

**A Novel Metadata Analysis Approaches for
Analyzing and Understanding Wood-Decaying
Mechanisms Exhibited by Fungi**

**A Thesis Presented to the
Faculty of Graduate Studies of
Lakehead University**

By

Ayyappa Kumar Sista Kameshwar

**Submitted in partial fulfillment of requirements for the degree of
Doctor of Philosophy in Biotechnology**

© Ayyappa Kumar Sista Kameshwar

Abstract

Fuel has become an essential commodity in our day to day life. Increase in global population and decreasing fuel reserves have forced mankind to look for other fuel alternatives. Forest biomass serves as a potential renewable resource for substituting the conventional fossil-based fuels. In the last few decades, several chemical, physical and microbial based methods were being developed for the breakdown and conversion of lignocellulosic components to commercially valuable products including bioethanol and other platform chemicals. The separation of lignocellulosic biomass plays a significant role in conversion of lignocellulosic biomass to ethanol and other valuable products respectively. Naturally, lignocellulosic components are arranged in intricate networks leading to its high recalcitrance nature. Over years research groups around the world have isolated and characterized several lignocellulose degrading microorganisms. Naturally, fungi play a crucial role in maintaining the geo-carbon cycle by decaying all the dead organic matter on the earth's surface. Majority of the wood-decaying fungi are grouped under Basidiomycota division. Based on their decay patterns the basidiomycetous fungi were classified into white-rot, brown-rot and soft-rot fungi. Understanding these natural fungal decaying mechanisms will benefit the growing biofuel, biorefining and bioremediation industries.

Next generation sequencing techniques have significantly enhanced our present day's knowledge about various biological mechanisms. *Phanerochaete chrysosporium* was the first basidiomycetous fungi with complete annotated genome sequence, which has inspired the whole-genome sequencing studies of different wood-decaying fungi. Increasing whole genome sequencing studies have led to the enrichment of public repositories especially JGI-MycoCosm, 1000 fungal genome project, Hungate 1000 projects have played a significant role in supporting these sequencing projects. As of today, there are 443 published and completely annotated fungal genome sequences in the JGI-MycoCosm repository. Thus, the availability of whole annotated genome sequences has significantly helped in designing genome-wide transcriptomic studies for understanding the molecular mechanisms underlying the process wood-decay. A total of 11 genome-wide transcriptomic studies of *P. chrysosporium* were reported and are publicly available under the NCBI-GEO repository. However, transcriptomic studies give a snapshot of gene expression at a given growth conditions and time period. Thus, we have developed a robust and efficient metadata analysis approach for re-analyzing gene expression datasets of *P. chrysosporium* and *Postia placenta* for understanding the common significant gene expression patterns of this model white and brown rot fungi cultured on different growth substrates (simple customized synthetic media and natural plant biomass media).

We have reported a significant list of genes encoding for various lignocellulolytic enzymes significantly expressed among all the gene expression datasets. Based on the common gene expression patterns obtained from our analysis we have tentatively derived the molecular network of genes and enzymes employed during the breakdown and conversion of lignin, cellulose, hemicellulose components of plant biomass. For the first time, we have reported, and classified lignin degrading genes expressed during the ligninolytic conditions of *P. chrysosporium*. It is well-known that fungi experience a significant amount of stress during the process wood-decay. Plant secondary metabolites such as quinones, tannins, stilbenes,

flavonoids and other phenolic compounds exhibits a fungicidal activity. Thus, wood-decaying fungi wood decaying fungi would have developed an efficient detoxification and stress responsive mechanisms for sustaining this effect. Using the results obtained in our study we have reported and classified as phase-I and phase-II metabolic genes involved in detoxification and stress responsive mechanisms respectively. Compared to *P. chrysosporium*, *P. placenta* lacks several copies of genes encoding for cellulolytic, hemicellulolytic, ligninolytic and pectinolytic enzymes. However, our gene expression metadata analysis has reported that *P. placenta* is strongly dependent on Fenton's reaction for the degradation of lignocellulosic components. We have also observed that genes encoding for several hemicellulolytic enzymes were differentially expressed even during cellulolytic conditions. Based on the present metadata analysis we have also tentatively developed cellulose and hemicellulose metabolic mechanism.

In the last few decades, genome sequencing studies of wood-decaying fungi have been extensively reported. Presently, the JGI-MycCosm database resides 1165 whole genome sequences of fungi, out of which about 443 fungal genome sequences are published. It is hard to choose a particular fungus specifically for the degradation of plant biomass. Thus, we have developed an efficient metadata analysis pipeline for comparing and understanding the genome-wide annotations of fungi. The metadata analysis pipeline reported in our study can be used for selecting a wood-decaying fungus for the *In vitro* degradation of studies. We have compared the genome-wide annotations of about 42 wood-decaying basidiomycetous fungi (white-rot, brown-rot and soft-rot fungi) and reported a tentative comparison method explaining the total cellulolytic, hemicellulolytic, ligninolytic and pectinolytic abilities. Similarly, we have also specifically compared and analyzed the genome wide annotations of anaerobic fungi belonging to *Neocallimastigomycota* division fungi. This study has reported the complete genetic makeup of these peculiar fungi and their carbohydrate degrading abilities (plant cell wall carbohydrates), as they completely lack lignin degrading enzymes. We have also delineated and compared the genes coding for structural and functional components of cellulosomes and hydrogenosomes.

We have also performed an extensive homology modeling and protein docking study of white-rot, brown-rot and soft-rot fungal laccases protein sequences using 6 different types of lignin model compounds. This study has revealed the structural and functional variations of white, brown and soft rot fungal laccases. Results obtained in this study reported that white and brown rot fungal laccases reported higher catalytic efficiencies compared to that of soft rot fungal laccases. This study also reported that soft rot fungal laccases exhibited small but significant variations in its structural and physicochemical properties. However, further molecular dynamic simulation and high throughput proteomic studies must be performed to understand the structural and functional properties of these laccases.

The metadata analysis work frame reported in this thesis can be extended to understand the natural wood-degrading mechanisms of various microorganisms (e.g. fungi, bacteria). The highly reactive significant list of proteins obtained in this study can be used *In vitro* for developing highly resistant enzyme mixes used for the breakdown and conversion of various organic compounds including plant biomass. The above reported strategies also can be used as a preliminary analysis for designing large scale experiments.

Acknowledgement

PhD is just not a degree, it's a major destination of my life. I sincerely thank everyone who made this dream into a reality. Thank you very much for fueling me with your immense motivation and encouragement.

Professor Dr. Wensheng Qin, I am grateful to have you as my supervisor. Your endless support, encouragement, care and guidance throughout my PhD term are immeasurable. Your ideology and perception of academic research is distinctive and has become a strong motive for me to achieve the hardest goals. Your thoughts and ideas sculpted my research in a better shape. Your support whether it may be personal or professional, you have always been there with me. During this journey I learned many things including scientific knowledge, research experience and administration. Thank you for extending your hand and supporting me in hardest twists and turns of my PhD journey.

Dr. Kam Leung and Dr. Apichart Linhananta - I am delighted to have you both in my doctoral committee. The suggestions and encouragement you have provided during my research proposal was instrumental. Your valuable suggestions were always vital. You have always been there at important turning points of my PhD.

Dr. Luiz Pereira Ramos – You have been very encouraging and supportive during my visit to your lab in Brazil. I thank your family and lab members for their love and support during my stay.

Lakehead University, Department of Biology- I sincerely thank **Dr. Todd Randall, Dr. Wely Floriano, Dr. Brenda Magajna, Dr. Susanne Walford** and **Micheal Moore** for their continuous support and encouragement.

Dr. Md. Shafiqur Rahman, Dr. Haipeng Guo, Dr. Yanwen Wu, Iman Almuharef, Hemkanta Sharma, Xuantong Chen (Tong), Richard Barber, Chris ChonLong Chio (Chris). I thank everyone of you for making the lab livelier and a better place to work.

Ontario Trillium Foundation – I sincerely thank Ontario Trillium Foundation for providing me with the prestigious Ontario Trillium Scholarship (OTS) to pursue my PhD in Lakehead University.

Mitacs® – I thank Mitacs corporation for awarding me with the “Globalink Researcher Award” to fund my visit and stay to Federal University of Parana for pursuing part of my research in Curitiba, Brazil. I am honoured to have such a wonderful opportunity to take my research abroad and have international insight.

Dr. Sudeep Rakshit, Dr. Mitchell Albert, Dr. Alla Reznik – Each one of you played a distinct role in shaping up my PhD. Especially **Dr. Rakshit** I am grateful for your teachings and thank you very much for sharing your ideology about science and life.

Suresh Kumar Konda, Sai Swaroop Dalli - Thank you will be a short word for all the support, encouragement, concern, motivation and guidance you have given me in the last decade. Our journey started as friends and transformed into family now.

Venkatesh Manikandan – Venky without you my PhD would have been boring and difficult. Thank you very much for introducing me an active and healthy life style. I always relish our coffee hours, discussions and physical training sessions. Our crazy and hilarious talks always replenished me during stressful times.

Suresh Kumar Konda, Sai Swaroop Dalli, Venkatesh Manikandan, Mohammed Ibrahim, Mohan Konduri, Balaji Venkateshgowda – It would have been a boring life without you guys. Thank you for giving me a reason to come out of my lab and I cherish all those fun-filled activities we had.

Sushma Karra – I couldn't have accomplished my research study without your love, concern and encouragement. Your strong belief in me has always rejuvenated and encouraged me during the hardest turns of my life. Thank you very much for believing in me and supporting me in my every step. I am thankful to my destiny for bringing you into my life. Finally, I will totally agree with the words of Groucho Marx "*Behind every successful man is a woman, behind her is his wife*".

My father **Sista Bala Kameshwar Rao** and mother **Sista Devi** for your endless love, nurture and affection. I will never forget all the selfless efforts and sacrifices you have made for supporting my life and career. Your faith and confidence in me always boost me to achieve higher goals every time.

I sincerely thank all my teachers, lecturers and professors for sharing their valuable knowledge and for their support and encouragement over the years!

I thank all my friends in India, Canada and Brazil for their encouragement, love and support.

I am grateful to God for this life.

DEDICATION

To my supervisor Professor Wensheng Qin

To my parents Sista Bala Kameshwar Rao, Sista Devi

Table of contents

ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
DEDICATION	vi
Table of contents.....	vii
List of Figures.....	xiii
List of Tables	xx
INTRODUCTION.....	1
I-1. Microorganisms as possible degraders of lignocellulose	2
I-2. Genome sequencing studies for the identification and characterization of Microorganisms	3
I-3. CAZy and FOLy Databases for Genome wide Enzyme Analysis.....	5
I-4. Whole genome sequence studies of lignocellulose degrading bacteria.....	7
I-5. Genome and transcriptome studies of lignocellulose degrading fungi.....	8
I-6. Whole microbiome studies of lignin degrading higher organisms	13
I-7. Summary and Outlook.....	14
Rationale behind this study	15
Novelty of current research	15
References.....	18
Chapter-1. Fungal Enzymes Involved in Lignocellulose Degradation.....	22
Lignocellulolytic Enzymes	23
Lignin Oxidizing Enzymes (LO).....	25
Laccases (EC 1.10.3.2, benzenediol: oxygen Oxidoreductase)	26
Peroxidases (EC:1.11.1.x).....	29
Lignin Peroxidases (E.C. 1.11.1.14).....	29
Manganese Peroxidases (EC 1.11.1.13)	32
Versatile Peroxidases	34
Cellobiose dehydrogenase.....	37
Lignin Degrading Auxiliary Enzymes (LDA).....	39

Aryl alcohol Oxidase.....	40
Vanillyl alcohol Oxidase	42
Glyoxal Oxidase	44
Pyranose Oxidase	45
Galactose Oxidase	46
Glucose Oxidase.....	47
Benzoquinone reductase	48
Plant cell wall Deacetylating Carbohydrate esterases.....	50
Acetyl xylan esterases	50
Pectin Carbohydrate esterases	52
Lignin-Carbohydrate de-esterases	53
References.....	55
Chapter-2. Metadata Analysis of Phanerochaete chrysosporium Gene Expression Data Identified Common CAZymes Encoding Gene Expression Profiles Involved in Cellulose and Hemicellulose Degradation.....	67
2.1. Abstract.....	67
2.2. Introduction.....	68
2.3. Data Analysis Methodology	70
2.3.1. Data collection.....	70
2.3.2. Data Analysis.....	70
2.3.3. Overview of Data analysis	72
2.4. Results.....	73
2.4.1. Distribution of CAZymes in P. chrysosporium Genome	73
2.4.2. Expression of cellulose degrading glycoside hydrolases	77
2.4.3. Expression of hemicellulose degrading CAZymes	78
2.4.4. Glycoside hydrolases involved in fungal cell wall synthesis	79
2.4.5. Expression of glycosyl transferases encoding genes.....	80
2.4.6. Gene expression of polysaccharide lyases	81
2.4.7. Major Facilitator Superfamily encoding genes.....	82
2.4.8. Genes encoding for carbohydrate metabolism.....	82

2.4.9. Genes encoding for carbohydrate binding modules (CBM).....	83
2.5. Discussion	83
References.....	89

Chapter 3. Gene Expression Metadata Analysis Reveals Molecular Mechanisms Employed by Phanerochaete chrysosporium During Lignin Degradation and Detoxification of Plant Extractives

3.1. Abstract.....	94
3.2. Introduction.....	95
3.3. Data retrieval and Methodology	99
3.3.1. Data retrieval.....	99
3.3.2. Data Analysis.....	100
3.3.3. Summary of data analysis	101
3.4. Results.....	102
3.4.1. Lignin oxidizing and auxiliary enzymes	102
3.4.2. Detoxification and stress responsive genes	106
3.4.3. Phase I Metabolic Enzymes.....	108
3.4.3. Phase II Metabolic Enzymes.....	109
3.4.4. Effect of growth substrate and incubation period.....	110
3.5. Discussion	113
References.....	121

Chapter 4. Analyzing Phanerochaete chrysosporium Gene Expression Patterns Controlling the Molecular Fate of Lignocellulose Degrading Enzymes

4.1. Abstract.....	125
4.2. Introduction.....	126
4.3. Methods	129
4.3.1. Data Retrieval	129
4.3.2. Data analysis	130
4.4. Results.....	131
4.4.1. RNA Processing and Modification (KOG ID: A)	133
4.4.2. Chromatin structure and dynamics (KOG ID: B).....	135

4.4.3. Translation ribosomal structure and biogenesis (KOG ID: J)	138
4.4.4. Transcription (KOG ID: K)	140
4.4.5. Replication, repair and recombination (KOG ID: L)	143
4.5. Discussion	146
References.....	151
Chapter 5. Molecular Networks of <i>Postia placenta</i> Involved in Degradation of Lignocellulosic Biomass Revealed from Metadata Analysis of Open Access Gene Expression Data.....	
5.1. Abstract.....	155
5.2. Introduction.....	156
5.3. Data retrieval and Analysis.....	159
5.3.1. Data retrieval.....	159
5.3.2 Data analysis methodology.....	160
5.4. Results.....	162
5.4.1. Genes encoding for Carbohydrate Active Enzymes (CAZymes) and Metabolism	163
5.4.2. Enzymes coding Fenton’s Reaction.....	166
5.4.3. Lignin Degrading and Detoxifying Enzyme systems	167
5.4.4. Genes encoding for Information, Storage and Processing processes	169
5.5. Discussion	171
5.6. Conclusion	179
References.....	179
Chapter 6. Comparative Study of Genome-Wide Plant Biomass Degrading CAZymes in White Rot, Brown Rot and Soft Rot Fungi.....	
6.1. Abstract.....	183
6.2. Introduction.....	184
6.2. Data retrieval and Analysis.....	186
6.2.1 Data Retrieval	186
6.2.2. Data Analysis.....	187
6.3. Results and Discussions	188
6.3.1. Distribution of CAZymes among white rot, brown rot and soft rot fungi	190
6.3.2. Lignin degrading CAZymes.....	192

6.3.3. Cellulose degrading CAZymes.....	193
6.3.4. Hemicellulose degrading CAZymes	194
6.3.5. Pectin degrading CAZymes.....	195
6.3.6. Total lignocellulolytic abilities of selected fungi	196
6.4. Conclusions.....	197
References.....	198
Chapter 7. Genome Wide Analysis Reveals the Extrinsic Cellulolytic and Biohydrogen Generating Abilities of Neocallimastigomycota Fungi.....	202
7.1. Abstract.....	202
7.2. Introduction.....	203
7.3. Data retrieval and Analysis.....	205
7.3.1. Data retrieval.....	205
7.3.2. Data Analysis.....	206
7.4. Results and Discussions	207
7.5. Conclusion	218
References.....	219
Chapter 8. Comparative Modeling and Molecular Docking Analysis of White, Brown and Soft Rot Fungal Laccases Using Lignin Model Compounds for Understanding the Structural and Functional Properties of Laccases	222
8.1. Abstract.....	222
8.2. Introduction.....	223
8.3. Materials and Methods	224
8.3.1. Selection and retrieval of laccase protein sequences.....	224
8.3.2. Phylogenetic analysis	225
8.3.3. Physico-chemical Properties of selected laccases	225
8.3.4. Structural and functional properties of laccases	226
8.3.5. Initial protein model generation and refinement.....	226
8.3.6 Preparation of ligands (Lignin model compounds.....	227
8.3.7 Protein docking of refined models.....	227
8.4. Results.....	228
8.4.1. Sequence retrieval, Analysis and physicochemical properties	228

8.4.2. Homology Modeling and Model validation	232
8.4.3. Molecular Docking of modelled laccases with lignin model compounds	235
8.5. Discussion	240
References.....	247
Chapter 9.	
9.1. Summary of conclusions	251
9.2. Recommendation for future work	254
PUBLICATIONLIST.....	255

List of Figures

- Figure I-1:** Schematic representation of plant cell wall components having high commercial importance a) Cellulose b) Hemicellulose c) Cellobiose d) xylose e) Lignin 1
- Figure I-2:** Represents the potential groups of bacteria and fungi, capable of degrading plant cell wall components (cellulose, hemicellulose and lignin)..... 3
- Figure I.3:** Illustrates the methods used for the characterization of lignocellulose degrading bacteria and fungi using traditional and next generation sequencing methods 5
- Figure I-4:** (A) Phylogeny and main lifestyles of Agaricomycetes with a published genome sequence. The major orders Agaricales and Polyporales are indicated. The majority of these species are wood decayers and can be further classified as either white rot fungi (which degrade all components of the plant cell wall) or brown rot fungi (which modify lignin, but do not break it down to a large extent). *Schizophyllum commune*, *Jaapia argillacea* and *Botryobasidium botryosum* are also wood decayers, but cannot be easily classified as either white or brown rot fungi. *Coprinopsis cinerea*, *Agaricus bisporus* and *Volvariella volvacea* are saprotrophs growing on non-woody substrates. The ectomycorrhizal fungus *Laccaria bicolor* and the endophyte *Piriformospora indica* both form interactions with plant roots. Species with an asterisk (*) are predominantly plant pathogens. The genomes of *G. marginata*, *P. ostreatus*, *J. argillacea* and *B. botryosum* have been submitted for publication. (B) Number of predicted genes for each genome. Each bar lines up with a species from the tree in (A). The total number of genes varies per genome, but the number of genes with at least one PFAM domain is more constant. Genes without a PFAM domain outnumber those with a PFAM domain, showing that much remains to be learned about these organisms. Reprinted with performance from ref [67], Copyright © 2014 Elsevier. 11
- Figure I-5:** The number of available genome sequences of the Agaricomycetes has dramatically increased over the past decade. For each year, the total number of published and publicly available genomes are given. The number of publicly available genomes (red line) is higher than the number of published genomes (blue line), since in the case of genomes sequenced by the Joint Genome Institute those genomes are publicly available in MycoCosm well before being published. Reprinted with performance from ref [67], Copyright © 2014 Elsevier. 12
- Figure 1.1:** Schematic representation of different lignin oxidizing enzymes namely, laccases (PDB ID: 3FPX), lignin peroxidase (LiP) (PDB ID: 1B85) [34], manganese peroxidase (MnP) (PDB ID: 1YYD) [35], versatile peroxidase (VP) (PDB ID: 3FKG) , cellobiose dehydrogenase (PDB ID: 1KDG) [36]. All the enzyme structures were obtained from the PDB RCSB repository.. 25
- Figure 1.2:** Catalytic cycle of laccase. Reprinted with permission from ref [64], Copyright © 2008 Springer Science+Business Media B.V. 29
- Fig. 1.3:** Chemical structures and reactions discussed in the text. (a) The principal β -O-4 structure of lignin and pathway for its C_{α} - C_{β} cleavage by LiP. (b) A phenylcoumaran lignin structure. (c) A resinol lignin structure. (d) LiP-catalyzed oxidation of the fungal metabolite veratryl alcohol. Gymnosperms contain lignin's in which most subunits have $R_1 = OCH_3$ and $R_2 = H$. Angiosperm lignin's also contain these structures but have in addition subunits in

which $R_1 = \text{OCH}_3$ and $R_2 = \text{OCH}_3$. Grass lignin's contain both types of structures but have in addition some subunits in which $R_1 = \text{H}$ and $R_2 = \text{H}$. These nonmethoxylated lignin structures are more difficult to oxidize than those that contain one or two methoxyl groups. In the predominating nonphenolic structures of lignin, $R_3 = \text{lignin}$, whereas $R_3 = \text{H}$ in the minor phenolic structures. Reprinted with permission from ref [74], Copyright © 2008, Elsevier..31

Figure 1.4: Catalytic cycle of manganese peroxidase. Reprinted with permission from ref [86], Copyright © 2002, Elsevier.33

Figure 1.5: Schemes of VP catalytic cycle. (a) Basic cycle described by [102] including two-electron oxidation of the resting peroxidase (VP, containing Fe^{3+}) by hydroperoxide to yield compound I (C-I, containing Fe^{4+} -oxo and porphyrin cation radical), whose reduction in two one-electron reactions results in the intermediate compound II (C-II, containing Fe^{4+} -oxo after porphyrin reduction) and then the resting form of the enzyme. As shown in the cycle, VP can oxidize both: (i) aromatic substrates (AH) to the corresponding radicals ($\text{A}\cdot$); and (ii) Mn^{2+} to Mn^{3+} , the latter acting as a diffusible oxidizer. (b) Extended cycle including also compounds I_B (C- I_B , containing Fe^{4+} -oxo and Trp radical) and II_B (C- II_B , containing Fe^{3+} and Trp radical) involved in oxidation of veratryl alcohol (VA) and other high redox potential aromatic compounds (C- I_B and C- II_B are in equilibrium with C- I_A and C- II_A respectively, which correspond to C-I and C-II in (a) (other low redox potential aromatic compounds are probably oxidized by both the A and B forms but they are not included for simplicity). The active Trp in C- I_B and C- II_B would be Trp164 (the part of the cycle showing aromatic substrate oxidation would be also applicable to LiP, being Trp171 the active amino acid). Reprinted with permission from ref [67], Copyright © 2005, Elsevier.35

Figure 1.6: Reactions of cellobiose dehydrogenase based on [118]. 'Fe' represents the heme iron, 'A' represents the one-electron acceptor. Reprinted with permission from ref [119], Copyright ©2008, Oxford University Press.38

Figure 1.7: Schematic representation of lignin degrading auxiliary enzymes namely, aryl alcohol oxidase (PDB ID: 3FIM) [145], vanillyl alcohol oxidase (PDB ID:1W1J) [146], glucose oxidase (PDB ID: 1CF3) [147], galactose oxidase (PDB ID: 2WQ8) [148], pyranose oxidase (PDB ID: 4MIF) [149]. benzoquinone reductase (PDB ID: 4LA4). All the above enzyme structures were obtained from PDB RCSB repository.40

Figure 1.8: Chemical structure of various substrates of *Pleurotus* AAO (I, benzyl alcohol; II, panisyl alcohol; III, cinnamyl alcohol; IV, 2-naphthalenemethanol; and V, 2,4-hexadien-1-ol) and relative activity estimated as O_2 consumption [155]. Scheme for H_2O_2 production by anisaldehyde redox-cycling involving extracellular AAO and intracellular AAD. Reprinted with permission from ref [162], Copyright © 2000, Elsevier.42

Figure 1.9: The reaction mechanism for the oxidation of 4-(methoxymethyl) phenol. In the first step, the substrate is oxidised via a direct hydride transfer from the substrate $\text{C}\alpha$ atom to the N5 of flavin. The reduced cofactor is then re-oxidised by molecular oxygen with the production of a hydrogen peroxide molecule. In the next step, the p-quinonemethoxymethide intermediate is hydroxylated by a water molecule, possibly activated by Asp170. The resulting 4-hydroxy-benzaldehyde and methanol products are released. Reprinted with permission from ref [166], Copyright © 1997, Elsevier.44

Figure 1.10: Reaction mechanism of glucose oxidase (GOD) [188]. Reprinted with permission from ref [185], Copyright © 2009, Elsevier. 48

Figure 2.1: Schematic representation of CAZymes distribution in *Phanerochaete chrysosporium* genome, GH (glycoside hydrolase), GT (glycosyl transferases), AA (Auxiliary activities), CBM (Carbohydrate binding modules), CE (Carbohydrate esterases) and PL (Polysaccharide lyases); the numbers represented on top of each box represents the number of genes encoding for that particular class of enzymes respectively. 74

Figure 2.2: Three way and four-way Venn diagrams showing the common differentially expressed genes of *P. chrysosporium* involved in cellulose and hemicellulose degradation;(A) Differentially expressed genes and (B) Differentially expressed CAZymes, resulted from the datasets GSE14734, GSE14735 and GSE27941; (C) Differentially expressed genes and (D) Differentially expressed CAZymes resulted from the datasets GSE54542, GSE52922, GSE69008 and GSE69461. 76

Figure 2.3: Venn diagram showing the cellulose degrading glycoside hydrolases (left), oligosaccharide metabolizing (center) hemicellulose degrading (right) and other CAZY enzyme classes. 85

Figure 2.4: Tentative network of genes coding for *P. chrysosporium* cellulose degrading enzymes and cellulose degradation mechanism; O (CAZymes involved in oligosaccharide degrading), C (cellulolytic CAZymes), CDH (cellobiose dehydrogenase encoding CAZymes) and LPMO (CAZymes coding for lytic polysaccharide monoxygenases). 86

Figure 2.5: Tentative network of genes coding for *P. chrysosporium* hemicellulose degrading enzymes and hemicellulose degradation mechanism. 88

Figure 3.1: Hierarchical clusters showing the differentially expressed genes obtained, 1st and 2nd columns lists the fold change expression values up regulated in cellulose and glucose (A) GSE14734, Carbon and nitrogen limited in (B) GSE14735 and 3rd, 4th and 5th columns in A and B lists the log transformed gene expression values of BMA, cellulose, glucose, carbon-limited, nitrogen-limited and replete growth mediums respectively 112

Figure 3.2: Venn diagrams of differentially expressed genes obtained from the gene expression datasets A) GSE14734 B) GSE14735 C) GSE69008 (where A, B, C represents high lignin-low glucose, high glucose-low lignin, average lignin-average glucose, 10, 20, 30 represents number of days) D) GSE52922 (where P717, 82 and 64 represents parent and transgenic lines of Poplar trichocarpa species respectively) E) GSE27941, GSE52922, GSE54542, GSE69008, GSE69461 (*Picea glauca* species) 113

Figure 3.3: Shows the primary enzymatic reactions of *P. chrysosporium* involved in lignin Degradation LiP (lignin peroxidase), MnP (Manganese peroxidase), VP (Versatile peroxidase), VA (Veratryl Alcohol), PAL (Phenylalanine ammonia lyase), MCO (Multicopper oxidase)..... 115

Figure 3.4: Bird's eye view of tentative and proposed general molecular mechanisms and pathways involved in lignin degradation..... 116

Figure 3.5: Tentative network of *P. chrysosporium* genes and enzymes involved in lignin degradation mechanisms.....118

Figure 3.6: Tentative network of *P. chrysosporium* genes and enzymes involved in detoxification mechanisms involving phase-I, phase-II and stress responsive pathways.....121

Figure 4.1: Pie diagram showing the distribution of gene models A- RNA processing and modification, B-chromatin structure and dynamics, J- translation, ribosomal structure and biogenesis, K- transcription and L- replication, recombination and repair. And heatmaps showing the gene models occurring more than 3 copies in the *P. chrysosporium* genome...127

Figure 4.2: Violin plots for the selected datasets GSE14734, GSE14735, GSE27941, GSE52922, GSE54542 (normalized and baselined) and GSE69461 (only baselined) briefly showing the distribution of samples.....132

Figure 4.3: Profile plots (GSE14734 and GSE14735) and volcano plots (GSE27941, GSE54542 and GSE69461) of the significant and differentially expressed genes with fold change cut-off >2.0 among the selected datasets.....133

Figure 4.4: List of commonly expressed and statistically significant genes among the selected gene expression datasets encoding for RNA processing and Modification group (KOG functional ID: A).....135

Figure 4.5: List of commonly expressed and statistically significant genes among various datasets encoding for chromatin structure and dynamics group (KOG functional ID: B)....137

Figure 4.6: List of commonly expressed and statistically significant genes among various datasets encoding for translation, ribosomal structure and biogenesis group (KOG functional ID: J).....140

Figure 4.7: List of commonly expressed and statistically significant genes among various datasets encoding for transcription group (KOG functional ID: K)143

Figure 4.8: List of commonly expressed and statistically significant genes (A), (B) among the selected gene expression datasets encoding for Replication, repair and recombination group (KOG functional ID: L)145

Figure 4.9: Heat map (Color scales) listing the important groups of genes encoding for A) transcription factors B) translation ribosomal structure and biogenesis C) chromatin structure and dynamics D) DNA repair mechanisms in genome of *P. chrysosporium*.....148

Figure 5.1: Pictorial representation of CAZymes distribution (GH-glycoside hydrolases, GT-glycosyl transferases, AA-auxiliary activity, CBM-carbohydrate binding domains, CE-carbohydrate esterases, EXPN-expansin like proteins, PL-polysaccharide lyases) in *P. placenta* MAD-698- Rv1.0 (A) and the top number denotes for the number of genes coding for particular class of enzymes and each bar internally shows different sub-classes of enzymes and the number of genes encoding for the corresponding enzymes, (B) Comparison of genome wide CAZymes between *P. placenta* and *P. chrysosporium*.....157

Figure 5.2: Customized step by step workflow used for the metadata analysis of <i>Postia</i> placenta gene expression datasets.....	162
Figure 5.3: Profile plot (GSE12540, GSE69004 and GSE84529) and volcano plots (GSE29656) of the significant and differentially expressed genes among the conditions.....	163
Figure 5.4: Four-way and three-way Venn diagrams showing the commonly expressed statistically significant CAZymes among the gene expression datasets A) glucose-cellulose, C B D x A 165 P a g e glucose-BMA, BMA-cellulose and GSE29656 B) high lignin and low glucose (A) high glucose and low lignin (B) average lignin and average glucose at incubation periods of 10-days, 20- days and 30-days, C) 15mm-20mm vs 0mm-5mm, 30mm-35mm vs 0mm-5mm and 30mm- 35mm vs 15mm-20mm D) GSE29656, GSE69004, GSE12540 and GSE84529 datasets.....	164
Figure 5.5: Four-way and three-way Venn diagrams showing the commonly expressed statistically significant enzymes among the gene expression datasets A) glucose-cellulose, glucose-BMA, BMA-cellulose and GSE29656 B) high lignin and low glucose (A) high glucose and low lignin (B) average lignin and average glucose at incubation periods of 10-days, 20-days and 30-days, C) 15mm-20mm vs 0mm-5mm, 30mm-35mm vs 0mm-5mm and 30mm-35mm vs 15mm-20mm D) GSE29656, GSE69004, GSE12540 and GSE84529 datasets.....	168
Figure 5.6: Four-way Venn diagrams showing the commonly expressed statistically significant protein-Ids among the gene expression datasets A) CAZymes among all the datasets B) All InterPro-IDs among all the datasets.....	169
Figure 5.7: Four-way Venn diagrams showing the commonly expressed statistically significant genes encoding for KOG groups A (RNA Processing and modification), B (Chromatin structure and dynamics), J (Translation, ribosomal structure and biogenesis), K (Transcription) L (Replication, recombination and repair) and heatmap showing total list of significant information storage and processing groups among the datasets.....	171
Figure 5.8: Tentative representation of Fenton's reaction system observed in <i>P. placenta</i> mainly includes H ₂ O ₂ generating and Iron reduction-homeostasis enzyme systems found to be commonly observed among the gene expression datasets.....	173
Figure 5.9: Tentative network of genes involved in coding for enzymes involved during cellulose and hemicellulose degradation.....	177
Figure 6.1: Tentative network of CAZymes involved in depolymerization of lignin, cellulose, pectin and hemicellulose observed in selected popular white rot, brown rot and soft rot fungi	186
Figure 6.2: Heatmaps showing the genome wide distribution of A) Metabolism (C = energy production and conversion, G = carbohydrate transport and metabolism and Q = secondary metabolites biosynthesis, transport and catabolism) and B) number of cytochrome P450 encoding genes in selected popular white rot, brown rot and soft rot fungi	190
Figure 6.3: Heatmap showing the genome wide distribution of CAZymes from selected popular white rot, brown rot and soft rot fungi.....	191

Figure 6.4: Heatmaps showing the genome wide distribution of total ligninolytic, cellulolytic, hemicellulolytic and pectinolytic CAZymes in selected popular white rot, brown rot and soft rot fungi and tentative overall.....197

Figure 7.1: Workflow pipeline implemented for analyzing and comparing the genomes of Neocallimastigomycota division fungi..... 207

Figure 7.2: (A) Hierarchical delineation of Neocallimastigomycota division and well-studied fungi with available annotated genomes, (B) phylogenetic relationship of the Neocallimastigomycota fungi based on the conserved ITS (Inter Transcribed Spacer regions) and other morphological characteristics, where the legend shows the phylogenetic relatedness among the selected fungi.....209

Figure 7.3: A) Genome wide distribution of KOG classified proteins under CSP (cellular signaling and processing), ISP (Information storage and processing), M (metabolism) and PC (poorly characterized) B) total number of KOG classified proteins C) unique genes D) proteins with multiple gene copies (>2 gene copies).....211

Figure 7.4: Genome wide distributions of A) CAZymes, B) cellulolytic, hemicellulolytic and pectinolytic activities (GH- Glycoside hydrolases, CE- Carbohydrate esterases and PL- polysaccharide lyases) where Anasp1 (Anaeromyces robustus), Neosp1 (Neocallimastix californiae), Orpsp1 (Orpinomyces sp), PirE2 (Piromyces sp. E2) and Pirfi3 (Piromyces finnis) respectively.....213

Figure 7.5: Genome wide distributions of A) KOG (Eukaryotic orthologous groups), B) SM (secondary metabolite) clusters C) distribution of proteins among KEGG classified pathway groups and D) clan-based classification of proteolytic enzymes..... 214

Figure 7.6: A) Carbohydrate binding modules (CBM) and the corresponding carbohydrate interacting residues, B) total tentative number of genes encoding for cellulose, xylan, alphan-glucans, plant cell walls and chitin binding CBM.....218

Figure 8.1: Shows the analysis of fungal laccase protein sequences A) phylogenetic analysis of laccase protein sequence (P. brevispora, D. squalens, F. pinicola, W. cocos, C. globosum) and experimentally determined laccases T. versicolor (1GYC) and M. albomyces (1GW0), B) sequence logos of MCO signature 1 (PS00079) and 2 (PS00080) patterns.230

Figure 8.2: QMEAN scores for the 3D modelled laccase structures obtained from SWISSMODEL server for fungal protein sequences P. brevispora, D. squalens, F. pinicola, W. cocos, C. globosum and Cadophora sp..... 233

Figure 8.3: Homology models of laccase protein sequences A) Phlebia brevispora B) Fomitopsis pinicola C) Dichomitus squalens D) Wolfiporia cocos E) Chaetomium globosum F) Cadophora DSE1049 v1.0.234

Figure 8.4: Protein docking of laccase protein molecular models with syringyl β -O-4 syringyl β -O-4 sinapyl alcohol (Trimer) and guaiacyl β -O-4 syringyl β - β syringyl β -O-4 guaiacyl (Tetramer), A) Phlebia brevispora, B) Wolfiporia cocos, C) Dichomitus squalens, D) Fomitopsis pinicola, E) Chaetomium globosum and F) Cadophora DSE1049 v1.0240

List of Tables

Table I-1: Whole genome sequencing studies of different bacterial strains	8
Table I-2: List of some of the sequencing studies performed on different classes of fungi strains.	9
Table I-3: Illustration of Whole microbiome sequencing studies performed	13
Table 1.1: Illustrates potential wood degrading fungal phylum and their properties.....	23
Table 1.2: Lists the catalytic mechanism and structural studies of different lignin oxidizing enzymes.....	39
Table 1.3: Catalytic mechanisms and structural studies of different lignin degrading auxiliary enzymes (LDA).....	49
Table 1.5: Different classes of carbohydrate esterase (CE) family and their corresponding representing enzymes with note on their protein 3D structure	51
Table 1.6: Different types of Pectic substances and pectinolytic enzymes responsible for its degradation	53
Table 2.1: Details of the <i>P. chrysosporium</i> transcriptome metadata retrieved from NCBI GEO and NCBI SRA	72
Table 2.2: Glycoside hydrolases differentially expressed among different gene expression datasets	78
Table 2.3: Glycosyl transferases and carbohydrate esterases differentially expressed among different gene expression datasets	81
Table 3.1: Details of the <i>P. chrysosporium</i> transcriptome metadata retrieved from NCBI GEO and NCBI SRA	99
Table 3.2: Lists differentially classified lignin degrading enzymes obtained from different gene expression datasets of <i>P. chrysosporium</i>	105
Table 3.3: Lists differentially classified lignin degrading enzymes obtained from different gene expression datasets of <i>P. chrysosporium</i>	107
Table 4.1: List of the <i>P. chrysosporium</i> gene expression datasets retrieved from NCBI GEO repository	129
Table 4.2: The total number of differentially expressed and statistically significant genes (unique) belonging to information storage and processing group among the selected datasets	131
Table 4.3: Transcription factors involved in regulation of plant cell wall utilizing enzymes studied in various commercially important fungi	149
Table 5.1: Platform details of <i>Postia placenta</i> gene expression data	160

Table 5.2: Common differentially expressed significant class of enzymes among different growth conditions of gene expression datasets.....	174
Table 8.1: Lists the physico-chemical properties of laccase protein sequences calculated from ExPASy ProtParam and EDBCP tools	229
Table 8.2: Computationally predicted motifs in <i>P. brevispora</i> , <i>D. squalens</i> , <i>F. pinicola</i> , <i>W. cocos</i> , <i>C. globosum</i> and <i>Cadophora</i> laccase protein sequences obtained from the MOTIF SCAN server	231
Table 8.3: Computationally predicted secondary structure elements of laccase protein sequences calculated using SOPMA web server	232
Table 8.4: Comparison of results obtained from the Swiss model server (BLAST and HHblits) and BLASTP (NCBI PSI-BLAST against PDB server)	234
Table 8.5: Ramachandran plot scores of laccase modelled protein structures obtained after customized refining pipeline using different model refining softwares and results from RAMPAGE, PROQ, ERRAT and RMSD (initial and final protein structures) webserver.....	235
Table 8.6: Lists the final predicted minimum binding energy scores (kcal/mol) of predicted laccase models with lignin model compounds obtained from AutoDock Vina software	236
Table 8.7: Lists the aminoacid residues of modelled fungal laccases in contact with the lignin-based model compounds Monomers (Sinapyl alcohol, Coniferyl alcohol and p-coumaryl alcohol), Dimer, Trimer and Tetramer (Note: aminoacids represented in bold are involved in hydrogen bonding between protein and ligand)	238
Table 8.8: Lists the root mean square deviation (RMSD Å) values obtained from the comparison studies of white, brown and soft rot fungal laccases	245

Chapter-1

Fungal Enzymes Involved in Lignocellulose Degradation

Naturally, lignocellulose is degraded by a large group of fungi and bacteria [1]. Fungi have evolved progressively with their dominant degrading abilities to decay organic debris including plant biomass by penetrating through their hyphae and spores (for long distance dispersal) [2]. Wood rotting fungi are categorized into white, brown and soft rot fungi based on their growth substrate preferences and wood decaying patterns [3]. Moreover, white rot fungi exhibit excellent decaying abilities and solely responsible for the degradation of lignin and polysaccharides in plant biomass. Microscopy based studies have differentiated the white rot decay patterns morphologically into a) simultaneous degradation of lignin and wood polysaccharides. For e.g. *Phanerochaete chrysosporium*, *Trametes versicolor*. b) selective degradation of plant biomass components. For e.g. *Phlebia radiata* [1, 4, 5]. However, some fungi like *Heterobasidium annosum* exhibits both simultaneous and selective decay patterns [6].

Brown rot fungi are well characterized as rapid cellulose and hemicellulose degraders, they access plant polysaccharides by potentially modifying or degrading lignin [7]. These fungi are the major invaders of forest biomass and wood-based constructions. Studies have reported that brown rot fungi have evolved from the saprotrophic white rot fungi by losing several essential genes encoding for lignocellulose degrading enzymes [8]. It was reported that hyphae of the brown rot fungi penetrates the cell lumen, colonizes the ray cells and axial parenchymal cells to access carbohydrates [8].

Most of the ascomycetes and fungi imperfecti cause soft rot decay in the presence of excessive moisture, soft rot decayed wood exhibits a greyish discoloration and fragmentation which is similar as brown rot. Previous morphological studies have divided the soft rot fungi decay into a) type-I (where hyphae penetrates secondary cell walls by forming characteristic cavities) b) type-II (attacks similarly as ascomycetes and white rot fungi leading to wood cell wall thinning) [1]. Wood decaying fungi and its secreted enzymes are being used commercially in biopulping, kraft pulping (xylanase bleaching), cellulases based refining, pitch removal (lipases), slime removal (using enzyme cocktail), fiber modification (pulp and paper industries) etc. Thus, finding its applications in biodegradation of plant polymers, detoxification and bioremediation of several toxic aromatic compounds and also in bio based

industries [1]. The depolymerizing abilities of the wood rotting fungi are directly proportional to its ability to secrete an array of lignocellulolytic enzymes, aromatic compound and detoxifying enzymes.

Table 1.1: Illustrates potential wood degrading fungal phylum and their properties

Type of Wood Rot	Phyla and Order	Wood Degradation Property	Decaying wood	Fungal strain
White Rot	<u>Basidiomycota</u> Agaricales Aphyllorphorales	Causes cell wall erosion in cell lumina by occupying large spaces with its mycelium. Efficiently degrade lignin.	Moist, spongy appearance white or yellow	<i>Phanerochaete chrysosporium</i> , <i>Ceriporiopsis subvermispora</i> ,
Brown Rot	<u>Basidiomycota</u> Agaricales Aphyllorphorales	Penetrates through cell wall pores, by effecting the S2 layer of cell wall in lumen. Efficiently degrades cellulose and hemicellulose.	Dry, shrunken, cracked, in brown colored fragments	<i>Gleophyllum trabeum</i> , <i>Postia placenta</i> , <i>Serpula lacrymans</i> ,
Soft Rot	<u>Ascomycota</u> Deuteromycota	Type I fungi forms cylindrical, biconical cavities in secondary cell walls. Type II fungi are erosive wood degraders.	Decayed wood is brown in color with soft look which further cracks and becomes dry.	<i>Fusarium solani</i> , <i>Penicillium chrysogenum</i> , <i>Daldinia concentrica</i> ,

1.1. Lignocellulolytic Enzymes

Cellulose is the earth's most abundant plant polysaccharide containing large array of glucose units linked through β (1 \rightarrow 4) linkages by existing in both crystalline and amorphous forms. Cellulose is widely distributed in plants by constituting up to 40 to 50% overall dry weight of the plant biomass. Microbial degradation of cellulose (and other plant biomass units) is considered as environmentally friendly, cost effective and potentially efficient method. Majorly cellulose degrading enzymes secreted by microorganisms can be classified as endoglucanases or 1,4- β -d-glucan-4-glucanohydrolases (EC 3.2.1.4) which are involved in random fragmentation of amorphous cellulose units resulting in oligosaccharides units of varying lengths. Exoglucanases which includes cellodextrinases or 1,4- β -D-glucan glucanohydrolases (EC 3.2.1.74) and cellobiohydrolases or 1,4- β -d-glucan cellobiohydrolases (EC 3.2.1.91) these set of enzymes act on reducing and non-reducing ends of cellulose or microcrystalline cellulose and β -glucosidases or β -glucoside glucohydrolases

(EC 3.2.1.21) are required for solubilizing cellodextrins and cellobiose to simple glucose residues [1]. Several extensive and informative reviews are already available on microbial degradation of cellulose units [1-7]

It was well-established that complete hemicellulose degradation requires a combination of hemicellulolytic enzymes such as endo- β -1,4-xylanase, β -xylosidase and other accessory enzymes, like α -arabinofuranosidase, α -glucuronidase, acetyl xylan esterase and ferulic acid esterases [9]. Zhang et al (2011), have reported that even after the application of hydrothermal, steam explosion pre-treatments on plant biomass most of the substituents such as acetyl residues might remain intact with the xylan chain and obstruct the action of xylanases during the enzyme hydrolysis [10]. Xylan hydrolysis is significantly increased by the removal of acetyl side chains and the hydrolysis is hindered by the degree of acetylation [11]. Grohmann et al (1989) have reported that, chemical deacetylation of aspen wood and wheat straw xylan units have enhanced the enzymatic hydrolysis of xylan and thus increased the cellulose accessibility [12].

Naturally, most of the fungi, bacteria and yeast secretes wide range of pectin methyl esterases and pectin depolymerizing enzymes for the degradation of pectin. Previous studies have extensively reported about various endogenous pectinases secreted by plants [13-16]. Based on their specific location of activity protopectinases were classified as A-type (inner site/reacts at the polygalacturonic acid region) and B-type (outer site/polysaccharide chains connected to polygalacturonic acid chain). A-type proto-pectinases were majorly reported to be secreted in the cultures of yeast and yeast like fungi, whereas B-type proto-pectinases were majorly reported in the cultures of *Bacillus* strains and especially in *Bacillus subtilis* cultures [13-15]. Polygalacturonases are class of pectinolytic enzymes which performs the hydrolytic cleavage of polygalacturonic acid by introducing water across the oxygen bridge. Based on its reactivity polygalacturonases are divided into endo (widely reported among fungi, bacteria and yeast, and were also reported in higher plants and parasitic nematodes) [17-22] and exo-polygalacturonases (well-studied in *Erwinia carotovora*, *Agrobacterium tumefaciens*, *Bacteroides thetaiotamicron*, *E. chrysanthemi*, *Alternaria mali*, *Fusarium oxysporum*, *Ralstonia solanacearum*, *Bacillus* sp) [23-28]. Pectin lyases catalyze non-hydrolytic cleavage of pectates or pectinates, lyases cleaves the glycosidic linkages at C-4 by simultaneously eliminating the H at C-5 by producing 4:5 double bonded unsaturated products. Polygalacturonate lyases were majorly reported to be secreted by bacteria and

some pathogenic fungi especially soft rot fungi (Figure 1.1). Pectin esterases are a class of carbohydrate esterases which are involved in de-esterification of methyl ester linkages present on the galacturonan chains of pectic substances present in the plant cell wall [29-32]. The de-esterified pectin is further degraded by the polygalacturonases and lyases (Figure 1.1) [13, 31]. The mode of action of the pectin esterases differs significantly based on its origin, pectin esterases secreted by fungi acts through a multichain mechanism to cleave methyl groups randomly. Whereas, pectin methyl esterases originated from plant acts either on the non-reducing ends or it acts on the groups next to free carboxyl groups by a single chain mechanism [32, 33].

1.2. Lignin Oxidizing Enzymes (LO): Non-specificity and high oxidation potential are the main attributes of lignin oxidizing enzymes. Lignin oxidizing enzymes are categorized into four classes LO1 (Laccases), LO2 (Lignin peroxidases, Manganese peroxidases, Versatile peroxidases and Chloroperoxidases) and LO3 (Cellobiose dehydrogenase) (Figure 1.1). The most thoroughly studied fungal enzymes involved in lignin attack are described below:

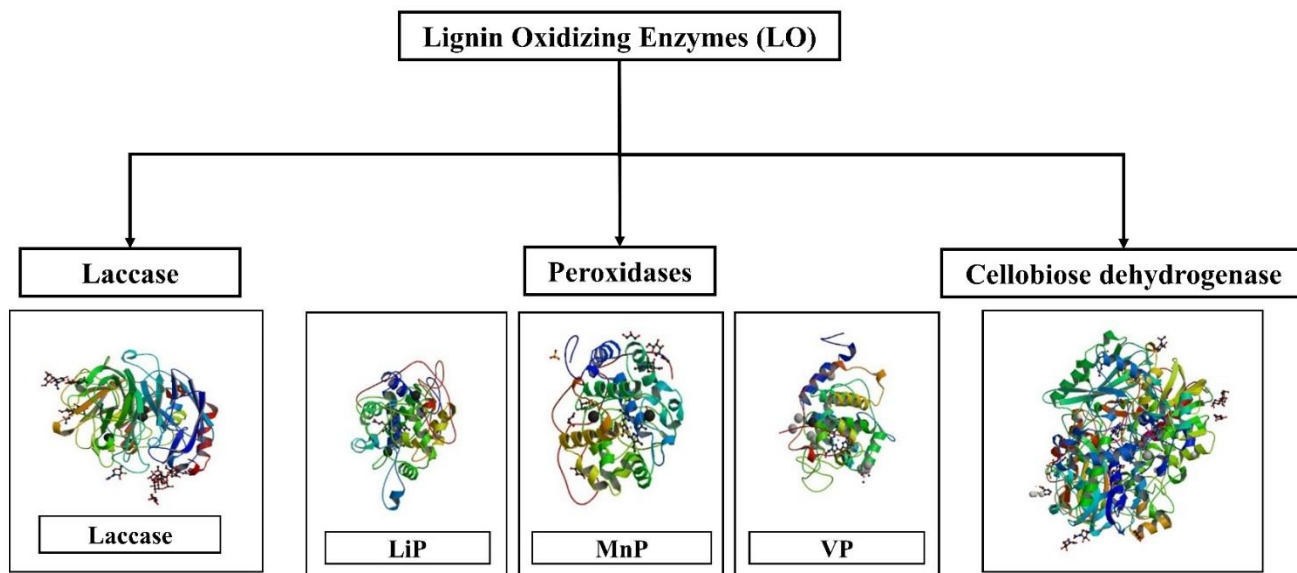


Figure 1.1: Schematic representation of different lignin oxidizing enzymes namely, laccases (PDB ID: 3FPX), lignin peroxidase (LiP) (PDB ID: 1B85) [34], manganese peroxidase (MnP) (PDB ID: 1YYD) [35], versatile peroxidase (VP) (PDB ID: 3FKG) , cellobiose dehydrogenase (PDB ID: 1KDG) [36]. All the enzyme structures were obtained from the PDB RCSB repository.

1.2.1. Laccases (EC 1.10.3.2, benzenediol: oxygen Oxidoreductase)

Laccases represents the largest sub group of blue multicopper oxidases (MCO) and are widely distributed among eukaryotes (fungi, plants) prokaryotes (bacteria) [37] They perform varied functions based on the source organism [37]. Laccase was first discovered in the sap of the Japanese lacquer tree *Rhus vernicifera* [38] and then it was also demonstrated in fungi [39]. Although laccases were discovered during early 19th century they have received much attention during the last five decades for their application to biofuel and biorefinery fields. The involvement of laccase in the degradation of wood by fungal groups such as basidiomycetes, ascomycetes has attracted scientific communities to study the structure, function and mechanisms of laccases [40]. Many fungal species belonging to the basidiomycetes phylum such as *Abortiporus biennis*, *Agaricus bisporus*, *Agaricus brunnescens*, *Armillaria mellea*, *Aspergillus nidulans*, *Botrytis cinerea*, *Ceriporiopsis subvermispora*, *Ganoderma lucidum*, *Lentinus edodes*, *Myceliophthora thermophile*, *Neurospora crassa*, *Penicillium crysogenum*, *Phanerochaete chrysosporium*, *Phlebia brevispora*, *Phlebia radiata*, *Pleurotus erygii*, *Pleurotus ostreatus*, *Pleurotus sojar-caju* *Polyporus species*, *Rhizoctonia Solani*, *Trametes hirsuta*, *Trametes versicolor* and *Trichoderma* were reported to secrete laccase [41]. Laccases are widely studied for two major functions a) their role in lignin polymerization (lignification) in plants, b) lignin depolymerization by fungi [42]. The contrasting role of laccases on lignin depolymerization was proved in vitro by Hatakka 1994 and Youn et al 1995, showing the oxidative reaction of laccases on lignin, resulting in loss of an electron from phenolic hydroxyl groups of lignin resulting in phenoxy radicals [43, 44]. These studies have also showed that these radicals can spontaneously reorganize leading to the cleavage of alkyl side chains of polymer. At the same time, the polymerizing activity of the laccase might result in the polymerization of low molecular weight compounds [41]. These studies suggested that lignin degradation by fungi in nature occurs by the synergistic effect of other lignin degrading enzymes and non-enzymatic components which establishes a balanced environment between lignin depolymerization and enzymatic polymerization [41]. Although studies have reported the involvement of laccases in both lignin polymerization and depolymerization, the exact role of laccases and other partnering enzymes in the degradation and modification of lignin were still under investigation [41, 42]. Apart from wood decay, laccases play important role in fungal physiological processes such as morphogenesis, fungal plant pathogen/host interactions, stress defense and lignin degradation [41, 45]. In fungi, laccases are expressed

during different stages of fungal development (morphogenesis, growth of rhizomorphs, sporulation, pathogenesis and virulence). According to Leatham and Stahmann 1981, increased laccase activity was observed in the developing fruiting bodies of *Lentinus edodes* (a commercially cultivable mushroom) [46]. The role of laccases on mushroom development was proved by Ikegaya et al (1993), in this study the developing fruiting bodies of *L. edodes* were treated with diethyldithiocarbamate (a potential inhibitor of laccase) which resulted in the decreased growth of *L. edodes* fruiting bodies, thus proving the role of laccase in fungal development [47]. A similar study was conducted on *Armillaria mellea* by Worrall et al (1986) which showed the requirement of laccase for the development and growth of rhizomorphs [48]. Laccases are also involved in imparting specific virulence properties to the fungi, *Botrytis cinerea* (common plant infecting fungi) secretes laccases which causes infection in some plants especially carrot and cucumber by triggering plant toxins such as cucurbitacins and tetracyclic triterpenoids. However, the virulence of these laccases was inhibited in EDTA pretreated plant tissues [49]. Thus, fungal laccases play three major functions: lignin degradation, detoxification and pigment formation. Industrially laccases are important in paper and pulp, bio bleaching, textile industries etc.

Mechanism: Laccases use their distinctive redox ability of copper ions for catalyzing the oxidation of various aromatic substrates concurrently reducing the molecular oxygen to water [50]. Laccases are able to catalyze direct oxidation of ortho, para-diphenols, aminophenols, polyphenols, polyamines, aryl diamines and also some inorganic ion [41, 51-55]. Laccases depends on copper (Cu) for their catalytic action, based on the number of copper ions laccases can be classified as dimeric or tetrameric glycoproteins. In addition, based on the types of copper ion centers they are classified as: a) Type-I (blue copper center) b) Type-II (normal copper center) c) Type-III (coupled binuclear copper center) that differ in their characteristic electronic paramagnetic resonance (EPR) signals [56, 57]. Type-I copper coordinates with four amino acids as ligands: two histidines, one cysteine and one methionine. Type-I copper containing laccases are generally a deep blue color, which can be detected by its absorbance at 600 nm wavelength. However, laccases which fail to absorb at 600 nm were reported in *Pleurotus ostreatus* (called white laccase) [58] *Panus tirinus* (called yellow laccases) [59]. Type-II copper coordinates with two histidine and water as ligands, Type-III copper coordinates with three histidines and a hydroxyl bridge which imparts strong anti-ferromagnetic coupling between the type-III copper atoms [60]. Type-II copper atoms do not absorb in the visible spectrum, while type-III copper atoms have an electron absorption at a

wavelength of 330 nm. Based on the structural properties of type of copper ions laccases are divided into high and low redox potential enzymes. Bacteria and plants secrete low redox potential laccases, whereas white rot fungi and some basidiomycetes secrete high-redox potential laccases [61, 62].

Different copper centers present in the laccase participate and completes the enzymatic reaction. Unlike peroxidases laccases does not require hydrogen peroxide for the oxidation of monolignols. Enzyme catalysis can be divided into three main stages: the copper ion of type-I is reduced by the reducing substrate followed by internal electron transfer between the type-I, type-II and type-III Cu clusters [37]. Finally, the reduction of oxygen takes place at the type-II and III Cu's resulting in water formation (Figure 1.2). *In vitro* lignin degradation by laccase primarily oxidizes phenolic hydroxyl groups of lignin to form phenoxy radicals which further reorganize to cleave the alkyl side chains. Laccase can degrade β -1 and β -O-4 dimer linkages between $C\alpha$ - $C\beta$ and cause $C\alpha$ oxidation and aryl-alkyl cleavages [37]. Thus the generated reactive radicals further release monomers by breaking down covalent bonds [63]. Due to the steric hindrance of laccase it cannot directly contact large polymers, thus small organic compounds or metals such as veratryl alcohol, manganese and 3-hydroxy anthranilic acid are oxidized and further activated to mediate radical catalyzed depolymerization of lignin [63, 64].

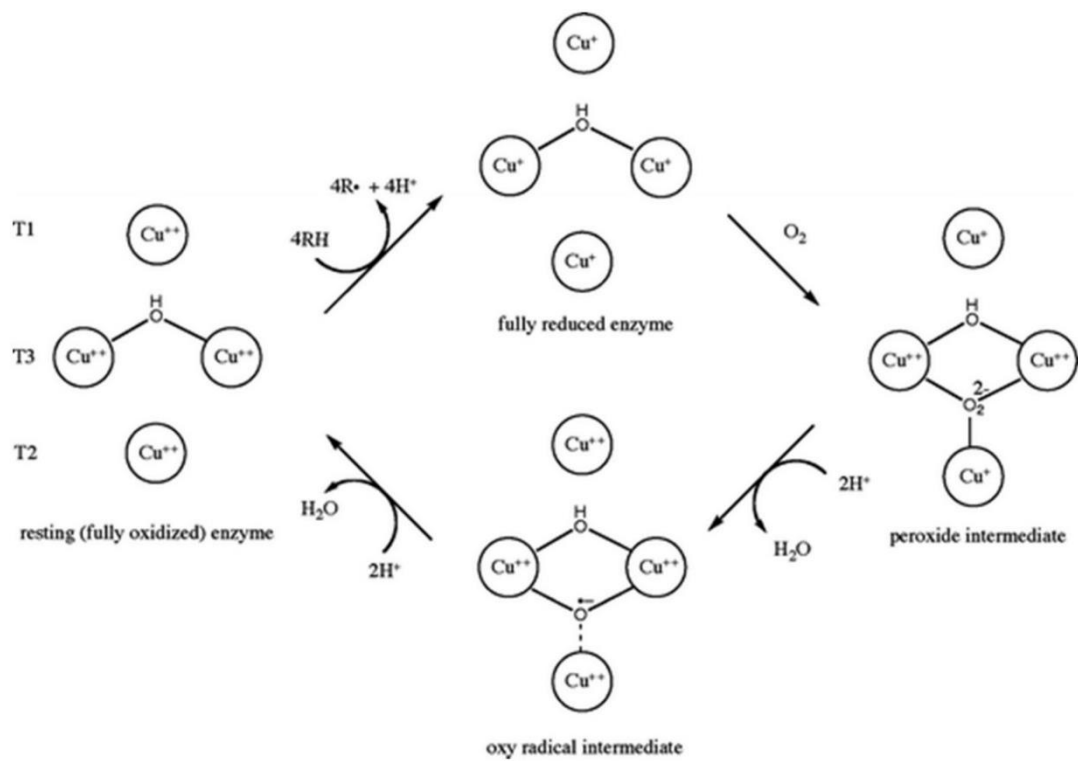


Figure 1.2: Catalytic cycle of laccase. Reprinted with permission from ref [64], Copyright © 2008 Springer Science+Business Media B.V.

1.2.2. Peroxidases (EC:1.11.1.x): Peroxidases are large group of enzymes widely distributed among plants, animals and microbes. Peroxidases play a wide variety of activities based on the source of the organism. Peroxidases are involved in several physiological processes such as plants defense mechanisms (response to pathogens), wound healing, auxin catabolism, lignification and suberization [65]. Microbes such as fungi and bacteria are well known for their ability of delignification which is efficiently fulfilled by the different types of peroxidases such as (LiP, MnP and VP). Peroxidases can also efficiently decolorize synthetic dyes and bioremediation of waste water and degradation of several toxic chemicals such as phenolic contaminants, polychlorinated biphenyls, chlorinated alkanes and alkenes, chlorinated dioxins, chlorinated insecticides and removal of endocrine disruptive chemicals etc, thus playing variety of roles in the environment [66]. Molecular structures of lignin degrading peroxidases share several common characteristics such as [67], Ligninolytic peroxidases generally contain a haem cofactor located internally in a cavity (haem pocket), which is connected to the protein by two small access channels [67-71]. Larger channel are common among all haem peroxidases, they are required for the hydrogen peroxide to reach the haem and react with (Fe^{+3}) forming an activated two electron enzyme form called compound I [67-71]. The entrance of this channel forms the substrate binding site in some peroxidases. A second channel extends to the heme propionate substrate where some specific lignolytic enzymes oxidize Mn^{2+} and Mn^{3+} which acts a diffusible oxidizer of phenolic lignin and other organic molecules [67-71]. In this section we will be focusing on the delignification mechanisms of lignin peroxidases, manganese peroxidases, and versatile peroxidase.

1.2.3. Lignin Peroxidases (E.C. 1.11.1.14)

Lignin peroxidases (LiP) the most studied lignin depolymerizing enzymes, LiP was first discovered in the extracellular medium of *P. chrysosporium* under nitrogen limited conditions [72]. Similar to classic peroxidases, LiP are dependent on hydrogen peroxide. The overall reaction mechanism of LiP is 1,2-bis(3,4-dimethoxyphenyl) propane-1,3-diol + $\text{H}_2\text{O}_2 \rightleftharpoons$ 3,4-dimethoxybenzaldehyde + 1-(3,4-dimethoxyphenyl)ethane- 1,2-diol + H_2O [64]. LiP can oxidize a wide range of phenolic compounds, organic compounds and also different lignin

model non-phenolic compounds by using hydrogen peroxide with a redox potential up to 1.4 V, thus showing its non-specificity towards substrates [73].

Mechanism: Lignin peroxidase resembles horse radish peroxidase (a classical peroxidase highly studied) by containing Fe (III) as a cofactor which is pentacoordinated to four heme tetrapyrrole nitrogens and to a histidine residue [74]. Lignin peroxidases are dependent on H_2O_2 for their reaction. H_2O_2 oxidize LiP resulting a two electron-oxidized intermediate (Compound I) in which iron is present as Fe (IV) leaving a free radical on the tetrapyrrole ring or on a nearby amino acid. Compound I then oxidizes a donor substrate to form a second intermediate (Compound II) and a substrate free radical (Figure 1.3) [74]. Later reduction of the enzyme to its resting state can be accomplished either by the same substrate molecule or with a second substrate molecule by giving off substrate-free radical [74]. An important functional difference between LiP and other classical peroxidases is that lignin peroxidases can oxidize aromatic rings that are moderately activated by electron donating substituents, at the same time classical peroxidases act only on strongly activated aromatic substrates. Therefore, LiP and horseradish peroxidase can oxidize 1, 2, 4, 5-tetra-methoxybenzene, phenols and anilines, at the same time LiP are capable of abstracting an electron from aromatics that carry only two or three ether like the major nonphenolic structures of lignin [75]. Primary products of this oxidation are temporary cation radical intermediates which certainly breakdown. Majorly $C\alpha-C\beta$ bonds of propyl side chains are broken down to give benzaldehydes which are the precursors of benzoic acid molecules, these benzoic acid molecules are mainly observed in lignin decaying white rot fungi (Figure 1.3) [76]. The unusual activity of lignin peroxidases is due to two structural differences, an electron-deficient iron atom in the porphyrin compared to classical peroxidases which makes LiP a stronger oxidant [77] and an invariant Trp171 in the isozyme of LiPA. This residue is present on the enzyme surface and is known to participate in a wide range electron transfers from aromatic substrates since they cannot contact the oxidized haem directly [78]. This important feature of LiP is responsible for oxidizing complex lignin and its related substrates directly. This function of Trp171 in LiPA was proved by a site directed mutagenesis in which the Trp171 was replaced by serine, which resulted in the loss of activity [79]. It was shown that the efficiency of LiP catalyzed oxidation of lignin molecules markedly decreases with an increase in size of the lignin molecule. LiP catalyzed oxidation of lignin trimers was found to be only 4% of the rate of oxidation of a monomer model [80]. Oxidation of lignin molecules by LiP takes place in the presence of veratryl alcohol, and the role of VA in oxidation by LiP are given below

[74]. Studies have showed that the VA cation radical oxidizes has a long half-life of 40ms even at acidic conditions [81, 82]. VA is the substrate of LiP. It was suggested that the VA cation radical oxidizes lignin molecules at remote locations [83]. VA acts as an efficient electron donor to protect LiP from oxidative inactivation by hydrogen peroxide. As LiP oxidizes large and complex lignin substrates, which is a slow reaction, VA prevents oxidization of LiP [84]. VA is also essential for the reduction of LiP compound II. Compound I is reduced by non-methoxylated lignin structures. As these lignin structures are difficult to oxidize since they carry only one electron donating ether group. Compound II of LiP is a comparatively weaker oxidant than compound I [85].

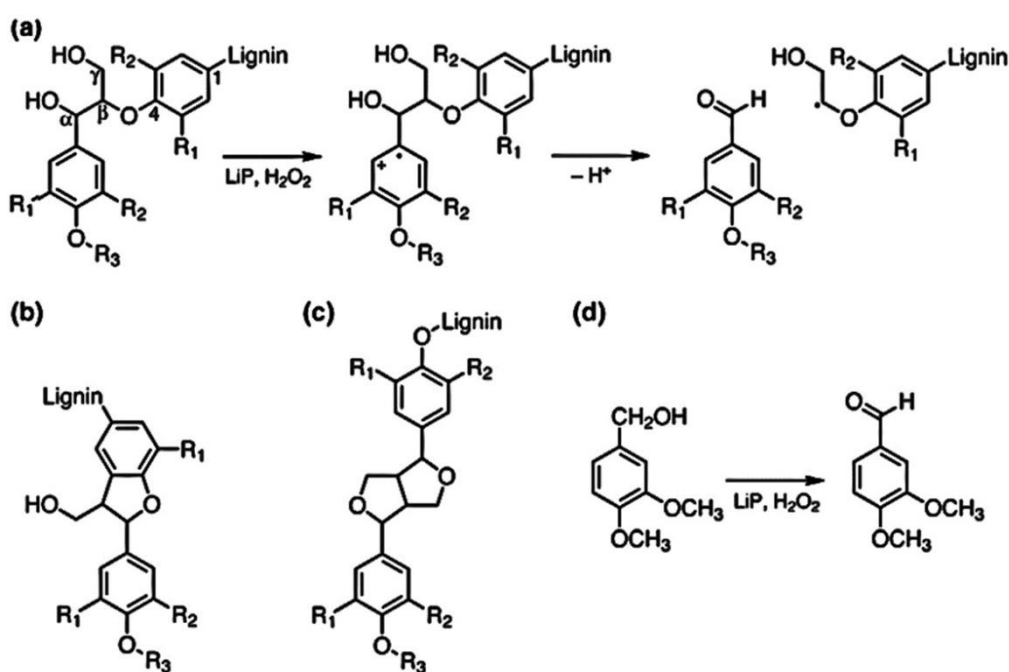


Figure 1.3: Chemical structures and reactions discussed in the text. (a) The principal β -O-4 structure of lignin and pathway for its C $_{\alpha}$ -C $_{\beta}$ cleavage by LiP. (b) A phenylcoumaran lignin structure. (c) A resinol lignin structure. (d) LiP-catalyzed oxidation of the fungal metabolite veratryl alcohol. Gymnosperms contain lignin's in which most subunits have R $_1$ = OCH $_3$ and R $_2$ = H. Angiosperm lignin's also contain these structures but have in addition subunits in which R $_1$ = OCH $_3$ and R $_2$ = OCH $_3$. Grass lignin's contain both types of structures but have in addition some subunits in which R $_1$ = H and R $_2$ = H. These nonmethoxylated lignin structures are more difficult to oxidize than those that contain one or two methoxyl groups. In the predominating nonphenolic structures of lignin, R $_3$ = lignin, whereas R $_3$ = H in the minor phenolic structures. Reprinted with permission from ref [74], Copyright © 2008, Elsevier.

1.2.4. Manganese Peroxidases (EC 1.11.1.13):

Wood decaying white rot fungus and other litter decomposing fungi efficiently degrade lignin in wood. These fungi secrete several non-specific oxidoreductases, among them manganese peroxidase plays an important role [86]. Manganese peroxidase (MnP) was first discovered in *P. chrysosporium* two decades ago [87, 88], however it received less attention than lignin peroxidase in beginning. Later it was found that LiP is not produced by all white rot fungi [43, 89, 90]. Production of MnP is limited only to basidiomycetes. Mainly two ecophysiological groups of fungi i.e. wood degrading fungi causing white rot and soil litter decomposing fungi secrete manganese peroxidase [90]. Wood decaying fungi belonging to families such as *Meruliaceae*, *Coriolaceae*, *Polyporaceae* and soil litter decomposing fungi such as *Strophoriaceae*, *Tricholomataceae* are known fungal families, which secrete MnP. Some prominent MnP producing fungi are *Abortiporus biennis*, *Agaricus bisporus*, *Armillaria mellea*, *Auricularia sp. M37*, *Bjerkandera adusta*, *Ceriporiopsis subvermispora*, *Coriolopsis polyzona*, *Dichomitus squalens*, *Ganoderma lucidum*, *Heterobasidion annosum*, *Hypholoma fasciculare*, *Lentinula (Lentinus) edodes*, *Panus tigrinus*, *Phaeolus schweinitzii*, *Phallus impudicus*, *Phanerochaete chrysosporium*, *Phanerochaete sordida*, *Phlebia brevispora*, *Phlebia radiata*, *Pleurotus enryngii*, *Pleurotus sajor-caju*, *Stropharia aeruginosa*, *Stropharia coronilla*, *Trametes hirsuta*, *Trametes versicolor* [86].

Mechanism: MnP is different from other peroxidases as it uses Mn (II) as the reducing substrate. MnP oxidizes Mn (II) to Mn (III), which then catalyzes the oxidation of a wide range of monomeric phenols, lignin model phenolic compounds and dyes [87, 91, 92]. The reaction mechanism of MnP proceeds as: first oxidation of Mn (II) by compound I (MnP-I), followed by oxidation of compound II (MnP-II) yielding Mn (III). MnP is a strong oxidizing agent like LiP, it cannot oxidize nonphenolic lignin related compounds because it lacks the invariant Trp171 residue which is required for electron transfer to aromatic substrates [74]. MnP has a manganese binding site which contains many acidic amino acids and also a heme propionate group. Thus, one electron transfer to compound I of MnP takes place from bound Mn^{+2} . Further Mn^{+3} is released from the active site in presence of the bidentate chelators such as oxalate, which helps prevent the disproportionation to Mn^{+2} and insoluble Mn^{+4} . This reaction is required for the transfer of oxidizing power of MnP to Mn^{3+} , which diffuses into the lignified cell wall thus attacking it from inside [74]. An important feature of MnP is to oxidize the low permeable lignocellulose network making it different from other peroxidases [64]. Chelators

such as oxalate increase the electron density on Mn^{+3} which makes it a weak oxidant, thus Mn^{+3} organic acid chelates produced by MnP cannot oxidize the nonphenolic substrates of lignin. Mn^{+3} chelates cannot cause extensive ligninolysis as they can only attack rare phenolic structures of lignin, which often are the end groups of lignin. The catalytic cycle of MnP begins with the binding of hydrogen peroxide or an organic peroxide to the native ferric enzyme resulting in the formation of an iron-peroxide complex (Figure 1.4). Further the breakdown of the oxygen-oxygen peroxide bond depends on a 2-electron transfer reaction from the heme resulting in the formation of MnP compound I (i.e. a Fe^{4+} -oxo-porphyrin radical complex). The dioxygen bond is cleaved resulting in removal of water and further reduction proceeds via MnP compound II. The Mn^{2+} ion (monochelated) donates one electron to the porphyrin intermediate and is oxidized to Mn^{3+} . Similarly compound II is reduced by releasing another Mn^{3+} and a second water molecule, thus leading to the resting state of the enzyme (Figure 1.4) [93-95].

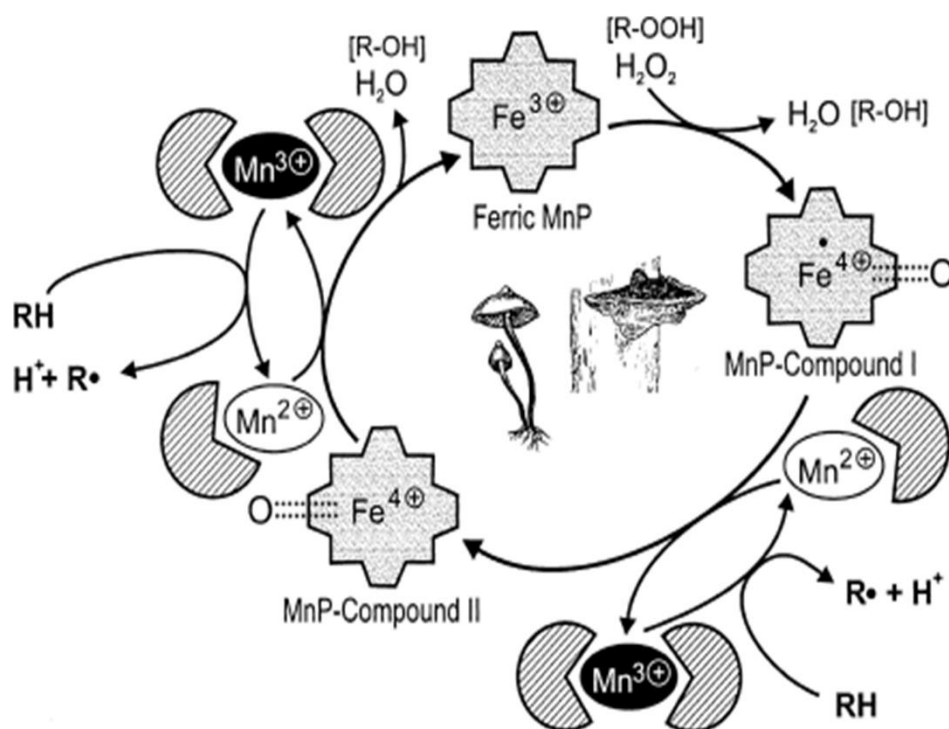


Figure 1.4: Catalytic cycle of manganese peroxidase. Reprinted with permission from ref [86], Copyright © 2002, Elsevier.

The oxidation of phenolic compounds by MnP occurs by Mn (III) chelator complexes, which diffuses and catalyzes one electron oxidation of phenolic compounds producing a phenoxy radical intermediate. The phenoxy radical intermediate undergoes bond cleavages, rearrangements and degradation of compounds non-enzymatically to produce different

breakdown products [92, 96, 97]. In contrast unchelated Mn (III) causes the formation of reactive radicals as second mediators for the oxidation of non-phenolic compounds. Oxidation of non-phenolic compounds by MnP is different from LiP, as LiP oxidizes by abstracting electrons from the aromatic ring resulting in a radical cation. In presence of thiols like glutathione, Mn (III) causes the oxidation of benzyl alcohol and diarylpropane structures to their corresponding aldehydes [98, 99].

1.2.5. Versatile Peroxidases:

Versatile peroxidases a new family of lignolytic peroxidases were reported for the first time in *P. chrysosporium* along with other lignolytic enzymes such as LiP and MnP [67]. Several fungi belonging to genera such as *Pleurotus*, *Bjerkandera*, *Lepista*, *Panus* and *Trametes* species were reported to produce versatile peroxidase (VP). Versatile peroxidase have important properties which combines the substrate specificity characteristics of the three fungal peroxidases such as manganese peroxidase, lignin peroxidase and *Coprinus cinereus* peroxidase [67]. Two well-known studies have revealed the occurrence of versatile peroxidases in nature, in the first study a Mn²⁺ binding site was introduced into the LiP of *P. chrysosporium* by site directed mutagenesis, the resulting enzyme had MnP activity [100]. In the second study a tryptophan residue similar to that in LiP was introduced into the MnP of *P. chrysosporium* and the enzyme acquired LiP activity [101]. Versatile peroxidase coding genes were first cloned and sequenced from *Pleurotus eryngii*. Studies of the catalytic properties of VP suggested that they were due to its hybrid molecular construction combining different oxidation and substrate binding sites[102, 103].

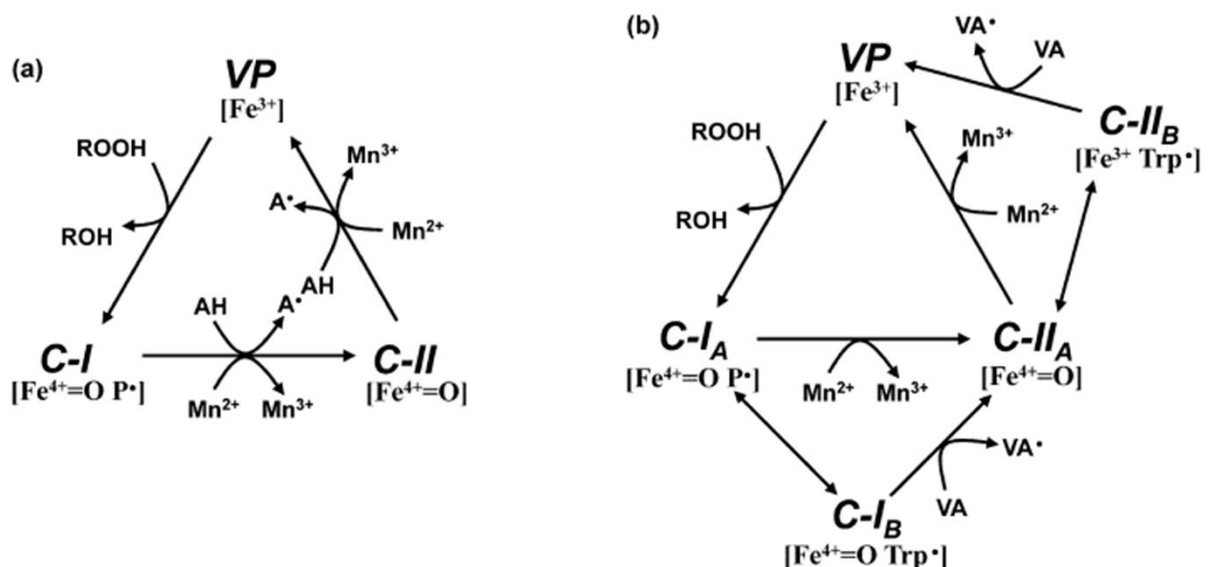


Figure 1.5: Schemes of VP catalytic cycle. (a) Basic cycle described by [102] including two-electron oxidation of the resting peroxidase (VP, containing Fe^{3+}) by hydrogen peroxide to yield compound I (C-I, containing Fe^{4+} -oxo and porphyrin cation radical), whose reduction in two one-electron reactions results in the intermediate compound II (C-II, containing Fe^{4+} -oxo after porphyrin reduction) and then the resting form of the enzyme. As shown in the cycle, VP can oxidize both: (i) aromatic substrates (AH) to the corresponding radicals ($\text{A}\cdot$); and (ii) Mn^{2+} to Mn^{3+} , the latter acting as a diffusible oxidizer. (b) Extended cycle including also compounds I_B (C- I_B , containing Fe^{4+} -oxo and Trp radical) and II_B (C- II_B , containing Fe^{3+} and Trp radical) involved in oxidation of veratryl alcohol (VA) and other high redox potential aromatic compounds (C- I_B and C- II_B are in equilibrium with C- I_A and C- II_A respectively, which correspond to C-I and C-II in (a) (other low redox potential aromatic compounds are probably oxidized by both the A and B forms but they are not included for simplicity). The active Trp in C- I_B and C- II_B would be Trp164 (the part of the cycle showing aromatic substrate oxidation would be also applicable to LiP, being Trp171 the active amino acid).. Reprinted with permission from ref [67], Copyright © 2005, Elsevier.

Mechanism: Basic features of versatile peroxidase are similar to those of all other classical peroxidases, however it is unique as far as the substrates that it is able to oxidize. A complete catalytic cycle combining those of other fungal peroxidases such as LiP and MnP was proposed by Ruiz-Duenas *et al.* Similar to LiP, versatile peroxidase also initiates the LRET pathway (Long range electron transfer) at an exposed tryptophan residue [67, 104]. Studies have examined the catalytic mechanism of VP using veratryl alcohol (reducing substrate) and its transitory states in the catalytic cycle. On reaction with one molecule of hydrogen peroxide the ferric group of VP (resting state) was converted to Compound I (Fe^{4+} -oxo-porphyrin⁺ complex) causing spectral changes (Figure 1.5) [67, 105, 106]. Compound I oxidizes a molecule of veratryl alcohol resulting in Compound II (Fe^{4+} -oxo), which will further oxidize another molecule of veratryl alcohol further reducing the enzyme back to its resting state [67, 106]. VP can oxidize high redox potential dyes like reactive black 5 (RB5) and also can oxidize low redox potential compounds such as phenolic monomers, simple amines, Mn^{2+} etc [104]. Compared to LiP and MnP the oxidization capacity of VP is higher for phenolic compounds; this ability might be due to its relatively more accessible distal main solvent channel allowing a third lower redox potential substrate oxidation site as in CiP. A research study conducted by Perez-Boada *et al* reported that the spectral changes occurring during the oxidation of

phenolic compounds by VP shows that VP in its resting state has a higher absorbance at 407 nm. Similarly, during charge transfer, the transient states such as Compound I and II have an absorbance at 505 nm and 637 nm respectively [67, 106]. Two major enzymes MnP and VP are known for their ability to oxidize Mn^{2+} to Mn^{3+} , the Mn^{2+} oxidation site of *P. eryngii* VP is similar to that of *P. chrysosporium* MnP. In VP, the Mn^{2+} binding site is formed by the side chains of Glu36, Glu40 and Asp175 located in front of the internal propionate of heme. Carboxylate groups of the amino acids and heme propionate are responsible for Mn^{2+} binding and for succeeding electron transfer to the activated heme of VP compounds I and II. Studies of the VP crystal structure showed a variable orientation of the Glu36, and Glu40 sidechains by interaction with Asp175 [104]. The position of these amino acids in recombinant VP shows an open gate conformation before exposure to Mn^{2+} , thus enabling the oxidation of the Mn^{2+} . At the same time native *P. eryngii* VP shows that the two glutamate side chains are pointed towards the Mn^{2+} corresponding to a closed gate conformation. In this conformation the carboxylate groups of Glu36, Glu40, Asp 175 and the propionate heme groups are at a distance from Mn^{2+} . VP also oxidizes high redox potential substrates similar to LiP (a classic ligninolytic enzyme) through the LRET pathway. This pathway occurs in several redox proteins like cytochrome-c-peroxidase, which oxidizes cytochrome-c on its surface by transferring electrons to tryptophan residues [104-106]. The LRET pathway was known earlier for its involvement in lignin degradation by different ligninolytic enzymes, thus overcoming steric hindrance which prevent the direct interaction of the heme group and the lignin polymer. Structural studies of VP show that three possible LRET pathways are involved during the oxidation of aromatic substrates by VP [104]. Oxidation of aromatic substrates starts at Trp 164 or His232 of VPL and at His82 or Trp170 of VPS1. VP can also efficiently oxidize low reduction potential compounds like ABTS, p-hydroquinone and 2, 6 dimethoxy phenol. Enzyme kinetics studies have showed that VP has two independent oxidation sites characterized by high and low specificities. Site directed mutagenesis of VP Trp164 performed by Ruiz-Duenas *et al* showed that in Trp164 mutants the high specificity active site was removed while the low specificity site remained intact [104, 105]. Studies have confirmed a similar effect of a W164S mutation on VP oxidation of phenols. Based on these studies we conclude that the catalytic features of VP are due to its hybrid molecular architecture which includes different oxidation sites for Mn^{2+} , high redox potential substrates (aromatic compounds) and low redox potential substrates (phenols and dyes) [104, 105].

1.2.6. Cellobiose dehydrogenase

Cellobiose dehydrogenase is an extracellular enzyme involved in carbohydrate metabolism that was shown to be involved in lignin degradation [107, 108]. It was first isolated from an imperfect form of *P. chrysosporium* (*Sporotrichum pulverulentum*) [109]. Cellobiose dehydrogenase is a flavocytochrome enzyme which can oxidize various carbohydrates such as cellobiose (major product of cellulose degradation) and mannobiose (product of mannose degradation) [110]. Several fungi were reported to produce cellobiose dehydrogenase, mostly white rot fungi such as *P. chrysosporium* (*Sporotrichum pulverulentum*), *Trametes versicolor*, *Pycnoporus cinnabarinus*, *Polyporus dichrous*, *Merulius tremellosus*, *Phlebia radiata*, *Pleurotus ostreatus* and *Fomes annosus*. Coniophora puteana (brown rot fungi), soft rot fungi, such as *Sporotrichum thermophile* (*Myceliophthora thermophile*), *Schizophyllum commune*, *Humicola insolens*, *Sclerotium Rolfsii*, *Chaetomium cellulolyticum*, imperfect soft rot fungi such as *Monilla sitophila*, *Agaricus bisporus* (Mushroom), *Stachybotrys* (Mold), *Cladodporium* (Mold) [109]. CDH degrades cellobiose and mannobiose to lactones by removing two electrons, which can be further transported to electron acceptors such as quinones, phenoxyradicals and dioxygen [110]. In CDH two prosthetic groups, FAD and heme, makes the enzyme suitable for the reduction of one electron acceptors such as radicals and metal ions. CDH has a high specificity for amorphous cellulose and less towards microcrystalline cellulose a unique property among non-hydrolytic enzymes [110]. CDH can produce hydroxyl radicals by reducing Fe^{3+} to Fe^{2+} and O_2 to H_2O_2 , These reactive species depolymerize cellulose, xylan and to some extent lignin polymers [110].

Mechanism: CDH has the properties of a typical dehydrogenase with both oxidative and reductive reactions. CDH oxidizes the C1 position of a saccharide to a lactone which is spontaneously hydrolyzed to a carboxylic acid. The electrons taken up by the enzyme are later transferred to one or two electron acceptors [111, 112]. Substrate specificity of CDH is higher for cellobiose, cellodextrins, lactose, mannobiose and galactosylmannose. However, the later substrates have higher K_m -values, the true substrates for CDH are di or oligosaccharides with reducing ends containing glucose or mannose residues. Monosaccharides such as glucose, mannose and maltose have very high K_m values suggesting that there is binding of two glucose residues to the active site in separate subsites, at the same time monosaccharides have lower K_{cat} values than the di or oligo saccharides which suggests that binding of the β -dihexosides to the active site stimulates the catalysis

creating an induced fit [113]. CDH also generates highly reactive hydroxyl radicals by a Fenton type reaction in the presence of an electron donor. Several studies were conducted to study the individual roles of the two prosthetic groups (flavin and heme domains) in the oxidation of compounds, which showed that oxidation of cellobiose (electron donor) is carried out by the FAD group which is further converted to FADH₂ and later transfers the electrons to the heme group (Figure 1.6) [111, 112, 114]. Several groups have proposed a role for CDH in lignin depolymerization by reducing phenoxy radicals thus preventing repolymerization of the radicals. Studies conducted by Henriksson et al (1995) have showed that the CDH could stop the repolymerization of lignin model compounds, this was showed by incubating CDH with cellobiose, ferric ions, hydrogen peroxide and lignin model compounds. CDH generates highly reactive hydroxyl molecules that depolymerized the polymers showing that CDH does not depolymerize lignin or its subunits directly but hydroxyl radical groups are involved in the degradation of lignin related compounds [107, 115-117].

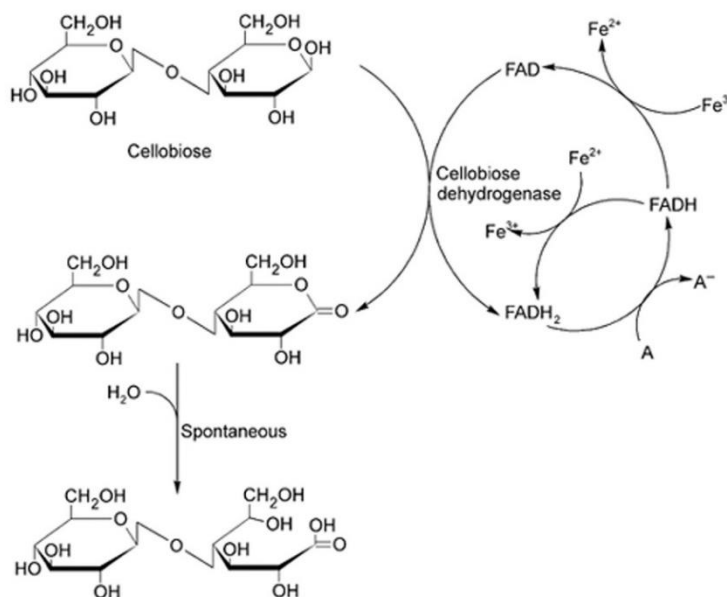


Figure 1.6: Reactions of cellobiose dehydrogenase based on [118]. 'Fe' represents the heme iron, 'A' represents the one-electron acceptor. Reprinted with permission from ref [119], Copyright ©2008, Oxford University Press.

Table 1.2: Lists the catalytic mechanism and structural studies of different lignin oxidizing enzymes.

Enzyme, FOLy, E.C number	Catalytic Reaction	Popular Strain Reference
Laccase (LO1) (EC 1.10.3.2)	$4\text{-benzendiol} + \text{H}_2\text{O}_2 \rightarrow 4\text{-benzosemiquinone} + 2\text{H}_2\text{O}$	<i>Trametes versicolor</i> [60] <i>Coprinus cinereus</i> [120] <i>Melanocarpus albomyces</i> [121] <i>Cerrena maxima</i> [122] <i>Thielavia arenaria</i> [123] <i>Lentinus tigrinus</i> [124]
Lignin Peroxidase (LO2) (EC 1.11.1.14)	LiP oxidizes alkyl side chains and benzyl alcohol, it is involved in breakdown of C-C side chains and aromatic rings of lignin	<i>Trametes cervina</i> [125] <i>Phanerochaete chrysosporium</i> [126, 127] [128-132]
Manganese Peroxidase (LO2) (EC 1.11.1.13)	MnP's catalytic mechanism is dependent on hydrogen peroxide and Mn^{2+} ions.	<i>Phanerochaete chrysosporium</i> [35, 133-136]
Versatile Peroxidase (LO2) (EC 1.11.1.16)	VP has substrate specificity features similar to that of MnP and LiP	<i>Pleurotus eryngii</i> [105, 106, 137-139]
Cellobiose Dehydrogenase (LO2) (EC 1.1.99.18)	CDH catalyzed reactions [110] $\text{Cellobiose} + 2 \text{Fe}^{3+} \rightarrow \text{Cellobionolactone} + 2 \text{Fe}^{2+}$ $\text{Cellobiose} + \text{O}_2 \rightarrow \text{Cellobionolactone} + \text{H}_2\text{O}_2$ (Spontaneous reaction) $\text{Fe}^{2+} + \text{H}_2\text{O}_2 \rightarrow \text{Fe}^{3+} + \text{OH}^- + \text{OH}^*$	<i>Phanerochaete chrysosporium</i> [36, 140-142]

1.3. Lignin Degrading Auxiliary Enzymes (LDA): Lignin degrading auxiliary enzymes are mostly H_2O_2 producers, as lignin degrading enzymes such as laccase, LiP, MnP, VP require the presence of extracellular H_2O_2 . Currently there are 7 enzymes classified as lignin degrading auxiliary enzymes (LDA): aryl alcohol oxidase (LDA1), vanillyl alcohol oxidase (LDA2), glyoxal oxidase (LDA3), pyranose oxidase (LDA4), galactose oxidase (LDA5), glucose oxidase (LDA6) and benzoquinone reductase (LDA7) (Figure 1.7) [143]. Among these 7 different enzymes aryl alcohol oxidase, glyoxal oxidase are the most active H_2O_2 generating enzymes [108, 144].

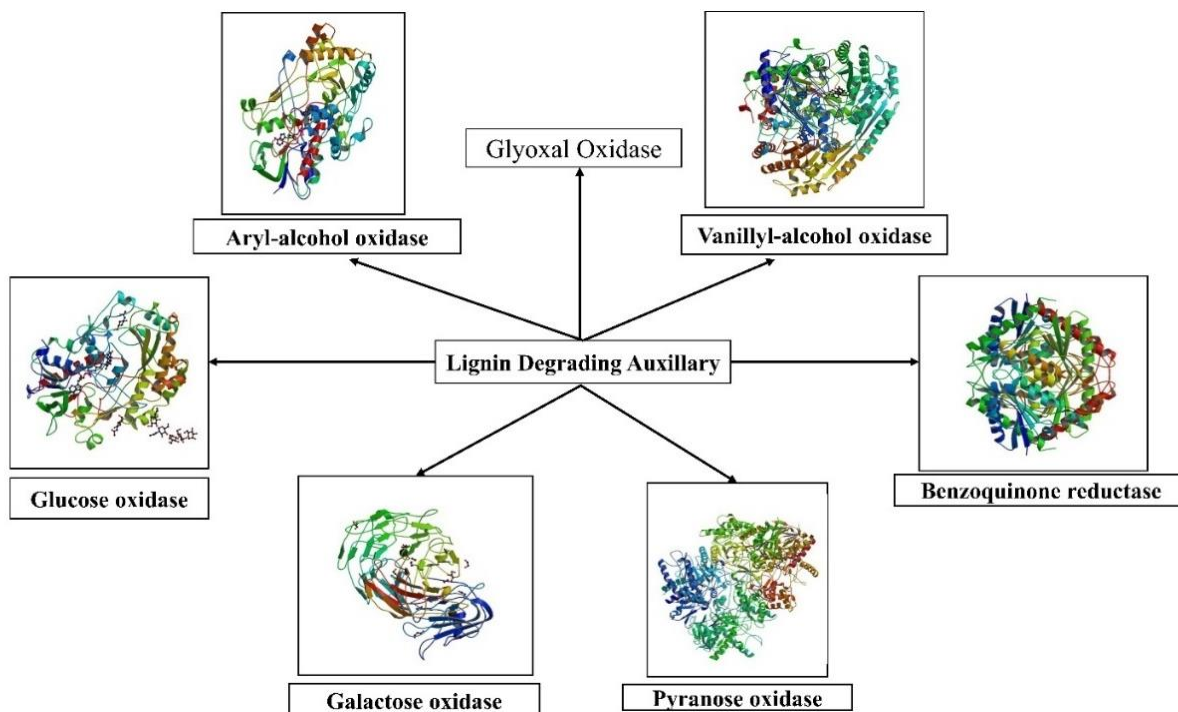


Figure 1.7: Schematic representation of lignin degrading auxiliary enzymes namely, aryl alcohol oxidase (PDB ID: 3FIM) [145], vanillyl alcohol oxidase (PDB ID:1W1J) [146], glucose oxidase (PDB ID: 1CF3) [147], galactose oxidase (PDB ID: 2WQ8) [148], pyranose oxidase (PDB ID: 4MIF) [149]. benzoquinone reductase (PDB ID: 4LA4). All the above enzyme structures were obtained from PDB RCSB repository.

1.3.1. Aryl Alcohol Oxidase:

Aryl alcohol oxidase (AAO) (EC.1.1.3.7) was first observed in *Polystictus versicolor* (or *Trametes versicolor*) during the 1960s. Aryl alcohol oxidase was detected and characterized in white rot basidiomycetes such as *Pleurotus* species (*P. eryngii*), *Bjerkandera adusta* and a few ascomycetous fungi [150-154]. White rot fungi were found to be involved in efficient degradation of lignin, aryl alcohol oxidase was found to be involved in lignin depolymerization process by generating H_2O_2 and fueling ligninolytic peroxidases [150]. AAO is an FAD containing enzyme belonging to the glucose-methanol-choline oxidase (GMC) family of oxidoreductases. It was reported that AAO of *Pleurotus eryngii* was found to be involved in generation of peroxide by redox cycling of *p*-anisaldehyde (a fungal extracellular metabolite), in addition AAO also was found to be involved in oxidation of polyunsaturated primary alcohols [155]. Redox cycling of *p*-methoxylated benzylic metabolites by *P. eryngii* takes places through an oxygen activation reaction by AAO. Amino acid sequence

comparisons of AAO revealed homology with glucose oxidase. AAO genes from *P. eryngii* and *Pleurotus pulmonarius* were cloned and sequenced [156]. For several years only the AAO sequence from *P. eryngii* was available, however recent advancements in genome sequencing and the sequencing of basidiomycetes genomes has revealed the sequence of around 40 AAO sequences and 112 GMC (glucose-methanol-choline oxidases) superfamily sequences were reported [157]. Kinetic isotope studies have showed that alcohol oxidation by AAO occurred by hydride transfer to the flavin domain and then hydroxyl proton transfer to the base [157]. At the same time site directed mutagenesis studies of AAO have showed that His502 is involved in activation of alcohol substrates by proton abstraction, this mechanism was later extended to other GMC oxidoreductases [157].

Mechanism: Structural and functional studies of AAO isolated from *P. eryngii* show that it has a variety of substrates, catalyzing the oxidation of primary and polyunsaturated alcohols [155]. The overall reaction mechanism of AAO can be divided into an oxidative and a reductive reaction, first AAO catalyzes the oxidative dehydrogenation of the substrate (reductive reaction) later the flavin adenine dinucleotide is reoxidized by molecular oxygen, generating H₂O₂ (Figure 1.8) [150]. Comprehensive studies of the substrate specificities of AAO revealed that it catalyses the oxidation of aromatic alcohols such as *p*-anisyl alcohol and aliphatic polyunsaturated primary alcohols to their corresponding aldehydes [155]. It was reported that phenolic hydroxyls strongly inhibits the enzymatic activity of AAO. The redox cycling of *p*-Anisaldehyde (important extracellular metabolite of *P. eryngii*) involves intracellular aryl-alcohol dehydrogenase along with AAO which results in hydrogen peroxide generation (Figure 1.8) [158]. AAO seems to have a similar catalytic mechanism to choline oxidase (GMC oxidoreductase family) which catalyzes the oxidization of alcohol substrates resulting in the production of aldehydes. Earlier studies on AAO of *P. eryngii* shows that it catalyzes the conversion of primary alcohols of varied structural properties. AAO exhibits a wide range of electron donor substrate specificity by catalyzing the oxidation of aromatic and π -system containing primary alcohols such as benzylic alcohol, naphthyl alcohol and aliphatic polyunsaturated alcohols [150, 155, 159]. The π -systems cause an increase in electron availability at the benzylic position causing hydride abstraction by the flavin N5 atom. The structural of the AAO active site prevents the oxidation of secondary alcohols as they cannot be accommodated at the appropriate distance from the catalytic histidine and flavin N5 atom due to the presence of Phe501 [160]. Bisubstrate kinetic analysis with different benzylic alcohols shows the overall AAO catalytic cycle is highly influenced by the nature of

substituents on the benzene ring. AAO catalysis is divided into a reductive and oxidative reactions, when it is treated with electron withdrawing substrates such as 3-chloro and 3-fluorobenzyl alcohols both the half reactions become independent resulting in aldehyde product dissociation before the oxygen reaction by a ping-pong steady state mechanism [161]. In electron donor substituents such as methoxylated benzyl alcohols, oxygen reacts with the reduced AAO-aldehyde complex resulting in a ternary complex prior to aldehyde product release [161]. The catalytic cycle of AAO depends on the stacking and stabilizing interactions of aromatic substrate and product at the active site Tyr92 residue (involved in stabilization of alcohol substrate) which occur by switching between ternary and ping-pong mechanisms [161].

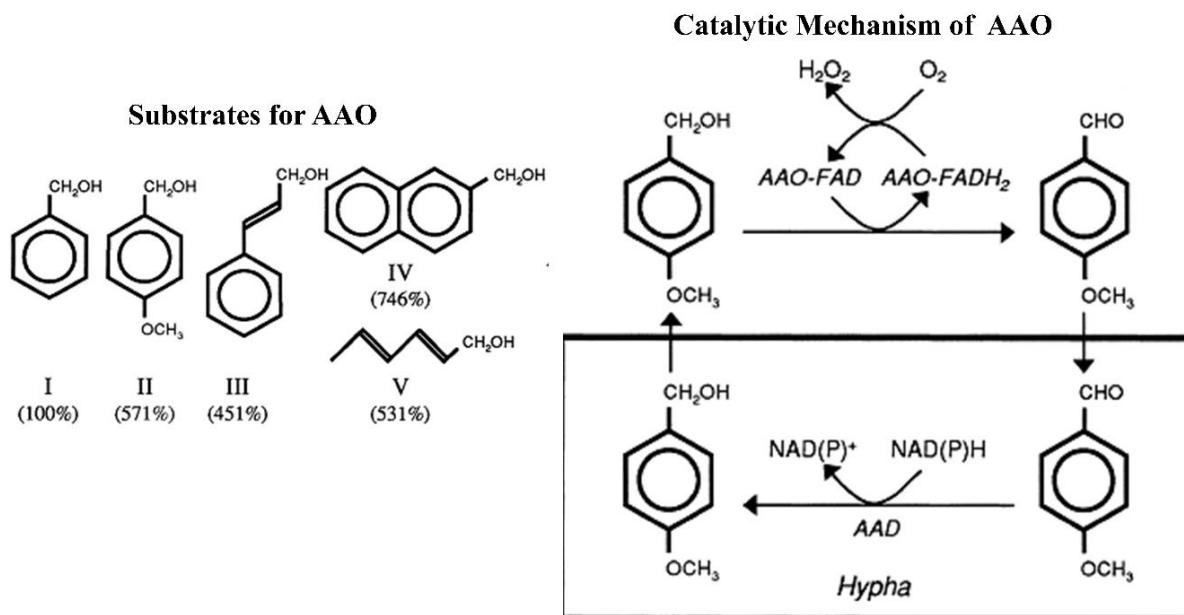


Figure 1.8: Chemical structure of various substrates of *Pleurotus* AAO (I, benzyl alcohol; II, p-anisyl alcohol; III, cinnamyl alcohol; IV, 2-naphthalenemethanol; and V, 2,4-hexadien-1-ol) and relative activity estimated as O_2 consumption [155]. Scheme for H_2O_2 production by anisaldehyde redox-cycling involving extracellular AAO and intracellular AAD. Reprinted with permission from ref [162], Copyright © 2000, Elsevier.

1.3.2. Vanillyl Alcohol Oxidase:

Vanillyl alcohol oxidase (EC.1.1.3.38) is a flavin containing protein which was first isolated from *Penicillium simplicissimum* based on its ability to oxidize vanillyl alcohol to vanillin, 4(methoxymethyl) phenol to 4-hydroxybenzaldehyde [163]. It was also studied for its ability to degrade lignin. Vanillyl alcohol oxidase (VAO) can convert phenolic compounds by

different catalytic processes such as oxidation, deamination, demethylation, hydroxylation and dehydrogenation [164]. The reaction mechanism of VAO for oxidation of 4(methoxy methyl) phenol involves a primary transfer of hydride from the substrate to the flavin leading to the formation of a two electron reduced enzyme complex with a *p*-quinone methide compound as an intermediate. Further the reduced flavin is reoxidized by oxygen associated with hydration of *p*-quinone methide [165] VAO is an industrially important enzyme for the production of the compounds: vanillin, 4-hydroxybenzaldehyde, coniferyl alcohol and pure phenolic derivatives [164].

Mechanism: Based on spectroscopic and kinetic studies it was shown that substrate oxidation commences via direct hydride transfer from the C α atom to N5 of flavin adenine dinucleotide. As a result a *p*-quinone methide (intermediate) is formed which is further activated by the preferential binding of the phenolate form of the substrate, this is supported by the three dimensional structure of the VAO [166]. Studies of VAO binding with VAO-isoeugenol, VAO-2-nitro-*p*-cresol complexes shows that VAO achieves hydride transfer from the C α atom 3.5Å^o from N5 atom. The hydroxyl oxygen is bound to three residues: Arg504, Tyr503 and Tyr108 through hydrogen bonds which stabilize the negative charge of the phenolate ion [166]. Under anaerobic conditions, VAO reaction with 4-methoxymethyl phenol results in a stable reduced enzyme-*p*-quinone methide complex, however the final product is synthesized and released immediately after exposure to oxygen, following FAD reoxidation. From three dimensional structures of VAO, it is suggested that charge stabilizations between the flavin, quinone intermediate and Arg-504 regulate the catalytic cycle. Besides its role in interacting with the phenolate oxygen, Arg-504 is involved in balancing the negative charge on the N1-C2=O₂ locus of the anionic reduced cofactor. The C2 atom of flavin deviates from its expected position due to the oxygen atom of *p*-quinonemethide molecule binding to the reduced enzyme [166]. Thus in the reduced enzyme, the negative charge of the flavin C2 atom causes electrostatic repulsion which prevents the formation of a phenolate ion resulting in the stabilization of the quinone intermediate form. Upon reoxidation of flavin, Arg-504 lacks an anionic partner which triggers the development of negative charge on the oxygen atom of the quinone group. The electrophilicity of the methide carbon is increased enabling hydroxylation of 4-methoxymethyl phenol or deprotonation of the intermediate (vanillyl-alcohol) thus generating the final product (Figure 1.9) [166, 167].

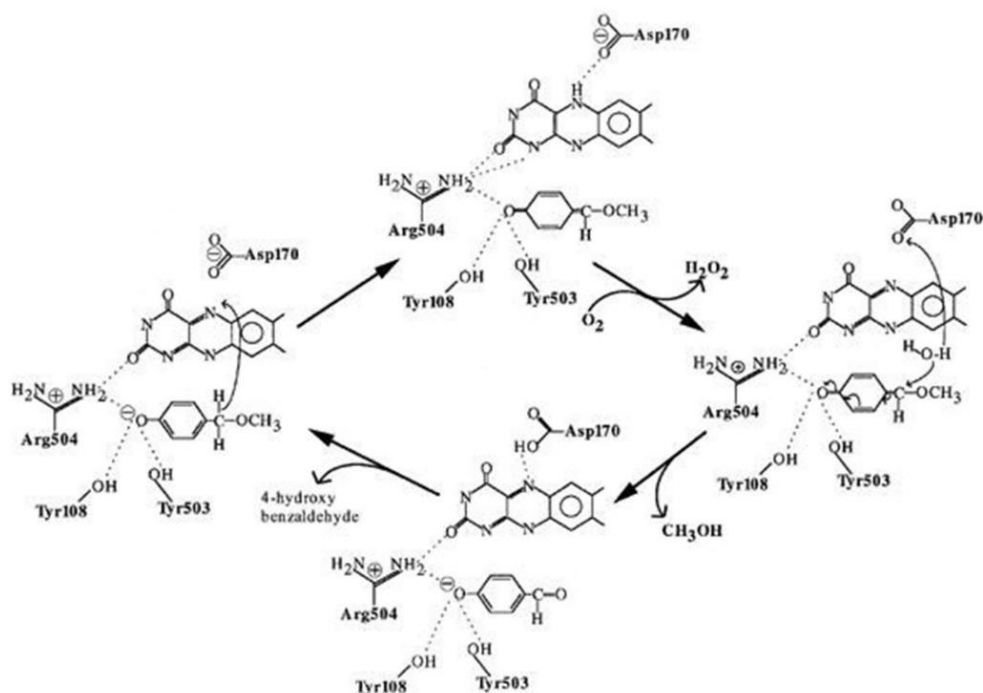


Figure 1.9: The reaction mechanism for the oxidation of 4-(methoxymethyl) phenol. In the first step, the substrate is oxidised via a direct hydride transfer from the substrate C α atom to the N5 of flavin. The reduced cofactor is then reoxidised by molecular oxygen with the production of a hydrogen peroxide molecule. In the next step, the p-quinone-methoxymethide intermediate is hydroxylated by a water molecule, possibly activated by Asp170. The resulting 4-hydroxy-benzaldehyde and methanol products are released. Reprinted with permission from ref [166], Copyright © 1997, Elsevier.

1.3.3. Glyoxal Oxidase:

Glyoxal oxidase an extracellular hydrogen peroxide producing enzyme secreted by lignolytic cultures of *P. chrysosporium* [168]. Glyoxal oxidases catalyzes the oxidation of wide range of aldehydes and α -hydroxyl carbonyl compounds by reducing O_2 to H_2O_2 , thus glyoxal oxidase fuels the process of lignin degradation by generating H_2O_2 which is used by ligninolytic peroxidases (such as lignin peroxidase, manganese peroxidase) [168, 169]. Glyoxal ($OHCCHO$) and methylglyoxal (CH_3COCHO) are two well known substrates for glyoxal oxidase in the extracellular fluids of lignolytic cultures [170].

Mechanism: Studies have revealed that *P. chrysosporium* secretes three extracellular enzymes: lignin peroxidase, manganese peroxidase and glyoxal oxidase. Glyoxal oxidase fuels the complete ligninolytic mechanism by generating extracellular H_2O_2 , which is required for the functioning of ligninolytic peroxidases [171]. Pure glyoxal oxidase is inactive,

however it is activated by peroxidases and peroxidase substrates) [171]. Though glyoxal and methylglyoxal (substrates) of glyoxal oxidase were observed in ligninolytic cultures, there are other substrates such as formaldehyde, acetaldehyde, glycoaldehyde, glyoxalic acid, dihydroxyacetone, glyceraldehyde. In addition, downstream lignin degradation products act as substrates for glyoxal oxidase [171]. Glyoxal oxidase has an efficient sequential oxidation process by converting glycolaldehyde to oxalate (glycoaldehyde → glyoxal → glyoxalate → oxalate) The catalytic mechanism behind oxidation of aldehydes by glyoxal oxidase is not known, however it was suggested that it oxidizes substrates similarly to galactose oxidase.

1.3.4. Pyranose Oxidase:

Pyranose oxidase (EC.1.1.3.10; oxygen 2-oxidoreductase) is a hydrogen peroxide producing enzyme. Pyranose oxidase catalyzes the oxidation of the C-2 of several aldopyranoses, D-glucose is a ideal substrate for the enzyme [172-174]. Several fungi belonging to basidiomycetes and particularly members of the order *Aphyllorphorales* secrete extracellular pyranose oxidase [175]. The structure and catalytic mechanism of pyranose oxidase were extensively studied in *Trametes multicolor* fungi,. Its amino acid sequence suggests that it belongs to the glucose-methanol-choline (GMC) family of flavin adenine dinucleotide (FAD) dependent oxidoreductases [176]. Pyranose oxidase is a large flavoproteins which can oxidize a number of monosaccharides at their carbon-2 position in the presence of molecular oxygen, producing 2-keto sugars and hydrogen peroxide [175].

Mechanism: Pyranose oxidase is an hydrogen peroxide generating enzyme which catalyzes the oxidation of D-glucose and other aldopyranoses at the C-2 position resulting in the production of 2-keto sugars it was also found to be involved in lignin depolymerization. It catalyses the regioselective oxidation of different aldopyranoses at their C-2 position using molecular oxygen resulting in 2-keto aldoses and H₂O₂. The whole reaction can be divided into an oxidative and a reductive reaction, in the reductive half reaction the sugar is oxidized to a keto sugar followed by reduction of FAD. The oxidative reaction involves the reduction of O₂ to H₂O₂ and reoxidation of the FAD [175, 177]. Pyranose oxidase also oxidizes certain compounds at the C-3 position such as 2-deoxy-D-glucose, 2-keto-D-glucose and methyl-β-D-glucosides [178, 179]. The (k_{cat}/K_m) is highest for β-D-glucose. Studies have reported that pyranose oxidase also oxidizes monosaccharides such as D-xylose, D-galactose and L-arabinose (constituents of hemicellulose) with lower catalytic efficiencies, which may extend the enzymes ability to generate hydrogen peroxide from the lignincellulose derived sugars.

The optimum pH of the enzyme varies based on the type of electron acceptors used ie oxygen, various quinones and radicals. Quinones and radicals are the best substrates of pyranose oxidase, suggesting its role in lignin depolymerization is as a hydrogen peroxide generating and quinone reducing enzyme. It was reported that pyranose oxidase from *Phlebiopsis gigantea* has the ability to hydrolyze β 1 \rightarrow 4 linked disaccharides (cellobiose and lactose) and α 1 \rightarrow 4 linked disaccharides (such as maltose) to the corresponding monosaccharides at their C2 position [180]. β glycosides of higher alcohols such as hexyl, phenyl, o-nitrophenyl and p-nitrophenyl) are converted to disaccharides by pyranose oxidase through a glycosyl transferase reaction [180].

1.3.5. Galactose Oxidase:

Galactose oxidase (EC 1.1.3.9) an extracellular enzyme secreted by *Fusarium spp.* Galactose oxidase is a monomeric enzyme containing a single copper ion, catalyzing the oxidation of primary alcohol substrates (D-isomers) such as D-galactose and other polysaccharides containing D-galactose on their reducing ends resulting in the production of aldehydes and hydrogen peroxide [181, 182]. Galactose oxidase belongs to the alcohol oxidoreductase family (also known as alcohol oxidase), enzymes belonging to this generally use molecular oxygen as electron acceptors for generating hydrogen peroxide [183]. Most alcohol oxidoreductases are flavoproteins that use FAD⁺ as primary electron acceptors, however some of these enzymes are copper radical containing oxidases (CROs) such as galactose oxidase, glyoxal oxidase and hexose-1-oxidase [183].

Mechanism: Galactose oxidizes primary alcohols resulting in the production of aldehydes and hydrogen peroxide. This is a two electron reaction with only one copper ion at the active site and a second redox active center, a tyrosine residue. Tyr-272 also acts as ligand to the copper ion [184]. The catalytic mechanism of galactose oxidase can be divided into two reactions a) proton transfer from the O-6 position of galactose to the axial tyrosine anion (hydrogen atom transfer) then from the C6 of galactose to the Tyr-Cys radical cofactor followed by electron transfer from the carbohydrate, generating an aldehyde and Cu⁺ [183]. In the second half of the reaction electron transfer continues from Cu⁺ to oxygen by producing superoxide then through hydrogen transfer, a proton is transferred from the phenolic hydroxyl group of the Tyr-Cys cofactor to superoxide, producing a metal bound hydroperoxide. The final proton transfer from the axial tyrosine to hydroperoxide generates hydrogen peroxide and Cu²⁺ (resting state of the enzyme) [183].

1.3.6. Glucose Oxidase:

Glucose oxidase (E.C.1.1.3.4) is an important H_2O_2 generating oxidoreductase produced by ligninolytic cultures of *P. chrysosporium*. Glucose oxidase catalyzes the oxidation of β -D-glucose to gluconic acid, using molecular oxygen (as electron acceptor) thus producing H_2O_2 [185, 186]. Glucose oxidase has several commercial applications such as increasing the quality of food materials (color, flavor and shelf life), oxygen removal from fruit juices and canned food etc [186]. Apart from these applications, glucose oxidase also inhibits different food-borne pathogens such as *Salmonella infantis*, *Staphylococcus aureus*, *Clostridium perfringens*, *Bacillus cereus*, *Campylobacter jejuni* and *Listeria monocytogens* [187].

Mechanism: The reaction of GOD can be divided into an oxidative step and a reductive step. The reductive step of GOD oxidizes β -D-glucose to D-glucono- δ -lactone which is further hydrolyzed to gluconic acid (non-enzymatically). In *A. niger* a lactonase catalyzes the hydrolysis of D-glucono- δ -lactone to gluconic acid. It also reduces the FAD domain of GOD to $FADH_2$ [188]. Reduced GOD is re-oxidized by molecular oxygen to H_2O_2 in the the oxidative reaction, H_2O_2 from the above rection is cleaved by catalase producing water and oxygen [189]. The flavin domains of GOD are involved in the redox reaction, during the oxidative reaction of GOD and electrons from electron donors are transferred to the isoalloxazine nucleous of flavin doman (FMN) and then to the electron acceptor [190]. GOD catalyzes the reaction by transferring the electrons from glucose to oxygen, producing H_2O_2 , thus placing GOD in the oxidoreductase class of enzymes. Overall enzyme catalysis of GOD depends on oxidation and reduction reaction steps of its flavin group (FAD) primarily glucose reduces the FAD to $FADH_2$ by producing gluconic acid (product) without forming free radical containing semiquinone (intermediate). At the same molecular oxygen (electron acceptor) reduces the $FADH_2$ back to FAD generating H_2O_2 as a product (Figure 1.10) [190].

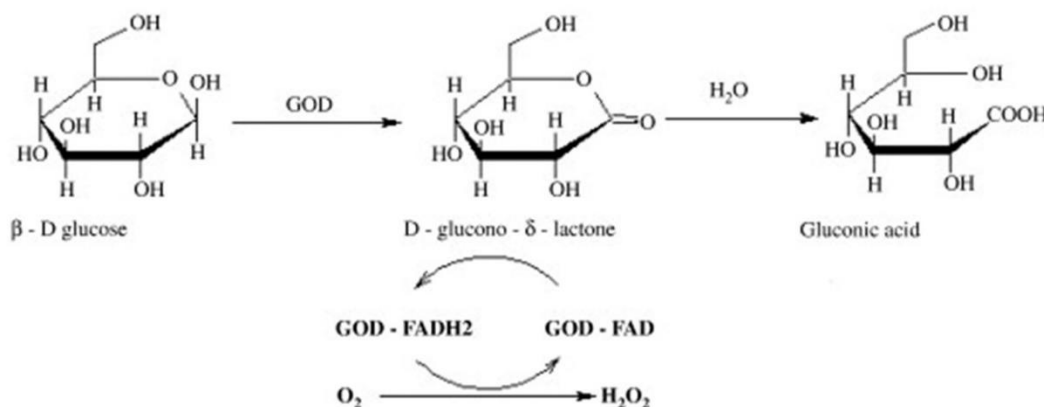


Figure 1.10: Reaction mechanism of glucose oxidase (GOD) [188]. Reprinted with permission from ref [185], Copyright © 2009, Elsevier.

1.3.7. Benzoquinone Reductase:

The 1,4-Benzoquinone reductase (EC.1.6.5.6) is an intracellular enzyme which was purified and characterized from the agitated cultures of *P. chrysosporium*. 1,4-Benzoquinone reductase was expressed in both nitrogen sufficient and limited conditions [191, 192]. *P. chrysosporium*, one of the highly studied lignin degrading fungi, secretes two classes of ligninolytic peroxidases: lignin peroxidase (LiP) and manganese peroxidase (MnP) along with several H₂O₂ generating enzymes. These enzymes catalyze the primary steps of lignin depolymerization resulting in a wide variety of intermediate products such as substituted quinones, hydroquinones, benzaldehydes and other ring opened fragments. Methoxylated lignin derived quinones are reduced by intracellular quinone reductases [191, 192].

Mechanism: Benzoquinone or quinone reductases are significant enzymes secreted by several fungi especially *P. chrysosporium*. 1,4-Benzoquinone reductase is a NADPH dependent intracellular enzyme, it contains flavin mononucleotide (FMN). 1,4-Benzoquinone reductase was active during both primary and secondary metabolism but the enzyme inducers are stronger during the primary metabolic processes [191-193]. Studies showed that when vanillate or methoxy-*p*-quinone are added to cells, carrying out primary metabolism, enzyme expression was increased. However, the effect was small when the same compounds were added to secondary metabolic cells, which suggests that quinone reductase is regulated independently of lignin and manganese peroxidase [191]. LiP and MnP are expressed only during the secondary metabolic stage of the growth and their expression is not induced by aromatic substrates. The regulation of quinone reductases is similar to that of vanillate hydroxylase, which suggests its involvement in vanillate metabolism [191-193]. Quinone reductase was expressed during the lignolytic phase of *P. chrysosporium*, suggesting a role in the reduction of quinones generated during lignin degradation. It was reported that quinone reductases are induced upon quinone addition, suggesting the involvement of quinone reductase in lignin and quinone degradation [191-193]. Besides degrading of quinone and lignin derived compounds, it is also reported that quinone reductases protect *P. chrysosporium* from oxidative stress by acting as redox active toxins. Quinones obtained by metabolic conversion are reduced by one electron generating semiquinone radicals, which are then oxidized by oxygen generated superoxide anion, this superoxide anion is further

converted to H₂O₂ through superoxide dismutase and later in the presence of suitable electron donors it results in production of highly reactive hydroxyl radicals from H₂O₂ [191-193]. The detailed mechanisms of the regulation and catalytic mechanism of quinone reductase need to be explored (Table 1.3).

Table 1.3: Catalytic mechanisms and structural studies of different lignin degrading auxiliary enzymes (LDA).

Enzyme and FOLy class	Catalytic mechanism	Structural studies and Reference
Aryl alcohol oxidase (LDA1) EC 1.1.3.7	Aromatic primary alcohol + O ₂ → Aromatic aldehyde + H ₂ O ₂	<i>Pleurotus eryngii</i> [145, 194-196]
Vanillyl alcohol oxidase (LDA2) EC 1.1.3.38	Vanillyl alcohol + O ₂ → vanillin + H ₂ O ₂	<i>Penicillium simplicissimum</i> [163, 166, 197-199]
Glyoxal oxidase (LDA3) EC 1.1.3.-	Glyoxal oxidase catalyzes oxidation of wide range of simple aldehydes, α-hydroxy carbonyl compounds by producing hydrogen peroxide	<i>Aspergillus nidulans</i> , <i>Pichia pastoris</i> , <i>Phanerochaete chrysosporium</i> [169, 200, 201]
Pyranose oxidase (LDA4) EC 1.1.3.10	FAD _(Oxidized) + D-Glucose → FAD _(Reduced) + 2-keto-D-glucose FAD _(Oxidized) + O ₂ → FAD _(Reduced) + H ₂ O ₂	<i>Trametes multicolor</i> [202-205], <i>Peniophora</i> sp. [206, 207] <i>Phanerochaete chrysosporium</i> [149]
Galactose oxidase (LDA5) EC 1.1.3.9	D-Galactose + O ₂ → D-Galacto-hexodialdose + H ₂ O ₂	<i>Aspergillus nidulans</i> , <i>Pichia pastoris</i> , [181-184, 208, 209]
Glucose oxidase (LDA6) EC 1.1.3.4	β-D-glucose + O ₂ → D-glucono-1,5-lactone + H ₂ O ₂	<i>Aspergillus niger</i> [210, 211]
Benzoquinone reductase (LDA7) EC 1.6.5.6	NADPH + H ⁺ + p-benzoquinone → NADP ⁺ + hydroquinone	<i>Phanerochaete chrysosporium</i> [191-193]

1.4. Plant Cell-Wall Deacetylating Carbohydrate Esterases:

1.4.1. Acetyl Xylan Esterases: Acetyl and methyl esterifications are two major naturally found substitutions in the plant cell wall polysaccharides. The non-cellulosic plant cell wall polysaccharides such as pectin and hemicellulose are differentially esterified by the *O*-acetyl and methyl groups to cease the action of various hydrolytic enzymes secreted by different fungi and bacterial species. Thus, microorganisms have emerged with a special class of enzymes known as carbohydrate esterases. Carbohydrate esterases catalyze *O*-de, *N*-deacetylation of acetylated saccharide residues (esters or amides, where sugars play the role of alcohol /amine / acid). Carbohydrate active enzyme (CAZy) database, has classified carbohydrate esterases into 16 classes, of which hemicellulose deacetylating carbohydrate esterases (CE) were grouped into 8 classes (CE-1 to CE-7, CE-16). Various plant biomass degrading fungi and bacteria secrete acetyl xylan esterases, however these enzymes exhibit varied substrate specificities. Acetyl xylan esterases and xylanases coupled pre-treatment methods exhibit significant applications such as enhancing animal feedstock, baking industry, production of food additives, paper and pulp, xylitol production and biorefinery industries-based industries respectively. Thus, understanding the structural and functional properties of acetyl xylan esterase will significantly aid in developing the efficient acetyl xylan esterases with wide range of industrial applications.

Acetyl xylan esterases (AXE) are secreted by microorganisms for the deacetylation of xylan polymers and xylooligosaccharides (E.C. 3.1.1.72) [212]. Acetyl xylan esterases are widely distributed among different carbohydrate esterase classes (CE-1 to CE-7). Biely et al (1985), have found the occurrence of acetyl xylan esterase in the cellulolytic and hemicellulolytic microorganisms [213]. After that several studies have reported the occurrence of the acetyl xylan esterase as they have found to act on acetyl glucuronoxylan, however later it was found that AcXE is active on other acetylated polysaccharides other than xylan [212]. The microbial endoxylanases were found to work in synergy with carbohydrate esterase, especially the activity of endoxylanases on acetyl xylan increased with the presence of acetyl xylan esterases [214]. Acetyl xylan (*O*-acetyl-4-*O*-methyl-Dglucurono- D-xylan) the naturally occurring form of hemicellulose in hardwood, alternate xylopyranosyl units of the polymeric xylan contains one acetyl group [212, 214]. Selig et al (2008), have showed that combinatorial action of endoxylanases and acetyl xylan esterases have significantly improved the hydrolysis of xylan polymer [215]. Selig et al (2009) have reported that when AXE's were

used in combination with endoxylanases on different corn stover substrates, a linear relationship in removal of acetyl groups and depolymerization of xylan was observed [216]. The hydrolysis of xylan from the pretreated wheat straw and giant reed using xylanolytic enzymes and AXE was conducted by Zhang et al (2011). This study has showed clearly that removal of acetyl groups by AXE enhanced the accessibility of xylan by the xylanolytic enzymes, and solubilization of xylan have progressively increased the availability of cellulose to cellulases, resulting in hydrolysis of cellulose [10]. Cellulolytic and xylanolytic enzymes function in synergistic effect and the combined use of cellulases, xylanases and AXE has resulted in higher hydrolysis of cellulose revealing the occurrence of acetylated xylan in the cellulose matrix [10]. In this article, we have extensively reviewed the structural and functional properties of acetyl xylan esterases occurring in different carbohydrate esterase classes (Table 1.5).

Table 1.5: Different classes of carbohydrate esterase (CE) family and their corresponding representing enzymes with note on their protein 3D structure:

CE- Class	Representing Enzymes	E.C. Number	3D Structure Status
CE-1	acetyl xylan esterase, cinnamoyl esterase, feruloyl esterase, carboxylesterase, S-formylglutathione hydrolase, diacylglycerol O-acyltransferase, trehalose 6-O-mycolyltransferase	(EC 3.1.1.72) (EC 3.1.1.-) (EC 3.1.1.73) (EC 3.1.1.1) (EC 3.1.2.12) (EC 2.3.1.20) (EC 2.3.1.122)	(α/β / α)-sandwich
CE-2	acetyl xylan esterase	(EC 3.1.1.72)	α/β + β -sheet
CE-3	acetyl xylan esterase	(EC 3.1.1.72)	($\alpha/\beta/\alpha$)-sandwich
CE-4	acetyl xylan esterase, chitin deacetylase, chitooligosaccharide deacetylase, peptidoglycan GlcNAc deacetylase, peptidoglycan N-acetylmuramic acid deacetylase	(EC 3.1.1.72) (EC 3.5.1.41) (EC 3.5.1.-) (EC 3.5.1.-) (EC 3.5.1.-)	(β/α) 7 barrel
CE-5	acetyl xylan esterase, Cutinase,	(EC 3.1.1.72)	($\alpha/\beta/\alpha$)-sandwich
CE-6	acetyl xylan esterase	(EC 3.1.1.72)	($\alpha/\beta/\alpha$)-sandwich
CE-7	acetyl xylan esterase, Cephalosporin-C deacetylase	(EC 3.1.1.72) (EC 3.1.1.41)	($\alpha/\beta/\alpha$)-sandwich

1.4.2. Pectin Carbohydrate Esterases: Enzymes required for the breakdown of pectin can be majorly classified into three categories as protopectinases (involved in breaking insoluble protopectin and results in soluble polymerized pectin), depolymerizing enzymes (required

for breaking down $\alpha(1\rightarrow4)$ glycosidic linkages of pectin) and esterases (required for the de-esterification and de-acetylation of pectin) [217]. Pectinolytic enzymes are widely observed among plants, bacterial and fungal species and most of the pectin methyl esterases can be divided based on their optimum pH, bacteria and plant PME exhibit an optimum pH range between 6 to 8 and PME secreted by fungi exhibit a pH 4 to 6 [217, 218]. Pectin degrading enzymes have attained high commercial importance since early 1930's in wine and fruit juice industries, pectinolytic enzymes secreted by *Aspergillus* species is highly used in industries [219]. Pectin present in vegetable tissues and majorly in fruits, contains complex hetero polysaccharides at a molecular weight ranging between 25 to 360kDa. Calcium and magnesium pectate forms the major constituent of the plant cell walls especially in middle lamella [218]. The gelling property of pectin majorly employed in the food industries is directly dependent on its degree of esterification, pectin's with higher degree of esterification gel around pH 3.0 in the presence of sugar, whereas pectin's with low degree of esterification gels in the presence of calcium ions under wide pH ranges and with or without sugar. [217, 220, 221] (Table 1.6).

Pectin methyl esterases or alkaline reagents were majorly used to demethoxylate large galacturonic chain for reducing the overall pectin methoxylation content [217]. Majorly pectinolytic enzymes were highly applied in fruit juice and wine industries for the clarification of the fruit juices, modification of fruits and vegetables [222]. Apart from these applications, pectinolytic enzymes were also used for extracting oils from germ, palm, coconut, sunflower seed and kernel rape seeds, by replacing the conventionally used carcinogenic solvents like hexane. These pectinolytic enzymes extract oil from different crops by liquefying the structural components of the cell walls. Commercial pectinase preparations called Olivex® were applied in olive oil industries for the extraction of oil and to increase the quality [223, 224]. Rhamnogalacturonan a complex polysaccharide unit present in the primary cell walls and middle lamella of higher plants, with alternating rhamnose and galacturonic acid residues acetylated majorly at C-2 and C-3 positions [225] As the acetylation of these residues sterically hinders the catalytic function of the corresponding lyases and hydrolases on the glycosidic linkages thus deacetylation facilitates the action of the lyases and hydrolases. Thus, rhamnogalacturonan acetyl esterase belonging to CE-12 family has gained significance in deacetylation of these residues and also been used industrially for the production of β -lactam antibiotics and paper bleaching purposes [226] (Table 1.6).

Table 1.6: Different types of pectic substances and pectinolytic enzymes responsible for its degradation:

S.no	Different types of pectic substances
1	Protopectin: Protopectin is present in the inner tissues of plant cell walls which is insoluble in water. Upon restricted hydrolysis yields pectin or pectic acids.
2	Pectic acid: Pectic acids are soluble pectic substances (galacturonans) with lesser number of methoxyl groups. Normal and pectic acid salts are called as pectates
3	Pectinic acids are long polygalacturonans with <75% methylated galacturonate units, salts of pectinic acids are called pectinates
4	Pectin: (or) polymethyl galacturonate is a polymeric material with 75% of the carboxyl groups are esterified with methanol. Pectin provides rigidity to the plant cell walls.
Pectinolytic enzymes	
1	Pectin methyl esterases: These esterases catalyzes the de-esterification pectin by releasing methoxy esters, resulting in pectic acids and methanol.
2	Pectin Depolymerizing Enzymes: a) Protopectins are enzymatically hydrolyzed by set of enzymes called as protopectinases (PPase). PPase are classified into two types a) A-type PPase, which reacts with polygalacturonic acid regions and b) B-type PPase reacts with the polysaccharide chains on outer region. (Protopectin (insoluble) + H₂O--(PPase)--> Pectin (soluble)) b) The pectin depolymerizing enzymes can be majorly classified as hydrolases divided into: Endo and Exo polygalacturonases such as (Exo-polygalacturonan-digalacturono hydrolase, Oligo galacturonate hydrolase, Delta 4:5 Unsaturated oligo galacturonate hydrolases, Endo-polymethyl-galacturonases, Endo-polymethyl-galacturonases). Lyases which majorly contains enzymes such as Endo and Exo polygalacturonase lyases.

Lignin-Carbohydrate De-Esterases: Plants contain a range of hydroxycinnamic acids like caffeic, *p*-coumaric, ferulic and sinapic acids which can be broadly classified as phenolic compounds and highly abundant among foods [227]. Feruloyl esterases (or ferulic acid esterases) or cinnamoyl esterases are carbohydrate esterase class of enzymes which hydrolyze the ester linkages between hydroxycinnamic acids and plant cell wall carbohydrates by releasing ferulic and cinnamic acid [228]. Combinatorial usage of feruloyl esterase or cinnamoyl esterases with glycoside hydrolases for the liberation of free carbohydrate residues and phenolic acids can significantly aid in different preprocessing steps of biofuel and biorefining industries [229]. The activity of FAE and CAE are chiefly limited to the position and conformations of the feruloyl groups present in the feruloylated polysaccharides and other surrounding cell wall components. Recent studies conducted by Faulds et al. (2003, 2006) have revealed the preferential partnership between glycoside hydrolase class-11 (GH-11) xylanases and FAE for liberating ferulic acid from the insoluble biomass, while partnership between GH-10 xylanases and FAE will liberate 5,5'-dimers [230,

231]. Glucuronoyl esterases are the class of carbohydrate esterases involved in hydrolysis of the ester linkages present between 4-O-methyl-d-glucuronic acid residues of glucuronoxylans and aromatic alcohols of lignin [232]. Glucuronoyl esterase which is involved in plant cell wall degradation was discovered in *Schizophyllum commune* for the first time [232]. Duranova et al. (2009) have purified and characterized using a series of synthetic substrates containing methyl esters of uronic acids and their glycoside derivatives [233]. These studies have revealed the specificity of GE towards 4-O-methyl-D-glucuronic acid, its methyl esters and D-glucuronic acid containing 4-nitophenyl aglycon, showing that GE attack the ester bonds between 4-O-methyl-D-glucuronic acid of glucuronoxylan and alcohols of lignin [233, 234]. Glucuronoyl esterase finds its applications in growing biofuel and biorefinery industries as it breaks down and separates the hemicellulose and lignin.

Pretreatment step is currently being used in bioethanol industries for releasing free carbohydrate residues from the other aromatic components of the cell wall. The heterophenolic lignin compounds interacting with the polysaccharide units increases the recalcitrant nature of the plant cell wall and the percentage of lignin in plant tissues is directly proportional to its digestibility. It has been assumed that FAE's and CAE's are required for breaking the lignin and carbohydrate linkages. According to Benoit et al. (2006), type-C and type-B FAE's isolated from *Aspergillus niger* release higher proportions of ferulic acid and *p*-coumaric acid from the steam exploded wheat straw [235]. The Type-A FAE from *A. niger* was found to be effective against the steam exploded wheat straw in the presence of cellulases and xylanases, and at 50°C the rate hydrolysis increased significantly [236]. Similarly, Selig et al. (2008), have used a combination of cellobiohydrolase Cel7A, xylanase, feruloyl esterase and acetyl xylan esterase and reported enhanced breakdown of the hot-water treated corn stover cellulose [215]. Apart from its long list of applications, FAE's were also used for the utilization of straws in paper industries [237, 238], detoxification of animal feed [239], for the removal of cinnamic acids and *p*-coumaric acids from coffee pulp [235].

References:

1. Daniel G. Fungal and bacterial biodegradation: white rots, brown rots, soft rots, and bacteria. Deterioration and Protection of Sustainable Biomaterials: ACS Publications; 2014. p. 23-58.
2. Kendrick B. Fungi: ecological importance and impact on humans. eLS. 2001.
3. Kameshwar AKS, Qin W. Lignin Degrading Fungal Enzymes. Production of Biofuels and Chemicals from Lignin: Springer; 2016. p. 81-130.
4. Rowell RM, Barbour RJ. Archaeological wood: properties, chemistry, and preservation: ACS Publications; 1989.
5. Daniel G. Use of electron microscopy for aiding our understanding of wood biodegradation. FEMS microbiology reviews. 1994; 13: 199-233.

6. Daniel G. Microview of wood under degradation by bacteria and fungi. ACS Publications; 2003.
7. Hatakka A. Biodegradation of lignin. Biopolymers Online. 2005.
8. Arantes V, Goodell B. Current understanding of brown-rot fungal biodegradation mechanisms: a review. Deterioration and protection of sustainable biomaterials: ACS Publications; 2014. p. 3-21.
9. Saba BC, Botbast RJ. Enzymology of xylan degradation. 1999.
10. Zhang J, Siika-aho M, Tenkanen M, Viikari L. The role of acetyl xylan esterase in the solubilization of xylan and enzymatic hydrolysis of wheat straw and giant reed. Biotechnology for Biofuels. 2011; 4: 1.
11. Poutanen K, Rättö M, Puls J, Viikari L. Evaluation of different microbial xylanolytic systems. Journal of Biotechnology. 1987; 6: 49-60.
12. Grohmann K, Mitchell D, Himmel M, Dale B, Schroeder H. The role of ester groups in resistance of plant cell wall polysaccharides to enzymatic hydrolysis. Applied Biochemistry and Biotechnology. 1989; 20: 45-61.
13. Sakai T, Sakamoto T, Hallaert J, Vandamme EJ. [Pectin, Pectinase, and Protopectinase: Production, Properties, and Applications. Advances in applied microbiology: Elsevier; 1993. p. 213-94.
14. Sakai T, Okushima M. Purification and crystallization of a protopectin-solubilizing enzyme from *Trichosporon penicillatum*. Agricultural and Biological Chemistry. 1982; 46: 667-76.
15. Sakamoto T, Hours RA, Sakai T. Purification, characterization, and production of two pectic transeliminases with protopectinase activity from *Bacillus subtilis*. Bioscience, biotechnology, and biochemistry. 1994; 58: 353-8.
16. Whitaker JR. Microbial pectolytic enzymes. Microbial enzymes and biotechnology: Springer; 1990. p. 133-76.
17. Luh B, Phaff H. Studies on polygalacturonase of certain yeasts. Archives of biochemistry and biophysics. 1951; 33: 212-27.
18. Sakai T, Okushima M, Yoshitake S. Purification, crystallization and some properties of endopolygalacturonase from *Kluyveromyces fragilis*. Agricultural and biological chemistry. 1984; 48: 1951-61.
19. Marcus L, Barash I, Sneh B, Koltin Y, Finkler A. Purification and characterization of pectolytic enzymes produced by virulent and hypovirulent isolates of *Rhizoctonia solani* Kuhn. Physiological and Molecular Plant Pathology. 1986; 29: 325-36.
20. De Lorenzo G, Salvi G, Degra L, D'ovidio R, Cervone F. Induction of extracellular polygalacturonase and its mRNA in the phytopathogenic fungus *Fusarium moniliforme*. Microbiology. 1987; 133: 3365-73.
21. Maria de Lourdes T, Jorge JA, Terenzi HF. Pectinase production by *Neurospora crassa*: purification and biochemical characterization of extracellular polygalacturonase activity. Microbiology. 1991; 137: 1815-23.
22. Manachini P, Fortina M, Parini C. Purification and properties of an endopolygalacturonase produced by *Rhizopus stolonifer*. Biotechnology letters. 1987; 9: 219-24.
23. Raymond P, Deléage G, Rasclé C, Fèvre M. Cloning and sequence analysis of a polygalacturonase-encoding gene from the phytopathogenic fungus *Sclerotinia sclerotiorum*. Gene. 1994; 146: 233-7.
24. Rodriguez-Palenzuela P, Burr T, Collmer A. Polygalacturonase is a virulence factor in *Agrobacterium tumefaciens* biovar 3. Journal of Bacteriology. 1991; 173: 6547-52.
25. Tierny Y, Bechet M, Joncquiert JC, Dubourguier HC, Guillaume JB. Molecular cloning and expression in *Escherichia coli* of genes encoding pectate lyase and pectin methylesterase activities from *Bacteroides thetaiotaomicron*. Journal of Applied Microbiology. 1994; 76: 592-602.
26. Kobayashi T, Higaki N, Yajima N, Suzumatsu A, Hagihara H, KAWAI S, et al. Purification and properties of a galacturonic acid-releasing exopolygalacturonase from a strain of *Bacillus*. Bioscience, biotechnology, and biochemistry. 2001; 65: 842-7.
27. Nozaki K, Miyairi K, Hozumi S, Fukui Y, Okuno T. Novel exopolygalacturonases produced by *Alternaria mali*. Bioscience, biotechnology, and biochemistry. 1997; 61: 75-80.
28. Garcíá Maceira FI, Di Pietro A, Roncero MIG. Purification and characterization of a novel exopolygalacturonase from *Fusarium oxysporum* f. sp. *lycopersici*. FEMS Microbiology Letters. 1997; 154: 37-43.

29. Whitaker JR. Pectic substances, pectic enzymes and haze formation in fruit juices. *Enzyme and Microbial Technology*. 1984; 6: 341-9.
30. Cosgrove DJ. Assembly and enlargement of the primary cell wall in plants. *Annual review of cell and developmental biology*. 1997; 13: 171-201.
31. Prade RA, Zhan D, Ayoubi P, Mort AJ. Pectins, pectinases and plant-microbe interactions. *Biotechnology and genetic engineering reviews*. 1999; 16: 361-92.
32. Micheli F. Pectin methylesterases: cell wall enzymes with important roles in plant physiology. *Trends in plant science*. 2001; 6: 414-9.
33. Förster H. Pectinesterases from *Phytophthora infestans*. *Methods in Enzymology*: Elsevier; 1988. p. 355-61.
34. Blodig W, Smith AT, Doyle WA, Piontek K. Crystal structures of pristine and oxidatively processed lignin peroxidase expressed in *Escherichia coli* and of the W171F variant that eliminates the redox active tryptophan 171. Implications for the reaction mechanism. *Journal of molecular biology*. 2001; 305: 851-61.
35. Sundaramoorthy M, Youngs HL, Gold MH, Poulos TL. High-resolution crystal structure of manganese peroxidase: substrate and inhibitor complexes. *Biochemistry*. 2005; 44: 6463-70.
36. Hallberg BM, Henriksson G, Pettersson G, Divne C. Crystal structure of the flavoprotein domain of the extracellular flavocytochrome cellobiose dehydrogenase. *Journal of molecular biology*. 2002; 315: 421-34.
37. Dwivedi UN, Singh P, Pandey VP, Kumar A. Structure–function relationship among bacterial, fungal and plant laccases. *Journal of Molecular Catalysis B: Enzymatic*. 2011; 68: 117-28.
38. Yoshida H. LXIII.-Chemistry of lacquer (Urushi). Part I. Communication from the Chemical Society of Tokio. *Journal of the Chemical Society, Transactions*. 1883; 43: 472-86.
39. Bertrand G. Sur la presence simultanee de la laccase et de la tyrosinase dans le suc de quelques champignons. *CR Hebd Seances Acad Sci*. 1896; 123: 463-5.
40. Heinzkill M, Messner K. The ligninolytic system of fungi. *Fungal biotechnology*. 1997: 213-27.
41. Gianfreda L, Xu F, Bollag J-M. Laccases: a useful group of oxidoreductive enzymes. *Bioremediation Journal*. 1999; 3: 1-26.
42. O'Malley DM, Whetten R, Bao W, Chen CL, Sederoff RR. The role of of laccase in lignification. *The Plant Journal*. 1993; 4: 751-7.
43. Hatakka A. Lignin-modifying enzymes fungi: production and role. *FEMS microbiology reviews*. 1994; 13: 125-35.
44. Youn H-D, Hah YC, Kang S-O. Role of laccase in lignin degradation by white-rot fungi. *FEMS Microbiology Letters*. 1995; 132: 183-8.
45. Thurston CF. The structure and function of fungal laccases. *Microbiology*. 1994; 140: 19-26.
46. Leatham GF, Stahmann MA. Studies on the laccase of *Lentinus edodes*: specificity, localization and association with the development of fruiting bodies. *Journal of General Microbiology*. 1981; 125: 147-57.
47. Ikegaya N, Goto M, Hayashi Y. Effect of phenolic compounds and urovides on the activities of extracellular enzyme during vegetative growth and fruit-body formation of *Lentinus edodes*. *Transactions of the Mycological Society of Japan (Japan)*. 1993.
48. Worrall J, Chet I, Hüttermann A. Association of rhizomorph formation with laccase activity in *Armillaria* spp. *Journal of General Microbiology*. 1986; 132: 2527-33.
49. Viterbo A, Staples RC, Yagen B, Mayer AM. Selective mode of action of cucurbitacin in the inhibition of laccase formation in *Botrytis cinerea*. *Phytochemistry*. 1994; 35: 1137-42.
50. Giardina P, Faraco V, Pezzella C, Piscitelli A, Vanhulle S, Sannia G. Laccases: a never-ending story. *Cellular and Molecular Life Sciences*. 2010; 67: 369-85.
51. Solomon EI, Sundaram UM, Machonkin TE. Multicopper oxidases and oxygenases. *Chemical reviews*. 1996; 96: 2563-606.
52. Yaropolov A, Skorobogat'Ko O, Vartanov S, Varfolomeyev S. Laccase. *Applied Biochemistry and Biotechnology*. 1994; 49: 257-80.
53. Sakurai T. Anaerobic reactions of *Rhus vernicifera* laccase and its type-2 copper-depleted derivatives with hexacyanoferrate (II). *Biochem J*. 1992; 284: 681-5.
54. Höfer C, Schlosser D. Novel enzymatic oxidation of Mn²⁺ to Mn³⁺ catalyzed by a fungal laccase. *FEBS letters*. 1999; 451: 186-90.

55. Schlosser D, Höfer C. Laccase-catalyzed oxidation of Mn²⁺ in the presence of natural Mn³⁺ chelators as a novel source of extracellular H₂O₂ production and its impact on manganese peroxidase. *Applied and Environmental Microbiology*. 2002; 68: 3514-21.
56. Bento I, Carrondo MA, Lindley PF. Reduction of dioxygen by enzymes containing copper. *JBIC Journal of Biological Inorganic Chemistry*. 2006; 11: 539-47.
57. Solomon EI, Baldwin MJ, Lowery MD. Electronic structures of active sites in copper proteins: contributions to reactivity. *Chemical Reviews*. 1992; 92: 521-42.
58. Palmieri G, Cennamo G, Faraco V, Amoresano A, Sannia G, Giardina P. Atypical laccase isoenzymes from copper supplemented *Pleurotus ostreatus* cultures. *Enzyme and Microbial Technology*. 2003; 33: 220-30.
59. Leontievsky AA, Vares T, Lankinen P, Shergill JK, Pozdnyakova NN, Myasoedova NM, et al. Blue and yellow laccases of ligninolytic fungi. *FEMS Microbiology Letters*. 1997; 156: 9-14.
60. Piontek K, Antorini M, Choinowski T. Crystal structure of a laccase from the fungus *Trametes versicolor* at 1.90-Å resolution containing a full complement of coppers. *Journal of Biological Chemistry*. 2002; 277: 37663-9.
61. Gutiérrez A, del Río JC, Ibarra D, Rencoret J, Romero J, Speranza M, et al. Enzymatic removal of free and conjugated sterols forming pitch deposits in environmentally sound bleaching of eucalypt paper pulp. *Environmental science & technology*. 2006; 40: 3416-22.
62. Mikolasch A, Schauer F. Fungal laccases as tools for the synthesis of new hybrid molecules and biomaterials. *Applied microbiology and biotechnology*. 2009; 82: 605-24.
63. Claus H. Laccases: structure, reactions, distribution. *Micron*. 2004; 35: 93-6.
64. Wong DW. Structure and action mechanism of ligninolytic enzymes. *Applied biochemistry and biotechnology*. 2009; 157: 174-209.
65. Kawano T. Roles of the reactive oxygen species-generating peroxidase reactions in plant defense and growth induction. *Plant cell reports*. 2003; 21: 829-37.
66. Bansal N, Kanwar SS. Peroxidase (s) in environment protection. *The Scientific World Journal*. 2013; 2013.
67. Perez-Boada M, Ruiz-Duenas FJ, Pogni R, Basosi R, Choinowski T, Martínez MJ, et al. Versatile peroxidase oxidation of high redox potential aromatic compounds: site-directed mutagenesis, spectroscopic and crystallographic investigation of three long-range electron transfer pathways. *Journal of molecular biology*. 2005; 354: 385-402.
68. Martínez AT. Molecular biology and structure-function of lignin-degrading heme peroxidases. *Enzyme and Microbial Technology*. 2002; 30: 425-44.
69. Smith AT, Veitch NC. Substrate binding and catalysis in heme peroxidases. *Current opinion in chemical biology*. 1998; 2: 269-78.
70. Banci L. Structural properties of peroxidases. *Journal of biotechnology*. 1997; 53: 253-63.
71. Gold M, Youngs H, Gelpke M. Manganese peroxidase. *Metal ions in biological systems*. 2000; 37: 559.
72. Orth A, Denny M, Tien M. Overproduction of lignin-degrading enzymes by an isolate of *Phanerochaete chrysosporium*. *Applied and environmental microbiology*. 1991; 57: 2591-6.
73. Valli K, Wariishi H, Gold MH. Oxidation of monomethoxylated aromatic compounds by lignin peroxidase: role of veratryl alcohol in lignin biodegradation. *Biochemistry*. 1990; 29: 8535-9.
74. Hammel KE, Cullen D. Role of fungal peroxidases in biological ligninolysis. *Current opinion in plant biology*. 2008; 11: 349-55.
75. Kersten PJ, Kalyanaraman B, Hammel KE, Reinhammar B, Kirk TK. Comparison of lignin peroxidase, horseradish peroxidase and laccase in the oxidation of methoxybenzenes. *Biochem J*. 1990; 268: 475-80.
76. Hammel KE, Kalyanaraman B, Kirk TK. Substrate free radicals are intermediates in ligninase catalysis. *Proceedings of the National Academy of Sciences*. 1986; 83: 3708-12.
77. Millis CD, Cai D, Stankovich MT, Tien M. Oxidation-reduction potentials and ionization states of extracellular peroxidases from the lignin-degrading fungus *Phanerochaete chrysosporium*. *Biochemistry*. 1989; 28: 8484-9.
78. Doyle WA, Blodig W, Veitch NC, Piontek K, Smith AT. Two substrate interaction sites in lignin peroxidase revealed by site-directed mutagenesis. *Biochemistry*. 1998; 37: 15097-105.

79. Mester T, Ambert-Balay K, Ciofi-Baffoni S, Banci L, Jones AD, Tien M. Oxidation of a tetrameric nonphenolic lignin model compound by lignin peroxidase. *Journal of Biological Chemistry*. 2001; 276: 22985-90.
80. Baciocchi E, Fabbri C, Lanzalunga O. Lignin peroxidase-catalyzed oxidation of nonphenolic trimeric lignin model compounds: fragmentation reactions in the intermediate radical cations. *The Journal of organic chemistry*. 2003; 68: 9061-9.
81. Bietti M, Baciocchi E, Steenken S. Lifetime, reduction potential and base-induced fragmentation of the veratryl alcohol radical cation in aqueous solution. Pulse radiolysis studies on a ligninase "mediator". *The Journal of Physical Chemistry A*. 1998; 102: 7337-42.
82. Candeias LP, Harvey PJ. Lifetime and reactivity of the veratryl alcohol radical cation. Implications for lignin peroxidase catalysis. *Journal of Biological Chemistry*. 1995; 270: 16745-8.
83. Gilardi G, Harvey PJ, Cass AE, Palmer JM. Radical intermediates in veratryl alcohol oxidation by ligninase. NMR evidence. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*. 1990; 1041: 129-32.
84. Cai D, Tien M. Kinetic studies on the formation and decomposition of compounds II and III. Reactions of lignin peroxidase with H₂O₂. *Journal of Biological Chemistry*. 1992; 267: 11149-55.
85. Koduri RS, Tien M. Kinetic analysis of lignin peroxidase: explanation for the mediation phenomenon by veratryl alcohol. *Biochemistry*. 1994; 33: 4225-30.
86. Hofrichter M. Review: lignin conversion by manganese peroxidase (MnP). *Enzyme and Microbial technology*. 2002; 30: 454-66.
87. Glenn JK, Gold MH. Purification and characterization of an extracellular Mn (II)-dependent peroxidase from the lignin-degrading basidiomycete, *Phanerochaete chrysosporium*. *Archives of Biochemistry and Biophysics*. 1985; 242: 329-41.
88. Paszczyński A, Huynh V-B, Crawford R. Enzymatic activities of an extracellular, manganese-dependent peroxidase from *Phanerochaete chrysosporium*. *FEMS Microbiology Letters*. 1985; 29: 37-41.
89. Tien M, Kirk TK. Lignin-degrading enzyme from the hymenomycete *Phanerochaete chrysosporium* Burds. *Science(Washington)*. 1983; 221: 661-2.
90. Glenn JK, Morgan MA, Mayfield MB, Kuwahara M, Gold MH. An extracellular H₂O₂-requiring enzyme preparation involved in lignin biodegradation by the white rot basidiomycete *Phanerochaete chrysosporium*. *Biochemical and biophysical research communications*. 1983; 114: 1077-83.
91. Glenn JK, Akileswaran L, Gold MH. Mn (II) oxidation is the principal function of the extracellular Mn-peroxidase from *Phanerochaete chrysosporium*. *Archives of Biochemistry and Biophysics*. 1986; 251: 688-96.
92. Paszczyński A, Huynh V-B, Crawford R. Comparison of ligninase-I and peroxidase-M2 from the white-rot fungus *Phanerochaete chrysosporium*. *Archives of Biochemistry and Biophysics*. 1986; 244: 750-65.
93. Wariishi H, Akileswaran L, Gold MH. Manganese peroxidase from the basidiomycete *Phanerochaete chrysosporium*: spectral characterization of the oxidized states and the catalytic cycle. *Biochemistry*. 1988; 27: 5365-70.
94. Wariishi H, Dunford HB, MacDonald I, Gold MH. Manganese peroxidase from the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. Transient state kinetics and reaction mechanism. *Journal of biological chemistry*. 1989; 264: 3335-40.
95. Wariishi H, Valli K, Gold MH. Manganese (II) oxidation by manganese peroxidase from the basidiomycete *Phanerochaete chrysosporium*. Kinetic mechanism and role of chelators. *Journal of Biological Chemistry*. 1992; 267: 23688-95.
96. Wariishi H, Valli K, Gold MH. Oxidative cleavage of a phenolic diarylpropane lignin model dimer by manganese peroxidase from *Phanerochaete chrysosporium*. *Biochemistry*. 1989; 28: 6017-23.
97. Tuor U, Wariishi H, Schoemaker HE, Gold MH. Oxidation of phenolic arylglycerol. beta.-aryl ether lignin model compounds by manganese peroxidase from *Phanerochaete chrysosporium*: oxidative cleavage of an. alpha.-carbonyl model compound. *Biochemistry*. 1992; 31: 4986-95.
98. Reddy GVB, Sridhar M, Gold MH. Cleavage of nonphenolic β-1 diarylpropane lignin model dimers by manganese peroxidase from *Phanerochaete chrysosporium*. *European journal of biochemistry*. 2003; 270: 284-92.

99. Wariishi H, Valli K, Renganathan V, Gold MH. Thiol-mediated oxidation of nonphenolic lignin model compounds by manganese peroxidase of *Phanerochaete chrysosporium*. *Journal of Biological Chemistry*. 1989; 264: 14185-91.
100. Mester T, Tien M. Engineering of a manganese-binding site in lignin peroxidase isozyme H8 from *Phanerochaete chrysosporium*. *Biochemical and biophysical research communications*. 2001; 284: 723-8.
101. Timofeevski SL, Nie G, Reading NS, Aust SD. Addition of veratryl alcohol oxidase activity to manganese peroxidase by site-directed mutagenesis. *Biochemical and biophysical research communications*. 1999; 256: 500-4.
102. Ruiz-Dueñas FJ, Martínez MJ, Martínez AT. Molecular characterization of a novel peroxidase isolated from the ligninolytic fungus *Pleurotus eryngii*. *Molecular microbiology*. 1999; 31: 223-35.
103. Camarero S, Ruiz-Dueñas FJ, Sarkar S, Martínez MJ, Martínez AT. The cloning of a new peroxidase found in lignocellulose cultures of *Pleurotus eryngii* and sequence comparison with other fungal peroxidases. *FEMS microbiology letters*. 2000; 191: 37-43.
104. Ruiz-Dueñas FJ, Morales M, García E, Miki Y, Martínez MJ, Martínez AT. Substrate oxidation sites in versatile peroxidase and other basidiomycete peroxidases. *Journal of Experimental Botany*. 2009; 60: 441-52.
105. Ruiz-Dueñas FJ, Pogni R, Morales M, Giansanti S, Mate MJ, Romero A, et al. Protein radicals in fungal versatile peroxidase catalytic tryptophan radical in both compound I and compound II and studies on W164Y, W164H and W164S variants. *Journal of Biological Chemistry*. 2009; 284: 7986-94.
106. Camarero S, Sarkar S, Ruiz-Dueñas FJ, Martínez MaJ, Martínez ÁT. Description of a versatile peroxidase involved in the natural degradation of lignin that has both manganese peroxidase and lignin peroxidase substrate interaction sites. *Journal of Biological Chemistry*. 1999; 274: 10324-30.
107. Henriksson G, Ander P, Pettersson B, Pettersson G. Cellobiose dehydrogenase (cellobiose oxidase) from *Phanerochaete chrysosporium* as a wood-degrading enzyme. *Studies on cellulose, xylan and synthetic lignin. Applied microbiology and biotechnology*. 1995; 42: 790-6.
108. Kersten P, Cullen D. Extracellular oxidative systems of the lignin-degrading Basidiomycete *Phanerochaete chrysosporium*. *Fungal Genetics and Biology*. 2007; 44: 77-87.
109. Cameron MD, Aust SD. Cellobiose dehydrogenase—an extracellular fungal flavocytochrome. *Enzyme and microbial technology*. 2001; 28: 129-38.
110. Henriksson G, Zhang L, Li J, Ljungquist P, Reitberger T, Pettersson G, et al. Is cellobiose dehydrogenase from *Phanerochaete chrysosporium* a lignin degrading enzyme? *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*. 2000; 1480: 83-91.
111. Henriksson G, Johansson G, Pettersson G. Is cellobiose oxidase from *Phanerochaete chrysosporium* a one-electron reductase? *Biochimica et Biophysica Acta (BBA)-Bioenergetics*. 1993; 1144: 184-90.
112. Morpeth FF. Some properties of cellobiose oxidase from the white-rot fungus *Sporotrichum pulverulentum*. *Biochem J*. 1985; 228: 557-64.
113. Henriksson G, Sild V, Szabó IJ, Pettersson G, Johansson G. Substrate specificity of cellobiose dehydrogenase from *Phanerochaete chrysosporium*. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*. 1998; 1383: 48-54.
114. Henriksson G, Pettersson G, Johansson G, Ruiz A, Uzategui E. Cellobiose oxidase from *Phanerochaete chrysosporium* can be cleaved by papain into two domains. *European journal of biochemistry*. 1991; 196: 101-6.
115. Ander P. The cellobiose-oxidizing enzymes CBQ and CbO as related to lignin and cellulose degradation-- a review. *FEMS microbiology reviews*. 1994; 13: 297-312.
116. Archibald F, Bourbonnais R, Jurasek L, Paice M, Reid I. Kraft pulp bleaching and delignification by *Trametes versicolor*. *Journal of Biotechnology*. 1997; 53: 215-36.
117. Cameron MD, Aust SD. Degradation of chemicals by reactive radicals produced by cellobiose dehydrogenase from *Phanerochaete chrysosporium*. *Archives of biochemistry and biophysics*. 1999; 367: 115-21.
118. Henriksson G, Johansson G, Pettersson G. A critical review of cellobiose dehydrogenases. *Journal of Biotechnology*. 2000; 78: 93-113.

119. Baldrian P, Valášková V. Degradation of cellulose by basidiomycetous fungi. *FEMS microbiology reviews*. 2008; 32: 501-21.
120. Ducros V, Brzozowski AM, Wilson KS, Brown SH, Østergaard P, Schneider P, et al. Crystal structure of the type-2 Cu depleted laccase from *Coprinus cinereus* at 2.2 Å resolution. *Nature Structural & Molecular Biology*. 1998; 5: 310-6.
121. Hakulinen N, Kiiskinen L-L, Kruus K, Saloheimo M, Paananen A, Koivula A, et al. Crystal structure of a laccase from *Melanocarpus albomyces* with an intact trinuclear copper site. *Nature Structural & Molecular Biology*. 2002; 9: 601-5.
122. Lyashenko AV, Zhukhlistova NE, Gabdoulkhakov AG, Zhukova YN, Voelter W, Zaitsev VN, et al. Purification, crystallization and preliminary X-ray study of the fungal laccase from *Cerrena maxima*. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*. 2006; 62: 954-7.
123. Kallio JP, Gasparetti C, Andberg M, Boer H, Koivula A, Kruus K, et al. Crystal structure of an ascomycete fungal laccase from *Thielavia arenaria*—common structural features of asco-laccases. *FEBS Journal*. 2011; 278: 2283-95.
124. Ferraroni M, Myasoedova N, Schmatchenko V, Leontievsky A, Golovleva L, Scozzafava A, et al. Crystal structure of a blue laccase from *Lentinus tigrinus*: evidences for intermediates in the molecular oxygen reductive splitting by multicopper oxidases. *BMC Structural Biology*. 2007; 7: 60.
125. Miki Y, Calviño FR, Pogni R, Giansanti S, Ruiz-Dueñas FJ, Martínez MJ, et al. Crystallographic, kinetic, and spectroscopic study of the first ligninolytic peroxidase presenting a catalytic tyrosine. *Journal of Biological Chemistry*. 2011; 286: 15525-34.
126. Piontek K, Smith A, Blodig W. Lignin peroxidase structure and function. *Biochemical Society Transactions*. 2001; 29: 111-6.
127. Edwards SL, Raag R, Wariishi H, Gold MH, Poulos TL. Crystal structure of lignin peroxidase. *Proceedings of the National Academy of Sciences*. 1993; 90: 750-4.
128. Johjima T, Itoh N, Kabuto M, Tokimura F, Nakagawa T, Wariishi H, et al. Direct interaction of lignin and lignin peroxidase from *Phanerochaete chrysosporium*. *Proceedings of the National Academy of Sciences*. 1999; 96: 1989-94.
129. Blodig W, Smith AT, Winterhalter K, Piontek K. Evidence from spin-trapping for a transient radical on tryptophan residue 171 of lignin peroxidase. *Archives of biochemistry and biophysics*. 1999; 370: 86-92.
130. Choinowski T, Blodig W, Winterhalter KH, Piontek K. The crystal structure of lignin peroxidase at 1.70 Å resolution reveals a hydroxy group on the C β of tryptophan 171: a novel radical site formed during the redox cycle. *Journal of molecular biology*. 1999; 286: 809-27.
131. Poulos T, Edwards S, Wariishi H, Gold M. Crystallographic refinement of lignin peroxidase at 2 Å. *Journal of Biological Chemistry*. 1993; 268: 4429-40.
132. Piontek K, Glumoff T, Winterhalter K. Low pH crystal structure of glycosylated lignin peroxidase from *Phanerochaete chrysosporium* at 2.5 Å resolution. *FEBS letters*. 1993; 315: 119-24.
133. Sundaramoorthy M, Kishi K, Gold MH, Poulos TL. The crystal structure of manganese peroxidase from *Phanerochaete chrysosporium* at 2.06-Å resolution. *Journal of Biological Chemistry*. 1994; 269: 32759-67.
134. Sundaramoorthy M, Gold MH, Poulos TL. Ultrahigh (0.93 Å) resolution structure of manganese peroxidase from *Phanerochaete chrysosporium*: Implications for the catalytic mechanism. *Journal of inorganic biochemistry*. 2010; 104: 683-90.
135. Pfister TD, Mirarefi AY, Gengenbach AJ, Zhao X, Danstrom C, Conatser N, et al. Kinetic and crystallographic studies of a redesigned manganese-binding site in cytochrome c peroxidase. *JBIC Journal of Biological Inorganic Chemistry*. 2007; 12: 126-37.
136. Sundaramoorthy M, Kishi K, Gold MH, Poulos TL. Crystal structures of substrate binding site mutants of manganese peroxidase. *Journal of Biological Chemistry*. 1997; 272: 17574-80.
137. Pogni R, Baratto MC, Teutloff C, Giansanti S, Ruiz-Dueñas FJ, Choinowski T, et al. A Tryptophan Neutral Radical in the Oxidized State of Versatile Peroxidase from *Pleurotus eryngii* a combined multifrequency EPR and density functional theory study. *Journal of Biological Chemistry*. 2006; 281: 9517-26.

138. Moreira PR, Duez C, Dehareng D, Antunes A, Almeida-Vara E, Frère J-M, et al. Molecular characterisation of a versatile peroxidase from a Bjerkandera strain. *Journal of biotechnology*. 2005; 118: 339-52.
139. Ruiz-Duenas FJ, Morales M, Mate MJ, Romero A, Martínez MJ, Smith AT, et al. Site-directed mutagenesis of the catalytic tryptophan environment in *Pleurotus eryngii* versatile peroxidase. *Biochemistry*. 2008; 47: 1685-95.
140. Ferri S, Sode K. Amino acid substitution at the substrate-binding subsite alters the specificity of the *Phanerochaete chrysosporium* cellobiose dehydrogenase. *Biochemical and biophysical research communications*. 2010; 391: 1246-50.
141. Hallberg BM, Henriksson G, Pettersson G, Vasella A, Divne C. Mechanism of the reductive half-reaction in cellobiose dehydrogenase. *Journal of Biological Chemistry*. 2003; 278: 7160-6.
142. Hallberg BM, Bergfors T, Bäckbro K, Pettersson G, Henriksson G, Divne C. A new scaffold for binding haem in the cytochrome domain of the extracellular flavocytochrome cellobiose dehydrogenase. *Structure*. 2000; 8: 79-88.
143. Levasseur A, Piumi F, Coutinho PM, Rancurel C, Asther M, Delattre M, et al. FOLy: an integrated database for the classification and functional annotation of fungal oxidoreductases potentially involved in the degradation of lignin and related aromatic compounds. *Fungal genetics and biology*. 2008; 45: 638-45.
144. Shah V, Nerud F. Lignin degrading system of white-rot fungi and its exploitation for dye decolorization. *Canadian journal of microbiology*. 2002; 48: 857-70.
145. Fernández IS, Ruiz-Duenas FJ, Santillana E, Ferreira P, Martínez MJ, Martínez ÁT, et al. Novel structural features in the GMC family of oxidoreductases revealed by the crystal structure of fungal aryl-alcohol oxidase. *Acta Crystallographica Section D: Biological Crystallography*. 2009; 65: 1196-205.
146. van den Heuvel RH, van den Berg WA, Rovida S, van Berkel WJ. Laboratory-evolved vanillyl-alcohol oxidase produces natural vanillin. *Journal of Biological Chemistry*. 2004; 279: 33492-500.
147. Wohlfahrt G, Witt S, Hendle J, Schomburg D, Kalisz HM, Hecht H-J. 1.8 and 1.9 Å resolution structures of the *Penicillium amagasakiense* and *Aspergillus niger* glucose oxidases as a basis for modelling substrate complexes. *Acta Crystallographica Section D: Biological Crystallography*. 1999; 55: 969-77.
148. Rannes JB, Ioannou A, Willies SC, Grogan G, Behrens C, Flitsch SL, et al. Glycoprotein labeling using engineered variants of galactose oxidase obtained by directed evolution. *Journal of the American Chemical Society*. 2011; 133: 8436-9.
149. Hassan N, Tan T-C, Spadiut O, Pisanelli I, Fusco L, Haltrich D, et al. Crystal structures of *Phanerochaete chrysosporium* pyranose 2-oxidase suggest that the N-terminus acts as a propeptide that assists in homotetramer assembly. *FEBS open bio*. 2013; 3: 496-504.
150. Ferreira P, Medina M, Guillén F, Martinez M, Van Berkel W, Martinez A. Spectral and catalytic properties of aryl-alcohol oxidase, a fungal flavoenzyme acting on polyunsaturated alcohols. *Biochem J*. 2005; 389: 731-8.
151. Farmer V, Henderson ME, Russell J. Aromatic-alcohol-oxidase activity in the growth medium of *Polystictus versicolor*. *Biochemical Journal*. 1960; 74: 257.
152. Guillén F, Martinez AT, Martínez MJ. Production of hydrogen peroxide by aryl-alcohol oxidase from the ligninolytic fungus *Pleurotus eryngii*. *Applied microbiology and biotechnology*. 1990; 32: 465-9.
153. Muheim A, Waldner R, Leisola MS, Fiechter A. An extracellular aryl-alcohol oxidase from the white-rot fungus *Bjerkandera adusta*. *Enzyme and microbial technology*. 1990; 12: 204-9.
154. Kim SJ, Suzuki N, Uematsu Y, Shoda M. Characterization of Aryl Alcohol Oxidase Produced by Dye-Decolorizing Fungus, *Geotrichum candidum* Decl. *Journal of bioscience and bioengineering*. 2001; 91: 166-72.
155. Guillen F, Martinez AT, Martinez MJ. Substrate specificity and properties of the aryl-alcohol oxidase from the ligninolytic fungus *Pleurotus eryngii*. *European Journal of Biochemistry*. 1992; 209: 603-11.
156. Varela E, Martinez A, Martinez M. Molecular cloning of aryl-alcohol oxidase from the fungus *Pleurotus eryngii*, an enzyme involved in lignin degradation. *Biochem J*. 1999; 341: 113-7.

157. Hernández-Ortega A, Ferreira P, Martínez AT. Fungal aryl-alcohol oxidase: a peroxide-producing flavoenzyme involved in lignin degradation. *Applied microbiology and biotechnology*. 2012; 93: 1395-410.
158. Gutierrez A, Caramelo L, Prieto A, Martínez MJ, Martínez AT. Anisaldehyde production and aryl-alcohol oxidase and dehydrogenase activities in ligninolytic fungi of the genus *Pleurotus*. *Applied and Environmental Microbiology*. 1994; 60: 1783-8.
159. Romero E, Ferreira P, Martínez ÁT, Martínez MJ. New oxidase from *Bjerkandera arthroconidial* anamorph that oxidizes both phenolic and nonphenolic benzyl alcohols. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*. 2009; 1794: 689-97.
160. Hernández-Ortega A, Ferreira P, Merino P, Medina M, Guallar V, Martínez AT. Stereoselective Hydride Transfer by Aryl-Alcohol Oxidase, a Member of the GMC Superfamily. *ChemBioChem*. 2012; 13: 427-35.
161. Ferreira P, Hernández-Ortega A, Lucas F, Carro J, Herguedas B, Borrelli KW, et al. Aromatic stacking interactions govern catalysis in aryl-alcohol oxidase. *FEBS Journal*. 2015.
162. Varela E, Martínez MaJ, Martínez AT. Aryl-alcohol oxidase protein sequence: a comparison with glucose oxidase and other FAD oxidoreductases. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*. 2000; 1481: 202-8.
163. van den Heuvel RH, Fraaije MW, Mattevi A, van Berkel WJ. Asp-170 is crucial for the redox properties of vanillyl-alcohol oxidase. *Journal of Biological Chemistry*. 2000; 275: 14799-808.
164. van den Heuvel RH, Fraaije MW, Mattevi A, Laane C, van Berkel WJ. Vanillyl-alcohol oxidase, a tasteful biocatalyst. *Journal of Molecular Catalysis B: Enzymatic*. 2001; 11: 185-8.
165. Fraaije MW, van Berkel WJ. Catalytic Mechanism of the Oxidative Demethylation of 4-(Methoxymethyl) phenol by Vanillyl-Alcohol Oxidase evidence for formation of a p-quinone methide intermediate. *Journal of Biological Chemistry*. 1997; 272: 18111-6.
166. Mattevi A, Fraaije MW, Mozzarelli A, Olivi L, Coda A, van Berkel WJ. Crystal structures and inhibitor binding in the octameric flavoenzyme vanillyl-alcohol oxidase: the shape of the active-site cavity controls substrate specificity. *Structure*. 1997; 5: 907-20.
167. van den Heuvel RH, Fraaije MW, Mattevi A, van Berkel WJ. Structure, function and redesign of vanillyl-alcohol oxidase. *International Congress Series: Elsevier*; 2002. p. 13-24.
168. Whittaker MM, Kersten PJ, Nakamura N, Sanders-Loehr J, Schweizer ES, Whittaker JW. Glyoxal oxidase from *Phanerochaete chrysosporium* is a new radical-copper oxidase. *Journal of Biological Chemistry*. 1996; 271: 681-7.
169. Whittaker MM, Kersten PJ, Cullen D, Whittaker JW. Identification of catalytic residues in glyoxal oxidase by targeted mutagenesis. *Journal of Biological Chemistry*. 1999; 274: 36226-32.
170. Kersten PJ, Kirk TK. Involvement of a new enzyme, glyoxal oxidase, in extracellular H₂O₂ production by *Phanerochaete chrysosporium*. *Journal of Bacteriology*. 1987; 169: 2195-201.
171. Kersten PJ, Witek C, Vanden Wymelenberg A, Cullen D. *Phanerochaete chrysosporium* glyoxal oxidase is encoded by two allelic variants: structure, genomic organization, and heterologous expression of glx1 and glx2. *Journal of bacteriology*. 1995; 177: 6106-10.
172. Danneel HJ, Rossner E, Zeeck A, Giffhorn F. Purification and characterization of a pyranose oxidase from the basidiomycete *Peniophora gigantea* and chemical analyses of its reaction products. *European journal of biochemistry*. 1993; 214: 795-802.
173. Izumi Y, Furuya Y, Yamada H. Purification and properties of pyranose oxidase from basidiomycetous fungus no. 52. *Agricultural and biological chemistry*. 1990; 54: 1393-9.
174. Machida Y, Nakanishi T. Purification and properties of pyranose oxidase from *Coriolus versicolor*. *Agricultural and biological chemistry*. 1984; 48: 2463-70.
175. Daniel G, Volc J, Kubatova E. Pyranose oxidase, a major source of H₂O₂ during wood degradation by *Phanerochaete chrysosporium*, *Trametes versicolor*, and *Oudemansiella mucida*. *Applied and environmental microbiology*. 1994; 60: 2524-32.
176. Cavener DR. GMC oxidoreductases: a newly defined family of homologous proteins with diverse catalytic activities. *Journal of molecular biology*. 1992; 223: 811-4.
177. Giffhorn F. Fungal pyranose oxidases: occurrence, properties and biotechnical applications in carbohydrate chemistry. *Applied microbiology and biotechnology*. 2000; 54: 727-40.

178. Volc J, Sedmera P, Havlíček V, Přikrylová V, Daniel G. Conversion of D-glucose to D-erythrohexos-2, 3-diulose (2, 3-diketo-D-glucose) by enzyme preparations from the basidiomycete *Oudemansiella mucida*. *Carbohydrate research*. 1995; 278: 59-70.
179. Freimund S, Huwig A, Giffhorn F, Köpper S. Rare keto-aldoses from enzymatic oxidation: substrates and oxidation products of pyranose 2-oxidase. *Chemistry- A European Journal*. 1998; 4: 2442-55.
180. Giffhorn F, Köpper S, Huwig A, Freimund S. Rare sugars and sugar-based synthons by chemo-enzymatic synthesis. *Enzyme and microbial technology*. 2000; 27: 734-42.
181. Baron AJ, Stevens C, Wilmot C, Seneviratne KD, Blakeley V, Dooley DM, et al. Structure and mechanism of galactose oxidase. The free radical site. *Journal of Biological Chemistry*. 1994; 269: 25095-105.
182. Firbank S, Rogers M, Wilmot C, Dooley D, Halcrow M, Knowles P, et al. Crystal structure of the precursor of galactose oxidase: An unusual self-processing enzyme. *Proceedings of the National Academy of Sciences*. 2001; 98: 12932-7.
183. Yin D, Urresti S, Lafond M, Johnston EM, Derikvand F, Ciano L, et al. Structure-function characterization reveals new catalytic diversity in the galactose oxidase and glyoxal oxidase family. *Nat Commun*. 2015; 6.
184. Ito N, Phillips SE, Stevens C, Ogel ZB, McPherson MJ, Keen JN, et al. Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. 1991.
185. Bankar SB, Bule MV, Singhal RS, Ananthanarayan L. Glucose oxidase—an overview. *Biotechnology advances*. 2009; 27: 489-501.
186. Hatzinikolaou D, Macris B. Factors regulating production of glucose oxidase by *Aspergillus niger*. *Enzyme and Microbial Technology*. 1995; 17: 530-4.
187. Kapat A, Jung J, Park Y. Enhancement of glucose oxidase production in batch cultivation of recombinant *Saccharomyces cerevisiae*: optimization of oxygen transfer condition. *Journal of applied microbiology*. 2001; 90: 216-22.
188. Witt S, Wohlfahrt G, Schomburg D, Hecht H, Kalisz H. Conserved arginine-516 of *Penicillium amagasakiense* glucose oxidase is essential for the efficient binding of β -D-glucose. *Biochem J*. 2000; 347: 553-9.
189. Witteveen CF, Veenhuis M, Visser J. Localization of glucose oxidase and catalase activities in *Aspergillus niger*. *Applied and environmental microbiology*. 1992; 58: 1190-4.
190. Raba J, Mottola HA. Glucose oxidase as an analytical reagent. *Critical reviews in Analytical chemistry*. 1995; 25: 1-42.
191. Brock BJ, Rieble S, Gold MH. Purification and Characterization of a 1, 4-Benzoquinone Reductase from the Basidiomycete *Phanerochaete chrysosporium*. *Applied and environmental microbiology*. 1995; 61: 3076-81.
192. Akileswaran L, Brock BJ, Cereghino JL, Gold MH. 1, 4-Benzoquinone reductase from *Phanerochaete chrysosporium*: cDNA cloning and regulation of expression. *Applied and environmental microbiology*. 1999; 65: 415-21.
193. Brock BJ, Gold MH. 1, 4-Benzoquinone reductase from the basidiomycete *Phanerochaete chrysosporium*: spectral and kinetic analysis. *Archives of biochemistry and biophysics*. 1996; 331: 31-40.
194. Aitor H-O, Kenneth B, Patricia F, Milagros M, Angel TM, Victor G. Substrate diffusion and oxidation in GMC oxidoreductases: an experimental and computational study on fungal aryl-alcohol oxidase. *Biochemical Journal*. 2011; 436: 341-50.
195. Hernández-Ortega A, Lucas F, Ferreira P, Medina M, Guallar V, Martínez AT. Modulating O₂ reactivity in a fungal flavoenzyme involvement of aryl-alcohol oxidase PHE-501 contiguous to catalytic histidine. *Journal of Biological Chemistry*. 2011; 286: 41105-14.
196. Hernández-Ortega A, Lucas Ft, Ferreira P, Medina M, Guallar V, Martínez AT. Role of active site histidines in the two half-reactions of the aryl-alcohol oxidase catalytic cycle. *Biochemistry*. 2012; 51: 6595-608.
197. Fraaije MW, van den Heuvel RH, van Berkel WJ, Mattevi A. Covalent flavinylation is essential for efficient redox catalysis in vanillyl-alcohol oxidase. *Journal of Biological Chemistry*. 1999; 274: 35514-20.

198. van den Heuvel RH, Fraaije MW, Ferrer M, Mattevi A, van Berkel WJ. Inversion of stereospecificity of vanillyl-alcohol oxidase. *Proceedings of the National Academy of Sciences*. 2000; 97: 9455-60.
199. Fraaije MW, van den Heuvel RH, van Berkel WJ, Mattevi A. Structural analysis of flavinylation in vanillyl-alcohol oxidase. *Journal of Biological Chemistry*. 2000; 275: 38654-8.
200. Vaidyanathan M, Palaniandavar M, Gopalan RS. Copper (II) complexes of sterically hindered phenolate ligands as structural models for the active site in galactose oxidase and glyoxal oxidase: X-ray crystal structure and spectral and redox properties. *Inorganica Chimica Acta*. 2001; 324: 241-51.
201. Halfen JA, Jazdzewski BA, Mahapatra S, Berreau LM, Wilkinson EC, Que L, et al. Synthetic models of the inactive copper (II)-tyrosinate and active copper (II)-tyrosyl radical forms of galactose and glyoxal oxidases. *Journal of the American Chemical Society*. 1997; 119: 8217-27.
202. Hallberg BM, Leitner C, Haltrich D, Divne C. Crystal structure of the 270 kDa homotetrameric lignin-degrading enzyme pyranose 2-oxidase. *Journal of molecular biology*. 2004; 341: 781-96.
203. Kujawa M, Ebner H, Leitner C, Hallberg BM, Prongjit M, Sucharitakul J, et al. Structural basis for substrate binding and regioselective oxidation of monosaccharides at C3 by pyranose 2-oxidase. *Journal of Biological Chemistry*. 2006; 281: 35104-15.
204. Pitsawong W, Sucharitakul J, Prongjit M, Tan T-C, Spadiut O, Haltrich D, et al. A conserved active-site threonine is important for both sugar and flavin oxidations of pyranose 2-oxidase. *Journal of Biological Chemistry*. 2010; 285: 9697-705.
205. Spadiut O, Tan TC, Pisanelli I, Haltrich D, Divne C. Importance of the gating segment in the substrate-recognition loop of pyranose 2-oxidase. *FEBS journal*. 2010; 277: 2892-909.
206. Heckmann-Pohl DM, Bastian S, Altmeyer S, Antes I. Improvement of the fungal enzyme pyranose 2-oxidase using protein engineering. *Journal of biotechnology*. 2006; 124: 26-40.
207. Bannwarth M, Heckmann-Pohl D, Bastian S, Giffhorn F, Schulz GE. Reaction geometry and thermostable variant of pyranose 2-oxidase from the white-rot fungus *Peniophora* sp. *Biochemistry*. 2006; 45: 6587-95.
208. Rogers MS, Tyler EM, Akyumani N, Kurtis CR, Spooner RK, Deacon SE, et al. The stacking tryptophan of galactose oxidase: a second-coordination sphere residue that has profound effects on tyrosyl radical behavior and enzyme catalysis. *Biochemistry*. 2007; 46: 4606-18.
209. Ito N, Phillips SE, Yadav KD, Knowles PF. Crystal structure of a free radical enzyme, galactose oxidase. *Journal of molecular biology*. 1994; 238: 704-814.
210. Hecht H, Kalisz H, Hendle J, Schmid R, Schomburg D. Crystal structure of glucose oxidase from *Aspergillus niger* refined at 2.3 Å resolution. *Journal of molecular biology*. 1993; 229: 153-72.
211. Hecht H, Schomburg D, Kalisz H, Schmid R. The 3D structure of glucose oxidase from *Aspergillus niger*. Implications for the use of GOD as a biosensor enzyme. *Biosensors and Bioelectronics*. 1993; 8: 197-203.
212. Biely P. Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology advances*. 2012; 30: 1575-88.
213. Biely P, Puls J, Schneider H. Acetyl xylan esterases in fungal cellulolytic systems. *Febs Letters*. 1985; 186: 80-4.
214. Biely P, MacKenzie C, Puls J, Schneider H. Cooperativity of esterases and xylanases in the enzymatic degradation of acetyl xylan. *Nature Biotechnology*. 1986; 4: 731-3.
215. Selig MJ, Knoshaug EP, Adney WS, Himmel ME, Decker SR. Synergistic enhancement of cellobiohydrolase performance on pretreated corn stover by addition of xylanase and esterase activities. *Bioresource Technology*. 2008; 99: 4997-5005.
216. Selig MJ, Adney WS, Himmel ME, Decker SR. The impact of cell wall acetylation on corn stover hydrolysis by cellulolytic and xylanolytic enzymes. *Cellulose*. 2009; 16: 711-22.
217. Gonzalez SL, Rosso ND. Determination of pectin methylesterase activity in commercial pectinases and study of the inactivation kinetics through two potentiometric procedures. *Food Science and Technology (Campinas)*. 2011; 31: 412-7.
218. Jayani RS, Saxena S, Gupta R. Microbial pectinolytic enzymes: a review. *Process Biochemistry*. 2005; 40: 2931-44.
219. Alkorta I, Garbisu C, Llama MJ, Serra JL. Industrial applications of pectic enzymes: a review. *Process Biochemistry*. 1998; 33: 21-8.

220. Fu J-T, Rao M. The influence of sucrose and sorbitol on gel–sol transition of low-methoxyl pectin+ Ca²⁺ gels. *Food Hydrocolloids*. 1999; 13: 371-80.
221. Fu J, Rao M. Rheology and structure development during gelation of low-methoxyl pectin gels: the effect of sucrose. *Food Hydrocolloids*. 2001; 15: 93-100.
222. Kohli P, Kalia M, Gupta R. Pectin methylesterases: A review. *Journal of Bioprocessing & Biotechniques*. 2015; 5: 1.
223. Kashyap D, Vohra P, Chopra S, Tewari R. Applications of pectinases in the commercial sector: a review. *Bioresource technology*. 2001; 77: 215-27.
224. Vierhuis E, Korver M, Schols HA, Voragen AG. Structural characteristics of pectic polysaccharides from olive fruit (*Olea europaea* cv moraiolo) in relation to processing for oil extraction. *Carbohydrate Polymers*. 2003; 51: 135-48.
225. Ishii T. O-acetylated oligosaccharides from pectins of potato tuber cell walls. *Plant physiology*. 1997; 113: 1265-72.
226. Navarro-Fernández J, Martínez-Martínez I, Montoro-García S, García-Carmona F, Takami H, Sánchez-Ferrer Á. Characterization of a new rhamnogalacturonan acetyl esterase from *Bacillus halodurans* C-125 with a new putative carbohydrate binding domain. *Journal of bacteriology*. 2008; 190: 1375-82.
227. Guglielmetti S, De Noni I, Caracciolo F, Molinari F, Parini C, Mora D. Bacterial cinnamoyl esterase activity screening for the production of a novel functional food product. *Applied and environmental microbiology*. 2008; 74: 1284-8.
228. Topakas E, Vafiadi C, Christakopoulos P. Microbial production, characterization and applications of feruloyl esterases. *Process Biochemistry*. 2007; 42: 497-509.
229. Faulds CB. What can feruloyl esterases do for us? *Phytochemistry Reviews*. 2010; 9: 121-32.
230. Faulds CB, Zanichelli D, Crepin VF, Connerton IF, Juge N, Bhat MK, et al. Specificity of feruloyl esterases for water-extractable and water-unextractable feruloylated polysaccharides: influence of xylanase. *Journal of cereal science*. 2003; 38: 281-8.
231. Faulds C, Mandalari G, Curto RL, Bisignano G, Christakopoulos P, Waldron K. Synergy between xylanases from glycoside hydrolase family 10 and family 11 and a feruloyl esterase in the release of phenolic acids from cereal arabinoxylan. *Applied microbiology and biotechnology*. 2006; 71: 622-9.
232. Špáníková S, Biely P. Glucuronoyl esterase–novel carbohydrate esterase produced by *Schizophyllum commune*. *FEBS letters*. 2006; 580: 4597-601.
233. ĎURANOVÁ M, Hirsch J, Kolenova K, Biely P. Fungal glucuronoyl esterases and substrate uronic acid recognition. *Bioscience, biotechnology, and biochemistry*. 2009; 73: 2483-7.
234. Topakas E, Moukouli M, Dimarogona M, Vafiadi C, Christakopoulos P. Functional expression of a thermophilic glucuronoyl esterase from *Sporotrichum thermophile*: identification of the nucleophilic serine. *Applied microbiology and biotechnology*. 2010; 87: 1765-72.
235. Benoit I, Navarro D, Marnet N, Rakotomanomana N, Lesage-Meessen L, Sigoillot J-C, et al. Feruloyl esterases as a tool for the release of phenolic compounds from agro-industrial by-products. *Carbohydrate Research*. 2006; 341: 1820-7.
236. Tabka M, Herpoël-Gimbert I, Monod F, Asther M, Sigoillot J. Enzymatic saccharification of wheat straw for bioethanol production by a combined cellulase xylanase and feruloyl esterase treatment. *Enzyme and Microbial Technology*. 2006; 39: 897-902.
237. Tapin S, Sigoillot J-C, Asther M, Petit-Conil M. Feruloyl esterase utilization for simultaneous processing of nonwood plants into phenolic compounds and pulp fibers. *Journal of agricultural and food chemistry*. 2006; 54: 3697-703.
238. Record E, Asther M, Sigoillot C, Pages S, Punt P, Delattre M, et al. Overproduction of the *Aspergillus niger* feruloyl esterase for pulp bleaching application. *Applied microbiology and biotechnology*. 2003; 62: 349-55.
239. Laszlo JA, Compton DL, Li X-L. Feruloyl esterase hydrolysis and recovery of ferulic acid from jojoba meal. *Industrial crops and products*. 2006; 23: 46-53.

The literature review (chapter-1) has been published, please find the article information below:

1. **Ayyappa Kumar Sista Kameshwar**, and Wensheng Qin. "Lignin Degrading Fungal Enzymes." In **Production of Biofuels and Chemicals from Lignin**, pp. 81-130. Springer Singapore, 2016. (2016)
2. **Ayyappa Kumar Sista Kameshwar & Wensheng Qin** (2018) Understanding the structural and functional properties of carbohydrate esterases with a special focus on hemicellulose deacetylating acetyl xylan esterases, **Mycology**, DOI: 10.1080/21501203.2018.1492979
3. **Ayyappa Kumar Sista Kameshwar**, and Wensheng Qin. "Structural and functional properties of pectin and lignin–carbohydrate complexes de-esterases: a review." **Bioresources and Bioprocessing** 5, no. 1 (2018): 43.

Chapter-2

Metadata Analysis of *Phanerochaete chrysosporium* Gene Expression Data Identified Common CAZymes Encoding Gene Expression Profiles Involved in Cellulose and Hemicellulose Degradation

[This work has been published in "International journal of biological sciences"13, no. 1 (2017): 85-99.]

Ayyappa Kumar Sista Kameshwar and Wensheng Qin*

2.1. Abstract

In literature, extensive studies have been conducted on popular wood degrading white rot fungus, *Phanerochaete chrysosporium* about its lignin degrading mechanisms compared to the cellulose and hemicellulose degrading abilities. This study delineates cellulose and hemicellulose degrading mechanisms through large scale metadata analysis of *P. chrysosporium* gene expression data (retrieved from NCBI GEO) to understand the common expression patterns of differentially expressed genes when cultured on different growth substrates. Genes encoding glycoside hydrolase classes commonly expressed during breakdown of cellulose such as GH-5,6,7,9,44,45,48 and hemicellulose are GH-2,8,10,11,26,30,43,47 were found to be highly expressed among varied growth conditions including simple customized and complex natural plant biomass growth mediums. Genes encoding carbohydrate esterase class enzymes CE (1,4,8,9,15,16) polysaccharide lyase class enzymes PL-8 and PL-14, and glycosyl transferases classes GT (1,2,4,8,15,20,35,39,48) were differentially expressed in natural plant biomass growth mediums. Based on these results, *P. chrysosporium*, on natural plant biomass substrates was found to express lignin and hemicellulose degrading enzymes more than cellulolytic enzymes except GH-61 (LPMO) class enzymes, in early stages. It was observed that the fate of *P. chrysosporium* transcriptome is significantly affected by the wood substrate provided. We believe, the gene expression findings in this study plays crucial role in developing genetically efficient microbe with effective cellulose and hemicellulose degradation abilities.

Keywords: *Phanerochaete chrysosporium*, Transcriptome, Lignocellulose, Gene Expression Omnibus (GEO), GEO2R, Bioconductor, Carbohydrate Active Enzyme database (CAZy)

2.2. Introduction

Recently, production of second generation fuel substrate (lignocellulose) gained much attention as the first generation fuel substrate (corn starch) faces the food vs fuel challenge [1]. Abundant availability of lignocellulosic biomass and its potential for fuel production, encourage the research interest in biorefining scientists to substitute renewable substrates for fossil based products [1]. However, major bottlenecks include the highly recalcitrant nature of lignocellulosics due to which various steps such as thermochemical pre-treatment followed by enzymatic hydrolysis and fermentation in upstream processing are required [2-4]. These tedious process increases the costs involved in the production of fuels [2-4]. Though chemical pre-treatment methods of polysaccharide degradation do exist, but world-wide biorefining strategies are enzyme based [5]. Usually the pretreated biomass is depolymerized by supplementing the enzyme cocktail which breakdowns the polysaccharides to pentoses and hexoses. Enzyme cocktails produced from the *Trichoderma reesei* are highly used for the degradation of polysaccharides [6].

Structurally plant cell wall components can be majorly divided into polysaccharide units and polyphenolic lignin units. Plant polysaccharide units are divided into cellulose, hemicellulosic polysaccharides (xyloglucans, xylans, glucomannans and mixed linkage glucans) [7] and pectic polysaccharides (homogalacturonan and rhamnogalacturonan I, II) [8, 9]. Cellulose is the characteristic chemical constituent in all the plant cell walls. Structurally, cellulose is comprised of glucose molecules with β -1,4 glycosidic linkages [9, 10]. These cellulose chains are bound by hydrogen bonds and van der Waals interactions resulting in microfibrils which are crystalline in nature and poses difficulty in enzymatic saccharification [11]. Naturally cellulose exists in crystalline (well-ordered) and amorphous (disordered) forms [11]. Hemicelluloses are hetero polysaccharides including xylans, glucans, mannans and glucomannans with (β -1,4) and (β -1,3) glycosidic linkages. Enzymatic breakdown of hemicellulose was considered easy and simple when compared to cellulose. However, due to the presence of certain recalcitrant oligomeric structures and its complex branching and acetylation patterns make the depolymerization difficult [12]. The enzymatic breakdown of cellulose results only in glucose whereas hemicellulose degradation results in a mixture of different sugars which significantly contain pentoses which are difficult to ferment further [13, 14].

Fermentation of cellulose and hemicellulose monomeric units of plant biomass using microorganisms for the production of bioethanol has gained significance in the recent years. Among different microorganism's, fungi are the efficient plant biomass degraders, especially Basidiomycota phylum. Based on their wood decaying properties, Basidiomycetes fungi are classified into white, brown and soft rot. *Phanerochaete chrysosporium*, a white rot fungi, can efficiently degrade all the components of plant biomass cellulose, hemicellulose and particularly lignin [15]. Complex structure of lignin enables its degradation by specific microorganisms, thus several studies conducted on *P. chrysosporium* were based on lignin degradation while few studies were based on cellulose and hemicellulose degradation mechanisms. Although, whole genome sequence of *P. chrysosporium* showed that it harbors around 240 carbohydrate active enzymes (CAZymes) [15]. Secretome and computational analysis performed by Wymelenberg et al. (2006) has revealed eighteen putative peptide sequences which are allotted to eight specific glycosyl hydrolases [16]. Glycosyl hydrolases were found to be associated with hemicellulose and pectin degradation, putative xylanase and exo glucanase encoding genes *xyn10D*, *exg55A* were detected in ground wood submerged cultures [17, 18]. Three unidentified peptides encoding for glycosyl hydrolases were noticed in carbon limited growth conditions [18]. Availability of annotated whole genome sequence of *P. chrysosporium* have led to the current understanding of its degradative patterns. Based on the PubMed results, 12 large scale gene expression studies were conducted on *P. chrysosporium*, out of which 10 microarray studies, one RNA-Seq and one Long SAGE studies. These studies have revealed various significant facts about the genes and enzymes involved in cellulose, hemicellulose and lignin degradation mechanisms.

Development of online based genome and enzyme databases like CAZy (Carbohydrate Active Enzyme database) [19], FOLy (Fungal Lignin Oxidizing Enzymes) [20], DOE-JGI (Joint Genome Institute) [21] and FungiDB [22] etc., has significantly influenced the present understanding of fungal genomic and proteomic studies. The CAZy database was developed by Vincent et al. (2014), to consolidate various enzymes that are involved in synthesis and breakdown of carbohydrates and other glycoconjugates [19]. CAZy database was divided into five major classes and distributed into 135 glycoside hydrolases, 99 glycosyl transferases, 24 polysaccharide lyases, 16 carbohydrate esterases and 13 auxiliary activities, it also consists of carbohydrate binding module (CBM) [19]. Levasseur, A et al. (2008) has developed FOLy based on the CAZy database structure and is dependent on family based management of sequence information and their respective accession from different public

repositories and its descriptions [20]. FOLy database is divided into two major sections as LO's (Lignin Oxidizing enzymes) and LDA's (Lignin Degrading Auxiliary enzymes) [20].

The availability of whole genome sequences and recently performed genome wide transcriptome studies on *P. chrysosporium* have inspired us for the current metadata analysis. In the last decade, several studies were conducted to understand the molecular mechanisms underlying lignocellulose degradation by *P. chrysosporium*. Although, these gene expression studies have explained about the expression of several cellulolytic and hemicellulolytic enzymes, understanding the common significant genes expressed under varied growth conditions will play a crucial role in biofuel production. In our current study, we have rigorously analyzed the whole transcriptome metadata retrieved from NCBI GEO using GEO2R and Bioconductor packages. To the best of our knowledge this is the first metadata analysis report on *P. chrysosporium* to understand the common gene expression patterns employed for cellulose and hemicellulose degradation.

2.3. Data Analysis Methodology

2.3.1. Data Collection: Microarray datasets were retrieved from Gene Expression Omnibus NCBI-GEO (<https://www.ncbi.nlm.nih.gov/geo/>), a public repository for gene expression datasets. All the microarray datasets collected were based on white rot fungus *P. chrysosporium* cultured on different growth substrates. Till date there are 6 Microarray, 1 RNA sequencing and 1 Long-SAGE (Serial Analysis for Gene Expression) studies based on *P. chrysosporium*, out of which we have analyzed the microarray and RNA sequencing studies to understand the cellulolytic and hemicellulolytic degradation mechanisms. Accession IDs of microarray datasets retrieved from NCBI GEO are GSE14734 [23, 24], GSE14735 [23, 24] GSE54542 [25], GSE27941[26], GSE52922 [27], GSE69008 [28] GSE69461[29] details of these gene expression datasets were shown in the Table 2.1.

2.3.2. Data Analysis: The microarray datasets were analyzed using GEO2R an interactive online tool (<https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html>) and using Bioconductor packages GEOquery and limma based on R software version 3.2.2 and. Following settings were used for analyzing the microarray datasets using GEO2R data analysis tool a) auto detect option used for the log transformation of the data, b) samples and value distribution were obtained using the box-whisker plot c) submitter provided annotations were used for the current analysis. The samples were grouped based on their experimental conditions and further differentially expressed genes were obtained using "Top250" function which

internally uses limma (linear models for microarray data). The top differentially expressed genes are obtained after performing multiple testing correction using Benjamini and Hochberg false discovery rate (FDR) method with a p-value 0.05. The supplier provided annotations for the microarray platform were used for analyzing the obtained differentially expressed genes. Gene annotations mainly InterPro Hits, protein ID and genome position were retrieved for the current analysis. The *P. chrysosporium* RP-78 v2.2 genome annotations were obtained from the MycoCosm (fungal genome repository) [30, 31]. Different analysis tool options such as Gene Ontology (GO), EuKaryotic Orthologous Groups (KOG) and CAZy were used for understanding and biological contextualization of the results. The expression values of differentially expressed genes were used to develop hierarchical clusters using the Cluster 3.0 software [32] (<http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm>) with options selected cluster for both genes and arrays using the complete linkage. The obtained cluster output files were used as input for the Java Treeview [33] software to develop the dendrograms using the standard conditions.

The differentially expressed genes obtained from each datasets were compared using Venny 2.1 [34] (<http://bioinfogp.cnb.csic.es/tools/venny/>) online software to obtain the common genes list. We have retrieved RPKM (Reads Per Kilobase Million sample value for the conditions 96-hour and 40-hour were obtained from the supplementary data provided for GSE69461[29]. The RPKM values from all the samples were further subjected to statistical analysis using R-Bioconductor packages: limma [35], Glimma and edgeR [36-39] Bioconductor packages. The statistically significant genes among different samples were obtained based on the fold change values more than 2.0 also subjected to false discovery rate correction at a p-value 0.05. From the obtained gene list, genes encoding for cellulolytic and hemicellulolytic enzymes and CAZymes were specifically recovered based on the InterPro annotations provided. The fold change values were used for the clustering analysis and the differentially expressed gene list was further compared with other datasets to obtain commonly expressed genes among the above-mentioned gene expression datasets.

Table 2.1: Details of the *P. chrysosporium* transcriptome metadata retrieved from NCBI GEO and NCBI SRA

GEO-ID's	Platform and Technology	Substrate	# Samples	References
GSE54542	NimbleGen <i>Phanerochaete chrysosporium</i> arrays	Oak acetonetic extractives	6	[25]
GSE27941	NimbleGen <i>Phanerochaete chrysosporium</i> arrays	Ball milled aspen, Ball milled pine	6	[26]
GSE52922	NimbleGen <i>Phanerochaete chrysosporium</i> arrays	P717 hybrid line, Transgenic line 82 Transgenic line 64	9	[27]
GSE14734	NimbleGen <i>Phanerochaete chrysosporium</i> arrays	Cellulose, Glucose, Ball milled aspen	9	[23, 24]
GSE14735	NimbleGen <i>Phanerochaete chrysosporium</i> arrays	Replete medium, Carbon limited Nitrogen limited	9	[23, 24]
GSE69008	NimbleGen <i>Phanerochaete chrysosporium</i> arrays	Poplar wood substrates	24	[28]
GSE69461	Illumina HiSeq 2000	<i>Picea glauca</i> (spruce sapwood)	18	[29]

2.3.3. Overview of Data Analysis

The whole transcriptome datasets considered for our present study can be divided into customized growth medium (Highley's basal medium supplemented with cellulose, glucose or other commercially available nutrients) and complex natural plant biomass medium (ball milled aspen, ball milled pine, spruce wood and poplar wood substrates) based on the media composition used for culturing *P. chrysosporium*. The customized growth mediums were used in gene expression studies with accession Ids GSE14734 and GSE14735. Growth medium consisting HBM supplemented with 0.5 % (wt/vol) of BMA or cellulose or glucose as sole carbon source was used for the culturing of *P. chrysosporium* (GSE14734) [23, 24]. For GSE14735 dataset *P. chrysosporium* was cultured on three different growth mediums a) replete B3 medium (with adequate carbon and nitrogen source), b) carbon limited medium c) nitrogen limited medium [23, 24]. The complex natural plant biomass growth medium was used in gene expression studies with accession numbers GSE27941, GSE52922, GSE54542, GSE69008 and GSE69461. Growth medium used for GSE27941 consists of 0.5% of ball milled aspen and ball milled pine as the sole carbon source supplemented with HBM [26]. Similarly, growth medium used for GSE52922 consisted Wiley milled chemically distinct wood substrates of *Populus trichocarpa* P717 (parental hybrid clone line) with 65 mol% of syringyl units and the two transgenic lines 64 and 82 with 94 and 85 mol % of syringyl units, supplemented with HBM [27]. GSE69461 medium consists of microtomed tangential sections

of *Picea glauca* (40mm long, 10mm wide and 40 mm thick with a dry weight of 7mg) covered with 90µl of agar supplemented with nitrogen mineral salt medium [29]. GSE54542 medium contained Oak heartwood fine powdered samples extracted using acetone and resuspended further in DMSO followed by a set of extraction processes [25]. Finally, GSE69008 growth medium consisted of chemically distinct *Populus trichocarpa* wood substrates, which can be divided in to A (high lignin-low glucose) B (low lignin-high glucose) and C (average lignin-average glucose) conditions [28].

2.4. Results

2.4.1. Distribution of CAZymes in *P. chrysosporium* Genome: Several studies have already proved the eccentric lignin degrading abilities of *P. chrysosporium*. However, very few studies were conducted till today to understand the cellulose and hemicellulose degradation mechanisms employed by this organism. Whole genome studies of *P. chrysosporium* conducted in 2004 by Martinez et.al, have reported that it encodes around 240 putative carbohydrate active enzymes, which mainly encodes 66 glycoside hydrolases, 14 carbohydrate esterases and 57 glycosyltransferases [15]. Present day annotated genome of *P. chrysosporium* RP-78 version 2.2, genome codes for around 440 putative carbohydrate active enzymes divided into 89 Auxiliary activity enzymes, 65 carbohydrate binding modules, 20 carbohydrate esterases, 181 glycoside hydrolases, 70 glycosyl transferases and 6 polysaccharide lyases (Figure 2.1). Extracellular cellulases and hemicellulases secreted by fungi can be grouped under glycoside hydrolases [40]. According to Martinez et al. (2004), the genome of *P. chrysosporium* encodes at least 40 genes for putative endoglucanases (GH5, GH9, GH12, GH61, GH74) 7 exo-cellobiohydrolases (GH6, GH7) 9 β-glucosidases (GH1, GH3) and 5 polygalacturonase (GH28) [15].

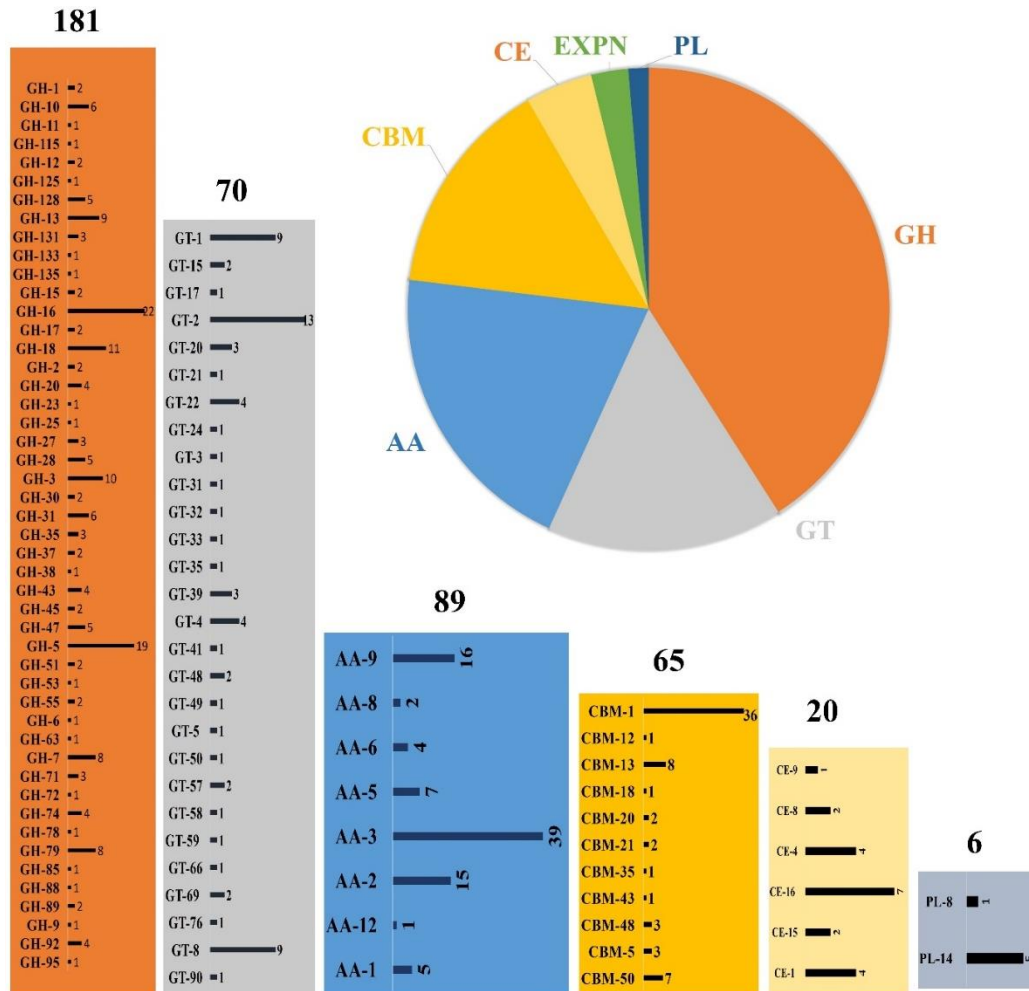


Figure 2.1: Schematic representation of CAZymes distribution in *Phanerochaete chrysosporium* genome, GH (glycoside hydrolase), GT (glycosyl transferases), AA (Auxiliary activities), CBM (Carbohydrate binding modules), CE (Carbohydrate esterases) and PL (Polysaccharide lyases); the numbers represented on top of each box represents the number of genes encoding for that particular class of enzymes respectively.

Pedro M et al. (2003) have reported the global correlation studies of CAZymes and their corresponding total number of open reading frames among the bacterial and eukaryotic genomes [41]. The number of genes encoding glycosyl transferases exceeds occasionally by large factor compared to glycoside hydrolases encoding genes observed in eukaryote genome sequences such as *Saccharomyces cerevisiae*, *Saccharomyces pombe*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Homo sapiens* and *Drosophila melanogaster*. Contrastingly, the genomic studies of *P. chrysosporium* revealed large number of glycosyl hydrolases encoding genes rather than glycosyl transferases [15].

Some of the significant findings from the *P. chrysosporium* gene expression studies with respect to CAZymes were discussed as follows. Wymelenberg et al. (2009) have reported the

gene expression data of *P. chrysosporium* cultured on cellulose, glucose and ball milled aspen (GSE14736) [23, 24]. It revealed several transcripts encoding various cellulases belonging to GH-5, 6, 7 and hemicellulose depolymerizing enzymes such as xylanases, mannanases, α - β -galactosidases, xyloglucanase, arabinofuranosidase, polygalacturonase and feruloyl esterases found to be highly upregulated in ball milled aspen along with GH-61 proteins. CAZy encoding genes were not expressed in nitrogen limited medium [23, 24]. Wymelenberg et al. (2011) have conducted the *P. chrysosporium* gene expression study by culturing it on ball milled aspen and ball milled pine (GSE27941) [26]. The obtained results have supported earlier investigations on the cellulolytic system of *P. chrysosporium*, various transcripts coding for endo and exoglucanases (CBH1, CEL7D, CBH2, CEL5A) and GH-61 were significantly expressed in BMA samples. While genes encoding cellobiose dehydrogenase (CDH) and aldose-1-epimerase transcripts were highly expressed in BMP but not in BMA [26]. Thuillier, A et al. (2014) have performed the whole transcriptomic study of *P. chrysosporium* cultured on oak acetic extractives (GSE54542) [25]. Genes encoding various CAZymes such as cellobiohydrolases, endoglucanases, β -glucosidase, endoxylanases, endoglucanases and mannanases were found to be expressed in control conditions and the same genes were downregulated in oak extractives [25]. In 2014, Gaskell et al. has performed gene expression studies on *P. chrysosporium* to understand the influence of *Populus* genotype (P717, 82 and 64 transgenic lines) on its gene expression (GSE52922) [27]. Results from this study showed that genes coding for GH-6 (CBH2), AA-3 (CDH), AA9 (LPMO), GH5 (endoglucanases) were highly upregulated in P717. It was also reported that transcripts coding for LPMO's, cellobiose dehydrogenase and aldose-1-epimerase were highly expressed in all the three cultures. In order to understand the lignocellulolytic abilities of *P. chrysosporium* on spruce wood samples, Korripally et al. (2015) have conducted a gene expression study (GSE69461) [29] and reported that 23 transcripts coding for putative hemicellulases showed more than 4-fold expression. Along with the hemicellulases, genes coding for carbohydrate esterases were also found to be highly expressed in 40 hour samples, which show that these hemicellulases and carbohydrate esterases are mutually involved in depolymerization [29]. Recently, Skyba et al. (2016) has conducted a gene expression study to understand the involvement of specific genes and enzymes involved in lignocellulose degradation by culturing *P. chrysosporium* on chemically distinct *P. trichocarpa* wood substrates [28]. Results from this study revealed the genes coding for AA9 (LPMO), GH13 were down regulated when cultured on low glucose substrates [28]. The metadata

analysis results from the above-mentioned gene expression studies have resulted in differentially expressed common gene lists involved in cellulose and hemicellulose degradation (Figure 2.2).

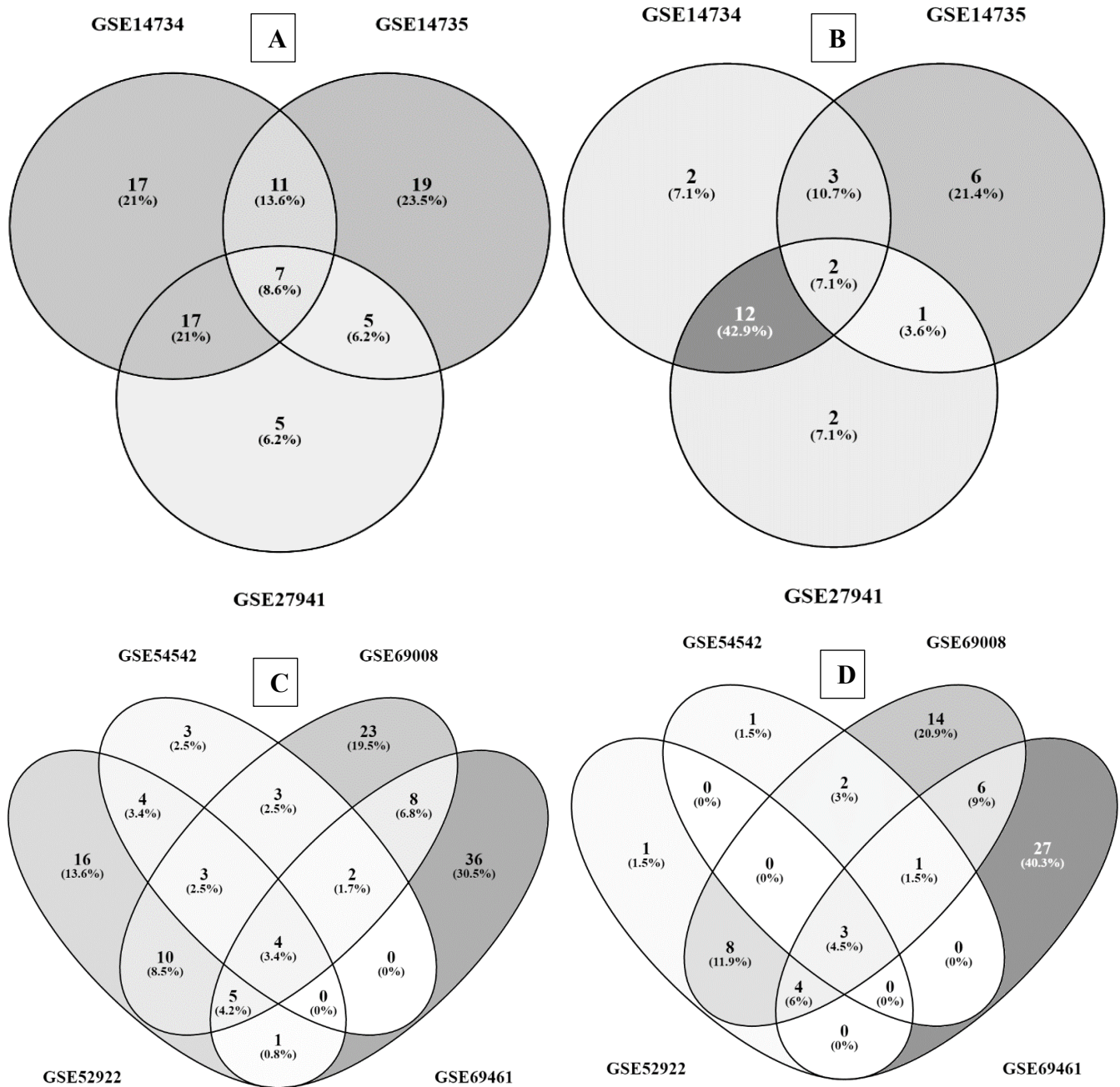


Figure 2.2: Three way and four-way Venn diagrams showing the common differentially expressed genes of *P. chrysosporium* involved in cellulose and hemicellulose degradation; (A) Differentially expressed genes and (B) Differentially expressed CAZymes, resulted from the datasets GSE14734, GSE14735 and GSE27941; (C) Differentially expressed genes and (D) Differentially expressed CAZymes resulted from the datasets GSE54542, GSE52922, GSE69008 and GSE69461.

2.4.2. Expression of Cellulose Degrading Glycoside Hydrolases

Glycoside hydrolases are present in large numbers among carbohydrate active enzymes with wide variety of functional properties. Enzymes belonging to this class cleaves the glycosidic bond linkages of glycosides, glycans and other glycoconjugates [42]. Metadata analysis of *P. chrysosporium* gene expression studies have resulted in common differentially expressed glycoside hydrolase classes involved in cellulose degradation such as GH-5, GH-6, GH-7, GH-9, GH-44, GH-45, GH-48 and GH-61 [Table 2.2]. Genes encoding GH-5, GH-6, GH-7 and GH-61 (AA9) were found to be differentially expressed by *P. chrysosporium* cultured on cellulose (GSE14734), ball milled aspen (GSE14734, GSE27941). It was reported that, in *Trichoderma reesei* genes coding for cellulolytic and xylanolytic enzymes are regulated by XYR1 and CRE1 transcriptional factors corresponding to the carbon source used for its growth [43]. It was also seen that, presence of glucose in the growth media represses the expression of cellulolytic and xylanolytic genes [44-47]. Above mentioned glycoside hydrolases significantly down regulated in nitrogen limited, carbon limited growth mediums (GSE14735). When *P. chrysosporium* was cultured on (*P. trichocarpa*) P717, 82 transgenic lines (GSE52922), genes encoding for GH-2, GH-3, GH-5 were differentially expressed when compared to 64 transgenic line. In growth mediums containing chemically distinct *P. trichocarpa*, glycoside hydrolase encoding genes GH-5, GH-6, GH-7, GH-61 were found to be highly expressed in high glucose-low lignin and average lignin-average glucose (GSE69008) samples. Most of the cellulose degrading glycoside hydrolase encoding genes were found to be down-regulated in oak acetic extracts (GSE54542). Glycoside hydrolases involved in cellulose breakdown such as GH-2, 5, 6, 7, 45, 48, 61, were significantly expressed in 96-hour spruce wood samples (GSE69461).

Table 2.2: Glycoside hydrolases differentially expressed among different gene expression datasets

Glycoside Hydrolases	Representing enzyme (Cellulose degrading)	Glycoside Hydrolases	Representing enzyme (Hemicellulose degrading)	Glycoside Hydrolases	Representing enzyme (Oligosaccharide)
GH-1	β-Glucosidase β-Galactosidase	GH-1	β-glucosidase β-galactosidase β-mannosidase β-glucuronidase	GH-1	β-glucosidase β-galactosidase β-mannosidase β-glucuronidase
		GH-2	β-Galactosidase β-Mannosidase		

		GH-5	Endo- β -1,4-glucanase, β -glucosidase, Endo- β -1,4-xylanase		
GH-3	β -glucosidase xylan 1,4- β - xylosidase	GH-8	Endo-1,4- β -xylanase Chitosanase Cellulase		
GH-5	Endo- β -1,4- glucanase, β - glucosidase, Endo- β -1,4- xylanase				
GH-6	Endoglucanase Cellobiohydrolase	GH-10	Endo-1,4- β -xylanase Endo-1,3- β -xylanase	GH-2	β -galactosidase β -mannosidase β -glucuronidase
GH-7	Endo- β -1,4- glucanase, endo- β -1,3-1,4- glucanase	GH-11	Endo-1,4- β -xylanase Endo-1,3- β -xylanase	GH-3	β -glucosidase xylan-1,4- β - xylosidase α -L- arabinofuranosidase
GH-9	Endoglucanase Endo- β -1,4- glucanase β -glucosidase	GH-12	Endoglucanase Xyloglucan hydrolase	GH-28	Polygalacturonase Exo- polygalacturonase Rhamnogalacturonase
GH-44	Endoglucanase Xyloglucanase	GH-26	β -Mannanase exo- β - 1,4mannobiohydrolase β -1,3-xylanase Mannobiose producing exo- β - mannanase	GH-29	α -L-fucosidase α -1,3-1,4-L-fucosidase
GH-45	Endoglucanase	GH-38	α -mannosidase mannosyl- oligosaccharide, α -1,2- mannosidase	GH-35	β -galactosidase exo- β - glucosaminidase exo- β -1,4-galactanase
GH-48	Reducing end- acting cellobiohydrolase, endo- β -1,4- glucanase	GH-43	β -xylosidase α -L- arabinofuranosidase		
		GH-47	α -mannosidase		
GH-61	Lytic Polysaccharide Monooxygenase (LPMO)	GH-61	Lytic Polysaccharide Monooxygenase (LPMO)	GH-39	β -xylosidase α -L-iduronidase
				GH-42	β -galactosidase, α -L- arabinopyranosidase

2.4.3. Expression of Hemicellulose Degrading CAZymes

Most of the naturally occurring plant polysaccharides are partly esterified by acetic acid to protect from microbial glycoside hydrolases as they cannot breakdown acetylated glycosyl units. As a result, microorganisms secrete a set of enzymes known as carbohydrate esterases which targets the acetyl groups of plant polysaccharides [48]. Carbohydrate esterases (CE) catalyze O-de or N-deacetylation of saccharides substituted with either esters or amides, where these sugars play the role of alcohol or amine [48]. Substrates for CEs can be majorly

classified into pectin methyl esters (acid form of sugar) and acetylated xylan (sugar alcohol). Presently carbohydrate esterases are classified into 16 different classes in the CAZy database. Therefore, in contrast to cellulose, hemicellulose degradation is commenced by a set of glycoside hydrolase and carbohydrate esterase class enzymes. *P. chrysosporium* secretes several hemicellulose degrading glycoside hydrolase class enzymes such as endo-1,4- β -xylanase, β -mannosidase, β -xylosidase and several other enzymes. Metadata analysis of the above mentioned publicly available gene expression datasets have resulted in following hemicellulose depolymerizing glycoside hydrolase classes GH-5, GH-8, GH-10, GH-11, GH-12, GH-26, GH-38, GH-43, GH-47 [Table 2.2]. When *P. chrysosporium* was cultured on ball milled aspen (GSE14734, GSE27941), genes encoding glycoside hydrolase such as GH-1, GH-2, GH-39, GH-43 which was also expressed in cellulose, and GH-47 were differentially expressed along with acetyl xylan esterases (CE-1). When cultured on replete nutrient medium (GSE14735) genes encoding GH-1, GH-5, GH-7, GH-39, GH-61 glycoside hydrolases were highly expressed. Glycoside hydrolase encoding genes GH-2, GH-16, GH-20 were found to be highly expressed in oak acetic extractives and GH-3, GH-43, GH-47 were highly expressed in control samples (GSE54542). In GSE52922, *P. chrysosporium* differentially expressed glycoside hydrolase classes such as GH-2, GH-10, GH-39 in transgenic line 82 and P717, whereas GH-43 encoding genes were expressed in transgenic line 64 growth substrates. We have also observed that hemicellulose degrading glycoside hydrolase GH-5, GH-10, GH-13, GH-43, GH-47 were expressed when *P. chrysosporium* cultured on high lignin and low glucose samples of *P. trichocarpa* wood substrates (GSE69008). In GSE69461, when *P. chrysosporium* cultured on spruce wood samples glycoside hydrolase encoding genes such as GH-5, GH-10, GH-47, GH-53, GH-115, GH-125 and GH-131 were highly upregulated in 96-hour growth samples. Along with these glycoside hydrolases, genes encoding carbohydrate esterases such as acetyl xylan esterase (CE-1) (CE-4), pectin methyl esterase (CE-8), 4-O-methyl-glucuronoyl methylesterase (CE-15) and acetylerase (CE-16) were also highly expressed in 96-hours samples. At the same time, acetyl xylan esterase (CE-4) and N-acetylglucosamine 6-phosphate deacetylase (CE-9) were highly expressed by *P. chrysosporium* in 40-hours spruce wood samples.

2.4.4. Glycoside Hydrolases Involved in Fungal Cell-Wall Synthesis

Composition of fungal cell walls are dynamic and differs significantly from that of cellulose based plant cell walls [49]. Fungal cell wall majorly consists of polysaccharide units and glycoproteins in which chitin and glucan chains are significantly high. However, other cell wall components vary among the fungal species. Principally, fungal glucan component occurs in long chains of β -1,3-linked glucose units and chitin is synthesized as chains of beta-1,4-linked N-acetylglucosamine units, however chitin is less abundant in fungal cell wall when compared to glucan chains or the glycoprotein portion [49]. Thus, several genes encoding for the enzymes involved in synthesis of β -1,3-glucans and chitin are constantly expressed in fungal cell by protecting the fungi from changes in osmotic pressure and various environmental stress [49]. Glycoside hydrolases encoding genes such as GH-16 (chitin β -1,6-glucanosyltransferase), GH-18 (chitinase), GH-37 (α,α -trehalose), GH-55 (exo-endo- β -1,3-glucanases), GH-71 (α -1,3-glucanase) GH-128 (β -1,3-glucanase). Genes coding for the GH-18 was found to be differentially expressed by *P. chrysosporium* in all the datasets among various growth conditions. We have also found that GH-16 encoding genes were highly upregulated in oak acetonic extractives. While genes coding for glycoside hydrolase classes such as GH-16, GH-37, GH-55, GH-71, GH-128 were found to be highly expressed in complex natural plant biomass growth substrates (40-hours spruce wood, high lignin-low glucose, average lignin-average glucose), which might be due to the exposure of fungal cell to the various stressful conditions.

2.4.5. Expression of Glycosyl Transferases Encoding Genes

Glycosyl transferase enzymes catalyze glycosidic linkages utilizing activated sugar donor containing a phosphate group (mostly nucleoside diphosphate sugars e.g. UDP Gal, GDP Man). In simple terms GT's catalyzes the reactions for joining sugars to aglycone, thus playing a significant role in synthesis of oligosaccharides [50]. From the current metadata analysis, we have found that differentially expressed common glycosyl transferase classes were GT-1, GT-2, GT-4, GT-8, GT-15, GT-20, GT-35, GT-39, GT-48. The type of growth mediums on which *P. chrysosporium* was cultured, have significant effect on the expression of glycosyl transferase classes. Genes encoding GT-4, GT-8, GT-48 enzyme classes were highly expressed on cellulose growth medium whereas GT-1, GT-2, GT-20, GT-39 were expressed on ball milled aspen, and GT-8, GT-48 genes were upregulated on glucose and replete growth mediums. GT-2 class of enzymes were highly expressed on Oak acetonic extractives and on control conditions GT-8 encoding genes were upregulated. *P. chrysosporium* when cultured on chemically distinct *P. trichocarpa* medium, discrete glycosyl transferase classes

were highly expressed, they are a) GT-1, GT-4, GT-20 on high lignin- low glucose conditions b), GT-2, GT-8, GT-15, GT-35, GT-39, GT-48 on low lignin-high glucose mediums c) GT-2, GT-8, GT-15, GT-20, GT-35, GT-39, GT-48 on average lignin-average glucose samples. Similarly, class of GT-39 encoding genes were expressed on P717 and 82 transgenic lines. Genes encoding glycosyl transferases such as GT-1, GT-2, GT-15, GT-48 were highly upregulated in 40-hour spruce wood samples.

2.4.6. Gene Expression of Polysaccharide Lyases

Polysaccharide lyases are class of enzymes which are involved in breakdown of activated glycosidic bonds involved in joining certain acidic polysaccharide units [51]. Polysaccharide lyases cleave the polysaccharide units through a eliminase mechanism rather than a hydrolysis resulting in oligosaccharide units [51]. Present day CAZy database has classified polysaccharide lyases into 24 different classes. The metadata analysis of gene expression datasets has shown that, hyaluronate lyase or chondroitin AC lyase (PL-8) encoding genes were down regulated in oak acetic extracts and upregulated in 96-hour spruce wood, ball milled aspen, nitrogen and carbon limited and 64, 82 transgenic line samples. Genes coding for β -1,4-glucuronan lyase or alginate lyase (PL-14) is highly upregulated in 40-hour spruce wood samples.

Table 2.3: Glycosyl transferases and carbohydrate esterases differentially expressed among different gene expression datasets

Glycosyl Transferases	Representing enzyme	Carbohydrate Esterases	Representing enzymes	Polysaccharide Lyases	Representing enzyme
GT-1	UDP-glucuronosyltransferase, zeatin O- β -xylosyltransferase, indole-3-acetate β -glucosyltransferase	CE-1	acetyl xylan esterase, cinnamoyl esterase, feruloyl esterase, carboxylesterase, S-formylglutathione hydrolase,	PL-8	Hyaluronate lyase, Chondroitin AC lyase, Xanthan lyase
GT-2	cellulose synthase, chitin synthase, dolichyl-phosphate β -D-mannosyltransferase,	CE-4	acetyl xylan esterase, chitin deacetylase, chitooligosaccharide deacetylase,		
GT-4	sucrose synthase, sucrose-phosphate synthase, α -glucosyltransferase, lipopolysaccharide N-acetylglucosaminyltransferase	CE-8	Pectin methylesterase		
GT-8	lipopolysaccharide α -1,3-galactosyltransferase,	CE-9	N-acetylglucosamine 6-phosphate deacetylase,		

	UDP-lipopolysaccharide α -1,2-glucosyltransferase, lipopolysaccharide glucosyltransferase 1,			PL-14	Alginate lyase, exo-oligo alginate lyase, β -1,4-glucuronan lyase
GT-15	glycolipid 2- α -mannosyltransferase, GDP-Man: α -1,2-mannosyltransferase,	CE-15	4-O-methyl-glucuronoyl methyltransferase,		
GT-20	α , α -trehalose-phosphate synthase, Glucosylglycerol-phosphate synthase, trehalose-6-P phosphatase,	CE-16	Acetyltransferase		
GT-35	glycogen or starch phosphorylase				
GT-39	protein α -mannosyltransferase				
GT-48	1,3- β -glucan synthase				

2.4.7. Major Facilitator Superfamily Encoding Genes

Major facilitator superfamily is one of the largest membrane transporter proteins involved in the intracellular transport of a wide variety of chemical compound by solute uniport, solute/solute antiport, solute/cation antiport and solute/cation symport [52]. MFS proteins are involved in transport of simple sugars, oligosaccharides, amino-acids, nucleosides, organophosphate esters and many other compounds [52]. Sugar transporters are class of membrane proteins involved in binding and transport of different carbohydrates, alcohols and acid compounds [53]. Efficient degradation and metabolism of plant cell wall polysaccharides by *P. chrysosporium* requires sugar transporters along with extracellular glycoside hydrolases [54]. Genes encoding for sugar transporters were highly expressed in cellulose, glucose, replete, ball milled aspen, high glucose-low lignin, average lignin-average glucose, 96-hours culture samples.

2.4.8. Genes Encoding for Carbohydrate Metabolism

Several genes involved in expression and regulation of glycolysis were also highly expressed in the actively growing cultures of *P. chrysosporium*. Genes encoding for aldose-1-epimerase were highly expressed in cellulose, glucose, control, high glucose-low lignin, average lignin-average glucose, 40-hour spruce wood cultures. Genes coding for enzymes involved in regulation of the glycolysis such as hexokinase, phosphofructokinase and pyruvate kinase were differentially expressed in cellulose, carbon-limited, high glucose-low lignin and average glucose-average lignin, 40-hour spruce wood samples. Fructose-2,6 bisphosphatase,

phosphoglucomutase genes were upregulated in glucose, replete, ball milled aspen samples. Enzymes involved in pentose phosphate pathway such as fructose-6-phosphoketolase, transketolase, transaldolase encoding genes were expressed in replete, high glucose-low lignin, average lignin-average glucose and 40-hour spruce wood growth samples. Alpha amylase, aldo/keto reductase, aldehyde dehydrogenase, zinc-containing alcohol dehydrogenase encoding genes were expressed in cellulose, glucose, nitrogen limited, carbon-limited, ball milled aspen, control, 64 transgenic line, high glucose-low lignin, average lignin-average glucose, 40-hours samples. Genes coding for pyruvate decarboxylase, pyruvate kinase was highly upregulated in carbon-limited, nitrogen limited, oak acetic extractive and 40-hour spruce wood samples.

2.4.9. Genes Encoding for Carbohydrate Binding Modules (CBM)

The cellulose and hemicellulose degrading hydrolytic enzymes secreted by polysaccharide degrading microorganisms contains a discrete module with composite molecular architecture linked by unstructured sequences, which consists of a catalytic module and a carbohydrate binding module (CBM) [55]. Carbohydrate binding modules were firstly classified as cellulose binding modules (CBD) as various modules were found to bind with cellulose [56-58]. However, several studies have reported the modules binding to the carbohydrates other than cellulose, which was the main reason behind renaming these modules as CBMs [55]. Currently, the carbohydrate binding modules (CBMs) are divided into 80 major classes in CAZy database. CBM-12 encoding genes were highly expressed in cellulose (GSE14734) ball milled aspen (GSE27941). Genes encoding CBM-1 and CBM-13 were expressed in 96-hour samples and CBM12, CBM-21, CBM-48 and CBM-50 were found to be expressed in 40-hours samples respectively (GSE69461). Similarly, genes encoding CBM-12, CBM-48, were expressed in high glucose-low lignin and average lignin-average glucose growth conditions. The genes coding for carbohydrate binding WSC domain were differentially expressed in cellulose, nitrogen limited, ball milled aspen, control, high lignin-low glucose growth conditions. According to Wymelenberg et al (2006), careful inspection of *cro3*, *cro4* and *cro5* cDNA sequences revealed 2 to 4 tandem copies of WSC domain which might be having a role in cell wall integrity and stress component [59]. The exact function of WSC-domain containing proteins in *P. chrysosporium* is not clear till today [59]. In *Trichoderma harzianum*, β -1,3- exoglucanase protein which is associated with host cell wall degradation was found to contain 2 copies of WSC domain [60], contrastingly in

Saccharomyces cerevisiae, the WSC-containing proteins are involved in maintaining cell wall integrity and heat shock response [61] .

2.5. Discussion

Analyzing the common gene expression patterns of *P. chrysosporium* involved in degradation of plant biomass will considerably enhance our current understanding. We have analyzed the gene expression datasets retrieved from public repositories based on *P. chrysosporium* growth conditions by using GEO2R and Bioconductor packages, to find the common differentially expressed genes among various datasets. Several genes encoding for carbohydrate active enzymes were found to be commonly expressed in *P. chrysosporium* gene expression datasets (Figure 2.3). Classically, it was well known that the process of cellulose degradation by *P. chrysosporium* (filamentous fungus) occurs through combination of hydrolytic reactions caused by a) endo-1,4- β -glucanases b) exo-1,4- β -glucanases (cellobiohydrolases) c) β -glucosidases [6]. Initial cellulose degradation by the filamentous fungus occurs through cellulases which are majorly classified under glycoside hydrolases, resulting in cellobiose [62]. Cellobiose, the major product obtained during the cellulose biodegradation is hydrolyzed into two molecules of glucose by cellobiose dehydrogenase (CDH) and β -glucosidase (BGL) [62]. It was reported that *P. chrysosporium* secretes a single isozyme coding for both BGL and CDH, initially CDH acts on cellobiose resulting in glucose and gluconolactone which are hydrolyzed by BGL [63]. Gene expression studies of *P. chrysosporium* conducted in the recent years have revealed several significant facts on expression patterns of cellulose degrading glycoside hydrolases and other carbohydrate active enzymes. Endoglucanase (endo-1,4- β -glucanase or 1,4- β -D-glucan 4-glucanohydrolases) is required for the initial hydrolysis of cellulose for breaking internal glycosidic bonds. Majority of the genes encoding for endo and exo glucanases belonging to GH-5, GH-6, GH-7, GH-9, GH-44, GH-45 and GH48 glycoside hydrolase classes were differentially expressed when cultured in cellulose, ball milled aspen as the carbon source (GSE14734), average lignin- average glucose (B10, C10, B20 and C20) (GSE69008) ball milled aspen (GSE27941). While these genes were down regulated in nitrogen limited, carbon limited (GSE14735), ball milled pine (GSE27941) transgenic line 64 (GSE52922), oak acetic extracts (GSE54542) spruce wood 40-hour samples. In fact, several genes encoding cellulolytic and hemicellulolytic enzymes were not expressed in nitrogen and carbon limited conditions (GSE14735). Lytic polysaccharide monooxygenases (LPMO) are class of enzymes

which can potentially breakdown the recalcitrant plant polysaccharide units by cooperating with cellobiose dehydrogenase enzyme (not necessarily as some organisms lack the CDH coding genes) resulting in production of oxidized and non-oxidized chains [64-67]. Glycoside hydrolase class 13 enzymes with alpha-amylase activity functions similar to AA-9 class enzymes which also act on polysaccharide units. Genes coding the AA-9 (LPMO) class enzymes were downregulated in low glucose-high lignin and also in only glucose (GSE14734) conditions, which might be due to its preference for lignocellulosic substrates [68]. The genes encoding these enzymes were also found to be down regulated in 40- hours samples, while these genes were highly expressed in 96-hours (GSE69461), which can be explained by expression of LPMO's during the initial days of fungal incubation [69].

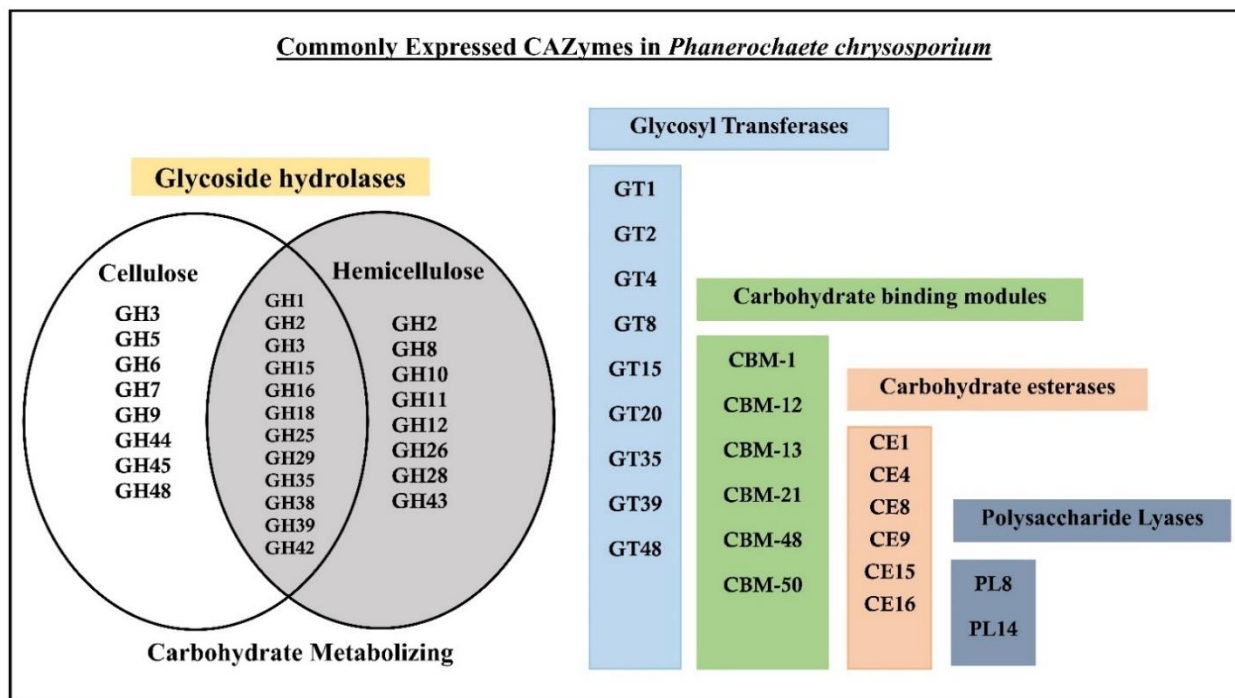


Figure 2.3: Venn diagram showing the cellulose degrading glycoside hydrolases (left), oligosaccharide metabolizing (center) hemicellulose degrading (right) and other CAZY enzyme classes.

Several studies have reported that occurrence of plant polymers such as cellulose, xylan in the growth medium of *T. reesei* induces the increased production of cellulolytic and hemicellulolytic enzymes [70-73]. The oligosaccharides such as sophorose [74, 75], β -cellobiono-1,5-lactone, xylobiose, D-galactose, D-xylose and lactose in the growth medium have also been reported to increase the genes encoding cellulolytic and hemicellulolytic enzymes [76-81]. It was always a fascinating fact that how these large insoluble polymers

cellulose and xylan can induce the production of cellulases and hemicellulases, as the fungal cells usually do not incorporate these polymers [82]. Researchers investigated on this fact and postulated that soluble low molecular weight compounds derived from cellulose induce the hydrolytic enzyme production by the fungal cells. Another proposed mechanism is that, minimal levels of extracellular cellulases especially CEL7A and CEL6A produced by fungal cells act on cellulose liberating a soluble inducer which enter the cell and induces the production of hydrolytic enzymes [83, 84]. These facts can explain the higher expression of endoglucanases and exo-glucanases in growth conditions mentioned above (Figure 2.4).

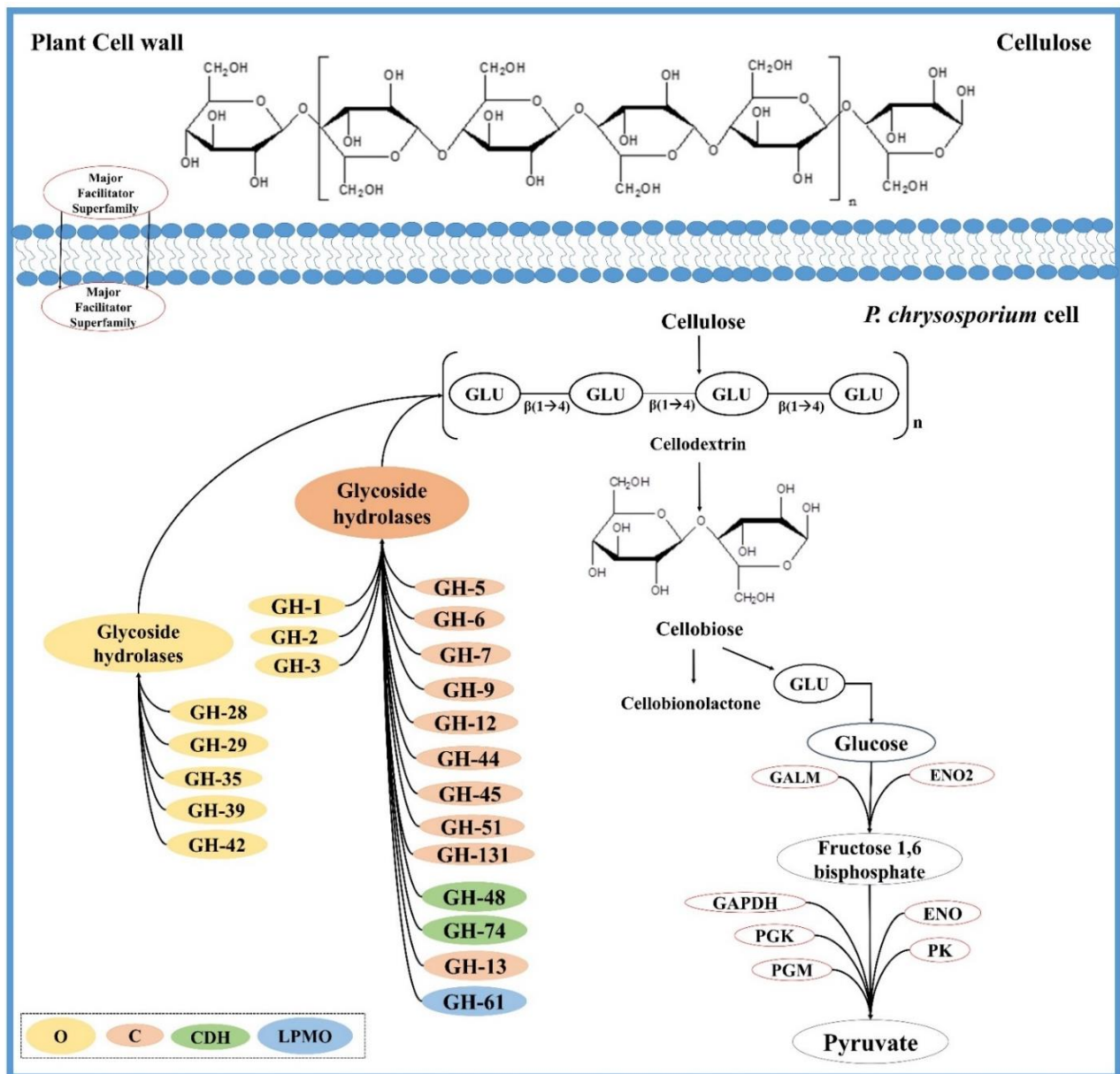


Figure 2.4: Tentative network of genes coding for *P. chrysosporium* cellulose degrading enzymes and cellulose degradation mechanism; O (CAZymes involved in oligosaccharide degrading), C (cellulolytic CAZymes), CDH (cellobiose dehydrogenase encoding CAZymes) and LPMO (CAZymes coding for lytic polysaccharide monoxygenases).

Most of the polysaccharide degrading microorganisms breakdown the hemicellulose units into monomeric sugars and acetic acid [85]. Most of the known hemicellulases are frequently classified based on their action on discrete substrates. Xylan forms the major carbohydrate constituent of hemicellulose and its degradation requires the parallel action of various hydrolytic enzymes [85]. Initially, xylan is degraded by endo-1,4- β -xylanase generating oligosaccharide chains and later xylan 1,4- β -xylosidase breaks the xylan oligosaccharides to xylose units [86]. Enzymes such as acetyl xylan esterases, α -L-arabinofuronosidases, α -4-O-methyl glucuronosidases and ferrulic, *p*-coumaric esterases are also efficiently required for the breakdown of wood xylan and mannans. along with the above mentioned major xylan degrading enzymes [85]. As mentioned by Pérez et al. (2002) degradation of highly occurring hemicellulosic polymer, O-acetyl-4-O-methylglucuronxylan takes place through synergistic action of endo-1,4- β -xylanases (which breaks down the whole polymer to oligosaccharide units) acetyl esterases (eliminates acetyl groups), α -glucuronidase or α -galactosidases removes galactose residues and finally β -mannosidase or β -glycosidase (cleaves the endomannose generated oligomer β -1,4-bonds) [85]. It was known that *P. chrysosporium* produces several endoxylanase encoding genes [87]. Glycoside hydrolases such as endoxylanases, β -xylosidases, β -galactosidases, β -mannosidases and other hemicellulases encoding genes such as GH-1, GH-2, GH-8, GH-10, GH-11, GH-26, GH-38, GH-43 and GH-47 were majorly found to be differentially expressed when *P. chrysosporium* was cultured in plant biomass containing growth mediums. We have observed that, process of hemicellulose degradation by *P. chrysosporium* occurs by combinatorial action of glycoside hydrolases and carbohydrate esterases (CE-1, 4, 8, 9, 15 and 16) as observed in the gene expression profiles of GSE69008, GSE69461 datasets. Based on our metadata analysis we have found that *P. chrysosporium* secrete a wide range of hydrolytic enzymes involved in the degradation of lignin and hemicellulose followed by secretion of cellulolytic enzymes when cultured on complex natural plant biomass (Figure 2.5).

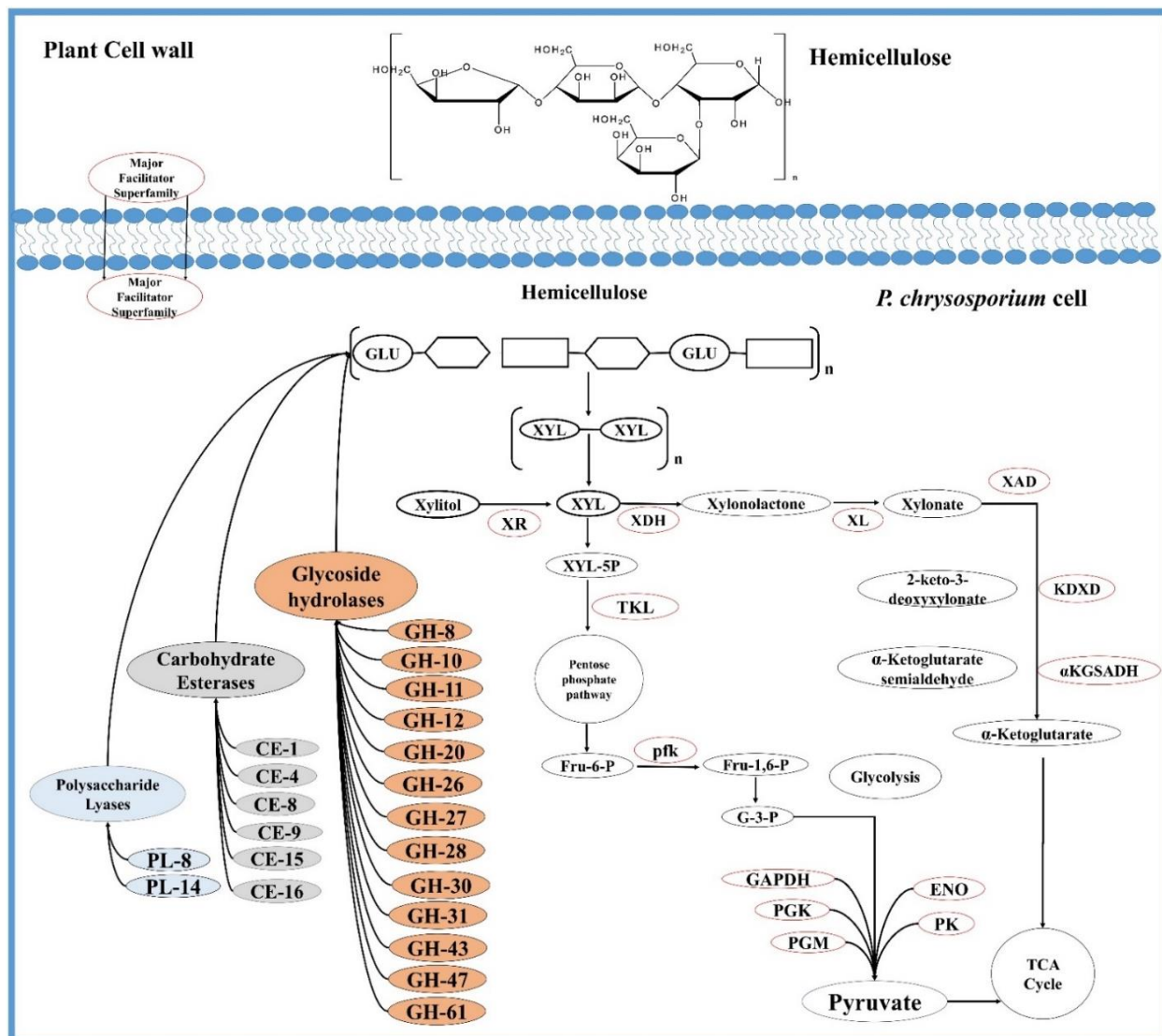


Figure 2.5: Tentative network of genes coding for *P. chrysosporium* hemicellulose degrading enzymes and hemicellulose degradation mechanism

Whole genome and transcriptome studies of *P. chrysosporium* conducted in the recent years have revealed several significant facts on the molecular mechanisms employed by this white rot fungus during degradation of plant cell wall components. Our present study based on metadata analysis of *P. chrysosporium* gene expression data provides a list of common cellulolytic and hemicellulolytic enzymes expressed among various datasets under different growth conditions (customized medium and complex plant biomass containing medium). This study proves that gene expression of *P. chrysosporium* is strongly influenced by the growth substrates and the incubation periods. The gene expression profiles observed in GSE69008 and GSE69461 show that, when *P. chrysosporium* is cultured on plant biomass it initially

secretetes various lignin and hemicellulolytic enzymes followed by secretion of cellulolytic enzymes in the later growth stages. The future prospects of this work would be determining the precise functions of various significantly expressed hypothetical proteins, as most of them were classified as proteins with unknown functions. We believe that our present work will potentially enhance the present day understanding about *P. chrysosporium* and its lignocellulosic degrading patterns and will also play a crucial role in developing overexpressed gene products of *P. chrysosporium* for targeting the cellulosic and hemicellulosic feedstocks.

References

1. Losordo Z, McBride J, Rooyen JV, Wenger K, Willies D, Froehlich A, et al. Cost competitive second-generation ethanol production from hemicellulose in a Brazilian sugarcane biorefinery. *Biofuels, Bioproducts and Biorefining*. 2016.
2. Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, et al. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *science*. 2007; 315: 804-7.
3. Zhao X, Zhang L, Liu D. Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose. *Biofuels, Bioproducts and Biorefining*. 2012; 6: 465-82.
4. Lynd LR. Overview and evaluation of fuel ethanol from cellulosic biomass: technology, economics, the environment, and policy. *Annual review of energy and the environment*. 1996; 21: 403-65.
5. Chandra RP, Bura R, Mabee W, Berlin dA, Pan X, Saddler J. Substrate pretreatment: The key to effective enzymatic hydrolysis of lignocellulosics? *Biofuels*: Springer; 2007. p. 67-93.
6. Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VG. Novel enzymes for the degradation of cellulose. *Biotechnology for biofuels*. 2012; 5: 1.
7. Scheller HV, Ulvskov P. Hemicelluloses. *Plant Biology*. 2010; 61: 263.
8. Harholt J, Suttangkakul A, Scheller HV. Biosynthesis of pectin. *Plant physiology*. 2010; 153: 384-95.
9. Keegstra K. Plant cell walls. *Plant Physiology*. 2010; 154: 483-6.
10. Somerville C. Cellulose synthesis in higher plants. *Annu Rev Cell Dev Biol*. 2006; 22: 53-78.
11. Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink V. Novel enzymes for the degradation of cellulose. *Biotechnology for biofuels*. 2012; 5: 1.
12. Gilbert HJ, Stålbrand H, Brumer H. How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. *Current opinion in plant biology*. 2008; 11: 338-48.
13. Agger J, Viksø-Nielsen A, Meyer AS. Enzymatic xylose release from pretreated corn bran arabinoxylan: differential effects of deacetylation and deferuloylation on insoluble and soluble substrate fractions. *Journal of Agricultural and food Chemistry*. 2010; 58: 6141-8.
14. Várnai A, Huikko L, Pere J, Siika-Aho M, Viikari L. Synergistic action of xylanase and mannanase improves the total hydrolysis of softwood. *Bioresource technology*. 2011; 102: 9096-104.
15. Martinez D, Larrondo LF, Putnam N, Gelpke MDS, Huang K, Chapman J, et al. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nature biotechnology*. 2004; 22: 695-700.
16. Wymelenberg AV, Minges P, Sabat G, Martinez D, Aerts A, Salamov A, et al. Computational analysis of the *Phanerochaete chrysosporium* v2. 0 genome database and mass spectrometry identification of peptides in ligninolytic cultures reveal complex mixtures of secreted proteins. *Fungal Genetics and Biology*. 2006; 43: 343-56.
17. Abbas A, Koc H, Liu F, Tien M. Fungal degradation of wood: initial proteomic analysis of extracellular proteins of *Phanerochaete chrysosporium* grown on oak substrate. *Current genetics*. 2005; 47: 49-56.

18. Wymelenberg AV, Sabat G, Martinez D, Rajangam AS, Teeri TT, Gaskell J, et al. The Phanerochaete chrysosporium secretome: database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium. *Journal of biotechnology*. 2005; 118: 17-34.
19. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research*. 2014; 42: D490-D5.
20. Levasseur A, Piumi F, Coutinho PM, Rancurel C, Asther M, Delattre M, et al. FOLy: an integrated database for the classification and functional annotation of fungal oxidoreductases potentially involved in the degradation of lignin and related aromatic compounds. *Fungal genetics and biology*. 2008; 45: 638-45.
21. Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, et al. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic acids research*. 2014; 42: D26-D31.
22. Stajich JE, Harris T, Brunk BP, Brestelli J, Fischer S, Harb OS, et al. FungiDB: an integrated functional genomics database for fungi. *Nucleic acids research*. 2011: gkr918.
23. Wymelenberg AV, Gaskell J, Mozuch M, Sabat G, Ralph J, Skyba O, et al. Comparative transcriptome and secretome analysis of wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium*. *Applied and environmental microbiology*. 2010; 76: 3599-610.
24. Wymelenberg AV, Gaskell J, Mozuch M, Kersten P, Sabat G, Martinez D, et al. Transcriptome and secretome analyses of *Phanerochaete chrysosporium* reveal complex patterns of gene expression. *Applied and environmental microbiology*. 2009; 75: 4058-68.
25. Thuillier A, Chibani K, Belli G, Herrero E, Dumarçay S, Gérardin P, et al. Transcriptomic Responses of *Phanerochaete chrysosporium* to Oak Acetonic Extracts: Focus on a New Glutathione Transferase. *Applied and environmental microbiology*. 2014; 80: 6316-27.
26. Wymelenberg AV, Gaskell J, Mozuch M, BonDurant SS, Sabat G, Ralph J, et al. Significant alteration of gene expression in wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium* by plant species. *Applied and environmental microbiology*. 2011; 77: 4499-507.
27. Gaskell J, Marty A, Mozuch M, Kersten PJ, BonDurant SS, Sabat G, et al. Influence of *Populus* genotype on gene expression by the wood decay fungus *Phanerochaete chrysosporium*. *Applied and environmental microbiology*. 2014; 80: 5828-35.
28. Skyba O, Cullen D, Douglas CJ, Mansfield SD. Gene expression patterns of wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium* are influenced by wood substrate composition during degradation. *Applied and Environmental Microbiology*. 2016: AEM. 00134-16.
29. Korripally P, Hunt CG, Houtman CJ, Jones DC, Kitin PJ, Cullen D, et al. Regulation of Gene Expression during the Onset of Ligninolytic Oxidation by *Phanerochaete chrysosporium* on Spruce Wood. *Applied and environmental microbiology*. 2015; 81: 7802-12.
30. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otilar R, et al. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research*. 2013: gkt1183.
31. Grigoriev IV, Cullen D, Goodwin SB, Hibbett D, Jeffries TW, Kubicek CP, et al. Fueling the future with fungal genomics. *Mycology*. 2011; 2: 192-209.
32. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics*. 2004; 20: 1453-4.
33. Saldanha AJ. Java Treeview—extensible visualization of microarray data. *Bioinformatics*. 2004; 20: 3246-8.
34. Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams. 2007.
35. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. 2015: gkv007.
36. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26: 139-40.
37. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic acids research*. 2012: gks042.
38. Robinson MD, Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*. 2008; 9: 321-32.
39. Zhou X, Lindsay H, Robinson MD. Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic acids research*. 2014; 42: e91-e.

40. Dashtban M, Schraft H, Qin W. Fungal bioconversion of lignocellulosic residues; opportunities & perspectives. *Int J Biol Sci.* 2009; 5: 578-95.
41. Coutinho PM, Stam M, Blanc E, Henrissat B. Why are there so many carbohydrate-active enzyme-related genes in plants? *Trends in plant science.* 2003; 8: 563-5.
42. Vuong TV, Wilson DB. Glycoside hydrolases: catalytic base/nucleophile diversity. *Biotechnology and bioengineering.* 2010; 107: 195-205.
43. dos Santos Castro L, Antoniêto ACC, Pedersoli WR, Silva-Rocha R, Persinoti GF, Silva RN. Expression pattern of cellulolytic and xylanolytic genes regulated by transcriptional factors XYR1 and CRE1 are affected by carbon source in *Trichoderma reesei*. *Gene Expression Patterns.* 2014; 14: 88-95.
44. Kubicek CP, Messner R, Gruber F, Mandels M, Kubicek-Pranz EM. Triggering of cellulase biosynthesis by cellulose in *Trichoderma reesei*. Involvement of a constitutive, sophorose-inducible, glucose-inhibited beta-diglucoside permease. *Journal of Biological Chemistry.* 1993; 268: 19364-8.
45. Ilmen M, Saloheimo A, Onnela M-L, Penttilä ME. Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*. *Applied and Environmental Microbiology.* 1997; 63: 1298-306.
46. Zeilinger S, Schmoll M, Pail M, Mach RL, Kubicek CP. Nucleosome transactions on the *Hypocrea jecorina* (*Trichoderma reesei*) cellulase promoter *cbh2* associated with cellulase induction. *Molecular Genetics and Genomics.* 2003; 270: 46-55.
47. Ilmén M, Onnela M-L, Klemsdal S, Keränen S, Penttilä M. Functional analysis of the cellobiohydrolase I promoter of the filamentous fungus *Trichoderma reesei*. *Molecular and General Genetics MGG.* 1996; 253: 303-14.
48. Biely P. Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology advances.* 2012; 30: 1575-88.
49. Bowman SM, Free SJ. The structure and synthesis of the fungal cell wall. *Bioessays.* 2006; 28: 799-808.
50. Lairson L, Henrissat B, Davies G, Withers S. Glycosyltransferases: structures, functions, and mechanisms. *Biochemistry.* 2008; 77: 521.
51. Linhardt R, Galliher P, Cooney C. Polysaccharide lyases. *Applied biochemistry and biotechnology.* 1987; 12: 135-76.
52. Pao SS, Paulsen IT, Saier MH. Major facilitator superfamily. *Microbiology and molecular biology reviews.* 1998; 62: 1-34.
53. Walmsley AR, Barrett MP, Bringaud F, Gould GW. Sugar transporters from bacteria, parasites and mammals: structure-activity relationships. *Trends in biochemical sciences.* 1998; 23: 476-81.
54. Suzuki H, MacDonald J, Syed K, Salamov A, Hori C, Aerts A, et al. Comparative genomics of the white-rot fungi, *Phanerochaete carnosus* and *P. chrysosporium*, to elucidate the genetic basis of the distinct wood types they colonize. *BMC genomics.* 2012; 13: 444.
55. Shoseyov O, Shani Z, Levy I. Carbohydrate binding modules: biochemical properties and novel applications. *Microbiology and Molecular Biology Reviews.* 2006; 70: 283-95.
56. Gilkes N, Warren R, Miller R, Kilburn DG. Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. *Journal of Biological Chemistry.* 1988; 263: 10401-7.
57. Tomme P, Tilbeurgh H, Pettersson G, Damme J, Vandekerckhove J, Knowles J, et al. Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. *European Journal of Biochemistry.* 1988; 170: 575-81.
58. Van Tilbeurgh H, Tomme P, Claeysens M, Bhikhabhai R, Pettersson G. Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*. *FEBS letters.* 1986; 204: 223-7.
59. Wymelenberg AV, Sabat G, Mozuch M, Kersten PJ, Cullen D, Blanchette RA. Structure, organization, and transcriptional regulation of a family of copper radical oxidase genes in the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. *Applied and environmental microbiology.* 2006; 72: 4871-7.
60. Cohen-Kupiec R, Broglie KE, Friesem D, Broglie RM, Chet I. Molecular characterization of a novel β -1, 3-exoglucanase related to mycoparasitism of *Trichoderma harzianum*. *Gene.* 1999; 226: 147-54.

61. Lodder AL, Lee TK, Ballester R. Characterization of the Wsc1 protein, a putative receptor in the stress response of *Saccharomyces cerevisiae*. *Genetics*. 1999; 152: 1487-99.
62. Yoshida M, Igarashi K, Kawai R, Aida K, Samejima M. Differential transcription of β -glucosidase and cellobiose dehydrogenase genes in cellulose degradation by the basidiomycete *Phanerochaete chrysosporium*. *FEMS microbiology letters*. 2004; 235: 177-82.
63. Eriksson KE. Enzyme mechanisms involved in cellulose hydrolysis by the rot fungus *Sporotrichum pulverulentum*. *Biotechnology and Bioengineering*. 1978; 20: 317-32.
64. Quinlan RJ, Sweeney MD, Leggio LL, Otten H, Poulsen J-CN, Johansen KS, et al. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proceedings of the National Academy of Sciences*. 2011; 108: 15079-84.
65. Westereng B, Ishida T, Vaaje-Kolstad G, Wu M, Eijsink VG, Igarashi K, et al. The putative endoglucanase PcGH61D from *Phanerochaete chrysosporium* is a metal-dependent oxidative enzyme that cleaves cellulose. *PloS one*. 2011; 6: e27807.
66. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for biofuels*. 2013; 6: 1.
67. Bey M, Zhou S, Poidevin L, Henrissat B, Coutinho PM, Berrin J-G, et al. Cello-oligosaccharide oxidation reveals differences between two lytic polysaccharide monooxygenases (family GH61) from *Podospora anserina*. *Applied and environmental microbiology*. 2013; 79: 488-96.
68. Žifčáková L, Baldrian P. Fungal polysaccharide monooxygenases: new players in the decomposition of cellulose. *Fungal ecology*. 2012; 5: 481-9.
69. Bennati-Granier C, Garajova S, Champion C, Grisel S, Haon M, Zhou S, et al. Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by *Podospora anserina*. *Biotechnology for biofuels*. 2015; 8: 1.
70. Bisaria VS, Mishra S, Eveleigh DE. Regulatory aspects of cellulase biosynthesis and secretion. *Critical reviews in biotechnology*. 1989; 9: 61-103.
71. Kubicek C. From cellulose to cellulase inducers: facts and fiction. *Proceedings of the second TRICEL symposium on Trichoderma reesei cellulases and other hydrolytic enzymes: Foundation of Biotechnical and Industrial Fermentation Research, Espoo*; 1993. p. 181-8.
72. Zeilinger S, Mach R. Xylanolytic enzymes of *Trichoderma reesei*: properties and regulation of expression. *Curr Topics Cer Chem*. 1998; 1: 27-35.
73. Mach R, Zeilinger S. Regulation of gene expression in industrial fungi: *Trichoderma*. *Applied microbiology and biotechnology*. 2003; 60: 515-22.
74. Hrmová M, Petráková E, Biely P. Induction of cellulose- and xylan-degrading enzyme systems in *Aspergillus terreus* by homo- and heterodisaccharides composed of glucose and xylose. *Microbiology*. 1991; 137: 541-7.
75. Mandels M, Parrish FW, Reese ET. Sophorose as an inducer of cellulase in *Trichoderma viride*. *Journal of Bacteriology*. 1962; 83: 400-8.
76. Aro N, Pakula T, Penttilä M. Transcriptional regulation of plant cell wall degradation by filamentous fungi. *FEMS microbiology reviews*. 2005; 29: 719-39.
77. Karaffa L, Fekete E, Gamauf C, Szentirmai A, Kubicek CP, Seiboth B. D-Galactose induces cellulase gene expression in *Hypocrea jecorina* at low growth rates. *Microbiology*. 2006; 152: 1507-14.
78. Kubicek C, Penttilä M. Regulation of production of plant polysaccharide degrading enzymes by *Trichoderma*. *Trichoderma and Gliocladium*. 1998; 2: 49-71.
79. Morikawa Y, Ohashi T, Mantani O, Okada H. Cellulase induction by lactose in *Trichoderma reesei* PC-3-7. *Applied microbiology and biotechnology*. 1995; 44: 106-11.
80. Stricker AR, Grosstessner-Hain K, Würleitner E, Mach RL. Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*. *Eukaryotic cell*. 2006; 5: 2128-37.
81. Stricker AR, Steiger MG, Mach RL. Xyr1 receives the lactose induction signal and regulates lactose metabolism in *Hypocrea jecorina*. *FEBS letters*. 2007; 581: 3915-20.
82. Amore A, Giacobbe S, Faraco V. Regulation of cellulase and hemicellulase gene expression in fungi. *Current genomics*. 2013; 14: 230-49.
83. El-Gogary S, Leite A, Crivellaro O, Eveleigh D, El-Dorry H. Mechanism by which cellulose triggers cellobiohydrolase I gene expression in *Trichoderma reesei*. *Proceedings of the National Academy of Sciences*. 1989; 86: 6138-41.

84. Carle-Urioste JC, Escobar-Vera J, El-Gogary S, Henrique-Silva F, Torigoi E, Crivellaro O, et al. Cellulase induction in *Trichoderma reesei* by cellulose requires its own basal expression. *Journal of Biological Chemistry*. 1997; 272: 10169-74.
85. Pérez J, Muñoz-Dorado J, de la Rubia T, Martínez J. Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. *International Microbiology*. 2002; 5: 53-63.
86. Jeffries TW. Biodegradation of lignin and hemicelluloses. *Biochemistry of microbial degradation*: Springer; 1994. p. 233-77.
87. Kirk TK, Cullen D. *Enzymology and molecular genetics of wood degradation by white-rot fungi. Environmentally friendly technologies for the pulp and paper industry* Wiley, New York. 1998: 273-307.

Chapter-3

Gene Expression Metadata Analysis Reveals Molecular Mechanisms Employed by *Phanerochaete chrysosporium* During Lignin Degradation and Detoxification of Plant Extractives

[This work has been published in “**Current Genetics**” (2017): 1-18]

Ayyappa Kumar Sista Kameshwar and Wensheng Qin*

3.1. Abstract

Lignin, most complex and abundant biopolymer on the earth's surface, attains its stability from intricate polyphenolic units and non-phenolic bonds, making it difficult to depolymerize or separate from other units of biomass. Eccentric lignin degrading ability, availability of annotated genome makes *Phanerochaete chrysosporium* ideal for studying lignin degrading mechanisms. Decoding and understanding the molecular mechanisms underlying the process of lignin degradation will significantly aid the progressing biofuel industries and lead to the production of commercially vital platform chemicals. In this study, we have performed a large-scale metadata analysis to understand the common gene expression patterns of *P. chrysosporium* during lignin degradation. Gene expression datasets were retrieved from NCBI GEO database and analyzed using GEO2R and Bioconductor packages. Commonly expressed statistically significant genes among different datasets were further considered to understand their involvement in lignin degradation and detoxification mechanisms. We have observed three sets of enzymes commonly expressed during ligninolytic conditions which were later classified into primary ligninolytic, aromatic compound degrading and other necessary enzymes. Similarly, we have observed three sets of genes coding for detoxification and stress responsive, phase-I and phase-II metabolic enzymes. Results obtained in this study indicate the coordinated action of enzymes involved in lignin depolymerization and detoxification-stress responses under ligninolytic conditions. We have developed tentative network of genes and enzymes involved in lignin degradation and detoxification mechanisms by *P. chrysosporium* based on the literature and results obtained in this study. However, ambiguity raised due to higher expression of several uncharacterized proteins necessitates for further proteomic studies in *P. chrysosporium*.

Keywords: *Phanerochaete chrysosporium*; Transcriptome; Lignocellulose; Gene Expression Omnibus (GEO); GEO2R; Detoxification responses; Phase-I and Phase-II metabolic enzymes

3.2. Introduction

Naturally, wood is composed of two organic compound groups a) carbohydrates (65-75%) b) lignin (20-30%) along with organic extraneous compounds (4-10%) and inorganic minerals (calcium, potassium etc.) [1]. Lignin constitutes second most abundant biopolymer, found in closer associations with cellulose and hemicellulosic units, complex polyphenolic units and non-phenolic linkages provides lignin with high stability. Lignin is considered as major bottleneck in biofuel industries, as it is necessary to separate lignin from other units of biomass for the efficient production of biofuels. Also, if degraded efficiently lignin can be used to produce commercially important platform chemicals. Only few microorganisms were reported till today with efficient lignin degrading abilities. *Phanerochaete chrysosporium* is a wood degrading white rot fungus belonging to basidiomycetes fungi, well known for its eccentric lignin degrading ability [2, 3]. Apart from lignin, *P. chrysosporium* also exhibits great ability in degrading and mineralizing various synthetic dyes, organic pollutants such as 2,4-dichlorophenol, 2,4-dinitrotoluene, endosulfan, pentachlorophenol, phenanthrene and several other harmful organic chemicals [4-7]. Degrading ability of *P. chrysosporium* is credited to its highly efficient enzyme system, it secretes a wide range of oxidative and hydrolytic enzymes for the successful degradation of various organic compounds. Several studies have reported about the lignin degrading enzyme system of *P. chrysosporium*, which majorly consists of ligninolytic peroxidases such as lignin peroxidase (LiP), manganese peroxidase (MnP) and hydrogen peroxide generating enzymes. *P. chrysosporium* induces its enzyme arsenal under nutrient deficient culture conditions during its secondary metabolism [8]. Studies have reported that *P. chrysosporium* enzyme system can catalyze the primary oxidation of various persistent xenobiotic compounds such as chloroaromatic, polyaromatic compounds and dioxins [9].

Martinez et al (2004) has sequenced and annotated the whole genome of *P. chrysosporium*, in fact it was first whole genome sequence from phylum Basidiomycota [10]. The 30 Mb haploid genome of *P. chrysosporium* with 11, 777 protein coding genes, has revealed several significant facts about the genes and enzymes involved in wood decaying process [10]. Genome of *P. chrysosporium* majorly codes for cytochrome P450 superfamily, glucose methanol choline oxidoreductases, protein kinases, alcohol oxidases, short chain reductases, aspartyl proteases, von Willebrand factor, lectin type proteins and several other ligninolytic enzymes. Present day annotated genome of *P. chrysosporium* RP-78 v2.2 encodes

about 444 CAZymes (glycoside hydrolase (181), glycosyl transferases (70), auxiliary activity (89), carbohydrate binding domains (65), carbohydrate esterases (20) and polysaccharide lyases (6)) out of which auxiliary activity, carbohydrate esterases and glycosyl transferases were found during lignin degradation. [10, 11]. Genome level transcriptome studies conducted during last decade has revealed several significant facts about the differential expression patterns of several genes involved in lignin depolymerization.

Minami et al (2007) have performed a Long SAGE analysis on *P. chrysosporium* to understand the changes in transcriptome during the initiation of manganese peroxidase and lignin peroxidase enzymes. Long SAGE analysis have revealed about several candidate gene sequences involved in regulation of the LiP and MnP enzyme production [12] (GSE6649). Wymelenberg et al (2009) have studied the extracellular protein products of *P. chrysosporium* grown under nutrient limited and replete conditions. Results obtained by them has confirmed the expression of several lignocellulose degrading enzymes under nutrient limited conditions and also reinforced the role of novel proteins [13](GSE14735). To study the extracellular proteins secreted by *P. chrysosporium* under standard cellulolytic conditions Wymelenberg et al (2009) have performed a whole transcriptome study. This gene expression study has confirmed the significance of carbohydrate active enzymes and supported the function of many novel proteins involved in lignocellulose degradation [13, 14] (GSE14734). Wymelenberg et al (2011) have conducted transcriptome study to analyze the gene expression patterns of *Postia placenta* and *Phanerochaete chrysosporium* colonized on (aspen) *Populus grandidentata* and (pine) *Pinus strobus*. Results have showed that transcriptome of these fungi are significantly influenced by wood species and this study has majorly differentiated molecular mechanisms employed by white and brown rot decay patterns [15] (GSE27941). Thuillier, Anne, et al (2014) have performed a transcriptome study of *Phanerochaete chrysosporium* cultured on oak acetic extracts and reported that *P. chrysosporium* employs both intracellular antioxidative detoxification mechanisms along with extracellular enzymes for lignin degradation. This study has also revealed the functional characteristics of PcGTT2.1 a glutathione-s-transferase isoform, involved in reducing the cellular toxicity caused by lipid peroxidation and also reported the loss of GTT2.1 isoform in some of the non-wood decaying fungi [16] (GSE54542). Gaskell, J et al (2014) have conducted experiments to understand the gene expression patterns of *P. chrysosporium*, colonized on hybrid poplar (*Populus alba x tremula*), 82 and 64 transgenic derivatives (syringyl rich). The microarray results have showed that gene expression patterns of *P. chrysosporium* are

considerably influenced by lignin composition of the growth substrate, especially peptides corresponding to various oxidoreductases were found to be highly expressed in 82 and 64 transgenic line substrates [17] (GSE52922). Korripally et al (2015) have performed a whole transcriptome study of *P. chrysosporium* by culturing it on spruce wood samples with an efficient oxidant sensing beads at 40 and 96-hours incubation periods. This study has also revealed the functional properties of 72 unknown proteins available under the *P. chrysosporium* genome database v2.2 , cytochrome P450 monooxygenases and transporters [18] (GSE69461). Skyba et al (2016), have performed a whole transcriptome study of *P. chrysosporium* and *P. placenta*, cultured on three *Populus trichocarpa* (poplar) wood substrates with different chemical compositions [19]. This study has clearly showed the influence of growth substrate (wood composition) and incubation period on the gene expressions of *P. chrysosporium* and *P. placenta* [19] (GSE69008). All the above genome wide studies have significantly influenced the present day understanding about the plant cell wall degrading abilities of *P. chrysosporium* and revealed functional properties of several uncharacterized proteins.

During the process of wood degradation *P. chrysosporium* is exposed to various highly toxic phenolic plant extractives such as flavonoids, quinones, stilbenes, tannins [16, 20, 21]. Among these plant extractives, flavonoids and stilbenes possess strong antifungal properties and are also required for the durability of wood [20, 22, 23]. However, the ligninolytic white rot fungi have developed an efficient enzyme systems involved in antioxidant and detoxification mechanisms [16]. Majorly the detoxification system of white rot fungi can be classified into phase I (cytochrome P450 group, epoxide hydrolases) and phase II (Glutathione-S-transferase, quinone oxidoreductase, UDP-glucuronosyltransferases) enzyme systems. These enzyme systems are specifically induced by a variety of xenobiotics and plant extractive compounds [24]. Thus, most of the white rot fungi maintain a complementary system of extracellular enzymes involved in wood degradation and a simultaneous intracellular antioxidant and detoxification systems.

Studies have reported that *P. chrysosporium* is equipped with large number of genes coding for cytochrome P450 monooxygenases, phase I and phase II metabolic enzymes and signaling cascade genes [24]. In the last decade, extensive research has been performed on the cytochrome P450 monooxygenases and its role in several physiological and catalytic processes of *P. chrysosporium* such as ligninolysis, secondary metabolism and xenobiotic

degradation processes. Doddapaneni and Yadav (2005) has performed global gene expression studies of *P. chrysosporium* to explain the differential expression patterns of cytochrome P450 monooxygenases using a custom designed 70-mer oligonucleotide microarray. For the first time this study has proved the involvement of cAMP and MAP kinase signaling pathways during the biodegradation and secondary metabolism of *P. chrysosporium* [24]. Later, studies have also reported that cAMP and calmodulin (CaM) signaling mechanisms play crucial role in expression of ligninolytic peroxidases, as expression of calmodulin inhibitor W-7 in *P. chrysosporium* has resulted in regulation of *lip* and *mnp* genes and their isoforms [25, 26]. Subramanian and Yadav (2009) have proved the significance of P450 monooxygenases in degradation of nonylphenol (endocrine disrupting chemical) under different nutrient conditions using a custom designed microarray. This study has shown the involvement of P450 monooxygenases in nonylphenol degradation [27]. In the year 2009 Jiang et al, has performed genome wide expression analysis specifically for identifying the genes involved in secondary metabolism of *P. chrysosporium*. This study has revealed the expression of genes coding for enzymes such as aryl alcohol dehydrogenase, cytochrome P450, alkyl hydroperoxide reductase, catalase, ABC transporters etc. [28]. Subramanian and Yadav (2009) have reported the transcriptome profiles of cytochrome P450 in *P. chrysosporium* under varied nutrient conditions [29]. Expression of P450 enzymes under different nutrient conditions suggests the role of P450 enzymes in the catalytic activity of *P. chrysosporium*. Differential expression of certain P450 enzymes during low and high nutrient conditions reveals the specific role played by these enzymes under ligninolytic and non-ligninolytic conditions [29]. Chigu et al (2010) have performed transcriptomic profiling of *P. chrysosporium* cytochrome P450 monooxygenases (PcCYPS) involved in anthracene metabolism [30]. This study has revealed that 14 PcCYP genes are involved in step by step conversion of anthracene to anthraquinone. 12 PcCYPS are upregulated upon exogenous addition of anthracene. Out of 12 PcCYPS, five genes showed high catalytic activity against anthracene and also reported that these genes play major role in *in vivo* anthracene metabolism [30].

In this study, we have conducted a large-scale metadata analysis on *P. chrysosporium* gene expression datasets, specifically to demonstrate the common gene expression patterns involved during extracellular lignin degradation and intracellular detoxification mechanisms. To the extent of our knowledge this is the first report on metadata analysis of *P. chrysosporium* for demonstrating lignin degradation and detoxification, stress responsive mechanisms.

3.3. Data Retrieval and Methodology

3.3.1. Data retrieval: We have used the term *Phanerochaete chrysosporium* to search for the gene expression datasets available in NCBI Gene expression omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>) a public repository for gene expression datasets. All the gene expression dataset corresponding to *P. chrysosporium* listed under “GEO Datasets” window was carefully analyzed by accessing experimental information provided by corresponding research team under “Accession display” window. Totally, there are 10 gene expression datasets which are currently available in GEO database out of which we have retrieved total of 8 *P. chrysosporium* gene expression datasets (six were microarray datasets, one RNA sequencing and one Long-SAGE dataset). The NCBI GEO accession IDs of gene expression datasets retrieved were GSE14734, GSE14735, GSE54542, GSE27941, GSE52922, GSE69008, GSE69461 and GSE6649. Substrate and platform level details of these gene expression datasets were shown in the Table 3.1. We have specifically considered *P. chrysosporium* gene expression datasets which were based on the natural plant biomass and simple synthetic compounds containing growth mediums to monitor the change in gene expression patterns under ligninolytic conditions.

Table 3.1: Details of the *P. chrysosporium* transcriptome metadata retrieved from NCBI GEO and NCBI SRA.

GEO- ID	Platform and Technology	Substrate	#Samples	Ref
GSE54542	NimbleGen <i>Phanerochaete chrysosporium</i> arrays	Oak acetic extractives	6	[16]
GSE27941	NimbleGen <i>Phanerochaete chrysosporium</i> arrays	Ball milled aspen, Ball milled pine	6	[15]
GSE52922	NimbleGen <i>Phanerochaete chrysosporium</i> arrays	P717 hybrid line, Transgenic line 82 Transgenic line 64	9	[17]
GSE14734	NimbleGen <i>Phanerochaete chrysosporium</i> arrays	Cellulose, Glucose, Ball milled aspen	9	[13, 14]
GSE14735	NimbleGen <i>Phanerochaete chrysosporium</i> arrays	Replete medium Carbon limited Nitrogen limited	9	[13, 14]
GSE69008	NimbleGen <i>Phanerochaete chrysosporium</i> arrays	Poplar wood substrates	24	[19]
GSE6649	Long Serial analysis of gene expression	Basal III medium (1% (v/v), 1% (w/v) glucose, 20 mM Na-2,2dimethylsuccinate	2	[12]

GSE69461	Illumina HiSeq 2000	(pH 4.5), 0.0001% thiamine, and 1.2 mM ammonium tartrate) 3 mM veratryl alcohol <i>Picea glauca</i> (spruce sapwood)	18	[18]
----------	---------------------	---	----	------

3.3.2. Data Analysis: The microarray datasets retrieved were analyzed using GEO2R (an R based interactive online tool) (<https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html>) and Bioconductor packages GEOquery and limma based on R software version 3.2.2. The settings used in GEO2R for analyzing the microarray datasets were listed below autodetect option (for log transformation of the data), box-whisker plot (samples and value distributions) and submitter provided annotations (for gene level annotations). The experimental design and sample grouping information was obtained from the gene expression datasets and the corresponding literature. Top 250 function was used to obtain the differentially expressed genes statistically significant genes, the Top 250 function internally uses limma (linear models for microarray data) for the statistical analysis and the genes are ranked based on their P-values. The differentially expressed genes were obtained after performing Benjamini and Hochberg false discovery rate multiple testing correction method with a p-value 0.05. As we have mentioned earlier, the process of biological contextualization was based on the supplier provided annotations and supplementary information which included mainly InterPro Hits and Protein IDs. The gene and protein level annotations of *P. chrysosporium* were also obtained from MycoCosm (fungal genome repository) [31, 32]. We have also used other analysis options available in JGI-MycoCosm such as Gene Ontology (GO)[33, 34], EuKaryotic Orthologous Groups (KOG) [35] and CAZy [36, 37] for analyzing the results obtained. The differentially expressed genes were also represented as hierarchical clusters using the Cluster 3.0 software [38] based on cluster for both genes and arrays using the complete linkage options, thus obtained cluster output files were used for visualization and development of dendrograms using Java Treeview software using the standard conditions [39]. We have used Venny 2.1 [40] and Jvenn [41] softwares to get the common differentially expressed gene lists among different datasets. The GSE69461 dataset was analyzed by retrieving the sample level RPKM (Reads Per Kilobase Million) values from supplementary data file provided of GSE69461[18]. Further analysis was performed using RPKM values using limma [42], Glimma (<http://bioconductor.org/packages/release/bioc/html/Glimma.html>) and edgeR [43-46] Bioconductor packages, to get statistically differentially expressed genes

with a fold change value 2.0 and p-value 0.05 with Benjamini and Hochberg FDR multiple correction method. From the obtained results genes encoding for ligninolytic and detoxification, stress responsive pathways were specifically retrieved based on their InterPro annotations. Long SAGE samples were analyzed using the Identification of Differentially Expressed Genes 6 (IDEG6) orphan tags were removed (sequential errors) further statistical analysis was performed using Audic-Claverie test [47]. We have reported the process of data analysis earlier in our previous work which reported the metabolic and molecular gene networks employed by *P. chrysosporium* during cellulose and hemicellulose degradation [48].

3.3.3. Summary of Data Analysis

Based on the growth substrates used for culturing of *P. chrysosporium*, the gene expression datasets considered for the present study were divided as customized synthetic growth medium (containing cellulose, glucose or other commercially available nutrients supplemented with Highley's basal medium) and complex natural plant biomass medium (containing ball milled aspen, ball milled pine, spruce wood and poplar wood substrates) reported earlier [48]. The accession IDs of gene expression datasets belonging to customized synthetic growth mediums were GSE14734 (HBM supplemented with 0.5 % (wt/vol) of BMA or cellulose or glucose as sole carbon source), GSE14735 (replete B3 medium, carbon limited medium and nitrogen limited medium) [13, 14] and GSE6649 [12]. The GEO accession IDs of complex natural plant biomass medium were GSE27941(0.5% of ball milled aspen and ball milled pine as the sole carbon source supplemented with HBM) [15], GSE52922 (parental hybrid clone line- *Populus trichocarpa* P717 with [65 mol% of syringyl], transgenic line 64 [94% syringyl] and transgenic line 82 with [85 mol % of syringyl]) [17], GSE54542 (fine powdered oak heartwood samples extracted using acetone and further resuspended in DMSO followed by a set of extraction processes) [16], GSE69461 (microtomed tangential sections of *Picea glauca* coated with 90 μ l of agar supplemented with nitrogen mineral salt medium) [18]. Similarly GSE69008 contained chemically distinct *Populus trichocarpa* wood substrates: high lignin-low glucose (A), low lignin-high glucose (B) and average lignin-average glucose (C) [19]. The detailed explanation about the experimental conditions and growth protocols performed for the gene expression study protocol can be followed from the corresponding literature cited.

3.4. Results

3.4.1. Lignin Oxidizing and Auxiliary Enzymes: Degradation of lignin and its derivatives are subject of interest since several years. Advancement of high throughput genomic and proteomic methods in the recent years have revealed various significant facts about lignin degradation mechanisms employed by *P. chrysosporium*. According to Kirk, T.K et al (1987), an array of oxidases and peroxidases are secreted by white rot fungi for the initial degradation of lignin, these reactions release highly reactive and non-specific free radicals which leads to a complex series of spontaneous cleavage reactions [10, 49, 50]. Most of the Basidiomycetes fungi and especially white rot fungi secrete extracellular laccases, which are involved in the single electron oxidation of phenols, phenoxy radicals, aromatic amines and electron rich compounds, ultimately transferring four electrons to O₂ and reducing it to H₂O molecule [49, 51]. The whole genome sequencing studies have showed that *P. chrysosporium* doesn't code for any conventional laccase however, it codes for a cluster of four multi copper oxidases (MCO) and ferroxidase enzymes. Thus, multi copper oxidases and ferroxidases secreted by *P. chrysosporium* are involved in extracellular oxidation of lignin along with other lignin oxidizing enzymes [10]. *P. chrysosporium* genome consists of 10 *lip* genes coding for lignin peroxidases, 5 *mnp* genes coding for manganese peroxidase and 1 hybrid peroxidase (pc.91.32.1) encoding gene sequences [10]. From the metadata analysis, we have observed that expression of peroxidase encoding genes in *P. chrysosporium* varies differentially based on the source of nutrients and time of infection. Gene expression studies mainly GSE14735, GSE69008, GSE69461 and GSE52922 have provided a significant evidence on differential expression of the ligninolytic peroxidases. Earlier studies have reported the differential expression of lignin peroxidase under nutrient limited conditions, but the mechanism behind its expression is not clear. Current metadata analysis supports previous findings on the expression of lignin degrading peroxidases in *P. chrysosporium* both in synthetic and natural supplemented medium.

Fungal lignin peroxidase coding genes were highly expressed when *P. chrysosporium* was cultured on ball milled aspen, nitrogen and carbon limited medium, hybrid line P717, transgenic line 82, transgenic line 64, control, high lignin–low glucose, day 3 cultures of GSE6649 and spruce wood at 96-hours of incubation. The genes encoding manganese peroxidases were only found to be highly expressed in spruce wood 96-hour incubation sample. Animal haem peroxidases were expressed in cellulose, carbon and nitrogen limited,

ball milled aspen, transgenic line 82, hybrid line P717, and low lignin-high glucose, average lignin-average glucose mediums. Chloroperoxidases were highly expressed in ball milled aspen, cellulose growth medium, transgenic line 64. GMC oxidoreductases coding genes were highly expressed in ball milled aspen, cellulose medium, transgenic line 64, high lignin-low glucose and 40-hour incubation period. Glyoxal coding genes were highly expressed in cellulose, glucose, carbon limited nutrient mediums. Multicopper oxidases encoding genes were differentially expressed in spruce wood 96 hours incubated, high lignin- low glucose conditions. Genes encoding amine oxidase or flavin amine oxidase, copper amine oxidase were differentially expressed in high lignin-low glucose, average lignin-average glucose, nitrogen limited conditions, 40-hour incubation period growth samples. Genes encoding for copper radical oxidase and aryl alcohol dehydrogenase were found to be highly expressed in 40-hour incubation period samples.

Delta-9 acyl-CoA desaturase which is involved in unsaturated fatty acid biosynthesis was found to be highly expressed in nitrogen limited, glucose, cellulose, ball milled pine, 64 and 82 transgenic lines, oak acetic extractives and low lignin-low glucose, average glucose- average lignin conditions. Major intrinsic protein genes were expressed in nitrogen and carbon limited, ball milled aspen, spruce wood 96-hour incubated samples. ABC transporter genes were expressed in glucose, transgenic line 82, parent line P717, low lignin-high glucose, average lignin-average glucose, control, spruce wood 40-hour incubated conditions. Major facilitator superfamily coding genes were highly expressed in ball milled aspen, carbon, nitrogen limited medium, high lignin- low glucose, transgenic line 64, 96-hour incubation and control growth conditions. Tetra/Oligopeptide transporters were differentially expressed in nitrogen limited medium, ball milled aspen, low lignin-high glucose, average lignin- average glucose growth conditions. Fumaryl acetoacetase encoding genes were expressed in high lignin-low glucose, fumarate reductase gene was expressed in spruce wood 40-hour incubation samples. Short chain dehydrogenase /reductases encoding genes were expressed in carbon and nitrogen limited, spruce wood 96-hours samples, control samples (GSE54542), transgenic line 64, low lignin-high glucose, average lignin-average glucose samples. Aromatic ring hydroxylase encoding genes are differentially expressed in cellulose, spruce wood 40 hours of incubation, average lignin-average glucose, low lignin- high glucose conditions, transgenic line 64 growth conditions. Dienelactone hydrolase coding genes are highly expressed in glucose, ball milled aspen, average lignin-average glucose conditions. Due to their little reactivity, aromatic compounds derived from

lignin degradation are usually attacked with the help of oxygen by oxygenases which results in intermediates like catechol or protocatechuate [52, 53]. Mainly the process of peripheral degradation is commenced through central ring cleavage which is oxidatively catalyzed by ring cleavage dioxygenases [54, 55]. Extradiol ring hydroxylase genes are highly expressed in parent line P717, transgenic line 82, intradiol ring cleavage dioxygenase is highly expressed in high lignin-low glucose conditions.

Aldehyde dehydrogenase encoding genes were differentially expressed in glucose, replete, nitrogen limited, ball milled aspen, transgenic line 64, low lignin-high glucose and spruce wood 96 hour incubated conditions. Zinc alcohol dehydrogenase encoding genes were highly expressed in replete, ball milled aspen, transgenic line 64, low lignin-high glucose (20 days) growth conditions. Aldo/keto reductase encoding genes were highly expressed in glucose, carbon limited, transgenic line 64. According to Robson et al, aldo/keto reductases support the process of lignin degradation by providing hydrogen peroxide and ROS through oxidation of NADPH [56]. Beta keto acyl synthase gene is highly expressed in glucose, ball milled aspen, low lignin- high glucose, average lignin- average glucose growth conditions. Acetamidase or formamidase is expressed in nitrogen limited, low lignin- high glucose, average lignin- average glucose conditions. It was reported that genes encoding for enzymes acetamidase, formamidase, uricase and amidohydrolase were found to be highly expressed by *P. chrysosporium* during nitrogen limitations [57]. Acetate kinase gene was expressed in replete, ball milled aspen conditions, earlier studies have reported that metabolism of hydroxycinnamic acids (vanillin, *p-coumaric* and ferrulic acids) in *Streptomyces setonii* resulted in accumulation of acetic acid by causing shift in activities of alcohol dehydrogenase to acetate kinase [58, 59]. Taurine catabolic dioxygenase is highly expressed in ball milled aspen growth condition, 3-hydroxyacyl-CoA dehydrogenase gene is highly expressed in glucose medium. Genes coding for 2-nitropropane dioxygenase were highly expressed in low lignin-high glucose, average lignin average glucose conditions. Carbonic anhydrase genes were highly expressed in transgenic line 64 and average lignin-average glucose conditions.

Haloacid dehalogenase genes were highly expressed in transgenic line 82, average lignin- average glucose and ball milled aspen conditions. Phenylalanine ammonia lyase encoding genes were highly expressed in nitrogen and carbon limited growth mediums, ball milled aspen, transgenic line 64, oak acetonic extracts, average lignin-average glucose growth conditions. Homogentisate 1,2-dioxygenase encoding genes were highly expressed

in control, transgenic line 64 conditions and high gene expression values were observed for ball milled aspen, replete, carbon, nitrogen limited growth mediums. Generic methyltransferases (or) O-methyltransferase encoding genes were highly upregulated in average lignin-average glucose, low lignin-high glucose, spruce wood 96 hour incubated and replete samples. Gene expression studies conducted earlier have proved that lignin and its derivatives induces the higher expression of cytochrome P450 monooxygenase encoding genes which are also involved in further degradation of lignin and its derivatives [24, 29, 60]. Studies have already proved that plant cell wall polymers lignin and hemicellulose occur in deacetylated forms, it was also reported that these polymers are linked with *p*-coumaric acid and ferulic acid which are deacetylated by carbohydrate esterases such as feruloyl esterases respectively [61-63]. While deacetylases belonging to carbohydrate esterase classes are of high significance as the acetylated plant cell wall polymers cease the action of microbial enzymes. Genes encoding carbohydrate esterases (CE) classes CE-4 and CE-9 were found to be highly expressed in natural plant biomass growth substrates 40-hour spruce wood samples and CE-1, CE-4, CE-8, CE-15 and CE-16 in 96-hours spruce wood samples [48] (Table 3.2).

Table 3.2: Lists differentially classified lignin degrading enzymes obtained from different gene expression datasets of *P. chrysosporium*

<i>P. chrysosporium</i> Genes involved in lignin degradation	
Lignin degrading enzymes (First level)	Lignin peroxidase, Manganese peroxidase, Glucose oxidase, Glyoxal oxidase, Benzoquinone reductase, Amine oxidase, Aryl alcohol oxidase, Chloroperoxidase, Copper radical oxidase, Multi copper oxidase, Pyranose-2-oxidase, copper amine oxidase, Phenylalanine ammonia lyase
Aromatic compound degrading enzymes (Second level)	Intradiol dioxygenases, Extradiol dioxygenases, Aromatic ring hydroxylase, Homogentisate 1,2-dioxygenase, Epoxide hydrolase, Cytochrome P450 Monooxygenase, Alcohol dehydrogenase, Dioxygenase, 2-nitropropane dioxygenase, Acireductone dioxygenase, Ferredoxin, Flavin containing monooxygenase, Iron reductases, fumarate reductase, catalase, Alcohol/methanol oxidases, Formate dehydrogenase, Haloacid dehalogenase hydrolase, Oxidoreductase, Prenyl transferase /Squalene oxidase, Acetamidase, Formamidase, Uricase
Other necessary enzymes (Tertiary level)	Esterase/ lipase/thioesterase, GMC oxidoreductase, Metallophosphoesterase, Short-chain dehydrogenase/ reductase, D-isomer specific 2-hydroxyacid dehydrogenase, Beta-ketoacyl synthase, Beta lactamase, 2-oxo acid dehydrogenase, Aldo/keto reductase, Aldehyde dehydrogenase, Alkyl hydroperoxide reductase, Amidohydrolase, delta-1-pyrroline-5-carboxylate dehydrogenase, 2-hydroxyacid dehydrogenase, FAD- linked oxidase, Thiolase, Hydroxymethylglutaryl-CoA synthase, Carbohydrate esterases, Glycosyl transferases

3.4.2. Detoxification and Stress Responsive Genes

During the process of initial infection *P. chrysosporium* and other basidiomycetes fungi produce extracellular reactive oxygen species (ROS) which are involved in breakdown of lignin, while studies have reported that expression of lignin peroxidase is linked with ROS production [16]. However, during the process of wood decay fungi comes across with highly reactive and toxic plant extractives which also exhibit a strong antifungal activity. So, efficient ligninolytic fungi should also exhibit the strong capacity to resist from the anti-fungicidal properties of plant extractives and lignin derivatives [16]. Thus, fungi have developed a significant detoxification system which majorly includes cytochrome P450 complex enzymes, glutathione-s-transferases [16]. Gene expression studies especially GSE52922 and other growth substrates containing natural plant biomass growth mediums (GSE69461, GSE54542, GSE69008, GSE27941) have showed higher expression of various genes encoding for detoxification and stress responsive enzymes. The cellular ROS level significantly influences the redox state controlled by the degree of oxidation/reduction of active redox species, further regulating the cellular metabolism among which pyridine nucleotides and thiol/disulfide compounds play a crucial role as they bridge enzymes of intricate metabolic networks [64]. Genes coding for 2-oxo acid dehydrogenase were highly expressed in cellulose, nitrogen limited, 64-transgenic line, average lignin-average glucose (10-days) growth mediums. 2-oxoglutarate dehydrogenase was highly expressed in 96-hours spruce wood growth medium. Thioredoxin encoding genes were highly expressed in carbon and nitrogen limited, BMA, oak acetic extractives, P717, transgenic line 82, high lignin-low glucose, 40-hours spruce wood growth samples, at the same time peroxiredoxin (ubiquitous group of antioxidant enzymes) encoding genes were expressed in 96-hours growth samples. Genes encoding thaumatin pathogenesis related protein were highly expressed in BMA, low lignin-high glucose growth mediums. The constant rates of NAD(P)H/NAD(P)⁺ and SH/S-S facilitate the redox reaction by directly effecting on the proteins [64]. Alkylhydroperoxide reductase and thiol specific antioxidant are involved in reducing the reduced dithiol form of organic hyperoxides and protect against sulphur containing radicals, genes encoding these enzymes were differentially expressed in glucose, replete, BMA, P717, transgenic line 82 growth samples. The three-major cell damaging units such as hydrogen peroxide, ROS and superoxide dismutase (SOD) are efficiently tackled by the fungal cells, if not leads to strong toxic stress to the cell and completely damages the cellular material, enzymes pertaining to it are catalase, peroxiredoxin, superoxide dismutase. Genes encoding to catalase and

superoxide dismutase enzymes are BMA, replete, high lignin-low glucose growth mediums, and transgenic line 64, control, high glucose-low lignin growth samples respectively. Heat shock proteins (HSP) and Ubiquitin conjugating systems were highly expressed gene systems among the *P. chrysosporium* detoxification and stress responsive mechanisms. Heat shock protein 20 (Hsp20) encoding genes were highly expressed in replete, BMA, high glucose-low lignin, average lignin-average glucose growth conditions. Similarly, several ubiquitin conjugating enzymes and complexes were highly expressed in plant biomass growth conditions such as P717, transgenic line 64, high lignin-low glucose, average lignin-average glucose growth conditions. Abortive infection protein was found to be highly expressed in high lignin-low glucose, control and stress responsive proteins were highly expressed in 40-hours spruce wood samples. According to Robson et al, aldo/keto reductases are required for various metabolic reactions such as degradation of β -aryl ethers present in lignin, degradation of carbohydrates and detoxification of xenobiotic compounds [56] (Table 3.3).

Table 3.3: Lists differentially classified lignin degrading enzymes obtained from different gene expression datasets of *P. chrysosporium*

<i>P. chrysosporium</i> cells genes involved in detoxification and stress responsive pathways	
Detoxification and stress responsive enzymes	2-oxo acid dehydrogenase, Alkylhydroperoxide reductase, Peroxiredoxin, Manganese/iron superoxide dismutase, Amidohydrolase Beta lactamase, Urease, Isoflavone reductase, Dihydroorotate dehydrogenase, Abortive infection protein, Heat shock protein, Ubiquitin enzyme complex, Thaumatin, Thioredoxin, Ferredoxin, Flavodoxin, Thiolase, Catalase, Aldo/Keto reductases, etc.,
Phase-I metabolic enzymes encoding genes	Cytochrome P450 monooxygenases, Epoxide hydrolases
Phase-II metabolic enzymes encoding genes	UDP-glucuronosyltransferases, Sulfotransferases, N-acyl transferases, Glutathione-S-transferases, Thioredoxin

3.4.3. Phase I Metabolic Enzymes

The biotransformation of xenobiotic compounds in fungal cells commences majorly through phase-I and phase-II reactions. The phase-I metabolic reactions majorly include

transformation of parent compound to polar metabolites through de novo formation of functional groups such as -OH, -NH₂, -SH [60, 65]. Phase-I metabolic enzymes such as cytochrome P450 monooxygenases, epoxide hydrolases and dioxygenases are involved in N- and O-dealkylation, hydroxylation of aliphatic and aromatic compounds, N- and S- oxidation and deamination reactions [65]. The ability to protect against harmful toxic external xenobiotic compounds can be majorly imparted to the complex enzymatic defense systems majorly including cytochrome P450 monooxygenases. In 2005 Doddapaneni and Yadav, have performed a customized genome wide microarray of *P. chrysosporium* for studying the global expression of cytochrome P450 monooxygenases under nutrient rich and nutrient limited growth conditions. Out of 150 cytochrome P450 encoding genes expressed, 23 genes were differentially expressed by 2.0 to 9.0-fold in nitrogen rich conditions and 4 genes expressed by 2.0 to 20-fold in low nitrogen conditions respectively. Subramanian V and Yadav, J.S (2009), have performed a genome wide role of cytochrome P450 monooxygenases in nonylphenol degradation by *P. chrysosporium*. This study has revealed that nonylphenol has induced multiple P450 monooxygenases out of which 18 genes were expressed with a fold change of 2 to 195 in nutrient-rich conditions, in low-nutrient growth conditions 17 genes with fold change of 2 to 6 and 3 genes were found to be expressed common among both these conditions. The current metadata analysis study has revealed the expression of cytochrome P450 monooxygenase encoding genes especially in natural plant biomass containing growth substrates. In GSE14736, cytochrome P450 encoding genes were found to be expressed in BMA (GSE14734, GSE27941), Replete and nitrogen-limited conditions with two-fold expression. In 52922 dataset genes encoding cytochrome P450 monooxygenases were differentially expressed in 64 and 82 transgenic lines respectively. In GSE54542, GSE69008 and GSE69461 datasets cytochrome P450 encoding genes were highly expressed in control, low lignin-high glucose, average lignin-average glucose and 40-hours spruce wood growth samples. Epoxide hydrolases one of the important phase-I metabolic enzymes are required for the cellular epoxide or oxiranes transformation. These enzymes majorly exhibit three functions a) detoxification b) catabolism and regulation of signaling molecules [66]. Oxiranes or epoxides are highly toxic compounds effecting the cellular growth and development, epoxide hydrolases released by fungal cells catalytically add water molecules to the epoxides by resulting in corresponding 1,2-diols or glycols [66, 67]. Genes encoding epoxide hydrolases were found to be highly expressed in 40-hours spruce wood, high glucose-low lignin, average lignin-average glucose transgenic lines, control, BMA growth conditions.

Based on the normalized values epoxide hydrolase was found to be differentially expressed in transgenic line 82 and glucose growth samples.

3.4.3. Phase II Metabolic Enzymes

The phase-II enzymes also play a crucial role biotransformation of xenobiotic and endogenous compounds by inactivating the active substances and converting it to easily excretable forms[65]. Phase-II enzymes majorly perform conjugating reactions by employing transferases such as glutathione-S-transferases, UDP-glucuronosyltransferases, sulfotransferases, GCN5-acyltransferases, O-methyltransferases, NAD(P)H quinone oxidoreductases, MAPEG (Membrane-associated proteins in eicosanoid and glutathione metabolism) and GFA (glutathione dependent formaldehyde activating enzyme) [65]. As reported in our previous work glycosyl transferases belonging to GT-1, GT-2, GT-4, GT-8, GT-20, GT-35, GT-39 and GT-48 were found commonly expressed among various datasets. Glycosyl transferases were found to be highly upregulated in growth mediums containing complex plant biomass. Genes encoding glycosyl transferases expressed among the datasets were BMA (GT-1, GT-2, GT-20 and GT-39), glucose and replete (GT-8, GT-48), oak acetic extracts (GT-2), high lignin-low glucose (GT-1, GT-4, GT-20), low lignin-high glucose (GT-2, GT-8, GT-15, GT-35, GT-39, GT-48), average glucose-average lignin (GT-2, GT-8, GT-15, GT-20, GT-35, GT-39, GT-48), P717 hybrid line and transgenic line 82 (GT-39), 40-hours spruce wood samples (GT-1, GT-2, GT-15, GT-48) respectively. Genes encoding GCN5-acyltransferase were highly expressed in carbon limited and replete growth mediums, transgenic line 64, low lignin-high glucose, average lignin-average glucose and 96-hour spruce wood samples. O-methyltransferases encoding genes were highly expressed in low lignin-high glucose, average lignin-average glucose, 40-hours and 96-hours spruce wood samples, transgenic line 64, carbon and nitrogen limited, BMA, control growth samples. Glutathione-s-transferases encoding genes were differentially expressed in glucose, nitrogen limited, BMA, oak acetic extracts, transgenic line 64, low lignin-high glucose and average lignin-average glucose growth conditions 96-hour spruce wood samples. Other enzymes constituting for the glutathione system such as glutathione dependent formaldehyde activator were differentially expressed in control and 96-hour spruce wood samples and high lignin-low glucose based on their expression values. MAPEG encoding genes were expressed in transgenic line 64 and oak acetic extract growth samples. Folate cycle plays a crucial role in the maintenance of glutathione levels by sequestering formaldehyde (a toxic

compound) formed from endo and exogenous compounds. Our present data analysis has revealed that genes encoding for methylenetetrahydrofolate reductase and dihydrofolate reductase enzymes were found commonly expressed by *P. chrysosporium* when cultured on complex plant biomass growth mediums. MTHFR encoding gene was highly expressed in glucose, nitrogen-limited, low lignin-high glucose, average lignin-average glucose and 96-hour spruce wood growth samples, DHFR encoding gene were differentially expressed in high lignin-low glucose. In oak acetic extract samples tetrahydrofolate dehydrogenase encoding gene was highly expressed. Thiolase encoding gene was highly expressed in oak acetic extract, replete, BMA and 40-hours spruce wood samples.

3.4.4. Effect of Growth Substrate and Incubation Period

Based on the gene expression studies and present metadata analysis we have found that *P. chrysosporium* gene expression is strongly influenced by the growth substrate. The gene expression profiles of *P. chrysosporium* were significantly different when cultured on synthetic (such as glucose, cellulose, carbon-limited, nitrogen-limited and replete growth mediums) and natural plant biomass growth substrates (such as ball milled aspen, ball milled pine, oak acetic extracts, natural and genetically modified *Poplar* wood substrates). When *P. chrysosporium* was cultured on simple synthetic growth substrates (glucose, cellulose and replete) genes encoding for CAZymes were highly expressed along with genes required for the normal cell progression, growth and metabolism were found to be highly expressed. Importantly, genes encoding were animal haem peroxidases, chloroperoxidases were down regulated in carbon, nitrogen limited and replete mediums but highly expressed in glucose, cellulose and ball milled aspen growth substrates (Figure 3.1). At the same time, several other genes encoding for lignin degrading enzymes such as lignin peroxidase, glyoxal oxidase, copper amine oxidase, flavin amine oxidase, cytochrome P450 monooxygenase, glycosyl transferases 2OG-Fe(II) oxygenase family and several other enzymes were up regulated when cultured in carbon and nitrogen limited growth mediums. When *P. chrysosporium* was cultured in natural plant biomass growth substrates various genes encoding aromatic compound degrading enzymes were highly expressed (Table 3.2). Genes encoding stress and detoxification responsive enzymes were found to be highly expressed in these datasets, as natural plant biomass substrates also contain different plant extractives like flavonoids, tannins, quinones and stilbenes which cause severe toxicity to the fungal cells. Apart from the toxic plant extractives lignin and its degraded products also induce toxicity and stress on the

fungal cells, due to which several stress responsive genes were highly expressed by the fungal cells such as cytochrome P450 monooxygenases, glutathione-s-transferase, thaumatin, abortive infection, heat shock and ubiquitin complex proteins. Genes involved in DNA, RNA and protein modification genes especially proteases (serine/threonine, aspartic) were also found to be highly expressed in natural plant biomass growth substrates.

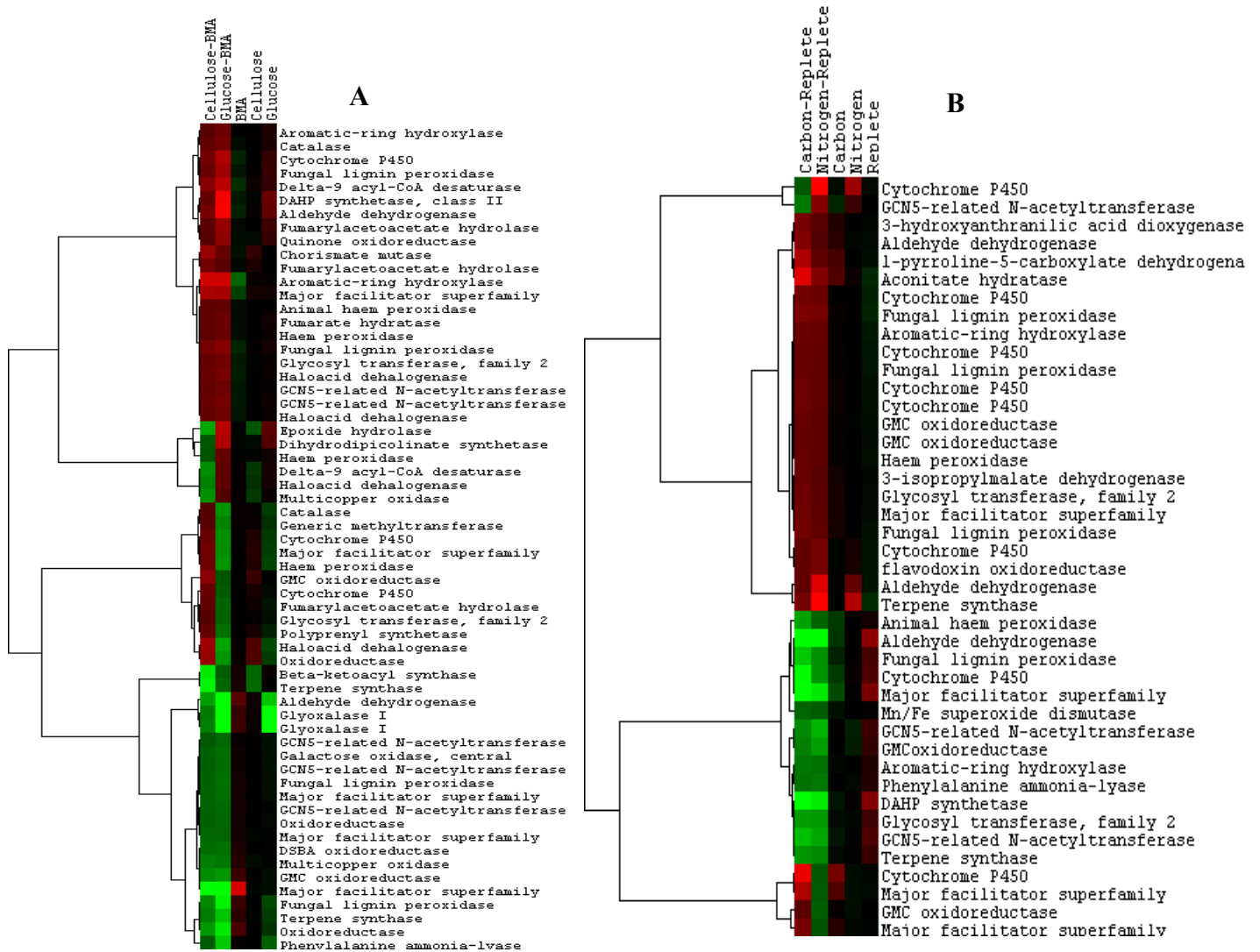
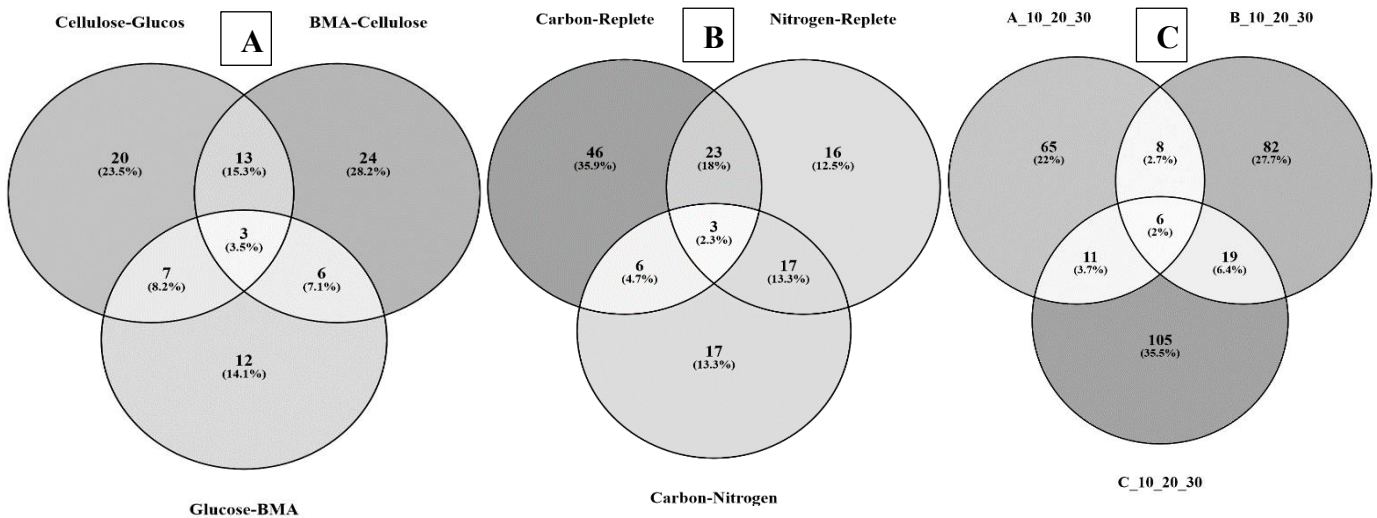


Figure 3.1: Hierarchical clusters showing the differentially expressed genes obtained, 1st and 2nd columns lists the fold change expression values up regulated in cellulose and glucose (A) GSE14734, Carbon and nitrogen limited in (B) GSE14735 and 3rd, 4th and 5th columns in A and B lists the log transformed gene expression values of BMA, cellulose, glucose, carbon-limited, nitrogen-limited and replete growth mediums respectively.

The gene expression pattern of *P. chrysosporium* is also significantly influenced by the incubation period, gene expression studies especially GSE69008, GSE69461 and GSE6649

confirms it. Lignin being a large heterophenolic polymers requires a wide range of enzymes and complex systems for its degradation. In studies GSE69008, GSE69461 and GSE6649, *P. chrysosporium* gene expression was monitored for 10, 20, 30-days, 40-hours and 90-hours and Day-2, Day-3 respectively. Results obtained from GSE69461 dataset show that genes encoding for lignin and manganese, chloro-peroxidases, alcohol oxidase, amine oxidases, multicopper oxidases, phenyl ammonia lyase (forming lignin degrading enzyme system) were highly expressed in 90-hours incubation periods. Where as in 40-hours incubation period samples genes encoding for cytochrome P450 monooxygenase, fumarate reductase, aryl alcohol dehydrogenase, alternative oxidase, copper radical oxidase, FAD binding, phenol 1,2 monooxygenases were highly expressed. In GSE69008 dataset gene expression was significantly influenced by both the substrate and incubation period. The genes encoding for lignin degradation and genes involved in detoxification responses were also expressed among the 10 and 20-days incubation period samples (Figure 3.2).



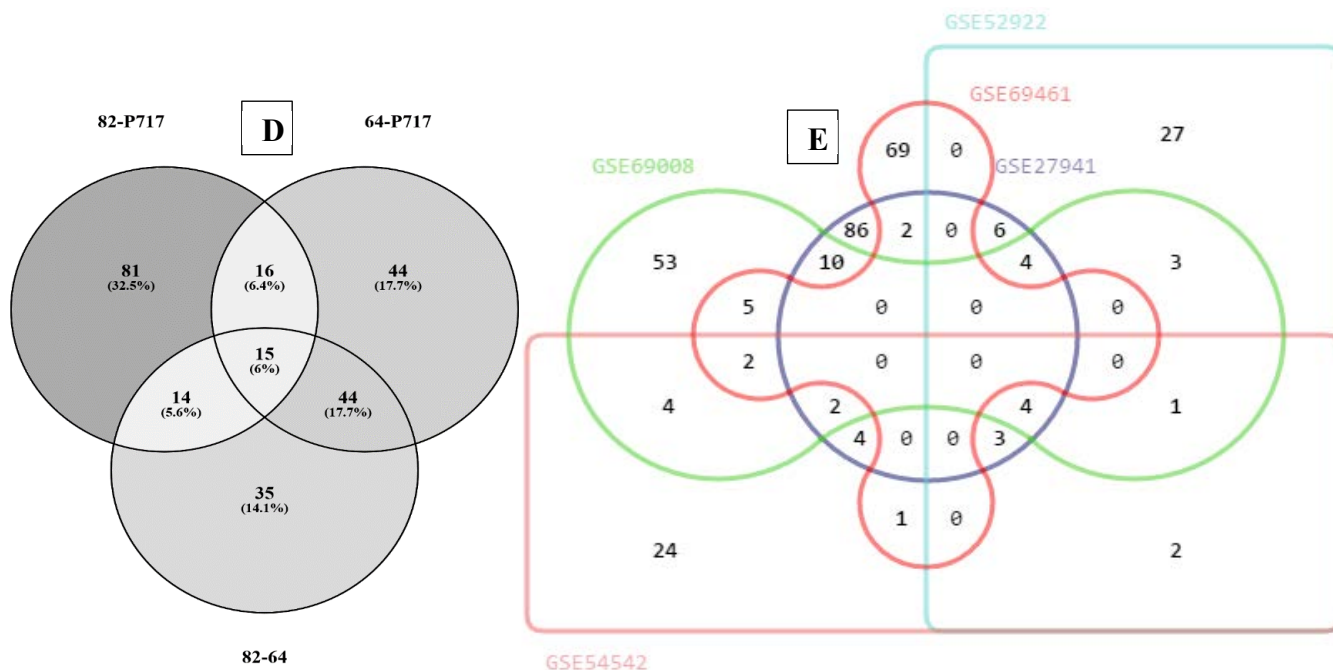


Figure 3.2: Venn diagrams of differentially expressed genes obtained from the gene expression datasets A) GSE14734 B) GSE14735 C) GSE69008 (where A, B, C represents high lignin-low glucose, high glucose-low lignin, average lignin-average glucose, 10, 20, 30 represents number of days) D) GSE52922 (where P717, 82 and 64 represents parent and transgenic lines of *Poplar trichocarpa* species respectively) E) GSE27941, GSE52922, GSE54542, GSE69008, GSE69461 (*Picea glauca* species).

3.5. Discussion

The complex and stable structure of lignin makes its degradation process slow, except few aerobic fungi, aerobic and anaerobic bacteria even plants lack the metabolic pathways required for recycling lignin [52]. Lignin oxidizing enzymes of *P. chrysosporium* includes lignin peroxidase (10 lip), manganese peroxidase (5 mnp) and l(NoP/ VP). These peroxidases are haem containing proteins by containing a ferric heme group [Fe (III)] at its resting state which reacts with hydrogen peroxide (H₂O₂) resulting in Compound-I oxo-ferryl intermediate (two-electron oxidized) containing [Fe (IV)]. The compound-I further oxidizes the donor substrate forming a second intermediate Compound-II, both the reactions releases free radicals [68, 69]. The catalytic action of lignin peroxidase is aided by veratryl alcohol (diffusible oxidant) produced by *P. chrysosporium* (as a metabolite) VA cation radical reacts with lignin molecule on remote locations. VA also supports in functioning and higher

expression of LiP enzyme [68, 70]. Veratryl alcohol is produced by phenylalanine ammonia lyase (PAL) which catalyzes the first step in the VA formation by oxidizing L-phenylalanine to cinnamic acid and free ammonium ion through non-oxidative deamination [71]. Another enzyme O-methyltransferase which transfers two-ring methoxyl group on VA [18]. Whereas Mn(II) being a highly oxidant acts as diffusible oxidant reacts with the lignin molecule even on the remote locations of lignin without the catalytic active center getting involved through oxidizing the lipid peroxidation reactions where in delta-9 fatty acid desaturases supports the catalytic function of MnP [69].

Ligninolytic peroxidases are similar to classic peroxidases in their function as they are dependent on hydrogen peroxide for their function. The process of hydrogen peroxide generation in *P. chrysosporium* is controlled by set of enzymes glyoxal oxidase, pyranose oxidase, aryl alcohol oxidase, veratryl alcohol oxidase, amine (copper amine) copper radical oxidase and alcohol oxidases. Genes encoding glyoxal oxidase, copper amine oxidase, alcohol oxidase, copper radical oxidase were commonly expressed among various datasets as explained above. Two unconventional genes encoding for chloroperoxidase, animal haem peroxidase was also being highly expressed on cellulose, BMA, replete, transgenic line 64 and low lignin-average glucose growth substrates. Based on the expression of genes coding for ferric reductase, ferroxidase, cellobiose dehydrogenase, LPMO and quinone reductase in plant biomass containing growth substrates proposes that *P. chrysosporium* depends on these enzymes for generation of toxic hydroxy radicals and in iron homeostasis. The hydroxy radicals generated upon reaction of hydrogen peroxide and iron-oxalate complex, these highly toxic radicals (OH*) attack the complex lignin molecules [69]. The above-mentioned enzymes can be classified to participate in primary ligninolytic reactions based on their functional properties and due to their common expression among all the growth substrates (Figure 3.3). Along with the above-mentioned lignin oxidizing enzymes, a wide range of aromatic compound degrading enzymes, stress responsive and detoxifying enzymes operate in coordination to degrade lignin and its derived products.

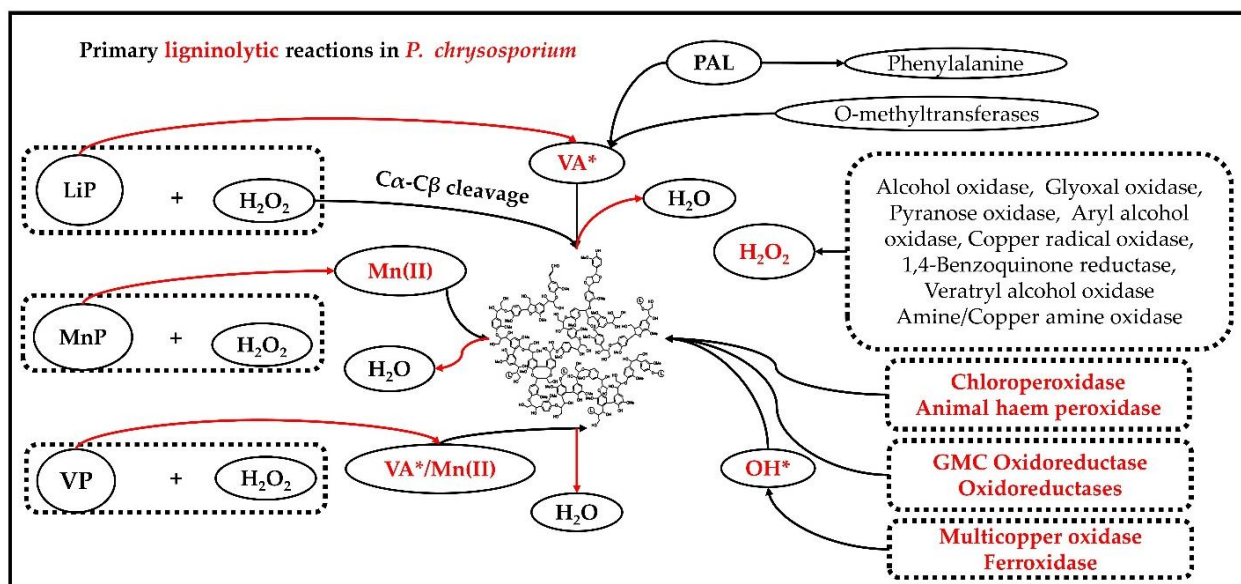


Figure 3.3: Shows the primary enzymatic reactions of *P. chrysosporium* involved in lignin Degradation LiP (lignin peroxidase), MnP (Manganese peroxidase), VP (Versatile peroxidase), VA (Veratryl Alcohol), PAL (Phenylalanine ammonia lyase), MCO (Multicopper oxidase).

Degradation of lignin molecules by *P. chrysosporium* results in various low-molecular weight chemical compounds, if progressive and controlled lignin breakdown strategies are developed it can lead to the production useful renewable and green platform chemicals [69]. Studies conducted previously on spruce wood degradation by *P. chrysosporium* has resulted in total 28 low molecular weight chemical compounds out of which 10 were aromatic carboxylic acids [72] and 13 were acyclic 2,4-hexadiene-1,6-dioic acids obtained through oxidative ring cleavage [73]. The chemical compounds obtained from lignin degradation were derivatives of benzoic acids by oxidative cleavage of C α -C β of lignin components. The biphenyl and diphenyl ether dicarboxylic acids obtained were derivatives of biphenyl and diphenyl ether components of lignin [69]. Studies have also reported that metabolism of β -aryl ether model compound by *P. chrysosporium* involves C α -C β oxidative cleavage resulting in vanillin [74]. Lignin peroxidase of *P. chrysosporium* catalyzes the oxidative cleavage of C α -C β bond of various lignin derivative compounds [49] such as diarylpropane [75], β -aryl ether model [74], phenylcoumarane [76] compounds resulting in aromatic aldehyde products, vanillin compounds respectively. It is well known that lignin peroxidase is also involved in degradation of non-phenolic units of lignin polymer (which constitutes to 90% of lignin polymer) at the same time MnP is involved in degradation of both phenolic and non-phenolic

units of lignin compound [68, 69]. According to Fumiaki et al (1981), the degradation process of alkylated phenylcoumarane by *P. chrysosporium* is directed via primary oxidation of side chains followed by the oxidation of heterocyclic ring to furan and then performs the oxidative cleavage of C α -C β bond [76].

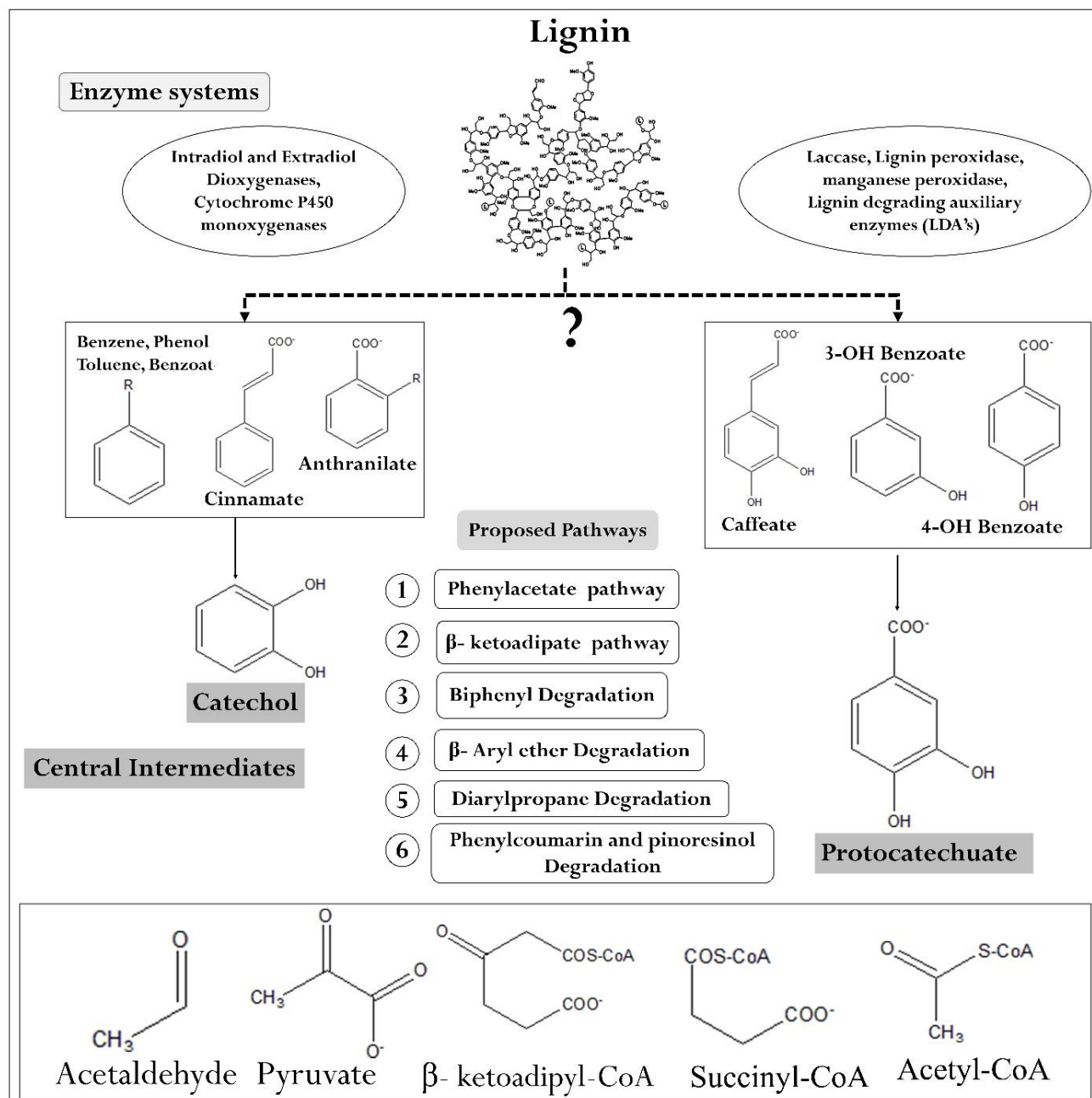


Figure 3.4: Bird's eye view of tentative and proposed general molecular mechanisms and pathways involved in lignin degradation.

The process of lignin degradation results in a wide range of aromatic compounds which are further degraded by a set of aromatic compound degrading enzymes. Enzyme systems involving oxygenases attack these aromatic compounds in the presence of oxygen resulting in two central intermediate compounds catechol and protocatechuate which are further degraded by intra and extradiol dioxygenases which catalyze the oxidative cleavage of central ring [52]. Microorganisms employ variety of metabolic pathways divided as upper pathways (resulting in catechol and protocatechuate) and lower pathways (acetyl-CoA, succinyl-CoA and pyruvate) to efficiently utilize wide range of aromatic substrates generated due to the process of degradation (Figure 3.4) [52]. The up regulation of genes encoding for central ring cleaving dioxygenases (Intra and extradiol dioxygenase), cytochrome P450 monooxygenases, aromatic ring hydroxylase dioxygenases, catechol 1,2-dioxygenase and other dioxygenases in complex plant biomass growth substrates supports the involvement of upper and lower pathways by *P. chrysosporium*.

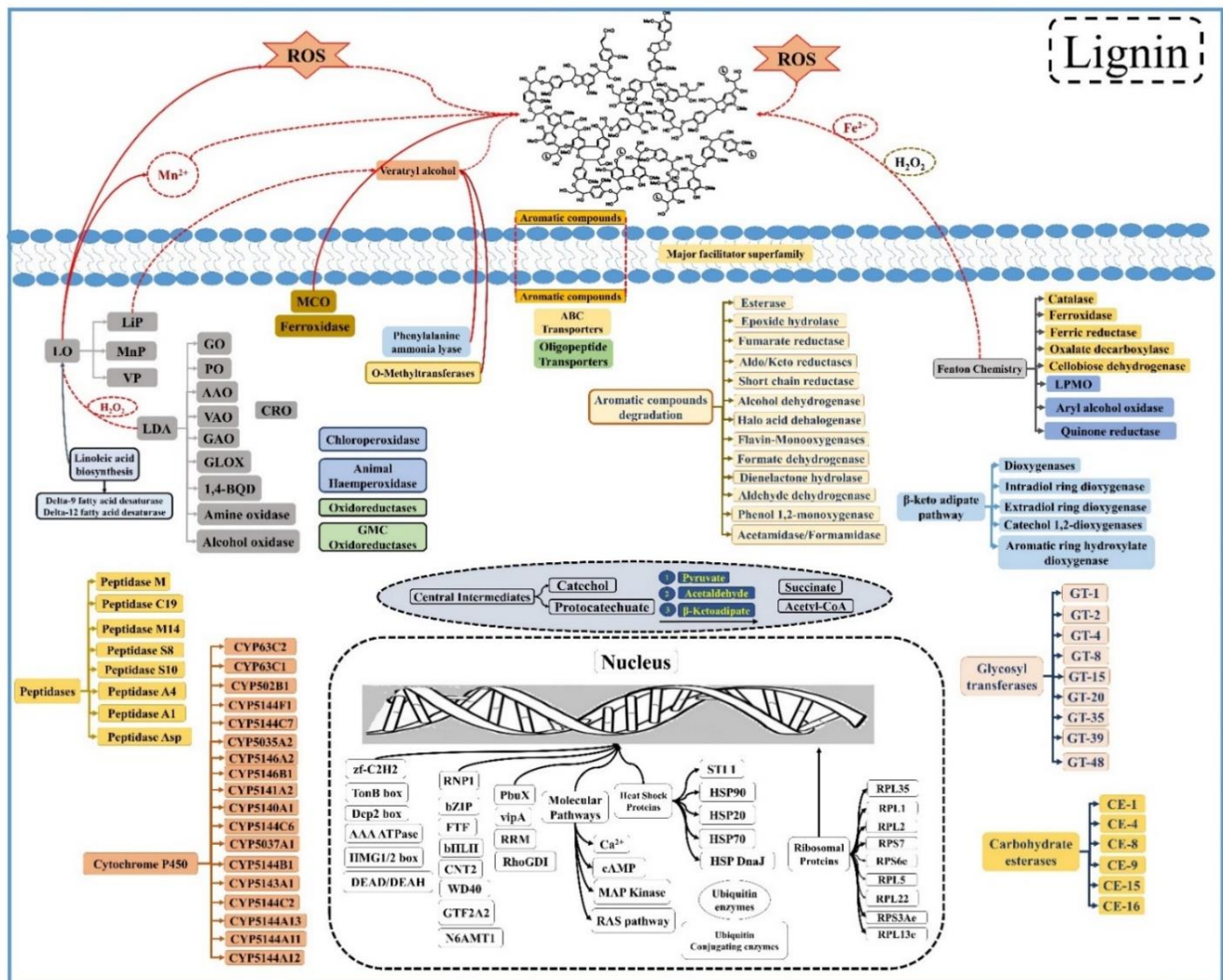


Figure 3.5: Tentative network of *P. chrysosporium* genes and enzymes involved in lignin degradation mechanisms.

Naturally, plant cell wall components especially hemicellulose and lignin occurs in acetylated forms to cease the activity of various hydrolyzing enzymes of microbial origin [61]. Higher expression of genes encoding for carbohydrate esterases endorses earlier reports on deacetylation of lignin units. According to Del Rio et al (2007) in lignin, acetyl groups were found to be associated with gamma carbon of aliphatic side chains situated on the syringyl and guaiacyl monomers of lignin with highest degree of acetylation observed in jute fibers, abaca and kenaf with 0.8 DA, acetylation of lignin in hardwood varies between the range of 1 to 50% (w/w) [77]. Along with carbohydrate esterases genes encoding for various glycosyl transferase class enzymes were found to be highly expressed in both customized synthetic and complex natural plant biomass growth substrates suggesting their role in controlling cellular toxicity. During the process of plant biomass degradation *P. chrysosporium* reacts with the toxic properties of lignin and its derivatives, thus there might be potential involvement of glycosyl transferases in changing the toxic properties of lignin derivative compounds. According to Julien et al (2016), glycosyl transferases are significantly involved in detoxification of plants during the secondary metabolite synthesis (flavonoids, lignin and phenylpropanoids) and process of glycosylation modifies their solubility, stability and toxicity [78]. In our analysis, we have also found that several genes involved in the process of DNA synthesis, modification and repair were highly expressed along with a set of transcription factors among the top differentially expressed genes. Genes encoding for various proteases especially serine, cysteine carboxyproteases were found to be commonly and differentially expressed among all the datasets (Figure 3.5).

As discussed earlier, fungal cells are exposed to highly toxic and stressful environments during the process of degradation fungal cells encounter polyphenolic lignin units and its degraded products along with other wood extractives [1, 21]. It was also reported that various chemical bioproducts such as aldehydes, aliphatic acids, phenolic and furan derivatives are obtained because of lignocellulosic biotreatment methods which majorly inhibit the action of ligninolytic enzymes and further microbial fermentation [79]. In order to protect from these toxic substances (degraded products of lignin and its derivatives, plant extractives and ROS, hydroxy radicals and superoxides /free radicals) *P. chrysosporium* secretes a wide range of antioxidant and stress responsive enzymes such as cytochrome P450

monooxygenases, glutathione-S-transferases, catalases, superoxide dismutases majorly classified as phase-I and phase-II metabolic systems. We have clearly observed the differential expression of genes encoding for phase-I and phase-II enzyme systems by *P. chrysosporium* when cultured on complex plant biomass growth substrates (Figure 3.6). The process of lignin degradation in *P. chrysosporium* is associated with ROS production, reactive oxygen and hydroxyl radicals which also cause additional toxicity. Results obtained from our present study and gene expression studies reported earlier strictly convey that the process of lignin degradation is interdependent on the intracellular detoxification systems and cellular redox states. Cellular redox states were characterized by redox-active species (such as thiol disulfide and pyridine nucleotide compounds) and their degree of oxidation/reduction reactions [64]. While the balanced ratios of NADPH/NADP⁺ and SH/S-S facilitate the process of redox regulation by effecting the proteins directly, at the same time NADPH/NADP⁺ and SH/S-S ratios are directly related to the cellular ROS level [64]. Genes coding for 2-oxoacid dehydrogenase multienzyme complexes which play a crucial role in regulating the cellular redox states were found to be highly expressed on complex natural plant biomass-based growth substrates. Similarly, higher expression of thioredoxin, glutathione-S-transferases, peroxiredoxin, NADPH oxidases etc. explains that *P. chrysosporium* continuously maintains the cellular redox state and controls the toxic conditions developed due to lignin degradation. The present metadata analysis is in complete accordance with previous reporting's made on production of ligninolytic enzyme machinery upon nitrogen repression and natural plant biomass containing growth substrates. We have also observed the expression of genes encoding for ligninolytic enzymes and intracellular antioxidant mechanisms simultaneously by *P. chrysosporium* when cultured on synthetic growth substrates mimicking ligninolytic conditions GSE6649, GSE14735 (nitrogen limited, carbon limited). Especially lignin and manganese peroxidases and lignin degrading auxiliary enzymes, were highly expressed in day 3 cultures and nitrogen limited mediums of GSE6649 and GSE14735 datasets respectively, followed by expression of enzymes such as cytochrome P450, thioredoxin, Mn²⁺superoxide dismutase and other intracellular anti-oxidant enzymes support the above process of degradation. Similar gene expression patterns were observed in *P. chrysosporium* when cultured on natural plant biomass growth substrates with clear expression of various enzymes involved in lignin degrading and detoxification, stress responsive mechanisms.

Through our present analysis, we have demonstrated the functional involvement of various enzymes in lignin degradation and detoxification by *P. chrysosporium*. These outcomes clearly indicate that unlike cellulose, hemicellulose *P. chrysosporium* invests high proportions of molecular and metabolic systems in the process of lignin degradation. Based on the results obtained we propose that process of lignin degradation in *P. chrysosporium* commences through synchronous expression of both ligninolytic enzymes and detoxification-stress responsive systems. Both previous reports and present results convey that *P. chrysosporium* involves highly toxic ROS, free radicals etc. along with conventional ligninolytic enzymes to create random cuts in the lignin structure, by creating the platform for detoxification and stress responsive enzymes. However, there is certain ambiguity related with lignin degradation and detoxification-stress responsive pathways proposed, as we have observed that several highly-expressed genes were annotated as uncharacterized proteins. Further proteomic studies must be continued to understand their involvement in lignin degradation and detoxification-stress responsive mechanisms of *P. chrysosporium*.

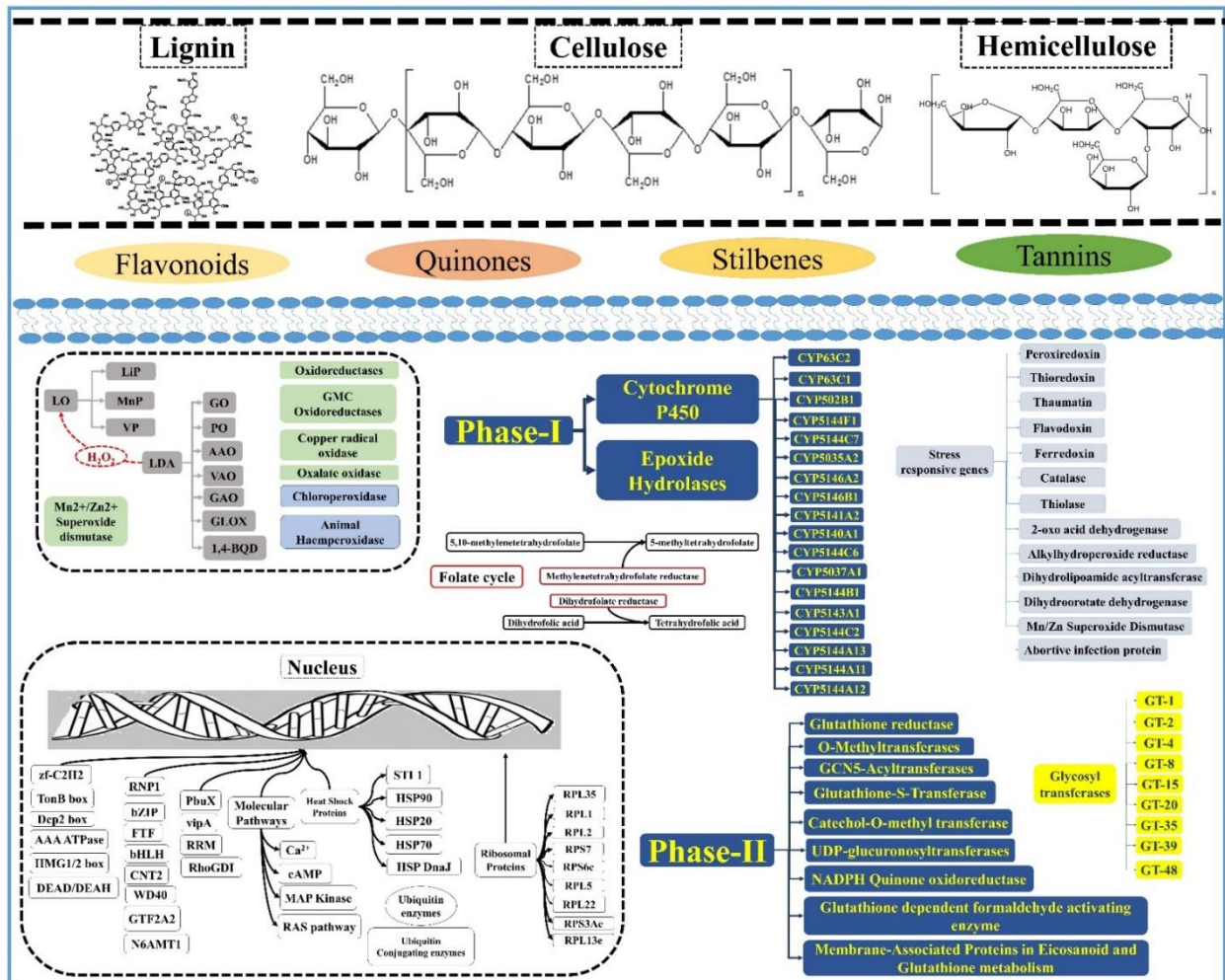


Figure 3.6: Tentative network of *P. chrysosporium* genes and enzymes involved in detoxification mechanisms involving phase-I, phase-II and stress responsive pathways.

References:

1. Nascimento M, Santana A, Maranhão C, Oliveira L, Bieber L. Phenolic extractives and natural resistance of wood. <http://www.intechopen.com/books/biodegradation-life-of-science>. 2013.
2. Singh D, Chen S. The white-rot fungus *Phanerochaete chrysosporium*: conditions for the production of lignin-degrading enzymes. *Applied microbiology and biotechnology*. 2008; 81: 399-417.
3. Randall TA, Reddy CA. The nature of extra-chromosomal maintenance of transforming plasmids in the filamentous basidiomycete *Phanerochaete chrysosporium*. *Current genetics*. 1992; 21: 255-60.
4. Valli K, Gold MH. Degradation of 2, 4-dichlorophenol by the lignin-degrading fungus *Phanerochaete chrysosporium*. *Journal of bacteriology*. 1991; 173: 345-52.
5. Valli K, Brock BJ, Joshi DK, Gold M. Degradation of 2, 4-dinitrotoluene by the lignin-degrading fungus *Phanerochaete chrysosporium*. *Applied and environmental microbiology*. 1992; 58: 221-8.
6. Kullman SW, Matsumura F. Metabolic pathways utilized by *Phanerochaete chrysosporium* for degradation of the cyclodiene pesticide endosulfan. *Applied and Environmental Microbiology*. 1996; 62: 593-600.
7. Reddy GVB, Gold MH. Degradation of pentachlorophenol by *Phanerochaete chrysosporium*: intermediates and reactions involved. *Microbiology*. 2000; 146: 405-13.
8. Keyser P, Kirk T, Zeikus J. Ligninolytic enzyme system of *Phanaerochaete chrysosporium*: synthesized in the absence of lignin in response to nitrogen starvation. *Journal of bacteriology*. 1978; 135: 790-7.
9. Hammel KE, Kalyanaraman B, Kirk TK. Oxidation of polycyclic aromatic hydrocarbons and dibenzo [p]-dioxins by *Phanerochaete chrysosporium* ligninase. *Journal of Biological Chemistry*. 1986; 261: 16948-52.
10. Martinez D, Larrondo LF, Putnam N, Gelpke MDS, Huang K, Chapman J, et al. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nature biotechnology*. 2004; 22: 695-700.
11. Wymelenberg AV, Mingos P, Sabat G, Martinez D, Aerts A, Salamov A, et al. Computational analysis of the *Phanerochaete chrysosporium* v2. 0 genome database and mass spectrometry identification of peptides in ligninolytic cultures reveal complex mixtures of secreted proteins. *Fungal Genetics and Biology*. 2006; 43: 343-56.
12. Minami M, Kureha O, Mori M, Kamitsuji H, Suzuki K, Irie T. Long serial analysis of gene expression for transcriptome profiling during the initiation of ligninolytic enzymes production in *Phanerochaete chrysosporium*. *Applied microbiology and biotechnology*. 2007; 75: 609-18.
13. Wymelenberg AV, Gaskell J, Mozuch M, Kersten P, Sabat G, Martinez D, et al. Transcriptome and secretome analyses of *Phanerochaete chrysosporium* reveal complex patterns of gene expression. *Applied and environmental microbiology*. 2009; 75: 4058-68.
14. Wymelenberg AV, Gaskell J, Mozuch M, Sabat G, Ralph J, Skyba O, et al. Comparative transcriptome and secretome analysis of wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium*. *Applied and environmental microbiology*. 2010; 76: 3599-610.
15. Wymelenberg AV, Gaskell J, Mozuch M, BonDurant SS, Sabat G, Ralph J, et al. Significant alteration of gene expression in wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium* by plant species. *Applied and environmental microbiology*. 2011; 77: 4499-507.
16. Thuillier A, Chibani K, Belli G, Herrero E, Dumarçay S, Gérardin P, et al. Transcriptomic Responses of *Phanerochaete chrysosporium* to Oak Acetonic Extracts: Focus on a New Glutathione Transferase. *Applied and environmental microbiology*. 2014; 80: 6316-27.
17. Gaskell J, Marty A, Mozuch M, Kersten PJ, BonDurant SS, Sabat G, et al. Influence of *Populus* genotype on gene expression by the wood decay fungus *Phanerochaete chrysosporium*. *Applied and environmental microbiology*. 2014; 80: 5828-35.
18. Korripally P, Hunt CG, Houtman CJ, Jones DC, Kitin PJ, Cullen D, et al. Regulation of Gene Expression during the Onset of Ligninolytic Oxidation by *Phanerochaete chrysosporium* on Spruce Wood. *Applied and environmental microbiology*. 2015; 81: 7802-12.

19. Skyba O, Cullen D, Douglas CJ, Mansfield SD. Gene expression patterns of wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium* are influenced by wood substrate composition during degradation. *Applied and environmental microbiology*. 2016; AEM. 00134-16.
20. Nascimento M, Santana A, Maranhão C, Oliveira L, Bieber L. Phenolic extractives and natural resistance of wood. *Intech*. 2013; 13: 349-70.
21. Shalaby S, Horwitz BA. Plant phenolic compounds and oxidative stress: integrated signals in fungal-plant interactions. *Current genetics*. 2015; 61: 347-57.
22. Harborne JB, Williams CA. Advances in flavonoid research since 1992. *Phytochemistry*. 2000; 55: 481-504.
23. Feraydoni V, Hosseinihashemi SK. Effect of walnut heartwood extractives, acid copper chromate, and boric acid on white-rot decay resistance of treated beech sapwood. *BioResources*. 2012; 7: 2393-402.
24. Doddapaneni H, Yadav J. Microarray-based global differential expression profiling of P450 monooxygenases and regulatory proteins for signal transduction pathways in the white rot fungus *Phanerochaete chrysosporium*. *Molecular Genetics and Genomics*. 2005; 274: 454-66.
25. Sakamoto T, Kitaura H, Minami M, Honda Y, Watanabe T, Ueda A, et al. Transcriptional effect of a calmodulin inhibitor, W-7, on the ligninolytic enzyme genes in *Phanerochaete chrysosporium*. *Current genetics*. 2010; 56: 401-10.
26. Suetomi T, Sakamoto T, Tokunaga Y, Kameyama T, Honda Y, Kamitsuji H, et al. Effects of calmodulin on expression of lignin-modifying enzymes in *Pleurotus ostreatus*. *Current genetics*. 2015; 61: 127-40.
27. Subramanian V, Yadav JS. Role of P450 monooxygenases in the degradation of the endocrine-disrupting chemical nonylphenol by the white rot fungus *Phanerochaete chrysosporium*. *Applied and environmental microbiology*. 2009; 75: 5570-80.
28. Jiang M, Li X, Zhang L, Feng H, Zhang Y. Gene expression analysis of *Phanerochaete chrysosporium* during the transition time from primary growth to secondary metabolism. *The Journal of Microbiology*. 2009; 47: 308-18.
29. Subramanian V, Yadav JS. Regulation and heterologous expression of P450 enzyme system components of the white rot fungus *Phanerochaete chrysosporium*. *Enzyme and microbial technology*. 2008; 43: 205-13.
30. Chigu NL, Hirose S, Nakamura C, Teramoto H, Ichinose H, Wariishi H. Cytochrome P450 monooxygenases involved in anthracene metabolism by the white-rot basidiomycete *Phanerochaete chrysosporium*. *Applied microbiology and biotechnology*. 2010; 87: 1907-16.
31. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otilar R, et al. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research*. 2013; gkt1183.
32. Grigoriev IV, Cullen D, Goodwin SB, Hibbett D, Jeffries TW, Kubicek CP, et al. Fueling the future with fungal genomics. *Mycology*. 2011; 2: 192-209.
33. Botstein D, Cherry JM, Ashburner M, Ball C, Blake J, Butler H, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000; 25: 25-9.
34. Consortium GO. Gene ontology consortium: going forward. *Nucleic acids research*. 2015; 43: D1049-D56.
35. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. *BMC bioinformatics*. 2003; 4: 41.
36. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic acids research*. 2009; 37: D233-D8.
37. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research*. 2014; 42: D490-D5.
38. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics*. 2004; 20: 1453-4.
39. Saldanha AJ. Java Treeview—extensible visualization of microarray data. *Bioinformatics*. 2004; 20: 3246-8.
40. Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams. 2007.
41. Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. jvenn: an interactive Venn diagram viewer. *BMC bioinformatics*. 2014; 15: 1.

42. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. 2015: gkv007.
43. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26: 139-40.
44. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic acids research*. 2012: gks042.
45. Robinson MD, Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*. 2008; 9: 321-32.
46. Zhou X, Lindsay H, Robinson MD. Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic acids research*. 2014; 42: e91-e.
47. Minami M, Suzuki K, Shimizu A, Hongo T, Sakamoto T, Ohyama N, et al. Changes in the gene expression of the white rot fungus *Phanerochaete chrysosporium* due to the addition of atropine. *Bioscience, biotechnology, and biochemistry*. 2009; 73: 1722-31.
48. Kameshwar AKS, Qin W. Metadata Analysis of *Phanerochaete chrysosporium* Gene Expression Data Identified Common CAZymes Encoding Gene Expression Profiles Involved in Cellulose and Hemicellulose Degradation. *International Journal of Biological Sciences*. 2017; 13: 85-99.
49. Kirk TK, Farrell RL. Enzymatic "combustion": the microbial degradation of lignin. *Annual Reviews in Microbiology*. 1987; 41: 465-501.
50. Alic M, Mayfield MB, Akileswaran L, Gold MH. Homologous transformation of the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. *Current genetics*. 1991; 19: 491-4.
51. Thurston CF. The structure and function of fungal laccases. *Microbiology*. 1994; 140: 19-26.
52. Fuchs G, Boll M, Heider J. Microbial degradation of aromatic compounds—from one strategy to four. *Nature Reviews Microbiology*. 2011; 9: 803-16.
53. Dagley S, Evans WC, Ribbons D. New pathways in the oxidative metabolism of aromatic compounds by micro-organisms. 1960.
54. Gibson DT, Parales RE. Aromatic hydrocarbon dioxygenases in environmental biotechnology. *Current Opinion in Biotechnology*. 2000; 11: 236-43.
55. Vaillancourt FH, Bolin JT, Eltis LD. The ins and outs of ring-cleaving dioxygenases. *Critical Reviews in Biochemistry and Molecular Biology*. 2006; 41: 241-67.
56. Tramontina R, Cairo JPLF, Liberato MV, Mandelli F, Sousa A, Santos S, et al. The *Coptotermes gestroi* aldo-keto reductase: a multipurpose enzyme for biorefinery applications. *Biotechnology for Biofuels*. 2017; 10: 4.
57. Tonon F, de Castro CP, Odier E. Nitrogen and carbon regulation of lignin peroxidase and enzymes of nitrogen metabolism in *Phanerochaete chrysosporium*. *Experimental Mycology*. 1990; 14: 243-54.
58. Filannino P, Gobetti M, De Angelis M, Di Cagno R. Hydroxycinnamic acids used as external acceptors of electrons: an energetic advantage for strictly heterofermentative lactic acid bacteria. *Applied and environmental microbiology*. 2014; 80: 7574-82.
59. Sutherland JB, Crawford DL, Pometto III AL. Metabolism of cinnamic, p-coumaric, and ferulic acids by *Streptomyces setonii*. *Canadian journal of microbiology*. 1983; 29: 1253-7.
60. Anzenbacher P, Anzenbacherova E. Cytochromes P450 and metabolism of xenobiotics. *Cellular and Molecular Life Sciences CMLS*. 2001; 58: 737-47.
61. Pawar PM-A, Koutaniemi S, Tenkanen M, Mellerowicz EJ. Acetylation of woody lignocellulose: significance and regulation. *Frontiers in plant science*. 2013; 4: 118.
62. Biely P. Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology advances*. 2012; 30: 1575-88.
63. Špániková S, Biely P. Glucuronoyl esterase—novel carbohydrate esterase produced by *Schizophyllum commune*. *FEBS letters*. 2006; 580: 4597-601.
64. Bunik VI. 2-Oxo acid dehydrogenase complexes in redox regulation. *European Journal of Biochemistry*. 2003; 270: 1036-42.
65. Jancova P, Anzenbacher P, Anzenbacherova E. Phase II drug metabolizing enzymes. *Biomedical Papers*. 2010; 154: 103-16.
66. Morisseau C, Hammock BD. Epoxide hydrolases: mechanisms, inhibitor designs, and biological roles. *Annu Rev Pharmacol Toxicol*. 2005; 45: 311-33.

67. Oesch F. Mammalian epoxide hydrases: inducible enzymes catalysing the inactivation of carcinogenic and cytotoxic metabolites derived from aromatic and olefinic compounds. *Xenobiotica*. 1973; 3: 305-40.
68. Kameshwar AKS, Qin W. Lignin Degrading Fungal Enzymes. *Production of Biofuels and Chemicals from Lignin*: Springer; 2016. p. 81-130.
69. Bugg TD, Ahmad M, Hardiman EM, Rahmanpour R. Pathways for degradation of lignin in bacteria and fungi. *Natural product reports*. 2011; 28: 1883-96.
70. Jensen KA, Evans KM, Kirk TK, Hammel KE. Biosynthetic pathway for veratryl alcohol in the ligninolytic fungus *Phanerochaete chrysosporium*. *Applied and environmental microbiology*. 1994; 60: 709-14.
71. Hyun MW, Yun YH, Kim JY, Kim SH. Fungal and plant phenylalanine ammonia-lyase. *Mycobiology*. 2011; 39: 257-65.
72. Chen C-L, Chang H-M, Kirk TK. Aromatic acids produced during degradation of lignin in spruce wood by *Phanerochaete chrysosporium*. *Holzforschung-International Journal of the Biology, Chemistry, Physics and Technology of Wood*. 1982; 36: 3-9.
73. Chen C-L, Chang H-M, Kirk TK. Carboxylic acids produced through oxidative cleavage of aromatic rings during degradation of lignin in spruce wood by *Phanerochaete chrysosporium*. *Journal of Wood Chemistry and Technology*. 1983; 3: 35-57.
74. Enoki A, Goldsby GP, Gold MH. Metabolism of the lignin model compounds veratrylglycerol- β -guaiacyl ether and 4-ethoxy-3-methoxyphenylglycerol- β -guaiacyl ether by *Phanerochaete chrysosporium*. *Archives of Microbiology*. 1980; 125: 227-31.
75. Enoki A, Gold MH. Degradation of the diarylpropane lignin model compound 1-(3', 4'-diethoxyphenyl)-1, 3-dihydroxy-2-(4"-methoxyphenyl)-propane and derivatives by the basidiomycete *Phanerochaete chrysosporium*. *Archives of Microbiology*. 1982; 132: 123-30.
76. Nakatsubo F, Kirk TK, Shimada M, Higuchi T. Metabolism of a phenylcoumaran substructure lignin model compound in ligninolytic cultures of *Phanerochaete chrysosporium*. *Archives of Microbiology*. 1981; 128: 416-20.
77. Del Río JC, Marques G, Rencoret J, Martínez ÁT, Gutiérrez A. Occurrence of naturally acetylated lignin units. *Journal of Agricultural and Food Chemistry*. 2007; 55: 5461-8.
78. Le Roy J, Huss B, Creach A, Hawkins S, Neutelings G. Glycosylation Is a Major Regulator of Phenylpropanoid Availability and Biological Activity in Plants. *Frontiers in Plant Science*. 2016; 7: 735.
79. Ko JK, Um Y, Park Y-C, Seo J-H, Kim KH. Compounds inhibiting the bioconversion of hydrothermally pretreated lignocellulose. *Applied microbiology and biotechnology*. 2015; 99: 4201-12.

Chapter-4

Analyzing *Phanerochaete chrysosporium* Gene Expression Patterns Controlling the Molecular Fate of Lignocellulose Degrading Enzymes

[This work has been published in "**Process Biochemistry**" 64 (2018): 51-62]

Ayyappa Kumar Sista Kameshwar and Wensheng Qin*

4.1. Abstract

The outstanding degrading abilities of *Phanerochaete chrysosporium* is solely dependent on its lignocellulolytic, aromatic compound degrading and detoxifying enzymes. However, the gene expression and protein turnover of lignocellulolytic enzymes are controlled at cellular level by various genes involved in information storage and processing KOG group. Understanding the gene expression patterns and mechanisms involved in regulation of lignocellulose degrading enzymes will significantly help in strain improvement and developing recombinant strains. To study the common expression patterns, we have retrieved *P. chrysosporium* gene expression datasets from NCBI GEO and analyzed using GeneSpring® software based on the genome wide KOG annotations retrieved from JGI-MycoCosm database. Statistically significant genes obtained from our analysis were separated into replication, repair and recombination, chromatin structure and dynamics, transcription factors, RNA processing and modification and translation, ribosomal structure and biogenesis processes. We have observed various genes encoding for DNA damage, repair and recombination, mRNA splicing, amidases, polyadenylate binding factors, heat shock, helix loop helix, HMG-box, CCAAT (HAP5), CRE-B transcription factors, histone acetyl transferases (MYST, SAGA) commonly expressed among the datasets of natural plant biomass growth substrates. Further studies must be conducted to understand the role and involvement of these significant genes in plant biomass degradation by *P. chrysosporium*.

Keywords: *Phanerochaete chrysosporium*, Information, storage and processing, Eukaryotic orthologous groups (KOG), Transcription factors, Gene expression datasets, NCBI GEO

4.2. Introduction

In the past few years, genome and transcriptome of *Phanerochaete chrysosporium* were intensively studied because of its efficient lignocellulose degrading abilities and availability of highly annotated genome. These studies have mostly delineated the genes and mechanisms responsible for the process of lignocellulose degradation [1-7]. However, mechanisms involved in cellular regulation of genetic material coding for lignocellulolytic enzymes were not clearly explained till today. Recent genomic studies conducted by various research groups and development of fungal genome repositories such as JGI (Joint Genome Institute) MycoCosm [8] and 1000 fungal genome project have revealed various significant facts. The eukaryote specific classification of KOG (clusters of orthologous groups) is significantly used for finding the ortholog and paralog proteins [8]. All the sequenced genomes deposited in JGI-MycoCosm database are provided with their respective KOG classification or KOG ID. The present day JGI sequencing protocol predicts and classifies the sequenced genome into four major classes a) Cellular processes and signaling b) Information storage and processing c) Metabolism d) poorly characterized.

The KOG group information, storage and processing is further classified into five groups as RNA processing and modification (KOG-ID: A), chromatin structure and dynamics (KOG-ID: B), translation, ribosomal structure and biogenesis KOG-ID: J), transcription (KOG-ID: K) and replication, recombination and repair (KOG-ID: L). Present day KOG classification of *P. chrysosporium* harbors 1713 gene models coding for Information storage and processing group which were further divided into 489 (RNA processing and modification), 201 (chromatin structure and dynamics), 366 (translation, ribosomal structure and biogenesis), 415 (transcription) and 242 (replication, recombination and repair) gene models respectively. Large number of gene models were mostly present in single copies with few numbers of gene models occur in multiple copies (Figure 4.1). Gene models involved in RNA processing and modification plays a crucial role in converting cellular genetic information from genes to proteins, thus determining the fate of cellular function and structure. However, RNA undergoes prior modifications and processing before performing the above functions. Majorly, RNA processing steps can be classified into three types a) trimming of RNA end segments resulting in a mature RNA form b) RNA splicing and c) sequence level modification of RNA segments.

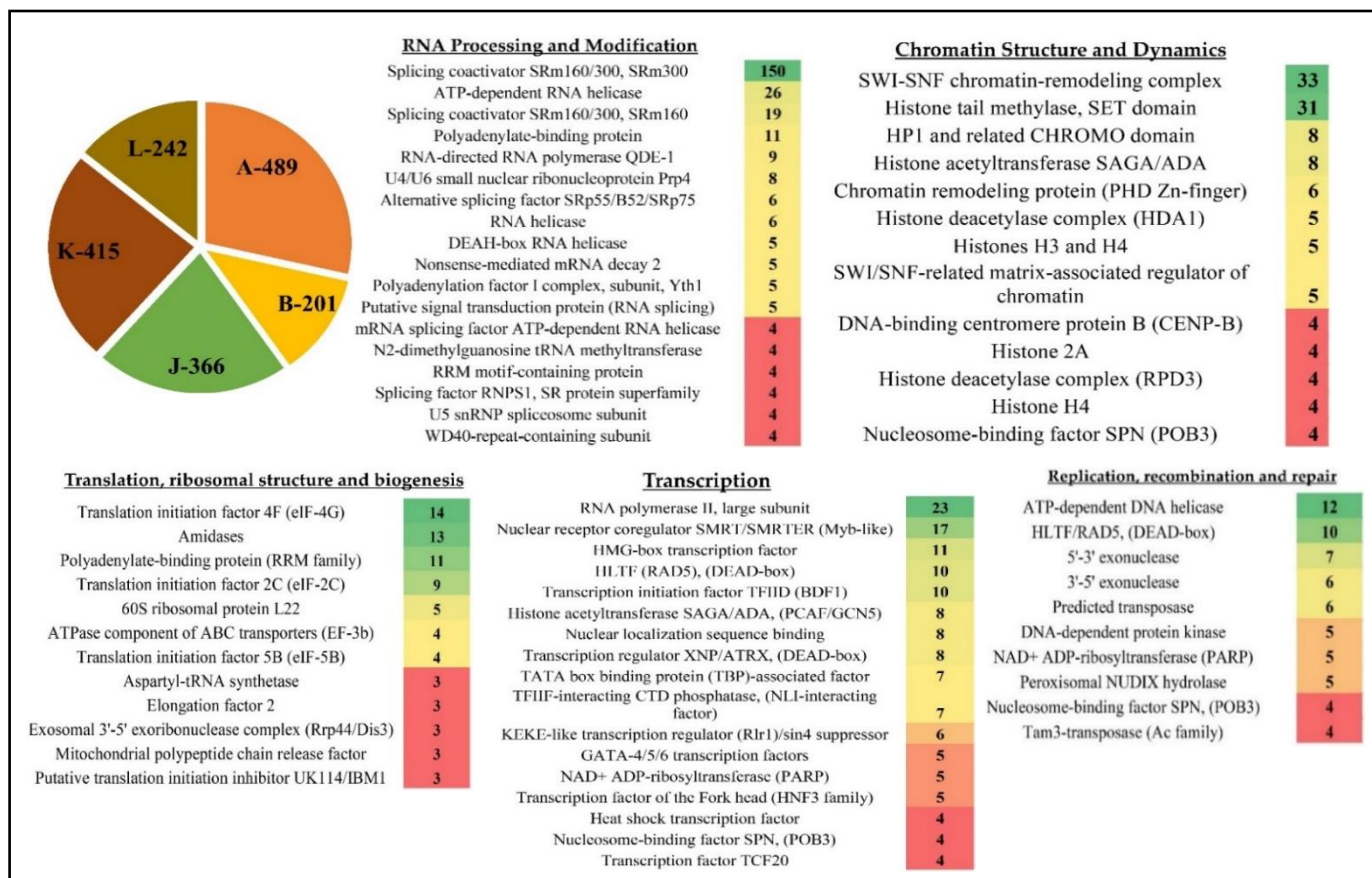


Figure 4.1: Pie diagram showing the distribution of gene models **A**- RNA processing and modification, **B**-chromatin structure and dynamics, **J**- translation, ribosomal structure and biogenesis, **K**- transcription and **L**- replication, recombination and repair. And heatmaps showing the gene models occurring more than 3 copies in the *P. chrysosporium* genome.

Gene regulation is an important physiological process and was extensively studied in fungi. The process of gene regulation assures the up and down regulation of genes based on the growth conditions. According to Richard et al (2014), in fungi, thirty-seven classes of gene regulators have been identified as zinc finger transcription factor proteins (C2H2 and binuclear zinc cluster protein (Zn2Cys6)), fungal specific transcription factors, bZIP, histone-like transcription, basic helix-loop-helix (HLH), heat shock factor (HSF), Myb-DNA binding, transcription enhancer factor (TEA) and GATA factors [9]. The production and secretion of extracellular fungal enzymes and their gene regulatory mechanisms especially cellulases and hemicellulases encoding genes were extensively studied in *Trichoderma* and *Aspergillus* species [10, 11]. These transcription factors are required for coordination various cellular

processes when cultured on different growth substrates. Occurrences of the gene sequences encoding for the above-mentioned transcription factors also might have been strongly related to the fungal diversity. *P. chrysosporium* genome contains the gene sequences coding for the above-mentioned crucial transcription factors.

The enormous genomic DNA of eukaryotes is well packed in the nucleus by the conserved histone proteins to form a complex ordered structure such as chromatin. After numerous stages of organization chromatin is further organized into chromosomes which allows precise cellular divisions [12]. Chromatin dynamics in eukaryotes can be introduced majorly by three ways a) ATP-dependent remodeling which might also lead to the interchanges in primary structures of histones [12-14] b) histones are also subjected to post translational modifications resulting in structural and functional outcomes [12, 15] and c) DNA is also subjected to methylations [12, 16]. All three processes chromatin modifications, chromatin remodeling and DNA methylation were found to be strongly interdependent. Steinfeld et al (2007), have revealed the crucial role of chromatin modifiers in transcriptional regulation of common yeast, *Saccharomyces cerevisiae* [17].

Gene models classified under translation, ribosomal structure and biogenesis are required for the accurate conversion of genetic code to proteins. Fungal ribosomes (cellular nanomachines) majorly comprises two ribonucleoprotein subunits 40S (in turn contains 33 ribosomal proteins and a 18S ribosomal rRNA) and 60S subunits (consists of three ribosomal rRNA 25S, 5.8S, 5S and 46 ribosomal proteins) [18]. *P. chrysosporium* genome codes for 254 genes involved in ribosome structure, biogenesis and translation processes. DNA replication, repair and recombination events are complex and basic molecular processes in the living organisms. Process of DNA replication in eukaryotes is one of the highly studied molecular process and various genes and proteins involved in this key process have been clearly explained. The genome of *P. chrysosporium* harbors 155 unique gene models which includes various genes involved in the process of DNA replication, repair and recombination. The proteome, transcriptome and secretome studies conducted in the past have revealed the involvement of various oxidases, hydrolytic enzymes in degradation of plant biomass, however additional research must be conducted to reveal the regulation of the lignocellulolytic enzymes at molecular level [2, 3, 7, 19].

In our present study, we have analyzed the gene expression patterns of *P. chrysosporium* to understand the common and significant expression of genes involved in RNA

processing and modification, chromatin structure and dynamics, translation, ribosomal structure and biogenesis, transcription and replication, recombination and repair processes among *P. chrysosporium* gene expression datasets. To the best of our knowledge, this is the first comprehensive report on *P. chrysosporium* genes involved in information, storage and processing processes and their expression patterns.

4.3. Methods

4.3.1. Data Retrieval

The gene expression datasets of *P. chrysosporium* cultured on different growth substrates, were retrieved from NCBI-GEO (Gene expression omnibus database) (<https://www.ncbi.nlm.nih.gov/geo/>) using the term *P. chrysosporium*. Details about these gene expression datasets analyzed were reported in Table 4.1. All the relevant experimental metadata corresponding to the gene expression datasets were retrieved using the “Accession display” option of NCBI GEO website and from the corresponding literature available. In NCBI GEO database, there are 8 gene expression datasets which are specifically studied on *P. chrysosporium*, out of which we have selected 7 (GSE14734, GSE14735, GSE27941, GSE54542, GSE52922, GSE69008 and GSE69461) gene expression datasets for our current analysis (Table 4.1). These datasets were considered mainly because of the varied substrate conditions (from simple synthetic growth medium to complex plant biomass medium) used for the culture of *P. chrysosporium*. The experimental metadata accession ID's, gene expression platform details substrate used for the growth of *P. chrysosporium* and sample information were reported in Table 4.1.

Table 4.1: List of the *P. chrysosporium* gene expression datasets retrieved from NCBI GEO repository:

GEO- ID	Platform and Technology	Substrate	#Samples	Ref
GSE54542	NimbleGen <i>P. chrysosporium</i> arrays	Oak acetonc extractives	6	[20]
GSE27941	NimbleGen <i>P. chrysosporium</i> arrays	Ball milled aspen, Ball milled pine	6	[21]
GSE52922	NimbleGen <i>P. chrysosporium</i> arrays	P717 hybrid line, Transgenic line 82 Transgenic line 64	9	[22]

GSE14734	NimbleGen <i>P. chrysosporium</i> arrays	Cellulose, Glucose, Ball milled aspen	9	[3, 4]
GSE14735	NimbleGen <i>P. chrysosporium</i> arrays	Replete, Carbon-limited Nitrogen-limited	9	[3, 4]
GSE69008	NimbleGen <i>P. chrysosporium</i> arrays	Poplar wood substrates	24	[1]
GSE69461	Illumina HiSeq 2000	<i>Picea glauca</i> (spruce sapwood)	18	[23]

4.3.2. Data Analysis

All the datasets were analyzed using GeneSpring v14.8 (<http://genespring-support.com/get-gs>) software. The expression experiments were created using generic single-color workflow, by creating a prior generic single-color technology with the available supplementary information. Gene expression datasets with accession IDs GSE14734, GSE14735, GSE27941, GSE54542, GSE52922, GSE69008 were log transformed with 2 log-base, normalized using 75th percentile normalization, with a threshold value of raw signals set to 1.0 and baseline transformation to median of all samples. The experimental details of all the samples were retrieved from the corresponding literature and experimental metadata and later this data was used for grouping the samples. The probe sets were filtered using the option “Filter probe sets by expression” by selecting the raw data and filtered using the filter by percentile option (upper and lower percentiles set to 100 and 20 respectively).

Differentially expressed significant genes were obtained using the “Statistical analysis” option and based on the experimental grouping and created interpretations T-test or one-way Anova was performed. However, for GSE69461 dataset, we have not performed any of the pre-processing steps such as normalization, log transformation or thresholding, as we have retrieved RPKM data for the individual samples. Based on the sample grouping T-test was performed with asymptotic p-value computation and Benjamini-Hochberg False discovery rate for multiple testing correction. Differentially expressed significant list of genes were retrieved from all the datasets and compared using Venny and Jvonn online softwares for obtaining common gene lists. We have retrieved the list of gene level annotations of *P. chrysosporium* RP78v2.2 encoding for information, storage and processing (KOG Group) which includes RNA processing and modification, chromatin structure and dynamics,

translation, ribosomal structure and biogenesis, transcription and replication, recombination and repair processes, from JGI MycoCosm database.

4.4. Results

The extrinsic plant biomass degrading properties of *P. chrysosporium* is majorly credited to its lignocellulolytic CAZymes and wide range of aromatic compound degrading and detoxifying enzymes [6, 24]. The expression of these enzymes is majorly regulated by a wide range of enzymes belonging to the information storage and processing group. Statistical analysis of GSE14734, GSE14735, GSE27941, GSE54542, GSE52922, GSE69008 and GSE69461 gene expression dataset has resulted in 691, 583, 146, 320, 275, 865 and 1235 unique differentially expressed genes belonging to information storage and processing KOG group (Table 4.2). Several genes encoding for various lignocellulolytic CAZymes, aromatic compound degrading and large array of detoxifying enzymes were also found to be highly upregulated which were extensively discussed in our previous works [6, 24].

Table 4.2: The total number of differentially expressed and statistically significant genes (unique) belonging to information storage and processing group among the selected datasets:

Unique genes after statistical analysis							(Log 2) Fold change ≥ 2.0					
Dataset	A	B	J	K	L	Total	A	B	J	K	L	Total
GSE14734	172	90	145	164	120	691	3	0	2	3	0	8
GSE14735	129	73	192	114	75	583	0	2	41	8	2	53
GSE27941	25	12	53	36	20	146	0	1	0	2	0	3
GSE54542	99	36	50	90	45	320	99	36	50	90	45	320
GSE52922	87	20	69	75	24	275	17	8	15	16	2	58
GSE69008	213	91	205	204	152	865	36	14	67	51	13	181
GSE69461	346	146	261	302	180	1235	310	127	248	274	160	1119

(Note: **A:** RNA processing and modification, **B:** chromatin structure and dynamics, **J:** translation, ribosomal structure and biogenesis, **K:** transcription and **L:** replication, recombination and repair)

We have observed a total of 8, 53, 3, 320, 58, 181 and 1119 unique genes encoding for information, storage and processing group were highly up regulated with higher fold change values ($\log_2 FC > 2.0$) among GSE14734, GSE14735, GSE27941, GSE54542, GSE52922, GSE69008 and GSE69461 datasets respectively (Table 4.2). The violin plots (sample and group level data on x-axis and normalized intensity values on y-axis) showing the distribution of normalized and baseline transformed samples, which briefly represents the differential expression of the normalized genes of all the samples at dataset level were shown (Figure 4.2). Similarly, the volcano plots (\log_2 fold change values on x-axis and $-\log_{10}$ corrected P-values on y-axis) and profile plots (experimental condition information on x-axis and normalized intensity values on y-axis) briefly representing the differentially expressed genes (with higher fold change values) among individual experimental conditions were shown (Figure 4.3).

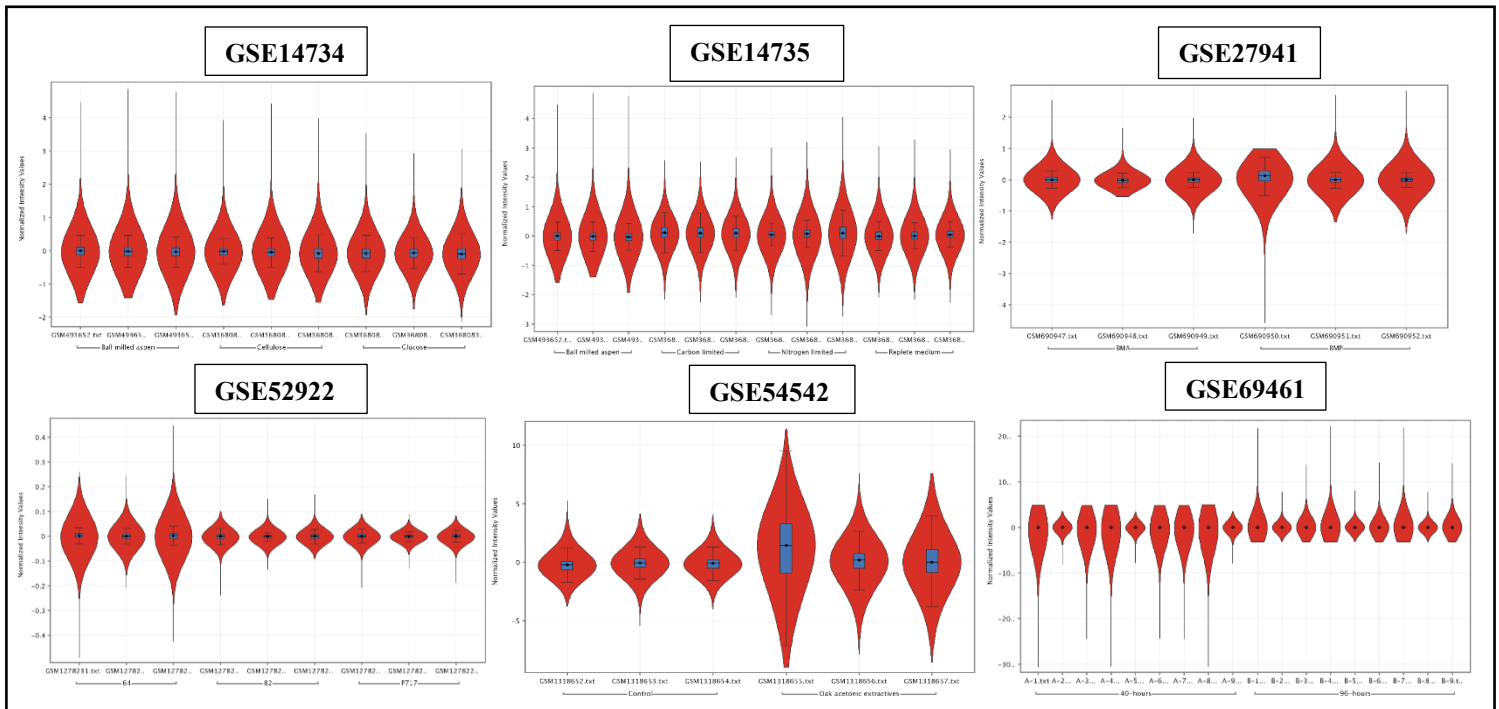


Figure 4.2: Violin plots for the selected datasets GSE14734, GSE14735, GSE27941, GSE52922, GSE54542 (normalized and baselined) and GSE69461 (only baselined) briefly showing the distribution of samples

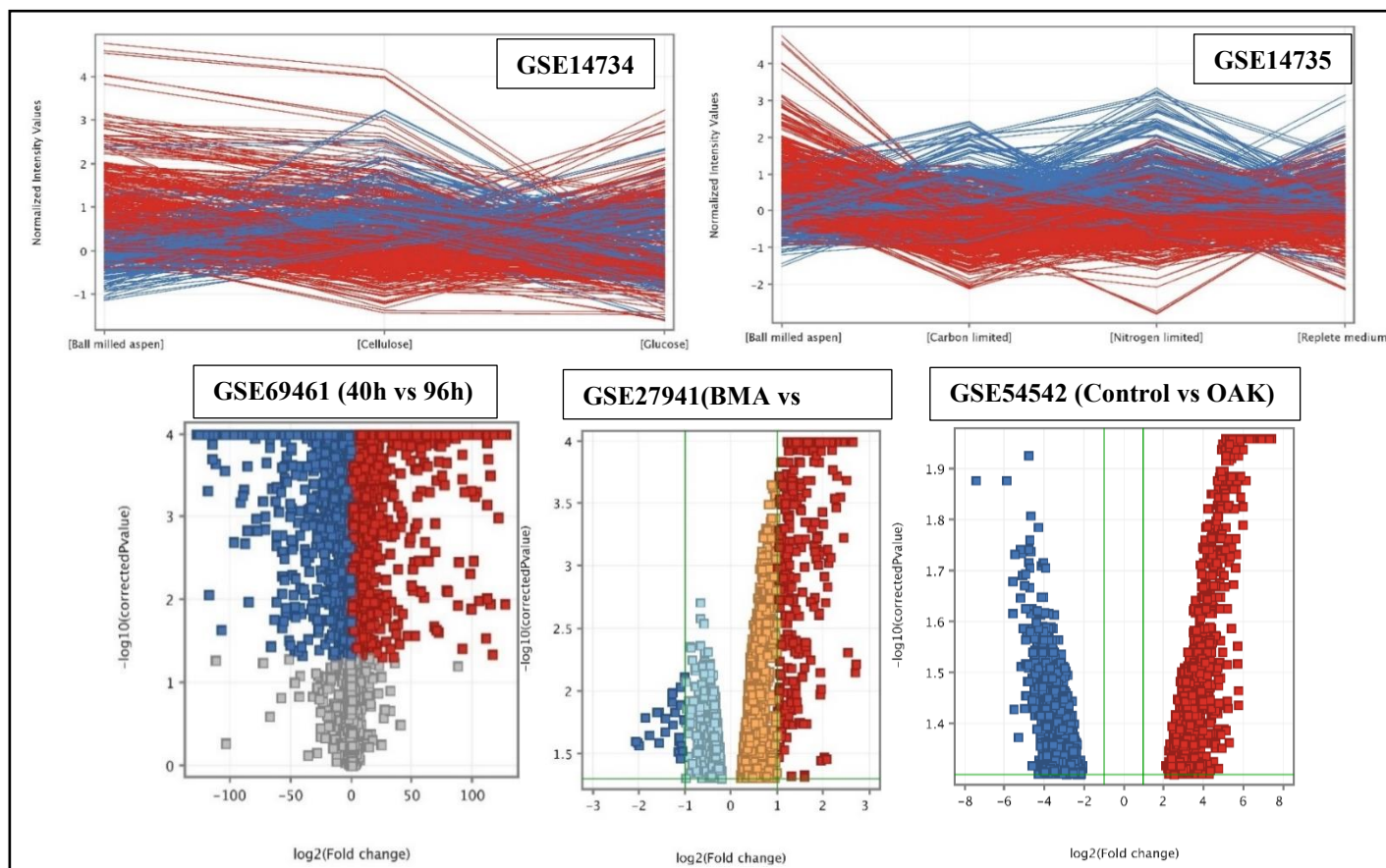


Figure 4.3: Profile plots (GSE14734 and GSE14735) and volcano plots (GSE27941, GSE54542 and GSE69461) of the significant and differentially expressed genes with fold change cut-off >2.0 among the selected datasets.

4.4.1. RNA Processing and Modification (KOG ID: A)

The eukaryotic mRNA is substantially processed through 5' capping, splicing and 3' end processing before it gets exported, thus, these processes play a crucial role in determining the fate of transcripts [25]. Several studies conducted in the past have clearly explained that a wide range of enzymes coordinately function together to enhance the regulation information which affects in transcript export, localization and stability [25]. As mentioned earlier genome of *P. chrysosporium* encodes 489 genes involved in the RNA processing and modification class. Majorly, mRNA guanylyl transferases, methyltransferases are involved in 5' capping which happens immediately after early transcription by RNA polymerase II [26]. The 3' end processing involves mRNA cleavage and polyadenylation factors, polyadenylate polymerase, polyadenylate binding protein-II, polyadenylation factor complex, mRNA cleavage factor I

and II [27, 28]. Spliceosomes performs the complex splicing reactions on pre-mRNA in eukaryotes, a large set of proteins along with U1 to U6 small nuclear RNPs (snRNPs) [25, 27, 29]. In *P. chrysosporium* genome 19, 36 and 237 gene models were found to be involved in 5'end capping, 3' end processing and mRNA splicing processes.

Genes encoding for ATP dependent helicases, DEAD-box superfamily, RNA binding proteins, mRNA guanylyl and methyl transferases (5'end capping) and mRNA cleavage and polyadenylation factor, polyadenylation binding protein and other associated enzymes were found to be statistically significant and commonly expressed among all the gene expression datasets. However, in dataset GSE14734 and GSE14735 genes encoding for ATP-dependent RNA helicase (Phch1_5340), Fibrillar and related nucleolar RNA-binding (Phch1_610), splicing factor hnRNP-F and related RNA-binding proteins (Phch1_4943) and ribosomal protein RPL1 (Phch1_128104) were found to be highly upregulated in cellulose, glucose, replete and nitrogen limited mediums and down regulated in ball milled aspen and carbon limited growth mediums. Whereas in GSE27941 dataset few genes encoding for RNA processing and modification were found to be statistically significant but were not highly expressed. In GSE54542, GSE52922, GSE69008 and GSE69461 gene expression datasets a large set of genes encoding for RNA processing and modification processes were found to be highly expressed. Total of 46 genes encoding for RNA processing and modification were found to be common among GSE69008 and GSE69461 datasets. In GSE54542 dataset 99 genes coding for RNA processing and modification were differentially expressed, transcripts encoding for spliceosomal proteins, U1, U4, U5 snRNP, WD40 repeats, ribosomal proteins (RPL1 and RPL2) and mRNA export factors and SWAP mRNA splicing regulator were found to be up regulated (with fold change ≥ 2.0) in oak acetic extractive samples and down regulated in control samples. In GSE52922 dataset, 17 genes (ATP dependent RNA helicase, decapping enzyme complex, mRNA cleavage and polyadenylation factor and splicing coactivator and other splicing factors) were up regulated in 82 and 64 growth substrates. In GSE69008 dataset genes encoding for various RNA processing and modification processes were down regulated in A (high lignin and low glucose) growth substrates, were as found to be differentially up regulated in B (low lignin and high glucose) samples. Especially transcripts encoding for alternative splicing factor (SRp55/SRp75), mRNA cleavage and polyadenylation factor, mRNA deadenylase, RNA binding protein (p54nrb) up regulated in C (average lignin and average glucose) substrates. Finally, in GSE69461 dataset total of 310 genes encoding for RNA processing and modification were found to be expressed with fold

change ≥ 2.0 . We have found that 199 genes were highly up regulated in 96 hours and 111 genes were highly upregulated in 40 hours growth substrates. Out of which 40 genes encoding for DEAD-box family, decapping enzyme complex, histone H3 (Lys4) and mRNA methyl transferase, mRNA cleavage and polyadenylation factors, polyadenylating polymerase, RNA binding proteins, spliceosomes and splicing factors U1 to U6 snRNP were highly up regulated in 40 hours growth substrate (Figure 4.4).

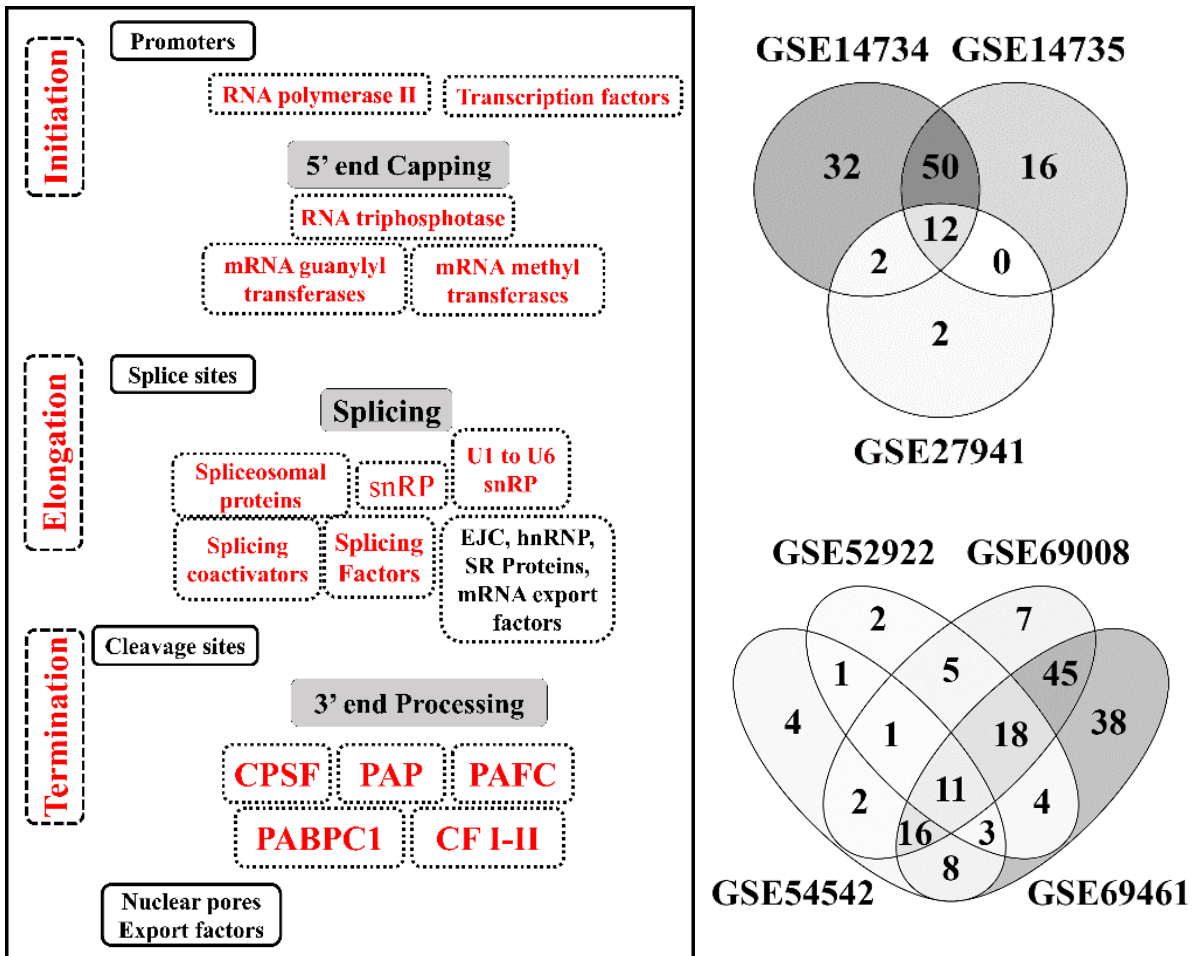


Figure 4.4: List of commonly expressed and statistically significant genes among the selected gene expression datasets encoding for RNA processing and Modification group (KOG functional ID):

4.4.2. Chromatin Structure and Dynamics (KOG ID: B)

It is a well-known fact that vast eukaryotic genomes are efficiently packaged in to chromatin and further into stable chromosomes. Several state-of-the-art reviews were already available

on eukaryotic chromatin structure and dynamics and few were listed here [30-35]. As discussed, earlier genome of *P. chrysosporium* codes 201 genes classified under chromatin structure and dynamics. Genes encoding for chromatin remodeling protein (PHD-Zn finger), chromosome condensation complex, histone acetyltransferase, and SWI-SNF chromatin remodeling complex proteins were found to be statistically significant and commonly expressed among GSE14734, GSE14735 and GSE27941 datasets. And in GSE54542, GSE52922, GSE69008 and GSE69012 datasets genes encoding for DNA binding centromere protein B, heterochromatin associated protein (HP1), histone acetyltransferase (SAGA) and type b catalytic unit, histones H3 and H4 and SWI-SNF chromatin remodeling complex proteins were significantly common. In GSE14734, GSE14735 and GSE2794 datasets total of 90, 73 and 12 genes were found to be statistically significant with a p-value <0.05 after multiple testing correction. Total of 57 transcripts were found to be common between GSE14734 and GSE14735 datasets which majorly contain histone proteins (H1, H2A, H2B, H3, H4, H5), histone methyltransferases, histone acetyl transferases (MYST, SAGA/ADA, PCAF/SAGA), sirutin 5, structural maintenance of chromosome, SWI-SNF remodeling protein encoding genes. Chromatin remodeling complexes, heterochromatin associated protein, nucleosome assembly protein, histone deacetylase complexes, histone tail methylase, SWI-SNF (Snf5 subunit) chromatin remodeling, telomerase catalytic unit were up regulated in ball milled aspen samples. Only genes encoding for DNA-binding centromere protein B (Phchr1_5151) (in GSE14735), SWI-SNF chromatin remodeling complex protein (Phchr1_5319) (both in GSE14735 and GSE27941 datasets) were highly upregulated in ball milled aspen samples with fold change ≥ 2.0 . However, no genes involved in chromatin structure and dynamics were found to be expressed with fold change <2.0 (not very highly expressed).

In GSE54542 dataset, chromatin remodeling complex (Phchr_7694), heterochromatin associated protein (Phchr_1386), histone acetyltransferase (MYST) (Phchr_1138, 42216), histone acetyltransferase (SAGA) (Phchr_932), histone H3 methyltransferase complex (Phchr_136167), histone tail methylase (Phchr_136398) and telomerase length regulating protein kinase (Phchr_2237) were highly up regulated with fold change ≥ 2.0 in control samples. At the same time genes encoding for chromatin remodeling complex (Phchr_130373, 6773, 7315), DNA binding centromere (Phchr_5151), histone acetyl transferase (SAGA/ADA) (Phchr_2529, 3745 and 29427), histone deacetylase (Phchr_5894), histone H3 methyl transferase (Phchr_820, 3326), histone tail methylase (Phchr_5894) and SWI-SNF chromatin remodeling complex proteins were highly up regulated in oak acetic

extract samples. In GSE52922 dataset, genes encoding for DNA binding centromere protein, histone proteins (H2A, H2B, H3, H4) and SWI-SNF chromatin remodeling complex protein were highly upregulated in P717 growth samples and down regulated in 64 and 82 growth substrates. Similarly, in dataset GSE69008, genes encoding for heterochromatin associated protein (HP1), histone acetyl transferase (SAGA/ADA) and sirtuin 5 (SIR2 family) proteins were highly expressed in A (high lignin-low glucose) proteins. In B (low lignin -high glucose) growth substrate genes encoding for chromatin remodeling complex, histone acetyl transferase, histone (H3 and H4), nucleosome binding factor, sirtuin 5 (SIR2) and SWI-SNF chromatin remodeling complex proteins were highly up regulated. In dataset GSE69461, genes encoding for chromatin assembly factor, cell cycle regulated histone H1, chromatin remodeling complex, chromosome condensation complex, DNA topoisomerase, DNA binding centromere binding, heterochromatin associated protein, histone proteins (H1, H2A, H2B, H3, H4), histone acetyl transferases (SAGA/ADA, MYST), histone deacetylase complex (HDA1, RPD3, SIN3), nucleosome assembly and remodeling factors, histone tail methylase, sirtuin 4, 5 , chromosome structural maintenance protein and SWI-SNF chromatin remodeling complex were highly up regulated in 96-hour growth substrates (Figure 4.5).

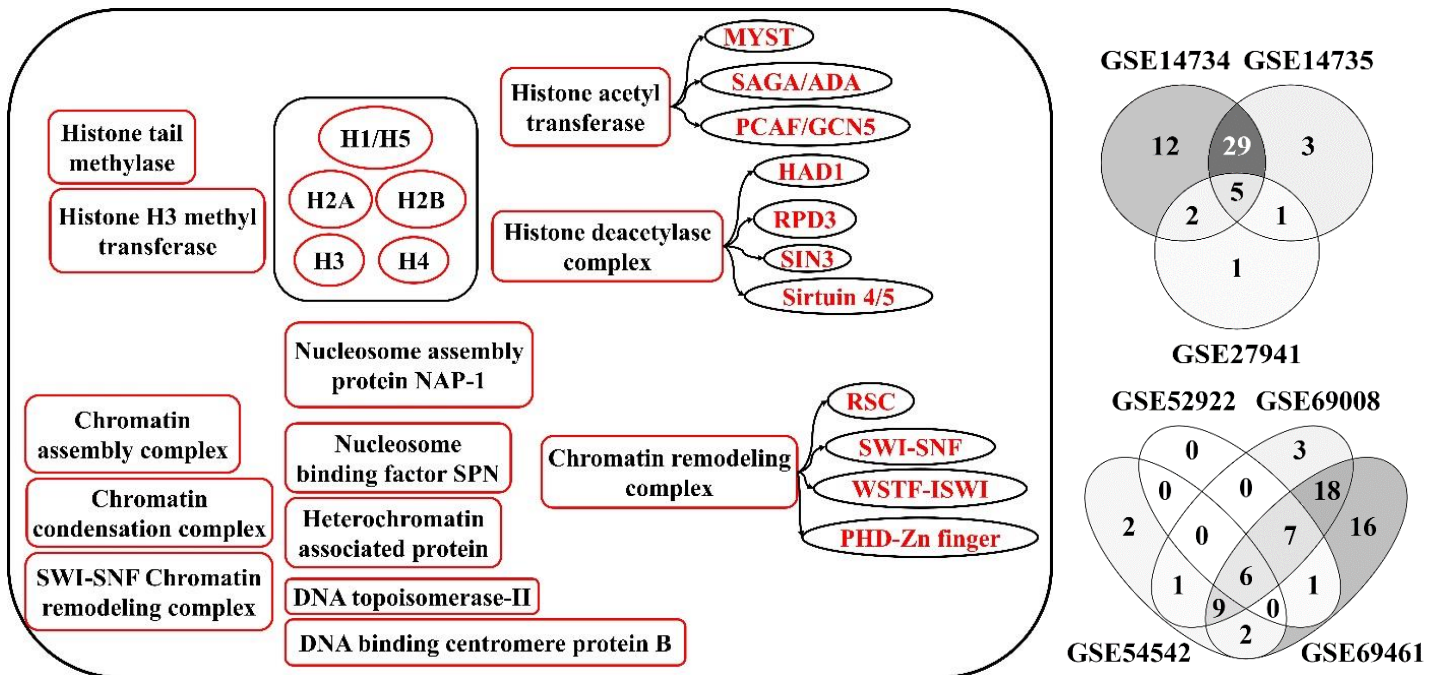


Figure 4.5: List of commonly expressed and statistically significant genes among various datasets encoding for chromatin structure and dynamics group (KOG functional ID: B)

4.4.3. Translation Ribosomal Structure and Biogenesis (KOG ID: J)

Translation is the central and crucial process for the conversion of nucleotide sequence to cellular expressing units. The process of translation is majorly dependent on ribosomes, which are highly conserved cellular nanomachines. Biogenesis of ribosomes happens in nucleolus of eukaryotic cells, where it synthesizes rRNA molecules (precursors) through DNA directed RNA polymerase I and III, these nascent rRNA molecules are further subjected to a set of RNA cleavage reactions and other chemical modifications [18, 36]. A large set of ribosome biogenesis proteins are required for the proper rRNA folding and enzymatic processing to produce active ribosomal subunits [18, 36]. Thus, a serious coordination of all the DNA dependent RNA polymerases is involved in synthesizing structurally active ribosomal subunits [36]. As mentioned above genome of *P. chrysosporium* encodes for 366 gene models classified to be involved in translation, ribosomal structure and biogenesis. Our present metadata analysis has revealed that total of 15 and 8 commonly expressed statistically significant genes among GSE14734, GSE14735, GSE27941 and GSE54542, GSE52922, GSE69008 and GSE69461 datasets, with various other genes found to be common between the datasets (Figure 4.3). Genes encoding for 60S acidic ribosomal protein P2 was up regulated in glucose substrate and translation initiation factor 4F (eIF-4G) was highly upregulated in ball milled aspen and cellulose with fold change ≥ 2.0 . In GSE14735 dataset, various genes encoding for translation ribosomal structure and biogenesis were significantly expressed with fold change values ≥ 2.0 . Genes encoding for 40S ribosomal proteins (S2, S3, S3A, S4, S6, S8, SA, S12, S15, S26, SA/P40), 60S acidic ribosomal (P0, P2), 60S ribosomal protein (L3, L5, L6, L7, L7A, L10A, L11, L13, L14, L15/27, L18, L22, L24, L39), ABC transporters with ATPase domain, glutamyl tRNA synthetase, translation initiation factor (eIF-2B, eIF-4A and EF-1), ubiquitin/40S ribosomal protein S27A and ribosomal proteins S18 and S7 were highly up regulated in replete and ball milled aspen growth substrates. Several transcripts encoding for amidases were found to be highly up regulated in nitrogen limited followed by carbon-limited growth substrates, no genes were found to be expressed above fold change ≥ 2.0 in GSE27941 dataset.

In GSE54542 dataset, most of the genes involved in translation, ribosomal structure and biogenesis were found to be down regulated in control samples with genes encoding for amidases (Phchr_3719 and 6738), mitotic cell division protein, mitochondrial ribosomal protein L16, NMD protein affecting stability, RNaseP, translation initiation factor 6, tRNA

splicing endonuclease and tRNA methyl transferase (GCD10). However, in oak acetic extractives samples genes encoding for 40S (S14, S26, SA) 60S acidic protein P1, 60S (L13, L22) ribosomal proteins, amidases, tRNA synthetase (asparaginyl, aspartyl, glutamyl, glycyl, leucyl, lysyl), mitochondrial ribosomal proteins (L28, S14/S29, polyadenylate binding protein, RNA binding protein, translation initiation inhibitor, translation initiation factor (eIF-2c, eIF-3, eIF4E, eIF-4G, eIF-6), translation repressor MUT5/PUF4 and tRNA methyl transferases were highly up regulated with fold change ≥ 2.0 . In dataset GSE52922, genes encoding for 40S (S10, S3A, S7, SA), 60S (L15/L27, L3, L30, L39) ribosomal proteins, S4 and S7 ribosomal proteins were comparatively highly expressed in P717 growth substrate. Amidase, ABC transporters with ATPase domain, polyadenylate binding protein, translation initiation factor 4F and translation repressor MPT5/PUF4 were comparatively up regulated in 82 and 64 growth substrates (Figure 4.3).

In GSE69008 dataset, most of the translation, ribosomal structure and biogenesis proteins were comparatively down regulated in A (high lignin-low glucose) and highly upregulated in C (average lignin-average glucose) samples. Genes encoding for 40S (S2, S3, S3A, S4, S6, S8, S15, S15/22, S23), 60S (L2/L8, L3, L5, L6, L7, L10, L10A, L13, L14, L15) and 60S acidic ribosomal protein P0, amidases, tRNA synthetase (asparaginyl, methionyl, threonyl), polyadenylate binding protein, ATP-dependent RNA helicase, translation initiation factor (EF-1, eIF-1/SUI1, eIF-2 alpha, beta, gamma, eIF-4G, eIF-5A and ubiquitin/40S ribosomal protein S27a) were highly up regulated in C growth substrates. Although genes encoding for 40S (S2/S26, SA/P40), 60S (L15/L27, L19, L22, L44), 60S acidic ribosomal protein P2, tRNA synthetase (alanyl, aspartyl, glycyl, prolyl), ABC transporter with ATPase domain, Exosomal 3'-5' exoribonuclease complex, RNA binding protein (translation regulation), translation initiation factor (eIF-2, eIF-3a, EF-1, eIF-2C, eIF-3C, 3D, eIF-4F) and tRNA isopentenyl pyrophosphate transferase were significantly up regulated in B (low lignin- high glucose) samples (Figure 4.3). Finally, in GSE69461 dataset a large set of genes encoding for translation, ribosomal structure and biogenesis were highly down regulated (with fold change ≥ 2.0) in 40-hour samples. Several transcripts encoding for 40S and 60S ribosomal proteins, amidases, tRNA synthetases (asparaginyl, aspartyl, glutaminyl, glutamyl, isoleucyl, leucyl, lysyl, methionyl, phenylalanyl, tyrosyl), translation initiation factors, tRNA methyltransferases, ribosomal proteins, mitochondrial ribosomal proteins, ABC transporters, Exosomal 3'-5' exoribonuclease complex and WD40 nucleolar protein (translation) were highly up regulated in 96-hour growth substrate (fold change ≥ 2.0). In 40-hours growth

substrates genes encoding for 40S (S2), 60S (P2, L10, L11, L30) ribosomal proteins, amidases, aspartyl, asparaginyl and cysteinyl tRNA synthetases and various translation initiation factors were found to be highly up regulated (Figure 4.6).

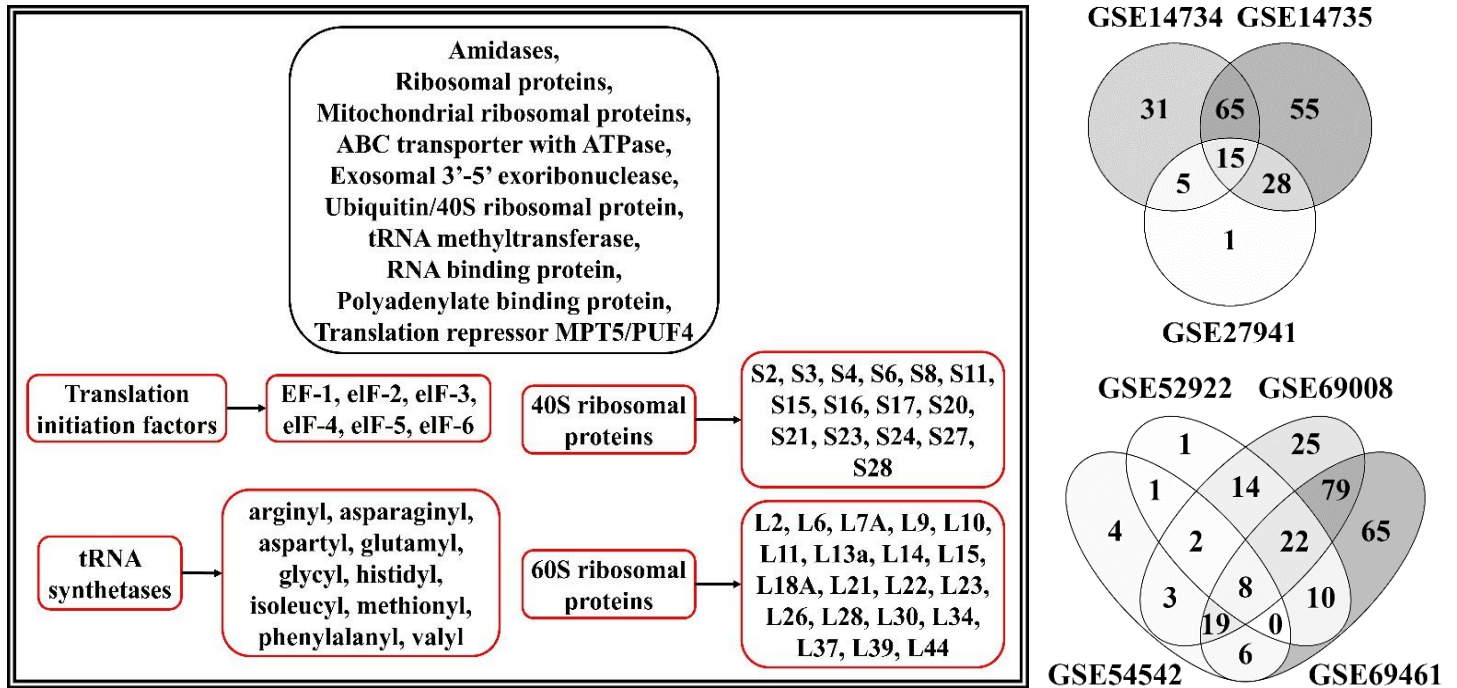


Figure 4.6: List of commonly expressed and statistically significant genes among various datasets encoding for translation, ribosomal structure and biogenesis group (KOG functional ID: J)

4.4.4. Transcription (KOG ID: K)

Previous studies have extensively determined that a wide range of transcription factors are involved in regulating the expression of lignocellulolytic enzymes [9, 37-39]. Genome of *P. chrysosporium* encodes 415 gene models classified under the transcription group. Our present metadata analysis has revealed a large set of commonly expressed statistically significant genes coding for the essential transcription factors. We have observed that a total of 14 and 21 genes were found to be common among GSE14734, GSE14735, GSE27941 and GSE54542, GSE52922, GSE69008, GSE69461 datasets respectively. In GSE14734 dataset, genes encoding for CREB/ATF transcription factor, TATA box binding protein and transcription coactivator encoding genes were highly down regulated in both ball milled aspen and cellulose growth substrates, while comparatively highly expressed in

glucose substrates. In dataset GSE14735, genes encoding for TATA box binding protein, HMG box transcription factors were found to be highly up regulated in carbon limited, ball milled aspen and nitrogen limited growth substrates, while genes encoding for GATA-4/5/6 transcription factors and elongation factor 1 were highly upregulated in cellulose containing growth substrates. And genes encoding for elongation factor 1 beta, protein arginine N-methyl transferases (PRMT1) related enzymes, RNA polymerase II were highly up regulated in replete growth substrates. In GSE27941 dataset, only genes encoding for TATA binding protein (RNA polymerase II) and component of TFIID and TFIIB were found to be highly up regulated with fold change ≥ 2.0 in ball milled aspen samples, however genes encoding for a large set of transcription factors especially CREB/ATF, CCAAT (HAP5), class transcription repressor, HMG box, heat shock transcription factor were found to be statistically significant and highly up regulated in ball milled aspen with fold change ≥ 1.5 (Figure 4.3).

Interestingly we have observed that in dataset GSE54542, a large set of genes encoding for various transcription factors were highly up regulated in oak acetonic extractive than control samples. Transcripts encoding for CAAT-binding factor, CREB/P300, CREB/ATF, calcium responsive transcription coactivator, chromatin remodeling complex, GATA-4/5/6, HMG-box, heat shock (HSF), helix loop helix (HLH), KEKE containing transcription regulator, Mlx related, nuclear receptor coregulator, transcription factor DATF1, RFX, TATA- binding factor, forkhead/HNF3, CCR4, upstream (L-myc-2) transcription factors were highly up regulated in oak acetonic extractive samples. However, genes encoding for CCAAT (HAP2), Cdk activating kinase, HMG box, leucine permease transcription regulator, TBP associated transcription factor, transcription factors (MEIS1, TCF20, Myb, CCAAT displacement CDP1), transcription repressor, Zn-finger transcription factors were highly up regulated in control samples with fold change ≥ 2.0 . In GSE52922 dataset, genes encoding for GATA-4/5/6, helix-loop-helix (HLH), HMG-box, NAD⁺ADP-ribosyltransferase Parp, RFX, RNA polymerase II, TATA box binding protein, TFIIF interacting CTD phosphatase, transcription coactivator, upstream (L-myc-2) transcription factors and ubinuclein nuclear proteins were highly up regulated in 82 growth samples.

In GSE69008 dataset we have observed that genes classified under transcription were highly up regulated in B (low lignin-high glucose) followed by C (average lignin-average glucose) samples. Genes encoding for calcium responsive transcription coactivator, casein kinase II, cell cycle control protein, chromodomain helicase, dosage compensation

regulatory complex, E3 ubiquitin ligase, GATA-4/5/6, glucose repressible alcohol dehydrogenase transcriptional effector (CCR4), heat shock (HSF), helix-loop-helix (HLH), transcription factors, negative regulation of transcription, nuclear receptor coregulator, transcription factors 5qNCA, RFX, OCT1, transcription coactivator(FOSB/c-Fos) and transcription initiation factor (TFIIB, TFIID) were comparatively highly up regulated in B (low lignin-high glucose). While genes encoding for HMG-box, KEKE-transcription regulatory, NAD⁺ADP-ribosyltransferase Parp, elongation factor 1, nuclear localization sequence binding protein, ubinuclein nuclear protein and upstream L-myc-2 transcription factors were down regulated in B samples. Total of 177 transcripts were highly regulated when cultured on *P. glauca* wood species of GSE69461 dataset, most of the genes encoding for transcription were found to be highly down regulated in 40-hours samples and highly up regulated in 96-hour samples. Especially genes coding for transcription factors bZIP, CCAAT (HAP5), CCR4 (NOT5), GATA-4/5/6, leucine zipper transcription factors, Glucose-repressible alcohol dehydrogenase transcriptional effector, HSF, HLTF/DNA helicase, HMG-box, MADS-box, MOT2, RFX, TATA-box binding protein, BLIMP-1, CA150, E2F, MBF1, MEIS1, NERF, OCT-1, HNF3-forkhead, PHOX2, TCF20, Myb, CCAAT (CDP1) transcription factors and nuclear receptor coregulator (SMRT, Myb domains) were highly up regulated in 96-hour samples. Simultaneously, wide range of transcripts encoding for alpha-1,2-glucosyltransferase, CCR4 (NOT5), cell cycle control protein, CREB/ATF, GATA-4/5/6, glucosyltransferase-Alg8p, HSF, HLH, HMG-box, homeobox, CCR4 associated factor, 5qNCA, FET5, RFX, TATA-box binding protein, MES1, NERF, NF-X1, HNF3-forkhead, PRD, TCF20, Myb, L-myc-2 transcription factors were highly up regulated in 40-hour samples with fold change ≥ 2.0 (Figure 4.3) (Figure 4.7).

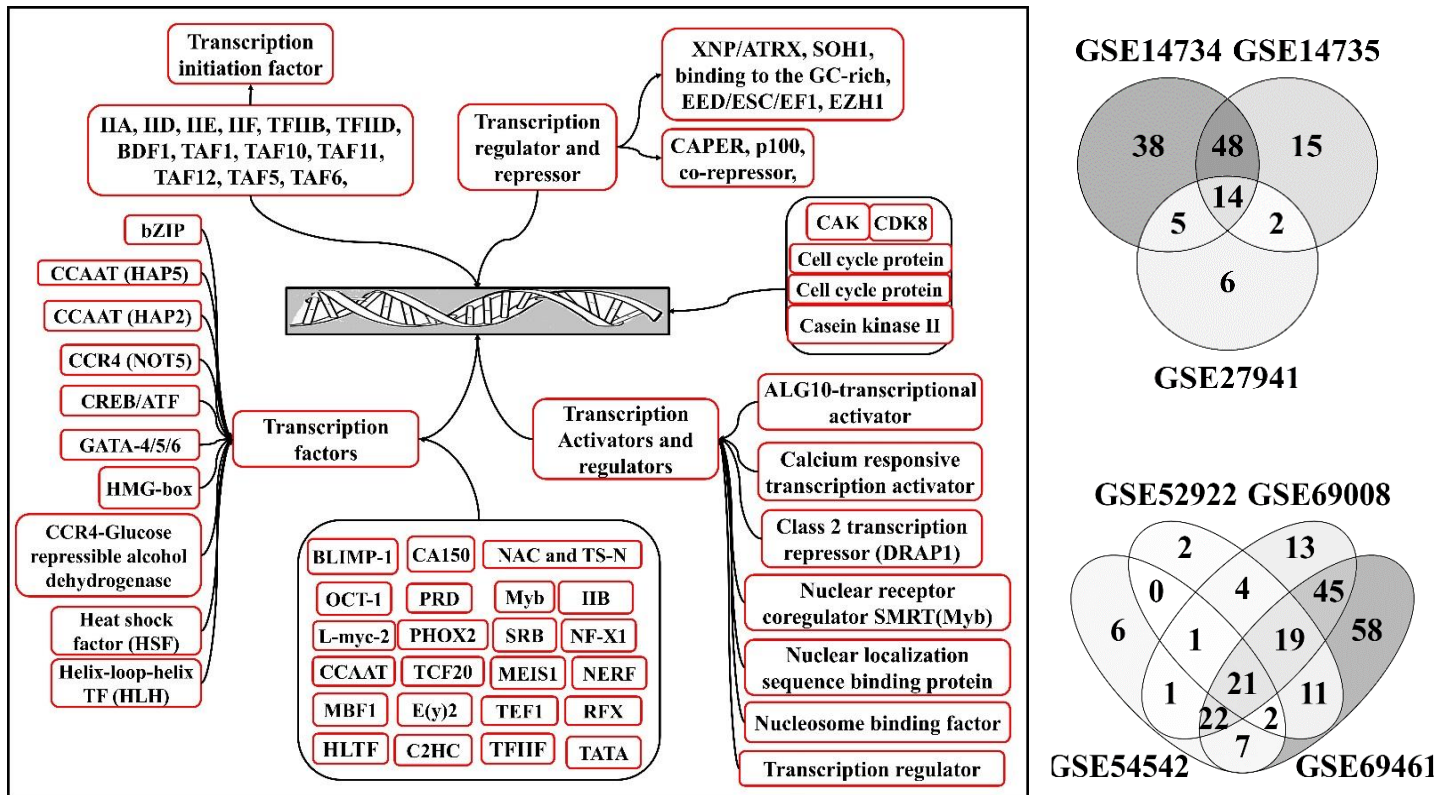


Figure 4.7: List of commonly expressed and statistically significant genes among various datasets encoding for transcription group (KOG functional ID: **K**)

4.4.5. Replication, Repair and Recombination (KOG ID: **L**)

DNA is the storehouse for the cellular genetic material from which other molecular components such as RNA and protein are derived. However, eukaryotic DNA experiences various DNA-damaging events every moment and the damaged DNA inhibits various necessary cellular mechanisms and pathways such as DNA replication and transcription, thus challenging the cellular fate [40]. Thus, eukaryotic cells have evolved with various repair mechanisms to reserve DNA from different types of damages. Eukaryotic cells employ five different types of repair mechanisms a) base excision repair, b) mismatch repair c) nucleotide excision repair d) homologous recombination and e) non-homologous end joining, state of the art review articles are already available on eukaryotic repair mechanisms [40]. As mentioned earlier *P. chrysosporium* genome encodes 242 genes classified under replication, repair and recombination group, based on the annotations these genes we have further tentatively grouped the genes based on their function as 60 (base excision repair), 20 (mismatch repair), 17 (nucleotide excision repair), 8 (homologous recombination) and 6 (non-homologous end joining). In GSE14734, GSE14735 and GSE27941 datasets, various genes

encoding for replication, repair and recombination were expressed statistically significant, however no genes were expressed with fold change ≥ 2.0 . Total of 52, 6, 5 genes were found to statistically significant and commonly expressed among GSE14734-GSE14735, GSE14734-GSE14735-GSE27941 and GSE14734 -GSE27941 datasets respectively. Most of the genes encoding for replication, repair and recombination were up regulated in cellulose grown samples followed by ball milled aspen. Transcripts encoding for 33 genes (5'-3' exonuclease (HKE1/RAT1, XRN1/KEM1/SEP1) ATP-dependent DNA helicase, DNA mismatch repair protein (MLH2, MutS), DNA polymerase, DNA repair protein RAD18, DNA licensing factor (MCM3, MCM4, MCM6, MCM7), DNA damage check point RHP9/CRB2/53BP1, Replication factor C, Signaling protein SWIFT and BRCT domain proteins, Single-stranded DNA-binding replication protein A) were found to be up regulated in glucose and down regulated in BMA and cellulose samples. Similarly, most of the replication, repair and recombination genes were up regulated in carbon and nitrogen limited growth conditions followed by replete medium.

In GSE54542 dataset, total of 45 genes were differentially expressed with a fold change ≥ 2.0 , 28 genes (5'-3' exonuclease, ATP-dependent DNA helicase, Cdk activating kinase (CAK), DNA damage checkpoint protein RHP9/CRB2/53BP1, DNA mismatch repair protein - MLH1, DNA polymerase, Eukaryotic-type DNA primase, DNA repair protein XPA-interacting protein, Mismatch repair ATPase (MSH4, MSH5, MSH6), origin of recognition complex, DNA damage inducible protein, Replication factor C (RFC2, RFC4), SNF2 family DNA-dependent ATPase and Tam3-transpose) were highly up regulated in control. In oak acetic extract samples, 17 genes were highly up regulated (3'-5' DNA helicase, 3'-5' exonuclease, 5'-3' exonuclease, ATPase related to holliday junction resolvase, damage specific DNA binding complex (DDB1), DNA polymerase, DNA repair and recombination protein (RAD52/RAD22, RHP57), DNA replication licensing factor (MCM2, MCM3, MCM4), endonuclease-III, HLTf/DNA helicase RAD5, NER factor (NEF2), origin recognition subunit1). In GSE52922 dataset, total of 24 replication, repair and recombination genes were found to be statistically significant out of which only 2 genes (DNA polymerase sigma, NAD⁺ ATP ribosyltransferase Parp) were significantly up regulated in 82 followed by P717 growth substrates. In GSE69008 dataset, genes encoding for 5'-3' exonuclease, transposase, ribonuclease HI, tam3-transposase (AC family) were significantly up regulated in A (high lignin-low glucose) and genes encoding for DNA topoisomerase I, nucleosome binding factor SPN, tam3-transposase were up regulated in B (low lignin-high glucose). In C (average lignin-average glucose) samples,

5'-3' exonuclease, NAD⁺ ADP-ribosyltransferase Parp, nucleotide excision repair (NEF2), ribonuclease HI were up regulated and down regulated in A and B samples. Finally, in GSE69461 dataset total of 120 and 40 genes were found to be highly upregulated with a fold change ≥ 2.0 . In 40-hour samples genes encoding for 3-methyladenine DNA glycosidase, ATP-dependent helicase and ligase, damage specific DNA binding complex, DNA damage check point, damage inducible protein, DNA polymerase, topoisomerase I, G/T mismatch DNA glycosylase, DNA repair protein XPA, MSH2, NAD⁺ADP ribosyltransferase Parp were highly expressed. Genes involved in DNA replication and various repair mechanisms were found to be highly up regulated in 96-hour samples (Figure 4.8).

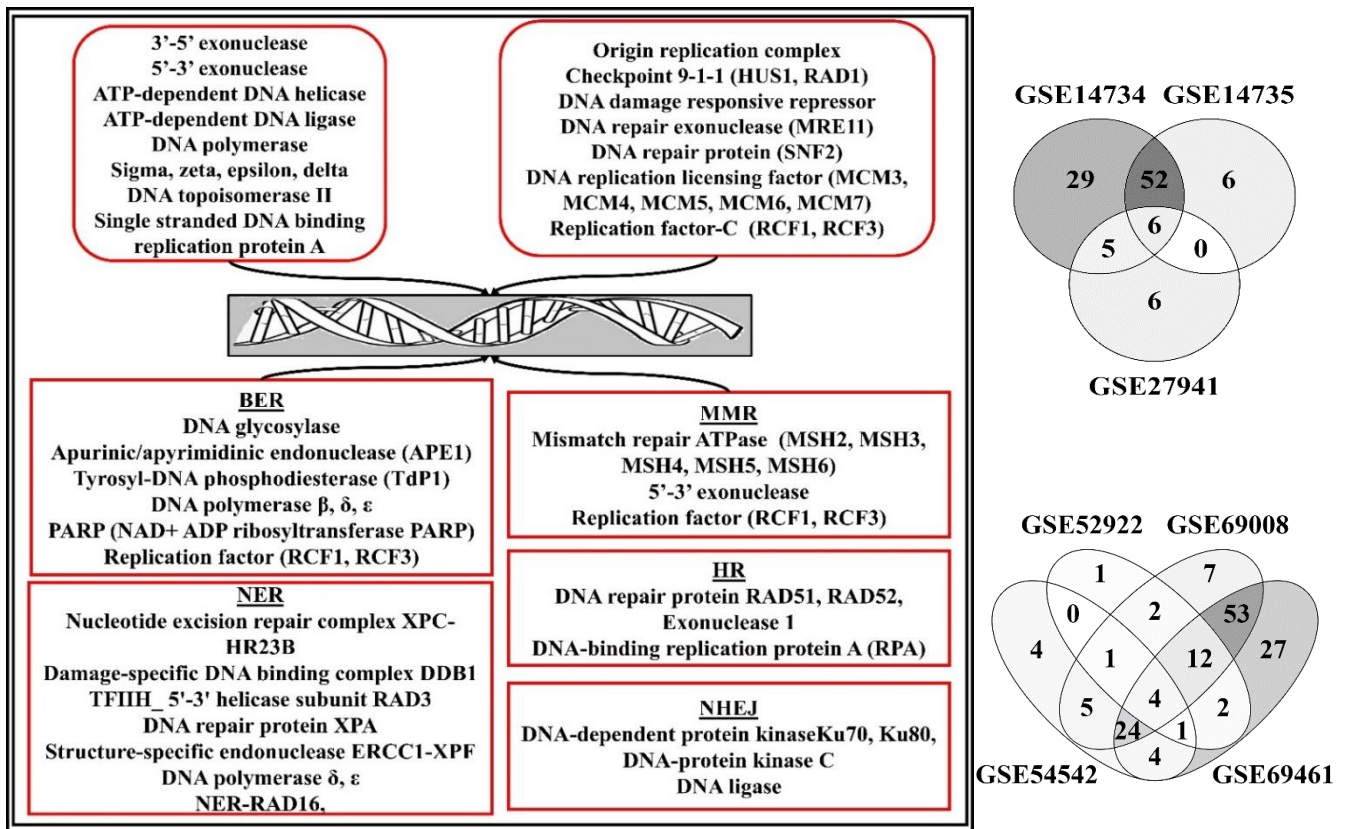


Figure 4.8: List of commonly expressed and statistically significant genes (A), (B) among the selected gene expression datasets encoding for Replication, repair and recombination group (KOG functional ID: L)

4.5. Discussion

Extracellular digestion of the growth substrate by secreting highly oxidative and degradative enzymes, before the absorption of nutrients is one of the significant characteristics of fungi. White rot fungi such as *P. chrysosporium* secretes an array of oxidative and hydrolytic enzymes for the degradation of plant cell wall components, as a primary degrader of lignin. Thus, white rot fungi play a crucial role in maintaining the global carbon cycle [7]. However, the expression and turnover of *P. chrysosporium*'s lignocellulolytic enzymes are directly controlled by wide range of genes at DNA and RNA level. We have clearly listed some important set of genes encoding for different protein models classified under information, storage and processing in *P. chrysosporium*'s genome (Figure 4.9). The 30 million base pair genome of *P. chrysosporium* has revealed variation within complex gene families encoding for oxidases, peroxidases, cytochrome P450 monooxygenases and glycosyl hydrolases [7].

Maintaining the integrity and stability of eukaryotic genomic DNA is of primary importance to the cell, as the DNA damage compromises some of the essential cellular processes (transcription and replication) and sometimes it might even challenge cell's survival [40]. The genomic DNA experiences different types of damages such as endogenous or DNA replication related damage (DNA mismatches introduced at a rate of 10^{-4} to 10^{-6} , chemically altered nucleotides ex: 8-oxo-dUTP and dGTP), exogenous or environmental damage (ultraviolet rays, ionizing radiations) and chemical damage (alkylating agents, heterocyclic amines, polycyclic aromatic hydrocarbons) [40]. Thus, eukaryotic cells have developed a five different types of specific repair mechanisms (BER, MMR, NER, HR and NHEJ) [41]. Degradation of polyphenolic compounds (lignin) and other plant components present in the growth substrates (GSE54542, GSE52922, GSE69008 and GSE69461) might result in chemical damage to the DNA. Thus, resulting in higher expression of genes involved in DNA repair mechanisms. Various studies have confirmed the involvement of reactive oxygen species (ROS), hydroxy radicals (OH^*) and hydrogen peroxide (H_2O_2) by ligninolytic peroxidases during the degradation of lignin [6, 42-45]. Interaction of ROS, OH^* , H_2O_2 and superoxide anion with DNA and other macromolecules results in various forms of DNA lesions challenging the cell survival by causing various detrimental effects [46]. According to Evan and Littlewood 1998, multicellular eukaryotic organisms employ programmed cell death process known as apoptosis for the elimination of cells possessing irreparable damaged DNA

[47]. However, unicellular organisms like *Saccharomyces cerevisiae* adapts the cells with irreparable damaged DNA and reenter into the cell cycle [48]. According to Glass et al (2000), filamentous fungi might use apoptosis for removal of hyphal cells which have undergone anastomosis with a discordant pattern [49]. Recent studies conducted by Jung et al (2016) on *Cryptococcus neoformans* (radiation resistant basidiomycetes fungus, usually found in radioactive habitats) have reported a special transcription factor Bdr1 containing basic leucine zipper domain which regulates the expression of genes coding for DNA repair genes [50]. However, the expression of Bdr1 is dependent on DNA damage response protein kinase (Rad53) [50]. Significantly common expression of genes encoding DNA repair protein (RAD1, RAD3, RAD5, RAD14, RAD16, RAD17, RAD18, RAD51 and RAD52) among the natural plant biomass growth substrates explains the occurrence of DNA lesions in *P. chrysosporium*. Thus, DNA repair is a vital procedure for protecting the cellular genetic information.

A	HMG-box	11	B	tRNA-Synthetase	42	
	HLTF	10		Translation Initiation factor	57	
	GATA-4/5/6	5		Translation elongation factor	4	
	HST	4		Mitochondrial/chloroplast ribosomal proteins	28	
	CCAAT(HAP5)	3		Mitochondrial ribosomal protein	10	
	CCR4-NOT5	3		Exosomal 3'-5' exoribonuclease complex	12	
	HLH	3		Elongation factor	6	
	C2HC (Zn finger)	3		Amidases	13	
	Myb	3		60S ribosomal protein subunits	51	
	C2H2-(Zn-finger)	1		40S ribosomal protein subunits	31	
	bZIP	1		C	Histone Acetyltransferases (HAT)	15
	CCAAT (HAP3)	1			Histone Deacetylases (HDAC)	16
	CCAAT (HAP2)	1			Histone Methyltransferases (HKMT)	5
	CREB	1			Histone Methylases	32
	CREB/ATF	1			Chromatin assembly complex (CAF)	4
	CCR4	1			Chromatin remodeling complex (CRF)	14
	Alg8p	1			Chromosome condensation complex (CCC)	3
	HST	1		D	Base Excision Repair (BER)	60
	HLH (BF/Olf-1)	1			Mismatch Repair (MMR)	20
	HMG-box (Capicua)	1			Nucleotide Excision Repair (NER)	17
TEA (r SPT4)	1	Homologous Recombination (HR)	8			
TEA (TAT-SF1)	1	Non-Homologous end Joining (NHEJ)	6			
TEA (TFIIS)	1					
Zn-finger TF	1					

Figure 4.9: Heat map (Color scales) listing the important groups of genes encoding for A) transcription factors B) translation ribosomal structure and biogenesis C) chromatin structure and dynamics D) DNA repair mechanisms in genome of *P. chrysosporium*.

DNA repair biochemical studies are usually performed using naked DNA substrates. However, this procedure is not physiologically applicable as the naked DNA is susceptible to nuclease digestion. Naked DNA will require more space and prone to other cellular insults until and unless it is organized [51]. Thus, eukaryotic naked DNA is structurally organized into nucleosomes (147 base pair DNA is packed around core histone proteins H2A, H2B, H3 and H4) which are further packed into chromatin. Subsequently, this chromatin condenses into chromosomes which enable accurate cellular division [52]. Chromatin's dynamic structure controls gene expression by controlling the nuclear processes such as DNA replication, transcription and DNA repair. During DNA replication, eukaryotic chromatin is exposed to proof reading and DNA repair mechanisms to ensure accurate transfer of genetic information. Present data analysis of the gene expression data revealed the genes involved in chromatin remodeling such as chromatin remodeling protein, histone acetyltransferase (MYST, SAGA/ADA, TRRAP), histone methyltransferases, chromatin remodeling complex (SWI-SNF, RSC, WSTF-ISWI and PHD Zn-finger), histone deacetylase complex (SIN3, RPD3). Along with these enzymes, genes coding for sirtuin 4, 5 (SIR2 family) structure maintenance of chromosome protein 4, nucleosome assembly protein and remodeling factors were common among the datasets. Based on our observations above-mentioned enzymes involved in chromatin structure, remodeling and dynamics were found to be highly expressed in natural plant biomass growth substrates.

Gene regulation is of higher priority in physiology of all the organisms which ensures the up and down regulation of genes by responding to the growth conditions [9]. Studies have reported that 37 different classes of regulatory proteins were identified in fungi which control and coordinate fungal growth [9]. According to Tiziano et al (2017), initial recognition of the growth substrate by fungi initiates the expression and secretion of plant biomass utilizing enzymes and the respective metabolic networks [53]. Plant cell wall components such as polysaccharides and lignin cannot act as inducers due to their large size as they cannot enter the fungal cell. Fungi identify the presence or absence of polymers by secreting polymer derived low molecular weight components (ex: mono (or) disaccharides). Majorly, the transcription factors regulating fungal plant cell wall degrading enzymes belong to zinc

cluster family (contains Zn-fingers and cysteine/histidine residues). Studies have reported that most positive and negative regulators of fungal plant cell wall degrading enzymes belong to Zn₂Cys₆ and Cys₂His₂ classes [53]. Reports on commercially important fungi such as *Aspergillus nidulans*, *Trichoderma reesei*, *Neurospora crass*, *Fusarium sp*, *Colletotrichum gloeosporioides*, *Botrytis cinerea*, *Magnaportha oryzae*, *Aspergillus aculeatus* etc., have revealed about important transcription factors regulating the expression of plant cell wall polysaccharide utilizing enzymes (Table 4.3) [53].

Table 4.3: Transcription factors involved in regulation of plant cell wall utilizing enzymes studied in various commercially important fungi:

Plant biomass component	Transcription factors	Enzymes
Cellulose	CLR-1, CLR-2, ClrA/1 (Cellulase regulators-Zn ₂ Cys ₆ -class) XlnR (Xylanolytic transcriptional activator-Zn ₂ Cys ₆) ACE2, ACE3 (Activators of cellulase expression-Zn ₂ Cys ₆) ClbR (Cellobiose response regulator-Zn ₂ Cys ₆) McmA (MADS-box protein-MADS-box) HAP2, HAP3, HAP5 (Multimeric protein complex) ACE1, CreA/1, BglR/COL-26 (Cellulase enzyme repressors) CRE1, CREA, CREB, CREC	Cellulases
Hemicellulose	XlnR/XYR1 (Xylanolytic transcriptional activator-Zn ₂ Cys ₆) AraR, ARA1 (Arabinose responsive regulators Zn ₂ Cys ₆) GalR (GalX) (Galactose responsive regulator-Zn ₂ Cys ₆)	Xylanases, Arabinases, Galactases
Starch, Inulin	InuR (Inulinolytic regulation-Zn ₂ Cys ₆) AmyR, MalR (Amylolytic regulation-Zn ₂ Cys ₆)	Inulases, Amylases, Maltases,
Pectin	RhaR, GaaR (GaaX) (Pectinolytic regulation-Zn ₂ Cys ₆)	Rhamnosidases, Pectinases

Genes encoding for various transcription factors were statistically significant and commonly expressed among *P. chrysosporium* gene expression datasets. Genes encoding for transcription factors CREB/ATF, HMG-box, MADS-box, CCAAT (HAP5, HAP2), CCR4,

calcium-responsive transcription coactivator, GATA/4/5/6 and L-myc2 were statistically significant and commonly expressed among all the datasets. Previous reports suggest that carbon catabolite repression (CCR) regulates the expression and secretion of plant cell wall utilizing enzymes when cultured on certain carbon sources [10, 38, 54-56]. Foreman et al (2003) have revealed that genes encoding for plant cell wall utilizing enzymes were coregulated by XYR1, ACE2, HAP-2/3/5 complex (positive regulators) with repressors being ACE1 and CRE1 [57]. The cis acting element CCAAT motif is present on promoter and enhancer regions in most of the eukaryotic genes. In filamentous fungi, all the CCAAT box binding proteins were mostly classified under HAP-like factors [56, 57]. Previous reports suggest that HAP-2/3/5 complex is involved in generation of open chromatin structure which subsequently required for the total transcriptional activation of cellulases [58, 59]. Studies conducted on *Aspergillus sp.* have reported that CRE-A, CRE-B and CRE-C are involved in carbon catabolite repression regulatory mechanisms [60-64]. According to de Vries, Ronald P, and Jaap Visser (2001), expression of genes encoding for cellulase, xylanase, arabinase, and various endoxylanase, xylosidase, feruloyl esterase and few pectinase are affected by CRE-A mediated repression [54]. Significant expression of heat shock, HMG-box, helicase like, RFX, Myb, L-myc2 transcription factors only among the natural plant biomass growth substrates might reveal the increased cellular stress on *P. chrysosporium* during degradation of lignin and other plant extractives. According to Jessica. M et al (1997), expression of *P. chrysosporium* genes encoding for manganese peroxidase (MnP) is strongly regulated by manganese (Mn), heat shock and hydrogen peroxide (H₂O₂) when cultured on nitrogen limited growth substrates [19, 65].

Before the process of translation in the cellular cytoplasm, messenger RNA is subjected to 5' capping, mRNA-splicing, 3'end processing and mRNA export, these cellular events are strongly interdependent and influence cellular fate of a transcript [25]. Genes encoding for ATP-dependent RNA helicase, DEAH-box RNA helicase, fibrillar related nucleolar RNA binding proteins, polyadenylation factor complex, splicing coactivator SRM 160/300, splicing factor RNPS1, hnRNP-F, 3b-subunit 4, mRNA splicing factor were statistically significant and commonly expressed among all the datasets. Serine-arginine nuclear matrix protein (SRM 160/300 protein) play crucial role in regulating alternative splicing, possibly by localizing splicing machinery components to the transcription active site. SR proteins are regulated by phosphorylation and dephosphorylation reactions as these are phosphoproteins in nature [66]. Higher expression of genes encoding for splicing factors in the natural plant

biomass growth substrates might be due to its involvement in regulation of lignocellulolytic enzymes. Besides the abovementioned common enzymes genes encoding for various 40S and 60S ribosomal proteins, amidases, translation initiation factor 2C, 4F, PUF3, were expressed commonly among GSE14734, GSE14735 and GSE27941 datasets. In natural plant biomass substrates (GSE54542, GSE52922, GSE69008 and GSE69461) Exosomal 3'-5' exoribonuclease, polyadenylate binding protein, translation initiation inhibitor UK114, initiation factor 4F and repressor MPT5 proteins were commonly expressed. Studies have confirmed the involvement of the enzymes in regulation and maintenance of the protein turnover, however the exact involvement and role of the abovementioned enzymes in expression and regulation of lignocellulolytic and detoxification enzymes in *P. chrysosporium* is not explained till today.

In this study, we have revealed the common gene expression patterns of *P. chrysosporium* involved in regulation of protein expression and turnover of lignocellulolytic enzymes. We have extensively reported about various genes involved in information storage and processing of *P. chrysosporium*. Higher and significant expression of various genes encoding for information storage and processing especially DNA damage, repair and recombination mechanisms, mRNA splicing, histone acetyltransferases by *P. chrysosporium* were common and highly expressed among datasets cultured on natural plant biomass growth substrate. Thus, degradation of natural plant biomass containing lignin and other plant extractives along with plant polysaccharides causes various DNA level lesions. These results also convey that expression of lignocellulolytic genes internally depends on expression or repression of various genes involved in information storage and processing. However, further investigations must be performed to understand and analyze the exact involvement of the abovementioned genes in expression and regulation of *P. chrysosporium* lignocellulolytic enzymes. Application of next generation sequencing techniques such as ChIP (chromatin immunoprecipitation), RNA sequencing and mRNA splicing studies might reveal the functional involvement of the reported genes which would significantly help in strain improvement and production of recombinant strains.

References

1. Skyba O, Cullen D, Douglas CJ, Mansfield SD. Gene expression patterns of wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium* are influenced by wood substrate composition during degradation. *Applied and environmental microbiology*. 2016: AEM. 00134-16.
2. Wymelenberg AV, Gaskell J, Mozuch M, BonDurant SS, Sabat G, Ralph J, et al. Gene expression of wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium* is significantly altered by plant species. *Applied and Environmental Microbiology*. 2011: AEM. 00508-11.

3. Wymelenberg AV, Gaskell J, Mozuch M, Kersten P, Sabat G, Martinez D, et al. Transcriptome and secretome analyses of *Phanerochaete chrysosporium* reveal complex patterns of gene expression. *Applied and environmental microbiology*. 2009; 75: 4058-68.
4. Wymelenberg AV, Gaskell J, Mozuch M, Sabat G, Ralph J, Skyba O, et al. Comparative transcriptome and secretome analysis of wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium*. *Applied and environmental microbiology*. 2010; 76: 3599-610.
5. Kameshwar AKS, Qin W. Metadata Analysis of *Phanerochaete chrysosporium* Gene Expression Data Identified Common CAZymes Encoding Gene Expression Profiles Involved in Cellulose and Hemicellulose Degradation. *International Journal of Biological Sciences*. 2017; 13: 85-99.
6. Kameshwar AKS, Qin W. Gene expression metadata analysis reveals molecular mechanisms employed by *Phanerochaete chrysosporium* during lignin degradation and detoxification of plant extractives. *Current Genetics*. 2017: 1-18.
7. Martinez D, Larrondo LF, Putnam N, Gelpke MDS, Huang K, Chapman J, et al. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nature biotechnology*. 2004; 22: 695-700.
8. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otilar R, et al. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research*. 2013: gkt1183.
9. Todd RB, Zhou M, Ohm RA, Leeggangers HA, Visser L, De Vries RP. Prevalence of transcription factors in ascomycete and basidiomycete fungi. *BMC genomics*. 2014; 15: 214.
10. Amore A, Giacobbe S, Faraco V. Regulation of cellulase and hemicellulase gene expression in fungi. *Current genomics*. 2013; 14: 230-49.
11. Hortschansky P, Haas H, Huber EM, Groll M, Brakhage AA. The CCAAT-binding complex (CBC) in *Aspergillus* species. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*. 2017; 1860: 560-70.
12. Brosch G, Loidl P, Graessle S. Histone modifications and chromatin dynamics: a focus on filamentous fungi. *FEMS microbiology reviews*. 2008; 32: 409-39.
13. Lusser A, Kadonaga JT. Chromatin remodeling by ATP-dependent molecular machines. *Bioessays*. 2003; 25: 1192-200.
14. Eberharter A, Becker PB. ATP-dependent nucleosome remodelling: factors and functions. *Journal of cell science*. 2004; 117: 3707-11.
15. Turner BM. Defining an epigenetic code. *Nature cell biology*. 2007; 9: 2-6.
16. Dobosy J, Selker E. Emerging connections between DNA methylation and histone acetylation. *Cellular and molecular life sciences*. 2001; 58: 721-7.
17. Steinfeld I, Shamir R, Kupiec M. A genome-wide analysis in *Saccharomyces cerevisiae* demonstrates the influence of chromatin modifiers on transcription. *Nature genetics*. 2007; 39: 303-9.
18. Woolford JL, Baserga SJ. Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics*. 2013; 195: 643-81.
19. Gold MH, Alic M. Molecular biology of the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. *Microbiological reviews*. 1993; 57: 605-22.
20. Thuillier A, Chibani K, Belli G, Herrero E, Dumarçay S, Gérardin P, et al. Transcriptomic Responses of *Phanerochaete chrysosporium* to Oak Acetonic Extracts: Focus on a New Glutathione Transferase. *Applied and environmental microbiology*. 2014; 80: 6316-27.
21. Wymelenberg AV, Gaskell J, Mozuch M, BonDurant SS, Sabat G, Ralph J, et al. Significant alteration of gene expression in wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium* by plant species. *Applied and environmental microbiology*. 2011; 77: 4499-507.
22. Gaskell J, Marty A, Mozuch M, Kersten PJ, BonDurant SS, Sabat G, et al. Influence of *Populus* genotype on gene expression by the wood decay fungus *Phanerochaete chrysosporium*. *Applied and environmental microbiology*. 2014; 80: 5828-35.
23. Korripally P, Hunt CG, Houtman CJ, Jones DC, Kitin PJ, Cullen D, et al. Regulation of Gene Expression during the Onset of Ligninolytic Oxidation by *Phanerochaete chrysosporium* on Spruce Wood. *Applied and environmental microbiology*. 2015; 81: 7802-12.
24. Kameshwar AKS, Qin W. Metadata Analysis of *Phanerochaete chrysosporium* gene expression data identified common CAZymes encoding gene expression profiles involved in cellulose and hemicellulose degradation. *International journal of biological sciences*. 2017; 13: 85.

25. Hocine S, Singer RH, Grünwald D. RNA processing and export. *Cold Spring Harbor perspectives in biology*. 2010; 2: a000752.
26. Shuman S. Structure, mechanism, and evolution of the mRNA capping apparatus. *Progress in nucleic acid research and molecular biology*. 2000; 66: 1-40.
27. Busch H, Reddy R, Rothblum L, Choi Y. SnRNAs, SnRNPs, and RNA processing. *Annual review of biochemistry*. 1982; 51: 617-54.
28. Wahle E, Rügsegger U. 3'-End processing of pre-mRNA in eukaryotes. *FEMS microbiology reviews*. 1999; 23: 277-95.
29. Stark H, Lührmann R. Cryo-electron microscopy of spliceosomal components. *Annu Rev Biophys Biomol Struct*. 2006; 35: 435-57.
30. Bradbury EM. Chromatin structure and dynamics: state-of-the-art. *Molecular cell*. 2002; 10: 13-9.
31. Zlatanova J, Leuba SH. Chromatin structure and dynamics: state-of-the-art: Gulf Professional Publishing; 2004.
32. Kouzarides T. Chromatin modifications and their function. *Cell*. 2007; 128: 693-705.
33. Widom J. Structure, dynamics, and function of chromatin in vitro. *Annual review of biophysics and biomolecular structure*. 1998; 27: 285-327.
34. Tessarz P, Kouzarides T. Histone core modifications regulating nucleosome structure and dynamics. *Nature reviews Molecular cell biology*. 2014; 15: 703-8.
35. Strauss J, Reyes-Dominguez Y. Regulation of secondary metabolism by chromatin structure and epigenetic codes. *Fungal Genetics and Biology*. 2011; 48: 62-9.
36. Brown SJ, Cole MD, Erives AJ. Evolution of the holozoan ribosome biogenesis regulon. *BMC genomics*. 2008; 9: 442.
37. Ilmen M, Thrane C, Penttilä M. The glucose repressor gene *cre1* of *Trichoderma*: Isolation and expression of a full-length and a truncated mutant form. *Molecular and General Genetics MGG*. 1996; 251: 451-60.
38. Ruijter GJ, Visser J. Carbon repression in *Aspergilli*. *FEMS Microbiology Letters*. 1997; 151: 103-14.
39. Borin GP, Sanchez CC, Santana ES, Zanini GK, Santos RAC, Pontes AO, et al. Comparative transcriptome analysis reveals different strategies for degradation of steam-exploded sugarcane bagasse by *Aspergillus niger* and *Trichoderma reesei*. *BMC genomics*. 2017; 18: 501.
40. Dexheimer TS. DNA repair pathways and mechanisms. *DNA repair of cancer stem cells*: Springer; 2013. p. 19-32.
41. Altieri F, Grillo C, Maceroni M, Chichiarelli S. DNA damage and repair: from molecular mechanisms to health implications. *Antioxidants & redox signaling*. 2008; 10: 891-938.
42. Belinky PA, Flikshtein N, Lechenko S, Gepstein S, Dosoretz CG. Reactive oxygen species and induction of lignin peroxidase in *Phanerochaete chrysosporium*. *Applied and environmental microbiology*. 2003; 69: 6500-6.
43. Morel M, Ngadin AA, Jacquot J-P, Gelhaye E. Reactive oxygen species in *Phanerochaete chrysosporium* relationship between extracellular oxidative and intracellular antioxidant systems. *Advances in botanical research*. 2009; 52: 153-86.
44. Belinky PA, Flikshtein N, Dosoretz CG. Induction of lignin peroxidase via reactive oxygen species in manganese-deficient cultures of *Phanerochaete chrysosporium*. *Enzyme and Microbial Technology*. 2006; 39: 222-8.
45. Zacchi L, Palmer JM, Harvey PJ. Respiratory pathways and oxygen toxicity in *Phanerochaete chrysosporium*. *FEMS microbiology letters*. 2000; 183: 153-7.
46. Azzam EI, Jay-Gerin J-P, Pain D. Ionizing radiation-induced metabolic oxidative stress and prolonged cell injury. *Cancer letters*. 2012; 327: 48-60.
47. Evan G, Littlewood T. A matter of life and cell death. *Science*. 1998; 281: 1317-22.
48. Toczyski DP, Galgoczy DJ, Hartwell LH. CDC5 and CKII control adaptation to the yeast DNA damage checkpoint. *Cell*. 1997; 90: 1097-106.
49. Glass NL, Jacobson DJ, Shiu PK. The genetics of hyphal fusion and vegetative incompatibility in filamentous ascomycete fungi. *Annual review of genetics*. 2000; 34: 165-86.
50. Jung K-W, Yang D-H, Kim M-K, Seo HS, Lim S, Bahn Y-S. Unraveling Fungal Radiation Resistance Regulatory Networks through the Genome-Wide Transcriptome and Genetic Analyses of *Cryptococcus neoformans*. *mBio*. 2016; 7: e01483-16.

51. Downs JA, Nussenzweig MC, Nussenzweig A. Chromatin dynamics and the preservation of genetic information. *Nature*. 2007; 447: 951.
52. Mekhail K, Moazed D. The nuclear envelope in genome organization, expression and stability. *Nature reviews Molecular cell biology*. 2010; 11: 317.
53. Benocci T, Aguilar-Pontes MV, Zhou M, Seiboth B, Vries RP. Regulators of plant biomass degradation in ascomycetous fungi. *Biotechnology for Biofuels*. 2017; 10: 152.
54. de Vries RP, Visser J. *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. *Microbiology and molecular biology reviews*. 2001; 65: 497-522.
55. Aro N, Pakula T, Penttilä M. Transcriptional regulation of plant cell wall degradation by filamentous fungi. *FEMS microbiology reviews*. 2005; 29: 719-39.
56. Kubicek CP, Mikus M, Schuster A, Schmoll M, Seiboth B. Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*. *Biotechnology for Biofuels*. 2009; 2: 19.
57. Foreman PK, Brown D, Dankmeyer L, Dean R, Diener S, Dunn-Coleman NS, et al. Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. *Journal of Biological Chemistry*. 2003; 278: 31988-97.
58. Zeilinger S, Schmoll M, Pail M, Mach RL, Kubicek CP. Nucleosome transactions on the *Hypocrea jecorina* (*Trichoderma reesei*) cellulase promoter *cbh2* associated with cellulase induction. *Molecular genetics and genomics*. 2003; 270: 46-55.
59. Bolotin-Fukuhara M. Thirty years of the HAP2/3/4/5 complex. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*. 2017; 1860: 543-59.
60. Arst HN, Cove DJ. Nitrogen metabolite repression in *Aspergillus nidulans*. *Molecular and General Genetics MGG*. 1973; 126: 111-41.
61. Dowzer C, Kelly JM. Analysis of the *creA* gene, a regulator of carbon catabolite repression in *Aspergillus nidulans*. *Molecular and Cellular Biology*. 1991; 11: 5701-9.
62. Hynes M, Kelly JM. Pleiotropic mutants of *Aspergillus nidulans* altered in carbon metabolism. *Molecular and General Genetics MGG*. 1977; 150: 193-204.
63. Lockington RA, Kelly JM. Carbon catabolite repression in *Aspergillus nidulans* involves deubiquitination. *Molecular microbiology*. 2001; 40: 1311-21.
64. Todd R, Lockington R, Kelly J. The *Aspergillus nidulans creC* gene involved in carbon catabolite repression encodes a WD40 repeat protein. *Molecular and General Genetics MGG*. 2000; 263: 561-70.
65. Gettemy JM, Li D, Alic M, Gold MH. Truncated-gene reporter system for studying the regulation of manganese peroxidase expression. *Current genetics*. 1997; 31: 519-24.
66. Eldridge A, Issner R, Li Y, Reifenberg E, Sharp PA, Blencowe BJ. Role of the SRm160/300 splicing coactivator in exon-enhancer function. *Biochemistry and Cell Biology*. 1999; 77: 393.

Chapter-5

Molecular Networks of *Postia placenta* Involved in Degradation of Lignocellulosic Biomass Revealed from Metadata Analysis of Open Access Gene Expression Data

[This work has been published in “**International journal of biological sciences**” 2018; 14(3):237-252. doi:10.7150/ijbs.22868.]

Ayyappa Kumar Sista Kameshwar and Wensheng Qin*

5.1. Abstract

To understand the common gene expression patterns employed by *P. placenta* during lignocellulose degradation, we have retrieved genome wide transcriptome datasets from NCBI GEO database and analyzed using customized analysis pipeline. We have retrieved the top differentially expressed genes and compared the common significant genes among two different growth conditions. Genes encoding for cellulolytic (GH1, GH3, GH5, GH12, GH16, GH45) and hemicellulolytic (GH10, GH27, GH31, GH35, GH47, GH51, GH55, GH78, GH95) glycoside hydrolase classes were commonly up regulated among all the datasets. Fenton's reaction enzymes (iron homeostasis, reduction, hydrogen peroxide generation) were significantly expressed among all the datasets under lignocellulolytic conditions. Due to the evolutionary loss of genes coding for various lignocellulolytic enzymes (including several cellulases), *P. placenta* employs hemicellulolytic glycoside hydrolases and Fenton's reactions for the rapid depolymerization of plant cell wall components. Different classes of enzymes involved in aromatic compound degradation, stress responsive and detoxification mechanisms (cytochrome P450 monooxygenases) were found highly expressed in complex plant biomass substrates. We have reported the genome wide expression patterns of genes coding for information, storage and processing (KOG), tentative and predicted molecular networks involved in cellulose, hemicellulose degradation and list of significant protein-ID's commonly expressed among different lignocellulolytic growth conditions.

Keywords: *Postia placenta*, Brown-rot decay, Gene expression, NCBI-Gene Expression Omnibus (GEO), Lignocellulose, Fenton's reaction

5.2. Introduction

Postia placenta a filamentous basidiomycete, which causes brown rot decay of the wood, and is one of the major destructors of wood-based constructions. *P. placenta* also exhibits the unique decaying property of brown rot fungi by rapidly breaking down cellulosic units of plant biomass without efficiently depolymerizing lignin units [1]. Whereas white rot fungi such as *Phanerochaete chrysosporium* attack plant biomass by efficiently degrading lignin followed by carbohydrate units of the plant biomass. Collectively these fungi are major inhabitants of forest biomass thus, playing a major role in carbon cycling and recycling of nutrients [1, 2]. Apart from their contrasting decay patterns, both *P. chrysosporium* and *P. placenta* belong to order Polyporales and are classified under Phlebia clade [3]. Phylogenetic studies conducted earlier have reported that brown rot fungi have progressively evolved from white rot fungi, the morphological properties and their choice of substrates also supports the above fact [4]. Martinez et al (2009), have completed the total genome sequence of *P. placenta* and performed genome wide transcriptome and secretome analysis, through which he has revealed the functional properties of various genes regulated during the process of wood decay [1]. This study has revealed a surprising fact that *P. placenta* genome codes for fewer number of cellulases (glycoside hydrolases) and other CAZymes when compared to *P. chrysosporium*. Present day JGI MycoCosm database harbours the genomes of *P. placenta* MAD 698-R v1.0 and *P. placenta* MAD-698-R-SB12 v1.0, coding for about 242 CAZymes and 324 CAZymes respectively [1] (Figure 5.1.). Though brown-rot fungi are well known for extensive cellulose depolymerization, *P. placenta* genome reveals that number of genes coding for cellulases are less when compared to *P. chrysosporium*. Thus, it is highly significant to understand the molecular mechanisms responsible for lignocellulose degradation by *P. placenta*.

Studies were also being conducted to understand and reveal the genes encoding enzymes involved in Fenton's reactions employed during lignocellulose degradation [5]. Brown rot fungi generally involve several non-enzymatic pathways to commence the Fenton's reactions which are required for the generation of ROS, OH^{*}, H₂O₂ etc. The probable mechanisms involved are a) location and solubilization of iron (Fe (oxyhydroxide) complexes in plant wood cell lumen, which is reduced to ferrous iron) [6, 7], b) hydrogen peroxide generation (reduction of molecular oxygen and oxidation of methanol)[8-10], c) iron reducing agents (reduction of Fe³⁺ to Fe²⁺ by reductants like 2,5-dimethoxy hydroquinone) [5, 6, 11,

12]. When compared to commercial cellulase producers like *Trichoderma reesei*, brown rot fungi such as *P. placenta* lacks genes coding for cellulases/carbohydrate binding domains, exo-cellobiohydrolases [1]. According to Ratto et al (1997), free radicals generated from Fenton's reactions will initially attack the cell walls, followed by breakdown of native and holocellulose happens through action of hydrolytic enzymes (hemicellulases and endoglucanases) [5, 13]. It is well-known that brown-rot fungi only modify lignin without significantly altering or degrading the structure of lignin, recent studies show that brown rot fungi cause extensive demethylation, oxidation of side chains, depolymerization and also can potentially repolymerize lignin, however brown-rotted lignin is observed in polymeric form supporting earlier statement [5, 14, 15]. However, exact mechanisms involved in iron reduction and homeostasis, extracellular H₂O₂ and free radical generation and enzymes involved in lignocellulose degradation were unclear.

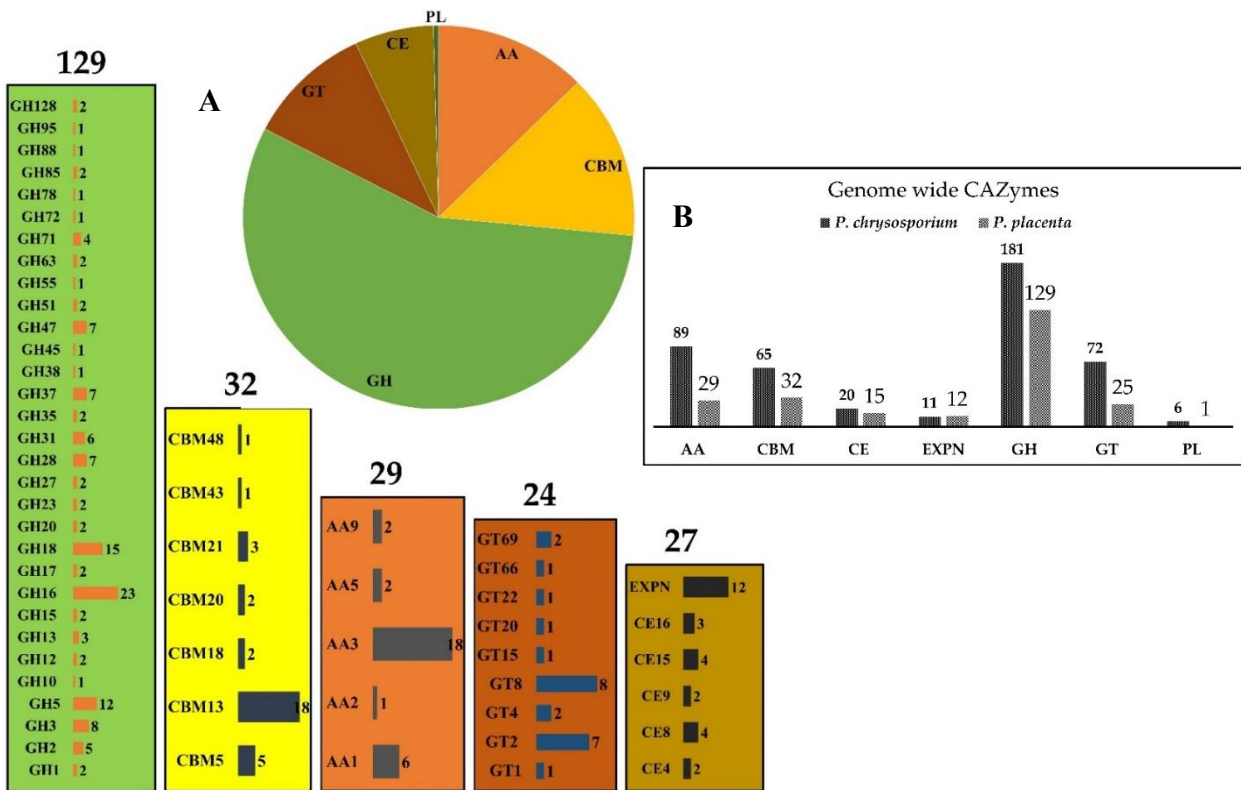


Figure 5.1: Pictorial representation of CAZymes distribution (GH-glycoside hydrolases, GT-glycosyl transferases, AA-auxiliary activity, CBM-carbohydrate binding domains, CE-carbohydrate esterases, EXPN-expansin like proteins, PL-polysaccharide lyases) in *P. placenta* MAD-698- Rv1.0 (**A**) and the top number denotes for the number of genes coding for particular class of enzymes and each bar internally shows different sub-classes of enzymes and the number of genes encoding for the corresponding enzymes, (**B**) Comparison of genome wide CAZymes between *P. placenta* and *P. chrysosporium*

Martinez et al (2009) and Wymelenberg et al (2010) have conducted a genome level transcriptome and proteome studies on *P. placenta* by culturing it on wood-derived microcrystalline cellulose, glucose (sole carbon source) and ball milled aspen, supplemented with Highley's basal salt medium (GSE12540). These studies have reported following facts A) *P. placenta* genome lacks genes encoding for exocellobiohydrolases and carbohydrate binding domains (commonly observed in cellulolytic microorganisms), B) several genes encoding for hemicellulases β -1-4 endoglucanases, iron and quinone reductases, extracellular iron Fe (II) and H₂O₂ generating oxidases were found to be highly expressed when *P. placenta* was cultured in cellulose medium. Wymelenberg et al (2011) has performed a microarray study to reveal the gene expression of *P. chrysosporium* and *P. placenta* colonized on ball milled aspen and pine. This study has reported that gene expression patterns of these fungi were significantly influenced by type of wood species (BMA or BMP) it colonizes and differences in gene expression patterns of these fungi reveal their preferences for carbon sources and their central decaying properties [16]. Daniel et al (2011) has developed a method for solubilization and multidimensional ¹H-¹³C NMR spectroscopy study to analyze the degradation of aspen wood by *P. placenta*. Obtained results showed that *P. placenta* majorly degraded the content of principal aryl glycerol- β -aryl ether interunit linkages of the lignin, by employing reactive oxygen species (ROS) free hydroxy radicals (OH*) obtained through extracellular Fenton system [17]. According to Micales (1991), oxalic acid which is highly produced and found accumulated in huge quantities by brown rot fungi, also might involve in iron chelation, pH gradient, free radical formation and acid hydrolysis of cellulose and hemicellulosic units etc [18]. In order to estimate the quantity of reactive oxygen species (ROS) produced by *P. placenta* laccase during the process of wood degradation, Wei et al (2010) has performed proteomic and gene expression studies of laccase and reported that oxidation of every hydroquinone molecule results in production of single perhydroxyl radical [19]. Oleksandr et al (2016) has performed, a genome wide transcriptome study to understand the effect of chemical composition of wood (*Populus trichocarpa*) on the gene expression of *P. placenta* and *P. chrysosporium*. Results obtained by them has proved that gene expression of these fungi is significantly influenced by the wood substrate composition and the incubation periods [20]. Recently Jiwei et al (2016) has performed a genome wide transcriptome study to examine differential expression of lignin oxidizing components involved in ROS generation earlier than differential expression of genes encoding for glycoside hydrolases [21]. Higher expression of ROS generating lignin

oxidizing components in hyphal front, followed by higher expression of genes encoding for glycoside hydrolases (GH) in lower hyphal front regions confirms the fact [21]. Zhang et al (2016), has developed a spatial mapping method to resolve temporal sequence in *P. placenta* by considering thin sections of directionally colonized wood wafers [21].

Understanding the fundamental molecular mechanisms employed by *P. placenta* during wood decay will significantly help the growing biofuel industries to develop novel and efficient methods for the breakdown of carbohydrate units (cellulose and hemicellulose) by selective modification of lignin (by demethoxylation and limited ring cleavage). In this study, we have performed gene expression metadata analysis on publicly available microarray and RNA-Seq datasets performed on *P. placenta* to understand the common gene expression patterns required during the progression of Fenton's reactions, oxalate metabolism, H₂O₂ and free radical generations, lignocellulose degradation (CAZymes and oxidoreductases).

5.3.1. Data Retrieval

Gene expression datasets used in our present study were retrieved from NCBI GEO (Gene Expression Omnibus is a public repository for gene expression datasets) by using the term "*Postia placenta*". Present day GEO repository resides four functional genomics datasets of *P. placenta* out of which three were microarray datasets (GSE12540, GSE29656, GSE69004) and one was RNA-Seq (GSE84529) dataset. All the microarray datasets were based on NimbleGen *P. placenta* MAD-698 whole genome microarray platform and cultured on different growth substrates. Dataset GSE12540 was based on *P. placenta* cultured on microcrystalline cellulose (Avicel), glucose, 0.5% (w/v) ball milled aspen (BMA) as sole carbon source and supplemented with Highley's basal medium [1, 22]. Similarly, GSE29656 dataset, *P. placenta* was cultured on 0.5% (w/v) ball milled white pine (*Pinus strobus*), 0.5% (w/v) ball milled bigtooth aspen (*Populus grandidentata*) as the sole carbon source and supplemented with Highley's basal medium for macro and micronutrients [16]. In dataset GSE69004, *P. placenta* was cultured on chemically distinct A (high lignin-low glucose), B (high glucose-low lignin) and C (average lignin-average glucose) *Populus trichocarpa* wood substrates [20]. The dataset GSE84529 *P. placenta* was cultured on wood wafers cut in dimensions (60 x 25 x 2.5mm) where the largest face of the wood wafers is cross sectioned and the tangential plane was arranged to be in contact with *P. placenta* mycelium, hyphal growth was allowed till 50 mm up on the wafers and later sectioned to A) 0-5 mm B) 15-20 mm C) 30-35 mm RNA-Seq analysis [21]. All the experimental conditions in these datasets were

designed with three replicates each and complete details about the gene expression platforms and sample information were listed in Table 5.1.

Table 5.1: Platform details of *Postia placenta* gene expression data:

Accession ID	Substrate	Platform	Samples	Ref
GSE84529	Wood wafers	Illumina HiSeq (<i>Postia placenta</i>)	9	[21]
GSE69004	Poplar wood stems	NimbleGen_UW/FPL <i>Postia placenta</i> MAD-698 whole genome 37K expression array version 1	23	[20]
GSE29656	Ball milled aspen Ball milled pine	NimbleGen_UW/FPL <i>Postia placenta</i> MAD-698 whole genome 37K expression array version 1	6	[16]
GSE12540	Glucose Cellulose Ball milled aspen	NimbleGen_UW/FPL <i>Postia placenta</i> MAD-698 whole genome 37K expression array version 1	9	[1, 22]

5.3.2 Data Analysis Methodology

Microarray GSE12540, GSE29656 and GSE69004 datasets were analyzed using GEO2R and customized programming scripts. The literature and supplementary information for the corresponding datasets were studied to understand the experimental design for performing the statistical analysis. Following options were applied when using GEO2R website A) autodetect option (for log transformation of the data), B) box-whisker plot (samples and value distributions), C) submitter provided annotations (for gene level annotations) and D) Benjamini & Hochberg (False discovery rate correction) for multiple testing correction of p-values was used. Once the samples are grouped according to their experimental conditions, Top 250 option was used for obtaining statistically significant genes expressed in each dataset. We have retrieved Top 1000 differentially expressed genes to understand *P. placenta* gene expression by using the function “Save all results” and the obtained results were analyzed further. The ranking of differentially expressed genes (Top 250) and statistical analysis in GEO2R internally happens through limma package and the statistically significant genes were ranked based on their corrected p-value (0.05). Thus, obtained statistically significant genes were analyzed by the supplier provided annotations Protein-ID and Best hit BlastP. Gene and protein level annotations of *P. placenta* MAD-698R were retrieved from MycoCosm (fungal genome repository) [23, 24]. Custom linux based scripts were written to retrieve the annotations for differentially expressed gene list using KOG, GO and InterPro

annotations of *P. placenta* MAD-698R v1. We have also used other analysis options available in JGI-MycoCosm such as Gene Ontology (GO)[25, 26], EuKaryotic Orthologous Groups (KOG) [27] and CAZy [28, 29] for analyzing the results obtained. The differentially expressed gene lists of the respective experimental conditions were compared using Venny 2.1 [30] and Jvenn [31] softwares. We have retrieved sample level FPKM (Fragments Per Kilobase of transcripts per Million) values, P-values and log fold change values from supplementary files provided for GSE84529 dataset, gene list was sorted based on their P-values and top 1000 genes were retrieved and compared among the conditions. The data analysis was similar as explained earlier in our previous studies [32-34].

Simultaneously, the datasets were also analyzed using GeneSpring® v.14.8 software. Gene expression datasets were retrieved using the option “Import NCBI GEO experiment” by saving the GEO sample files in the local folder. The experiments were created as generic single color by applying the following preprocessing conditions “Threshold value set at 1.0”, “Normalization using shift 75th percentile”, sample values were log base 2 transformed and baselined to the median of all samples. The experimental conditions were retrieved from the corresponding GEO experiment and literature, was used for the grouping the samples. The samples were filtered using “Probesets by expression” with the parameters set to data filter on normalized data and filter by percentile (upper percentile set to 100.0 and lower percentile 20.0) respectively. Based on the experimental conditions one-way Anova and Moderated T-test was performed respectively. However, for the dataset GSE84529, the samples were retrieved, and the experiment was created without any preprocessing steps. We have performed fold change analysis on the grouped samples using the FPKM values and the transcripts differentially expressed >2.0 were retrieved for the analysis. The detailed step by step workflow used for the metadata analysis of *P. placenta* gene expression datasets were showed in Figure 5.2.

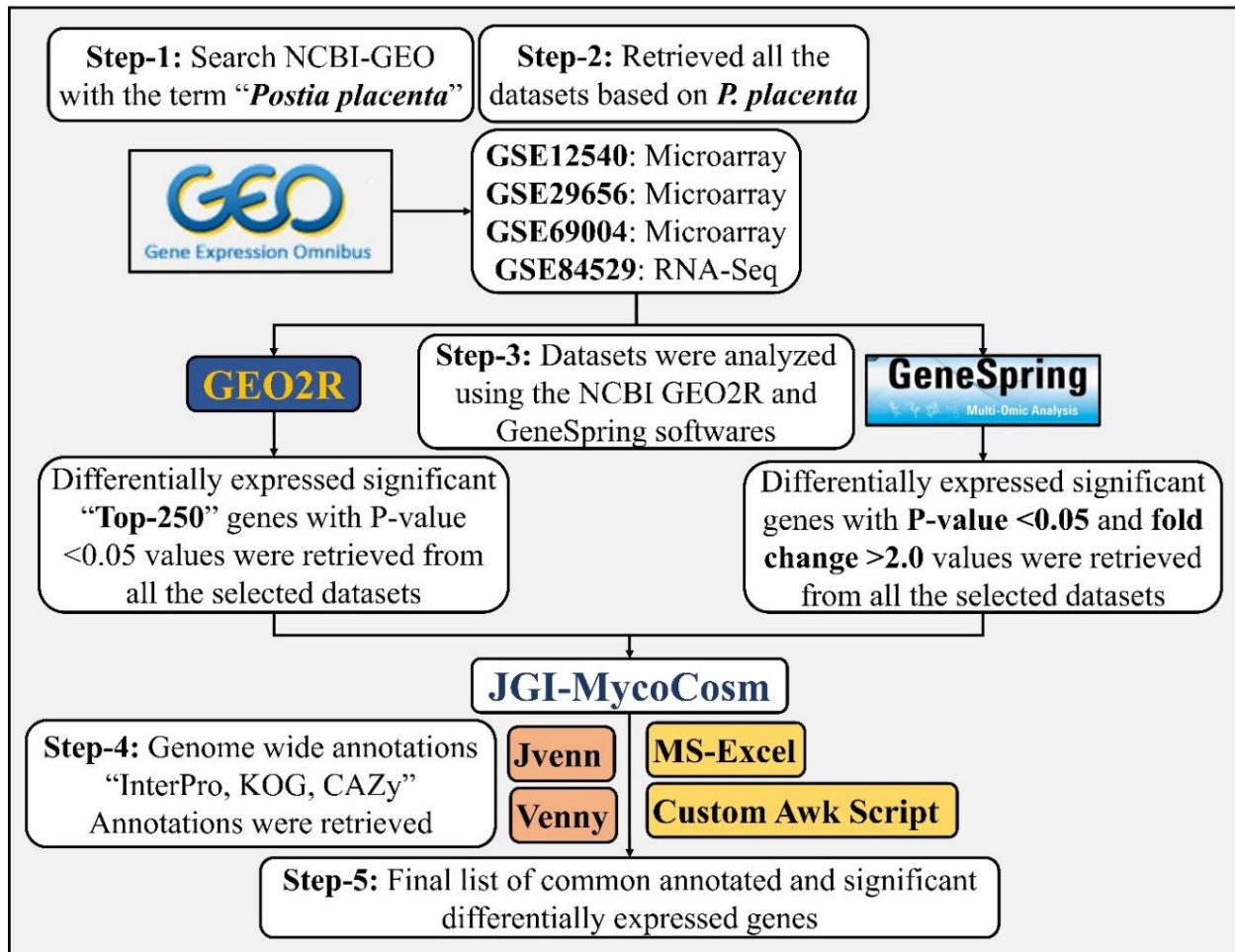


Figure 5.2: Customized step by step workflow used for the metadata analysis of *Postia placenta* gene expression datasets.

5.4. Results

The rapid cellulolytic capacity of the brown rot fungi *P. placenta* surely will be credited to the potential cellulolytic enzymes and uncompromising Fenton's reactions. Apart from these enzymes *P. placenta* also secretes a large list of aromatic compound degrading and detoxifying enzymes. Statistical analysis of the *P. placenta* gene expression datasets based on the provided experimental conditions has resulted in 5174, 7519, 6390 and 10,754 differentially expressed transcripts among GSE12540, GSE29656, GSE69004 and GSE84529 datasets respectively (Figure 5.3).

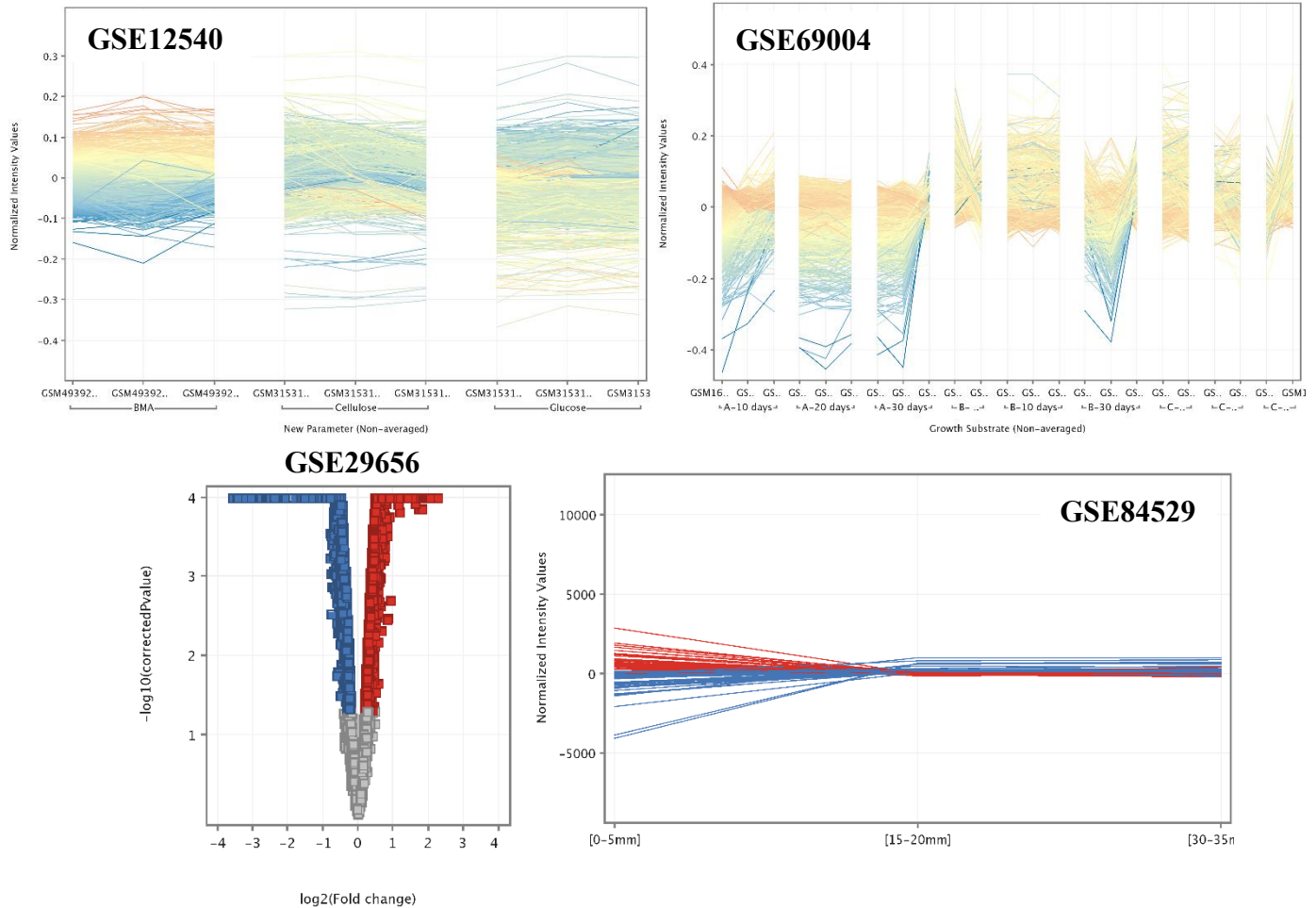


Figure 5.3: Profile plot (GSE12540, GSE69004 and GSE84529) and volcano plots (GSE29656) of the significant and differentially expressed genes among the conditions.

5.4.1. Genes Encoding for Carbohydrate Active Enzymes (CAZymes) and Metabolism

The genome of the *P. placenta* MAD-698 Rv1 harbors around 245 CAZymes, when compared to the genome of *P. chrysosporium* RP78 (450 CAZymes), several genes encoding for glycoside hydrolase and glycosyl transferase classes were reduced to 1 or absent. Genes encoding for various classes of glycoside hydrolases were found to be absent in the genome of *P. placenta* such as GH6, GH7, GH9, GH11, GH25, GH30, GH43, GH53, GH74, GH79, GH89, GH-92, GH115, GH125, GH131, GH133 and GH135. Several other GH-classes were reduced to one gene such as GH-10, GH38, GH45, GH55, GH72, GH78, GH88 and GH95 [1]. Similarly, *P. placenta* genome codes only for 24 glycosyl transferase encoding genes which were classified under GT (1), GT15 (1), GT20 (1), GT22 (1), GT66 (1), GT4 (2), GT69 (2), GT2 (7) and GT8 (8) classes and several GT-classes were found to be absent such as GT3, GT5, GT17, GT21, GT24, GT31, GT32, GT33, GT35, GT39, GT41, GT48, GT49, GT50, GT57, GT58, GT59,

GT76 and GT90 [1]. Also, genes coding for polysaccharide lyase class 8, carbohydrate esterase class 1 and carbohydrate binding module class 1, 35 and 50 were found to be absent in *P. placenta* genome [1]. The statistically significant differentially expressed genes obtained under different growth conditions were compared to find the genes which were expressed commonly. When *P. placenta* was cultured on simple culture medium containing cellulose, glucose and ball-milled aspen growth conditions in GSE12540 dataset. Several genes encoding CAZymes such as glycoside hydrolases (GH-1, 2, 3, 5, 8,10, 16, 18, 27, 28, 31, 35, 37, 51, 55, 71, 88, 95 and 128), glycosyl transferases (GT-8), carbohydrate esterase (CE-4 and CE-8), auxiliary activity (GMC oxidoreductases, ferroxidases) EXPN related proteins were commonly expressed. Similarly, when *P. placenta* was cultured on ball milled aspen (BMA) and ball milled pine (BMP), CAZymes such as glycoside hydrolase classes 16,18, 28, 55, 71, glycosyl transferases class-1, and laccase encoding genes were highly and differentially expressed. In gene expression datasets GSE69004 and GSE84529, *P. placenta* was cultured on natural plant biomass growth substrates. In GSE69004, genes encoding for CAZymes GH-5, GH-16, GH-18, GH-71 and GH-128, carbohydrate binding module CBM-18 and CBM-21 and carbohydrate esterase class CE-4 were found be commonly expressed among all the conditions of A-B-C-10 days, A-B-C-20 days and A-B-C-30 days. Similarly, in GSE84529 dataset, genes encoding for CAZymes GH-1, 2, 3, 5, 8, 12 13, 16, 17, 18, 28, 51 and 78 GT-20, CE-8, 9 and 15, CBM-13 and GMC oxidoreductases, AA-3 were found to be commonly expressed among the growth conditions (Figure 5.3). On comparison of Top-1000 differentially expressed genes from the datasets GSE12540, GSE29656, GSE69004 and GSE84529, CAZymes encoding genes for GH-2, GH-3, GH-5, GH-16, GH-18, GH-27, GH-28, GH-55, GH-71 and GH-95, GT-1 were found to be commonly expressed among the datasets.

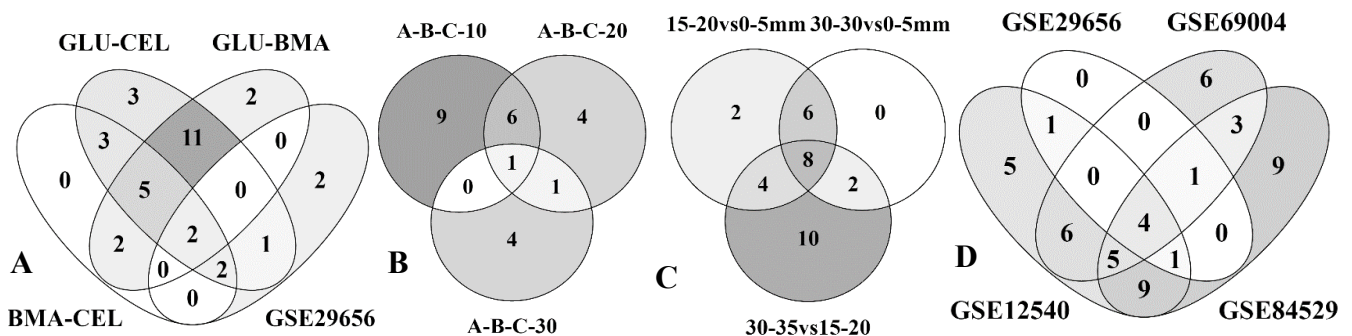


Figure 5.4: Four-way and three-way Venn diagrams showing the commonly expressed statistically significant CAZymes among the gene expression datasets A) glucose-cellulose,

glucose-BMA, BMA-cellulose and GSE29656 B) high lignin and low glucose (A) high glucose and low lignin (B) average lignin and average glucose at incubation periods of 10-days, 20-days and 30-days, C) 15mm-20mm vs 0mm-5mm, 30mm-35mm vs 0mm-5mm and 30mm-35mm vs 15mm-20mm D) GSE29656, GSE69004, GSE12540 and GSE84529 datasets.

In dataset GSE12540, when *P. placenta* was cultured on glucose, cellulose and ball milled aspen several glycoside hydrolases involved in cellulose and hemicellulose degradation were found to be highly expressed. Genes encoding for cellulases GH-1, GH-3, GH-5, GH-12, GH-16 were found to be highly up-regulated in ball milled aspen followed by cellulose growth mediums. Similarly, genes encoding for hemicellulases GH-10, GH-27, GH-31, GH-35, GH-47, GH-51, GH-55 and GH-95 were also found to be up-regulated mostly in ball milled aspen followed by cellulose growth mediums. Based on the previous reports it is known that microorganisms release several carbohydrate esterases for the deacetylation of carbohydrates, genes encoding for acetyl xylan esterase CE-4 and pectin methyl esterase CE-8 and carbohydrate esterase type-B were highly expressed in ball milled aspen followed by cellulose growth medium. We have observed that in dataset GSE29656, where *P. placenta* was cultured on complex plant cell wall materials of ball milled aspen and ball milled pine, out of top-1000 differentially expressed genes only few CAZymes encoding genes were found. Genes encoding for glycosyl transferase class-1, glycoside hydrolases GH-18 (Pp1118230), GH-28 were highly up regulated in ball milled pine, GH-18 (Pp11107968), and GH-28, in ball milled aspen GH-16 GH55 and GH71, carboxyl esterase type B and laccase (AA1) were found to be highly expressed.

When cultured on composite natural plant biomass growth substrates (GSE69004), containing chemically modified *Populus trichocarpa* cell wall components. We have observed that the gene expression patterns of *P. placenta* fluctuated upon different incubation periods. The top-1000 genes obtained for the 10-days incubation period of all the three conditions showed genes encoding for cellulases GH-3, GH-5, GH-16, CBM-18 and CBM21, hemicellulases GH-2, GH-27, GH-47, GH-55, GH-95 and CE-4 were highly expressed in low lignin-high glucose and average lignin-average glucose conditions. While with 20- days incubation period samples genes encoding for cellulases GH-5, GH-16, CBM-18, CBM-21 and hemicellulases GH-35, GH-47, CE-4. Similarly, 30-days incubation samples genes coding for hemicellulases GH-37, GH-55 along with GH-16, GH-128 and GT-20 encoding genes.

5.4.2. Enzymes Coding for Fenton's Reaction

Absence of genes encoding for lignin (LiP) and manganese peroxidase (MnP) is one of the major reason for minimal the lignin degrading abilities of *P. placenta*. Interestingly, *P. placenta* is completely dependent on the hydroxy radical (OH^{*}) and ferric ion (Fe³⁺) generated through Fenton's reaction for the process of lignin modification (or) depolymerization [6, 35]. Previous studies have reported that Fenton's reactions in wood rotting basidiomycetes is deployed by three different mechanisms a) cellobiose dehydrogenase based reactions b) small glycopeptide catalyzed reactions generating hydroxy radicals c) cyclic redox based reactions by low molecular weight redox compounds (such as quinones and oxalates) [35]. Fenton's reaction is majorly dependent on a) extracellular hydrogen peroxide generation and b) metabolite and enzyme based reduction of iron from ferric state to ferrous state (cellobiose dehydrogenase) [1]. However, absence of genes encoding for cellobiose dehydrogenase makes *P. placenta* solely dependent on the extracellular fungal metabolites such as hydroquinones and low molecular weight glycopeptides to catalyze the reduction of iron [1]. Hydrogen peroxide generating enzymes mainly glucose oxidase, copper radical oxidase, alcohol oxidase, glycolate oxidase and polyphenol oxidase were observed among the *P. placenta* gene expression datasets. Similarly, enzymes involved in iron reduction and homeostasis mainly quinone reductases, quinone transporters, phenylalanine ammonia lyase, ferric reductases, iron permeases and ferroxidases were significantly expressed among the datasets. In GSE12540 dataset, genes encoding for copper radical oxidase, 1,4-benzoquinone reductase, polyphenol oxidase, phenylalanine ammonia lyase, iron permease, ferroxidase and multicopper oxidase were found to be highly expressed in cellulose, ball milled aspen growth substrates. Similarly, in GSE29656 dataset genes encoding copper radical oxidase, polyphenol oxidase, phenylalanine ammonia lyase, iron permease, multicopper oxidase and ferroxidase were found to be differentially expressed in ball milled aspen and 1,4-benzoquinone reductase, ferric reductase encoding genes were expressed in ball milled pine growth substrates. Results obtained from GSE69004 dataset showed that genes encoding for GMC oxidoreductases, iron permease were highly expressed in A 10, 20, 30 (low glucose-high lignin) growth substrates, alcohol oxidase, glucose oxidase and glycolate oxidase were highly expressed in low lignin- high glucose and average glucose-average lignin growth substrates.

5.4.3. Lignin Degrading and Detoxifying Enzyme systems

Due to the lack of genes encoding for ligninolytic peroxidases, the process of lignin modification or depolymerization is majorly dependent on other auxiliary activity enzymes such as multicopper oxidases, laccases and oxidoreductases. Genes encoding for auxiliary enzymes such as ferroxidase, multicopper oxidase, laccase, GMC Oxidoreductases, alcohol oxidases, glyoxal oxidase 1,4-benzoquinone reductases, polyphenol oxidase and lytic polysaccharide monooxygenases were found to be highly expressed under different culture conditions as mentioned above. Apart from the auxiliary enzymes several other enzymes involved in aromatic compound degradation and metabolism were also found to be highly expressed among the datasets. In GSE12540 dataset, genes encoding for aromatic ring hydroxylase, catechol dioxygenase, intradiol ring dioxygenase, Tannase and feruloyl esterase, taurine catabolism dioxygenase, 2-nitropropane dioxygenase, 4-coumarate coenzyme A ligase, O-methyl transferase, FAD-linked oxidoreductase, alpha aminoadipate reductase, aldo/keto reductases, alcohol dehydrogenases, zinc alcohol dehydrogenases were found to be highly expressed in cellulose and ball milled aspen growth substrates. In GSE29656 dataset, genes encoding for aldehyde dehydrogenase, acyl-coA thioesterase, homocitrate synthase, flavodoxin, epoxide hydrolase, HMG-CoA lyase, isoflavone reductase, FAD monooxygenase, NADH flavin dependent oxidoreductase, short chain dehydrogenase, transketolase, terpene synthase, zinc alcohol dehydrogenase, 4-coumarate coA ligase, O-methyl transferase and various genes encoding for cytochrome P450 class monooxygenases were highly expressed in ball milled pine growth substrate. However, genes encoding for esterases, Tannase and feruloyl esterase, phenylalanine ammonia lyase, polyphenol oxidase, laccase and copper radical oxidase were found to be highly expressed in ball milled aspen growth substrates. In GSE69004 dataset, where *P. placenta* was cultured on *Populus trichocarpa* with chemically distinct growth substrates genes encoding for aldo keto reductases, cytochrome c, several genes coding for cytochrome P450, diene lactone hydrolases, E-class P450 group IV, induced cAMP protein, iron permease, NADH flavin oxidoreductase, class-I auxiliary activity enzymes were found to be expressed in high lignin-low glucose conditions (A10, A20 and A30). Several genes encoding for acyl-CoA dehydrogenase, carboxyl esterase type B, several genes encoding for cytochrome c, cytochrome b5, diene lactone hydrolase, esterase, haloacid dehalogenase/epoxide hydrolase, mandelate racemase, Metallophosphoesterase, FAD monooxygenase, oxidoreductase, Thiolase, UbiA prenyltransferase, alcohol oxidase, 2-nitropropane

dioxygenase, flavin monooxygenase and peroxidase were found to be highly expressed among low lignin-high glucose and average lignin-average glucose growth conditions.

In GSE84529 dataset among the top 1000 differentially expressed genes, we have commonly observed genes encoding for aromatic ring hydroxylase, 2-nitropropane dioxygenase, flavoprotein monooxygenase, carboxylesterase type B, esterase, flavoprotein monooxygenase, aldo/keto reductase, zinc alcohol dehydrogenase, short chain dehydrogenase/reductase, 2OG-Fe(II) oxygenase superfamily, isoflavone reductase, GMC oxidoreductase, generic and O-methyltransferase, UbiA prenyltransferase and D-isomer specific 2-hydroxyacid dehydrogenase in the early (0mm to 5mm) and late (15mm-20mm) conditions. Along with above mentioned enzymes several genes encoding for cytochrome P450 monooxygenases were found to be highly expressed in early (0mm to 5mm) and few in late (15mm to 20 mm) conditions. According to Zhang et al (2016), lignin oxidizing and Fenton chemistry related enzymes required for the production of H₂O₂ and iron reduction and homeostasis were found to be highly expressed during early decay (0mm to 5mm) process [21]. This study also reported that during early decay phase (0mm to 5mm) genes coding for lignin oxidation, Fenton chemistry, cytochrome P450 enzymes were highly expressed when compared to late decay phases (15mm to 20mm and 30mm to 35mm) where genes coding for various CAZymes, sugar metabolism, Fenton chemistry and aldo keto reductases were found to be differentially expressed [21] (Figure 5.4).

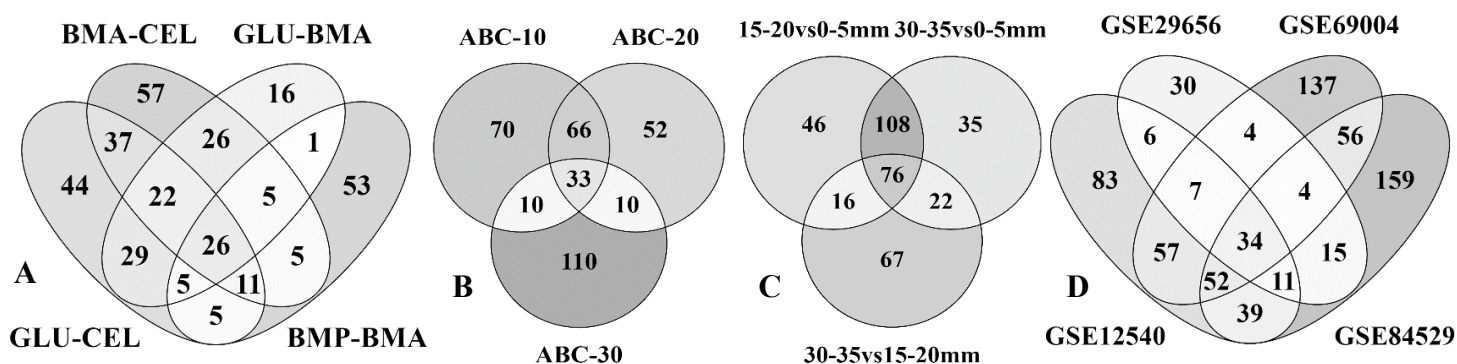


Figure 5.5: Four-way and three-way Venn diagrams showing the commonly expressed statistically significant enzymes among the gene expression datasets A) glucose-cellulose, glucose-BMA, BMA-cellulose and GSE29656 B) high lignin and low glucose (A) high glucose and low lignin (B) average lignin and average glucose at incubation periods of 10-days, 20-days and 30-days, C) 15mm-20mm vs 0mm-5mm, 30mm-35mm vs 0mm-5mm and 30mm-35mm vs 15mm-20mm D) GSE29656, GSE69004, GSE12540 and GSE84529 datasets.

The comparison of significant differentially expressed gene lists obtained from all the datasets have revealed that overall 40 genes encoding for CAZymes were found to be common with each dataset sharing few common CAZymes. Common auxiliary activity enzymes include laccase (Psp11_111314), ferroxidase (Psp11_109824), GMC oxidoreductase (Psp11_92024 and Psp11_27847). Similarly, common cellulolytic enzymes include GH-3 (Psp11_128500 and Psp11_46915), GH-5 (Psp11_121831, Psp11_116199, Psp11_121713 and Psp11_57386), GH-16 (Psp11_112941, Psp11_62300, Psp11_94601 and Psp11_51311). Finally, common hemicellulolytic enzyme classes include GH-27 (Psp11_120395), GH-31 (Psp11_117029), GH-35 (Psp11_127993), GH-51 (Psp11_100251 and Psp11_94557), GH-55 (Psp11_105490) and GH-95 (Psp11_105952). Several genes encoding for aromatic compound degrading and metabolizing enzymes were found to be commonly expressed among the datasets. Importantly, genes encoding for cytochrome P450 monooxygenases (Psp11_9739, Psp11_98329, Psp11_89741, Psp11_92219 and Psp11_21733), aromatic ring hydroxylase (Psp11_90902, Psp11_22746 and Psp11_23052), dioxygenases including 2-nitropropane dioxygenase (Psp11_24756, Psp11_28683), intradiol dioxygenase (Psp11_34850) and taurine catabolism dioxygenase (Psp11_89958) (Figure 5.5).

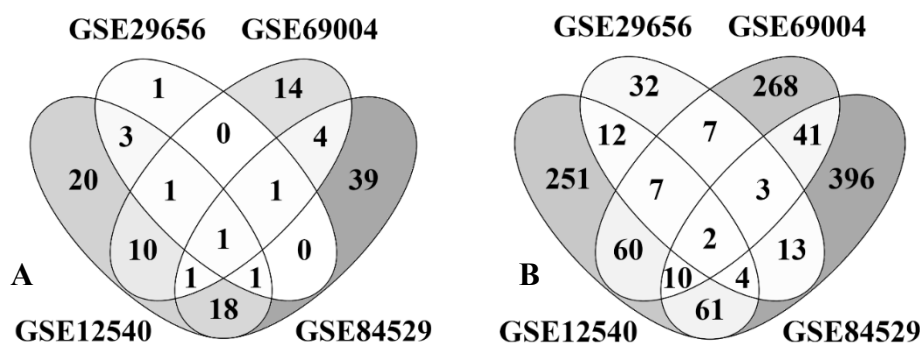


Figure 5.6: Four-way Venn diagrams showing the commonly expressed statistically significant protein-Ids among the gene expression datasets A) CAZymes among all the datasets B) All InterPro-IDs among all the datasets.

5.4.4. Genes Encoding for Information, Storage and Processing processes

Expression of the lignocellulolytic enzymes are controlled and regulated by genes encoding for the information, storage and processing processes. The eukaryotic orthologous group have classified these genes into RNA processing and modification (KOG: A), chromatin structure and dynamics (KOG: B), Translation, ribosomal structure and biogenesis (KOG: J), Transcription (KOG: K) and Replication, recombination and repair (KOG: L) groups. We have separated the differentially expressed gene list in to the KOG groups A, B, J, K and L using the

KOG classification list retrieved from MycoCosm database for *P. placenta*. We have observed that a total of 308, 496, 357 and 648 transcripts were significantly expressed among GSE12540, GSE29656, GSE69004 and GSE84529 datasets respectively (Figure 5.7). Total of 37 (A), 15 (B), 89 (J), 38 (K) and 36 (L) were found to be commonly expressed among the gene expression datasets (Figure 5.7).

Results obtained from these gene expression studies has revealed that total of 15 proteins encoding for RNA processing and modification were found to be common and significantly expressed RNA helicase nonsense mRNA reducing factor, splicing coactivator SRm160/300, RNA directed polymerase QDE1, polyadenylation factor, dsRNA-specific nuclease Dicer and ribonuclease, mRNA capping enzyme, mRNA decay protein, RNA helicase BRR2, RNA binding protein p54nrb and polyadenylate binding protein. Similarly, four genes encoding for chromatin structure and dynamics structural maintenance of chromosome protein 4, Zn-finger-MYND type, chromosome condensation complex and ubiquitin component Cue were common and significantly expressed. 21 genes encoding for translation, ribosome biogenesis and structure were common and significantly expressed 40S ribosomal protein S6, SA(P40), 60S ribosomal protein L10, L3, L28, Ribosomal proteins L7A, L18e, L23, L5, S25, L6E, S14, S10, L32, translation initiation factors eIF-5B, eIF-3a, eIF-4G and mRNA export factor. 9 genes encoding for chromodomain helicase, transcription initiation factor, HMG1/2, nuclear receptor coregulator SMRT, transcription regulator XNP, Ssu72, bZIP, TGF-beta and SMAD protein, HLH-transcription factor was commonly expressed in transcription class. Finally, in replication, recombination, repair class 8 genes encoding for 3-methyladenine DNA glycosidase, exonuclease, replication factor C, apurinic/apyrimidinic endonuclease, DNA replication-MCM7 factor, A/G-specific DNA glycosylase, excision repair protein RAD14/XPA and Nudix family hydrolase were commonly expressed among all the datasets. Majority of the genes encoding for information storage and processing were highly expressed in glucose, cellulose, ball milled aspens, A (low glucose-high lignin) and 0mm05mm and 30mm-35mm growth substrates.

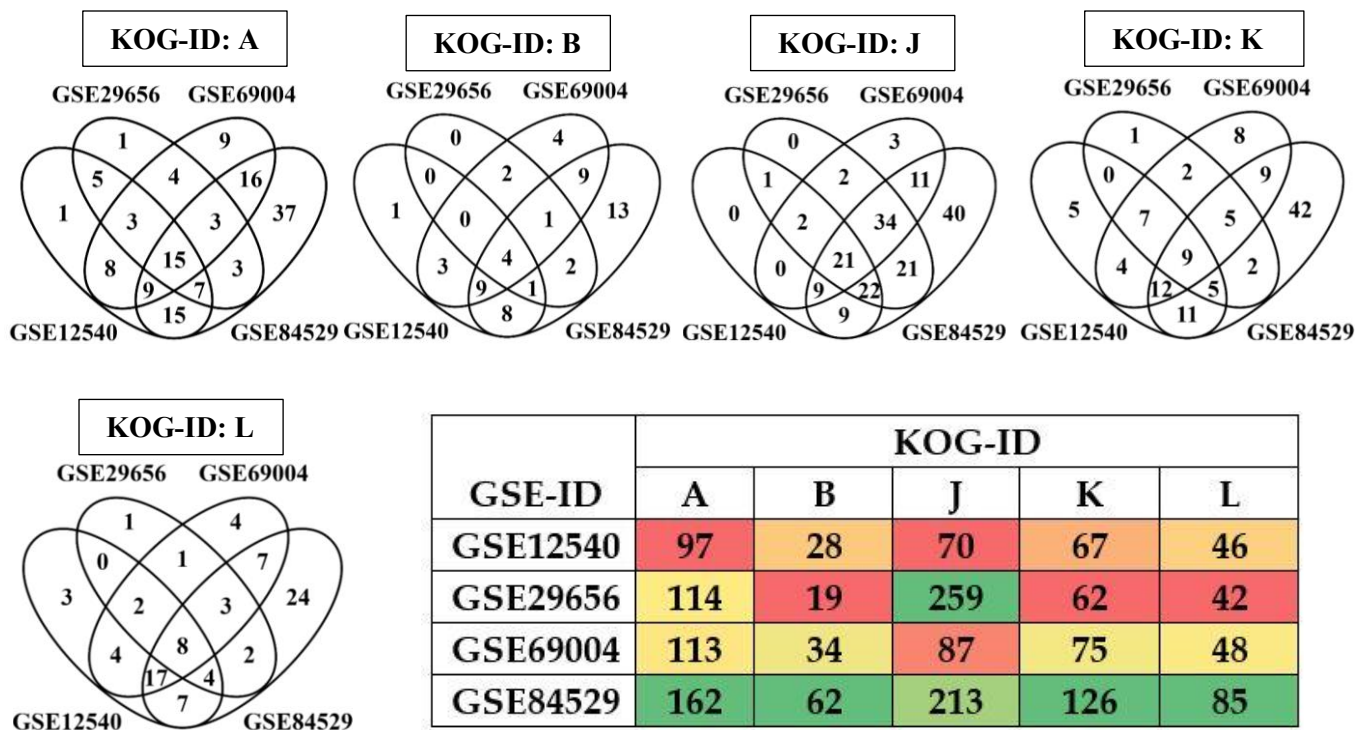


Figure 5.7: Four-way Venn diagrams showing the commonly expressed statistically significant genes encoding for KOG groups **A** (RNA Processing and modification), **B** (Chromatin structure and dynamics), **J** (Translation, ribosomal structure and biogenesis), **K** (Transcription) **L** (Replication, recombination and repair) and heatmap showing total list of significant information storage and processing groups among the datasets.

5.5. Discussion

Genome wide evolutionary studies conducted earlier have revealed that during transition from white-rot to brown-rot, basidiomycete fungi have experienced an extensive gene loss. The whole genome studies has revealed that *P. placenta* MAD-698-R v1 has suffered complete reduction of various genes coding for glycoside hydrolases such as GH-6, GH-7, GH-10, GH-11 and GH-61 [1]. Though the genome of *P. placenta* lacks genes encoding for various essential lignocellulolytic CAZymes the decay patterns of *P. placenta* suggest the rapid depolymerization of cellulose and hemicellulose units through modifying the lignin units. During initial degradation process *P. placenta* significantly reduces the degree of polymerization (DP) of cellulose from 1800-2000 glycosyl units to 150-200 units [36]. However, due to its large structure cellulases cannot enter through the pores of wood, thus brown rot fungi employ other oxidoreductases and Fenton's system for the depolymerization of plant cell wall structures. Unlike other cellulose degrading microorganisms, *P. placenta* lack genes coding for conserved CBH1(GH7) and CBH2 (GH6) exo-cellobiohydrolases, carbohydrate binding module class-1 and cellulase binding endoglucanases. Hence, *P. placenta* must

accomplish the process of cellulose and hemicellulose degradation through a set of endoglucanases, exo-glucanases, β -glucosidases and several other hemicellulases. Previous gene expression studies of *P. placenta* conducted on cellulose reported various genes coding for hemicellulases, laminarinases, chitinases, but it is still not clear whether these enzymes can attack crystalline cellulose [1]. It is well known that *P. placenta* genome lacks the sequences encoding for major ligninolytic such as lignin (LiP), manganese (MnP) and versatile (VP) peroxidases. However, the *P. placenta* genome was found to possess gene sequences coding for a low redox potential peroxidase (Protein-ID: 50226) which lacks the Mn(II) oxidation sites but possess a tryptophan residue which is involved in substrate oxidation similar to LiP and VP [1]. These genomic evidences support the stronger involvement of Fenton's chemistry during depolymerization or modification of the lignocellulosic biomass.

The process of Fenton's chemistry in *P. placenta* is employed by a set of genes encoding for enzymes involved in hydrogen peroxide generation and iron reduction and homeostasis. Generation of non-specific highly powerful oxidative hydroxy radicals with a half-life period around 9 to 10 seconds by Fenton's chemistry will play a crucial role in the depolymerization and modification of lignocellulosic components of plant cell wall [5, 6]. Studies have reported that Fenton's chemistry is highly dependent on three factors they are iron (Fe^{3+}), hydrogen peroxide (H_2O_2) and pH [5]. Earlier studies have reported that brown rot fungi might also depend on the insoluble iron oxyhydroxide complexes for the uptake of iron and reduction [6, 7]. We have commonly observed the genes encoding for hydrogen peroxide generating and iron homeostasis and reduction among various datasets. Expression of genes involved in the process of iron reduction and homeostasis such as ferroxidase, phenylalanine ammonia lyase, catalase, iron permease, flavin monooxygenases, dioxygenases quinone transporters and reductases proves the importance of Fenton's system. According to Berry et al (1997), brown rot fungi secretes oxalic acid extracellularly which might be involved in solubilizing the iron from its naturally occurring iron oxyhydroxide complexes of the plant wood [6]. As reported by Suong oh et al (2006), dissolution of iron oxides (protonating OH groups resulting in weak Fe-O bond) is significantly influenced by the cellular pH conditions [37]. Several theories have been proposed on the process of iron chelation with oxalic acid under lower pH conditions by forming soluble and stable oxalate-iron complexes which further diffuse into the plant cell wall through wood lumen [5-7]. Genes encoding for alcohol and methanol oxidases which are also involved in the extracellular

production of H₂O₂ were found to be highly expressed among the datasets. Studies have also reported that low molecular weight compound such as 2,5-dimethoxyhydroquinone are employed during the process of lignin depolymerization [38] (Figure 5.8).

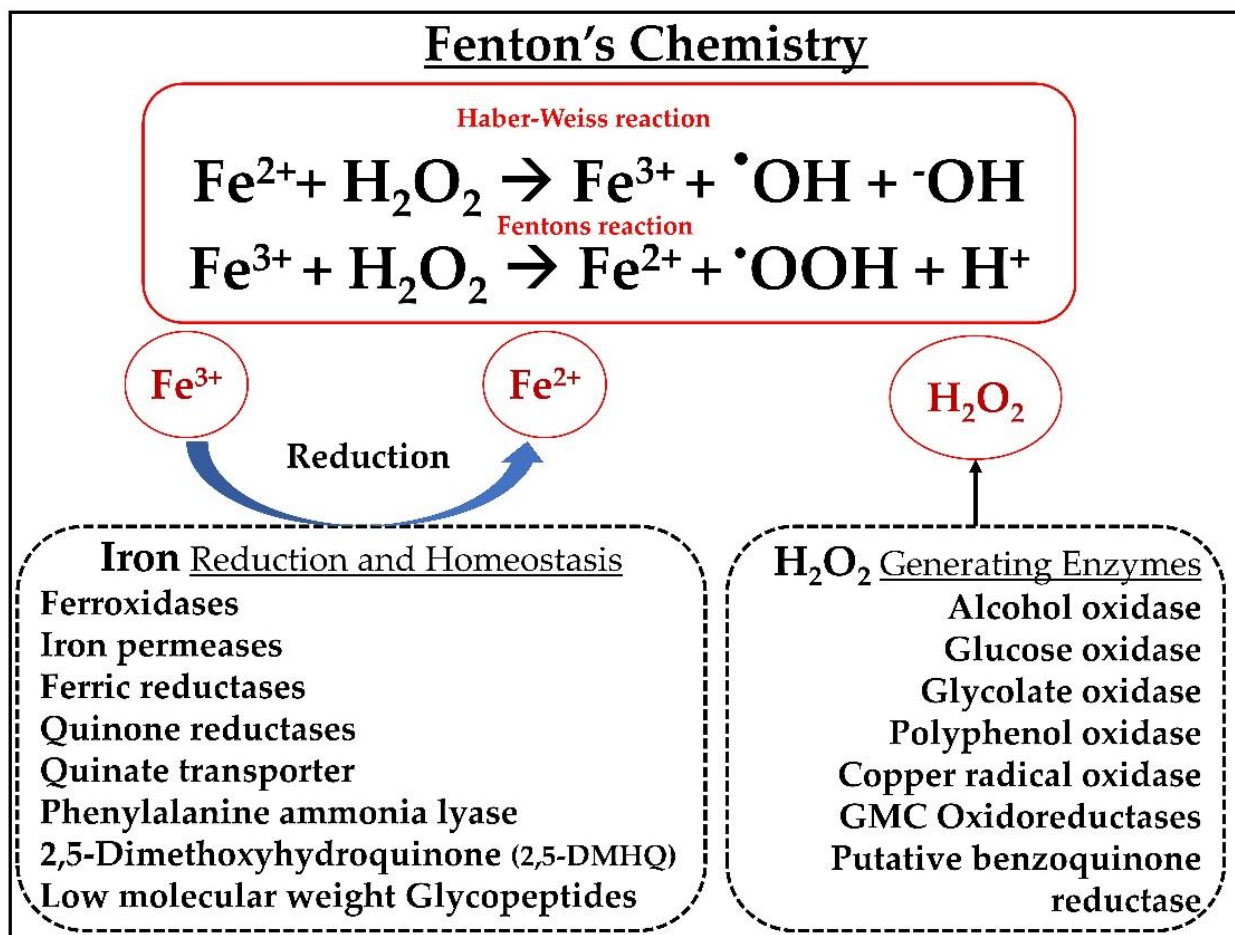


Figure 5.8: Tentative representation of Fenton's reaction system observed in *P. placenta* mainly includes H₂O₂ generating and Iron reduction-homeostasis enzyme systems found to be commonly observed among the gene expression datasets.

However, till date the exact reaction mechanisms involved during the depolymerization or modification of lignocellulosic biomass is not well understood. In our present analysis, we have clearly observed significant expression of genes encoding for enzymes involved in Fenton's reaction. Apart from the above mentioned enzymatic fungal Fe³⁺ reductants, non-enzymatic low molecular weight iron reductants such as 2,5-dimethoxyhydroquinone (2,5-DMHQ) were also reported to be significantly involved in depolymerization of lignocellulose components. Previous studies have observed the presence of 2,5-DMHQ in the cultures of *Gleophyllum trabeum*[38], *Postia placenta* [19], *Serpula lacrymans* [11] and also in various cultures of *Gleophyllum* species [11, 12, 39]. The

low molecular weight 2,5DMHQ reduces fungal Fe³⁺ to Fe²⁺ by simultaneously generating a radical of semiquinone, which can reduce oxygen to *OOH later upon dismutation reaction it can generate H₂O₂ or it can reduce an extra Fe³⁺ to Fe²⁺ resulting in quinone formed due to original Fe³⁺ reductant [9, 39]. Along with these mechanisms there are other two iron reduction mechanisms a) cellobiose dehydrogenase dependent and b) low molecular weight glycopeptides. However, the genome of *P. placenta* lacks the genes encoding for cellobiose dehydrogenase (CDH) which rules out the CDH dependent iron reduction [1]. According to Martinez et al (2009), Low molecular weight glycopeptides were found to be expressed in the cultures of *P. placenta* [1]. Based on the results obtained in our present study, we have listed the common significant enzymes obtained among all the datasets which were listed in the Table 5.2.

Table 5.2: Common differentially expressed significant class of enzymes among different growth conditions of gene expression datasets.

Enzyme class	Commonly Expressed <i>P. chrysosporium</i> Genes among the datasets
Cellulose Degradation	GH-1, GH-3, GH-5, GH-12, GH-16 and GH-45
Hemicellulose Degradation	GH-10, GH-27, GH-31, GH-35, GH-47, GH-51, GH-55, GH-78 and GH-95
Carbohydrate Metabolism	GH-2, GH-13, GH-15, GH-17, GH-18, GH-20, GH-23, GH-28, GH-37, GH-38, GH-63, GH-71, GH-71, GH-72, GH-85 and GH-88
Lignin Degradation	Auxiliary Activity Enzymes: Laccase, Ferroxidase, Multicopper oxidase, Low redox potential lignin peroxidase, Glucose-Methanol-Choline (GMC) oxidoreductase, Alcohol oxidase, Glyoxal oxidase, Lytic polysaccharide monooxygenase, Chloroperoxidase Aromatic compound degrading: Intradiol dioxygenases, Aromatic ring hydroxylase, Epoxide hydrolase, Cytochrome P450 Monooxygenase, Alcohol dehydrogenase, Dioxygenase, 2-nitropropane dioxygenase, Flavin containing monooxygenase, Iron reductases, Catalase, Alcohol/methanol oxidases, Haloacid dehalogenase, Oxidoreductase, Tannase and feruloyl esterase, Esterase/ lipase/thioesterase, Short-chain dehydrogenase/reductase, D-isomer specific 2-hydroxyacid dehydrogenase, Beta-ketoacyl synthase, 2-oxo acid dehydrogenase, Aldo/keto reductase, Aldehyde dehydrogenase, Alkyl hydroperoxide reductase, FAD-linked oxidase, Thiolase, Carbohydrate esterases, Glycosyl transferases
Fenton's Chemistry	Ferroxidase, Ferric reductase, Iron permease, Quinone reductase, Quinone transporters, phenylalanine ammonia lyase, low molecular weight glycopeptides, alcohol oxidase, glucose oxidase, glycolate oxidase, polyphenol oxidase, copper radical oxidase, 1, 4-Benzoquinone reductase

Detoxification & Stress Responsive enzymes	2OG-Fe(II) oxygenase superfamily, Argonaut and dicer complex, Cytochrome P450, E-class P450, Thioesterase, flavin monooxygenase, Glutathione-S-transferase, Thiolase, Ubiquitin, Universal stress protein, Flavodoxin, Epoxide hydrolase, Isoflavone reductase, FAD monooxygenase, NADH flavin oxidoreductase, Dienelactone hydrolase, Thioredoxin, Thaumatin, Ferredoxin, Beta lactamase, Catalase, UbiA prenyltransferase, Universal stress protein, Ubiquitin system component, Cytochrome b5, Cytochrome c, amidohydrolase
---	--

Wood decaying fungi majorly depend on the holocellulose for the carbon and energy requirements, however genome wide studies have revealed that cellulolytic systems of *P. placenta* and other brown rot fungi vary significantly when compared with popular fungi such as *Trichoderma reesei* and *Phanerochaete chrysosporium* (which employ an efficient set of hydrolytic enzymes involving endo and exo glucanases, beta-glucosidases and cellobiose dehydrogenases [5]. Despite the loss of various genes encoding for glycoside hydrolases in the genome *P. placenta* has found to cause rapid and significant depolymerization of holocellulose by significantly decreasing the degree of polymerization to 150-200 through reducing the amorphous regions of cellulose. Brown rot fungi likely opens the cell wall structure by extensive removal of hemicelluloses resulting in a 20% of weight loss which increases the accessibility of the cellulose [5, 12, 40, 41]. According to Cohen et al (2005), some fungi the absence of cellobiose dehydrogenase is compensated by the expression of functionally similar endoglucanase [42]. These earlier proposed reports suggest that hydrolytic enzymes including endoglucanases and hemicellulases contribute to the complete holocellulose depolymerization released after primary attack of the cell wall [5]. As reported by Ratto et al (1997), higher hydrolysis rates were observed when *P. placenta* and *T. reesei* were cultured on spruce sawdust containing growth substrates, after the initial Fenton reaction chemistry [13]. Ratto et al (1997) have also reported that initial oxidation will significantly increase the hydrolysable nature of cellulose by endoglucanases secreted by brown rot fungi [13]. Jung et al (2015), have conducted a study by mimicking Fenton reaction conditions on rice straw degradation, results obtained in this study have showed that composition of lignin and xylan present in the rice straw was significantly reduced by the Fenton reaction systems [43]. Results obtained in our present analysis is in accordance with previous studies and supports the highly dependent nature of brown rot fungi on Fenton's reaction system for the process of lignocellulose degradation. We have also observed that genes encoding for hemicelluloses were highly up-regulated during the cellulolytic

conditions, when cultured on both simple synthetic and complex natural growth substrates (Figure 5.9).

Studies conducted in the past have reported that reactive hydroxy radicals are involved in the polymerization and depolymerization of lignin [44, 45]. According to Goodell et al (1997), brown rot fungi exhibits similar mechanism for the process of lignin depolymerization or modification, using chelator mediated Fenton reaction [5, 6]. Various studies have strongly reported that brown rot fungi affected wood experience extensive oxidative demethylation [15, 46-48] and side chain oxidation [15, 48, 49]. During the process of lignin modification various important processes were reported earlier which includes a) partial aromatic ring cleavages [48] b) aromatic ring hydroxylation [48] c) limited side chain hydroxylation and C β -ether cleavage [49] d) formation of aryl derivative side chains (aryl-o-aryl, aryl-aryl etc.) [14]. It is well-known that unlike its counterparts *P. placenta* (brown rot fungi) do not secrete lignin degrading auxiliary enzymes such as LiP, MnP [1]. However it was reported that S2 layer of the secondary cell wall was found to contain brown rotted lignin whereas the above mentioned ligninolytic auxiliary enzymes were expected to fail in penetration and causing attack [50]. Although brown rot fungi (*P. placenta*) affected lignin undergoes the above-mentioned modifications, it was reported that lignin does not lose its polymeric nature, which proves the fact that brown rot fungi are weak lignin degraders. Studies have also reported that reactive hydroxy radicals released due to Fenton reaction system might also lead to the repolymerization of lignin [14, 15, 51-54]. These reporting's from the previous studies were observed in our study, genes encoding for various enzymes involved in aromatic compound degrading enzymes and iron reduction-homeostasis reactions of Fenton system were found to be highly expressed in *P. placenta* during the ligninolytic conditions. Along with the genes encoding for lignocellulolytic enzymes several detoxification and stress responsive related enzymes were found to be highly expressed during the ligninolytic conditions suggesting the release of highly toxic intermediates and conditions due to lignin modification.

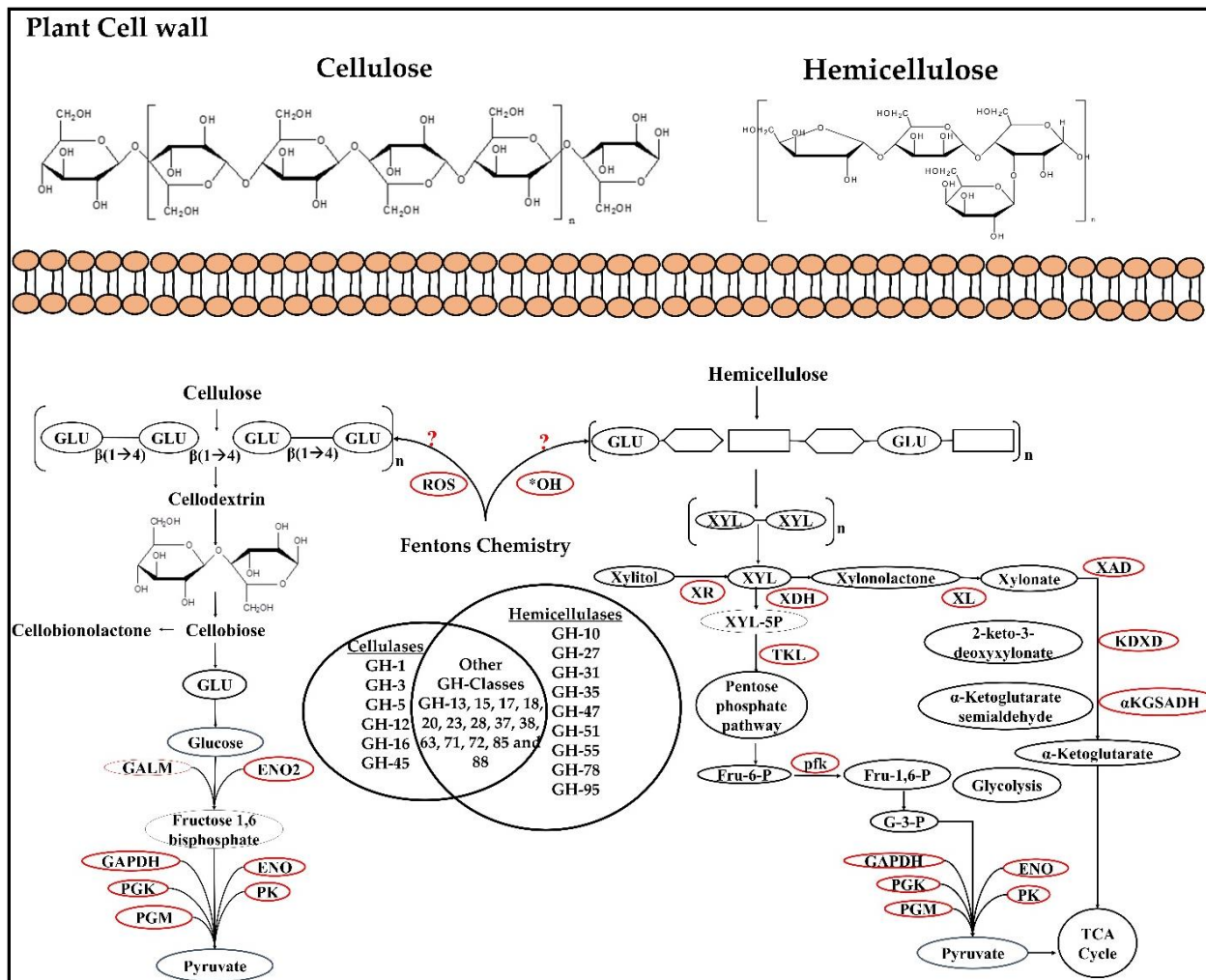


Figure 5.9: Tentative network of genes involved in coding for enzymes involved during cellulose and hemicellulose degradation.

Genes classified under information storage and processing are involved in controlling the expression of lignocellulolytic and detoxification enzymes. Present day genome of *P. placenta* encodes total of 1035 genes encoding for information storage and processing (KOG group) which are further classified into 296 (KOG: A), 103 (KOG: B), 253 (KOG: J), 244 (KOG: K) and 139 (KOG: L) [1]. The eukaryotic KOG process RNA processing and modification decides the cellular fate of transcripts. Several studies conducted in the past have clearly explained that mRNA is considerably pretreated by 5' capping, splicing and 3' end processing before the mRNA gets exported. The above reported commonly expressed genes RNA helicase, RNA polymerase QDE1 and RNA binding protein are involved in the initiation process. Genes encoding for mRNA guanylyl and methyl transferases are involved in 5' end capping, splicing coactivator, dsRNA specific nuclease Dicer and related ribonuclease, non-

sense mediated mRNA decay and polyadenylation factor complex protein are involved in elongation and termination of the transcripts [55-57]. Together these genes directly control the expression, localization, stability and export of the mature transcripts. Organization of large eukaryotic genomes into stable chromosomes requires a wide range of enzymes classified under chromatin structure and dynamics class. Maintaining the structural dynamics of eukaryotic chromatin and organizing it in stable chromosomes is of primary importance for the cell otherwise, it might bring various restrictions in various DNA related processes. Expression of histone methyl and acetyl transferases, chromatin condensation and SWI-SNF remodeling complex commonly among the gene expression datasets, expression of these genes is directly involved in regulation and control of various plant cell wall degrading enzymes. Similarly, genes involved in base excision repair, nucleotide excision repair, mismatch repair and homologous recombination were found to be commonly expressed among all the datasets. Results obtained from the gene expression studies that when *P. placenta* was cultured on natural plant biomass growth substrates expression of plant cell wall degrading and detoxifying enzymes are strongly influenced and regulated at various cellular and nuclear levels. However, the research must be conducted to understand the exact cellular mechanisms required for the expression of plant cell wall degrading enzymes. Previous studies have revealed that ascomycetes code for a wide range of transcription factors out of which zinc binuclear cluster proteins (XYR1, ACE2, XlnR and XLR1), multimeric proteins (HAP2, HAP3, HAP5), two zinc binuclear cluster (CLR-1/2), three Cys2-His2 Zn-finger (PacC) are well known as positive regulators of cellulolytic and hemicellulolytic enzymes. Whereas, ACE1, CRE1, CRE-A, CRE-B and CRE-C are considered as negative regulators of cellulolytic and hemicellulolytic enzymes [58]. Genes coding for the afore-mentioned transcription factors were not significantly expressed among the *P. placenta* gene expression datasets, however genes encoding for other transcription factors have been observed among the datasets. Studies must be conducted to understand the functional involvement of the common significantly expressed information storage and processing genes [34].

5.6. Conclusion

Genome wide transcriptome and proteome studies of *P. placenta* conducted in the last decade have revealed about the functional properties of lignocellulolytic enzymes and various CAZymes. We have performed an extensive metadata analysis on the *P. placenta* gene expression datasets (GSE12540, GSE29656, GSE84529 and GSE69004) to understand the

common gene expression patterns involved during the lignocellulose degradation. For the first time, we have reported the genome wide common expression patterns of *P. placenta* when cultured on different plant biomass growth substrates. We have clearly observed the strong dependency of *P. placenta* on the Fenton's reaction system for the modification and degradation of lignocellulosic components. The higher expression of genes encoding for hemicellulolytic glycoside hydrolases even under cellulolytic conditions proves the synchronized action of both cellulolytic and hemicellulolytic hydrolytic enzymes. Based on the reports from previous plant biomass degradation studies and the results obtained in our present study clearly show that, *P. placenta* exhibits stronger ability to degrade cellulose and hemicellulose by selectively modifying lignin. Even though *P. placenta* has experienced an evolutionary loss of genes encoding for CAZymes, it is well known for its rapid cellulolytic and lignin modifying abilities which can be credited to its outstanding oxidizing capacities. The significant differentially expressed genes from all the datasets were compared and the protein Id's of the common highly expressed genes were also reported. However, level of gene expression does not completely explain the lignocellulose degradation mechanisms employed by the *P. placenta*, as the gene expression studies are the snapshot of the cell at a particular moment. Simultaneously, several functionally uncharacterized proteins were also found to be highly expressed among the *P. placenta* datasets. Further studies must be conducted to exactly understand and reveal the functional roles of these commonly observed significant genes and various other uncharacterized proteins. Understanding the rapid degradation of cellulosic and hemicellulosic units by brown rot fungi will significantly benefit the growing biofuel and biorefining industries, through exploring its unprecedented oxidizing abilities.

References

1. Martinez D, Challacombe J, Morgenstern I, Hibbett D, Schmoll M, Kubicek CP, et al. Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proceedings of the National Academy of Sciences*. 2009; 106: 1954-9.
2. Schilling JS, Kaffenberger JT, Liew FJ, Song Z. Signature wood modifications reveal decomposer community history. *PloS one*. 2015; 10: e0120679.
3. Binder M, Hibbett DS, Larsson KH, Larsson E, Langer E, Langer G. The phylogenetic distribution of resupinate forms across the major clades of mushroom-forming fungi (Homobasidiomycetes). *Systematics and Biodiversity*. 2005; 3: 113-57.
4. Hibbett DS, Donoghue MJ. Analysis of character correlations among wood decay mechanisms, mating systems, and substrate ranges in homobasidiomycetes. *Systematic biology*. 2001: 215-42.
5. Arantes V, Goodell B. Current understanding of brown-rot fungal biodegradation mechanisms: a review. *Deterioration and protection of sustainable biomaterials*. 2014; 1158: 3-21.

6. Goodell B, Jellison J, Liu J, Daniel G, Paszczyński A, Fekete F, et al. Low molecular weight chelators and phenolic compounds isolated from wood decay fungi and their role in the fungal biodegradation of wood. *Journal of Biotechnology*. 1997; 53: 133-62.
7. Xu G, Goodell B. Mechanisms of wood degradation by brown-rot fungi: chelator-mediated cellulose degradation and binding of iron by cellulose. *Journal of biotechnology*. 2001; 87: 43-57.
8. Hirano T, Tanaka H, Enoki A. Extracellular substance from the brown-rot basidiomycete *Tyromyces palustris* that reduces molecular oxygen to hydroxyl radicals and ferric iron to ferrous iron. *Journal of the Japan Wood Research Society (Japan)*. 1995.
9. Kerem Z, Jensen KA, Hammel KE. Biodegradative mechanism of the brown rot basidiomycete *Gloeophyllum trabeum*: evidence for an extracellular hydroquinone-driven fenton reaction. *FEBS letters*. 1999; 446: 49-54.
10. Daniel G, Volc J, Filonova L, Plíhal O, Kubátová E, Halada P. Characteristics of *Gloeophyllum trabeum* alcohol oxidase, an extracellular source of H₂O₂ in brown rot decay of wood. *Applied and environmental microbiology*. 2007; 73: 6241-53.
11. Shimokawa T, Nakamura M, Hayashi N, Ishihara M. Production of 2, 5-dimethoxyhydroquinone by the brown-rot fungus *Serpula lacrymans* to drive extracellular Fenton reaction. *Holzforschung*. 2004; 58: 305-10.
12. Suzuki MR, Hunt CG, Houtman CJ, Dalebroux ZD, Hammel KE. Fungal hydroquinones contribute to brown rot of wood. *Environmental microbiology*. 2006; 8: 2214-23.
13. Rättö M, Ritschkoff A-C, Viikari L. The effect of oxidative pretreatment on cellulose degradation by *Poria placenta* and *Trichoderma reesei* cellulases. *Applied microbiology and biotechnology*. 1997; 48: 53-7.
14. Yelle DJ, Ralph J, Lu F, Hammel KE. Evidence for cleavage of lignin by a brown rot basidiomycete. *Environmental microbiology*. 2008; 10: 1844-9.
15. Martínez AT, Rencoret J, Nieto L, Jiménez-Barbero J, Gutiérrez A, del Río JC. Selective lignin and polysaccharide removal in natural fungal decay of wood as evidenced by in situ structural analyses. *Environmental microbiology*. 2011; 13: 96-107.
16. Wymelenberg AV, Gaskell J, Mozuch M, BonDurant SS, Sabat G, Ralph J, et al. Gene expression of wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium* is significantly altered by plant species. *Applied and Environmental Microbiology*. 2011: AEM. 00508-11.
17. Yelle DJ, Wei D, Ralph J, Hammel KE. Multidimensional NMR analysis reveals truncated lignin structures in wood decayed by the brown rot basidiomycete *Postia placenta*. *Environmental Microbiology*. 2011; 13: 1091-100.
18. Micales JA. Localization and induction of oxalate decarboxylase in the brown-rot wood decay fungus *Postia placenta*. *International biodeterioration & biodegradation*. 1997; 39: 125-32.
19. Wei D, Houtman CJ, Kapich AN, Hunt CG, Cullen D, Hammel KE. Laccase and its role in production of extracellular reactive oxygen species during wood decay by the brown rot basidiomycete *Postia placenta*. *Applied and environmental microbiology*. 2010; 76: 2091-7.
20. Skyba O, Cullen D, Douglas CJ, Mansfield SD. Gene expression patterns of wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium* are influenced by wood substrate composition during degradation. *Applied and environmental microbiology*. 2016: AEM. 00134-16.
21. Zhang J, Presley GN, Hammel KE, Ryu J-S, Menke JR, Figueroa M, et al. Localizing gene regulation reveals a staggered wood decay mechanism for the brown rot fungus *Postia placenta*. *Proceedings of the National Academy of Sciences*. 2016: 201608454.
22. Wymelenberg AV, Gaskell J, Mozuch M, Sabat G, Ralph J, Skyba O, et al. Comparative transcriptome and secretome analysis of wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium*. *Applied and environmental microbiology*. 2010; 76: 3599-610.
23. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otilar R, et al. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research*. 2013: gkt1183.
24. Grigoriev IV, Cullen D, Goodwin SB, Hibbett D, Jeffries TW, Kubicek CP, et al. Fueling the future with fungal genomics. *Mycology*. 2011; 2: 192-209.
25. Botstein D, Cherry JM, Ashburner M, Ball C, Blake J, Butler H, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000; 25: 25-9.
26. Consortium GO. Gene ontology consortium: going forward. *Nucleic acids research*. 2015; 43: D1049-D56.

27. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. *BMC bioinformatics*. 2003; 4: 41.
28. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic acids research*. 2009; 37: D233-D8.
29. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research*. 2014; 42: D490-D5.
30. Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams. 2007.
31. Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. jvenn: an interactive Venn diagram viewer. *BMC bioinformatics*. 2014; 15: 1.
32. Kameshwar AKS, Qin W. Metadata Analysis of Phanerochaete chrysosporium Gene Expression Data Identified Common CAZymes Encoding Gene Expression Profiles Involved in Cellulose and Hemicellulose Degradation. *International Journal of Biological Sciences*. 2017; 13: 85-99.
33. Kameshwar AKS, Qin W. Gene expression metadata analysis reveals molecular mechanisms employed by Phanerochaete chrysosporium during lignin degradation and detoxification of plant extractives. *Current Genetics*. 2017: 1-18.
34. Kameshwar AKS, Qin W. Analyzing Phanerochaete chrysosporium gene expression patterns controlling the molecular fate of lignocellulose degrading enzymes. *Process Biochemistry*. 2017.
35. Grinhut T, Salame TM, Chen Y, Hadar Y. Involvement of ligninolytic enzymes and Fenton-like reaction in humic acid degradation by *Trametes* sp. *Applied microbiology and biotechnology*. 2011; 91: 1131-40.
36. Highley TL, Illman BL. Progress in understanding how brown-rot fungi degrade cellulose. *Biodeterioration Abstracts*; 1991. p. 231-44.
37. Lee SO, Tran T, Park YY, Kim SJ, Kim MJ. Study on the kinetics of iron oxide leaching by oxalic acid. *International Journal of Mineral Processing*. 2006; 80: 144-52.
38. Korripally P, Timokhin VI, Houtman CJ, Mozuch MD, Hammel KE. Evidence from *Serpula lacrymans* that 2, 5-dimethoxyhydroquinone is a lignocellulolytic agent of divergent brown rot basidiomycetes. *Applied and environmental microbiology*. 2013; 79: 2377-83.
39. Newcombe D, Paszczyński A, Gajewska W, Kröger M, Feis G, Crawford R. Production of small molecular weight catalysts and the mechanism of trinitrotoluene degradation by several *Gloeophyllum* species. *Enzyme and Microbial Technology*. 2002; 30: 506-17.
40. Fackler K, Stevanic JS, Ters T, Hinterstoisser B, Schwanninger M, Salmén L. Localisation and characterisation of incipient brown-rot decay within spruce wood cell walls using FT-IR imaging microscopy. *Enzyme and microbial technology*. 2010; 47: 257-67.
41. Schilling JS, Tewart JP, Duncan SM. Synergy between pretreatment lignocellulose modifications and saccharification efficiency in two brown rot fungal systems. *Applied microbiology and biotechnology*. 2009; 84: 465.
42. Cohen R, Suzuki MR, Hammel KE. Processive endoglucanase active in crystalline cellulose hydrolysis by the brown rot basidiomycete *Gloeophyllum trabeum*. *Applied and environmental microbiology*. 2005; 71: 2412-7.
43. Jung YH, Kim HK, Park HM, Park Y-C, Park K, Seo J-H, et al. Mimicking the Fenton reaction-induced wood decay by fungi for pretreatment of lignocellulose. *Bioresource technology*. 2015; 179: 467-72.
44. Barr DP, Aust SD. Mechanisms white rot fungi use to degrade pollutants. *Environmental Science & Technology*. 1994; 28: 78A-87A.
45. Agosin E, Jarpa S, Rojas E, Espejo E. Solid-state fermentation of pine sawdust by selected brown-rot fungi. *Enzyme and microbial technology*. 1989; 11: 511-7.
46. Filley T, Cody G, Goodell B, Jellison J, Noser C, Ostrofsky A. Lignin demethylation and polysaccharide decomposition in spruce sapwood degraded by brown rot fungi. *Organic Geochemistry*. 2002; 33: 111-24.
47. Kirk TK. Effects of a brown-rot fungus, *Lenzites trabea*, on lignin in spruce wood. *Holzforschung-International Journal of the Biology, Chemistry, Physics and Technology of Wood*. 1975; 29: 99-107.
48. Kirk TK, Adler E. Methoxyl-deficient structural elements in lignin of sweetgum decayed by a brown-rot fungus. *Acta Chem Scand*. 1970; 24: 90.

49. Koenig AB, Sleighter RL, Salmon E, Hatcher PG. NMR structural characterization of *Quercus alba* (white oak) degraded by the brown rot fungus, *Laetiporus sulphureus*. *Journal of wood chemistry and technology*. 2010; 30: 61-85.
50. Eriksson K-EL, Blanchette R, Ander P. *Microbial and enzymatic degradation of wood and wood components*: Springer Science & Business Media; 2012.
51. Arantes V, Qian Y, Kelley SS, Milagres AM, Filley TR, Jellison J, et al. Biomimetic oxidative treatment of spruce wood studied by pyrolysis–molecular beam mass spectrometry coupled with multivariate analysis and ¹³C-labeled tetramethylammonium hydroxide thermochemolysis: implications for fungal degradation of wood. *JBIC Journal of Biological Inorganic Chemistry*. 2009; 14: 1253.
52. Arantes V, Jellison J, Goodell B. Peculiarities of brown-rot fungi and biochemical Fenton reaction with regard to their potential as a model for bioprocessing biomass. *Applied Microbiology and Biotechnology*. 2012; 94: 323-38.
53. Hyde SM, Wood PM. A mechanism for production of hydroxyl radicals by the brown-rot fungus *Coniophora puteana*: Fe (III) reduction by cellobiose dehydrogenase and Fe (II) oxidation at a distance from the hyphae. *Microbiology*. 1997; 143: 259-66.
54. Arantes V, Milagres AM, Filley TR, Goodell B. Lignocellulosic polysaccharides and lignin degradation by wood decay fungi: the relevance of nonenzymatic Fenton-based reactions. *Journal of industrial microbiology & biotechnology*. 2011; 38: 541-55.
55. Hocine S, Singer RH, Grünwald D. RNA processing and export. *Cold Spring Harbor perspectives in biology*. 2010; 2: a000752.
56. Shuman S. Structure, mechanism, and evolution of the mRNA capping apparatus. *Progress in nucleic acid research and molecular biology*. 2000; 66: 1-40.
57. Busch H, Reddy R, Rothblum L, Choi Y. SnRNAs, SnRNPs, and RNA processing. *Annual review of biochemistry*. 1982; 51: 617-54.
58. Amore A, Giacobbe S, Faraco V. Regulation of cellulase and hemicellulase gene expression in fungi. *Current genomics*. 2013; 14: 230-49.

Chapter-6

Comparative Study of Genome-Wide Plant Biomass Degrading CAZymes in White Rot, Brown Rot and Soft Rot Fungi

[This work has been published in *Mycology* (2017): 1-13.]

Ayyappa Kumar Sista Kameshwar and Wensheng Qin*

6.1. Abstract

We have conducted a genome-level comparative study of basidiomycetes wood rotting fungi (white, brown and soft rot) to understand the total plant biomass (lignin, cellulose, hemicellulose and pectin) degrading abilities. We have retrieved the genome level annotations of well-known 14 white rot fungi, 15 brown rot fungi and 13 soft rot fungi. Based on the previous literature and the annotations obtained from CAZy (carbohydrate active enzyme) database, we have separated the genome wide CAZymes of the selected fungi into lignin, cellulose, hemicellulose and pectin degrading enzymes. Results obtained in our study reveals that white rot fungi especially *Pleurotus eryngii* and *Pleurotus ostreatus* potentially possess high ligninolytic ability and soft rot fungi especially *Botryosphaeria dothidea* and *Fusarium oxysporum* sp potentially possess high cellulolytic, hemicellulolytic and pectinolytic abilities. The total number of genes encoding for cytochrome P450 monooxygenases and metabolic processes were high in soft and white rot fungi. We have tentatively calculated the overall lignocellulolytic abilities among the selected wood rotting fungi which suggests that white rot fungi possess higher lignin and soft rot fungi potentially possess higher cellulolytic, hemicellulolytic and pectinolytic abilities. This approach can be applied industrially to efficiently find lignocellulolytic and aromatic compound degrading fungi based on their genomic abilities.

Keywords: Plant biomass, Lignocellulose, CAZy, White rot fungi, Brown rot fungi, Soft rot fungi,

6.2. Introduction

Naturally, lignocellulose is degraded by a large group of fungi and bacteria [1]. Fungi have evolved progressively with their dominant degrading abilities to decay organic debris including plant biomass by penetrating through their hyphae and spores (for long distance dispersal) [2]. Wood rotting fungi are categorized into white, brown and soft rot fungi based on their growth substrate preferences and wood decaying patterns [3]. Moreover, white rot fungi exhibit excellent decaying abilities and solely responsible for the degradation of lignin and polysaccharides in plant biomass. Microscopy based studies have differentiated the white rot decay patterns morphologically into a) simultaneous degradation of lignin and wood polysaccharides. For e.g. *Phanerochaete chrysosporium*, *Trametes versicolor*. b) selective degradation of plant biomass components. For e.g. *Phlebia radiata* [1, 4, 5]. However, some fungi like *Heterobasidium annosum* exhibits both simultaneous and selective decay patterns [6].

Brown rot fungi are well characterized as rapid cellulose and hemicellulose degraders, they access plant polysaccharides by potentially modifying or degrading lignin [7]. These fungi are the major invaders of forest biomass and wood-based constructions. Studies have reported that brown rot fungi have evolved from the saprotrophic white rot fungi by losing several essential genes encoding for lignocellulose degrading enzymes [8]. It was reported that hyphae of the brown rot fungi penetrates the cell lumen, colonizes the ray cells and axial parenchymal cells to access carbohydrates [8].

Most of the ascomycetes and fungi imperfectii cause soft rot decay in the presence of excessive moisture, soft rot decayed wood exhibits a greyish discoloration and fragmentation which is similar as brown rot. Previous morphological studies have divided the soft rot fungi decay into a) type-I (where hyphae penetrates secondary cell walls by forming characteristic cavities) b) type-II (attacks similarly as ascomycetes and white rot fungi leading to wood cell wall thinning) [1]. Wood decaying fungi and its secreted enzymes are being used commercially in biopulping, kraft pulping (xylanase bleaching), cellulases based refining, pitch removal (lipases), slime removal (using enzyme cocktail), fiber modification (pulp and paper industries) etc. Thus, finding its applications in biodegradation of plant polymers, detoxification and bioremediation of several toxic aromatic compounds and also in bio based industries [1].

The depolymerizing abilities of the wood rotting fungi are directly proportional to its ability to secrete an array of lignocellulolytic enzymes, aromatic compound and detoxifying enzymes. The plant cell wall modifying and degrading enzymes secreted by microorganisms were been classified into six classes by CAZy database [9]. They are glycoside hydrolases (GH), glycosyl transferase (GT), auxiliary activity (AA), carbohydrate esterase (CE), polysaccharide lyases (PL) and carbohydrate binding domains (CBD) [9]. Cellulose, hemicellulose and pectin are the most important and major polysaccharides of the plant cell walls. The presence of lignin (heterophenolic aromatic polymer) along with these components make the plant cell wall recalcitrant [10]. Structurally and functionally plant cell walls are unique, and they can be divided into a) middle lamella b) primary cell wall and c) secondary cell wall. Chemical composition of plant cell walls varies considerably among monocots, dicots, softwood and hardwood. Primary cell walls of renewable energy crops (monocots, grasses etc,) contain cellulose and hemicellulose similarly secondary cell walls contains higher amounts of cellulose, varied compositions of hemicellulose and substantial amounts of lignin [10, 11]. Where as in dicots primary cell walls contain low xylan, high xyloglucan and mannan, secondary cell walls contain cellulose, hemicellulose and lignin, in dicot plant cell walls pectin is considerably higher [10, 11].

Most abundant plant polysaccharide cellulose, provides rigidity to the plant cell walls by constituting upto 40-50% of its dry weight. Cellulose is made up of β (1 \rightarrow 4) linear chains of D-glucose repeating units linked through hydrogen bonds, wherein the ratio of crystalline to amorphous regions differs between the layers of primary and secondary cell walls and also among the plant species [12, 13]. Hemicellulose constitutes to 20 to 30% dry weight of the plant biomass, it is mainly composed of xylan (β (1 \rightarrow 4) D-xylose units), xyloglucan, β -glucans (β (1 \rightarrow 3) (1 \rightarrow 4) D-glucose), and mannan (β 1 \rightarrow 4 D-mannose), it also contains oligomers of galactose, xylose, arabinose, fucose, glucuronic acid [10, 13]. The hemicellulose occurs in close association with cellulose, by supporting the microfibrillar structure of cellulose. In plant cell walls pectin occurs as homogalacturonan (α (1 \rightarrow 4) D-galacturonic acid), xylogalacturonan (galacturonan and β (1 \rightarrow 3) D-xylose), rhamnogalacturonan-I and rhamnogalacturonan-II. Thus, pectin is the non-cellulosic plant polysaccharide which occurs in intricate associations with other plant cell wall components [10, 13]. Fungi secretes an array of CAZymes and lignin degrading enzymes (which includes aromatic compound degrading and detoxifying enzymes) for the degradation of lignocellulose [10] (Figure 6.1).

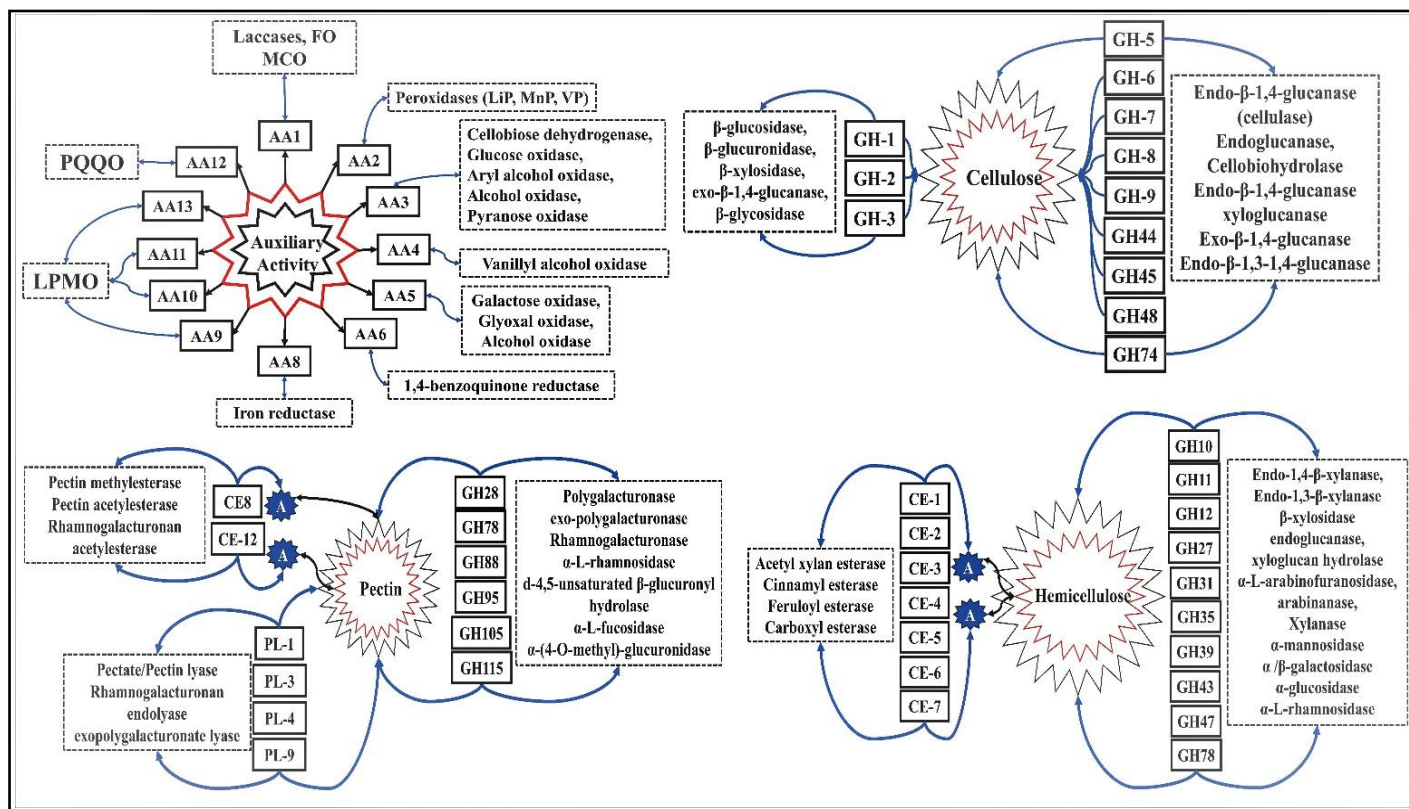


Figure 6.1: Tentative network of CAZymes involved in depolymerization of lignin, cellulose, pectin and hemicellulose observed in selected popular white rot, brown rot and soft rot fungi

Several genomic studies were conducted in the past to reveal the genes encoding for plant biomass degrading enzymes. In this study, we have selected popular white rot, brown rot and soft rot fungal strains and retrieved their genome wide annotations to reveal the number of cellulolytic, hemicellulolytic, pectinolytic, lignin degrading and detoxifying enzymes (especially cytochrome P450). Present comparative analysis approach can be applied industrially to efficiently find lignocellulose and xenobiotic compound degrading fungal strains, which can be applied in production of commercially important enzymes and growing biofuel and biorefinery industries.

6.2. Data retrieval and Analysis

6.2.1 Data Retrieval

We have selected **A**) 14 popular white rot fungal strains (*Ceriporiopsis subvermispora* B [14], *Heterobasidion annosum* v2.0 [15], *Fomitiporia mediterranea* v1.0 [16], *Phanerochaete carnosa* HHB-10118 [17], *Pycnoporus cinnabarinus* BRFM 137 [18], *Phanerochaete chrysosporium* R78

v2.2 [19, 20], *Dichomitus squalens* LYAD-421 SS1 [16], *Trametes versicolor* v1.0 [16], *Punctularia strigosozonata* v1.0 [16], *Phlebia brevispora* HHB-7030 SS6 [21], *Botrytis cinerea* v1.0 [22], *Pleurotus ostreatus* PC15 v2.0 [23-25], *Stereum hirsutum* FP-91666 SS1 v1.0 [16], *Pleurotus eryngii* ATCC90797 [26-29]. **B**) 15 popular brown rot fungal strains (*Postia placenta* MAD 698-R v1.0 [30], *Fibroporia radiculosa* TFFH 294 [31], *Wolfiporia cocos* MD-104 SS10 v1.0 [16], *Dacryopinax primogenitus* DJM 731 SSP1 v1.0 [16], *Daedalea quercina* v1.0 [32], *Laetiporus sulphureus* var v1.0 [32], *Postia placenta* MAD-698-R-SB12 v1.0 [30], *Neolentinus lepideus* v1.0 [32], *Serpula lacrymans* S7.9 v2.0 [33], *Calocera cornea* v1.0 [33], *Gloeophyllum trabeum* v1.0 [16], *Fistulina hepatica* v1.0 [34], *Fomitopsis pinicola* FP-58527 SS1 [34], *Hydnomerulius pinastri* v2.0 [35] and *Coniophora puteana* v1.0 [35]. **C**) 13 popular soft rot fungal strains (*Trichoderma reesei* v 2.0 [36], *Rhizopus oryzae* 99-880 from Broad [37], *Aspergillus wentii* v1.0 [38], *Penicillium chrysogenum* Wisconsin 54-1255 [39], *Daldinia eschscholzii* EC12 v1.0, *Hypoxylon* sp. CI-4A v1.0 [40], *Aspergillus niger* ATCC 1015 v4.0 [41], *Hypoxylon* sp. EC38 v3.0 [40], *Hypoxylon* sp. CO27-5 v1.0 [40], *Neurospora crassa* OR74A v2.0 [42], *Lecythophora* sp. AK0013 v1.0 [43, 44], *Botryosphaeria dothidea* [45-49], *Fusarium oxysporum* sp. lycopersici 4287 v2 [50] with available annotated genomes were retrieved from the JGI (Joint genome institute) MycoCosm database. Genome level annotations of the selected fungal strains especially InterPro, KOG and CAZy were retrieved from the JGI-MycoCosm database.

6.2.2. Data Analysis

Based on the previous literature and the available CAZy annotations we have classified the genome wide CAZymes of the above selected white, brown and soft rot fungi. List of CAZymes retrieved from the JGI-MycoCosm database were individually classified into cellulases, hemicellulases and ligninolytic enzymes. We have used Microsoft Excel 2016 to represent the number of genes coding for plant cell wall degrading enzymes present in the genome wide annotations of fungi. The images were generated using the option “conditional formatting followed by selecting option color scales”. Present CAZy database is classified into 145 glycoside hydrolases, 104 glycosyl transferases, 27 polysaccharide lyases, 16 carbohydrate esterases and 13 auxiliary activity enzymes. We have used the available annotations and literature, to separate the CAZymes into plant cell wall degrading (cellulolytic, hemicellulolytic, pectinolytic and ligninolytic) enzymes. Similarly, we have analyzed the number of protein models encoding for various significant cellular processes using the retrieved KOG (eukaryotic orthologous groups) and genome wide InterPro annotations

retrieved from JGI-MycoCosm database. We have also calculated the tentative overall lignocellulolytic abilities of white, brown and soft rot fungi based on the genome wide distribution of lignocellulolytic enzymes in selected fungi. We have performed the hierarchical clustering analysis of the genome level data (CAZy, InterPro and KOG) of the above selected fungi, using Cluster 3 [51] and visualized the obtained trees using Java Treeview softwares. Following options were used in Cluster3.0 software: we have uploaded the sample files (containing number of genes encoding for CAZymes, KOG) and selected “hierarchical” clustering, “cluster” options for both the genes and arrays with complete linkage clustering method. The CDT file obtained from the Cluster3.0 software was imported into the Java Treeview software and the corresponding images were generated and further exported. The hierarchical dendrograms of the plant cell wall degrading CAZymes, ligninolytic, cellulolytic, hemicellulolytic, pectinolytic and Eukaryotic orthologous groups (KOG).

6.3. Results and Discussions

Basidiomycetes fungi were highly studied and classified based on their plant biomass decaying abilities into white rot, brown rot and soft rot fungi. White rot fungi are the efficient plant biomass degraders with its specialty lying in degradation of aromatic compounds, thus giving a characteristic white appearance to the decayed wood. Brown rot fungi represents about 6 to 7 percent of the basidiomycete fungi. *Phanerochaete chrysosporium* genome was the first basidiomycete complete genome sequence to be published in the year 2004 by Martinez et al (2004), which has revealed various significant facts about lignocellulose degradation mechanisms [19, 20]. After this study, the complete genome sequences of several basidiomycetes fungi were revealed in the recent years [52]. Development and advancement of genome repositories such as 1000 fungal genome project and JGI MycoCosm have fastened various findings about the fungal metabolism, physiology and degrading mechanisms [53]. As mentioned above we have selected 14 white rot fungi, 15 brown rot fungi and 13 soft rot fungi whose complete annotated genome sequences are published and publicly available. We have retrieved the genome wide annotations such as InterPro, CAZy, KOG for all the selected fungal strains. Based on the available literature and CAZy architecture we have separated lignocellulose CAZymes into their respective cellulose, hemicellulose, pectin and lignin depolymerizing enzymes.

The JGI-MycoCosm database classifies and annotates the fungal genomes using the KOG tool (a eukaryotic version of cluster of orthologous groups) KOG tool is used for identifying the ortholog and paralog proteins. The KOG groups are divided into four functional groups a) cellular processes and signaling b) information storage and processing c) metabolism and d) poorly characterized. These four functional groups are further divided into different classes based on their functional characteristics. We have retrieved the classified genome KOG groups and their respective function level annotated gene numbers for all the selected wood rotting fungi. We have observed that *T. reesei*, *Lecythophora* sp, *P. placenta*, *F. mediterranea*, *C. subvermispora*, genomes contain lower number of genes classified under the above mentioned four functional KOG groups. At the same time, higher number of genes encoding for the KOG functional groups were observed among the *F. oxysporum*, *R. oryzae*, *P. brevispora*, *S. hirsutum*, *C. puteana*, *F. pinicola* fungi (Figure 6.2A). In this study, we have specifically compared the total number of genes encoding for KOG functional processes encoding for energy production and conversion (C), carbohydrate transport and metabolism (G) and secondary metabolite biosynthesis and transport (Q). The descending order of fungi based on the number of genes encoding for KOG processes C, G, Q were *F. oxysporum*, *B. dothidea*, *P. brevispora*, *S. hirsutum*, *C. puteana* and *F. pinicola* respectively. Similarly, lower number of genes were observed among *Lecythophora* sp, *T. reesei*, *N. crassa* (soft rot), *P. cinnabarinus*, *P. ostreatus* (white rot) and *P. placenta* MAD-698 R 1.0 respectively (Figure 6.2B).

Several studies have proved the strong involvement of cytochrome P-450 monooxygenases in the degradation of aromatic and xenobiotic compounds present in the environment by fungi[54, 55]. In fungi, cytochrome P450 monooxygenases occur in multiple copies as they play wide range of roles especially in detoxification and degradation [56]. Among the selected wood rotting fungi, *P. carnosus* (253), *P. brevispora* (238), *B. dothidea* (237), *C. puteana* (229), *F. oxysporum* (211) contains higher and *N. crassa* (39), *R. oryzae* (47), *T. reesei* (69), *Lecythophora* sp (76) lower number of genes encoding for cytochrome P450 monooxygenase encoding genes (Figure 6.2B). White rot fungi harbor higher number of genes encoding for cytochrome P450 monooxygenases followed by brown rot fungi and soft rot fungi.

A	SRF	T.re	R.or	A.we	P.ch	D.es	Hypo	A.ni	Hy.EC38	N.cr	Hy.CO27	F.ox	Lecy	B.do			
	C	316	380	463	411	382	392	474	401	312	403	813	319	487			
	G	346	430	506	457	419	442	482	454	419	451	923	315	551			
	Q	262	150	450	380	366	392	485	435	181	433	808	270	618			
	WRF	P.chr	P.ca	C.sub	H.an	F.me	P.cin	D.sq	T.ve	P.st	P.br	B.ci	P.os	S.hi	P.er		
	C	352	393	330	311	310	309	349	383	346	402	341	312	391	314		
	G	335	343	331	350	362	335	390	401	362	433	451	342	500	365		
	Q	344	490	392	305	276	250	369	384	315	493	343	314	486	288		
	BRF	P.pl	F.ra	W.co	D.pr	D.qu	L.su	P.pl	N.le	S.la	C.co	G.tr	F.he	F.pi	H.pi	C.pu	
	C	146	297	337	310	374	345	354	354	309	348	362	320	366	339	396	
	G	202	296	316	340	335	363	312	332	296	382	323	356	364	316	355	
	Q	227	344	396	308	336	348	397	322	337	342	300	267	388	334	467	
B	SRF	T.re	R.or	A.we	P.ch	D.es	Hypo	A.ni	Hy.EC38	N.cr	Hy.CO27	F.ox	Lecy	B.do			
	Cyt-P450	69	47	130	97	117	136	145	162	39	163	211	76	237			
	WRF	P.chr	P.ca	C.sub	H.an	F.me	P.cin	D.sq	T.ve	P.st	P.br	B.ci	P.os	S.hi	P.er		
	Cyt-P450	148	253	205	138	119	100	170	183	133	238	118	143	205	126		
	BRF	P.pl	F.ra	W.co	D.pr	D.qu	L.su	P.pl	N.le	S.la	C.co	G.tr	F.he	F.pi	H.pi	C.pu	
Cyt-P450	104	165	197	113	134	154	194	122	152	127	114	100	170	131	229		

Figure 6.2: Heatmaps showing the genome wide distribution of A) Metabolism (C = energy production and conversion, G = carbohydrate transport and metabolism and Q = secondary metabolites biosynthesis, transport and catabolism) and B) number of cytochrome P450 encoding genes in selected popular white rot, brown rot and soft rot fungi

6.3.1. Distribution of CAZymes among white rot, brown rot and soft rot fungi

Present day CAZy database comprises of 145 glycoside hydrolases (GH), 104 glycosyl transferases (GT), 27 polysaccharide lyases (PL), 16 carbohydrate esterases (CE), 13 auxiliary activity (AA) and 81 carbohydrate binding modules (CBM) [9, 57]. Further to this classification glycoside hydrolases classes GH-5, GH-13, GH-30, GH-43 is further divided into 53, 42, 8, 37 subfamilies respectively [58, 59]. Total number of CAZymes distributed among the selected fungi ranges between 370 (*C. subvermispora*) to 588 (*P. eryngii*) in white rot fungi, 245 (*P. placenta*) to 426 (*C. puteana*) in brown rot fungi and 408 (*T. reesei*) to 881 (*F. oxysporum*) in soft rot fungi respectively (Figure 6.1). On average, the total number of CAZymes distributed among the selected fungi were 366 in brown rot, 480 in white rot and 553 in soft rot fungi respectively. Genome wide distribution of CAZymes among the selected white rot, brown rot and soft rot fungi were clearly listed in the heatmaps (Figure 6.3).

	AA	CBM	CE	EXPN	GH	GT	PL	Total CAZY
White Rot Fungi								
<i>Ceriporiopsis subvermispora</i> B	62	37	17	11	169	68	6	370
<i>Heterobasidion annosum</i> v2.0	77	36	21	12	176	63	9	394
<i>Fomitiporia mediterranea</i> v1.0	77	26	17	11	198	78	6	413
<i>Phanerochaete carnosa</i> HHB-10118	83	43	15	13	186	75	8	423
<i>Pycnoporus cinnabarinus</i> BRFM 137	70	46	18	15	194	79	8	430
<i>Phanerochaete chrysosporium</i> R78 v2.2	89	71	20	11	181	72	6	450
<i>Dichomitus squalens</i> LYAD-421 SS1	88	45	23	13	224	70	11	475
<i>Trametes versicolor</i>	89	47	19	12	222	85	9	485
<i>Punctularia strigosozonata</i> v1.0	76	57	30	10	225	78	12	488
<i>Phlebia brevispora</i> HHB-7030 SS6	93	101	18	15	217	92	7	543
<i>Botrytis cinerea</i> v1.0	76	68	37	5	247	112	10	555
<i>Pleurotus ostreatus</i> PC15	114	82	28	8	235	67	23	557
<i>Stereum hirsutum</i> FP-91666SS1	103	46	33	18	271	76	14	561
<i>Pleurotus eryngii</i> ATCC90797	115	95	26	14	232	74	32	588
Brown rot Fungi								
<i>Postia placenta</i> MAD 698-R v1.0	29	33	15	12	130	25	1	245
<i>Fibroporia radiculosa</i> TFFH 294	33	28	11	16	132	62	2	284
<i>Wolfiporia cocos</i> MD-104 SS10 v1.0	27	18	13	13	147	69	2	289
<i>Dacryopinax primogenitus</i> DJM 731 SSP1 v1.0	21	14	18	6	153	92	3	307
<i>Daedalea quercina</i> v1.0	36	24	17	10	160	67	2	316
<i>Laetiporus sulphureus</i> var v1.0	46	23	14	11	152	70	3	319
<i>Postia placenta</i> MAD-698-R-SB12 v1.0	40	35	11	21	144	72	5	328
<i>Neolentinus lepideus</i> v1.0	41	22	13	10	183	70	6	345
<i>Serpula lacrymans</i> S7.9 v2.0	37	43	13	15	164	70	6	348
<i>Calocera cornea</i> v1.0	27	18	15	6	166	115	3	350
<i>Gloeophyllum trabeum</i> v1.0	41	19	17	13	195	68	9	362
<i>Serpula lacrymans</i> var SHA21-2 v1.0	49	39	12	13	193	67	6	379
<i>Fistulina hepatica</i> v1.0	39	27	13	10	205	83	7	384
<i>Fomitopsis pinicola</i> FP-58527 SS1	43	45	17	20	199	75	3	402
<i>Hydnomerulius pinastri</i> v2.0	53	64	22	6	198	70	6	419
<i>Coniophora puteana</i> v1.0	51	31	20	19	243	59	3	426
Soft Rot Fungi								
<i>Trichoderma reesei</i> v 2.0	32	55	16	7	199	94	5	408
<i>Rhizopus oryzae</i> 99-880 from Broad	13	51	45	19	131	152	6	417
<i>Aspergillus wentii</i> v1.0	67	53	25	4	229	110	10	498
<i>Penicillium chrysogenum</i> Wisconsin 54-1255	60	67	24	4	237	113	9	513
<i>Daldinia eschscholzii</i> EC12 v1.0	91	53	37	5	240	90	8	524
<i>Hypoxyylon</i> sp. CI-4A v1.0	95	48	35	4	240	94	10	526
<i>Aspergillus niger</i> ATCC 1015 v4.0	64	69	21	2	251	123	9	539
<i>Hypoxyylon</i> sp. EC38 v3.0	99	57	39	5	258	90	10	558
<i>Hypoxyylon</i> sp. CO27-5 v1.0	98	62	39	5	255	90	10	559
<i>Neurospora crassa</i> OR74A v2.0	51	71	24	3	192	96	4	441
<i>Lecytophthora</i> sp. AK0013 v1.0	80	138	42	3	292	101	7	663
<i>Botryosphaeria dothidea</i>	110	68	41	6	305	114	23	667
<i>Fusarium oxysporum</i> sp. lycopersici 4287 v2	115	152	54	11	390	136	23	881

Figure 6.3: Heatmap showing the genome wide distribution of CAZymes from selected popular white rot, brown rot and soft rot fungi.

6.3.2. Lignin degrading CAZymes

Fungi secretes an array of oxidative enzymes for the degradation of lignin and other various aromatic compounds. The FOly (Fungal oxidative lignin enzymes) database has classified lignin degrading enzymes into two major classes lignin oxidizing enzymes (LO) and lignin degrading auxiliary enzymes [3, 60]. Lignin oxidizing enzymes includes laccase (LO1), lignin peroxidase (LO2), manganese peroxidase (LO2), versatile peroxidase (LO2) and cellobiose dehydrogenase (LO3). Similarly, lignin degrading auxiliary enzymes majorly includes hydrogen peroxide generating enzymes such as aryl alcohol oxidase (LDA1), vanillyl alcohol oxidase (LDA2), glyoxal oxidase (LDA3), pyranose oxidase (LDA4), galactose oxidase (LDA5), glucose oxidase (LDA6) and benzoquinone reductase (LDA7) [3, 60].

In CAZy, lignin degrading enzymes are classified under the auxiliary activity (AA) enzyme class. We have used CAZy database structure and previous literature to determine the genome wide lignin degrading CAZymes and ligninolytic capacities of the selected wood rotting fungi [61]. Lignin oxidizing enzymes (Laccase, Peroxidases (LiP, MnP, VP) and cellobiose dehydrogenase) were classified among AA-1, AA-2, AA-3_1 enzyme classes. Lignin degrading auxiliary enzymes were classified among AA-3, AA-4, AA-5, AA-6, AA-8 enzyme classes. The genome wide distribution of lignin degrading auxiliary activity enzymes among the selected white rot, brown rot and soft rot fungi were listed. These results convey that white rot fungi possess higher number of laccase encoding genes when compared to brown rot and soft rot fungi (e.g. >10 laccase encoding genes were observed in *S. hirsutum*, *H. annosum*, *P. ostreatus*, *F. mediterranea*, *D. squalens*, *P. strigosozonata*, *B. cinerea*, *P. eryngii*). Genes encoding for ferroxidases were found to be mostly reduced to 1 to 2 copies in white and brown rot fungi. However, most of the soft rot fungi possessed 2 ferroxidases genes and other fungi harbored 3 to 4 ferroxidase genes.

Genes encoding for AA2 class enzymes (lignin (LiP), manganese (MnP), versatile (VP) peroxidases) were found to be reduced among brown and soft rot fungi to few copies (0 to 4 genes). While genomes of white rot fungi harbors more than 5 gene copies for AA2 class enzymes with higher number (26) of AA2 encoding genes were observed in *T. versicolor*. Total number of genes encoding for glucose methanol choline (GMC) oxidoreductases or auxiliary activity class 3 enzymes dominate in number when compared to other lignin degrading auxiliary activity enzymes. Especially the number of genes encoding for aryl alcohol oxidase and glucose-1-oxidase (AA-3_2 subclass) outnumber other AA-class

enzymes. Higher number of AA-3 class enzymes were observed in white rot fungi (48 in *S. hirsutum*) followed by soft rot fungi (46 in *B. dothidea*) and brown rot fungi (30 in *L. sulphureus*) respectively. In most of the white and brown rot fungi genes encoding for vanillyl alcohol oxidase (AA4) were reduced to 0, with some fungi comprises around 1 to 3 genes sequences. Except in *T. reesei*, *R. oryzae* and *N. crassa* all the selected soft rot fungi possessed the gene sequences coding for vanillyl alcohol oxidase. Genes encoding for AA5 class enzymes (galactose, glyoxal and alcohol oxidase), AA6 (benzoquinone reductases) were comparatively higher among the white rot fungi. These results convey that white rot fungi harbor multiple gene copies encoding for lignin degrading and its auxiliary enzymes when compared to brown and soft rot fungi (Figure 6.3). Finally, the number of genes encoding for lytic polysaccharide monooxygenases (LPMO) classified under AA-9, AA-10, AA-11 and AA-13 enzyme classes were higher in number among the white rot fungi followed by the soft rot fungi respectively (Figure 6.3). Based on the total number of genes encoding for lignin degrading auxiliary activity enzymes, we propose that ligninolytic ability of white rot fungi is higher followed by soft rot fungi.

Studies have revealed that fungi secretes a wide range of aromatic compound degrading and detoxifying enzymes during the process of lignin degradation [56]. Large group of enzymes encoding for aromatic ring and epoxide hydroxylases, intra and extra dioxygenases, alcohol dehydrogenases, (Fentons reagents) iron reductase, ferredoxin, catalase, oxidoreductases, cytochrome P450 monooxygenases and other large set of enzymes were involved in degradation of lignin [56]. Previous studies have reported the strong dependence of white rot and brown rot fungi on Fenton's chemistry in the degradation of lignocellulosic units of plant biomass. Apart from the lignin degrading auxiliary activity CAZymes, aromatic compound degrading, detoxifying enzymes and highly reactive free radicals such as hydrogen peroxide, hydroxy radicals, superoxide and reactive singlet oxygen ions also play a crucial role in degradation of lignocellulosic units.

6.3.3. Cellulose degrading CAZymes

Fungi mainly secretes three classical enzymes endoglucanases, exoglucanases and β -glucosidases for the hydrolysis of cellulose [10]. Endoglucanases/ β -(1 \rightarrow 4)-endoglucanases hydrolases cellulose chains by releasing glucooligosaccharides while cellobiohydrolases/exoglucanases liberates cellobiose from end chains of cellulose. Cellobiohydrolases (CBH) are divided into CBHI and CBHII, based on location of the cleavage

sites either on reducing or non-reducing ends of the cellulose. β -glucosidases releases individual glucose units from the shorter oligosaccharide chains [10]. We have reported the most commonly expressing cellulase encoding genes among the selected fungal strains. In most of the selected wood rotting fungi especially brown and soft rot fungi, we have commonly observed that, the genes encoding for GH-8, GH-44, GH-48 enzyme class (cellulases) were totally reduced to 0. Similarly, in brown and soft rot fungi genes encoding for GH-9, GH-45, GH-74 and GH-38 enzyme classes were reduced to single copies. In white and brown rot fungi genes encoding for GH-1, GH-2 class enzymes were found to occur in between the range of 1-5 gene copies. Most of the soft rot fungal genomes contain 2 -3 gene copies of GH-1 class enzymes and 5-10 gene copies of GH-2 class enzymes respectively. However, among all the cellulolytic glycoside hydrolases, the number of genes encoding for the GH-3, GH-5 enzyme classes outnumbered other cellulolytic GH enzyme classes among the white, brown and soft rot fungi. Among the selected wood rotting fungi, the total number of cellulases were higher among *P. eryngii* (69), *P. ostreatus* (68) (white rot fungi), *H. pinastri* (51), *C. puteana* (49) (brown rot fungi) and *F. oxysporum* (73) and *Lecythophora* sp (63) (soft rot fungi) respectively (Figure 6.3). Our analysis reports that among all the selected wood rotting fungi, glycoside hydrolase enzyme classes GH-1, GH-2, GH-3, GH-5 and GH-7 occurs in higher copy numbers. At the same time, lower number of cellulases were found in *H. annosum* (39), *C. subvermispora* (38) (white rot), *P. placenta* (29), *F. radiculosa* (29) (brown rot) and *N. crassa* (36), *R. oryzae* (26) (soft rot) respectively. Recent studies have reported that along with classical cellulases strong oxidoreductases such as cellobiose dehydrogenase (CDH) and lytic polysaccharide monoxygenases (LPMO) also partake in degradation of cellulose. Based on total number of genes encoding for cellulases among the selected white, brown and soft rot fungi, we report that the cellulolytic ability is lower in white rot fungi when compared to brown rot and soft rot fungi. Cellobiose dehydrogenase and LPMO enzymes work cooperatively for depolymerizing cellulose, as CDH produces highly reactive hydroxy radicals through Fenton's chemistry which plays a dual role by modifying lignin and providing electrons for LPMO based cellulose degradation [10].

6.3.4. Hemicellulose degrading CAZymes

Compared to cellulose, the microbial degradation of hemicellulose is performed by a specific set of CAZymes, which is majorly due to its complex structure. Classical enzymes such as β (1 \rightarrow 4) endoxylanases, xylobiohydrolase, β (1 \rightarrow 4) xylosidases which are involved in

hydrolysis of xylan backbone, xylan into xylobiose, releases D-xylose units from xylooligosaccharides and hydrolyzes xylobiose units to monomeric units respectively [10, 62-64]. Studies have reported that cellulases (endoglucanases, cellobiohydrolases and beta glucosidases) are involved in hydrolysis of cellulose like xyloglucan and β -glucan backbone structures [62-64]. Other enzymes such as β (1 \rightarrow 4) endo mannanases and β (1 \rightarrow 4) mannosidases cleaves mannan back bone structures by releasing monomeric D-mannose units [62-64]. Apart from these glycoside hydrolases enzymes such as cellobiose dehydrogenases, lytic polysaccharide monooxygenases are involve in oxidative cleavage of hemicellulose and carbohydrate esterases are involved in O-de-N-deacylation of acetylated plant cell wall residues especially hemicellulose, pectin and lignin. In most of the selected white and brown rot fungi, genes encoding for carbohydrate esterase classes CE-2, CE-3, CE-5, CE-6, CE-7 and glycoside hydrolase classes GH-11, GH-39 were completely reduced between 1 to 2 gene copies. Based on the total number of genes coding for hemicellulolytic enzymes, we have calculated the hemicellulolytic ability of the selected wood rotting fungi. In white rot fungi, total number of hemicellulolytic genes varies in between 31 *C. subvermispora* (low), 63 *B. cinerea* (high), where as in brown rot fungi number of genes varies between 23 *P. placenta* MAD-698 Rv.1.0 (low), 52 *C. puteana* (high). Contrastingly, soft rot fungi harbors higher number of genes encoding for hemicellulolytic enzymes varying in between 43 *T. reesei* (low) to 135 *F. oxysporum* (high). These results convey that soft rot fungi exhibit higher hemicellulolytic ability followed by white rot fungi.

6.3.5. Pectin degrading CAZymes

Pectin (a non-cellulosic polysaccharide present in plant cell walls) majorly comprises of galacturonic acid which is intricately connected with the cellulose and hemicellulose units. Majorly pectin occurs in primary and middle lamella of plant cell walls [10]. Structurally pectin can be classified as simple e.g. homogalacturonan (linear polymer of α (1 \rightarrow 4) D-galacturonic acids, methylated at C-6 and acetylated at C-3 positions), xylogalacturonan (chain of galacturonic acid is connected to β (1 \rightarrow 3) D-xylose units) and complex pectin e.g. rhamnogalacturonan-I and II (which contains glycosyl residues such as 2-O-methyl xylose, 2-O-methyl fucose, acetic acid, 2-keto-3-deoxy-D-lyxo heptulosaric acid, and 2-keto-3-deoxy-D-mannooctulosonic acid) [10, 13]. Wood rotting fungi secretes an arsenal of enzymes involved in depolymerization of pectin which includes endo-polygalacturonases, exo-polygalacturonases, xylogalacturonan hydrolases, endo-rhamnogalacturonase,

rhamnogalacturonan rhamnohydrolase, pectin and pectate lyases, rhamnogalacturonan hydrolases. Endo and exo polygalacturonases acts on starting and terminal ends by cleaving the linear chain of α (1 \rightarrow 4)-D-galacturonic acid present in the homogalacturonan and releases D-galacturonic acid. Similarly, enzymes xylogalacturonan hydrolases and endo-rhamnogalacturonase, rhamnogalacturonan rhamnohydrolase and α -rhamnosidase are involved in depolymerization of xylogalacturonan and rhamnogalacturonan respectively [10]. In most of the white rot fungi, genes encoding for polysaccharide lyase classes PL-1, PL-3 and PL-9 were completely reduced to 0. The number of pectinolytic enzymes encoding genes varied between 10 (*P. carnosus*) to 49 (*B. cinerea*). Genes encoding for the polysaccharide lyase (PL-1, PL-3, PL-4 and PL-9) and carbohydrate esterase class-12 were completely reduced to 0 or 1, in all the selected brown rot fungi. Compared to white rot and brown rot fungi, total number of pectinolytic enzymes encoding genes varies in between 11 in *T. reesei* (low) to 24 *F. hepatica* (high). Interestingly in *R. oryzae*, genes encoding for pectinolytic enzymes were completely reduced to 0 (except GH-28(18) and CE-8). These results convey that tentative overall pectinolytic ability of the soft rot fungi is higher than the selected white rot and brown rot fungi.

6.3.6. Total lignocellulolytic abilities of selected fungi

The total number of lignin, cellulose, hemicellulose and pectin degrading CAZymes in all the selected white rot, brown rot and soft rot fungi were separated. We have tentatively calculated the total ligninolytic, cellulolytic, hemicellulolytic and pectinolytic abilities by taking average of all the lignocellulolytic CAZymes individually. These results have revealed that, highest ligninolytic ability was observed in *P. eryngii*, *P. ostreatus*, *S. hirsutum* (white rot fungi), *H. pinastri*, *C. puteana* (brown rot fungi) and *B. dothidea*, *F. oxysporum* (soft rot fungi) respectively. Similarly, the highest cellulolytic ability was observed in *P. eryngii*, *P. ostreatus* (white rot fungi), *H. pinastri*, *C. puteana* (brown rot) and *F. oxysporum*, *Lecythophora* sp (soft rot fungi) respectively. Highest hemicellulolytic ability was observed in *B. cinerea*, *S. hirsutum* (white rot fungi), *C. puteana*, *F. pinicola* (brown rot fungi) and *B. dothidea*, *F. oxysporum* (soft rot fungi) respectively. Finally, highest pectinolytic ability was observed in *B. cinerea*, *P. eryngii* (white rot fungi), *F. hepatica*, *F. pinicola*, *G. trabeum* (brown rot fungi) and *B. dothidea*, *F. oxysporum* (soft rot fungi) respectively (Figure 6.4). We have averaged the total number of genes encoding for lignocellulolytic enzymes to tentatively, to find the overall highest lignocellulolytic abilities. Results obtained from this analysis suggests that white rot fungi (85)

possess highest ligninolytic capacity followed by soft rot fungi (71). At the same time, soft rot fungi exhibited potentially highest cellulolytic, hemicellulolytic and pectinolytic abilities by harboring higher number of genes (Figure 6.4). Interestingly, the total average of genes encoding for various enzymes involved in cellular processes and signaling were higher in white rot fungi followed by soft rot fungi. Whereas genes encoding for information storage and processing, metabolism processes were higher in soft rot fungi followed by white rot fungi respectively (Figure 6.4).

White rot fungi														
Total count	P.ch	P.ca	C.su	H.an	F.me	P.cin	D.sq	T.ve	P.st	P.br	B.ci	P.os	S.hi	P.er
Ligninolytic	89	84	63	76	77	70	84	89	75	93	70	111	100	113
Cellulolytic	50	52	38	39	46	42	44	51	49	48	46	68	53	69
Hemicellulolytic	39	38	31	41	36	34	42	39	52	37	63	58	62	58
Pectinolytic	11	10	13	23	34	19	23	21	39	16	49	32	41	46

Brown rot fungi															
Total count	P.pl	F.ra	W.co	D.pr	D.qu	L.su	P.pl	N.le	S.la	C.co	G.tr	F.he	F.pi	H.pi	C.pu
Ligninolytic	27	33	27	18	32	46	40	41	37	25	21	36	38	53	51
Cellulolytic	29	29	31	40	33	38	31	42	41	40	43	39	39	51	49
Hemicellulolytic	23	30	29	33	31	27	25	29	30	31	33	35	36	35	52
Pectinolytic	14	18	17	12	16	17	16	18	14	13	22	24	22	16	21

Soft rot fungi														
Total Count	T.re	R.or	A.we	P.ch	D.es	Hypo	A.ni	Hy.EC38	N.cr	Hy.CO27	F.ox	Lecy	B.do	
Ligninolytic	31	14	64	48	88	92	61	96	49	95	104	77	105	
Cellulolytic	37	26	40	44	54	55	44	62	36	60	76	63	60	
Hemicellulolytic	43	54	66	64	78	73	63	80	51	80	135	95	104	
Pectinolytic	11	24	35	26	20	24	49	22	13	22	79	30	58	

No. of genes	WR	BR	SR
Ligninolytic	85.2857	35	71.0769
Cellulolytic	49.6429	38.3333	50.5385
Hemicellulolytic	45	31.9333	75.8462
Pectinolytic	26.9286	17.3333	31.7692

KOG Processes	WR	BR	SR
Cellular processes and signaling	2442.714	2227	2362.769
Information, storage and processing	1624.071	1547.667	1651
Metabolism	2240.571	2103.133	2745.308
Poorly Characterized	1594.214	1462.267	1895.538

Figure 6.4: Heatmaps showing the genome wide distribution of total ligninolytic, cellulolytic, hemicellulolytic and pectinolytic CAZymes in selected popular white rot, brown rot and soft rot fungi and tentative overall

6.4. Conclusions

Degradation of plant biomass by fungi is a highly researched subject for several years. Fungi are the most efficient degraders of the plant biomass (most abundant carbon source on the earth's surface) and natural scavengers in the environment thus playing a key role in maintenance of the global carbon cycle. Enzyme systems secreted by fungi are commercially applied in various industries such as paper, pulp, detergents, textile, wine industries and

especially fungi are highly studied and applied in the growing biofuel, biorefinery and bioremediation industries [65]. The decaying ability of the wood rotting fungi is directly proportional to its ability to secrete plant biomass degrading enzymes. In this study, we have performed a comparative analysis to understand the genome wide distribution of lignocellulolytic CAZymes among well-known 14 white rot, 15 brown rot and 13 soft rot fungi. We have separated and classified genome wide wood rotting fungal CAZymes into lignin, cellulose, hemicellulose and pectin degrading enzymes. The total number of genes encoding for ligninolytic, cellulolytic, hemicellulolytic and pectinolytic enzymes calculated in this study reveals that white rot fungi are well equipped with efficient enzyme machinery for the degradation of lignin. The total ligninolytic abilities of white rot fungi (lignin degrading auxiliary activity enzymes and cytochrome P450 monooxygenases) was significantly higher than that of soft rot fungi and brown rot fungi. In contrast, total cellulolytic, hemicellulolytic and pectinolytic abilities were highest in soft rot fungi followed by white rot and brown rot fungi. These results suggest that white rot fungal strains are highly suitable for the degradation of lignin, other aromatic compounds and environmental pollutants, soft rot fungal strains are highly suitable in cellulose, hemicellulose and pectin degradation studies thus highly suitable in biofuel and biorefining industries. We understand that the number of protein encoding (lignocellulolytic enzymes) genes do not totally determine the complete lignocellulolytic capacity of the fungi as the expression and turnover of these lignocellulolytic enzymes is dependent on various factors and enzymes. However, this study provides preliminary genomic details which are enough to decide on a strain which is comparatively better from the other strains. We believe that understanding the genetic material coding for the lignocellulolytic enzymes will significantly benefit researchers to choose genetically better strain for their studies. However, further relevant studies must be conducted to optimize the appropriate growth and environmental conditions to enhance the expression and protein turnover of these lignocellulolytic enzymes.

References

1. Daniel G. Fungal and bacterial biodegradation: white rots, brown rots, soft rots, and bacteria. *Deterioration and Protection of Sustainable Biomaterials*: ACS Publications; 2014. p. 23-58.
2. Kendrick B. *Fungi: ecological importance and impact on humans*. eLS. 2001.
3. Kameshwar AKS, Qin W. Lignin Degrading Fungal Enzymes. *Production of Biofuels and Chemicals from Lignin*: Springer; 2016. p. 81-130.
4. Rowell RM, Barbour RJ. *Archaeological wood: properties, chemistry, and preservation*: ACS Publications; 1989.
5. Daniel G. Use of electron microscopy for aiding our understanding of wood biodegradation. *FEMS microbiology reviews*. 1994; 13: 199-233.

6. Daniel G. Microview of wood under degradation by bacteria and fungi. ACS Publications; 2003.
7. Hatakka A. Biodegradation of lignin. Biopolymers Online. 2005.
8. Arantes V, Goodell B. Current understanding of brown-rot fungal biodegradation mechanisms: a review. Deterioration and protection of sustainable biomaterials: ACS Publications; 2014. p. 3-21.
9. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic acids research. 2013; 42: D490-D5.
10. Rytioja J, Hildén K, Yuzon J, Hatakka A, de Vries RP, Mäkelä MR. Plant-polysaccharide-degrading enzymes from basidiomycetes. Microbiology and Molecular Biology Reviews. 2014; 78: 614-49.
11. Vogel J. Unique aspects of the grass cell wall. Current opinion in plant biology. 2008; 11: 301-7.
12. Harris PJ, Stone BA. Chemistry and molecular organization of plant cell walls. Biomass recalcitrance: deconstructing the plant cell wall for bioenergy. 2008: 61-93.
13. Sjostrom E. Wood chemistry: fundamentals and applications: Elsevier; 2013.
14. Fernandez-Fueyo E, Ruiz-Dueñas FJ, Ferreira P, Floudas D, Hibbett DS, Canessa P, et al. Comparative genomics of *Ceriporiopsis subvermispora* and *Phanerochaete chrysosporium* provide insight into selective ligninolysis. Proceedings of the National Academy of Sciences. 2012; 109: 5458-63.
15. Olson Å, Aerts A, Asiegbu F, Belbahri L, Bouzid O, Broberg A, et al. Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. New Phytologist. 2012; 194: 1001-13.
16. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, et al. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science. 2012; 336: 1715-9.
17. Suzuki H, MacDonald J, Syed K, Salamov A, Hori C, Aerts A, et al. Comparative genomics of the white-rot fungi, *Phanerochaete carnosa* and *P. chrysosporium*, to elucidate the genetic basis of the distinct wood types they colonize. BMC genomics. 2012; 13: 444.
18. Lévassieur A, Lomascolo A, Chabrol O, Ruiz-Dueñas FJ, Boukhris-Uzan E, Piumi F, et al. The genome of the white-rot fungus *Pycnoporus cinnabarinus*: a basidiomycete model with a versatile arsenal for lignocellulosic biomass breakdown. BMC Genomics. 2014; 15: 486.
19. Martinez D, Larrondo LF, Putnam N, Gelpke MDS, Huang K, Chapman J, et al. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. Nature biotechnology. 2004; 22: 695.
20. Ohm RA, Riley R, Salamov A, Min B, Choi I-G, Grigoriev IV. Genomics of wood-degrading fungi. Fungal Genetics and Biology. 2014; 72: 82-90.
21. Binder M, Justo A, Riley R, Salamov A, Lopez-Giraldez F, Sjökvist E, et al. Phylogenetic and phylogenomic overview of the Polyporales. Mycologia. 2013; 105: 1350-73.
22. Amselem J, Cuomo CA, Van Kan JA, Viaud M, Benito EP, Couloux A, et al. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS genetics. 2011; 7: e1002230.
23. Riley R, Salamov AA, Brown DW, Nagy LG, Floudas D, Held BW, et al. Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. Proceedings of the National Academy of Sciences. 2014; 111: 9923-8.
24. Castanera R, López-Varas L, Borgognone A, LaButti K, Lapidus A, Schmutz J, et al. Transposable elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles. PLoS genetics. 2016; 12: e1006108.
25. Alfaro M, Castanera R, Lavín JL, Grigoriev IV, Oguiza JA, Ramírez L, et al. Comparative and transcriptional analysis of the predicted secretome in the lignocellulose-degrading basidiomycete fungus *Pleurotus ostreatus*. Environmental microbiology. 2016; 18: 4710-26.
26. Camarero S, Sarkar S, Ruiz-Dueñas FJ, Martínez MaJ, Martínez ÁT. Description of a versatile peroxidase involved in the natural degradation of lignin that has both manganese peroxidase and lignin peroxidase substrate interaction sites. Journal of Biological Chemistry. 1999; 274: 10324-30.
27. Guillen F, Martinez AT, Martinez MJ. Substrate specificity and properties of the aryl-alcohol oxidase from the ligninolytic fungus *Pleurotus eryngii*. The FEBS Journal. 1992; 209: 603-11.
28. Matheny PB, Curtis JM, Hofstetter V, Aime MC, Moncalvo J-M, Ge Z-W, et al. Major clades of Agaricales: a multilocus phylogenetic overview. Mycologia. 2006; 98: 982-95.
29. Ruiz-Dueñas FJ, Martínez MJ, Martínez AT. Molecular characterization of a novel peroxidase isolated from the ligninolytic fungus *Pleurotus eryngii*. Molecular microbiology. 1999; 31: 223-35.

30. Martinez D, Challacombe J, Morgenstern I, Hibbett D, Schmoll M, Kubicek CP, et al. Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proceedings of the National Academy of Sciences*. 2009; 106: 1954-9.
31. Tang JD, Perkins AD, Sonstegard TS, Schroeder SG, Burgess SC, Diehl SV. Short-read sequencing for genomic analysis of the brown rot fungus *Fibroporia radiculosa*. *Applied and environmental microbiology*. 2012; 78: 2272-81.
32. Nagy LG, Riley R, Tritt A, Adam C, Daum C, Floudas D, et al. Comparative genomics of early-diverging mushroom-forming fungi provides insights into the origins of lignocellulose decay capabilities. *Molecular biology and evolution*. 2015; 33: 959-70.
33. Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, Aerts A, et al. The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. *Science*. 2011; 333: 762-5.
34. Floudas D, Held BW, Riley R, Nagy LG, Koehler G, Ransdell AS, et al. Evolution of novel wood decay mechanisms in Agaricales revealed by the genome sequences of *Fistulina hepatica* and *Cylindrobasidium torrendii*. *Fungal Genetics and Biology*. 2015; 76: 78-92.
35. Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, et al. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature genetics*. 2015; 47: 410-5.
36. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, et al. Corrigendum: Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nature Biotechnology*. 2008; 26: 1193-.
37. Ma L-J, Ibrahim AS, Skory C, Grabherr MG, Burger G, Butler M, et al. Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS genetics*. 2009; 5: e1000549.
38. de Vries RP, Riley R, Wiebenga A, Aguilar-Osorio G, Amillis S, Uchima CA, et al. Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. *Genome biology*. 2017; 18: 28.
39. Van Den Berg MA, Albang R, Albermann K, Badger JH, Daran J-M, Driessen AJ, et al. Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. *Nature biotechnology*. 2008; 26: 1161-8.
40. Wu W, Davis RW, Tran-Gyamfi MB, Kuo A, LaButti K, Mihaltcheva S, et al. Characterization of four endophytic fungi as potential consolidated bioprocessing hosts for conversion of lignocellulose into advanced biofuels. *Applied microbiology and biotechnology*. 2017; 101: 2603-18.
41. Andersen MR, Salazar MP, Schaap PJ, Van De Vondervoort PJ, Culley D, Thykaer J, et al. Comparative genomics of citric-acid-producing *Aspergillus niger* ATCC 1015 versus enzyme-producing CBS 513.88. *Genome Research*. 2011; 21: 885-97.
42. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature*. 2003; 422: 859.
43. Damm U, Fourie P, Crous P. *Coniochaeta* (Lecythophora), *Collophora* gen. nov. and *Phaeomoniella* species associated with wood necroses of *Prunus* trees. *Persoonia: Molecular Phylogeny and Evolution of Fungi*. 2010; 24: 60.
44. U'Ren JM, Lutzoni F, Miadlikowska J, Laetsch AD, Arnold AE. Host and geographic structure of endophytic and endolichenic fungi at a continental scale. *American Journal of Botany*. 2012; 99: 898-914.
45. Desprez-Loustau M-L, Marçais B, Nageleisen L-M, Piou D, Vannini A. Interactive effects of drought and pathogens in forest trees. *Annals of Forest Science*. 2006; 63: 597-612.
46. Piškur B, Pavlic D, Slippers B, Ogris N, Maresi G, Wingfield MJ, et al. Diversity and pathogenicity of *Botryosphaeriaceae* on declining *Ostrya carpinifolia* in Slovenia and Italy following extreme weather conditions. *European Journal of Forest Research*. 2011; 130: 235-49.
47. Schoch CL, Shoemaker RA, Seifert KA, Hambleton S, Spatafora JW, Crous PW. A multigene phylogeny of the Dothideomycetes using four nuclear loci. *Mycologia*. 2006; 98: 1041-52.
48. Slippers B, Crous PW, Denman S, Coutinho TA, Wingfield BD, Wingfield MJ. Combined multiple gene genealogies and phenotypic characters differentiate several species previously identified as *Botryosphaeria dothidea*. *Mycologia*. 2004; 96: 83-101.

49. Slippers B, Wingfield MJ. Botryosphaeriaceae as endophytes and latent pathogens of woody plants: diversity, ecology and impact. *Fungal biology reviews*. 2007; 21: 90-106.
50. Ma L-J, Van Der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*. 2010; 464: 367.
51. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics*. 2004; 20: 1453-4.
52. Kameshwar AKS, Qin W. Recent developments in using advanced sequencing technologies for the genomic studies of lignin and cellulose degrading microorganisms. *International journal of biological sciences*. 2016; 12: 156.
53. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otilar R, et al. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research*. 2013: gkt1183.
54. Črešnar B, Petrič Š. Cytochrome P450 enzymes in the fungal kingdom. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*. 2011; 1814: 29-35.
55. van Gorcom RF, van den Hondel CA, Punt PJ. Cytochrome P450 enzyme systems in fungi. *Fungal Genetics and Biology*. 1998; 23: 1-17.
56. Kameshwar AKS, Qin W. Gene expression metadata analysis reveals molecular mechanisms employed by *Phanerochaete chrysosporium* during lignin degradation and detoxification of plant extractives. *Current Genetics*. 2017: 1-18.
57. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic acids research*. 2009; 37: D233-D8.
58. Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal*. 1991; 280: 309-16.
59. Henrissat B, Davies G. Structural and sequence-based classification of glycoside hydrolases. *Current opinion in structural biology*. 1997; 7: 637-44.
60. Lévassieur A, Piumi F, Coutinho PM, Rancurel C, Asther M, Delattre M, et al. FOLy: an integrated database for the classification and functional annotation of fungal oxidoreductases potentially involved in the degradation of lignin and related aromatic compounds. *Fungal genetics and biology*. 2008; 45: 638-45.
61. Lévassieur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for biofuels*. 2013; 6: 41.
62. van den Brink J, de Vries RP. Fungal enzyme sets for plant polysaccharide degradation. *Applied microbiology and biotechnology*. 2011; 91: 1477.
63. Ghosh M, Nanda G. Purification and some properties of a xylanase from *Aspergillus sydowii* MG49. *Applied and environmental microbiology*. 1994; 60: 4620-3.
64. Polizeli M, Rizzatti A, Monti R, Terenzi H, Jorge JA, Amorim D. Xylanases from fungi: properties and industrial applications. *Applied microbiology and biotechnology*. 2005; 67: 577-91.
65. Mäkelä MR, Donofrio N, de Vries RP. Plant biomass degradation by fungi. *Fungal Genetics and Biology*. 2014; 72: 2-9.

Chapter-7

Genome Wide Analysis Reveals the Extrinsic Cellulolytic Abilities of *Neocallimastigomycota* Fungi

[This work has been published in “**Journal of Genomics**” 6 (2018): 74-87]

Ayyappa Kumar Sista Kameshwar and Wensheng Qin*

7.1. Abstract

Ruminating animals, especially cattle lack the carbohydrate active enzyme encoding genes which are required for the degradation of the glycosidic linkages of plant cell wall carbohydrates (such as cellulose, hemicellulose, lignin and pectin). Thus, ruminating animals are completely dependent on the microorganisms (anaerobic bacteria and fungi, methanogenic archaea and protozoa) residing in their rumen (hindgut). In this study, we have retrieved and analyzed the complete genome wide annotations of the *Neocallimastigomycota* division fungi such as *Anaeromyces robustus*, *Neocallimastix californiae*, *Orpinomyces* sp, *Piromyces finnis*, *Piromyces* sp E2. We have retrieved the InterPro, CAZy, KOG, KEGG, SM Clusters and MEROPS genome level data of these anaerobic fungi from JGI-MycoCosm database. Results obtained in our study reveals that, the genomes of anaerobic fungi completely lack genes encoding for lignin degrading auxiliary activity enzymes. Contrastingly, these fungi outnumbered other fungi by having highest number of CAZyme encoding genes. The genes encoding for dockerins and carbohydrate binding modules exaggerated other CAZymes which are involved in the structure and functioning of cellulosomes. Presence of cellulosomes and higher number of carbohydrate transport and metabolism genes also endorses the plant cell wall carbohydrate degrading abilities of these fungi. We also reported the tentative total cellulolytic, hemicellulolytic and pectinolytic abilities. And we have explicitly reported the genes, enzymes and the mechanisms involved in structure and functioning of the cellulosomes. Our present work reveals the genomic machinery underlying the extrinsic plant cell wall degrading abilities of the anaerobic fungi. Results obtained in our study can be significantly applied in improving the gut health of cattle and especially in the fields of biofuel, biorefining and bioremediation-based industries.

Keywords: Ruminating animals (cattle), *Neocallimastigomycota* (Anaerobic fungi), Cellulose, Plant biomass, Cellulosomes,

7.2. Introduction

Increasing global temperatures, over dependence and depletion of fossil fuels to meet the increasing fuel needs, have forced the mankind to produce sustainable renewable energy systems. Plant biomass which contains lignocellulose polymers constitutes the most abundant component on the earth's surface. Separation and production of renewable energy from the lignocellulosic biomass is a complex procedure and requires application of different chemical, physical and mechanical methods. However, biological way of biofuel production from lignocellulosic biomass is the most preferred and researched due to its ecofriendly and cost-effective nature. Naturally, enormous amounts of cellulose are digested by the herbivorous ruminating animals (having plant biomass containing diets e.g. cows, sheep, buffalo, sheep, deer, goats etc.) and this process of digestion is solely supported by microorganisms residing in its rumen. Gut of the ruminating animals can be divided into four chambers they are rumen, reticulum, omasum, abomasum (or stomach). The process of food digestion starts from the rumen, saliva mixed food is mechanically broken down to smaller pieces which is passed into the reticulum where it separates the food into digestible and non-digestible forms into cuds (partly degraded food). The partially digested cuds are regurgitated, which is further rechewed and swallowed by the ruminating animals. Thus, from the partially digested food, omasum absorbs water, nutrients, vitamins, fatty acids etc. Major part of the cellulose fermentation happens in the rumen due to the presence of cellulolytic microorganisms. Studies have reported that anaerobic bacteria such as *Ruminococcus* genus (e.g. *Ruminococcus albus*, *Ruminococcus flavefaciens*) *Streptococcus*, *Escherichia*, *Megasphaera*, *Fibrobacter* are significantly involved in cellulose degradation [1-5].

In the year 1970, Colin Orpin has identified the anaerobic fungi based on its zoospores and chitin containing cell walls, which are now classified under the division of *Neocallimastigomycota*. This peculiar group of anaerobic fungi has raised considerable curiosity among mycologists worldwide. Two major reasons which brings *Neocallimastigomycota* fungi under the lime light are its distinctive physiology and potential applications in the fields of biomass conversion and animal nutrition [6]. Naturally these anaerobic fungi reside in the gastrointestinal tracts of herbivores and significantly aids in digestion of different plant biomass components. Anaerobic fungi were observed among all the foregut fermenters (where most of the digestion occurs before the gastric digestion) e.g. ruminants, pseudo-ruminants and foregut non-ruminants and also in most of the hindgut fermenters (where the most of the fermentation occurs after gastric digestion in the large

intestine and caecum) [7]. Significantly, the anaerobic fungi residing in the gut of ruminating animals performs two major functions plant biomass digestion and forming a dedicated digestive chamber with neutral pH conditions [7]. Herbivorous animals lack the ability to digest plant biomass components as they cannot secrete cellulolytic/hemicellulolytic enzymes, in turn they depend on symbiotic gut microorganisms such as anaerobic bacteria and fungi, methanogenic archaea and protozoa [7]. Among the above-mentioned microorganisms, anaerobic fungi were found to play a crucial role in degradation of lignocellulosic components of plant biomass.

Studies conducted by Heath et al (1986), has revealed that reproduction in anaerobic fungi commences through asexual reproduction by releasing flagellated zoospores from sporangia[8]. Previous reports reveal that ingestion of food by the ruminating animals induces the anaerobic fungi to release its zoospores from the sporangia and it was also reported that within 30-60 minutes the density of zoospore peaks in the rumen [7, 9-12]. According to Orpin & Greenwood (1986), haem and other related porphyrin compounds released from the ingested plant materials trigger the sporangia of anaerobic fungi which induces the process of differentiation and maturation of zoospore in the rumen [13]. The flagellated zoospores of anaerobic fungi are motile and they colonize on the plant material based on the chemotactic responses from the surrounding sugars and phenolic compounds [14]. Flagellated zoospores transform to a cyst by shedding its flagella, once it attaches to the plant material. Formation and germination of cyst is involved by thickening of the cell wall and production of germ tube from the polar end and further development of cyst varies from monocentric to polycentric based on the organism. Endogenous cyst germination is observed in monocentric taxa as the nucleus is situated in the cyst forming a zoosporangium, by leaving rhizoids anucleate. Contrastingly, exogenous cyst germination is observed in polycentric taxa where the nucleus migrates to the rhizoids thus leading to the formation of multiple sporangia [7, 15]. The rhizomycelium (bulbous or filamentous) of the anaerobic fungi performs two major and important functions, it provides support to the growing sporangium (monocentric) or sporangia (polycentric) and performs the enzymatic digestion through penetrating into the plant material. Penetration of developing rhizoids into the rigid plant material, opens the internal plant tissues, making them susceptible to the enzymatic hydrolysis, which supports the developing sporangia with the nutrients for the development and maturation of the sporangia. Mature sporangia will produce zoospores ranging between 1 (less) to 80 (high), under suitable conditions (inducers), mature sporangium undergoes differentiation and

further releases zoospores by dissolving the sporangial cell wall. Previous studies report that it is tough to free host animals from anaerobic fungi, as they exhibit efficient dispersal of anaerobic fungi among the host animals through forming aerotolerant cellular structures with extreme survival abilities (e.g. 2-4 chambered spores of *Anaeromyces* sp) [16, 17]. Studies have also confirmed that, anaerobic fungi can be cultured from the faecal material of host animals followed by air drying, freezing and from the long-settled cow dung [7, 18-20].

Industrially, anaerobic fungi exhibit various biotechnological applications especially applied as microbial probiotic supplement to improve the process of digestion and thus utilizing the low-quality forages. Several studies were already being conducted on improving the feed intake, efficiency, development-growth rate of animals and especially in milk production [21-24]. Previous findings suggest that, dietary supplementation of anaerobic fungal enzymes (e.g. glycoside hydrolases) in cattle resulted to be more effective than the viable cultures (e.g. swine and poultry). According to Azain et al., (2002), dietary supplementation of glycoside hydrolases (to support in depolymerization of plant biomass components) obtained from anaerobic fungi improved the growth and development of broiler chicken by 25% [7, 25]. Anaerobic fungi were of high interest in brewing, food, paper, textile and biofuel industries, due to its ability to secrete polysaccharide degrading enzymes. Studies were also being conducted for developing anaerobic fungi based and fermentative production of cellulosic ethanol and renewable fuels from the agricultural residues. Youssef et al., (2013) has reported the lignocellulose degradation study using *Orpinomyces* sp strain C1A by simultaneously producing cellulosic ethanol, *Orpinomyces* sp was found to degrade 61.3% of corn stover (dry weight) resulting in 0.045 to 0.096 mg of ethanol per mg of biomass [26]. In this study we have specifically analyzed the genome wide architecture of five completely sequenced genomes of anaerobic fungi classified under *Neocallimastigomycota* division. We have compared the genome wide lignocellulolytic abilities of these fungi specifically by comparing genome wide annotations such as CAZy, InterPro, MEROPS and SM (secondary metabolite) Clusters databases. Our study will reveal lignocellulolytic, detoxifying and degrading abilities of the anaerobic fungi.

7.3. Data Retrieval and Analysis

7.3.1. Data retrieval: In our present study, we have selected and retrieved the genome level data of 5 anaerobic fungi classified under the *Neocallimastigomycota* division (*Anaeromyces robustus* v1.0 (Anaspl) [27], *Neocallimastix californiae* G1 v1.0 (Neospl) [27], *Orpinomyces*

sp. (Orpsp1_1) [26], *Piromyces finnis* v3.0 (Pirfi3) [27], *Piromyces* sp. E2 v1.0 (PirE2_1) [27]). The genome level data of all the above selected fungi were retrieved from the JGI-MycoCosm repository (<https://genome.jgi.doe.gov/programs/fungi/index.jsf>).

7.3.2. Data Analysis: From JGI-MycoCosm database we have specifically selected and retrieved the genome wide annotations such as InterPro (a database for the protein families, domains and functional sites), CAZy (database for carbohydrate active enzymes), KOG (eukaryotic orthologous group) and SM Clusters (database for secondary metabolite gene clusters) of the above listed fungi. To understand and reveal the distribution of plant cell wall degrading enzymes and their evolutionary gene losses among the selected fungi. We have segregated and compared the total number of genome wide InterPro annotations of the selected fungi into protein domains occurring in a) multiple copies and b) single copies. Similarly, we have segregated and compared the total number of genes encoding for the glycoside hydrolases (GH), glycosyl transferases (GT), carbohydrate binding modules (CBM), auxiliary activity (AA), polysaccharide lyases (PL), dockerin (DOC) and expansin (EXPN) classes and total number of genome wide carbohydrate active enzymes among the selected fungi. We have also retrieved and compared the KOG (eukaryotic orthologous) groups a) cellular signaling and processing (CSP) b) information storage and processing (ISP) c) metabolism and d) poorly characterized. Finally, we have retrieved and compared the SM (secondary metabolite) clusters such as DMAT (Di Methyl Allyl Tryptophan Synthase), HYBRID (hybrid genes), NRPS (Non-Ribosomal Peptide Synthetases), Poly-ketide Synthases (PKS) and MEROPS (database for the peptidases) for comparing the distribution of proteases among the selected fungal genomes We have used Microsoft excel program for comparing and representing the genomes of the selected fungi. The analysis pipeline implemented in our present study is extensively represented in the Figure 7.1.

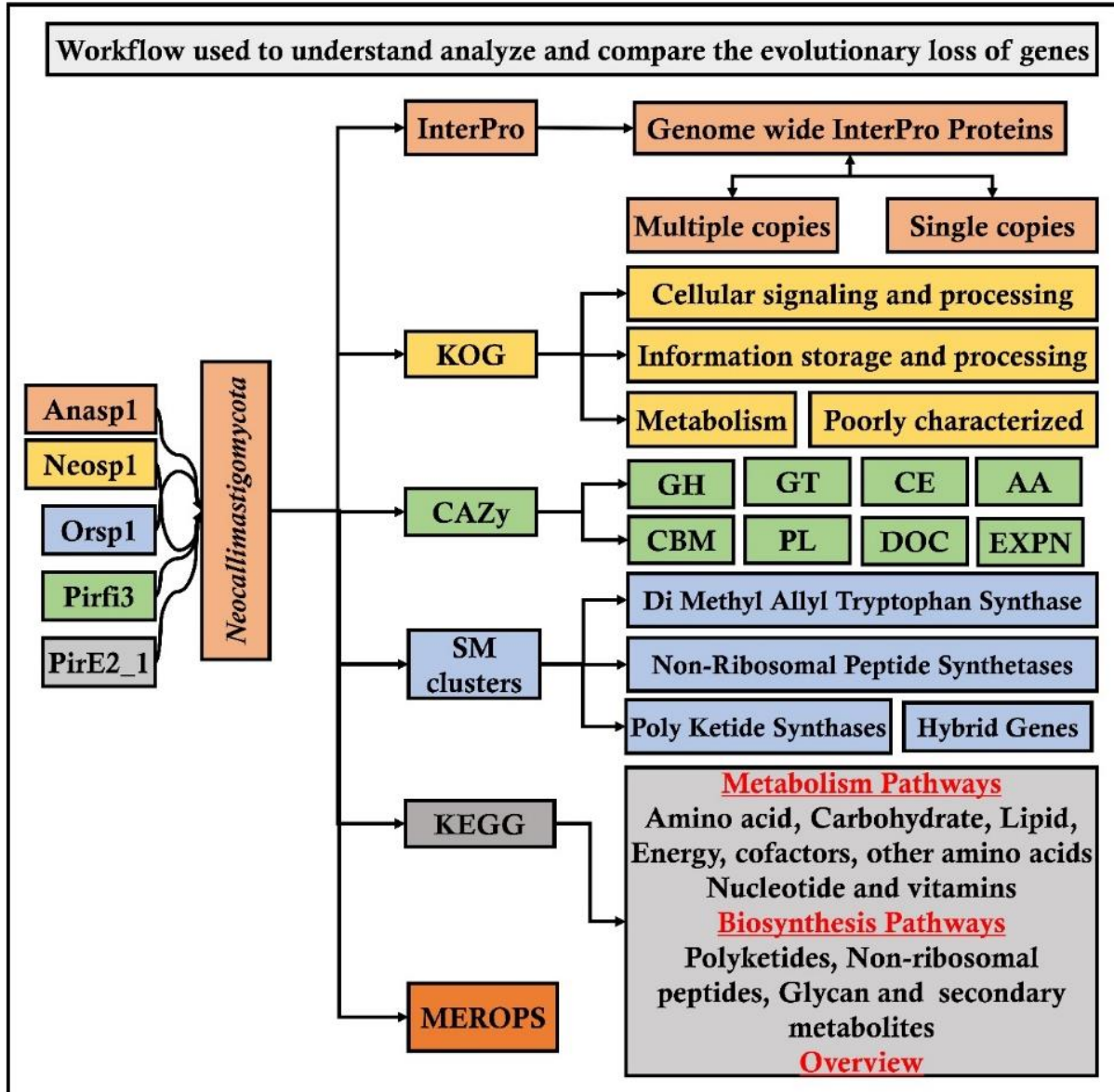


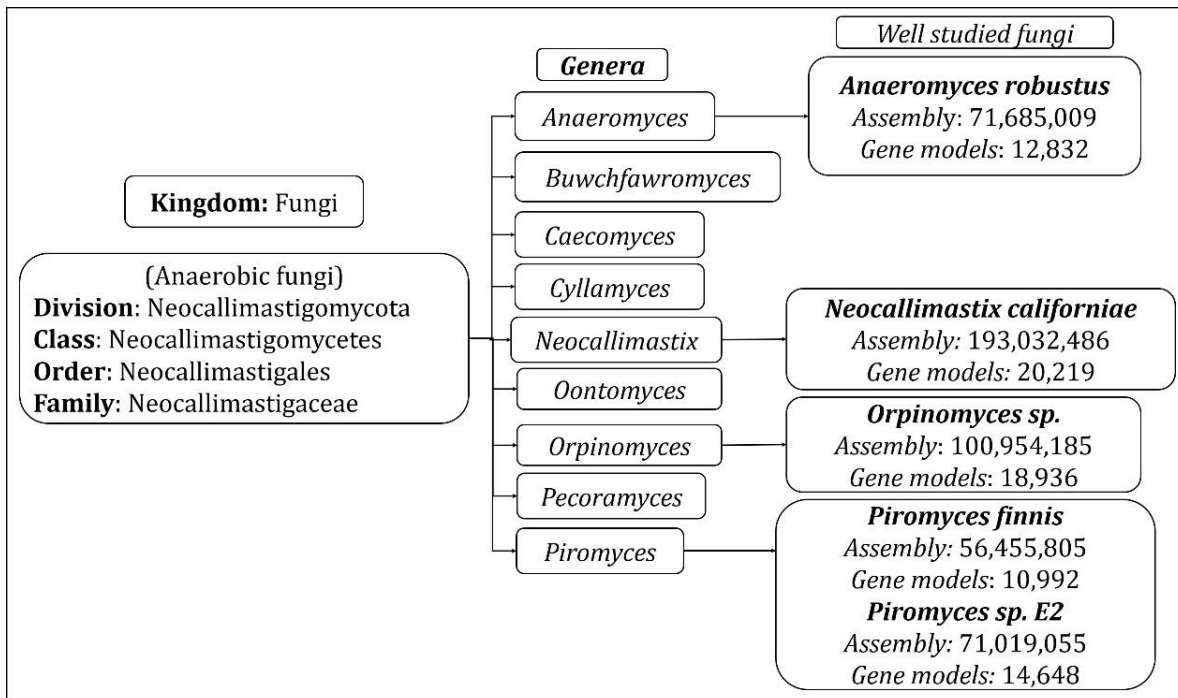
Figure 7.1: Workflow pipeline implemented for analyzing and comparing the genomes of *Neocallimastigomycota* division fungi

7.4. Results and Discussions

Genomic studies of *Neocallimastigomycota* fungi has revealed various interesting facts about its plant cell wall decaying abilities and about the evolutionary loss of several other genes involved in various metabolic processes. *Neocallimastigomycota* fungi were found to be closely related to the phylum *Chytridiomycota*. Though *Neocallimastigomycota* fungi share some of the key morphological characters as *Chytridiomycota* fungi, they exhibit some of the unique features such as their cellular physiology suitable for its anaerobic living and peculiar flagellar movement [7]. Later conducted genomic studies have endorsed the separation of

Neocallimastigomycota into a separate clade basal to *Chytridiomycota* fungi [28-30]. *Neocallimastigomycota* fungi comprises of six different genera which can be separated based on its distinguishable morphological properties such as rhizoidal and bulbous morphologies of thallus and zoospore flagellation (mono vs polyflagellate) [7, 31, 32]. Currently, *Neocallimastigomycota* division comprises of nine genera *Anaeromyces*, *Buwchfawromyces*, *Caecomyces*, *Cyllamyces*, *Neocallimastix*, *Oontomyces*, *Orpinomyces*, *Pecoromyces* and *Piromyces*, where each genus possesses some unique and distinguishable morphological characteristics. Phylogenetic studies conducted in the past, based on the conserved DNA sequences such as 18S RNA and ITS (Inter transcribed spacer regions) sequences have reported the close relationship among *Neocallimastix*, *Orpinomyces* and *Caecomyces*, *Cyllamyces* genera respectively [7, 33, 34]. Recent whole genome sequencing studies of anaerobic fungi have reported the complete assembled and annotated genomes, with total number of assembled gene models ranging between 10,992-*Piromyces finnis*, 12,832-*Anaeromyces robustus*, 14,648-*Piromyces* E2 18,936-*Orpinomyces* and 20,219-*Neocallimastix californiae* respectively [26, 27] (Figure 7.2).

Genomes of *Neocallimastigomycota* fungi exhibited nearly about 3221 (*Anasp1-Anaeromyces robustus* v1.0 [27]), 3313 (*Neosp1-Neocallimastix californiae* G1 v1.0 [27]), 2598 (*Orsp1-Orpinomyces* sp. [26]), 3190 (*Pirfi3-Piromyces finnis* v3.0 [27]) and 2795 (*Piromyces* sp. E2 v1.0 [27]) unique InterPro annotated domains respectively. Majority of these protein domains were found to occur in multiple copies in the genomes, making the genome wide distribution of InterPro protein domains to 30969 (*Anasp1*), 46794 (*Neosp1*), 35585 (*Orsp1*), 27050 (*Pirf3*) and 28577 (*PirE2*) respectively. Results obtained in our study reveals that genes encoding for ankyrin, WD-40 repeat, chitin-binding type-I, Src homology-3, cellulose binding region, protein kinases, leucine rich repeats, ABC transporters, dockerin (cellulose binding family 5), RNA recognition motif (RNP-1), calcium binding EF, Armadillo-type fold, serine/threonine protein kinase, zinc finger (RING, C2H2-type), tyrosine protein kinase, glycoside hydrolase, heat shock protein (DnaJ), tetratricopeptide region, AAA+ ATPase, spore coat protein and kinesin protein domain were observed in high numbers among the genomes of anaerobic fungi.



	Genera	Thallus	Rhizoids	Zoospore Flagellation
Phylogenetic relatedness (ITS region)	<i>Neocallimastix</i>	Monocentric	Filamentous	Polyflagellate
	<i>Piromyces</i>	Monocentric	Filamentous	Uni (bi or tetra) flagellate
	<i>Caecomyces</i>	Monocentric	Bulbous	Uni (bi or tetra) flagellate
	<i>Orpinomyces</i>	Polycentric	Filamentous	Polyflagellate
	<i>Anaeromyces</i>	Polycentric	Filamentous	Uni flagellate
	<i>Cyllamyces</i>	Polycentric	Bulbous	Uni (bi or tetra) flagellate

Legend:

Figure 7.2: (A) Hierarchical delineation of *Neocallimastigomycota* division and well-studied fungi with available annotated genomes, (B) phylogenetic relationship of the *Neocallimastigomycota* fungi based on the conserved ITS (Inter Transcribed Spacer regions) and other morphological characteristics, where the legend shows the phylogenetic relatedness among the selected fungi.

To understand the genomic distribution of proteins we have analyzed the genome wide KOG annotations of the selected anaerobic fungi. We have retrieved the total number of gene models classified under cellular signaling and processing, information storage and processing, metabolism and poorly characterized KOG categories. The total KOG classified gene models were further analyzed to understand the proteins occurring in unique and multiple copies among the genomes of anaerobic fungi. Genes encoding for chitinases, leucine rich repeat, serine/threonine protein kinases, RNA-binding protein, lipid exporter,

ankyrim have outnumbered all other protein encoding genes (Figure 7.3A). We have observed that Neosp1 genome harbors higher number of gene models compared to other anaerobic fungal genomes, followed by Orsp1genomes respectively (Figure 7.3A). The ascending order of the anaerobic fungi based on their total KOG classified gene models Pirfi3-8462 < Anasp1-9556 < PirE2-9810 < Orsp1-12559 < Neosp1-14,449 respectively (Figure 7.3B).

A		Protein Description	Neosp1	Anasp1	Orsp1	Pirfi3	PirE2
		Chitinase	684	372	393	317	461
		Uncharacterized conserved protein	541	377	396	362	333
		FOG: Leucine rich repeat	233	185	311	168	147
		Serine/threonine protein kinase	224	245	248	129	201
		RNA-binding protein (RRM, Pumilio-like repeats)	201	276	303	136	188
		Lipid exporter ABCA1 and related proteins	196	112	182	64	254
		Ankyrin	189	173	482	114	263
		FOG: Ankyrin repeat	175	176	222	63	214
		ER-Golgi vesicle-tethering protein p115	164	77	133	147	77
		Nucleolar GTPase/ATPase p130	163	68	111	84	147
		C-type lectin	156	126	132	66	114
		beta-1,6-N-acetylglucosaminyltransferase	154	72	130	76	122
		Predicted transporter (major facilitator superfamily)	154	48	49	34	45
		RNA-binding Ran Zn-finger protein, related proteins	137	82	97	44	77
		Uncharacterized conserved coiled-coil protein	120	53	84	87	53
		von Willebrand factor related coagulation proteins	116	112	122	92	154
		Extracellular protein SEL-1, related proteins	114	57	78	60	57
		FOG: Zn-finger	99	73	52	68	68
		Gluconate transport-inducing protein	94	64	120	50	82
		Ca ²⁺ -modulated nonselective cation channel polycystin	84	68	56	48	64
		Predicted chitinase	84	26	71	48	119
		Signaling protein SWIFT and BRCT domain proteins	80	44	48	64	60
		FOG: RRM domain	59	34	38	32	36
		Carboxylesterase and related proteins	57	33	50	29	42
		Mitotic checkpoint protein MAD1	56	38	54	28	35
		Molecular chaperone (DnaJ superfamily)	52	36	34	32	27
		WD40 repeat-containing protein	51	30	38	31	26
		Predicted E3 ubiquitin ligase	49	34	35	33	29
		Myosin class II heavy chain	45	28	44	30	24
		Subtilisin-related protease/Vacuolar protease B	45	121	97	25	71
		Kinesin-like protein	39	26	43	35	26

	CSP
	M
	ISP
	PC

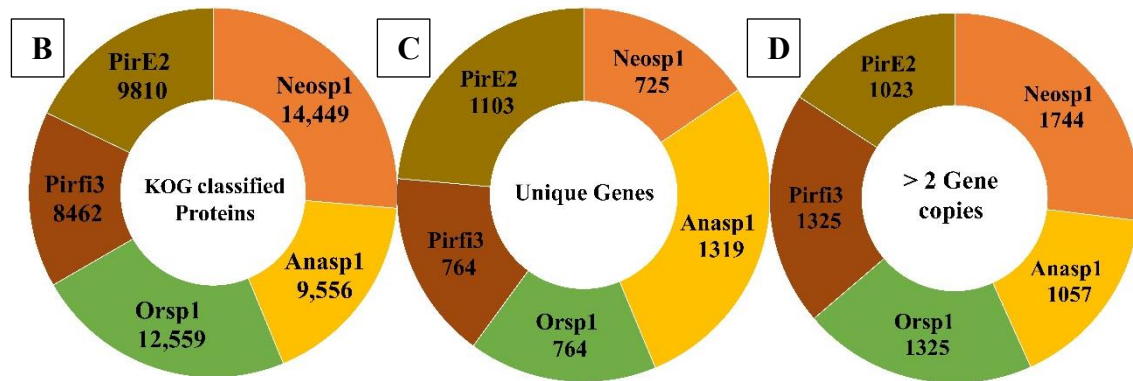


Figure 7.3: A) Genome wide distribution of KOG classified proteins under CSP (cellular signaling and processing), ISP (Information storage and processing), M (metabolism) and PC (poorly characterized) B) total number of KOG classified proteins C) unique genes D) proteins with multiple gene copies (>2 gene copies).

Results obtained from the classification of carbohydrate active enzymes (CAZymes) among *Neocallimastigomycota* (anaerobic) fungi, reveal that these fungi have suffered a severe evolutionary loss of genes encoding for lignin degrading enzymes. The total number of genes encoding for the auxiliary activity class enzymes such as lignin oxidizing enzymes (laccase, lignin peroxidase, manganese peroxidase, versatile peroxidase and cellobiose dehydrogenase) and lignin degrading auxiliary enzymes (aryl alcohol oxidase, vanillyl alcohol oxidases, glyoxal oxidases, pyranose oxidases, galactose oxidases, glucose oxidases and benzoquinone reductases) were completely reduced to zero. Importantly, genes encoding for GMC (glucose methanol choline) oxidoreductases, lytic polysaccharide monoxygenases (LPMO), cellobiose dehydrogenase (CDH) and iron reductase enzymes were completely reduced among the selected *Neocallimastigomycota* division fungi. Genes encoding for glycoside hydrolases, glycosyl transferases were found to range between 262-Anasp1(low) and 548-Neosp1 (high),104-Pirfi3(low) and 196-Neosp1 (high) respectively. While genes encoding for carbohydrate esterases, expansin and polysaccharide lyases ranges between 92-Pirfi3 (low) and 213-Neosp1 (high), 9-Pirfi3 (low) and 29-Neosp1 (high), 11-Anasp1(low) and 82-Neosp1(high) respectively (Figure 7.4A).

Contrastingly, the genomes of *Neocallimastigomycota* division fungi harbors higher number of genes encoding for carbohydrate binding domains (CBM) and dockerin proteins, which are found to be involved in the formation of cellulosomes (a complex structure involved in degradation of plant cell wall polysaccharides especially cellulose). The number of gene

models encoding for CBM and dockerin proteins ranges between 494-Pirfi3 (low) and 857-Neosp1(high), 469-Pirfi3 (low) and 838-Neosp1(high) respectively. Totally, the number of CAZymes encoding genes ranges between Neosp1-2763 (high) and Pirfi3-1468 (low) respectively (Figure 4A). Studies have reported that anaerobic fungi residing in the animal guts are majorly involved in digestion of plant biomass components (cellulose, hemicellulose, pectin and lignin). Previous studies have classified the CAZyme classes coding for cellulases, hemicellulases, ligninases and pectinolytic enzymes [35, 36]. The glycoside hydrolase enzyme classes GH-1, GH-2, GH-3, GH-5, GH-6, GH-8, GH-9, GH-38, GH-45, GH-48 and GH-74 code for cellulases respectively. Based on the total number of cellulase encoding genes Anasp1 exhibits lower and Neosp1 exhibits higher cellulolytic activity comparatively. Glycoside hydrolase classes GH-10, GH-11, GH-30, GH-31, GH-38, GH-39, GH-43, GH-45, GH-47, GH-53, GH-115, and carbohydrate esterase class enzymes CE-1, CE-2, CE-3, CE-4, CE-6 and CE-16 codes for hemicellulases respectively. Based on the total number of hemicellulase encoding genes Pirfi3 exhibits lower and Neosp1 exhibits higher hemicellulolytic activity. Similarly, polysaccharide lyase class enzymes PL-1, PL-3, PL-4, PL-9, PL-11 glycoside hydrolase class enzymes GH-28, GH-78, GH-95, GH-105, GH-115 and carbohydrate esterase class enzymes CE-8, CE-12 and CE-16 are involved in degradation of pectin respectively. We have theoretically predicted the total cellulolytic, hemicellulolytic and pectinolytic enzyme activities by calculating the total number of cellulases, hemicellulases and pectinolytic enzymes. Anasp1 and Pirfi3 exhibits lower total cellulolytic, hemicellulolytic and pectinolytic activities. Neosp1 exhibits higher total cellulolytic, hemicellulolytic and pectinolytic activities, whereas Orsp1 and PirE 2_1 exhibits total activities similarly (Figure 7.4B).

A	CAZy	Organisms				
		Anasp1	Neosp1	Orsp1	Pirfi3	PirE2
	AA	0	0	0	0	0
	GH	261	548	379	279	472
	GT	131	196	111	106	104
	CBM	650	857	744	494	821
	CE	121	213	119	92	139
	DOC	583	838	628	469	692
	EXPN	14	29	18	9	18
	PL	11	82	30	15	32
	Total	1771	2763	2029	1468	2282

B	Total activity	Anasp1	Neosp1	Orpsp1	PirE2	Pirfi3
	Cellulolytic	97	247	177	179	122
	Hemicellulolytic	211	394	271	297	182
	Pectinolytic	36	145	61	68	38

Hemicellulases						Cellulases						Pectinolytic Enzymes					
Class	Anasp1	Neosp1	Orpsp1	PirE2	Pirfi3	Class	Anasp1	Neosp1	Orpsp1	PirE2	Pirfi3	Class	Anasp1	Neosp1	Orpsp1	PirE2	Pirfi3
GH-10	15	60	32	29	21	GH-1	7	16	10	17	10	PL-1	6	33	19	17	10
GH-11	33	30	52	72	41	GH-2	1	7	1	1	1	PL-3	1	30	8	6	3
GH-30	2	4	3	3	1	GH-3	15	53	18	26	15	PL-4	3	17	3	9	2
GH-31	7	10	19	5	2	GH-5	26	72	51	46	29	PL-9	1	0	0	0	0
GH-38	2	1	2	2	1	GH-6	13	28	49	35	22	PL11	0	2	0	0	0
GH-39	5	9	3	5	2	GH-8	2	2	1	2	1	GH-28	0	1	0	1	0
GH-43	18	48	32	31	14	GH-9	9	14	13	12	12	GH-78	1	1	2	1	0
GH-45	14	29	16	20	15	GH-38	2	1	2	2	1	GH-95	2	2	0	1	1
GH-47	5	6	1	3	2	GH-45	14	29	16	20	15	GH-105	0	2	0	2	1
GH-53	1	3	0	2	1	GH-48	7	21	14	14	13	GH-115	1	9	3	4	3
GH-115	1	9	3	4	3	GH-74	1	4	2	4	3	CE-8	5	14	9	8	5
CE1	28	48	33	37	17							CE-12	6	8	2	7	6
CE2	11	8	6	10	1							CE16	10	26	15	12	7
CE3	1	1	1	3	0												
CE4	47	88	45	41	43												
CE6	11	14	8	18	11												
CE16	10	26	15	12	7												

C

Figure 7.4: Genome wide distributions of A) CAZymes, B) cellulolytic, hemicellulolytic and pectinolytic activities (GH- Glycoside hydrolases, CE- Carbohydrate esterases and PL- polysaccharide lyases) where Anasp1 (*Anaeromyces robustus*), Neosp1 (*Neocallimastix californiae*), Orpsp1 (*Orpinomyces sp.*), PirE2 (*Piromyces sp. E2*) and Pirfi3 (*Piromyces finnis*) respectively.

The cluster of orthologous groups (COG) is a prokaryotic database used for the identification of ortholog and paralog proteins. Similarly, KOG the eukaryotic version for cluster of orthologous groups detects the ortholog and paralog proteins in the given eukaryotic genome. The descending order of fungi based on their total number of KOG classified gene models are Neosp1>Orsp1>PirE2>Anasp1>Pirfi3 respectively (Figure 7.5A) The SM cluster database identifies the secondary metabolites of fungal genomes into three major classes as a) non-ribosomal peptide synthetases (NRPS) b) polyketide synthetases (PKS) and c) terpene synthetases (TS) respectively [37]. These key enzymes are involved in the production of the important fungal secondary metabolites a) non-ribosomal peptides and amino acid-derived compounds, (b) polyketides and fatty acid-derived compounds and (c) terpenes [37]. Using a set of modules (single ATC module or multimodule model with ATC

repeating units) NRPS catalyzes the biosynthesis of small peptides (a ribosome independent mechanism) [38]. According to Finking and Marahiel (2004), three core domains (adenylation (A), thiolation (T) and a condensation domain (C)) containing module catalyzes the peptide bond formation on the megasynthase complex [39]. Polyketide synthetase (PKS) fungal modules containing three domains a) ketoacyl synthase (KS), b) acyl transferase and c) phosphopantetheine site [40] (Figure 7.5C).

A

KOG	Anasp1	Neosp1	Orsp1	Pirf3	PirE2
CSP	3303	5217	4485	3021	3435
ISP	1792	2603	2155	1730	1764
M	1980	3077	2499	1670	2103
PC	2481	3552	3460	2041	2504

B

Genome	DMAT	HYBRID	NRPS	NRPS-Like	PKS	PKS-Like	TC	Total
Anasp1	0	0	28	4	6	8	0	46
Neosp1	0	0	10	6	14	9	0	39
Orsp1	0	0	97	49	2	7	0	155
Pirf3	0	1	1	1	8	2	0	13
PirE2	0	0	34	18	5	8	0	65

C

KEGG Classes	Anasp1	Neosp1	Orsp1	PirE2	Pirf3
Amino Acid Metabolism	270	418	266	229	224
Biosynthesis of Polyketides and Nonribosomal Peptides	43	63	54	31	41
Biosynthesis of Secondary Metabolites	126	203	137	118	121
Carbohydrate Metabolism	440	816	420	381	418
Energy Metabolism	90	153	87	89	78
Glycan Biosynthesis and Metabolism	217	309	199	149	196
Lipid Metabolism	201	332	177	175	185
Metabolism of Cofactors and Vitamins	250	368	219	183	224
Metabolism of Other Amino Acids	57	140	49	52	58
Nucleotide Metabolism	562	739	573	416	511
Overview	202	346	193	180	189
Xenobiotics Biodegradation and Metabolism	106	165	78	80	91

D

Type of Peptidase	Anasp1	Neosp1	Orsp1	PirE2	Pirf3
Aspartic Proteases	3	6	25	11	3
Cysteine Proteases	30	55	54	33	37
Metallo Proteases	37	49	33	34	27
Mixed Proteases	1	12	14	4	4
Serine Proteases	60	41	52	47	26
Zincin	48	45	116	66	30
Ntn-hydrolase	15	31	17	15	17
alpha/beta hydrolase	91	125	118	127	58
Unassigned	25	37	11	19	5

Figure 7.5: Genome wide distributions of A) KOG (Eukaryotic orthologous groups), B) SM (secondary metabolite) clusters C) distribution of proteins among KEGG classified pathway groups and D) clan-based classification of proteolytic enzymes

To broadly understand the genome wide distribution of genes involved in various cellular mechanisms, we have retrieved and compared the KEGG (Kyoto Encyclopedia of Genes and Genomes) annotations of the selected fungi. We have totally selected genes classified under 12 pathway classes (Figure 7.5B), out of these higher number of genes were classified under the carbohydrate and nucleotide metabolism pathways among all the selected anaerobic fungi. The descending order of the fungi based on the number of gene models classified among the 12 pathway classes were Neosp1>Orsp1>Anasp1>Pirfi3>PirE2_1 respectively (Figure 7.5B). The proteases (proteolytic enzymes or peptidases) play a crucial role in various molecular and biological processes. We have also analyzed the genome wide occurrence of proteolytic enzymes among the selected anaerobic fungi. MEROPS is public database for the proteolytic enzymes and their corresponding substrates and inhibitors respectively [41]. In MEROPS the proteolytic enzymes are classified based on its similitude of the protein structure at the tertiary and primary levels by specifically comparing the active and reactive sites of the proteases. The proteases are classified into families and clans respectively. Results obtained in our analysis shows that aspartic, mixed and zincin class proteases were high in Orsp1, cysteine, mixed, metallo, Ntn-hydrolase, alpha-beta hydrolase class proteases were high in Neosp1, serine proteases are high in Anasp1 and alpha/beta hydrolase proteases were high in PirE2 genomes respectively (Figure 7.5D).

Our analysis explicitly reports that the selected anaerobic fungi harbors large number of genes encoding for the carbohydrate transport and metabolism. Thus, we have specifically focused on the total number of genes involved in carbohydrate transport and metabolism. Interestingly, we have observed that genomes of the selected anaerobic fungi harbor higher number of chitinase encoding gene models ranging between 293 (Pirfi3) to 639 (Neosp1). Previous studies have already reported that amino acid tryptophan was found to conserved in the chitin binding domain in bacteria, these domains were also found to be involved during binding of cellulase with cellulose [42]. For our present analysis we have only considered only the enzymes with multiple gene copies (<2). We have observed that 20 protein encoding gene models involved in carbohydrate transport and metabolism were found to be occur in

multiple copies in the genomes of the selected anaerobic fungi genomes. Genes encoding for β -1,6-N-acetylglucosaminyltransferase, gluconate transport-inducing protein, β -glucosidase, Golgi mannosyltransferase, maltase glucoamylase (GH-31) were found to occur more than five copies in the genomes of anaerobic fungi respectively. Occurrence of higher number of gene models encoding for the carbohydrate transport metabolism proteins the ascending order based on the genomes were 492 (Pirfi3) > 578 (Anasp1) > 676 (Orsp1) > 693 (PirE2_1) > 1037 (Neosp1) respectively. These genomic evidences suggest that anaerobic fungi classified under *Neocallimastigomycota* has developed sophisticated organelles such as cellulosomes for the degradation of plant cell wall components [27, 43].

To understand the carbohydrate breakdown and metabolism by the *Neocallimastigomycota* division fungi, we have analyzed and compared the genomes of the selected anaerobic fungi to reveal the carbohydrate binding modules (CBM) and the corresponding carbohydrate interacting residues (Figure 7.6A). Our results report that genes encoding for carbohydrate binding modules (CBM18) interacting with chitin residues were found to be high in PirE2_1-640 and low in Pirfi3-288 respectively. We have analyzed and compared the total number of genes encoding for the carbohydrate binding module classes interacting with cellulose residues such as CBM-1, CBM-6 CBM-10, CBM-63. Similarly, we have analyzed and compared the CBM's interacting with xylan (CBM6, CBM13, CBM22 and CBM35) plant cell walls (CBM6, CBM13, CBM22, CBM29, CBM32, CBM35, CBM50, CBM52 and CBM61), alpha-glucans (CBM21, CBM25, CBM26 and CBM48) and chitin (CBM1, CBM12, CBM18 and CBM50) residues respectively. We have tentatively calculated the total number of genes encoding for CBM interacting with cellulose, xylan, chitin, alpha-glucans, plant cell wall and bacterial cell wall sugars. These results suggest that Neosp1 harbors higher number of cellulose, xylan, plant cell wall sugar binding CBM, whereas Orsp1 contains higher number of alpha glucan binding CBM and PirE2 contains higher number of chitin binding CBMs respectively (Figure 7.6B). Thus, anaerobic fungal resides higher number of genes encoding for the carbohydrate binding modules, dockerin proteins involved in maintenance of the structure and functioning of cellulosomes.

Recent genomic studies of the anaerobic fungi conducted by Haitjema, C.H et al (2017), have clearly proved the occurrence of 1600 dockerin domain proteins (DDP). About 20% of these dockerin domain proteins can be classified under CotH spore coat protein, these proteins were expected to be involved in binding with the plant cell wall components

however, the exact function of it is not known till today [27]. Studies have also reported that majority of the lignocellulolytic enzymes contain non-catalytic dockerin domains (NCDDs) which facilitate the assembly of multiprotein cellulosome complexes and further required for the binding of carbohydrates and degradation of plant biomass [27]. It was reported that anaerobic fungal cellulosomes exhibits 13% higher GH activity due to the presence of glycoside hydrolase classes GH-3, GH-6 and GH-45 compared to bacterial cellulosomes. Especially the supplementary GH-3 class enzyme (Beta glucosidase) activity empowers fungal cellulosomes in converting cellulose to single fermentable sugars (monosaccharides) when compared to low molecular weight oligosaccharides generating bacterial cellulosomes (eg: *Clostridium sp*) [27]. Thus, the complete genome sequencing studies of *Anaeromyces robustus* (Anasp1), *Neocallimastix californiae* (Neosp1) and *Piromyces finnis* (Pirf3) by Haitjema, C.H et al (2017) and *Orpinomyces sp Strain C1A* (Orsp1) by Youssef, N. H et al (2013), have clearly reported that genomic sequences of fungal dockerin and scaffoldin proteins are not like that of bacterial cellulosome components. Which explains that the gut residing anaerobic fungi have exclusively evolved (bacterial independent) in developing the cellulosome based degradation of plant cell wall components [26, 27].

A	Interacting residues	CBM	Anasp1	Neosp1	Orpsp1	PirE2	Pirf3
	Cellulose /Chitin	CBM1	94	145	104	102	103
	Cellulose /Xylan /Plant cell wall	CBM6	6	15	7	14	7
	Cellulose	CBM10	12	33	29	2	20
	Chitin	CBM12	1	2	2	3	6
	Xylan /Plant cell wall	CBM13	20	48	23	12	17
	Chitin	CBM18	447	521	500	640	288
	Alpha-glucans	CBM21	4	8	4	2	5
	Xylan /Plant cell wall	CBM22	1	1	0	0	1
	Alpha-glucans	CBM25	11	4	13	1	1
	Alpha-glucans	CBM26	20	10	26	15	1
	Plant cell wall	CBM29	2	15	12	7	18
	Plant cell wall	CBM32	0	5	1	2	0
	Xylan / Plant /Bacterial cell wall	CBM35	5	8	2	6	4
	Alpha-glucans	CBM48	8	17	5	5	9
	Chitin / Plant cell wall	CBM50	5	4	4	4	5
	Plant cell wall/Bacterial cell wall	CBM52	5	11	5	3	5
	Plant cell wall	CBM61	1	3	1	1	1
	Cellulose	CBM63	6	7	3	2	3
	Bacterial cell wall sugars	CBM66	2	0	3	0	0

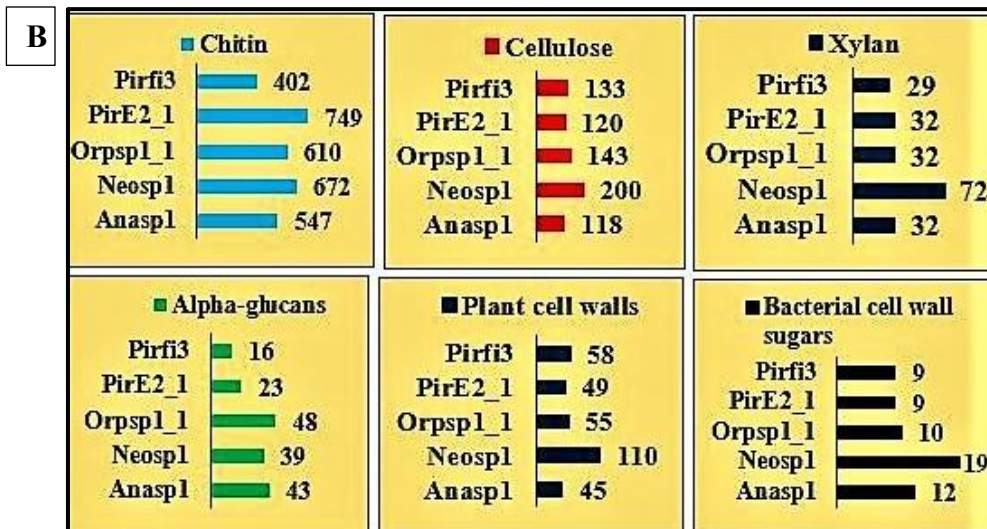


Figure 7.6: A) Carbohydrate binding modules (CBM) and the corresponding carbohydrate interacting residues, B) total tentative number of genes encoding for cellulose, xylan, alpha-glucans, plant cell walls and chitin binding CBM.

7.5. Conclusion

Genetically cattle lack the ability to encode for the lignocellulolytic enzymes, thus they are solely dependent on their rumen microbiota for the degradation of plant cell wall components. Improving the feeding efficiency and digestibility of the low-quality plant biomass components by manipulating the rumen microbiota and rumen fermentation is gaining its prominence in the recent times. Studies were being conducted to increase the number of lignocellulolytic microorganisms in rumen and their overall catalytic efficiency respectively [44]. Several microorganisms including bacteria (*Ruminococcus* genus, *Megasphaera*, *Fibrobacter*, *Streptococcus*, *Escherichia*), archaea (methanogens), fungi (*Chytridiomycetes* and *Neocallimastigomycetes*) were observed among the rumen. However, earlier studies have predicted that about 70% of the rumen microbiota is still unknown till today [45]. Anaerobic fungi potentially secrete higher number of lignocellulolytic enzymes such as cellulase (microcrystalline cellulose), xylanase, pectinase and proteases respectively. Anaerobic fungi degrade the plant biomass by breaking the fibrous plant components through penetrating its rhizoids and facilitate the access for the rumen microorganisms to the secondary cell wall components, thus playing a crucial role in degradation of the poor quality plant cell wall components.[46-48]. In this study we have analyzed and compared the genomic properties of the recently sequenced five anaerobic fungi. The genome level data and the corresponding annotations were retrieved from the JGI-

MycCosm database. Genome level annotations of the selected anaerobic fungi were compared among the *Neocallimastigomycota* division fungi. We have observed that several genes and protein domains occurred in multiple copies among the genomes of the selected anaerobic fungi. Genome of *Neocallimastix californiae* outnumbered other anaerobic fungi based on the total number of genes categorized under CAZy, SM Clusters, KOG and thus the KEGG pathways. Anaerobic fungi also encode for higher number of genes involved in carbohydrate transport and metabolism with most of the genes (carbohydrate binding modules and dockerins) involved in formation of complex multienzyme organelles employed for efficient degradation of plant cell wall carbohydrates. Using the results obtained we have tentatively calculated the degradation potentials of these fungi, which suggests that *Neosp1* possess highest cellulolytic, hemicellulolytic and pectinolytic abilities than other anaerobic fungi. We have also tentatively reported the mechanism involved during the carbohydrate metabolism by these fungal organelles respectively. In this study we have specifically emphasized on the genomic properties of the anaerobic fungi by analyzing the genome wide annotations of these fungi. Understanding the genomic complexities of anaerobic fungi will significantly enhance the gut health of the cattle (increases the total enzyme activity) and supports the biofuel and biorefining industries in making eccentric enzyme mix with efficient catalytic properties.

References:

1. Leschine SB. Cellulose degradation in anaerobic environments. *Annual Reviews in Microbiology*. 1995; 49: 399-426.
2. Brulc JM, Yeoman CJ, Wilson MK, Miller MEB, Jeraldo P, Jindou S, et al. Cellulosomics, a gene-centric approach to investigating the intraspecific diversity and adaptation of *Ruminococcus flavefaciens* within the rumen. *PLoS One*. 2011; 6: e25329.
3. Weimer PJ. Cellulose degradation by ruminal microorganisms. *Critical Reviews in Biotechnology*. 1992; 12: 189-223.
4. Wallace RJ. Gut microbiology—broad genetic diversity, yet specific metabolic niches. *Animal*. 2008; 2: 661-8.
5. Morrison M, Miron J. Adhesion to cellulose by *Ruminococcus albus*: a combination of cellulosomes and Pil-proteins? *FEMS microbiology letters*. 2000; 185: 109-15.
6. Griffith GW, Baker S, Fliegerova K, Ligginstoffer A, van der Giezen M, Voigt K, et al. Anaerobic fungi: *Neocallimastigomycota*. *IMA fungus*. 2010; 1: 181-5.
7. Gruninger RJ, Puniya AK, Callaghan TM, Edwards JE, Youssef N, Dagar SS, et al. Anaerobic fungi (phylum *Neocallimastigomycota*): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential. *FEMS microbiology ecology*. 2014; 90: 1-17.
8. Heath IB, Kaminskyj S, Bauchop T. Basal body loss during fungal zoospore encystment: evidence against centriole autonomy. *Journal of cell science*. 1986; 83: 135-40.
9. Orpin C. Studies on the rumen flagellate *Neocallimastix frontalis*. *Microbiology*. 1975; 91: 249-62.
10. Orpin C. Studies on the rumen flagellate *Sphaeromonas communis*. *Microbiology*. 1976; 94: 270-80.
11. Orpin C. The occurrence of chitin in the cell walls of the rumen organisms *Neocallimastix frontalis*, *Piromonas communis* and *Sphaeromonas communis*. *Microbiology*. 1977; 99: 215-8.

12. Orpin C. The rumen flagellate *Piromonas communis*: its life-history and invasion of plant material in the rumen. *Microbiology*. 1977; 99: 107-17.
13. Orpin CG, Greenwood Y. The role of haems and related compounds in the nutrition and zoosporogenesis of the rumen chytridiomycete *Neocallimastix frontalis* H8. *Microbiology*. 1986; 132: 2179-85.
14. Wubah D, Kim D. Isolation and characterisation of a free-living species of *Piromyces* from a pond. *Abstracts of the Mycological Society of America (Inoculum)*. 1996.
15. Trinci AP, Davies DR, Gull K, Lawrence MI, Nielsen BB, Rickers A, et al. Anaerobic fungi in herbivorous animals. *Mycological Research*. 1994; 98: 129-52.
16. Becker ER. Methods of rendering the rumen and reticulum of ruminants free from their normal infusorian fauna. *Proceedings of the National Academy of Sciences*. 1929; 15: 435-8.
17. Brookman JL, Ozkose E, Rogers S, Trinci AP, Theodorou MK. Identification of spores in the polycentric anaerobic gut fungi which enhance their ability to survive. *FEMS microbiology ecology*. 2000; 31: 261-7.
18. Lowe SE, Theodorou MK, Trinci AP. Isolation of anaerobic fungi from saliva and faeces of sheep. *Microbiology*. 1987; 133: 1829-34.
19. Milne A, Theodorou MK, Jordan MG, King-Spooner C, Trinci AP. Survival of anaerobic fungi in feces, in saliva, and in pure culture. *Experimental Mycology*. 1989; 13: 27-37.
20. Davies DR, Theodorou MK, Lawrence MI, Trinci AP. Distribution of anaerobic fungi in the digestive tract of cattle and their survival in faeces. *Microbiology*. 1993; 139: 1395-400.
21. Dey A, Sehgal JP, Puniya AK, Singh K. Influence of an anaerobic fungal culture (*Orpinomyces* sp.) administration on growth rate, ruminal fermentation and nutrient digestion in calves. *ASIAN AUSTRALASIAN JOURNAL OF ANIMAL SCIENCES*. 2004; 17: 820-4.
22. Lee S, Ha J, Cheng K-J. Influence of an anaerobic fungal culture administration on in vivo ruminal fermentation and nutrient digestion. *Animal Feed Science and Technology*. 2000; 88: 201-17.
23. Paul SS, Deb SM, Punia BS, Das KS, Singh G, Ashar MN, et al. Effect of feeding isolates of anaerobic fungus *Neocallimastix* sp. CF 17 on growth rate and fibre digestion in buffalo calves. *Archives of animal nutrition*. 2011; 65: 215-28.
24. Gao AW, Wang HR, Yang JL, Shi CX. The effects of elimination of fungi on microbial population and fiber degradation in sheep rumen. *Applied Mechanics and Materials: Trans Tech Publ*; 2013. p. 224-31.
25. Azain M, Li X, Shah A, Davies T. Separation of the effects of xylanase and b-glucanase addition on performance of broiler chicks fed barley based diets. *Int Poult Sci Forum*; 2002. p. 111.
26. Youssef NH, Couger M, Struchtemeyer CG, Ligginstoffer AS, Prade RA, Najjar FZ, et al. The genome of the anaerobic fungus *Orpinomyces* sp. strain C1A reveals the unique evolutionary history of a remarkable plant biomass degrader. *Applied and environmental microbiology*. 2013; 79: 4620-34.
27. Haitjema CH, Gilmore SP, Henske JK, Solomon KV, de Groot R, Kuo A, et al. A parts list for fungal cellulosomes revealed by comparative genomics. *Nature microbiology*. 2017; 2: 17087.
28. Fliegerova K, Hodrova B, Voigt K. Classical and molecular approaches as a powerful tool for the characterization of rumen polycentric fungi. *Folia microbiologica*. 2004; 49: 157-64.
29. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, et al. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*. 2006; 443: 818.
30. James TY, Letcher PM, Longcore JE, Mozley-Standridge SE, Porter D, Powell MJ, et al. A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). *Mycologia*. 2006; 98: 860-71.
31. Ho Y, Barr D. Classification of anaerobic gut fungi from herbivores with emphasis on rumen fungi from Malaysia. *Mycologia*. 1995: 655-77.
32. Ozkose E. Morphology and molecular ecology of anaerobic fungi. University of Wales, Aberystwyth. 2001.
33. Brookman J, Mennim G, Trinci A, Theodorou M, Tuckwell D. Identification and characterization of anaerobic gut fungi using molecular methodologies based on ribosomal ITS1 and 18S rRNA. *Microbiology*. 2000; 146: 393-403.
34. Hausner G, Inglis G, Yanke L, Kawchuk L, McAllister T. Analysis of restriction fragment length polymorphisms in the ribosomal DNA of a selection of anaerobic chytrids. *Canadian Journal of Botany*. 2000; 78: 917-27.

35. Sista Kameshwar AK, Qin W. Comparative study of genome-wide plant biomass-degrading CAZymes in white rot, brown rot and soft rot fungi. *Mycology*. 2017; 1-13.
36. Kameshwar AKS, Qin W. Metadata Analysis of *Phanerochaete chrysosporium* gene expression data identified common CAZymes encoding gene expression profiles involved in cellulose and hemicellulose degradation. *International journal of biological sciences*. 2017; 13: 85.
37. Hoffmeister D, Keller NP. Natural products of filamentous fungi: enzymes, genes, and their regulation. *Natural product reports*. 2007; 24: 393-416.
38. Bushley KE, Turgeon BG. Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC evolutionary biology*. 2010; 10: 26.
39. Finking R, Marahiel MA. Biosynthesis of nonribosomal peptides. *Annu Rev Microbiol*. 2004; 58: 453-88.
40. Kroken S, Glass NL, Taylor JW, Yoder O, Turgeon BG. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proceedings of the National Academy of Sciences*. 2003; 100: 15670-5.
41. Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic acids research*. 2013; 42: D503-D9.
42. Hamid R, Khan MA, Ahmad M, Ahmad MM, Abdin MZ, Musarrat J, et al. Chitinases: an update. *Journal of pharmacy & bioallied sciences*. 2013; 5: 21.
43. van der Giezen M, Sjollem KA, Artz RR, Alkema W, Prins RA. Hydrogenosomes in the anaerobic fungus *Neocallimastix frontalis* have a double membrane but lack an associated organelle genome. *FEBS letters*. 1997; 408: 147-50.
44. Paul SS, Deb SM, Punia BS, Singh D, Kumar R. Fibrolytic potential of anaerobic fungi (*Piromyces* sp.) isolated from wild cattle and blue bulls in pure culture and effect of their addition on in vitro fermentation of wheat straw and methane emission by rumen fluid of buffaloes. *Journal of the Science of Food and Agriculture*. 2010; 90: 1218-26.
45. Cho S-J, Cho K-M, Shin E-C, Lim W-J, Hong S-Y, Choi B-R, et al. 16S rDNA Analysis of Bacterial Diversity in Three Fractions of Cow Rumen. *Journal of Microbiology and Biotechnology*. 2006; 16: 92-101.
46. Orpin CG. Anaerobic fungi: taxonomy, biology and distribution in nature. *Anaerobic fungi: biology, ecology and function*. 1994; 12: 1-45.
47. Gordon G, Phillips MW. Degradation and utilization of cellulose and straw by three different anaerobic fungi from the ovine rumen. *Applied and Environmental Microbiology*. 1989; 55: 1703-10.
48. Grenet E, Breton A, Barry P, Fonty G. Rumen anaerobic fungi and plant substrate colonization as affected by diet composition. *Animal Feed Science and Technology*. 1989; 26: 55-70.

Chapter-8

Comparative Modeling and Molecular Docking Analysis of White, Brown and Soft Rot Fungal Laccases Using Lignin Model Compounds for Understanding the Structural and Functional Properties of Laccases

[This work has been published in "**Journal of Molecular Graphics and Modelling**"79 (2018): 15-26]

Ayyappa Kumar Sista Kameshwar, Richard Barber, Wensheng Qin*

8.1. Abstract

Extrinsic catalytic properties of laccase enable it to oxidize a wide range of aromatic (phenolic and non-phenolic) compounds which makes it commercially an important enzyme. In this study, we have extensively compared and analyzed the physico-chemical, structural and functional properties of white, brown and soft rot fungal laccases using standard protein analysis software. We have computationally predicted the three-dimensional comparative models of these laccases and later performed the molecular docking studies using the lignin model compounds. We also report a customizable rapid and reliable protein modelling and docking pipeline for developing structurally and functionally stable protein structures. We have observed that soft rot fungal laccases exhibited comparatively higher structural variation (higher random coil) when compared to brown and white rot fungal laccases. White and brown rot fungal laccase sequences exhibited higher similarity for conserved domains of *Trametes versicolor* laccase, whereas soft rot fungal laccases shared higher similarity towards conserved domains of *Melanocarpus albomyces* laccase. Results obtained from molecular docking studies showed that aminoacids PRO, PHE, LEU, LYS and GLN were commonly found to interact with the ligands. We have also observed that white and brown rot fungal laccases showed similar docking patterns (topologically monomer, dimer and trimer bind at same pocket location and tetramer binds at another pocket location) when compared to soft rot fungal laccases. Finally, the binding efficiencies of white and brown rot fungal laccases with lignin model compounds were higher compared to the soft rot fungi. These findings can be further applied in developing genetically efficient laccases which can be applied in growing biofuel and bioremediation industries.

Keywords: Multicopper oxidase, Laccases, White rot, Brown rot, Soft rot fungi, Homology Modeling, Molecular Docking

8.2. Introduction

Laccase (EC 1.10.3.2) is highly studied commercially important enzyme representing the major subgroup of multicopper oxidase (MCO) family, widely distributed among bacteria (prokaryotes), fungi and plants (eukaryotes) [1]. The function of laccases varies widely based on their host organisms, in plants it is involved in lignin biosynthesis, where as in fungi and bacteria it is involved in lignin degradation [1, 2]. It was first discovered in the sap of plants (*Rhus vernicifera*) [3] and later it was demonstrated in fungi [4]. Other enzymes belonging to multicopper oxidases (copper containing enzymes) family are ferroxidase (EC 1.16.3.1), ascorbate oxidase (EC 1.10.3.3), ceruloplasmin monooxygenases, dioxygenases and various manganese oxidases [5]. Multicopper oxidase family enzymes usually found to contain one to six copper atoms per molecule, with the aminoacids ranging between 100 to 1000 per a single peptide chain [5, 6]. Laccases are characterized by the presence of four catalytic copper atoms: the T1 copper site and the T2/T3 trinuclear copper cluster [7]. Substrate oxidation occurs at the T1 copper due to its high redox potential (up to +800 mV). The one electron substrate oxidation is coordinated with the four electron reduction of molecular oxygen at the T2/T3 cluster; oxidation of four substrates is necessary for complete reduction of molecular oxygen to water [7].

Laccases extensively uses the redox ability of copper ions for oxidation of various aromatic substrates concomitantly reducing the molecular oxygen to water [2, 8]. Laccases directly oxidize ortho, para-diphenols, aminophenols, polyphenols, polyamines, aryl diamines and also some inorganic ions [2]. The use of the laccase mediator system allows for oxidation of non-phenolic compounds and substrates too large to bind to the active site [9-12]. A mediator is a low molecular weight compound (acting as electron shuttle) with higher redox potential than the T1 copper (> 900 mV) [13]. The most common laccase mediators used are 2,2'-azino-bis (3-ethylbenzthiazoline-6-sulfonic acid) (ABTS) and triazole 1-hydroxybenzotriazole (HBT). The mediator is initially oxidized at the T1 site, generating a strong oxidizing intermediate, which then diffuses out of the active site and oxidizes the substrate [13]. In this way, the laccase mediator acts as an electron transport shuttle. Laccases typically show low substrate-specificity, and the range of substrates oxidized can vary between laccases. Oxidizing ability of laccases also depends on the nature of substrate whether it is monomeric, dimeric, or tetrameric [13]. Possible substrates of laccases include polyphenols, methoxy-substituted phenols, aromatic amines, and ascorbate [14].

The comparative modeling of fungal and bacterial laccases was reported in the past, however studies on fungal laccases typically focused on white rot basidiomycetes due to the extrinsic lignolytic abilities. Rivera-Hoyos et al.(2015), reported the three dimensional (3D) homology models of white rot fungal (*Ganoderma lucidum* and *Pleurotus ostreatus*) laccase proteins, which revealed the laccase interactions with ABTS [15]. The 3D homology models of white rot fungi *Pycnoporus cinnabarinus* [16], *Lentinula edodes* [17], were reported earlier. Tamboli et al (2015) has compared physio-chemical properties of bacterial and fungal (*Cryphonectria parasitica*, *Ganoderma lucidum*, *Phomopsis liquidambaris*, *Pycnoporus coccineus* and *Trametes sanguine*) laccases and generated the 3D comparative models of bacterial and fungal laccase proteins which can be used for molecular docking studies [18]. Molecular docking studies with fungal laccases were performed using various chemical substrates such as ABTS [15] and also with lignin model compounds such as sinapyl alcohol, dimer, trimer and tetramer [19] were reported earlier. However, studies comparing the structural and functional properties of white, brown and soft-rot fungal laccases were not been reported till today. As these fungi exhibit differential wood decaying properties white rot (can efficiently degrade lignin, cellulose and hemicellulose), brown rot fungi (efficient cellulose, hemicellulose degrading with lignin modifying) and soft rot fungi (exhibits partial decaying abilities). It would be interesting to understand the structural and functional differences among the laccases of these fungi.

In our present study, we have reported the three-dimensional homology models of the selected white, brown and soft fungal laccase protein sequences retrieved from public repositories and extensively discussed about their structural and functional properties using standard tools. Using a set of lignin model compounds (monomers, dimer, trimer and tetramer) we have performed the molecular docking experiments. Results obtained in our study demarcates the structural and functional properties of white, brown and soft rot fungi and highlights the significant aminoacids which are involved in its catalysis. These results can be further applied for designing and developing recombinantly efficient laccases having wide range applications in clinical, chemical, environmental and industrial sectors.

8.3. Materials and Methods

8.3.1. Selection and Retrieval of Laccase Protein Sequences

Laccase protein sequences of six different fungi viz., *Phlebia brevispora* HHB-7030 SS6 v1.0 [20], *Dichomitus squalens* CBS463.89 (White rot), *Fomitopsis pinicola* FP-58527 SS1 [21],

Wolfiporia cocos MD-104 SS10 [21] (Brown rot) and *Chaetomium globosum* v1.0 [22], *Cadophora* sp. DSE1049 (Soft rot), were retrieved from JGI (Joint Genome Institute) MycoCosm database. The *D. squalens* CBS463.89 and *Cadophora* sp. DSE1049 laccase protein sequences (Dsqual_59186 and Cadophora_560981) were produced by the “US Department of Energy Joint Genome Institute <http://www.jgi.doe.gov/> in collaboration with the user community”. We have used CAZy (Carbohydrate active enzymes), KOG (Eukaryotic orthologous groups) and GO (Gene Ontology) tools of JGI MycoCosm database during the retrieval of laccase protein sequences. Initially, we have retrieved a total of 56 laccase protein sequences (*P. brevispora* (5), *D. squalens* (12), *F. pinicola* (6), *W. cocos* (4), *C. globosum* (6) and *Cadophora* sp (22)) respectively, from JGI MycoCosm database. All the retrieved laccase protein sequences from each organism was queried through BLAST against protein data bank (PDB) database using PSI-BLAST algorithm a variation of BLAST (sensitive to low-similarity, provides biologically relevant sequences and three times faster than regular BLAST) [23]. Laccase protein sequences showing highest sequence similarity and query coverage was designated as the template for the comparative modeling studies.

8.3.2. Phylogenetic Analysis

All the retrieved laccase protein sequences of each organism were aligned using ClustalW algorithm (fast, accurate, and robust method, which uses a residue comparison matrix and position specific gap penalties to align sequences) of MEGA v7 software [24]. The ClustalW aligned sequences were considered for the construction of phylogenetic trees using Neighbour Joining method and Bootstrap resampling of 1000 replicates parameters were used for the estimation of phylogenetic tree topologies [25]. The phylogenetic trees were constructed for both intra and inter organism level to determine the laccase target sequences which are closely related to the template during the evolution.

8.3.3. Physico-Chemical Properties of Selected Laccases

Physico-chemical properties of above selected laccase protein sequences were determined using the ExPASy ProtParam tool [26]. Our analysis included the parameters such as aminoacid composition, number of positively (+R) and negatively (-R) charged aminoacid residues, predicted molecular weight, theoretical isoelectric point (pI), extinction coefficient (EC) [27], instability index (Ii) [28], aliphatic index (Ai) [29] and GRAVY (grand average hydropathicity) [30].

8.3.4. Structural and Functional Properties of Laccases

The above selected laccase protein sequences were studied for their structural and functional properties for which we have used SOPMA (Self-optimized prediction method with alignment) tool for determining secondary structure elements [31]. We have used Motif Scan web server to identify the well-known motif sequences using the motif sources such as PeroxiBase, HAMAP, PROSITE patterns and profiles, More profiles, Pfam HMM (both local and global) profiles [32]. To understand the cellular localization of selected laccases, the protein sequences were subjected CELLO v2.5 web server [33]. We have used EDBCP (Ensemble-based Disulfide Bonding Connectivity Pattern) for understanding the presence of cysteine residues and to predict the most possible disulfide (S-S) bonds [34]. To predict the location and presence of signal peptide cleavage sites the protein sequences of laccases were analyzed using SignalP v4.1 web server [35]. And to predict the presence of transmembrane helices we have analyzed the selected protein sequences using TMHMM v2.0 web server (<http://www.cbs.dtu.dk/services/TMHMM/>). Acetylation of the selected fungal laccase proteins are assessed using the NetAcet v1.0 web server [36].

8.3.5. Initial protein model generation and refinement

Laccase protein sequences exhibiting highest sequence similarity and highest query coverage (obtained from the BLAST analysis), were further considered as input to SWISS-MODEL protein prediction web server (generates tertiary protein structure from a queried protein sequence based on the SWISS-MODEL template library), for the generation of initial unrefined 3D modelled structure of the target protein [37]. The 3D structures of modelled laccase proteins were further refined using GalaxyRefine (method improves both local and global qualities of template-based predicted protein structures) [38] and KoBaMIN (knowledge based minimization method, which minimizes the potential of mean force derived from the experimental structures of PDB) [39] web servers. GalaxyRefine server optimizes the side chain conformations and performs the energy minimization on each conformation and subsequently relaxes the overall protein structure through molecular dynamic simulation methods. Whereas KoBaMIN refined structures are stereochemically optimized with MESHI software, KoBaMIN energy function indirectly includes the effects of both solvent interactions and the crystal environments. The refined laccase protein structures were validated to assess the stereochemical quality, using a set of software such as PROCHECK [40], RAMPAGE [41], ERRAT [42] and PROQ [43].

8.3.6 Preparation of Ligands (Lignin model compounds)

The NMR based structural studies of lignin and plant cell wall compounds explained by Ralph, S. A et al (2004), were derived for the present study [44]. The lignin model compounds considered for the present study were monomers (sinapyl alcohol, coniferyl alcohol and *p*-coumaryl alcohol), dimer (guaiacyl 4-O-5 guaiacyl), trimer (syringyl β -O-4 syringyl β -O-4 sinapyl alcohol) and tetramer (guaiacyl β -O-4 syringyl β - β syringyl β -O-4 guaiacyl) were sketched using ChemDraw Ultra v7.0. and later internally transferred to Chem3D Pro v7.0 software and further saved the structure in protein data bank format. The ligands were subjected for energy minimization before using it for the protein docking simulation experiments using AutoDock software [45].

8.3.7 Protein Docking of Refined Models

The refined and validated 3D modelled laccase proteins were used for further protein docking studies. We have used AutoDock Tools v1.5.6 and AutoDock Vina v1.1.2 [46] for the simulated protein docking experiments. The refined protein model in PDB file format were initially opened in AutoDock Tools and following functions were performed a) add all hydrogens b) merge non-polar hydrogens c) compute Gasteiger charges d) finally, save the protein model in PDBQT format. The ligand (lignin model compounds) are loaded into AutoDock Tools and following functions were performed a) detect root using the option torsion tree b) the number of torsions were set to maximum c) then finally save the ligand in PDBQT format. Later we have prepared the AutoDock Vina configuration files for all the laccase modelled structures and ligands, the above prepared files were used for performing the protein docking analysis. Ligand docking was performed using AutoDock Vina which is installed on an instance of the Galaxy platform based on In-house High-Performance Computing Cluster (LUHPCC). We have performed blind docking with all the lignin ligand models with the exhaustiveness set to 32. The best-fit ligand conformations were selected based on their minimum binding energies. Aminoacid residues which established a contact with ligand and the residues which are involved in hydrogen bonding with ligands were recorded using AutoDock Tools using the results obtained from AutoDock Vina. The validated 3D laccase modelled structures were compared using the root mean square deviation (RMSD) using SWISS-PDB viewer v4.1 [47]. We have used Edu PyMOL v1.7.4 (<https://pymol.org/educational>) for visualizing the interactions of the ligand and modelled protein structure and for developing the respective docked images.

8.4. Results:

8.4.1. Sequence Retrieval, Analysis and Physico-Chemical Properties

We have retrieved genome wide laccase encoding protein sequences from *P. brevispora* and *D. squalens* (white rot fungi), *F. pinicola* and *W. cocos* (brown rot fungi), and *C. globosum* and *Cadophora* sp (soft rot fungi). Total of 56 laccase protein sequences from *P. brevispora* (5), *D. squalens* (12), *F. pinicola* (6), *W. cocos* (4), *C. globosum* (6) and *Cadophora* sp (22), were retrieved from JGI MycoCosm database. The laccase protein sequences exhibiting highest sequence similarities with the known laccase structures from PDB database upon BLASTP search, were considered for the present study. The following protein sequences of fungi exhibiting higher sequence identities were used for the modeling studies: Phlbr1-25201, Dsqal-59186, Fompi3-45001, Wcocos-139080, Chagl1-12114 and Cadophora-560981.

The selected laccase protein sequences were found to contain about 520 to 619 aminoacid residues with an exception of 479 (*W. cocos*), with a theoretical molecular weight and pI ranging between 52277.43 to 68515.16 Daltons and 4.46 to 6.88 respectively (Table 8.1). All the above fungal laccase protein sequences were found to be stable with an instability index value ranging between 28.04 to 38.02, with an exception of *W. cocos* laccase protein instability index of (40.44) (Table 8.1). Aliphatic index values of the laccase proteins were under the range of 74.44 to 90.4. The grand average of hydropathicity (GRAVY) index values of the laccase proteins were mostly negative indicating the hydrophilic nature of these proteins, with an exception of *P. brevispora* laccase protein which gave a positive value indicating its hydrophobic nature and the hydropathicity plots generated using DiscoveryStudio ® 2016 Client® were also found to be in accordance with the GRAVY results (Table 8.1). The selected laccase protein sequences exhibited higher evolutionary similarities with the template protein sequences. All the above considered laccase protein sequences were found to contain sequences coding for cupredoxin superfamily conserved domains specifically trinuclear Cu binding site (CuRO_1_Tv-LCC_like), domain 3 interface (CuRO_2_Tv-LCC_like) and type-1 copper binding site (CuRO_3_Tv-LCC_like) with an exception of *W. cocos* laccase only CuRO_3_Tv-LCC_like domain sequences. Laccase protein sequences of soft rot fungi contain cupredoxin superfamily conserved domains matching to *Melanocarpus albomyces*, whereas white and brown rot fungi possessed cupredoxin superfamily conserved domains matching to *Trametes versicolor* (Figure 8.1).

Table 8.1: Lists the physico-chemical properties of laccase protein sequences calculated from ExPASy ProtParam and EDBCP tools:

Organism (protein-ID)	Length	M.W*	pI*	EC	-R	+R	Ii	Ai	GRAV _y	#.Cys	Predicted S-S bonds
<i>Phlebia brevispora</i> (25201)	520	55977.21	5.20	69705	45	23	31.55	90.04	0.082	7	106-509, 138-227
<i>Dichomitus squalens</i> (59186)	520	56477.01	4.63	71195	48	22	28.04	84.79	-0.044	6	106-509, 138-226
<i>Fomitopsis pinicola</i> (45001)	539	58990.52	4.75	67185	55	26	34.35	88.85	-0.055	7	109-513, 141-228, 349-352
<i>Wolfiporia cocos</i> (139080)	479	52277.43	4.46	68550	54	18	40.44	83.47	-0.135	5	64-468, 96-183
<i>Chaetomium globosum</i> (12114)	619	68515.16	6.11	120820	55	45	35.90	74.44	-0.295	8	51-344, 112-378, 161-586
<i>Cadophora</i> sp. DSE1049 (560981)	582	63262.99	6.88	127810	34	33	38.02	75.74	-0.231	8	26-34, 135-563, 327-361

(Note: M.W* = Molecular weight in Daltons, pI* = Theoretical pI values, EC = Extinction coefficient ($M^{-1} \text{ cm}^{-1}$), -R = number of negatively charged amino acids, +R = number of positively charged amino acids, Ii = Instability index, Ai = Aliphatic index, GRAV_y= grand average of hydropathy, #. Of Cys= Number of cysteine residues)

The selected laccase protein sequences were also subjected to series of analysis using SignalP (for the detection of signal peptides cleavage sites), TMHMM (for detecting transmembrane helices in proteins) and NetAcet (for predicting N-acetyltransferase A substrates) web servers. According to Nakai (2000), the signal peptide ranges between the 15 to 40 amino acid residues, required for the protein secretion and eventually these sites are cleaved from the mature protein [48]. We have observed that all the laccases except *W. cocos* (13980) possessed a signal peptide, which supports that these laccases are secretory proteins. The amino acid sequence involved in formation of transmembrane helices were only found to be present in *D. squalens* and these results explain that majority of amino acid residues occur in the outside region. Results obtained from CELLO v2.5 subcellular location predictor showed that the laccase proteins of *P. brevispora* (3.981), *D. squalens* (4.644), *F. pinicola* (3.688), *W. cocos* (3.871), *C. globosum* (3.995) and *Cadophora* (3.782) were located in the extracellular regions of fungi. All the selected laccase protein sequences showed N-Myristoylation, N-glycosylation, Amidation (except *Cadophora* sp), Casein kinase-II phosphorylation site, protein kinase c phosphorylation site and multi copper oxidase type-

1&2 (Cu_oxidase_1, 2 and 3) motif sites (Table 8.2). Analysis from NetAcet v1.0 server has showed that laccase proteins of *P. brevispora* and *Cadophora* sp possess the substrates for N-acetyltransferases with values of 0.511 and 0.522 respectively. Whereas laccase protein sequences of *D. squalens* (0.491), *F. pinicola* (0.478), *W. cocos* (0.483) and *C. globosum* (no Ala, Gly, Ser or Thr at positions 1-3) do not show any possible substrates for N-acetyltransferases. We have also analyzed the number of cysteine residues present and reported the number of predicted disulfide bridges in the selected laccase protein sequences using EDBCP web server. We have observed that laccase protein sequences of *P. brevispora*, *D. squalens* and *W. cocos* contain two disulfide bridges, laccase protein sequences of *F. pinicola*, *C. globosum* and *Cadophora* sp contain three disulfide bridges. The secondary structure elements of the selected laccase proteins were calculated using the SOPMA web server. These results show that higher percentage of aminoacids were found to be involved in formation of random coils, the alpha helical content of soft rot fungi is less when compared to brown and white rot fungi (Table 8.3).

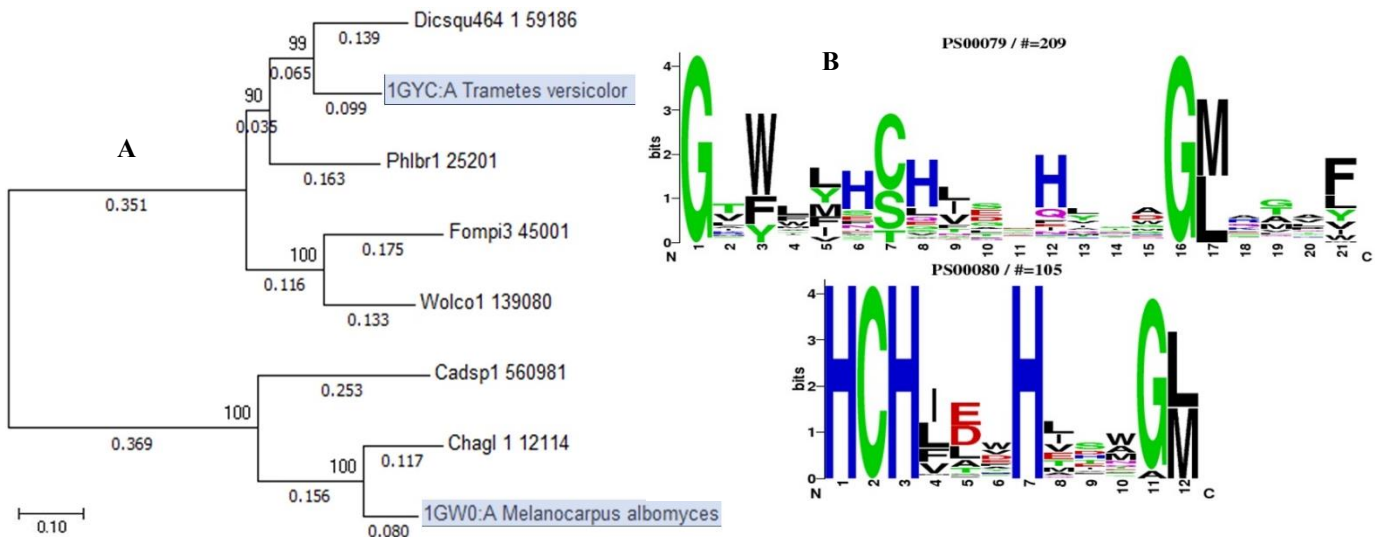


Figure 8.1: Shows the analysis of fungal laccase protein sequences A) phylogenetic analysis of laccase protein sequence (*P. brevispora*, *D. squalens*, *F. pinicola*, *W. cocos*, *C. globosum*) and experimentally determined laccases *T. versicolor* (1GYC) and *M. albomyces* (1GW0), B) sequence logos of MCO signature 1 (PS00079) and 2 (PS00080) patterns.

Table 8.2: Computationally predicted motifs in *P. brevispora*, *D. squalens*, *F. pinicola*, *W. cocos*, *C. globosum* and *Cadophora* laccase protein sequences obtained from the MOTIF SCAN server:

Organism, Protein-ID	Motif Description / Aminoacid residues (# of sites)
<i>P. brevispora</i> (25201)	Multi copper oxidase 1: 125-145 (1) Multi copper oxidase (Cu_oxidase): 163-307 (1) Multi copper oxidase (Cu_oxidase_2): 369-494 (1) Multi copper oxidase (Cu_oxidase_3): 30-152 (1)
<i>D. squalens</i> (59186)	Multi copper oxidase 1: 125-145 (1) Multi copper oxidase (Cu_oxidase): 163-305 (1) Multi copper oxidase (Cu_oxidase_2): 365-494 (1) Multi copper oxidase (Cu_oxidase_3): 30-152 (1)
<i>F. pinicola</i> (45001)	Multi copper oxidase 1: 128-148 (1) Multi copper oxidase (Cu_oxidase): 166-310 (1) Multi copper oxidase (Cu_oxidase_2): 372-497 (1) Multi copper oxidase (Cu_oxidase_3): 33-155 (1)
<i>W. cocos</i> (139080)	Multi copper oxidase 1: 83-103 (1) Multi copper oxidase (Cu_oxidase): 121-267 (1) Multi copper oxidase (Cu_oxidase_2): 327-453 (1) Multi copper oxidase (Cu_oxidase_3): 1-110, 10-110 (1)
<i>C. globosum</i> (12114)	Multi copper oxidase 1, 2: 543-563, 548-559 Multi copper oxidase (Cu_oxidase): 212-361 (1) Multi copper oxidase (Cu_oxidase_2): 427-569 (1) Multi copper oxidase (Cu_oxidase_3): 87-206 (1)
<i>Cadophora</i> sp (560981)	Multi copper oxidase 1, 2: 523-543, 528-539 (2) Multi copper oxidase (Cu_oxidase): 186-344 (1) Multi copper oxidase (Cu_oxidase_2): 410-549 (1) Multi copper oxidase (Cu_oxidase_3): 61-180 (1)

Table 8.3: Computationally predicted secondary structure elements of laccase protein sequences calculated using SOPMA web server:

Protein-ID	Alpha helix (%)	Extended Strand (%)	Beta turn (%)	Random coil (%)
Phlbr1_25201	15.96 (83)	30.77 (160)	10.96 (57)	42.31 (220)
Dsqual1_59186	15.38 (80)	31.92 (166)	11.73 (61)	40.96 (213)
Fompi3_45001	16.70 (90)	29.31 (158)	9.09 (49)	44.09 (242)
Wcocos_139080	10.23 (49)	31.94 (153)	10.65 (51)	47.18 (226)
Chagl1_12114	12.12 (75)	33.12 (205)	10.82 (67)	43.94 (272)
Cadoph_560981	9.79 (57)	28.01 (163)	9.28 (54)	52.92 (308)

8.4.2. Homology Modeling and Model validation: The selected laccase protein sequences were used to develop the corresponding 3D protein structures using SWISS-MODEL server. In SWISS-MODEL server the best suitable template protein structures for the target protein sequence were identified using BLAST and HHblits. The server will also evaluate the global and local qualities of the modelled protein structure using QMEAN (a complex function used for the estimation of both local and global qualities calculation which includes four parameters all atom, C β , solvation and torsion) and GMQE (global model quality estimation, which retrieves information from template and target alignment). All the selected laccase protein sequences were found by BLAST search with >60% of sequence identity. The 3D modelled protein structures of fungal laccase proteins were found to be statistically acceptable with higher QMEAN and GMQE scores, *P. brevispora* (1.08), *D. squalens* (0.55) and *C. globosum* (0.41), *F. pinicola* (-1.6), *W. cocos* (-0.19) and *Cadophora* sp (-1.66). Both the QMEAN and GMQE scores were expressed as a number in a range between 0 to 1, where the higher number represents higher quality (Figure 8.2) (Table 8.4). These initial protein models were refined using GalaxyRefine and KoBaMIN web servers which performs the side chain refinement, energy minimization and relaxes the overall modelled structure. The above obtained refined laccase protein structures were validated using PROCHECK software and Ramachandran plots. The results obtained from Ramachandran plots were found to be statistically acceptable, all the laccase refined models attained >90% residues in most favored regions (except for *F. pinicola*-89.3% and *Cadophora* sp 89.4%) (Table 8.4). The refined protein structures were also validated using PROQ, ERRAT and RAMPAGE. PROQ web tool provides mainly LG (a -log P-value, models are good if the score is >3 and very good if the score is >5) and Max sub scores (ranges between 0 to 1, 1 being highly reliable and 0 is insignificant). All the refined laccase structures achieved significant scores with an LG score

of >4 and Max sub score of > 0.3 (except *W. cocos*). Similarly, overall quality factor values obtained from ERRAT server for the refined structures were >84 (except *F. pinicola*). Generally, the predicted protein models attaining overall quality factor >50% infers that the homology models were stable and reliable. Results from RAMPAGE server showed that all the refined laccase structures exhibited >93% of residues in most favored regions. Results obtained from QMEAN4, Ramachandran plots, PROQ, ERRAT and RAMPAGE web servers convey that predicted fungal laccases were of good quality (Table 8.5) (Figure 8.2 & 8.3). The structural variation observed in the Ramachandran plots generated from PROCHECK and RAMPAGE web servers can be explained by the advanced refined and more reliable protein structure validation methods implemented by the RAMPAGE web server [41].

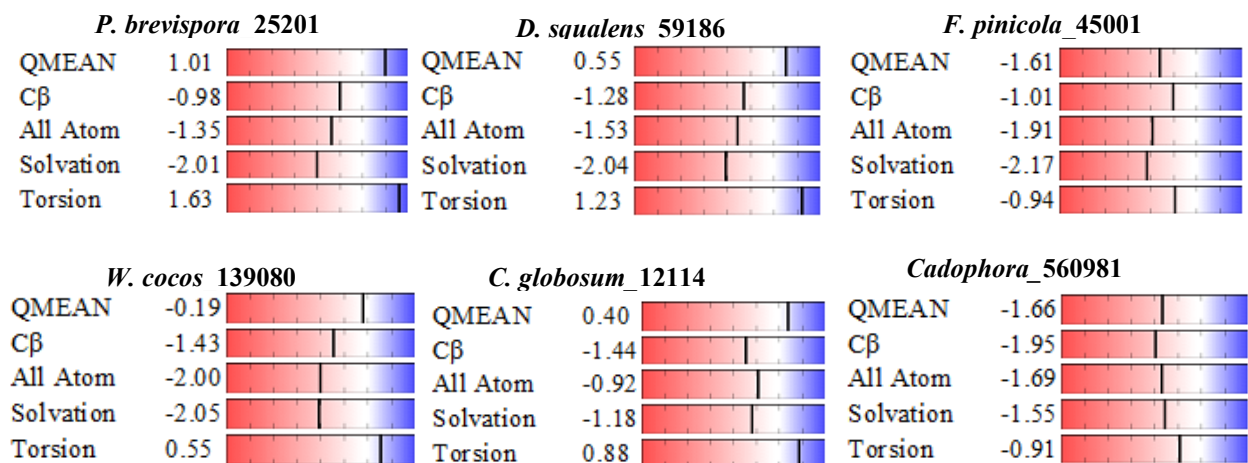


Figure 8.2: QMEAN scores for the 3D modelled laccase structures obtained from SWISS-MODEL server for fungal protein sequences *P. brevispora*, *D. squalens*, *F. pinicola*, *W. cocos*, *C. globosum* and *Cadophora* sp

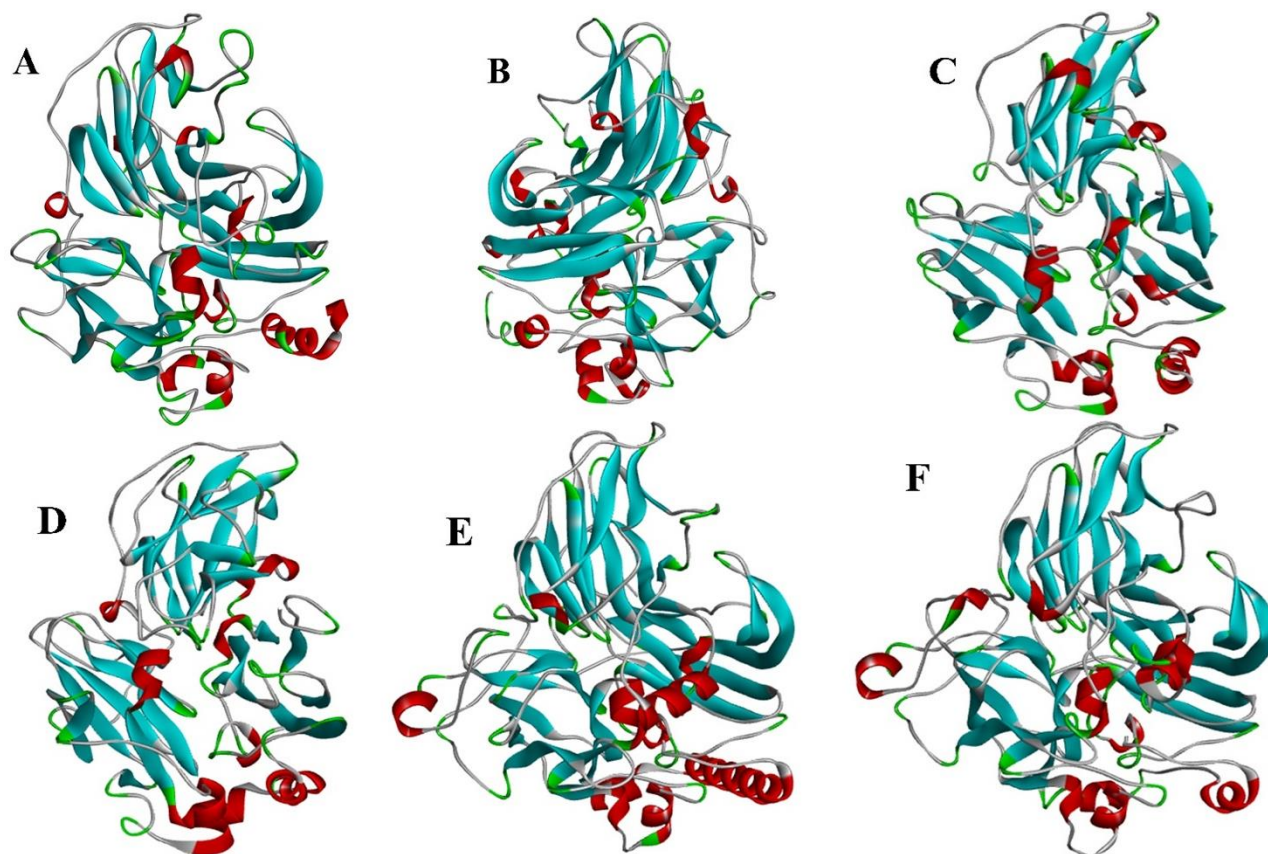


Figure 8.3: Homology models of laccase protein sequences A) *Phlebia brevispora* B) *Fomitopsis pinicola* C) *Dichomitus squalens* D) *Wolfiporia cocos* E) *Chaetomium globosum* F) *Cadophora* DSE1049 v1.0.

Table 8.4: Comparison of results obtained from the Swiss model server (BLAST and HHblits) and BLASTP (NCBI PSI-BLAST against PDB server):

Organism, Protein-ID	Template Swiss PDB-server	Sequence Identity	GQME	QMEAN	Template BLASTP	Sequence Identity
Phlbr1 (25201)	5a7e.1A	70.24	0.86	1.08	3KW7	71
Dsquall1 (59186)	1kya.1A	80.92%	0.97	0.55	2QT6	82
Fompi3 (45001)	5ehf.1A	60.57	0.78	-1.6	3KW7	63
Wcocos (139080)	5anh.1A	63.83%	0.84	-0.19	3KW7	66
Chagl1 (12114)	1gw0.1A	80.29	0.84	0.41	1GW0	80
Cadophora (560981)	3pps.1A	61.75%	0.79	-1.66	3PPS	59

Table 8.5: Ramachandran plot scores of laccase modelled protein structures obtained after customized refining pipeline using different model refining softwares and results from RAMPAGE, PROQ, ERRAT and RMSD (initial and final protein structures) webserver:

Method Used	Phlbr1 (25201)	Dsqual (59186)	Fompi3 (45001)	Wcocos (139080)	Chag11 (12114)	Cadophora (560981)
SWISS model (Initial quality)	87.0	87.1	82.7	82.4	85.9	81.6
Galaxy Refine (after SWISS)	90.5	90.4	88.3	90.3	91.3	89.4
KoBaMIN (after GR)	90.7	92.3	89.3	89.5	90.5	88.3
Only KoBaMIN	91.7	89.4	87.8	90.0	92.2	87.9
RAMPAGE (RFR, RAR, ROR)	97.6, 2, 0.4	97.4, 1.8, 0.8	94.8, 2.6, 2.6	93.9, 4, 2.1	98.4, 1.4, 0.2	96.2, 2.3, 1.4
RMSD (Initial vs Final Model)	0.36 Å°	0.40 Å°	0.34 Å°	0.33 Å°	0.35 Å°	0.40 Å°
RMSD (Final vs Template)	0.35 Å°	0.52 Å°	0.37 Å°	0.34 Å°	0.41 Å°	0.51 Å°
PROQ (LG and Max Sub)	4.631, 0.348	5.002, 0.359	5.228, 0.389	4.486, 0.295	4.562, 0.369	5.050, 0.325
ERRAT Server	87.780	86.735	75.102	87.152	89.091	84.432

Note: (The pipeline used for refining the protein models was **SWISS model** → **Galaxy Refine** → **KoBaMIN**; **only KoBaMIN**, RAMPAGE results were RFR= residues in favored regions, RAR= residues in allowed regions and ROR = residues in outlier regions; The highlighted regions represent the final refined laccase modelled structures used for the molecular docking analysis and their respective Ramachandran plot scores)

8.4.3. Molecular Docking of Modelled Laccases with Lignin Model Compounds

Molecular docking experiments of 3D modelled fungal laccases (white rot, brown rot and soft rot) with lignin model compounds that is monomers (sinapyl, coniferyl and *p*-coumaryl alcohol), dimer (guaiacyl 4-O-5 guaiacyl), trimer (syringyl β-O-4 syringyl β-O-4 sinapyl alcohol) and tetramer (guaiacyl β-O-4 syringyl β-β syringyl β-O-4 guaiacyl) was performed using AutoDock Tools and Vina software. Results obtained from this study were reported in Table 8.6, 8.7 and 8.8. Based on the results obtained we have observed a sharp increase in

binding efficiencies from monomers to tetramers with increase in size of the ligand with all the fungal (white, brown and soft rot) laccases (Table 8.6). The binding efficiencies of white rot fungal laccases (*P. brevispora* and *D. squalens*) has exhibited a clear ascending order of minimum binding efficiencies from monomers to tetramers (-6.0 to -8.2). Brown rot fungal laccases (*F. pinicola* and *W. cocos*) has also shown a clear increase in minimum binding efficiencies from monomers to tetramers however, unlike white rot laccases brown rot laccases exhibited selective specificity among the ligands. When compared to white rot and brown rot fungal laccases, soft rot fungal laccases (*C. globosum* and *Cadophora* sp) exhibited lesser minimum binding efficiencies and were variably specific among the ligands. Based on these results we can conclude that fungal laccases exhibit higher minimum binding efficiencies for lignin model compounds and out of which trimers and tetramers bind more efficiently to the fungal laccases. The higher binding efficiencies exhibited by white rot fungal laccases towards lignin model compounds are evident to their extrinsic lignin degrading abilities.

Table 8.6: Lists the final predicted minimum binding energy scores (kcal/mol) of predicted laccase models with lignin model compounds obtained from AutoDock Vina software:

Ligand (Modelled Protein-ID)	Phlbr1 (25201)	Dsqual (59186)	Fompi3 (45001)	Wcocos (139080)	Chagl1 (12114)	Cadophora (560981)
Sinapyl alcohol	-6.0	-5.4	-5.7	-5.8	-6.0	-5.6
Coniferyl alcohol	-6.0	-6.2	-7.0	-6.0	-5.6	-5.8
p-coumaryl alcohol	-6.1	-6.3	-6.7	-5.8	-5.5	-6.8
Guaiacyl 4-O-5 guaiacyl	-6.9	-6.7	-6.2	-7.1	-6.5	-6.8
Syringyl β-O-4 syringyl β-O-4 sinapyl alcohol	-7.8	-7.1	-7.0	-8.2	-6.8	-6.9
Guaiacyl β-O-4 syringyl β-β syringyl β-O-4 guaiacyl	-7.7	-7.4	-7.2	-7.0	-7.2	-6.1

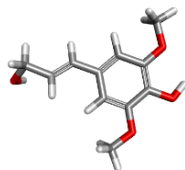
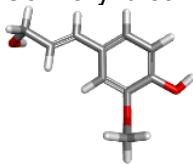
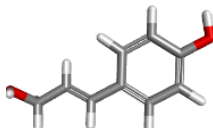
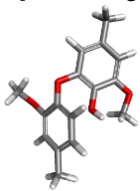
The lignin model compounds sinapyl (SA), coniferyl (CA), *p*-coumaryl alcohol (CoA), dimer, trimer and tetramer were found to interact with a total of 7, 7, 3, 6, 10,13 aminoacid residues of *P. brevispora* respectively. Out of which following aminoacid ligand interactions were found to be common among SER134, GLU481, PHE471, PHE369 (SA: CA), PRO371, HIS132 (SA: CA: CoA), ALA101 (SA: CoA) and PRO371, PHE102 (dimer and trimer). We have observed that hydrogen bond formations were between SA-ALA101, CA-HIS132 and GLU481,

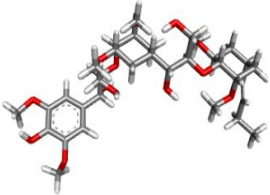
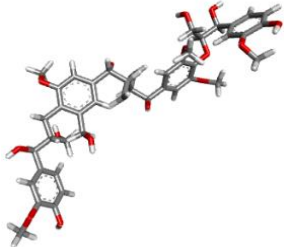
CoA-HIS132 and tetramer GLU325 residues of *P. brevispora* laccase. Similarly, lignin model compounds were found to interact with 4, 6, 7, 6, 10 and 11 aminoacid residues of *D. squalens* laccase protein. Following aminoacid ligand interactions were observed in common LEU480 (SA-Tetramer), ALA101 (CoA-Tetramer), PHE102, TYR512 (SA-Dimer-Tetramer), PHE365 (CA-CoA-Dimer), ASP481, HIS132 (CA-CoA-Tetramer), SER134 (CA-Dimer-Tetramer), LEU133 (SA-CA-Dimer-Tetramer) and PRO367 (CA-CoA-Dimer-Tetramer). And hydrogen bond formations were observed between SA-LEU480, CA-SER134 and PHE365, CoA- HIS132 and ASP481, Trimer- LEU115 and Tetramer-ASP481.

Brown rot fungal laccases considered for this study *F. pinicola* and *W. cocos* were found to interact with lignin model compounds (SA, CA, CoA, dimer, trimer and tetramer) through 4, 7, 6, 5, 12, 12 and 5, 6, 7, 8, 16, 11 aminoacids residues respectively. *F. pinicola* laccase protein and ligands were involved in hydrogen bond formations between SA-ARG95, CA-ALA187, CoA-ASP229, dimer-VAL169, trimer-SER374 and tetramer-ASP152, with commonly interacting residues ASP229, TYR176 and CYS141 between CA and CoA, ILE183 between CA-tetramer respectively. In *W. cocos* laccase protein following aminoacids were found in common GLU204 (CoA-trimer), ASN205(dimer-trimer), GLY203 (SA, CA, CoA, dimer), PHE85 (SA, CA, dimer, trimer) and GLN81, SER384, ALA82 (SA, CA, CoA, dimer and tetramer) with formation of hydrogen bonds between SA-SER384, CA-ALA82, CoA-ALA82, dimer-GLN81, trimer-ALA82, ASN424 and tetramer-GLN214 respectively (Table 8.7).

Soft rot fungal laccases considered for this study *C. globosum* and *Cadophora* sp were found to interact with lignin ligand model compounds SA, CA, CoA, dimer, trimer and tetramer by 7, 7, 8, 7, 9, 12 and 5,5,8,7,9, 10 aminoacid residues respectively. In *C. globosum* laccase following aminoacid residues were found to interact commonly GLY192, TYR222, ASP220, TYR221 with SA and CoA, ARG400, GLY402, ILE451, VAL452, GLN453, TYR462 with trimer and tetramer respectively. We have observed hydrogen bond formation between SA-TYR222, CoA-THR280, dimer-ARG513, trimer-ARG400, ILE451, GLN453 and tetramer-THRE447, ILE451, GLN459, TYR462 residues respectively. In *Cadophora* sp laccase protein following residues SER536, VAL400 and GLN537 found to be interact commonly with SA and CA, whereas THR144, ARG148, LYS143, ASN123, SER146, ASN126, ILE125 with trimer and tetramer respectively. And hydrogen bond formation was observed between CoA- ALA532, dimer-ALA336, trimer-THR144, ARG148 and tetramer-ASN562 residues (Table 8.7).

Table 8.7: Lists the aminoacid residues of modelled fungal laccases in contact with the lignin-based model compounds Monomers (Sinapyl alcohol, Coniferyl alcohol and *p*-coumaryl alcohol), Dimer, Trimer and Tetramer (Note: aminoacids represented in bold are involved in hydrogen bonding between protein and ligand):

Ligand	Phlbr1 (25201)	Dsqual (59186)	Fompi3 (45001)	Wcocos (139080)	Chagl1 (12114)	Cadophora (560981)
Sinapyl alcohol 	PRO371 HIS132 SER134 GLU481 PHE471 GLY483 PHE369	PHE102 LEU133 TYR512 LEU480	ARG95 PRO124 ASN125 GLN94	GLN81 PHE85 SER384 GLY203 ALA82	GLY192 TYR222 ASP220 VAL196 MET219 TYR221 THR278	ASP579 ALA397 SER536 VAL400 GLN537
Coniferyl alcohol 	SER134 HIS132 PRO371 GLU481 ALA101 PHE369 PHE471	SER134 LEU133 PRO367 ASP481 HIS132 PHE365	PHE481, ILE183 ASP229, ALA187 TYR176, ARG46 CYS141	ALA82 GLN81 PHE48 GLY203 SER384 PHE85	PHE598 VAL601 THR134 ASN135 TRP430 ASN596 ASP227	LYS401 VAL400 ARG410 GLN537 SER536
<i>p</i> -Coumaryl alcohol 	HIS132 PRO371 ALA101	PRO367 GLY483 PHE365 ALA101 ALA482 ASP481 HIS132	ASP229 TRP175 CYS141 CYS228 ASP174 TYR176	ASP202 ALA82 GLY203 GLU204 GLN81 ASP405 SER384	TYR221 LYS101 TRP553 TYR222 ASP220 LEU229 GLY192 THR280	VAL264 LEU582 TRP156 TYR165 VAL535 ALA532 SER170 PRO262
Guaiacyl 4-O-5 guaiacyl 	PRO371 TYR512 LEU133 SER134 PHE102 PHE369	LEU133 PHE365 PHE102 PRO367 SER134 TYR512	LEU59 GLU166 ILE211 VAL169 ASN195	SER384, LYS19 ALA82, PHE85 ASN205, GLY203 PHE48, GLN81	GLY490 VAL520 ARG513 ARG491 PRO519 PRO382 VAL488	TYR244 PRO237 ASN332 TRP317 ASN348 ALA336 VAL235

<p>Syringyl β-O-4 syringyl</p>  <p>β-O-4 sinapyl alcohol</p>	<p>GLU481, PRO371 PHE102, LEU480 LYS178, ALA101 PHE471, PRO182 VAL370, LYS361</p>	<p>LYS92 TRP96 ARG64 PRO502 ALA504 GLN119 ASP117 THR94 LEU115 TRP505</p>	<p>ASP82, GLU523 LEU483, SER181 LEU138, PRO110 SER374, ALA484 ILE183, ALA104 PHE105, GLY137</p>	<p>PHE48, ASN424 GLU204, ASN205 ALA82, HIS383 LEU291, ARG403 ASP422, PHE85 SER384, GLU292 PRO111, SER386 PRO297, GLN81</p>	<p>ARG400 ASN403 GLY402 ILE451 PHE397 VAL452 GLN453 TYR462 VAL398</p>	<p>THR150 THR144 ARG148 LEU121 LYS143 ASN123 SER146 ASN126 ILE125</p>
<p>Guaiacyl β-O-4 syringyl</p>  <p>β-β syringyl β-O-4 guaiacyl</p>	<p>ASN284, ASN324 TYR261, GLN264 LEU316, THR259 TYR266, THR232 GLU325, SER433 PHE293, SER234 LEU323</p>	<p>THR369 PRO367 ALA101 ASP481 SER134 LEU133 TYR512 ALA517 LEU480 HIS132 PHE102</p>	<p>THR340, PRO343 ASP152, GLY338 THR466, GLY341 ARG448, PHE465 PHE430, LYS64 ASN250, SER431</p>	<p>PRO282, GLN216 LEU283, LEU284 GLU285, SER408 PHE218, SER279 GLN214, ALA391 GLY392</p>	<p>VAL452 GLN459 ASP455 GLY402 TYR462 THR447 THR461 ASN422 ARG400 ILE451 GLN453 TRP463</p>	<p>ASN562 THR144 ARG148 LYS143 PHE147 ASN123 THR560 SER146 ASN126 ILE125</p>

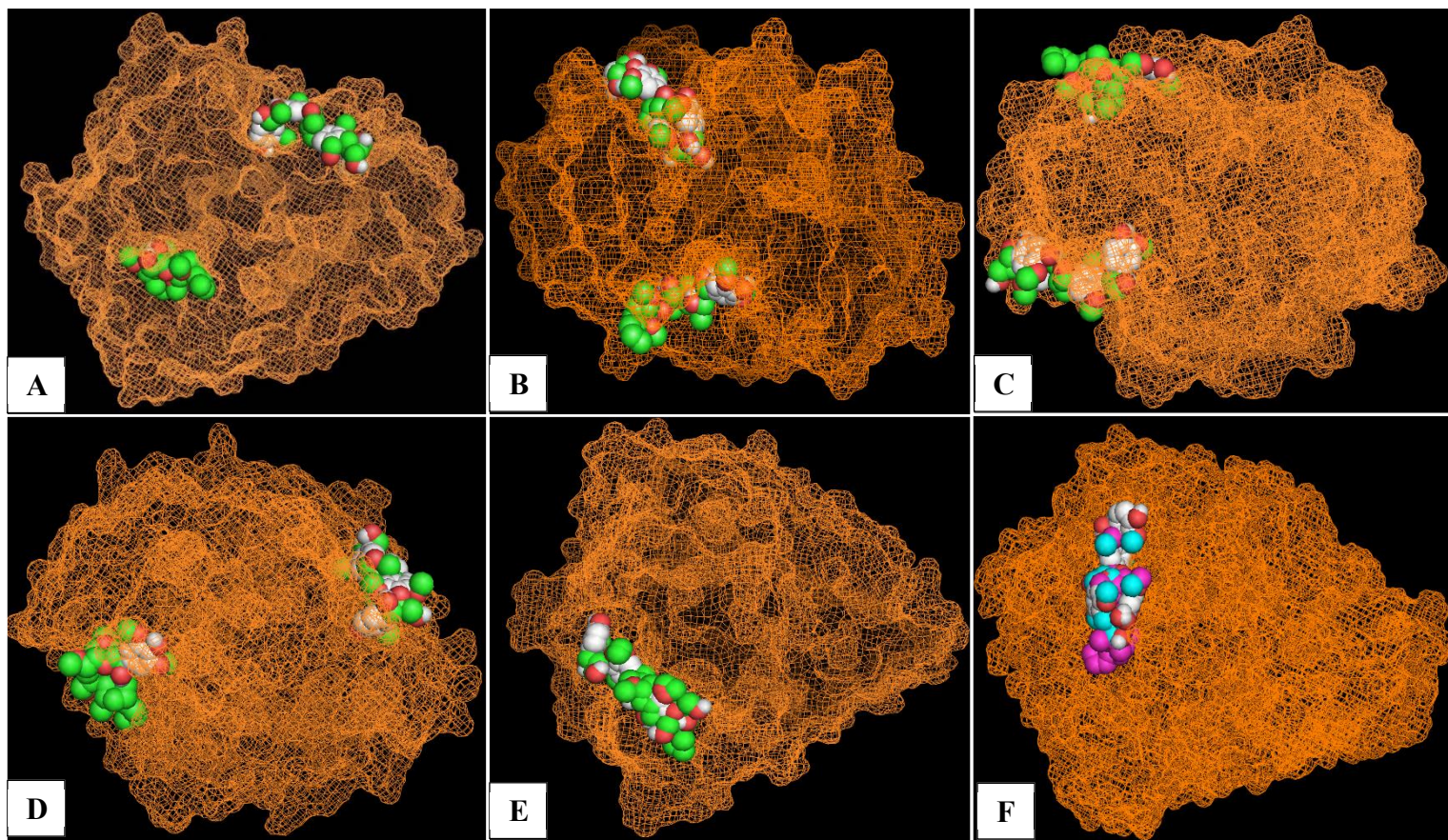


Figure 8.4: Protein docking of laccase protein molecular models with syringyl β -O-4 syringyl β -O-4 sinapyl alcohol (Trimer) and guaiacyl β -O-4 syringyl β - β syringyl β -O-4 guaiacyl (Tetramer), A) *Phlebia brevispora*, B) *Wolfiporia cocos*, C) *Dichomitus squalens*, D) *Fomitopsis pinicola*, E) *Chaetomium globosum* and F) *Cadophora* DSE1049 v1.0.

8.5. Discussion

In our present study, we have performed a comparative modeling and molecular docking study of fungal laccase protein sequences, to understand and reveal the lignin degrading abilities exhibited by white rot, brown rot and soft rot fungi. We have retrieved all the laccase protein sequences (genome wide) of the selected fungi however, based on the sequence similarity against experimentally validated protein structures (PDB) we have considered Phlbr1-25201, Dsqual-59186, Fompi3-45001, Wcocos-139080, Chagl1-12114 Cadophora-560981. The above considered laccases possess the essential domains belonging to cupredoxin superfamily and specifically contains Cu_oxidase type 1, 2 and 3 of multicopper oxidase domains. We have observed that *P. brevispora*, *D. squalens* (white rot)

laccase sequences were closely related to the experimentally determined laccase of *Trametes versicolor* (1GYC) and laccase sequences of *F. pinicola* and *W. cocos* fall under the same branch and share good similarity with *T. versicolor* (1GYC) and white rot fungal laccases. Whereas soft rot fungal laccases *C. globosum* and *Cadophora* sp were found to be closely related to the experimentally determined laccase of *Melanocarpus albomyces* (1GW0) (Figure 8.1A). Multicopper oxidases (MCO) are group of enzymes performing single electron oxidation of various substrate with an associated four electron reduction of molecular oxygen to water molecule [49]. The enzymes belonging to MCO class were further classified into laccase, ferroxidase, ascorbate oxidase and ceruloplasmin. Previous studies have revealed that MCOs consists of two active sites a) substrate oxidizing site (blue type-1 (T1) copper site) and b) oxygen binding site (tri nuclear copper site containing three type-2 (T2) or type-3 (T3) coppers) [49]. The electrons are transferred from T1 copper site to T2/T3 copper site through a set of highly conserved amino acid residues (MCO-specific patterns) [6, 50, 51] (Figure 8.1B). Except ceruloplasmin (six domains) and bacterial laccases (two domains), most of the MCOs contain three cupredoxin domains, based on the presence of these domains the length of MCOs ranges between 300 to 1000 residues and can contain up to six copper ions [6, 49].

All the fungal laccase protein sequences selected for the present study ranges between 479 to 619 (Table 8.1) containing four copper ions mostly in contact with histidine residues. The aminoacid composition of the selected laccases show the higher content of negatively charged amino acids (aspartic and glutamic acids) can explain about the acidic nature (theoretical pI values obtained) of the laccases. The concentration of tyrosine, tryptophan and cysteine residues reflects the extinction coefficients of fungal laccases [27]. The physico-chemical properties such as theoretical pI and molecular weight, extinction coefficient values for soft rot fungal laccases (*C. globosum* and *Cadophora* sp) were comparatively higher than brown and white rot laccases. The lower extinction coefficients of white and brown rot laccases might be due to the lower content of phenylalanine, tyrosine, tryptophan and cysteine residues [27]. Studies have reported that proteins exhibiting an instability index lesser than 40 possess an in vivo half-life of 5h and instability index greater than 40 has an in vivo half-life period of 16h [52]. The instability index (used for estimation of in vivo half-life of proteins) values report that the selected laccases have a long in vivo half-life period of 16 hours except for *W. cocos* laccase protein (in vivo half-life period of < 5 hours) [28, 52]. The aliphatic index (determined using the relative volume occupied by aliphatic side chains of alanine, valine, leucine and isoleucine) is used as a positive factor for increase in

thermal stability of globular proteins [29]. The selected fungal laccases exhibited aliphatic index values in the range of 74.44 to 90.04, which suggests the stability of selected laccases at wider temperature ranges. The hydrophobicity/hydrophilicity of protein can be defined using the GRAVY (grand average of hydropathicity) index values, where the positive and negative values denote for hydrophobic and hydrophilic natures of the protein. All the selected fungal laccase proteins except *P. brevispora* were found to be hydrophilic in nature reporting that these laccases interact better with water. The hydropathicity plots generated using Discovery studio visualizer® for the laccase protein sequences also support the above reported values.

The secondary structure analysis of the selected laccase protein sequences using SOPMA web server has revealed that the percentage of random coiled secondary structures content is higher followed by extended strand percentage (Table 8.3). Earlier studies have reported that random coiled secondary structures are involved in imparting flexibility, turnover and conformational changes of the enzymes [53]. Upon Motif scan analysis, all the selected laccase proteins were commonly found to contain motif sequences for phosphorylation, glycosylation, myristoylation and multi copper oxidase patterns (Table 8.2). Phosphorylation is significant process which effects the functional and structural activities of proteins and regulates the cell behavior in eukaryotes, through controlling its intrinsic biological activity, cellular localization and interaction with other proteins [54]. The laccase protein sequences commonly showed casein kinase-II, protein kinase C and cAMP/cGMP dependent protein kinase (*F. pinicola* and *W. cocos*) phosphorylation sites. Similarly, we have observed myristoylation patterns commonly among the laccases, myristoylation (modification of proteins with myristic acid) alters the conformational stability of proteins through interacting with hydrophobic membranes and domains of the protein, it also plays crucial role in cellular signaling and extracellular export of the proteins [18, 55-57]. Other commonly observed patterns include glycosylation and amidation these post translational modifications are involved in imparting thermal stability, copper retention and retains its biological activities [58]. According to Marion et al (1998), most of the laccases are extracellular glycoproteins [59], results obtained from CELLO v2.5 subcellular location predictor showed that the above selected laccases were extracellular. We have observed two disulfide bonds in *P. brevispora*, *D. squalens*, *W. cocos* and three disulfide bonds in *F. pinicola*, *C. globosum* and *Cadophora* sp. Previous studies have reported that extracellular proteins contain more cysteine residues and disulfide bonds, compared to intracellular proteins, thus

disulfide bonds in extracellular laccase proteins play crucial role during protein folding and stability of the protein [60, 61]. Presence of signal peptide cleavage sites (except *W. cocos*) and absence of transmembrane helices (except *D. squalens*), supports the extracellular nature of the selected fungal laccases.

Homology modeling studies are highly significant in determining protein structures for the proteins which lack the experimental structures, in the past few decades a wide range of efficient tools and servers were developed which can perform modeling studies of proteins even with 30% of sequence identity with an accuracy achieved from low resolution X-ray structures [18, 62]. Thus performing computational studies predicting the structural and functional properties of commercially important enzymes will significantly support in planning biological experiments based on these enzymes [18]. Homology modeling of the selected fungal laccases was performed using SWISS Model automated server, it selects best template based on the sequence identity results obtained through BLAST and HHblits [37]. SWISS-Model server performs a range of quality checks and refines the side chains and loops of the targets using template structures, it generates quality scores such as QMEAN4 and GMQE scores. According to Benkert et al (2009), QMEAN is a compound scoring function used for the estimation of global and local model qualities [63]. QMEAN uses four structural descriptors a) torsion angle potential b) all atom c)C-beta interactions d) solvation potential, for the estimation of local and global qualities of the modelled structures between 0 (unreliable) to 1(better model) [63]. Global model quality estimation (GMQE) combines properties from the template and target alignment, GMQE is estimated in between 0 (unreliable) to 1 (reliable mode) [37]. We have obtained positive QMEAN scores for *P. brevispora*, *D. squalens* and *C. globosum* whereas negative QMEAN scores for *F. pinicola*, *W. cocos* and *Cadophora* sp laccases. SWISS Model server provides the quality estimations in the form of a chart where 3D modelled structures have a color gradient in between blue to red which defines the resolutions between 1 and 3.5Å respectively, where the greater blue values represents a reliable structures [37]. Estimated global model qualities (GMQE) for the modelled laccases were found to be in between the range of 0.78 to 0.97, which indicates a good quality and reliable structures (Figure 8.3).

The above obtained 3D modelled structures of fungal laccases were further refined using GalaxyRefine and KoBaMIN web servers. KoBaMIN web server performs a knowledge based potential of mean force correction of the modeled structures, KoBaMIN internally uses

knowledge based potential derived from PDB structures which also uses the energy function which implements the effects of solvent and crystal environment and MESHI for optimizing stereochemistry of the models [39]. GalaxyRefine is online web server which primarily rebuilds the side chains and also performs repacking of side chains and later relaxes the overall structure through molecular dynamic simulation methods, It improves both local and global structural qualities of the predicted models [38]. We have applied a combination of above-mentioned model refining methods differentially to obtain a best protein model with reliable Ramachandran scores and higher acceptable residues. We have performed validation of the refined models using PROCHECK, RAMPAGE and ERRAT analysis (Table 8.4). The refined selected laccase models upon Ramachandran plot analysis revealed that 3D-modelled protein structures have >90% of residues in allowed regions. For the predicted protein structures, it is ideal to contain at least 90% of the residues must be in regions of allowed regions, which suggests the predicted laccase structures were of good quality. We have also validated the refined laccase structures using PROQ web server, it generates two quality metrics for the estimation of model quality they are LG score and Max sub-score [43]. Where LG score must be greater than 4 and max sub score must be greater than 0.8 (very good) for the reliable protein structures. All the predicted laccase models have attained LG and max sub scores between the range of (4.48-5.22) and (0.295-0.389) respectively, which suggests that the refined 3D models of fungal laccases were of good quality [43]. Finally, we have used ERRAT web server for the validation of the refined models, ERRAT analyses the non-bonded interactions among different atom types based on characteristic atomic interactions [42]. We have observed that all the predicted model structures have attained an overall quality factor >50 which confirms that the refined laccase models were of good quality [18, 42]. From the predicted 3D modelled fungal laccase structures, we can infer that soft rot fungal laccases *C. globosum* and *Cadophora* sp have shown significant differences in structural and physico-chemical properties when compared to white rot and brown rot fungal laccases. In order, to reveal these differences we have superimposed all the refined laccase models using SWISS PDB viewer. Results obtained from the superimposition studies have revealed that RMSD values obtained for soft rot fungal laccases (*C. globosum* and *Cadophora* sp) showed slight structural differences when compared to white and brown rot laccases (Table 8).

Table 8.8: Lists the root mean square deviation (RMSD Å) values obtained from the comparison studies of white, brown and soft rot fungal laccases:

Organism	<i>P. brevispora</i>	<i>D. squalens</i>	<i>F. pinicola</i>	<i>W. cocos</i>	<i>C. globosum</i>	<i>Cadophora</i> sp
<i>P. brevispora</i>	n/a	0.62	0.83	0.46	1.11	1.06
<i>D. squalens</i>	0.62	n/a	0.84	0.68	1.13	1.08
<i>F. pinicola</i>	0.83	0.84	n/a	0.80	1.23	1.17
<i>W. cocos</i>	0.46	0.68	0.80	n/a	1.07	1.12
<i>C. globosum</i>	1.11	1.13	1.23	1.07	n/a	0.68
<i>Cadophora</i> sp	1.06	1.08	1.17	1.12	0.68	n/a

Several structural studies were conducted in the past which has revealed the structural and functional properties of plant cell wall-based compounds and especially lignin-based compounds. Ralph S.A et al (2004) has conducted structural studies of lignin and other plant cell wall compounds, and developed a single source database for lignin model compounds [44]. We have considered six lignin model compounds for the present protein docking studies, lignin building monomers (sinapyl, coniferyl and *p*-coumaryl alcohol), dimer (guaiacyl 4-O-5 guaiacyl), trimer (syringyl β -O-4 syringyl β -O-4 sinapyl alcohol) and tetramer (guaiacyl β -O-4 syringyl β - β syringyl β -O-4 guaiacyl) [19]. We have used AutoDock Vina and Tools for achieving the protein docking studies using the above-mentioned lignin model compounds [46]. AutoDock Vina is a fast and accurate method of ligand protein docking tools which will facilitate flexible docking studies [46]. We have clearly observed that in *P. brevispora*, *W. cocos* laccase protein, monomers, dimer and trimer were mostly found to bind to the same pocket in different conformations, with tetramer binding at different (large) pocket. Whereas in *D. squalens* laccase protein we have observed that monomers, dimer and tetramer bind at the same pocket and contrastingly trimer binds separately at different pocket. The lignin model compounds (monomer, dimer, trimer and tetramers) were found bind separately at different pockets in *F. pinicola* laccase protein. Both soft rot fungal (*Cadophora* sp and *C. globosum*) laccases when docked with lignin model compounds, showed similar binding patterns monomers, dimer bound at different pockets however,

trimer and tetramer bound at the same pockets respectively. The binding patterns of trimer and tetramer by modelled fungal laccases were shown below (Figure 8.4).

We have compared the aminoacid residues of fungal laccases which are occurring in close interactions with lignin model compounds. Commonly found aminoacid residues between the white rot fungal laccases were PRO, HIS, SER, PHE, GLY, ALA, TYR, LEU, LYS, GLN, THR. Similarly, common aminoacid residues between brown rot fungal laccases were ARG, PRO, ASN, GLN, PHE, ASP, ALA, LEU, GLU, SER, GLY and LYS. Finally, common aminoacid residues between soft rot fungal laccases were TYR, ASP, VAL, THR, PHE, ASN, TRP, LYS, LEU, ARG, PRO, ILE and GLN. The commonly occurring aminoacid residues among the selected white, brown and soft rot fungal laccases were found to be PRO, PHE, LEU, LYS and GLN, other commonly found aminoacids were SER, GLY, ALA (white and brown rot), ARG, ASN, ASP (brown and soft rot) and TYR, THR (white and soft rot) respectively.

In support of our present study, protein docking studies using lignin model compounds (sinapyl alcohol, dimer, trimer and tetramer) earlier by Awasthi et al (2015), has revealed the, interactions of 11 aminoacid residues commonly (LEU, ASP, ASN, PHE, SER, PHE, GLY, ALA, PRO, ILE, and HIS) with all the lignin models [19]. Crystallographic studies of *Melanocarpus albomyces* laccase using lignin model compounds by Kallio et al (2009), has revealed the interactions of seven amino acid residues (ALA, PRO, GLU, LEU, PHE, TRP, and HIS) with the lignin model compounds [64]. According to Awasthi et al (2015), fungal laccases exhibiting higher redox potential were found to include phenylalanine in the active binding site of the lignin model compounds, which is in accordance with the present results [19]. Molecular docking studies of *G. lucidum* and *P. ostreatus* 3D-predicted laccases with ABTS, has revealed the close interactions of aminoacid residues (Phe, Asp, Ser, Pro, Gly, Ile, His and Gly, Val, Pro, Asp, Ser, Asn, Phe, Ile, Trp and His) with the ligand [15]. Similarly, studies conducted by Morozova et al (2007) have revealed that laccases exhibiting high potential were usually found to contain phenylalanine residue at its axial ligand of type-T1 copper binding site. [65]. According to Xu (1996), laccases containing PHE residue at type-T1 copper binding site has showed higher redox potential than MET containing laccases [66].

Results obtained in our present study highlights the structural and functional properties exhibited by white, brown and soft rot fungal laccase models. We have clearly observed that white and brown rot fungi exhibited clear and strong binding efficiencies towards lignin model compounds when compared to soft rot fungal laccases. We have also

seen that physico-chemical and structural properties of soft rot fungi exhibited little but significant differences upon comparison with white and brown rot fungi. Further molecular dynamic simulation and corresponding wet lab experiments must be performed to understand the catalytic efficiencies of fungal laccases. Increasing genome and transcriptome wide studies were continuously revealing the molecular complexities of several fungi. Efficient methods for fishing high potential laccases will play a crucial role in developing genetically efficient microorganisms for the degradation of lignin, which will significantly help the growing biofuel and bioremediation industries.

References

1. Dwivedi UN, Singh P, Pandey VP, Kumar A. Structure–function relationship among bacterial, fungal and plant laccases. *Journal of Molecular Catalysis B: Enzymatic*. 2011; 68: 117-28.
2. Kameshwar AKS, Qin W. *Lignin Degrading Fungal Enzymes. Production of Biofuels and Chemicals from Lignin*: Springer; 2016. p. 81-130.
3. Yoshida H. LXIII.—chemistry of lacquer (Urushi). Part I. communication from the chemical society of Tokio. *Journal of the Chemical Society, Transactions*. 1883; 43: 472-86.
4. Bertrand G. Sur la presence simultanee de la laccase et de la tyrosinase dans le suc de quelques champignons. *CR Hebd Seances Acad Sci*. 1896; 123: 463-5.
5. Singh D, Sharma KK, Dhar MS, Viridi JS. Molecular modeling and docking of novel laccase from multiple serotype of *Yersinia enterocolitica* suggests differential and multiple substrate binding. *Biochemical and biophysical research communications*. 2014; 449: 157-62.
6. Messerschmidt A, Huber R. The blue oxidases, ascorbate oxidase, laccase and ceruloplasmin Modelling and structural relationships. *European journal of Biochemistry*. 1990; 187: 341-52.
7. Yao B, Ji Y. Lignin biodegradation with laccase-mediator systems. *Frontiers in Energy Research*. 2014; 2: 12.
8. Giardina P, Faraco V, Pezzella C, Piscitelli A, Vanhulle S, Sannia G. Laccases: a never-ending story. *Cellular and Molecular Life Sciences*. 2010; 67: 369-85.
9. Eggert C, Temp U, Eriksson K-E. The ligninolytic system of the white rot fungus *Pycnoporus cinnabarinus*: purification and characterization of the laccase. *Applied and Environmental Microbiology*. 1996; 62: 1151-8.
10. Bourbonnais R, Paice M, Freiermuth B, Bodie E, Borneman S. Reactivities of various mediators and laccases with kraft pulp and lignin model compounds. *Applied and environmental microbiology*. 1997; 63: 4627-32.
11. Camarero S, Ibarra D, Martínez MJ, Martínez ÁT. Lignin-derived compounds as efficient laccase mediators for decolorization of different types of recalcitrant dyes. *Applied and environmental microbiology*. 2005; 71: 1775-84.
12. Rencoret J, Pereira A, José C, Martínez AT, Gutiérrez A. Laccase-mediator pretreatment of wheat straw degrades lignin and improves saccharification. *BioEnergy Research*. 2016; 9: 917-30.
13. Kunamneni A, Ballesteros A, Plou FJ, Alcalde M. Fungal laccase—a versatile enzyme for biotechnological applications. *Communicating current research and educational topics and trends in applied microbiology*. 2007; 1: 233-45.
14. Thurston CF. The structure and function of fungal laccases. *Microbiology*. 1994; 140: 19-26.
15. Rivera-Hoyos CM, Morales-Álvarez ED, Poveda-Cuevas SA, Reyes-Guzmán EA, Poutou-Piñales RA, Reyes-Montañón EA, et al. Computational analysis and low-scale constitutive expression of laccases synthetic genes *GILCC1* from *Ganoderma lucidum* and *POXA 1B* from *Pleurotus ostreatus* in *Pichia pastoris*. *PloS one*. 2015; 10: e0116524.
16. Meshram RJ, Gavhane A, Gaikar R, Bansode T, Maskar A, Gupta A, et al. Sequence analysis and homology modeling of laccase from *Pycnoporus cinnabarinus*. *Bioinformation*. 2010; 5: 150.

17. Wong K-S, Cheung M-K, Au C-H, Kwan H-S. A novel *Lentinula edodes* laccase and its comparative enzymology suggest guaiacol-based laccase engineering for bioremediation. *PloS one*. 2013; 8: e66426.
18. Tamboli AS, Rane NR, Patil SM, Biradar SP, Pawar PK, Govindwar SP. Physicochemical characterization, structural analysis and homology modeling of bacterial and fungal laccases using in silico methods. *Network Modeling Analysis in Health Informatics and Bioinformatics*. 2015; 4: 17.
19. Awasthi M, Jaiswal N, Singh S, Pandey VP, Dwivedi UN. Molecular docking and dynamics simulation analyses unraveling the differential enzymatic catalysis by plant and fungal laccases with respect to lignin biosynthesis and degradation. *Journal of Biomolecular Structure and Dynamics*. 2015; 33: 1835-49.
20. Binder M, Justo A, Riley R, Salamov A, Lopez-Giraldez F, Sjökvist E, et al. Phylogenetic and phylogenomic overview of the Polyporales. *Mycologia*. 2013; 105: 1350-73.
21. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, et al. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science*. 2012; 336: 1715-9.
22. Berka RM, Grigoriev IV, Otilar R, Salamov A, Grimwood J, Reid I, et al. Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. *Nature biotechnology*. 2011; 29: 922-7.
23. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*. 1997; 25: 3389-402.
24. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular biology and evolution*. 2016: msw054.
25. Larkin MA, Blackshields G, Brown N, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *bioinformatics*. 2007; 23: 2947-8.
26. Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExPASy server: Springer; 2005.
27. Gill SC, Von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. *Analytical biochemistry*. 1989; 182: 319-26.
28. Guruprasad K, Reddy BB, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein engineering*. 1990; 4: 155-61.
29. Atsushi I. Thermostability and aliphatic index of globular proteins. *Journal of biochemistry*. 1980; 88: 1895-8.
30. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*. 1982; 157: 105-32.
31. Combet C, Blanchet C, Geourjon C, Deleage G. NPS@: network protein sequence analysis. Elsevier Current Trends; 2000.
32. Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel CV, Hau J, et al. MyHits: improvements to an interactive resource for analyzing protein sequences. *Nucleic acids research*. 2007; 35: W433-W7.
33. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. *Proteins: Structure, Function, and Bioinformatics*. 2006; 64: 643-51.
34. Lin H-H, Hsu J-C, Hsu Y-N, Pan R-H, Chen Y-F, Tseng L-Y. Disulfide connectivity prediction based on structural information without a prior knowledge of the bonding state of cysteines. *Computers in biology and medicine*. 2013; 43: 1941-8.
35. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*. 2011; 8: 785-6.
36. Kiemer L, Bendtsen JD, Blom N. NetAcet: prediction of N-terminal acetylation sites. *Bioinformatics*. 2005; 21: 1269-70.
37. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic acids research*. 2003; 31: 3381-5.
38. Heo L, Park H, Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic acids research*. 2013; 41: W384-W8.
39. Rodrigues JP, Levitt M, Chopra G. KoBaMIN: a knowledge-based minimization web server for protein structure refinement. *Nucleic acids research*. 2012: gks376.

40. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*. 1993; 26: 283-91.
41. Lovell SC, Davis IW, Arendall WB, de Bakker PI, Word JM, Prisant MG, et al. Structure validation by $C\alpha$ geometry: ϕ , ψ and $C\beta$ deviation. *Proteins: Structure, Function, and Bioinformatics*. 2003; 50: 437-50.
42. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science*. 1993; 2: 1511-9.
43. Cristobal S, Zemla A, Fischer D, Rychlewski L, Elofsson A. A study of quality measures for protein threading models. *BMC bioinformatics*. 2001; 2: 5.
44. Ralph SA, Ralph J, Landucci L, Landucci L. NMR database of lignin and cell wall model compounds. US Forest Prod Lab, Madison, WI (<http://ars.usda.gov/Services/docs.htm>). 2004.
45. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*. 2009; 30: 2785-91.
46. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*. 2010; 31: 455-61.
47. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *electrophoresis*. 1997; 18: 2714-23.
48. Nakai K. Protein sorting signals and prediction of subcellular localization. *Advances in protein chemistry*. 2000; 54: 277-344.
49. Sirim D, Wagner F, Wang L, Schmid RD, Pleiss J. The Laccase Engineering Database: a classification and analysis system for laccases and related multicopper oxidases. *Database*. 2011; 2011: bar006.
50. Ouzounis C, Sander C. A structure-derived sequence pattern for the detection of type I copper binding domains in distantly related proteins. *FEBS letters*. 1991; 279: 73-8.
51. Kumar S, Phale PS, Durani S, Wangikar PP. Combined sequence and structure analysis of the fungal laccase family. *Biotechnology and Bioengineering*. 2003; 83: 386-94.
52. Rogers S, Wells R, Rechsteiner M. Amino acid sequences common to rapidly degrade proteins: the PEST hypothesis. *Science*. 1986; 234: 364-9.
53. Buxbaum E. *Fundamentals of protein structure and function*: Springer; 2007.
54. Cohen P. The regulation of protein function by multisite phosphorylation—a 25 year update. *Trends in biochemical sciences*. 2000; 25: 596-601.
55. Podell S, Gribskov M. Predicting N-terminal myristoylation sites in plant proteins. *Bmc Genomics*. 2004; 5: 37.
56. Zheng J, Knighton DR, Taylor SS, Xuong NH, Sowadski JM, Eyck LFT. Crystal structures of the myristylated catalytic subunit of cAMP-dependent protein kinase reveal open and closed conformations. *Protein Science*. 1993; 2: 1559-73.
57. Olsen HB, Kaarsholm NC. Structural effects of protein lipidation as revealed by LysB29-myristoyl, des (B30) insulin. *Biochemistry*. 2000; 39: 11893-900.
58. Walsh G. Post-translational modifications in the context of therapeutic proteins: An introductory overview. *Post-translational Modification of Protein Biopharmaceuticals*. 2009: 1-14.
59. Heinzkill M, Bech L, Halkier T, Schneider P, Anke T. Characterization of laccases and peroxidases from wood-rotting fungi (family Coprinaceae). *Applied and Environmental Microbiology*. 1998; 64: 1601-6.
60. Bradshaw RA. Protein translocation and turnover in eukaryotic cells. *Trends in biochemical sciences*. 1989; 14: 276-9.
61. Nakashima H, Nishikawa K. Discrimination of intracellular and extracellular proteins using amino acid composition and residue-pair frequencies. *Journal of molecular biology*. 1994; 238: 54-61.
62. Xiang Z. Advances in homology protein structure modeling. *Current Protein and Peptide Science*. 2006; 7: 217-27.
63. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic acids research*. 2009: gkp322.
64. Kallio J, Auer S, Jänis J, Andberg M, Kruus K, Rouvinen J, et al. Structure–function studies of a *Melanocarpus albomyces* laccase suggest a pathway for oxidation of phenolic compounds. *Journal of molecular biology*. 2009; 392: 895-909.

65. Morozova O, Shumakovich G, Gorbacheva M, Shleev S, Yaropolov A. "Blue" laccases. *Biochemistry (Moscow)*. 2007; 72: 1136-50.
66. Xu F. Oxidation of phenols, anilines, and benzenethiols by fungal laccases: correlation between activity and redox potentials as well as halide inhibition. *Biochemistry*. 1996; 35: 7608-14.

Chapter-9

Overall Discussion and Future Recommendations

The work presented in this dissertation focussed on understanding the fungal molecular mechanisms underlying the lignocellulose breakdown and conversion patterns. We have developed an efficient metadata analysis pipeline for understanding the genomic and transcriptomic datasets of model wood-decaying fungi. Rapidly increasing whole-genome sequencing and genome-wide transcriptomic studies of various wood-decaying Basidiomycetous fungi are revealing about the extrinsic decaying abilities of fungi. *Phanerochaete chrysosporium* and *Postia placenta* were the first complete annotated genome sequences to represent Basidiomycota division. Availability of complete annotated genomes, well-designed genome-wide transcriptomic studies and extensive literature talking about their growth patterns and decaying abilities have convinced us to choose these two-model wood-decaying fungi for our metadata analysis workflow.

The metadata analysis of *P. chrysosporium* and *P. placenta* gene expression datasets has provided us with a highly significant list of lignocellulolytic enzymes and several other carbohydrate and aromatic metabolizing enzymes commonly expressed among different datasets developed under different growth conditions (customized medium and complex plant biomass containing medium) respectively. Our study for the first time has reported a list of highly resistant list of lignocellulolytic enzymes significantly expressed commonly among all the *P. chrysosporium* and *P. placenta* gene expression datasets. Based on the obtained results we have reported tentative molecular network of genes, proteins and enzymes involved during the breakdown and conversion of plant cell-wall components. Our present study has also proved that gene expression of *P. chrysosporium* is strongly influenced by the growth substrates and the incubation periods. The gene expression profiles observed in GSE69008 and GSE69461 datasets showed that, when *P. chrysosporium* is cultured on plant biomass it initially secretes various lignin and hemicellulolytic enzymes followed by secretion of cellulolytic enzymes in the later growth stages. Previous studies have reported that lignin degradation by fungi is majorly dependent on lignin oxidizing enzymes especially laccases, ligninolytic peroxidases and other oxidoreductases. However, in our analysis we have observed that apart from these lignin oxidizing enzymes several other aromatic compound degrading enzymes were differentially expressed among all the *P. chrysosporium* and *P.*

placenta gene expression datasets. Thus, based on their degree of involvement during the process of lignin degradation we have tentatively classified the differentially expressed enzymes into primary-lignin degrading enzymes, secondary - aromatic compound degrading and tertiary other necessary enzymes respectively.

During the process of wood-degradation fungi encounters several antioxidant plant secondary metabolites such as flavonoids, quinones, stilbenes, tannins and other phenolic compounds. In order to escape this strong attack fungi have developed an efficient detoxification and stress responsive mechanisms to survive and continue the process of wood decay. In our analysis we have consistently observed a set of genes involved in fungal detoxification and stress responsive mechanisms were differentially expressed among all the gene expression datasets. Based on their involvement we have also classified the differentially expressed detoxification and stress genes into phase-I and phase-II metabolic enzymes respectively. We have also reported a tentative network of all the genes and enzymes involved during lignin degradation and detoxification -stress responsive mechanisms of *P. chrysosporium*. Results obtained in our analysis also reveals a coordinated action of enzymes involved in lignin depolymerization and detoxification-stress responses under ligninolytic conditions respectively. Our analysis has also reported the cellular regulation mechanisms controlling the expression and protein turnover of various lignocellulolytic enzymes respectively. This study has also emphasized that *P. chrysosporium* cultured on natural plant biomass highly expressed several genes involved in gene regulation processes such as information storage and processing especially DNA damage, repair and recombination mechanisms, mRNA splicing, histone acetyltransferases respectively.

Though *P. placenta* genome lacks several lignocellulolytic CAZymes, it has developed its own signature wood-decay mechanism by employing highly reactive Fenton's reaction for the breakdown and conversion of lignocellulosic components. Major findings of our analysis include a) the strong dependency of *P. placenta* on the Fenton's reaction system for the modification and degradation of lignocellulosic components, b) Higher expression of genes encoding for hemicellulolytic glycoside hydrolases even under cellulolytic conditions proves the synchronized action of both cellulolytic and hemicellulolytic hydrolytic enzymes. Based on the reports from previous plant biomass degradation studies and the results obtained in our present study clearly show that, *P. placenta* exhibits stronger ability to degrade cellulose

and hemicellulose by selectively modifying lignin. We have also reported a similar tentative network of genes and enzymes employed by *P. placenta* during wood-decaying mechanisms.

The metadata analysis of gene expression datasets has further encouraged us to perform a genome-wide comparative analysis of several wood-decaying fungi to understand the total plant biomass degrading abilities. We have primarily analyzed complete genome-wide annotations of 42 wood-decaying separated into white, brown and soft rot fungi. We have extensively analyzed CAZymes, KOG, KEGG, InterPro annotations, based on their genome-wide annotations we have tentatively calculated cellulolytic, hemicellulolytic, ligninolytic and pectinolytic abilities of the fungi. Results obtained in this study has suggested that white rot fungal strains are highly suitable for the degradation of lignin, other aromatic compounds and environmental pollutants, soft rot fungal strains are highly suitable in cellulose, hemicellulose and pectin degradation studies thus highly suitable in biofuel and biorefining industries. Further we have also extended our analysis to understand the lignocellulolytic abilities of *Neocallimastigomycota* division fungi by comparing CAZymes, KOG, KEGG, InterPro, SM-Clusters and MEROPS annotations. This study has revealed that genomes of anaerobic fungi completely lack genes encoding for lignin degrading auxiliary activity enzymes. Contrastingly, these fungi outnumbered other fungi by having highest number of CAZyme encoding genes. Also, we have explicitly reported the genes, enzymes and the mechanisms involved in structure and functioning of the cellulosomes and hydrogenosomes. Thus, understanding the genetic material coding for the lignocellulolytic enzymes will significantly benefit researchers to choose genetically better strain for their studies. However, further relevant studies must be conducted to optimize the appropriate growth and environmental conditions to enhance the expression and protein turnover of these lignocellulolytic enzymes.

Finally, we have performed an extensive homology modeling and molecular docking of laccases (multi copper oxidase) protein sequences of white, brown and soft rot fungi, using six lignin model compounds ranging from monomer, dimer, trimer and tetramer units to understand the catalytic process. This study has highlighted and compared the structural and functional properties of selected white, brown and soft rot fungal laccases. Results obtained in this study revealed that white and brown rot fungi exhibited clear and strong binding efficiencies towards lignin model compounds when compared to soft rot fungal laccases. Also, we have observed that physico-chemical and structural properties of soft rot fungi exhibited little but significant differences upon comparison with white and brown rot fungi. However,

further molecular dynamic simulation and corresponding wet lab experiments must be performed to understand the catalytic efficiencies of fungal laccases.

Metadata analysis work frame reported in this thesis can be used for studying the decaying patterns and molecular mechanisms of several other fungi and bacteria. The gene expression metadata analysis of *P. chrysosporium* and *P. placenta* has specifically reported the protein ID's of significant and highly reactive lignocellulolytic, carbohydrate and aromatic metabolizing enzymes. The information about these significant and highly active protein-ID's of *P. chrysosporium* and *P. placenta* can potentially be used at different scales to improve the protein turnover and expression to achieve higher degradation and conversion rates of plant biomass. Further high throughput genomic and proteomic studies must be conducted to understand and reveal the structural and functional properties of these highly reactive lignocellulolytic enzymes. The studies should also be conducted further to prove and confirm the involvement of tentative molecular networks reported in this dissertation. The comparative genomic data analysis pipeline should be applied further to several other wood-decaying microorganisms to retrieve a best lignocellulolytic fungal or bacterial strain based on their genome-wide cellulolytic, hemicellulolytic, ligninolytic and pectinolytic abilities respectively. The metadata analysis pipeline reported in this thesis can also be efficiently applied for analyzing the metagenomic datasets of gut and rumen for understanding the structural and functional involvement of microbial communities in metabolism of plant biomass components. Developing an automated data retrieval and data analysis for obtaining significant list of differentially expressed genes involved in breakdown and conversion of biomass and other organic compounds will potentially benefit the researchers in related fields in experimental design and execution.

PUBLICATION LIST

1. **Ayyappa Kumar Sista Kameshwar**, and Wensheng Qin. "Structural and functional properties of pectin and lignin-carbohydrate complexes de-esterases: a review." **Bioresources and Bioprocessing** 5, no. 1 (2018): 43.
2. **Ayyappa Kumar Sista Kameshwar & Wensheng Qin** (2018) Understanding the structural and functional properties of carbohydrate esterases with a special focus on hemicellulose deacetylating acetyl xylan esterases, **Mycology**, DOI: 10.1080/21501203.2018.1492979
3. **Ayyappa Kumar Sista Kameshwar**, and Wensheng Qin. "Genome Wide Analysis Reveals the Extrinsic Cellulolytic and Biohydrogen Generating Abilities of Neocallimastigomycota Fungi." **Journal of Genomics** 6 (2018): 74-87.
4. **Ayyappa Kumar Sista Kameshwar**, and Wensheng Qin. "Isolation and Screening of Cellulose-Degrading Microorganisms from Different Ecological Niches." In **Cellulases**, pp. 47-56. Humana Press, New York, NY, (2018).
5. **Ayyappa Kumar Sista Kameshwar**, and Wensheng Qin. "Purification and Characterization of the Total Cellulase Activities (TCA) of Cellulolytic Microorganisms." In **Cellulases**, pp. 255-269. Humana Press, New York, NY, (2018).
6. **Ayyappa Kumar Sista Kameshwar**, Qin W. Molecular Networks of Postia placenta Involved in Degradation of Lignocellulosic Biomass Revealed from Metadata Analysis of Open Access Gene Expression Data. **International journal of biological sciences** 2018; 14(3):237-252. doi:10.7150/ijbs.22868.
7. **Ayyappa Kumar Sista Kameshwar**, and Wensheng Qin. "Comparative study of genome-wide plant biomass-degrading CAZymes in white rot, brown rot and soft rot fungi." **Mycology** (2017): 1-13.
8. **Ayyappa Kumar Sista Kameshwar**, and Wensheng Qin. "Analyzing Phanerochaete chrysosporium gene expression patterns controlling the molecular fate of lignocellulose degrading enzymes." **Process Biochemistry** 64 (2018): 51-62.

9. **Ayyappa Kumar Sista Kameshwar**, Richard Barber, and Wensheng Qin. "Comparative modeling and molecular docking analysis of white, brown and soft rot fungal laccases using lignin model compounds for understanding the structural and functional properties of laccases." **Journal of Molecular Graphics and Modelling** 79 (2018): 15-26.
10. **Ayyappa Kumar Sista Kameshwar**, and Wensheng Qin. "Gene expression metadata analysis reveals molecular mechanisms employed by Phanerochaete chrysosporium during lignin degradation and detoxification of plant extractives." **Current Genetics** (2017): 1-18.
11. **Ayyappa Kumar Sista Kameshwar**, Wensheng Qin, "Metadata Analysis of Phanerochaete chrysosporium Gene Expression Data Identified Common CAZymes Encoding Gene Expression Profiles Involved in Cellulose and Hemicellulose Degradation." **International journal of biological sciences** 13, no. 1 (2017): 85-99.
12. Xiong, Lili, **Ayyappa Kumar Sista Kameshwar**, Xi Chen, Zhiyun Guo, Canquan Mao, Sanfeng Chen, and Wensheng Qin. "The ACEII recombinant Trichoderma reesei QM9414 strains with enhanced xylanase production and its applications in production of xylitol from tree barks." **Microbial Cell Factories** 15, no. 1 (2016): 215.
13. **Ayyappa Kumar Sista Kameshwar**, and Wensheng Qin. "Qualitative and Quantitative Methods for Isolation and Characterization of Lignin-Modifying Enzymes Secreted by Microorganisms." **BioEnergy Research** (2016): 1-19.
14. **Ayyappa Kumar Sista Kameshwar**, and Wensheng Qin. "Lignin Degrading Fungal Enzymes." In **Production of Biofuels and Chemicals from Lignin**, pp. 81-130. Springer Singapore, 2016.
15. **Ayyappa Kumar Sista Kameshwar**, and W. Qin. "Recent Developments in Using Advanced Sequencing Technologies for the Genomic Studies of Lignin and Cellulose Degrading Microorganisms." **International journal of biological sciences** 12, no. 2 (2015): 156-171.