Origin of fungal biomass degrading enzymes: Evolution, diversity and function of enzymes of early lineage fungi

Lange, Lene; Pilgaard, Bo; Herbst, Florian-Alexander; Busk, Peter Kamp; Gleason, Frank; Pedersen, Anders Gorm

Published in:
Fungal Biology Reviews

Link to article, DOI:
10.1016/j.fbr.2018.09.001

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Review

Origin of fungal biomass degrading enzymes: Evolution, diversity and function of enzymes of early lineage fungi

Lene Langea,*, Bo Pilgaardb, Florian-Alexander Herbstc, Peter Kamp Buskd, Frank GLEASONe, Anders Gorm PEDERSENF

aCenter for Bioprocess Engineering, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Kgs. Lyngby, 2800, Denmark
bDepartment of Biotechnology and Biomedicine, Technical University of Denmark, Kgs. Lyngby, 2800, Denmark
cDepartment of Chemistry and Bioscience Aalborg University, Denmark
dDepartment of Science and Environment, Roskilde University, Denmark
eUniversity of Sydney, Sydney, NSW, 2006, Australia
fDepartment of Bio and Health Informatics, Technical University of Denmark, Kgs. Lyngby, 2800, Denmark

Abstract

The aim of this study was to elucidate the evolution of enzyme secretome of early lineage fungi to contribute to resolving the basal part of Fungal Kingdom and pave the way for industrial evaluation of their unique enzymes. By combining results of advanced sequence analysis with secretome mass spectrometry and phylogenetic trees, we provide evidence for that plant cell wall degrading enzymes of higher fungi share a common ancestor with enzymes from aerobic ancient fungi. Sequence analysis (HotPep, confirmed by dbCAN-HMM models) enabled prediction of enzyme function directly from sequence. For the first time, oxidative enzymes are described here in early lineage fungi (Chytridiomycota & Cryptomycota), which supports the conceptually new understanding that fungal LPMOs were also present in the early evolution of the Fungal Kingdom. Phylogenetic analysis of fungal AA9 proteins suggests an LPMO-common-ancestor with Ascomycetes and Basidiomycetes and describes a new clade of AA9s. We identified two very strong biomass degraders, Rhizophlyctis rosea (soil-inhabiting) and Neocallimastix californiae (rumen), with a rich spectrum of cellulolytic, xylanolytic and pectinolytic enzymes, characteristically including several different enzymes with the same function. Their secretome composition suggests horizontal gene transfer was involved in transition to terrestrial and rumen habitats. Methods developed for recombinant production and protein characterization of enzymes from zoosporic fungi pave the way for biotechnological exploitation of unique enzymes from early lineage fungi with potential to contribute to improved biomass
1. Introduction

The aim of this study was to elucidate the enzyme secretome composition of zoosporic early lineage fungi, across all phyla (James et al., 2006b; Kämper et al., 2006; Powell and Letcher, 2014); to contribute to resolving the basal part of the fungal tree of life, and to pave the way for industrial use of their unique enzymes by developing methods for cloning, expression and protein characterization of enzymes from such fungi.

Zoosporic fungi have been described as having widespread ecological importance for global plant biomass conversion due to their role in recycling global carbon (Chang et al., 2015; Gleason et al., 2017; Jephcott et al., 2016). Yet early lineage fungi remain underexplored, especially at the molecular level. While evolution and phylogeny of the higher Dikarya fungi have been elucidated (Ruggiero et al., 2015) the basal lineages of the Fungal Tree of Life remain largely unresolved. Contributions to resolving its evolution will have to come from advanced sequence analysis and phylogeny (Chang et al., 2015; Taylor and Berbee, 2006).

James et al. (2006a) advanced understanding of the early evolution of fungi by the use of a six-gene phylogeny that enabled identification of the last common ancestor of crown group fungi. James et al. (2006b) described Blastocladiomycota as a new phylum. Hibbett et al., (2007) assembled a higher level classification of fungi, including the early lineages. They retained the phylum Chytridiomycota and proposed a new phylum, Neocallimastigomycota. Among the early lineage zoosporic fungi, Olpidium and Rozella long remained unclassified (Hibbett et al., 2007). Jones et al., (2011) described Cryptomycota as a new phylum, and James et al., (2013) analyzed the genome of Rozella allomycis - the first species of this new phylum to be genome sequenced. Based on morphological and functional characters, zoospore ultrastructure and DNA-based molecular methods Powell and Letcher, (2012) recognized the following phyla: Chytridiomycota, Monoblepharidomycota and Neocallimastigomycota. However, Ruggiero et al., (2015) grouped all zoosporic fungi into one phylum, and ranked the anaerobic rumen fungi only as a class. The paper by Chang et al., (2015) marked a breakthrough with regard to including enzymes in descriptions of classification and evolution of the early lineage fungi. Using expansion of pectinase genes and the phylogeny of the organism, Chang et al. (l.c.) estimated the genomic time clock of the plant-–fungus association. Such advanced studies were taken further by the work of Spatafora and Berbee research groups (Berbee et al., 2017; Spatafora et al., 2016).

In the current study, we focus on the molecular, functional and structural diversity of enzymes from early lineage fungi as key characteristics for their biology, ecology and evolution. The composition of the fungal secretome metabolizes substrates to provide nutrients, and is therefore a determining factor for competitiveness of the fungus, thus vital for growth and reproduction. However, the findings of Krijger et al. (2014) support the observation that the fungal secretome is influenced more by phylogeny than lifestyle.

The fast track methodological approach, used in the current study to elucidate the CAZy enzyme profiles (Cantarel et al., 2009) builds on a combined use of the non-alignment based HotPep (Homology-to-Peptides) sequence analysis program (Busk et al., 2017) and a dbCAN analysis (Huang et al., 2018; Yin et al., 2012). The basis for such analysis is the well-curated CAZy database, which groups enzymes into enzyme families (and subfamilies). Furthermore, HotPep can capture protein features of importance for enzyme function; mass spectrometry studies identify presence of proteins in the secretome; and HotPep-mediated prediction of function (going beyond annotation to enzyme family) enables biological and evolutionary interpretations across organismal habitats and physiological and taxonomic groupings.

2. Materials and methodologies for comparative secretome studies

Rhizophlyctis rosea (de Bary and Woronin) Fischer, NRBC 105426, was isolated from soil collected in Sydney University garden using filter paper as bait (Cleason et al., 2004) and deposited in the NITE Biological Resource Center (see Fig. 1, Supplementary Fig. A.1.). The medium used for inducing production of plant cell wall degrading enzymes consisted of 1 % Avicel (Sigma), 0.5 % CMC (Sigma), 0.5 % wheat branbroth (Finax, Denmark), 0.5 % yeast extract (Sigma), 0.1 % mycological peptone (Sigma) and 0.05 % glucose (Sigma). For full description of methods, see Supplementary Text A.1.

Eleven different AZCL substrates (Megazyme, Ireland) were prepared using Britten Robinson buffer. The plates were incubated at 30 °C overnight and the radius of the zones measured. Endoglucanase activity was measured as described earlier (Busk and Lange, 2013a).

DNA was extracted as described (Busk and Pages, 2002) and sequenced twice on an Illumina Hiseq 2000 in one multiplexed lane as Paired-End libraries using Truseq chemistry. The genome coverage of the two PE-libraries was estimated to 140x and 279x. Before assembly, the raw reads were assessed with FastQC v0.10.1 (Andrews, 2010) and processed according to (Lindgreen, 2012), (Martin, 2011) and (Kelley et al., 2010) and assembled with Velvet 1.2.10 (Zerbino and Birney, 2008) The genome was submitted to Genbank (https://www.ncbi.nlm.nih.gov/genbank/) and given the accession number GCA002214945.1. For GenBank submissions, see Supplementary Tab. A.1.
All annotations were performed on whole genome predicted protein sequences. For *R. rosea* and *H. polyrhiza* AUGUSTUS 2.7 (Stanke et al., 2008) was utilized with *R. oryzae* and *B. dendrobatitis* as model organisms for gene predictions from the whole genome. PPR (Busk and Lange, 2013b) based HotPep (Busk et al., 2017) and dbCAN (Huang et al., 2018) based HMMer3 (Eddy, 2011) were used for annotation and functional prediction of carbohydrate active enzymes. Only hits confirmed by both approaches were used for genome wide comparisons. All protein sequences submitted to Genbank from *R. rosea* (Supplementary Tab. A.1) were manually inspected. Signal peptides were predicted using the SignalP 4.1 (Petersen et al., 2011) and Phobius (Käll et al., 2007) servers.

Total RNA was extracted and purified using the RNeasy Plant Mini Kit (Qiagen, Germany). cDNA was synthesized with oligodT random primer and ImProm II reverse transcriptase (Promega, USA) as described (Madsen et al., 2009). The synthesized cDNA was used as template for the subsequent PCRs (Tab. A2).

Protein expression was performed using the PichiaPink™ Expression System protocol (Invitrogen, USA) (see Supplementary Fig. A.2).

Preparing for LC-MS/MS: The total protein content from the triplicate supernatants were precipitated by trichloroacetic acid (Sigma, Germany) overnight at –20 °C. The protein pellet was solubilized in digestion buffer (1% sodium deoxycholate, 50 mM triethylammonium bicarbonate, pH 8.0) and subsequently reduced, alkylated, in-solution digested, desalted and measured by LC-MS/MS as previously described (Huang et al., 2015). Protein identification was done using the open-source software MaxQuant v1.5.2.8 (Tyanova et al., 2016).

PCA analysis was performed and visualized in Rstudio 1.1.383 (RStudio Team, 2016) using R version 3.4.3 (R core team, 2017) and the factoextra package (Kassambara and Mundt, 2017) with the scaled family annotations for all the genomes as dataset.

For construction of phylogenetic trees, publically available protein sequences were obtained from Genbank. The catalytic domains of these sequences used for phylogeny were defined using the NCBI CDD batch search (Marchler-Bauer et al., 2015). For full list of accession numbers, see Supplementary Tab. A.3. Multiple alignments were performed with MAFFT version 7 (Katoh and Toh, 2008) using the E-INS-i method. Phylogenies were reconstructed with Bayesian inference using MrBayes v3.2.4 that averaged over 10 aa substitution models and assumed gamma-distributed rates over sites. The runs continued until convergence was reached across 2 sets of 3 chains: “R-hat” values for parameters near 1 and standard deviation of split frequencies below 0.01. The trees were visualized with Fig-Tree v1.4.2 (http://tree.bio.ed.ac.uk/software/Figtree/).

3. Results

The results from re-analyzing published genome data, put in perspective by new information, are grouped into four parts, 3.1–3.4.

3.1. *R. rosea* as a model species for studying enzyme secretome of early lineage fungi

The soil inhabiting chytrid, *R. rosea* (Chytridiomycota), was chosen for studying plant biomass degrading enzymes among the early lineage zoosporic fungi. The aim was to develop methods for cloning, expression and characterization of enzymes, to facilitate investigation of evolution of biotroph and saprotroph roles of such fungi in nature, and at the same time to pave the way for industrial use of the enzymes from this hitherto unexploited group of microorganisms.

The *R. rosea* isolate NRBC 105426 (Gleason et al., 2004; Pilgaard, 2014) was selected based on its activity profile characterized by testing the culture broth when grown under inducing conditions (see Table 1). The light micrographs (Fig. 1) show the morphology of the *R. rosea* isolate studied when grown in lab-culture. Genome sequencing of *R. rosea* was performed as paired-end sequencing on an Illumina Hiseq 2000. The draft genome assembly yielded 4476 contigs with sizes ranging from 200 b to 309 kb and a median size of 12.4 kb; total length of assembly approx. 47 Mb.
The single copy gene GH45 (predicted by HotPep analysis to have endo-1,4-β-D-glucanase, EC 3.2.1.4 function) was selected for cloning. It was recombinantly produced in Pichia pastoris. The predicted gene RrCel45 was 104 bp and 897 bp long with and without two introns (lengths confirmed by sequencing PCR products of the cDNA and gDNA). A CDD search confirmed the GH45 domain and revealed a CBM1 in N-terminal position. RrCel45 was successfully expressed in P. pastoris. MALDI-TOF analysis of the SDS-band confirmed the identity of the protein; and the purity of the protein was >95% (see Fig. A2). Two more enzymes from R. rosea (β-xylosidase, GH43, and endo-xylanase, GH11) were successfully expressed in P. pastoris (to be reported separately by Huang, Lange et al.).

The RrCel45 activity was tested on azurine crosslinked substrates. This revealed high activity towards carboxymethyl cellulose and less activity towards α-cellulose and β-glucan. No activity was found on crystalline cellulose or xylan. Optimum pH was pH5 at 35 °C; and optimum temperature was 35 °C at pH5 (Fig. 2). Thermostability analysis of RrCel45 revealed high thermostability, retaining approx. 40% residual activity after incubation at 100 °C for 1 h. The specific activity of RrCel45 was estimated to be low (0.5 U/mg of protein).

3.2. Elucidating enzyme secreteme composition of the earliest branches of the fungal tree of life

One thousand two hundred and twenty-eight proteins were found by aligning MS data (from R. rosea culture broth) to the genome predicted proteins. Of these, 124 proteins were recognized by HotPep/dbCAN analysis to belong to four CAZy enzyme categories, Auxiliary Activity (AA), Polysaccharide Lyases (PL), Carbohydrate Esterases (CE) and Glycohydrolases (GH). Domain searches resulted in 46 additional proteins recognized as putative carbohydrate active enzymes. Supplementary Tab. A.4 gives a comparison of carbohydrate active enzymes of genome and secreteme of R. rosea. Table 2 lists the top 20% of the most abundant carbohydrate active enzymes in R. rosea secreteme predicted by HotPep/dbCAN analysis. The 20% top scoring enzymes with regard to relative secreteme abundance were GH6, GH7, GH10 and GH11. Abundance was calculated as a relative value based on the total amount of predicted carbohydrate active enzymes in the MS analysis. The MS-based estimate of enzyme protein abundance, summed up for each type of function, is presented in Table 3. These data show that the three most dominating functions found in the secreteme of R. rosea, ranked from one to three, were endo 1,4 β xylanase (EC 3.2.1.8), LPMO AA9, and endo 1,4 β glucanase (EC 3.2.1.4). The high score of AA9 is found by summing up the 20 AA9 proteins found by MS analysis of the secreteme.

Available sequencing data from 11 published zoosporic early lineage species (Supplementary Tab. A.5) (Chang et al., 2015; Grigoriev et al., 2012; James et al., 2013; Joneson et al., 2011; Nordberg et al., 2014; Russ et al., 2016; Youssef et al., 2013) and the de novo genome sequencing data of R. rosea, were analyzed for carbohydrate active enzymes by HotPep/dbCAN. Only the hits confirmed by both HotPep and dbCAN are included in this study; for statistics of annotation by HotPep as compared to dbCAN, see Supplementary Tab. A.6. The 12 genomes represent all four fungal zoosporic early lineage phyla: Chytridiomycota, Neocallimastigomycota, Blastocladiomycota and Cryptomycota; see taxonomic overview, Supplementary Tab. A.7 (for taxonomic affiliation of R. rosea, see (Letcher et al., 2008) and for R. globosum see (Letcher et al., 2006)). A list of annotated CAZy hits, assigned to protein family based on HotPep/dbCAN is shown in Supplementary Tab. A.8.

The two strongest biomass degraders were R. rosea and the anaerobic cow rumen fungus, Neocallimastix carbonaria. Both HotPep and dbCAN annotated these two hotspots to have a high number of predicted CAZy genes (up to 225 and 667, respectively; Supplementary Tab. A.6; and Tables 4a,b and 5a,b; and to have several genes encoding enzymes with the same e.g. cellulolytic or xylanolytic activity (Busk et al., 2014). Gonapodyda prolifera also had a broad spectrum of CAZy enzymes, however, this species was richest not in

<table>
<thead>
<tr>
<th>Substrate</th>
<th>β-glucan</th>
<th>HE-Cellulose</th>
<th>Galacto-mannan</th>
<th>Galactan (potato)</th>
<th>Xyloglucan (tamarind)</th>
<th>Xylan (birch)</th>
<th>Arabino-xylan</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMC</td>
<td>56</td>
<td>50</td>
<td>33</td>
<td>28</td>
<td>50</td>
<td>50</td>
<td>129</td>
</tr>
<tr>
<td>Avicel</td>
<td>73</td>
<td>50</td>
<td>25</td>
<td>13</td>
<td>50</td>
<td>50</td>
<td>121</td>
</tr>
<tr>
<td>Mixed-medium (MS)</td>
<td>154</td>
<td>28</td>
<td>79</td>
<td>13</td>
<td>79</td>
<td>113</td>
<td>113</td>
</tr>
</tbody>
</table>

The activity profile (scored by measuring the blue halo clearing zones) was estimated by semi-quantitative AZCL-plate assays. Units calculated in mm².

**Fig. 2** — Characterization of optimum temperature and pH indicate a typical fungal enzyme profile. However, temperature stability was exceptionally high: Stable up to 70°C; 40% residual activity of intact protein, active at 100°C.
cellulolytic and xylanolytic enzymes but in pectinolytic enzymes. The three specialized parasitic species, B. dendrobatidis, C. anguillulae, and R. allomycis, had only few carbohydrate active enzymes in their genomes (57, 35 and 27, respectively; Supplementary Tab. A.8).

Overview of early lineage fungi cellulolytic, xylanolytic and pectinolytic enzyme capacity is shown in Tables 4a,b and 5a,b; (for detailed description, see Supplementary Text A.2 and Tab. A.8): β-glucan EC 3.2.1.6 active enzymes were predicted to be found only in R. rosea and S. punctatus (GH16). Enzymes with EC 3.2.1.91 function were found in R. rosea, R. globosum and R. allomycis. The defense related enzyme, lysozyme (EC 3.2.1.17) was found in R. globosum and R. allomycis (both belonging to GH19). Neocallimastix californiae also had a lysozyme but in this anaerobic rumen fungus it was a lysozyme belonging to GH73.

Table 2 – Top 20 % of the most abundant carbohydrate active enzyme proteins of the R. rosea secretome.

<table>
<thead>
<tr>
<th>Cazy family</th>
<th>Predicted function by EC</th>
<th>Abundance %</th>
<th>Trans-membrane domain</th>
<th>Signal Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>GH7</td>
<td>3.2.1.4</td>
<td>9.08</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>GH10</td>
<td>3.2.1.8</td>
<td>6.86</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>GH6</td>
<td>3.2.1.91</td>
<td>6.65</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>GH7</td>
<td>3.2.1.176</td>
<td>6.57</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>GH11</td>
<td>3.2.1.8</td>
<td>5.10</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>CE1</td>
<td>3.1.1.72</td>
<td>3.47</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>CE4</td>
<td>3.5.1.41</td>
<td>3.10</td>
<td>1</td>
<td>Y</td>
</tr>
<tr>
<td>AA9</td>
<td>0</td>
<td>3.10</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>CE4</td>
<td>3.5.1.41</td>
<td>2.79</td>
<td>1</td>
<td>Y</td>
</tr>
<tr>
<td>AA9</td>
<td>0</td>
<td>2.52</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>GH10</td>
<td>3.2.1.8</td>
<td>2.01</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>GH26</td>
<td>0</td>
<td>1.94</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>GH10</td>
<td>3.2.1.8</td>
<td>1.89</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>AA8</td>
<td>1.1.99.18</td>
<td>1.77</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>AA9</td>
<td>0</td>
<td>1.68</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>CE4</td>
<td>3.5.1.41</td>
<td>1.60</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>GH43</td>
<td>3.2.1.37</td>
<td>1.28</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>GH10</td>
<td>3.2.1.8</td>
<td>0.98</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>GH3</td>
<td>3.2.1.21</td>
<td>0.92</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>GH26</td>
<td>0</td>
<td>0.79</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>GH35</td>
<td>3.2.1.23</td>
<td>0.79</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>AA6</td>
<td>0</td>
<td>0.74</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>GH5</td>
<td>3.2.1.4</td>
<td>0.69</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>CE1</td>
<td>3.1.1.72</td>
<td>0.59</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

The abundance value is a relative value based on the total amount of predicted carbohydrate active enzyme proteins in the MS analysis. The top 20 % enzymes listed are mainly involved in cellulolytic (EC 3.2.1.4, 3.2.1.91, 3.2.1.176, 3.2.1.21, AA9, AA8) and xylanolytic (EC 3.2.1.8, 3.2.1.37, 3.1.1.72) degradation.

Table 3 – MS-based estimate of enzyme protein abundance, summed up for each type of function.

<table>
<thead>
<tr>
<th>Function</th>
<th>endo-1,4-β-xylanase</th>
<th>AA9 LPMO</th>
<th>Endo-glucanase</th>
<th>Acetylxylan esterase</th>
<th>1,4-β-cellobiosidase</th>
<th>1,4-β-cellobiosidase</th>
<th>β-glucosidase</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC #</td>
<td>3.2.1.8</td>
<td>-</td>
<td>3.2.1.4</td>
<td>3.1.1.72</td>
<td>3.2.1.91</td>
<td>3.2.1.176</td>
<td>3.2.1.21</td>
</tr>
<tr>
<td>Abundance %</td>
<td>17.9</td>
<td>17.8</td>
<td>10.7</td>
<td>7.2</td>
<td>6.7</td>
<td>6.8</td>
<td>3.1</td>
</tr>
</tbody>
</table>

Scoring highest, endo 1,4 β-xylanase (EC 3.2.1.8); scoring second highest LPMO AA9; scoring third, endo 1,4 β glucanase (EC 3.2.1.4). The relative abundance values shown represent abundancies of enzyme proteins found to be secreted in the R. rosea experimental studies.

Rhizophlyctis rosea, H. polyrhiza and the three rumen fungi had only the first three of these; B. dendrobatidis only the first two. Both species of Blastocladiomycota also had these first two functions but also a GH2 of the 3.2.1.52 function. Gonapodya prolifera had only one chitinase (viz. EC 3.2.1.14), but this function was represented by two different types of protein families, GH18 and GH19. The Cryptomycota, R. allomycis was found only to have one GH18 chitinase (EC 3.2.1.14). The defense related enzyme, lysozyme (EC 3.2.1.17) was found in R. globosum and R. allomycis (both belonging to GH19). Neocallimastix californiae also had a lysozyme but in this anaerobic rumen fungus it was a lysozyme belonging to GH73.

Unique CAZyme genes within the secretome of the five fungal phyla were primarily found in Neocallimastigomycota, R. rosea and G. prolifera (Chytridiomycota). As can be seen (Tables 4a,b and 5a,b), conserved patterns of similar enzyme protein family for similar enzyme function were found among the aerobic Chytridiomycota, Blastocladiomycota and Cryptomycota. The anaerobic rumen fungi had many unique
enzymes: GH 36 \( \alpha \)-galactosidase (EC 3.2.1.22); GH 78 \( \alpha \)-L-rhamnosidase (EC 3.2.1.40); GH48 1,4\( \beta \)-cellobiohydrolase, reducing-end active; GH26 1, 4\( \beta \)-endoglucanase (EC 3.2.1.4); CE6 acetyl-xyylan esterase (EC 3.1.1.72); GH8 endo 1,4\( \beta \)-xylosidase (EC 3.2.1.8); GH32 \( \beta \)-fructofuranosidase (EC 3.2.1.26); and GH8 chitosanase (EC 3.2.1.132). Interestingly, monoblepharidomycete species, \( G. \) prolifera, also had unique enzymes, namely GH28 with galacturan 1,4\( \alpha \)-galacturonidase activity (EC 3.2.1.67) and GH62 \( \alpha \)-N-arabinofuranosidase activity (EC 3.2.1.55), notably the latter not found in any other species studied.

A PCA analysis was made of all carbohydrate active enzyme hits. As shown in Supplementary Figs A.3-4, the \( R. \) rosea secretome occupied a truly outlier position as did the secretome of \( N. \) californiae, though to a lesser degree. The very significant composition of these two enzymatically strongest biomass degraders dominated the PCA plot to the extent that the relatedness and clustering of the groupings of the 12 species remained unresolved by PCA.

### 3.3. LPMO AA9 and AA11 are found also in early lineage fungi

We report here for the first time that lytic polysaccharide monooxygenases, LPMOs, are present in early lineage fungi Chytridiomycota (AA9 in \( R. \) rosea and \( G. \) prolifera), and that AA11 is found in Cryptomycota (\( R. \) allomycis). \( R. \) rosea has a high number (24) of AA9 in its genome; and \( G. \) prolifera was found to have two AA9 (see Table 4a,b). Further, wild type, homolog expression and secretion of AA9 proteins were confirmed by MS for \( R. \) rosea, thus experimentally demonstrating the presence of LPMO proteins in \( R. \) rosea culture broth. Notably, AA9 was documented by MS to be among the most abundant proteins in the secretome of \( R. \) rosea (Tables 2 and 3). These first records of LPMOs in early lineage fungi were made by using the bioinformatics approach described by Busk and Lange, (2015).

Table 6 gives an overview of the more detailed information about the LPMO, AA9 and AA11 found in the zoosporic early lineage fungi \( R. \) rosea, \( G. \) prolifera and \( R. \) allomycis. The functional overview of LPMO diversity was adapted from Busk and Lange (2015). As Table 6 shows, the LPMO genes found represent a rich diversity, illustrated by their distribution over a wide spectrum of different LPMO PPR groups, including both C1 and C4 oxidizing types. Further, the variation in CBM1 domain and signal peptide occurrences is also apparent from Table 6. No LPMO were found in the genomes of any of the three rumen fungi studied; and no LPMO AA11 was found in the skin-invading frog pathogen, \( B. \) dendrobatidis.

Fig. 3 presents a (proportional) radial phylogram tree of fungal AA9. The tree includes a broad selection of fungal, ascomycetous and basidiomycetous AA9 sequences, along with the 20 (MS confirmed) new AA9 LPMO sequences from \( R. \) rosea. The tree consists of two all-fungal clades (i.e. including all three Asco-, Basidio- and Chrytridiomycota AA9); one clade with Asco- and Basidiomycota AA9, but no \( R. \) rosea AA9; a cluster of highly diverse, unique, early branching Chytridiomycota AA9 can also be seen (to the right). The resolution of this part of the tree is improved in the presentation of the same data given in Supplementary Fig. A.5 (here including boot-strap values).
3.4. Constructing phylogenetic trees for selected enzyme proteins of early lineage fungi

Phylogenetic trees of homolog proteins from all Biological Kingdoms are shown in Figs 3, 4 and 6; and in Supplementary Figs A.5, A.6. The trees visualize the global diversity and relatedness of the enzymes from the zoosporic early lineage fungi to enzymes derived from other types of organisms (plants, animals, higher fungi, bacteria, and archaea).

In the phylogenetic tree of GH5 (EC3.2.1.4) presented in Fig. 4, sequences of *R. rosea* are embedded in fungal clades of the upper half of the tree, which also contain ascomycetous...
basidiomycetous and zygomycetous GH5. However, in contrast, the GH5 from the anaerobic species studied are found embedded in a clade dominated by bacterial proteins, originating predominantly from the cow rumen habitat. Besides these two major occurrences of early lineage fungal GH5, one copy of *R. rosea* GH5 is found in a predominantly bacterial GH5 clade together with one GH5 protein from a zygomycetous species (*Umbelopsis ramanniana*) and an Archaea (*Pyrococcus abyssi*) GH5 protein. No GH5 with predicted EC 3.2.1.4 endogluccanase activity was found in *G. prolifera* genome.

The GH5 phylogenetic tree in Fig. 4 and the sub-tree in Supplementary Fig. A.6, based on same data, both illustrate two possible horizontal transfer (HGT) events. The data provide strong evidence for HGT events between rumen bacteria (*Fibrobacter succinogenes*) and the rumen zoosporic fungus *O.* *joyonii*. There is likewise strong evidence of HGT between rumen bacterium *Fibrobacter succinogenes* and rumen protist *Epidinium ecaudatum*.

To investigate further the difference between the *R. rosea* GH5 copies found in the fungal clade and one of the sequences found embedded in the bacteria-dominated clade, a comparison was made of their gene structure (see Fig. 5). The bacterial related *R. rosea* GH5 gene had only one intron and no CBM. By contrast, the fungal related *R. rosea* GH5 had a CBM1 domain and up to many introns and additional gene domain inserts.

The phylogenetic tree of GH13 (EC 3.2.1.1), shown in Fig. 6 in two versions, illustrates the extremely broad diversity of enzymes derived from the different phyla of early lineage fungi: GH13 from early lineage fungi are found in most parts of the radial tree. GH13 of Neocallimastigomycota and Blastocladiomycota are embedded in two different bacterial clades; while GH13 of *R. rosea* are found on a separate, very basal part of an all-fungal clade. However, in the opposite part of the tree, embedded in a bacterial clade is another GH13 gene variant of *R. rosea*. Adjacent to this is GH13 of the Chytrid, *S. punctatus*. *G. prolifera* GH13 is found separately.

### 4. Discussion

The investigations reported here of secreted carbohydrate active enzymes (CAZy) from early lineage fungi provide strong evidence that fungal (Asco- and Basidiomycetes) plant cell wall degrading enzymes share a common ancestor with enzymes from aerobic zoosporic early lineage fungi. Phylogenetic analysis revealed that the major part of *R. rosea* and *G. prolifera* plant cell wall degrading enzymes (Chang et al., 2015) is embedded in or placed in clades closely associated to the clades of homolog enzymes from crown fungi (i.e. ascomycetous and basidiomycetous cellulolytic, xylanolytic and pectinolytic enzymes). However, we have also shown that the CAZy secretomes have developed very differently in the four phyla. The phylogenetic studies presented here show that plant biomass degrading enzymes from anaerobic rumen fungi are embedded in or placed in clades close to rumen bacteria (data shown for GH5 and GH13). This support the notion of the anaerobic Neocallimastigomycota being only more distantly related to the rest of the early lineage fungi. The phylogenetic studies presented here show that plant biomass degrading enzymes from anaerobic rumen fungi are embedded in or placed in clades close to rumen bacteria (data shown for GH5 and GH13). This support the notion of the anaerobic Neocallimastigomycota being only more distantly related to the rest of the early lineage fungi. The results here reported suggest the common ancestor of CAZy cell wall degrading enzymes of rumen fungi may most likely have been of bacterial origin, with the genes initially acquired through HGT in the aqueous rumen habitat. Such dynamic
interaction between rumen bacteria and the zoosporic rumen fungi may be an ongoing process. HGT is assumed to function in parallel with the fungal evolutionary mechanism of gene copy/gene loss (Braun et al., 2000; Wapinski et al., 2007). The current comparative study of the phylogenetic variation among the homolog variants of GHS (EC 3.2.1.4) shows that most of R. rosea GH5 proteins are placed within all-fungal clades (Fig. 4). However, one of the GHS copies is embedded in an all-bacterial clade, and the gene structure of this gene is closer to the structure of a bacterial than a fungal clade. This is an unexpected observation, as the current comparative study supports the idea that the rumen fungi are clearly related to the other fungal phyla. Comparison between the two primary carbohydrate metabolizing Chytridiomycota included in this study, the terrestrial R. rosea and the aquatic G. prolifera, shows that the full spectrum of cellulosytic and xylanolytic plant cell wall degrading enzymes was found only in R. rosea. Comparison between only one aquatic and one terrestrial species is however not sufficient for conclusions to be drawn about enzyme secretome evolution. More plant biomass metabolizing, zoosporic fungi, both terrestrial and aquatic (e.g. belonging to Cladochytriales (Mozeley-Standridge et al., 2009), must be genome sequenced and functionally analyzed before any hypothesis can be formulated. Such broader studies would be highly interesting as it could provide insight in how the key cellulosytic and xylanolytic plant cell wall degrading enzymes evolved: Whether such enzymes were developed and maintained through speciation; and/or acquired through HGT (from which types of organisms?) in the transition from aquatic to terrestrial plant biomass degrading lifestyle. By contrast, a significant part of the mannan and starch degrading enzymes, and fungal cell wall modifying chitinases, were shared across all four early lineage fungal phyla investigated in this study (Table 5a,b); notably for these types of enzymes the rumen fungi are clearly related to the other fungal phyla. The data support a similar interpretation of the occurrence of the fungal carbohydrate binding domain CBM1, which is known to be an add-on domain of many CAZy enzymes in higher fungi (http://www.cazy.org/, n. d.). CBM1 was also observed to be present widely in enzymes of early lineage fungi (Howlett et al., 2008). The current study includes observations of specific HGT events (see Fig. 4) which together with several other observations of comparative genome sequence analysis (Tables 4 and 5) suggest that HGT plays a prominent role in the evolution of the metabolic enzyme secretome of early lineage fungi. The
phylogenetic trees of selected proteins provide basis for suggesting that HGT events have also taken place between other types of organisms; for example, HGT of GHS from *Fibrobacter succinogenes* to the rumen fungus *O. joyonii* as well as to the dinoflagellate, *Epidinium*, and HGT of GHS from rumen bacteria to the rumen fungi, *Neocallimastix*, *Orpinomyces* and *Piromyces*. Evolution of early lineage fungi from an aquatic environment to a terrestrial lifestyle and metabolizing lignocellulosic plant cell wall materials may have involved and required disruptive evolutionary mechanisms such as HGT.

Based on the above we conclude that occurrence of HGT is likely to be more prominent among aquatic organisms. Interestingly, a series of studies on the role of HGT in another group of aquatic organisms (the Oomycetes) has been published recently (Richards et al., 2011; Savory et al., 2015); such studies support our hypothesis that HGT occur more frequently among aquatic microbial life forms.

In the present study we document for the first time that oxidative LPMO enzymes are present in the early lineage fungi and not only in the higher Dikarya fungi, the ascomycetes and basidiomycetes (Beeson et al., 2012; Busk and Lange, 2015). LPMO/AA9 was found in both the chytridiomycetes (*R. rosea*) and the monoblepharidiomycetes (*G. prolifera*) of the Chytridiomycota. Further, AA11 (Hemsworth et al., 2014) was found in *R. allomycis*, positioned at the most basal part of the fungal evolutionary tree, the Cryptomycota. Interestingly, *R. rosea* was predicted to have as many as 24 different AA9 proteins in its genome. With this result we show that LPMOs in early lineage fungi also share the pattern, typically observed in higher fungi, that strong lignocellulosic biomass degraders in particular have several (or even many) monooxygenase genes (Busk and Lange, 2015). The sequence diversity of AA9 found in *R. rosea* is striking, illustrated by the fact that the 20 AA9 enzyme proteins found by MS analysis to be present in

---

**Fig. 3** — A radial phylogenetic tree of fungal AA9 LPMOs, including a broad selection of fungal, Ascomycota and Basidiomycota, AA9 LPMO sequences, along with the 20 new AA9 LPMO sequences found expressed in this study of wild type *R. rosea*. The tree has five major clades: three all-fungal clades (including asco-basidio- and Chytridiomycota AA9); one clade with AA9 from only Ascomycota and Basidiomycota. And pointing to the right, a cluster of seven, early branches of *R. rosea* AA9s; these seven closely clustered branches are unique for *R. rosea* and as appears only distantly related to the Ascomycota and Basidiomycota LPMO AA9s. Improved resolution of this part of the tree is given in Fig. S4, where also bootstrap values are included. Color code: green, Ascomycota AA9; orange, Basidiomycota AA9; purple, Chytridiomycota AA9.
the secretome belong to 10 different AA9 PPR-groups. Notably, the two AA9 predicted proteins in \textit{G. prolifera} belong to two different PPR-groups, neither of which were found in \textit{R. rosea}; furthermore both of the two AA9 of \textit{G. prolifera} have a carbohydrate binding domain, CBM1, while only 8 of the 20 AA9 in \textit{R. rosea} have CBM1. Interestingly, none of the AA11 found in \textit{R. allomycis} in this study had CBM domains, but one of the AA11 was predicted (Table 6) to have a chitin-binding domain, CBM19 (see also (Eijsink et al., 2010; Várnai et al., 2014)). In summary, we conclude that the phylogenetic studies of AA9 proteins of \textit{R. rosea} and \textit{G. prolifera} (Fig. 3) suggest a common ancestor with AA9 found in ascomycetes and basidiomycetes. A shared common ancestor also among all fungal AA11 is presented only as a hypothesis due to insufficient data. Notably, no AA11 LPMO was found in the genome of the frog pathogen, \textit{B. dendrobatidis}, in contrast to fungal dermatophytes and \textit{Onygena coruina} (Busk and Lange, 2015; Lange et al., 2016).

The \textit{R. rosea} experimental studies together with bioinformatics and MS analysis of protein abundance in the culture broth gave another surprising result. The LPMO enzyme proteins constitute a very significant proportion of the entire \textit{R. rosea} enzyme secretome: LPMO AA9 proteins made up $\approx 18\%$ of all carbohydrate active enzyme protein found.

The discovery of LPMOs from two different phyla of early lineage fungi was made by the use of HotPep/dbCAN analysis. HotPep identification of conserved patterns of peptides enabled a range of new discoveries which again enabled formation of AA9 expanded family. Based on the expanded AA9 family, we also mined the genome of the protozoa

Fig. 4 — GH5 of early lineage fungi. The GH5 (EC 3.2.1.4) 1,4 $\beta$-endoglucanases of the chytrid \textit{R. rosea} are embedded in an ”all fungal” clade, containing also ascomycetes, basidiomycetes and zygomycetes GH5 endoglucanases. GH5 endoglucanases of anaerobic rumen fungi (Neocladomastigomycota) are found in a clade dominated by rumen bacterial GH5 proteins. Besides these two major occurrences of early lineage fungal GH5, one copy of \textit{R. rosea} GH5 is found in a predominantly bacterial GH5 clade (to the left) together with one GH5 protein from a zygomycetes species (\textit{Umbelopsis ramanniana}) and an Archaea (\textit{Pyrococcus abyssi}) GH5 protein. Two possible horizontal gene transfer events are indicated (between rumen bacteria and rumen fungi and between rumen bacteria and rumen protozoa). Color code: Purple, early lineage fungi; red, Zygomycota; green, Ascomycota. orange, Basidiomycota; blue, Bacteria; turquoise, Archaea; dark green, Planta; black, Animalia.

L. Lange et al.
Fig. 5 – *R. rosea* GH5 gene structure: The bacterial-related GH5 variant of *R. rosea* (top gene, MF432154; placed in an all-bacterial clade in Fig. 4) has one intron in the catalytic domain and no CBM. The fungal related *R. rosea* GH5 variants have a CBM1 domain (MF432153, MF432152) and up to many introns and additional gene domain inserts (MF432153, MF432158, MF432152). This means that the gene structure of the bacterial-related *R. rosea* GH5 (Fig. 4) has a gene structure more closely related to bacterial genes, suggesting that it has been more recently acquired through horizontal gene transfer from bacteria.

Fig. 6 – Phylogenetic tree of GH13, indicating the very broad diversity of enzymes of zoosporic fungi here found in all parts of the radial tree. GH13 of Neocallimastigomycota and Blastocladiomycota are both embedded in bacterial clade; GH13 of Chytridiomycota are placed in the all-fungal clade and in a bacterial clade. Color code: purple, early lineage fungi; red, Zygomycota; green, Ascomycota; orange, Basidiomycota; blue, bacteria; turquoise, Archaea; dark green, Planta; black, Animalia.
Dictyostelium discoideum (Eichinger et al., 2005). Interestingly, for this soil-inhabiting amoeba, which is described as a rich cellulose degrader, we did not find any LPMOs. This result could support the interpretation that LPMOs have developed (or been acquired) at some time between when the protozoa lineage and the fungal lineage split off from the eukaryotes. Notably, LPMOs are found also in bacteria and viruses, and are described from a range of aquatic organisms (algae and animals) (Vaaje-Kolstad et al., 2017). Both occurrences allow the possibility that LPMO can have been acquired or re-acquired by early lineage fungi through HGT, evolutionary paving the way for mobilizing cellulosic substrates.

Based on the findings that LPMOs are present in the early lineage fungi and that selected species of early lineage fungi have the fully fledged spectrum of cellulolytic, xylanolytic and pectinolytic cell wall degrading enzymes, we suggest the following evolutionary hypothesis: The challenge of decomposing the recalcitrant cellulose fibers of plant cell walls has been the driver for shaping the development of the entire spectrum of enzymes that degrade plant cell wall polymers. More specifically we suggest that the LPMOs could have played a key role in this evolution of microbial biomass conversion, essential for developing the biological Kingdoms. The radically different mechanism of LPMO enzymes creates synergy with the primarily hydrolytic cell wall degrading enzymes. Further, the many copies of LPMO could indicate that the LPMO mechanism is promiscuous, being active on many different types of plant (and animal) components going beyond cellulose and hemicellulose, possibly both by breaking bonds and by making room for the large enzyme proteins to establish contact with their substrates.

The outlier position of R. rosea is the most striking observation from the PCA plot (for strain description see (Gleason et al., 2004). This position is underpinned by its full spectrum of cellulolytic, xylanolytic, and pectinolytic enzymes, as well as a host of proteins degrading mannan and starch. R. rosea in its evolutionary transition to the terrestrial life form (where it is a biomass degrader widely present in agricultural soils) has acquired or maintained a range of plant cell wall degrading enzymes. Notably, five of the nine different cellulolytic enzymes and 10 of the 13 xylanolytic enzymes of R. rosea are in this study unique for R. rosea; more specifically not found in the genomes of the other carbohydrate metabolizing species included in this study, neither in the rumen fungi nor in the G. prolifera genome. The other outlier in the PCA plots (Figs S3 and S4) is N. californiae, an anaerobic rumen fungus (Solomon et al., 2016). Neocallimastix californiae (and other rumen fungi albeit to a lesser extent) has the most highly diversified spectrum of plant cell wall polymer degrading enzymes described, which is especially rich in cellulolytic and xylanolytic enzymes. Gonapodaya prolifera is an outstanding pectin-degrader (Chang et al., 2015). Its spectrum of pectinolytic enzymes is, however, less diverse with regard to number of enzyme functions as compared to R. rosea. Gonapodaya prolifera appears to gain its pectinolytic strength from the many copies and functions of the GH28 enzyme. Its few but prominent cellulose and hemicellulose degrading enzymes suggest that it acts in synergy with other organisms (e.g. bacteria) in breaking down such plant polymers.

Synechymium punctatus and R. globosum appear from the secreome analysis to be specialized to degrade starch, and for the latter species also to degrade chitin. The specialized biotrophic and pathogenic species, B. dendrobatidis, C. anguillulae, R. allomyces and H. polyhiza, have as expected from their lifestyle only few carbohydrate active enzymes, and basically no enzyme capacity for plant cell wall degradation (not relevant for its pathogenic specialization).

We document here that the enzymes of early lineage fungi represent a very rich pool of diversity of carbohydrate active enzymes of relevance also for industrial biomass conversion. A highly interesting aspect, possibly harboring significant additional enzyme diversity, is apparent from the enzyme variant/copy number heat map, Tables 4 and 5 This is however, only sporadically dealt with in the text above. From the phylogenetic trees, it appears that there is a high level of difference particularly between the R. rosea copies/variants of the same type of enzyme. This adds significant diversity to the enzyme discovery pool of early lineage fungi.

5. Conclusions and perspectives

The study of the enzymes of early lineage fungi, their diversity, function, phylogeny, evolution and applied potential is still only in its infancy. Seen from an evolutionary perspective aimed at contributing to resolving the lower lineages of the Fungal Tree of Life, the most interesting next steps would be to broaden such studies to include the enzyme profiles of Glomeromycota and Microsporidia. For elucidating the secreome evolution of zoosporic fungi, in the transition from aquatic to terrestrial life form the obvious next steps are to genome sequence more carbohydrate metabolizing aquatic and terrestrial zoosporic fungi. It further seems justified based on the current studies to suggest to include description of the enzyme activity profile to future descriptions of new fungal taxa as exemplified in the Aspergillus hankokkii paper (Pitt et al., 2017). An LPMO-related puzzle worth investigating is that we have not found any enzyme mechanism to explain how rumen fungi under anaerobic conditions cope with breaking down recalcitrant cellulose from plant cell walls when they lack LPMOs; oxygenase activity does not normally function under anaerobic conditions (Vaaje-Kolstad et al., 2017). A hypothesis could be that GH48 and/or rumen bacterial cellulosomes possess a stronger processivity hydrolysis mechanism.

Conflict of interest

No issues of conflicting interest have been declared or identified.

Role of authors

Lene Lange: Formulated the idea, conceptualized the study, supplied expert knowledge regarding laboratory handling of Early Lineage fungi and wrote the paper.

Bo Pilgaard: Carried out all the experimental work with R. rosea, cloning, expression and characterization of the proteins, including enzyme assaying; and was responsible for HotPep and dbCAN analysis of genomes of all of the 12 species studied.
Florian-Alexander Herbst: Responsible for the MS analysis of R. rosea secretome, including interpretation of the results and calculation of relative abundances.

Peter Kamp Busk: Developed—together with Lene Lange—the PPR-based sequence technology platform. More specifically developed the HotPep program used in the current paper, as basis for prediction of function directly from sequence.

Frank Gleason: Sampled, identified and described the strong biomass degrading strain of R. rosea; contributed information about ecology and lab-culturing of this zoosporic fungus.

Anders Gorm Pedersen: Constructed and interpreted the phylogenetic trees included in the study.

Acknowledgement

We thank Kristian Barrett, PhD student, for his kind assistance with the lay out of the phylogenetic trees; and Pia Nord-Larsen for technical assistance in preparing the manuscript. The study was supported by Innovation Fund Denmark, #0603- 005228 (BIOVALUE SPIR).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fbr.2018.09.001.

REFERENCES


Hibbett, D.S., Binder, M., Bischoff, J.F., Blackwell, M., Cannon, P.F., Eriksson, O.E., Huhndorf, S., James, T.,


Lene Lange, Professor, PhD et Dr. scient., Center for BioProcess Engineering, Technical University of Denmark, DK2800 Lyngby, Denmark. Lene Lange has held full professorship positions at three Danish Universities, KU, AAU and DTU, and been Director of Research in both public and private research organizations. All through her career, she had her own experimental mycology research group, focusing on fungal interaction with their hosts and substrates. Current methodological approach: fast track discovery of new enzymes, predicting function directly from sequence; combined with mass spectrometry, phylogenetic analysis and activity screening for elucidating (meta-) secretome composition of taxonomic and ecological biomass conversion hotspots in Nature. Her interest in zoosporic fungi dates back to being research assistant for Professor F.K. Sparrow.