paucimobilis enzyme that cleaves beta-aryl ether. *Journal of Bacteriology*, 173, 7950-7955.
- MASAI, E., KATAYAMA, Y., NISHIKAWA, S., YAMASAKI, M., MOROHOSHI, N. & HARAGUCHI, T. 1989. Detection and localization of a new enzyme catalyzing the beta-aryl ether cleavage in the soil bacterium (*pseudomonas-paucimobilis* syk-6). *Febs Letters*, 249, 348-352.
- MATE, D. M. & ALCALDE, M. 2017. Laccase: a multi-purpose biocatalyst at the forefront of biotechnology. *Microbial Biotechnology*, 10, 1457-1467.

References

- MCBRIDE, M. J. 2014. The Family Flavobacteriaceae. In: ROSENBERG, E., DELONG, E. F., LORY, S., STACKEBRANDT, E. & THOMPSON, F. (eds.) *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*. Berlin, Heidelberg: Springer Berlin Heidelberg, 643-676.
- MCGOVERN, T. A., LABER, L. J. & GRAM, B. C. 1979. Characteristics of the salts secreted by *Spartina alterniflora* Loisel and their relation to estuarine production. *Estuarine and Coastal Marine Science*, 9, 351-356.
- MCKENDRY, P. 2002. Energy production from biomass (part 1): overview of biomass. *Bioresource Technology*, 83, 37-46.
- MCKINNEY, W. Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference, 2010. Austin, TX, 51-56.
- MCLEOD, E., CHMURA, G. L., BOUILLON, S., SALM, R., BJORK, M., DUARTE, C. M., LOVELOCK, C. E., SCHLESINGER, W. H. & SILLIMAN, B. R. 2011. A blueprint for blue carbon: toward an improved understanding of the role of vegetated coastal habitats in sequestering CO₂. *Frontiers in Ecology and the Environment*, 9, 552-560.
- MCOWEN, C. J., WEATHERDON, L. V., VAN BOCHOVE, J.-W., SULLIVAN, E., BLYTH, S., ZOCKLER, C., STANWELL-SMITH, D., KINGSTON, N., MARTIN, C. S. & SPALDING, M. 2017. A global map of saltmarshes. *Biodiversity data journal*, 5, 1-13.
- MELLEROWICZ, E. J. & SUNDBERG, B. 2008. Wood cell walls: biosynthesis, developmental dynamics and their implications for wood properties. *Current Opinion in Plant Biology*, 11, 293-300.
- MENENDEZ, M. & SANMARTI, N. 2007. Geratology and decomposition of *Spartina versicolor* in a brackish Mediterranean marsh. *Estuarine Coastal and Shelf Science*, 74, 320-330.
- MENG, D. D., YING, Y., CHEN, X. H., LU, M., NING, K., WANG, L. S. & LI, F. L. 2015. Distinct Roles for Carbohydrate-Binding Modules of Glycoside Hydrolase 10 (GH10) and GH11 Xylanases from *Caldicellulosiruptor* sp Strain F32 in Thermostability and Catalytic Efficiency. *Applied and Environmental Microbiology*, 81, 2006-2014.
- MESA, L., MARTINEZ, Y., BARRIO, E. & GONZALEZ, E. 2017. Desirability function for optimization of Dilute Acid pretreatment of sugarcane straw for ethanol production and preliminary economic analysis based in three fermentation configurations. *Applied Energy*, 198, 299-311.
- MOHAMED, D. J. & MARTINY, J. B. H. 2011. Patterns of fungal diversity and composition along a salinity gradient. *Isme Journal*, 5, 379-388.
- MOHAMMAD, N., ALAM, M. Z., KABBASHI, N. A. & AHSAN, A. 2012. Effective composting of oil palm industrial waste by filamentous fungi: A review. *Resources Conservation and Recycling*, 58, 69-78.
- MOIN, N. S., NELSON, K. A., BUSH, A. & BERNHARD, A. E. 2009. Distribution and Diversity of Archaeal and Bacterial Ammonia Oxidizers in Salt Marsh Sediments. *Applied and Environmental Microbiology*, 75, 7461-7468.
- MONTEMAYOR, D. I., ADDINO, M., VALINAS, M., FANJUL, E., ALVAREZ, M. F. & IRIBARNE, O. O. 2015. Biomass dynamics of the two dominant SW Atlantic *Spartina* species and its implications on the saltmarsh organic matter accumulation/exportation. *Aquatic Botany*, 120, 201-204.
- MOREAU, J. W., ZIERENBERG, R. A. & BANFIELD, J. F. 2010. Diversity of Dissimilatory Sulfite Reductase Genes (*dsrAB*) in a Salt Marsh Impacted by Long-Term Acid Mine Drainage. *Applied and Environmental Microbiology*, 76, 4819-4828.
- MOREIRA-VILAR, F. C., SIQUEIRA-SOARES, R. D., FINGER-TEIXEIRA, A., DE OLIVEIRA, D. M., FERRO, A. P., DA ROCHA, G. J., FERRARESE, M. D. L., DOS SANTOS, W. D. & FERRARESE, O. 2014. The Acetyl Bromide Method Is Faster, Simpler and Presents Best Recovery of Lignin in Different Herbaceous Tissues than Klason and Thioglycolic Acid Methods. *Plos One*, 9, 7.
- MORRIS, J. T., BARBER, D. C., CALLAWAY, J. C., CHAMBERS, R., HAGEN, S. C., HOPKINSON, C. S., JOHNSON, B. J., MEGONIGAL, P., NEUBAUER, S. C., TROXLER, T. & WIGAND, C. 2016. Contributions of organic and inorganic matter to sediment volume and accretion in tidal wetlands at steady state. *Earths Future*, 4, 110-121.

References

- MORRIS, J. T., SUNDBERG, K. & HOPKINSON, C. S. 2013. Salt Marsh Primary Production and Its Responses to Relative Sea Level and Nutrients in Estuaries at Plum Island, Massachusetts, and North Inlet, South Carolina, USA. *Oceanography*, 26, 78-84.
- MUKHERJEE, I. & SOVACOOOL, B. K. 2014. Palm oil-based biofuels and sustainability in southeast Asia: A review of Indonesia, Malaysia, and Thailand. *Renewable & Sustainable Energy Reviews*, 37, 1-12.
- MULLER, G., VARNAI, A., JOHANSEN, K. S., EIJSINK, V. G. H. & HORN, S. J. 2015. Harnessing the potential of LPMO-containing cellulase cocktails poses new demands on processing conditions. *Biotechnology for Biofuels*, 8, 9.
- NAAS, A. E., MACKENZIE, A. K., DALHUS, B., EIJSINK, V. G. H. & POPE, P. B. 2015. Structural Features of a Bacteroidetes-Affiliated Cellulase Linked with a Polysaccharide Utilization Locus. *Scientific Reports*, 5, 12.
- NAIK, S. N., GOUD, V. V., ROUT, P. K. & DALAI, A. K. 2010. Production of first and second generation biofuels: A comprehensive review. *Renewable & Sustainable Energy Reviews*, 14, 578-597.
- NAKAMURA, A. M., NASCIMENTO, A. S. & POLIKARPOV, I. 2017. Structural diversity of carbohydrate esterases. *Biotechnology Research and Innovation*, 1, 35-51.
- NAYEK, A., SEN GUPTA, P. S., BANERJEE, S., MONDAL, B. & BANDYOPADHYAY, A. K. 2014. Salt-Bridge Energetics in Halophilic Proteins. *Plos One*, 9, 11.
- NEGRIN, V. L., TRILLA, G. G., KANDUS, P. & MARCOVECCHIO, J. E. 2012. DECOMPOSITION AND NUTRIENT DYNAMICS IN A Spartina alterniflora MARSH OF THE BAHIA BLANCA ESTUARY, ARGENTINA. *Brazilian Journal of Oceanography*, 60, 259-263.
- NELSON, J. L. & ZAVALETA, E. S. 2012. Salt Marsh as a Coastal Filter for the Oceans: Changes in Function with Experimental Increases in Nitrogen Loading and Sea-Level Rise. *Plos One*, 7, 14.
- NELSON, K. A., MOIN, N. S. & BERNHARD, A. E. 2009. Archaeal Diversity and the Prevalence of Crenarchaeota in Salt Marsh Sediments. *Applied and Environmental Microbiology*, 75, 4211-4215.
- NEWELL, S. Y. 2001. Spore-expulsion rates and extents of blade occupation by ascomycetes of the smooth-cordgrass standing-decay system. *Botanica Marina*, 44, 277-285.
- NICELL, J. A. & WRIGHT, H. 1997. A model of peroxidase activity with inhibition by hydrogen peroxide. *Enzyme and Microbial Technology*, 21, 302-310.
- NIKALJE, G. C. & SUPRASANNA, P. 2018. Coping With Metal Toxicity - Cues From Halophytes. *Frontiers in Plant Science*, 9, 11.
- NILADEVI, K. N., JACOB, N. & PREMA, P. 2008. Evidence for a halotolerant-alkaline laccase in Streptomyces psammoticus: Purification and characterization. *Process Biochemistry*, 43, 654-660.
- NOBLE, W. S. 2015. Mass spectrometrists should search only for peptides they care about. *Nature Methods*, 12, 605-608.
- NOLD, S. C., ZHOU, J. Z., DEVOL, A. H. & TIEDJE, J. M. 2000. Pacific northwest marine sediments contain ammonia-oxidizing bacteria in the beta subdivision of the Proteobacteria. *Applied and Environmental Microbiology*, 66, 4532-4535.
- NOVY, R. A. M. B. 2001. Novagen, Inc. Report use of glucose to control basal expression in the pET system. 13, 6-7.
- NWODO, U. U., GREEN, E. & OKOH, A. I. 2012. Bacterial Exopolysaccharides: Functionality and Prospects. *International Journal of Molecular Sciences*, 13, 14002-14015.
- OHTA, K., FUJIMOTO, H., FUJII, S. & WAKIYAMA, M. 2010. Cell-associated beta-xylosidase from Aureobasidium pullulans ATCC 20524: Purification, properties, and characterization of the encoding gene. *Journal of Bioscience and Bioengineering*, 110, 152-157.
- OLIPHANT, T. E. 2006. *A guide to NumPy*, Trelgol Publishing USA. URL: <http://web.mit.edu/dvp/Public/numpybook.pdf>.

References

- OLSEN, J. P., ALASEPP, K., KARI, J., CRUYS-BAGGER, N., BORCH, K. & WESTH, P. 2016. Mechanism of product inhibition for cellobiohydrolase Cel7A during hydrolysis of insoluble cellulose. *Biotechnology and Bioengineering*, 113, 1178-1186.
- ONI, O. E., SCHMIDT, F., MIYATAKE, T., KASTEN, S., WITT, M., HINRICH, K. U. & FRIEDRICH, M. W. 2015. Microbial Communities and Organic Matter Composition in Surface and Subsurface Sediments of the Helgoland Mud Area, North Sea. *Frontiers in Microbiology*, 6, 16.
- ORTNER, A., HUBER, D., HASKE-CORNELIUS, O., WEBER, H. K., HOFER, K., BAUER, W., NYANHONGO, G. S. & GUEBITZ, G. M. 2015. Laccase mediated oxidation of industrial lignins: Is oxygen limiting? *Process Biochemistry*, 50, 1277-1283.
- PANG, J., LIU, Z. Y., HAO, M., ZHANG, Y. F. & QI, Q. S. 2017. An isolated cellulolytic *Escherichia coli* from bovine rumen produces ethanol and hydrogen from corn straw. *Biotechnology for Biofuels*, 10, 10.
- PARADA, A. E., NEEDHAM, D. M. & FUHRMAN, J. A. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, 18, 1403-1414.
- PARK, T. H., CHOI, M. W., PARK, C. S., LEE, S. B., KANG, H. Y., SHON, K. J., PARK, J. S. & CHA, J. H. 2005. Substrate specificity and transglycosylation catalyzed by a thermostable beta-glucosidase from marine hyperthermophile *Thermotoga neapolitana*. *Applied Microbiology and Biotechnology*, 69, 411-422.
- PARK, Y. B. & COSGROVE, D. J. 2015. Xyloglucan and its Interactions with Other Components of the Growing Cell Wall. *Plant and Cell Physiology*, 56, 180-194.
- PAUL, S., BAG, S. K., DAS, S., HARVILL, E. T. & DUTTA, C. 2008. Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. *Genome Biology*, 9, 19.
- PAULY, M., GILLE, S., LIU, L. F., MANSOORI, N., DE SOUZA, A., SCHULTINK, A. & XIONG, G. Y. 2013. Hemicellulose biosynthesis. *Planta*, 238, 627-642.
- PAYNE, C. M., JIANG, W., SHIRTS, M. R., HIMMEL, M. E., CROWLEY, M. F. & BECKHAM, G. T. 2013. Glycoside Hydrolase Processivity Is Directly Related to Oligosaccharide Binding Free Energy. *Journal of the American Chemical Society*, 135, 18831-18839.
- PERKINS, D. N., PAPPIN, D. J. C., CREASY, D. M. & COTTRELL, J. S. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20, 3551-3567.
- PETERSEN, T. N., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8, 785-786.
- PETRO, C., STARNAWSKI, P., SCHRAMM, A. & KJELDSSEN, K. U. 2017. Microbial community assembly in marine sediments. *Aquatic Microbial Ecology*, 79, 177-195.
- PICART, P., MULLER, C., MOTTWEILER, J., WIERMANS, L., BOLM, C., DE MARIA, P. D. & SCHALLMEY, A. 2014. From Gene Towards Selective Biomass Valorization: Bacterial beta-Etherases with Catalytic Activity on Lignin-Like Polymers. *Chemsuschem*, 7, 3164-3171.
- POKORNA, D. & ZABRANSKA, J. 2015. Sulfur-oxidizing bacteria in environmental technology. *Biotechnology Advances*, 33, 1246-1259.
- POLI, A., DI DONATO, P., ABBAMONDI, G. R. & NICOLAUS, B. 2011. Synthesis, Production, and Biotechnological Applications of Exopolysaccharides and Polyhydroxyalkanoates by Archaea. *Archaea-an International Microbiological Journal*, 13.
- POTHIRAJ, C., KANMANI, P. & BALAJI, P. J. M. 2006. Bioconversion of lignocellulose materials. 34, 159-165.
- PRAESTGAARD, E., ELMERDAHL, J., MURPHY, L., NYMAND, S., MCFARLAND, K. C., BORCH, K. & WESTH, P. 2011. A kinetic model for the burst phase of processive cellulases. *Febs Journal*, 278, 1547-1560.

References

- PRAKASH, B., VIDYASAGAR, M., JAYALAKSHMI, S. K. & SREERAMULU, K. 2012. Purification and some properties of low-molecular-weight extreme halophilic xylanase from *Chromohalobacter* sp TPSV 101. *Journal of Molecular Catalysis B-Enzymatic*, 74, 192-198.
- PURAHONG, W., WUBET, T., LENTENDU, G., SCHLOTTER, M., PECYNA, M. J., KAPTURSKA, D., HOFRICHTER, M., KRUGER, D. & BUSCOT, F. 2016. Life in leaf litter: novel insights into community dynamics of bacteria and fungi during litter decomposition. *Molecular Ecology*, 25, 4059-4074.
- QUAST, C., PRUESSE, E., YILMAZ, P., GERKEN, J., SCHWEER, T., YARZA, P., PEPLIES, J. & GLOCKNER, F. O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41, D590-D596.
- RAJAN, K. & CARRIER, D. J. 2014. Characterization of Rice Straw Prehydrolyzates and Their Effect on the Hydrolysis of Model Substrates Using a Commercial endo-Cellulase, beta-Glucosidase and Cellulase Cocktail. *Acs Sustainable Chemistry & Engineering*, 2, 2124-2130.
- RAVANAL, M. C., CALLEGARI, E. & EYZAGUIRRE, J. 2010. Novel Bifunctional alpha-L-Arabinofuranosidase/Xylobiohydrolase (ABF3) from *Penicillium purpurogenum*. *Applied and Environmental Microbiology*, 76, 5247-5253.
- RAWESRI, P., RIANGRUNGROJANA, P. & PINPHANICHAKARN, P. 2008. alpha-L-Arabinofuranosidase from *Streptomyces* sp PC22: Purification, characterization and its synergistic action with xylanolytic enzymes in the degradation of xylan and agricultural residues. *Bioresource Technology*, 99, 8981-8986.
- REHM, B. H. A. 2010. Bacterial polymers: biosynthesis, modifications and applications. *Nature Reviews Microbiology*, 8, 578-592.
- REYES-ORTIZ, V., HEINS, R. A., CHENG, G., KIM, E. Y., VERNON, B. C., ELANDT, R. B., ADAMS, P. D., SALE, K. L., HADI, M. Z., SIMMONS, B. A., KENT, M. S. & TULLMAN-ERCEK, D. 2013. Addition of a carbohydrate-binding module enhances cellulase penetration into cellulose substrates. *Biotechnology for Biofuels*, 6, 13.
- REYES-SOSA, F. M., MORALES, M. L., GOMEZ, A. I. P., CRESPO, N. V., ZAMORANO, L. S., ROCHA-MARTIN, J., MOLINA-HEREDIA, F. P. & GARCIA, B. D. 2017. Management of enzyme diversity in high-performance cellulolytic cocktails. *Biotechnology for Biofuels*, 10, 10.
- RFA 2018. Ethanol Strong, 2018 Ethanol Industry Outlook. URL: <http://www.ethanolresponse.com/wp-content/uploads/2018/02/2018-RFA-Ethanol-Industry-Outlook.pdf>. *Renewable Fuels Association*.
- RILEY, T. V. 2012. 22 - Clostridium: Gas gangrene; tetanus; food poisoning; pseudomembranous colitis. In: GREENWOOD, D., BARER, M., SLACK, R. & IRVING, W. (eds.) *Medical Microbiology (Eighteenth Edition)*. Edinburgh: Churchill Livingstone, 1-794.
- ROCHA-MARTIN, J., MARTINEZ-BERNAL, C., PEREZ-COBAS, Y., REYES-SOSA, F. M. & GARCIA, B. D. 2017. Additives enhancing enzymatic hydrolysis of lignocellulosic biomass. *Bioresource Technology*, 244, 48-56.
- ROGNES, T., FLOURI, T., NICHOLS, B., QUINCE, C. & MAHE, F. 2016. VSEARCH: a versatile open source tool for metagenomics. *Peerj*, 4, 22.
- ROGOWSKI, A., BASLE, A., FARINAS, C. S., SOLOVYOVA, A., MORTIMER, J. C., DUPREE, P., GILBERT, H. J. & BOLAM, D. N. 2014. Evidence That GH115 alpha-Glucuronidase Activity, Which Is Required to Degrade Plant Biomass, Is Dependent on Conformational Flexibility. *Journal of Biological Chemistry*, 289, 53-64.
- ROMANENKO, L. A., SCHUMANN, P., ROHDE, M., MIKHAILOV, V. V. & STACKEBRANDT, E. 2004. *Reinekea marinisedimentorum* gen. nov., sp nov., a novel gammaproteobacterium from marine coastal sediments. *International Journal of Systematic and Evolutionary Microbiology*, 54, 669-673.
- ROSANO, G. L. & CECCARELLI, E. A. 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*, 5, 17.

References

- ROSENBERG, E. 2014. The Family Marinilabiaceae. In: ROSENBERG, E., DELONG, E. F., LORY, S., STACKEBRANDT, E. & THOMPSON, F. (eds.) *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*. Berlin, Heidelberg: Springer Berlin Heidelberg, 731-732.
- SABBADIN, F., HEMSWORTH, G. R., CIANO, L., HENRISSAT, B., DUPREE, P., TRYFONA, T., MARQUES, R. D. S., SWEENEY, S. T., BESSER, K., ELIAS, L., PESANTE, G., LI, Y., DOWLE, A. A., BATES, R., GOMEZ, L. D., SIMISTER, R., DAVIES, G. J., WALTON, P. H., BRUCE, N. C. & MCQUEEN-MASON, S. J. 2018. An ancient family of lytic polysaccharide monoxygenases with roles in arthropod development and biomass digestion. *Nature Communications*, 9, 12.
- SAEMAN, J. F. 1945. Kinetics of wood saccharification - hydrolysis of cellulose and decomposition of sugars in dilute acid at high temperature. *Industrial and Engineering Chemistry*, 37, 43-52.
- SAINTILAN, N., WILSON, N. C., ROGERS, K., RAJKARAN, A. & KRAUSS, K. W. 2014. Mangrove expansion and salt marsh decline at mangrove poleward limits. *Global Change Biology*, 20, 147-157.
- SALLES, J. F., SILVA, M., DINI-ANDREOTE, F., DIAS, A. C. F., GUILLAUMAUD, N., POLY, F. & VAN ELSAS, J. D. 2017. Successional patterns of key genes and processes involved in the microbial nitrogen cycle in a salt marsh chronosequence. *Biogeochemistry*, 132, 185-201.
- SANCHEZ-AMAT, A., SOLANO, F. & LUCAS-ELIO, P. 2010. Finding New Enzymes from Bacterial Physiology: A Successful Approach Illustrated by the Detection of Novel Oxidases in *Marinomonas mediterranea*. *Marine Drugs*, 8, 519-541.
- SANNI, G. O., COULON, F. & MCGENITY, T. J. 2015. Dynamics and distribution of bacterial and archaeal communities in oil-contaminated temperate coastal mudflat mesocosms. *Environmental Science and Pollution Research*, 22, 15230-15247.
- SANTIN, C., GONZALEZ-PEREZ, M., OTERO, X. L., VIDAL-TORRADO, P., MACIAS, F. & ALVAREZ, M. A. 2008. Characterization of humic substances in salt marsh soils under sea rush (*Juncus maritimus*). *Estuarine Coastal and Shelf Science*, 79, 541-548.
- SANTORO, A. L., BASTVIKEN, D., TRANVIK, L. & ENRICH-PRAST, A. 2013. Simultaneous measurements of dark carbon fixation and bacterial production in lake sediment. *Limnology and Oceanography-Methods*, 11, 298-303.
- SARKAR, N., GHOSH, S. K., BANNERJEE, S. & AIKAT, K. 2012. Bioethanol production from agricultural wastes: An overview. *Renewable Energy*, 37, 19-27.
- SATLEWAL, A., AGRAWAL, R., BHAGIA, S., DAS, P. & RAGAUSKAS, A. J. 2018. Rice straw as a feedstock for biofuels: Availability, recalcitrance, and chemical properties. *Biofuels Bioproducts & Biorefining-Biofpr*, 12, 83-107.
- SCHELLER, H. V. & ULVSKOV, P. 2010. Hemicelluloses. In: MERCHANT, S., BRIGGS, W. R. & ORT, D. (eds.) *Annual Review of Plant Biology*, Vol 61. Palo Alto: Annual Reviews, 61, 263-289.
- SCHNEIDER, T., KEIBLINGER, K. M., SCHMID, E., STERFLINGER-GLEIXNER, K., ELLERSDORFER, G., ROSCHITZKI, B., RICHTER, A., EBERL, L., ZECHMEISTER-BOLTENSTERN, S. & RIEDEL, K. 2012. Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. *Isme Journal*, 6, 1749-1762.
- SCHOCH, C. L., SEIFERT, K. A., HUHNDORF, S., ROBERT, V., SPOUGE, J. L., LEVESQUE, C. A., CHEN, W., BOLCHACOVA, E., VOIGT, K., CROUS, P. W., MILLER, A. N., WINGFIELD, M. J., AIME, M. C., AN, K. D., BAI, F. Y., BARRETO, R. W., BEGEROW, D., BERGERON, M. J., BLACKWELL, M., BOEKHOUT, T., BOGALE, M., BOONYUEN, N., BURGAZ, A. R., BUYCK, B., CAI, L., CAI, Q., CARDINALI, G., CHAVERRI, P., COPPINS, B. J., CRESPO, A., CUBAS, P., CUMMINGS, C., DAMM, U., DE BEER, Z. W., DE HOOG, G. S., DEL-PRADO, R., DENTINGER, B., DIEGUEZ-URIBEONDO, J., DIVAKAR, P. K., DOUGLAS, B., DUENAS, M., DUONG, T. A., EBERHARDT, U., EDWARDS, J. E., ELSHAHED, M. S., FLIEGEROVA, K., FURTADO, M., GARCIA, M. A., GE, Z. W., GRIFFITH, G. W., GRIFFITHS, K., GROENEWALD, J. Z., GROENEWALD, M., GRUBE, M., GRYZENHOUT, M., GUO, L. D., HAGEN, F., HAMBLETON, S., HAMELIN, R. C., HANSEN, K., HARROLD, P., HELLER, G., HERRERA, G., HIRAYAMA, K., HIROOKA, Y., HO, H. M., HOFFMANN, K., HOFSTETTER, V., HOGNABBA, F., HOLLINGSWORTH, P. M., HONG, S. B., HOSAKA, K., HOUBRAKEN, J., HUGHES,

References

- K., HUHTINEN, S., HYDE, K. D., JAMES, T., JOHNSON, E. M., JOHNSON, J. E., JOHNSTON, P. R., JONES, E. B., KELLY, L. J., KIRK, P. M., KNAPP, D. G., KOLJALG, U., KOVACS, G. M., KURTZMAN, C. P., LANDVIK, S., LEAVITT, S. D., LIGGENSTOFFER, A. S., LIIMATAINEN, K., LOMBARD, L., LUANGSA-ARD, J. J., LUMBSCH, H. T., MAGANTI, H., MAHARACHCHIKUMBURA, S. S., MARTIN, M. P., MAY, T. W., MCTAGGART, A. R., METHVEN, A. S., et al. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 6241-6246.
- SCHOOLMASTER, D. R. & STAGG, C. L. 2018. Resource competition model predicts zonation and increasing nutrient use efficiency along a wetland salinity gradient. *Ecology*, 99, 670-680.
- SELIG, M. J., ADNEY, W. S., HIMMEL, M. E. & DECKER, S. R. 2009. The impact of cell wall acetylation on corn stover hydrolysis by cellulolytic and xylanolytic enzymes. *Cellulose*, 16, 711-722.
- SELIG, M. J., KNOSHAUG, E. P., ADNEY, W. S., HIMMEL, M. E. & DECKER, S. R. 2008. Synergistic enhancement of cellobiohydrolase performance on pretreated corn stover by addition of xylanase and esterase activities. *Bioresource Technology*, 99, 4997-5005.
- SERVICE, R. F. 2009. GREEN ENERGY Another Biofuels Drawback: The Demand for Irrigation. *Science*, 326, 516-517.
- SEYLER, L. M., MCGUINNESS, L. M. & KERKHOF, L. J. 2014. Crenarchaeal heterotrophy in salt marsh sediments. *Isme Journal*, 8, 1534-1543.
- SHAHINPEI, A., AMOOZEGAR, M. A., FAZELI, S. A. S., SCHUMANN, P. & VENTOSA, A. 2014. *Salinispirillum marinum* gen. nov., sp nov., a haloalkaliphilic bacterium in the family 'Saccharospirillaceae'. *International Journal of Systematic and Evolutionary Microbiology*, 64, 3610-3615.
- SHANMUGAM, A. K. & NESVIZHSKII, A. I. 2015. Effective Leveraging of Targeted Search Spaces for Improving Peptide Identification in Tandem Mass Spectrometry Based Proteomics. *Journal of Proteome Research*, 14, 5169-5178.
- SHARMIN, F., WAKELIN, S., HUYGENS, F. & HARGREAVES, M. 2013. Firmicutes dominate the bacterial taxa within sugar-cane processing plants. *Scientific Reports*, 3, 7.
- SHEN, T. Y. & GNANAKARAN, S. 2009. The Stability of Cellulose: A Statistical Perspective from a Coarse-Grained Model of Hydrogen-Bond Networks. *Biophysical Journal*, 96, 3032-3040.
- SHOSEYOV, O., LEVY, I., SHANI, Z. & MANSFIELD, S. D. 2003. Modulation of wood fibers and paper by cellulose-binding domains. *Applications of Enzymes to Lignocellulosics*, 855, 116-131.
- SHOSEYOV, O., SHANI, Z. & LEVY, I. 2006. Carbohydrate binding modules: Biochemical properties and novel applications. *Microbiology and Molecular Biology Reviews*, 70, 283-+.
- SIMONES, M. P., CALADO, M. L., MADEIRA, M. & GAZARINI, L. C. 2011. Decomposition and nutrient release in halophytes of a Mediterranean salt marsh. *Aquatic Botany*, 94, 119-126.
- SINGH, G., VERMA, A. K. & KUMAR, V. 2015. Catalytic properties, functional attributes and industrial applications of beta-glucosidases. *3 Biotech*, 6, 14.
- SINGHVI, M. & GOKHALE, D. 2013. Biomass to biodegradable polymer (PLA). *Rsc Advances*, 3, 13558-13568.
- SINHA, R. & KHARE, S. K. 2014. Protective role of salt in catalysis and maintaining structure of halophilic proteins against denaturation. *Frontiers in Microbiology*, 5, 6.
- SOUSA, A. I., SANTOS, D. B., DA SILVA, E. F., SOUSA, L. P., CLEARY, D. F. R., SOARES, A. & LILLEBO, A. I. 2017. 'Blue Carbon' and Nutrient Stocks of Salt Marshes at a Temperate Coastal Lagoon (Ria de Aveiro, Portugal). *Scientific Reports*, 7, 11.
- SPRING, S., SCHEUNER, C., GOKER, M. & KLENK, H. P. 2015. A taxonomic framework for emerging groups of ecologically important marine gammaproteobacteria based on the reconstruction of evolutionary relationships using genome-scale data. *Frontiers in Microbiology*, 6, 17.
- STOECK, T., BASS, D., NEBEL, M., CHRISTEN, R., JONES, M. D. M., BREINER, H. W. & RICHARDS, T. A. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*, 19, 21-31.

References

- SUH, S. O., BLACKWELL, M., KURTZMAN, C. P. & LACHANCE, M. A. 2006. Phylogenetics of Saccharomycetales, the ascomycete yeasts. *Mycologia*, 98, 1006-1017.
- SUKSONG, W., JEHLEE, A., SINGKHALA, A., KONGJAN, P., PRASERTSAN, P., IMAI, T. & O-THONG, S. 2017. Thermophilic solid-state anaerobic digestion of solid waste residues from palm oil mill industry for biogas production. *Industrial Crops and Products*, 95, 502-511.
- SUN, F. F., HONG, J. P., HU, J. G., SADDLER, J. N., FANG, X., ZHANG, Z. Y. & SHEN, S. 2015a. Accessory enzymes influence cellulase hydrolysis of the model substrate and the realistic lignocellulosic biomass. *Enzyme and Microbial Technology*, 79-80, 42-48.
- SUN, F. F., ZHANG, X. L., ZHANG, Q. Q., LIU, F. H., ZHANG, J. P. & GONG, J. 2015b. Seagrass (*Zostera marina*) Colonization Promotes the Accumulation of Diazotrophic Bacteria and Alters the Relative Abundances of Specific Bacterial Lineages Involved in Benthic Carbon and Sulfur Cycling. *Applied and Environmental Microbiology*, 81, 6901-6914.
- SVARTSTROM, O., ALNEBERG, J., TERRAPON, N., LOMBARD, V., DE BRUIJN, I., MALMSTEN, J., DALIN, A. M., EL MULLER, E., SHAH, P., WILMES, P., HENRISSAT, B., ASPEBORG, H. & ANDERSSON, A. F. 2017. Ninety-nine de novo assembled genomes from the moose (*Alces alces*) rumen microbiome provide new insights into microbial plant biomass degradation. *ISME Journal*, 11, 2538-2551.
- SYVITSKI, J. P. M., VOROSMARTY, C. J., KETTNER, A. J. & GREEN, P. 2005. Impact of humans on the flux of terrestrial sediment to the global coastal ocean. *Science*, 308, 376-380.
- TABKA, M. G., HERPOEL-GIMBERT, I., MONOD, F., ASTHER, M. & SIGOILLOT, J. C. 2006. Enzymatic saccharification of wheat straw for bioethanol production by a combined cellulase xylanase and feruloyl esterase treatment. *Enzyme and Microbial Technology*, 39, 897-902.
- TAKASHIMA, S., NAKAMURA, A., HIDAKA, M., MASAKI, H. & UOZUMI, T. 1999. Molecular cloning and expression of the novel fungal beta-glucosidase genes from *Humicola grisea* and *Trichoderma reesei*. *Journal of Biochemistry*, 125, 728-736.
- TALAMANTES, D., BIABINI, N., DANG, H., ABDOUN, K. & BERLEMONT, R. 2016. Natural diversity of cellulases, xylanases, and chitinases in bacteria. *Biotechnology for Biofuels*, 9, 11.
- TALEBNIA, F., KARAKASHEV, D. & ANGELIDAKI, I. 2010. Production of bioethanol from wheat straw: An overview on pretreatment, hydrolysis and fermentation. *Bioresource Technology*, 101, 4744-4753.
- TANCA, A., PALOMBA, A., FRAUMENE, C., PAGNOZZI, D., MANGHINA, V., DELIGIOS, M., MUTH, T., RAPP, E., MARTENS, L., ADDIS, M. F. & UZZAU, S. 2016. The impact of sequence database choice on metaproteomic results in gut microbiota studies. *Microbiome*, 4, 13.
- TEDERSOO, L., BAHRAM, M., POLME, S., KOLJALG, U., YOROU, N. S., WIJESUNDERA, R., RUIZ, L. V., VASCO-PALACIOS, A. M., THU, P. Q., SUJIA, A., SMITH, M. E., SHARP, C., SALUVEER, E., SAIITA, A., ROSAS, M., RIIT, T., RATKOWSKY, D., PRITSCH, K., POLDMAA, K., PIEPENBRING, M., PHOSRI, C., PETERSON, M., PARTS, K., PARTEL, K., OTSING, E., NOUHRA, E., NJOUONKOU, A. L., NILSSON, R. H., MORGADO, L. N., MAYOR, J., MAY, T. W., MAJUAKIM, L., LODGE, D. J., LEE, S. S., LARSSON, K. H., KOHOUT, P., HOSAKA, K., HIIESALU, I., HENKEL, T. W., HAREND, H., GUO, L. D., GRESLEBIN, A., GRELET, G., GEML, J., GATES, G., DUNSTAN, W., DUNK, C., DRENKHAN, R., DEARNALEY, J., DE KESEL, A., DANG, T., CHEN, X., BUEGGER, F., BREARLEY, F. Q., BONITO, G., ANSLAN, S., ABELL, S. & ABARENKOV, K. 2014. Global diversity and geography of soil fungi. *Science*, 346, 1078-+.
- TEERI, T. T. 1997. Crystalline cellulose degradation: New insight into the function of cellobiohydrolases. *Trends in Biotechnology*, 15, 160-167.
- TEMUDO, M. F., MATO, T., KLEEREBEZEM, R. & VAN LOOSDRECHT, M. C. M. 2009. Xylose anaerobic conversion by open-mixed cultures. *Applied Microbiology and Biotechnology*, 82, 231-239.
- THOMAS, F., GIBLIN, A. E., CARDON, Z. G. & SIEVERT, S. M. 2014. Rhizosphere heterogeneity shapes abundance and activity of sulfur-oxidizing bacteria in vegetated salt marsh sediments. *Frontiers in Microbiology*, 5, 14.

References

- TOKUNAGA, H., ISHIBASHI, M., ARAKAWA, T. & TOKUNAGA, M. 2004. Highly efficient renaturation of beta-lactamase isolated from moderately halophilic bacteria. *Febs Letters*, 558, 7-12.
- TORZILLI, A. P., SIKAROODI, M., CHALKLEY, D. & GILLEVET, P. M. 2006. A comparison of fungal communities from four salt marsh plants using automated ribosomal intergenic spacer analysis (ARISA). *Mycologia*, 98, 690-698.
- TRILLA, G. G., DE MARCO, S., MARCOVECCHIO, J., VICARI, R. & KANDUS, P. 2010. Net Primary Productivity of *Spartina densiflora* Brong in an SW Atlantic Coastal Salt Marsh. *Estuaries and Coasts*, 33, 953-962.
- TRIVEDI, N., GUPTA, V., KUMAR, M., KUMARI, P., REDDY, C. R. K. & JHA, B. 2011. An alkali-halotolerant cellulase from *Bacillus flexus* Isolated from green seaweed *Ulva lactuca*. *Carbohydrate Polymers*, 83, 891-897.
- TRIVEDI, N., GUPTA, V., REDDY, C. R. K. & JHA, B. 2013. Detection of ionic liquid stable cellulase produced by the marine bacterium *Pseudoalteromonas* sp isolated from brown alga *Sargassum polycystum* C. Agardh. *Bioresource Technology*, 132, 313-319.
- TURNER, D. R., WHITFIELD, M. & DICKSON, A. G. 1981. The equilibrium speciation of dissolved components in fresh-water and seawater at 25-degrees-c and 1 atm pressure. *Geochimica Et Cosmochimica Acta*, 45, 855-881.
- UPDEGRAFF, D. M. 1969. Semimicro determination of cellulose in biological materials. *Analytical Biochemistry*, 32, 420-+.
- URBANOVA, M., SNAJDR, J. & BALDRIAN, P. 2015. Composition of fungal and bacterial communities in forest litter and soil is largely determined by dominant trees. *Soil Biology & Biochemistry*, 84, 53-64.
- VALDIVIA, M., GALAN, J. L., LAFFARGA, J. & RAMOS, J. L. J. M. B. 2016. Biofuels 2020: Biorefineries based on lignocellulosic materials. 9, 585-594.
- VALIELA, I., TEAL, J. M., ALLEN, S. D., VANETTEN, R., GOEHRINGER, D. & VOLKMANN, S. 1985. Decomposition in salt-marsh ecosystems - the phases and major factors affecting disappearance of above-ground organic-matter. *Journal of Experimental Marine Biology and Ecology*, 89, 29-54.
- VALIELA, I., WILSON, J., BUCHSBAUM, R., RIETSMA, C., BRYANT, D., FOREMAN, K. & TEAL, J. 1984. Importance of chemical-composition of salt-marsh litter on decay-rates and feeding by detritivores. *Bulletin of Marine Science*, 35, 261-269.
- VAN DER WEIJDE, T., KAMEI, C. L. A., TORRES, A. F., VERMERRIS, W., DOLSTRA, O., VISSER, R. G. F. & TRINDADE, L. M. 2013. The potential of C4 grasses for cellulosic biofuel production. *Frontiers in Plant Science*, 4, 18.
- VANHOLME, R., DEMEDTS, B., MORREEL, K., RALPH, J. & BOERJAN, W. 2010. Lignin Biosynthesis and Structure. *Plant Physiology*, 153, 895-905.
- VARNAI, A., SIIKA-AHO, M. & VIKARI, L. 2013. Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs. *Biotechnology for Biofuels*, 6, 11.
- VASILEIADIS, S., PUGLISI, E., ARENA, M., CAPPÀ, F., COCCONCELLI, P. S. & TREVISAN, M. 2012. Soil Bacterial Diversity Screening Using Single 16S rRNA Gene V Regions Coupled with Multi-Million Read Generating Sequencing Technologies. *Plos One*, 7, 11.
- VERA, F., GUTIERREZ, J. L. & RIBEIRO, P. D. 2009. Aerial and detritus production of the cordgrass *Spartina densiflora* in a southwestern Atlantic salt marsh. *Botany-Botanique*, 87, 482-491.
- VIGNERON, A., CRUAUD, P., ROUSSEL, E. G., PIGNET, P., CAPRAIS, J. C., CALLAC, N., CIOBANU, M. C., GODFROY, A., CRAGG, B. A., PARKES, J. R., VAN NOSTRAND, J. D., HE, Z. L., ZHOU, J. Z. & TOFFIN, L. 2014. Phylogenetic and Functional Diversity of Microbial Communities Associated with Subsurface Sediments of the Sonora Margin, Guaymas Basin. *Plos One*, 9, 11.
- VOGEL, C. & MARCOTTE, E. M. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, 13, 227-232.
- VOGEL, J. 2008. Unique aspects of the grass cell wall. *Current Opinion in Plant Biology*, 11, 301-307.

References

- VUONG, T. V. & WILSON, D. B. 2009. Processivity, Synergism, and Substrate Specificity of *Thermobifida fusca* Cel6B. *Applied and Environmental Microbiology*, 75, 6655-6661.
- WALIA, A., GULERIA, S., MEHTA, P., CHAUHAN, A. & PARKASH, J. 2017. Microbial xylanases and their industrial application in pulp and paper biobleaching: a review. *3 Biotech*, 7, 12.
- WALKER, A. K. & CAMPBELL, J. 2010. Marine fungal diversity: a comparison of natural and created salt marshes of the north-central Gulf of Mexico. *Mycologia*, 102, 513-521.
- WANG, B. J., JOHNSTON, E. M., LI, P. F., SHAIK, S., DAVIES, G. J., WALTON, P. H. & ROVIRA, C. 2018a. QM/MM Studies into the H₂O₂-Dependent Activity of Lytic Polysaccharide Monooxygenases: Evidence for the Formation of a Caged Hydroxyl Radical Intermediate. *Acs Catalysis*, 8, 1346-1351.
- WANG, D., KIM, D. H., SEO, N., YUN, E. J., AN, H. J., KIM, J. H. & KIM, K. H. 2016a. A Novel Glycoside Hydrolase Family 5 beta-1,3-1,6-Endoglucanase from *Saccharophagus degradans* 2-40(T) and Its Transglycosylase Activity. *Applied and Environmental Microbiology*, 82, 4340-4349.
- WANG, F., WU, J. & CHEN, S. 2018b. Preparation of gentiooligosaccharides using *Trichoderma viride* beta-glucosidase. *Food Chemistry*, 248, 340-345.
- WANG, G. H., TANG, M. X., WU, H. L., DAI, S. K., LI, T., CHEN, C. H., HE, H., FAN, J. W., XIANG, W. Z. & LI, X. 2015a. *Aliikangiella marina* gen. nov., sp. nov., a marine bacterium from the culture broth of *Picochlorum* sp 122, and proposal of *Kangiellaceae* fam. nov in the order *Oceanospirillales*. *International Journal of Systematic and Evolutionary Microbiology*, 65, 4488-4494.
- WANG, L. S., ZHANG, Y. Z. & GAO, P. J. 2008. A novel function for the cellulose binding module of cellobiohydrolase I. *Science in China Series C-Life Sciences*, 51, 620-629.
- WANG, M., CHEN, J. K. & LI, B. 2007. Characterization of bacterial community structure and diversity in rhizosphere soils of three plants in rapidly changing salt marshes using 16S rDNA. *Pedosphere*, 17, 545-556.
- WANG, Q., LUO, Y., HE, B., JIANG, L. S., LIU, J. X. & WANG, J. K. 2015b. Characterization of a Novel Xylanase Gene from Rumen Content of Hu Sheep. *Applied Biochemistry and Biotechnology*, 177, 1424-1436.
- WANG, T., YANG, H., KUBICKI, J. D. & HONG, M. 2016b. Cellulose Structural Polymorphism in Plant Primary Cell Walls Investigated by High-Field 2D Solid-State NMR Spectroscopy and Density Functional Theory Calculations. *Biomacromolecules*, 17, 2210-2222.
- WANG, X., WU, Y. B. & ZHOU, Y. 2017. Transglycosylation, a new role for multifunctional cellulase in overcoming product inhibition during the cellulose hydrolysis. *Bioengineered*, 8, 129-132.
- WANG, X. C., LIU, C., HUANG, L., BENGTSSON-PALME, J., CHEN, H. M., ZHANG, J. H., CAI, D. Y. & LI, J. Q. 2015c. ITS1: a DNA barcode better than ITS2 in eukaryotes? *Molecular Ecology Resources*, 15, 573-586.
- WANG, X. H. 2002. Tide-induced sediment resuspension and the bottom boundary layer in an idealized estuary with a muddy bed. *Journal of Physical Oceanography*, 32, 3113-3131.
- WARDEN, A. C., WILLIAMS, M., PEAT, T. S., SEABROOK, S. A., NEWMAN, J., DOJCHINOV, G. & HARITOS, V. S. 2015. Rational engineering of a mesohalophilic carbonic anhydrase to an extreme halotolerant biocatalyst. *Nature Communications*, 6, 10.
- WASKOM, M., BOTVINNIK, O., HOBSON, P., WARMENHOVEN, J., COLE, J., HALCHENKO, Y., VANDERPLAS, J., HOYER, S., VILLALBA, S. & QUINTERO, E. 2014. Seaborn: statistical data visualization. URL: <https://seaborn.pydata.org/>.
- WEBER, C. F., ZAK, D. R., HUNGATE, B. A., JACKSON, R. B., VILGALYS, R., EVANS, R. D., SCHADT, C. W., MEGONIGAL, J. P. & KUSKE, C. R. 2011. Responses of soil cellulolytic fungal communities to elevated atmospheric CO₂ are complex and variable across five ecosystems. *Environmental Microbiology*, 13, 2778-2793.
- WEINER, R. M., TAYLOR, L. E., HENRISSAT, B., HAUSER, L., LAND, M., COUTINHO, P. M., RANCUREL, C., SAUNDERS, E. H., LONGMIRE, A. G., ZHANG, H. T., BAYER, E. A., GILBERT, H. J., LARIMER, F., ZHULIN, I. B., EKBORG, N. A., LAMED, R., RICHARDSON, P. M., BOROVOK, I. &

References

- HUTCHESON, S. 2008. Complete genome sequence of the complex carbohydrate-degrading marine bacterium, *Saccharophagus degradans* strain 2-40(T). *Plos Genetics*, 4, 13.
- WEJSE, P. L., INGVORSEN, K. & MORTENSEN, K. K. 2003. Purification and characterisation of two extremely halotolerant xylanases from a novel halophilic bacterium. *Extremophiles*, 7, 423-431.
- WENG, J. K. & CHAPPLE, C. 2010. The origin and evolution of lignin biosynthesis. *New Phytologist*, 187, 273-285.
- WESTERENG, B., CANNELLA, D., AGGER, J. W., JORGENSEN, H., ANDERSEN, M. L., EIJSINK, V. G. H. & FELBY, C. 2015. Enzymatic cellulose oxidation is linked to lignin by long-range electron transfer. *Scientific Reports*, 5, 9.
- WHITE, D. A., TRAPANI, J. M., THIEN, L. B. & WEISS, T. E. 1978. PRODUCTIVITY AND DECOMPOSITION OF DOMINANT SALT-MARSH PLANTS IN LOUISIANA. *Ecology*, 59, 751-759.
- WILLIAMS, K. P. & KELLY, D. P. 2013. Proposal for a new class within the phylum Proteobacteria, Acidithiobacillia classis nov., with the type order Acidithiobacillales, and emended description of the class Gammaproteobacteria. *International Journal of Systematic and Evolutionary Microbiology*, 63, 2901-2906.
- WILSON, D. B. & KOSTYLEV, M. 2012. Cellulase processivity. *Biomass Conversion*. Springer, 908, 93-99.
- WILSON, J. O., BUCHSBAUM, R., VALIELA, I. & SWAIN, T. 1986. DECOMPOSITION IN SALT-MARSH ECOSYSTEMS - PHENOLIC DYNAMICS DURING DECAY OF LITTER OF SPARTINA-ALTERNIFLORA. *Marine Ecology Progress Series*, 29, 177-187.
- WONG, M. T., WANG, W. J., COUTURIER, M., RAZEQ, F. M., LOMBARD, V., LAPEBIE, P., EDWARDS, E. A., TERRAPON, N., HENRISSAT, B. & MASTER, E. R. 2017. Comparative Metagenomics of Cellulose- and Poplar Hydrolysate-Degrading Microcosms from Gut Microflora of the Canadian Beaver (*Castor canadensis*) and North American Moose (*Alces americanus*) after Long-Term Enrichment. *Frontiers in Microbiology*, 8, 14.
- WOO, H. L. & HAZEN, T. C. 2018. Enrichment of Bacteria From Eastern Mediterranean Sea Involved in Lignin Degradation via the Phenylacetyl-CoA Pathway. *Frontiers in Microbiology*, 9, 11.
- WRIGHT, A. J., EDWARDS, R. J., VAN DE PLASSCHE, O., BLAAUW, M., PARNELL, A. C., VAN DER BORG, K., DE JONG, A. F. M., ROE, H. M., SELBY, K. & BLACK, S. 2017. Reconstructing the accumulation history of a saltmarsh sediment core: Which age-depth model is best? *Quaternary Geochronology*, 39, 35-67.
- WU, M., MINTZ, M., WANG, M. & ARORA, S. 2009. Water Consumption in the Production of Ethanol and Petroleum Gasoline. *Environmental Management*, 44, 981-997.
- YAKIMOV, M. M., GIULIANO, L., GENTILE, G., CRISAFI, E., CHERNIKOVA, T. N., ABRAHAM, W. R., LUNSDORF, H., TIMMIS, K. N. & GOLYSHIN, P. N. 2003. *Oleispira antarctica* gen. nov., sp nov., a novel hydrocarbonoclastic marine bacterium isolated from Antarctic coastal sea water. *International Journal of Systematic and Evolutionary Microbiology*, 53, 779-785.
- YAMAGUCHI, A., ISOZAKI, K., NAKAMURA, M., TAKAYA, H. & WATANABE, T. 2016. Discovery of 12-mer peptides that bind to wood lignin. *Scientific Reports*, 6, 11.
- YIN, J., CHEN, J. C., WU, Q. & CHEN, G. Q. 2015a. Halophiles, coming stars for industrial biotechnology. *Biotechnology Advances*, 33, 1433-1442.
- YIN, Y. B., MAO, X. Z., YANG, J. C., CHEN, X., MAO, F. L. & XU, Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, 40, W445-W451.
- YIN, Y. R., ZHANG, F., HU, Q. W., XIAN, W. D., HOZZEIN, W. N., ZHOU, E. M., MING, H., NIE, G. X. & LI, W. J. 2015b. Heterologous expression and characterization of a novel halotolerant, thermostable, and alkali-stable GH6 endoglucanase from *Thermobifida* halotolerans. *Biotechnology Letters*, 37, 857-862.

References

- YU, T. T., LI, M., NIU, M. Y., FAN, X. B., LIANG, W. Y. & WANG, F. P. 2018. Difference of nitrogen-cycling microbes between shallow bay and deep-sea sediments in the South China Sea. *Applied Microbiology and Biotechnology*, 102, 447-459.
- YU, Z. Y., JAMEEL, H., CHANG, H. M., PHILIPS, R. & PARK, S. 2013. Quantification of bound and free enzymes during enzymatic hydrolysis and their reactivities on cellulose and lignocellulose. *Bioresource Technology*, 147, 369-377.
- YUAN, F., LENG, B. Y. & WANG, B. S. 2016. Progress in Studying Salt Secretion from the Salt Glands in Recrethahalophytes: How Do Plants Secrete Salt? *Frontiers in Plant Science*, 7, 12.
- ZAMOCKY, M., HOFBAUER, S., SCHAFFNER, I., GASSELHUBER, B., NICOLUSSI, A., SOUDI, M., PIRKER, K. F., FURAMULLER, P. G. & OBINGER, C. 2015. Independent evolution of four heme peroxidase superfamilies. *Archives of Biochemistry and Biophysics*, 574, 108-119.
- ZHANG, C., LIU, G. B., XUE, S. & WANG, G. L. 2016a. Soil bacterial community dynamics reflect changes in plant community and soil properties during the secondary succession of abandoned farmland in the Loess Plateau. *Soil Biology & Biochemistry*, 97, 40-49.
- ZHANG, G. L., BAI, J. H., JIA, J., WANG, X., WANG, W., ZHAO, Q. Q. & ZHANG, S. 2018a. Soil Organic Carbon Contents and Stocks in Coastal Salt Marshes with *Spartina alterniflora* Following an Invasion Chronosequence in the Yellow River Delta, China. *Chinese Geographical Science*, 28, 374-385.
- ZHANG, J., CHOI, Y. S., YOO, C. G., KIM, T. H., BROWN, R. C. & SHANKS, B. H. 2015. Cellulose-Hemicellulose and Cellulose-Lignin Interactions during Fast Pyrolysis. *Acs Sustainable Chemistry & Engineering*, 3, 293-301.
- ZHANG, J. H., SIIKA-AHO, M., TENKANEN, M. & VIKARI, L. 2011a. The role of acetyl xylan esterase in the solubilization of xylan and enzymatic hydrolysis of wheat straw and giant reed. *Biotechnology for Biofuels*, 4, 9.
- ZHANG, J. W., D'ROZARIO, A., DUAN, S. M., WANG, X. Y., LIANG, X. Q. & PAN, B. R. 2018b. Epidermal characters of *Tamarix* L. (Tamaricaceae) from Northwest China and their taxonomic and palaeogeographic implications. *Journal of Palaeogeography-English*, 7, 179-196.
- ZHANG, T., DATTA, S., EICHLER, J., IVANOVA, N., AXEN, S. D., KERFELD, C. A., CHEN, F., KYRPIDES, N., HUGENHOLTZ, P., CHENG, J. F., SALE, K. L., SIMMONS, B. & RUBIN, E. 2011b. Identification of a haloalkaliphilic and thermostable cellulase with improved ionic liquid tolerance. *Green Chemistry*, 13, 2083-2090.
- ZHANG, X. L., YANG, W. H. & BLASIAK, W. 2011c. Modeling Study of Woody Biomass: Interactions of Cellulose, Hemicellulose, and Lignin. *Energy & Fuels*, 25, 4786-4795.
- ZHANG, Y., SCHOCH, C. L., FOURNIER, J., CROUS, P. W., DE GRUYTER, J., WOUDEBERG, J. H. C., HIRAYAMA, K., TANAKA, K., POINTING, S. B., SPATAFORA, J. W. & HYDE, K. D. 2009. Multi-locus phylogeny of Pleosporales: a taxonomic, ecological and evolutionary re-evaluation. *Studies in Mycology*, 85-102.
- ZHANG, Y. J. & CREMER, P. S. 2006. Interactions between macromolecules and ions: the Hofmeister series. *Current Opinion in Chemical Biology*, 10, 658-663.
- ZHANG, Y. L., DU, J. Z., DING, X. P. & ZHANG, F. F. 2016b. Comparison study of sedimentary humic substances isolated from contrasting coastal marine environments by chemical and spectroscopic analysis. *Environmental Earth Sciences*, 75, 14.
- ZHANG, Y. L., DU, J. Z., ZHAO, X., WU, W. S., PENG, B. & ZHANG, J. 2014. A multi-proxy study of sedimentary humic substances in the salt marsh of the Changjiang Estuary, China. *Estuarine Coastal and Shelf Science*, 151, 295-301.
- ZHAO, D. L. & LI, Y. R. 2015. Climate Change and Sugarcane Production: Potential Impact and Mitigation Strategies. *International Journal of Agronomy*, 10.
- ZHAO, Y., DAMGAARD, A. & CHRISTENSEN, T. H. 2018. Bioethanol from corn stover - a review and technical assessment of alternative biotechnologies. *Progress in Energy and Combustion Science*, 67, 275-291.

References

- ZHONG, R. Q., CUI, D. T., PHILLIPS, D. R. & YE, Z. H. 2018. A Novel Rice Xylosyltransferase Catalyzes the Addition of 2-O-Xylosyl Side Chains onto the Xylan Backbone. *Plant and Cell Physiology*, 59, 554-565.
- ZHOU, M., GUO, P., WANG, T., GAO, L., YIN, H. J., CAI, C., GU, J. & LU, X. 2017. Metagenomic mining pectinolytic microbes and enzymes from an apple pomace-adapted compost microbial community. *Biotechnology for Biofuels*, 10, 15.
- ZHU, N., YANG, J. S., JI, L., LIU, J. W., YANG, Y. & YUAN, H. L. 2016. Metagenomic and metaproteomic analyses of a corn stover-adapted microbial consortium EMSD5 reveal its taxonomic and enzymatic basis for degrading lignocellulose. *Biotechnology for Biofuels*, 9, 23.