



In silico Proteomic Analysis Provides Insights Into Phylogenomics and Plant Biomass Deconstruction Potentials of the Tremelalles

Habibu Aliyu^{1*†}, Olga Gorte^{1†}, Xinhai Zhou^{1,2}, Anke Neumann¹ and Katrin Ochsenreither^{1*}

¹ Institute of Process Engineering in Life Science 2: Technical Biology, Karlsruhe Institute of Technology, Karlsruhe, Germany,

² State Key Laboratory of Materials-Oriented Chemical Engineering, College of Biotechnology and Pharmaceutical Engineering, Nanjing Tech University, Nanjing, China

OPEN ACCESS

Edited by:

Joao Carlos Setubal,
University of São Paulo, Brazil

Reviewed by:

Marco Antonio Seiki Kadowaki,
Université libre de Bruxelles, Belgium

Renato Graciano De Paula,
Federal University of Espírito Santo,
Brazil

Gabriel Paes,
Fractionnation of AgroResources
and Environment (INRA), France

*Correspondence:

Habibu Aliyu
habibu.aliyu@partner.kit.edu;
aliyuhabibu@gmail.com
Katrin Ochsenreither
katrin.ochsenreither@kit.edu

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

Received: 06 January 2020

Accepted: 05 March 2020

Published: 03 April 2020

Citation:

Aliyu H, Gorte O, Zhou X,
Neumann A and Ochsenreither K
(2020) *In silico* Proteomic Analysis
Provides Insights Into Phylogenomics
and Plant Biomass Deconstruction
Potentials of the Tremelalles.
Front. Bioeng. Biotechnol. 8:226.
doi: 10.3389/fbioe.2020.00226

Basidiomycetes populate a wide range of ecological niches but unlike ascomycetes, their capabilities to decay plant polymers and their potential for biotechnological approaches receive less attention. Particularly, identification and isolation of CAZymes is of biotechnological relevance and has the potential to improve the cache of currently available commercial enzyme cocktails toward enhanced plant biomass utilization. The order Tremellales comprises phylogenetically diverse fungi living as human pathogens, mycoparasites, saprophytes or associated with insects. Here, we have employed comparative genomics approaches to highlight the phylogenomic relationships among thirty-five Tremellales and to identify putative enzymes of biotechnological interest encoded on their genomes. Evaluation of the predicted proteomes of the thirty-five Tremellales revealed 6,918 putative carbohydrate-active enzymes (CAZymes) and 7,066 peptidases. Two soil isolates, *Saitozyma podzolica* DSM 27192 and *Cryptococcus* sp. JCM 24511, show higher numbers harboring an average of 317 compared to a range of 267–121 CAZymes for the rest of the strains. Similarly, the proteomes of the two soil isolates along with two plant associated strains contain higher number of peptidases sharing an average of 234 peptidases compared to a range of 226–167 for the rest of the strains. Despite these huge differences and the apparent enrichment of these enzymes among the soil isolates, the data revealed a diversity of the various enzyme families that does not reflect specific habitat type. Growth experiment on various carbohydrates to validate the predictions provides support for this view. Overall, the data indicates that the Tremellales could serve as a rich source of both CAZymes and peptidases with wide range of potential biotechnological relevance.

Keywords: biotechnology, CAZymes, peptidases, plant biomass degradation, phylogenomics, Tremellales

INTRODUCTION

Plant biomass is the most abundant carbon rich waste material and can be used by biorefineries for production of food, feed, building block chemicals and bioenergy, such as biofuels. The polymeric plant cell wall comprises cellulose, hemicellulose and pectin as main components while wooden plant material is cross-linked with the aromatic hetero polymer lignin during lignification (Harris and Stone, 2009; Kai et al., 2018). The degradation into monomeric compounds is difficult because

of the branched and complex structures (Harris and Stone, 2009; Kai et al., 2018). For instance, pre-treatment is required to deconstruct the complex association of the constituent polymers of the plant biomass prior to hydrolysis into monomers (Harris and Stone, 2009; Rosnow et al., 2017). However, pre-treatment increases processing costs and leads to lower competitiveness with standard fossil fuel (Sheridan, 2013). This process can be augmented biologically by application of various polysaccharide degrading and modifying enzymes, called carbohydrate-active enzymes (CAZymes) (Lombard et al., 2014; Silveira et al., 2015). Thus, finding and isolation of CAZymes is of biotechnological relevance and has the potential to improve the cache of currently available commercial enzyme cocktails toward enhanced plant biomass utilization.

Aside polysaccharides and lignin, plant biomass also consist of substantial amounts (~5% of the total biomass) of proteins, triglycerides and terpenes (De Schouwer et al., 2019). Considering the estimated global annual biomass production of 146 billion metric tons (Balat and Ayar, 2005), the amount of these secondary fractions is substantial and could be integrated into the overall scheme of biorefining for the production of important industrial chemicals (Bilal et al., 2017; de Paula et al., 2019). Deconstruction of proteins by proteolytic enzymes mainly yields shorter peptides and amino acids that can be utilized in various industries, including food, animal feed, cosmetics, pharmaceuticals, and agrochemicals (De Schouwer et al., 2019). Proteolytic enzymes are among the top industrial enzymes and have been previously reported to account for ~60% of global enzyme demand (Rao et al., 1998), a market that has been projected to reach a value of \$6.32 billion by 2021 (Chapman et al., 2018).

Basidiomycetes populate a wide range of ecological niches, including forest, crops, compost, and plant matter in soils, and are adapted to various substrates. In contrast to ascomycetes, basidiomycetes are less studied for their capabilities to decay plant polymers and their potential for industrial use. The apparent lack of interest in the Basidiomycetes has been linked to the long standing popularity and industrial relevance of the ascomycetes which is well-established rather than the lack of industrially relevant enzymes (Rytioja et al., 2014). Promising candidates for identification of new CAZymes and proteolytic enzymes are fungi isolated from such different biotopes. Recently, a new oleaginous yeasts *Saitozyma podzolica* DSM 27192 has been isolated from peat bog soil (Schulze et al., 2014). Initial analysis of the predicted proteome relative to other members of the order Tremellales revealed that DSM 27192 genome encodes a larger number of proteins linked to carbohydrate-active and proteolytic enzymes (Aliyu et al., 2019).

Previous genomic studies on the distribution of plant biomass hydrolyzing enzymes among fungi (Zhao et al., 2014) have largely paid less attention to the Tremellales. Similarly, the contribution and identification of putative enzymes for the hydrolysis of the protein component of the plant biomass remained understudied (De Schouwer et al., 2019) compared to polysaccharides. Here, we have employed phylogenomic and comparative genomic strategies to study thirty-five members of the order Tremellales. We used genome scale phylogenetic

analysis to decipher the genetic diversity among members of the order. In-depth comparison of the predicted proteomes with emphasis on sources of isolation identified a diverse repertoire of CAZymes and peptidases and revealed an enrichment of these features among the two-soil isolated *Saitozyma* species. This study, therefore, enhances the understanding of both the evolutionary and functional diversities of the Tremellales, which will be useful for the development of strategies for future biotechnological and industrial exploration of these fungi.

MATERIALS AND METHODS

Genome Sequences, Structural and Functional Annotations

The genomes of thirty-five members of the order Tremellales and one out group strain from the order Trichosporonales (Table 1), obtained from the NCBI or JGI databases, were structurally annotated using the Funannotate pipeline (v. 1.5.0-8f86f8c) (Love et al., 2018). The completeness of each genome was determined using BUSCO v3.0.3 (Waterhouse et al., 2018). To identify putative carbohydrate-active enzymes (CAZymes), the proteomes of the strains were annotated using run_dbcan 2.0, a dbCAN standalone algorithm comprising three tools; DIAMOND, Hotpep, and HMMER (Zhang et al., 2018). To improve CAZymes prediction, this study focuses only on sequences identified by at least two of these tools (Zhang et al., 2018). Proteolytic enzymes were identified using BLASTP search (*E*-value cut-off $\leq 1.00E-10$) with the proteomes of the strains as queries against MEROPS database release 12.1 (Rawlings et al., 2017). Major protein families of sugar and amino acid transporters were identified from Interproscan v.5.30-69.0 (Jones et al., 2014) annotation of the various proteomes.

Phylogenomic Analysis

Orthologous relationships among the predicted protein sequences of the thirty-five Tremellales and the outgroup strains were inferred using OrthoFinder v2.3.3 (Emms and Kelly, 2015) with default parameter settings. To construct the phylogeny of Tremellales, single copy orthologs, were aligned using T-coffee v11.00.8cbe486 (Notredame et al., 2000; Magis et al., 2014). The resultant alignment was concatenated and trimmed using Gblocks v0.9b (Castresana, 2000; Talavera and Castresana, 2007). The trimmed alignment was used to construct a Maximum likelihood (ML) tree using IQ-TREE version 1.6.7 (Schmidt et al., 2014) based on the LG + F + R6 model (predicted using IQ-TREE) and 1,000 bootstrap replicates. Average amino acid identity (AAI) and orthologous average nucleotide identity (orthoANI) among the compared strains were computed using CompareM¹ and OrthoANIu tool (Yoon et al., 2017).

Carbon Substrate Utilization

Saitozyma podzolica DSM 27192 was recently deposited at the DSMZ culture collection, Germany (Schulze et al., 2014). *Tremella mesenterica* DSM 1558 was obtained from

¹<https://github.com/dparks1134/CompareM>

TABLE 1 | Genomic features of thirty-five Tremellales and one outgroup strain included in this study.

Strains	Isolation source/locality	Assembly name	Assembly Size	# Scaffolds	Scaffold N50	% GC	# Genes	# Proteins	# tRNA	Com*	Dup [‡]
<i>Cryptococcus gattii</i> WM276	<i>Eucalyptus tereticornis</i> debris, Australia	ASM18594v1	18,374,760	14	1,333,124	47.85	6,570	6,429	141	97.6	0.0
<i>Cryptococcus neoformans</i> var. <i>grubii</i> H99	Human	CNA3	18,916,112	15	1,422,463	48.21	6,849	6,694	155	96.2	0.7
<i>Cryptococcus neoformans</i> var. <i>neoformans</i> JEC21	Generated from NIH B-3501 and NIH B-3502 α -types (progenies of ATCC 28957 \times ATCC 28958) (https://www.ncbi.nlm.nih.gov/pmc/articles/-PMC257671/pdf/iai00026-0294.pdf)	ASM9104v1	19,051,922	14	1,438,950	48.54	6,879	6,740	139	96.9	0.7
<i>Cryptococcus floricola</i> DSM 27421	Flower nectar collected in Tenerife, Canary Islands, Spain (https://mbio.asm.org/content/10/3/e00764-19)	ASM635230v1	21,743,217	16	1,629,715	53.47	7,703	7,511	192	93.1	0.7
<i>Cryptococcus wingfieldii</i> CBS 7118^T	Frass from scolytid beetles, South Africa	ASM614915v1	20,871,786	15	1,575,190	53.40	7,412	7,217	195	94.8	1.0
<i>Cryptococcus amyloletus</i> CBS 6039^T	Frass of larvae of <i>Enneadesmus forficulus</i> (buprestid beetle), South Africa	Cryp_amyl_CBS6039_V3	20,254,996	15	1,545,655	53.36	8,357	8,164	193	96.2	1.0
<i>Cryptococcus amyloletus</i> CBS 6273	Frass of <i>Sinoxylon ruficorne</i> in <i>Dichrostachys cinerea</i> , South Africa	Cryp_amyl_CBS6273_V2	20,280,838	18	1,529,651	53.17	8,350	8,175	175	93.1	0.3
<i>Cryptococcus depauperatus</i> CBS 7841^T	Dead spider, Canada	Filo_depa_CBS7841_V1	15,797,590	68	501,291	44.36	6,251	6,125	126	96.9	0.3
<i>Cryptococcus depauperatus</i> CBS 7855	Surface of colony of <i>Verticillium lecanii</i> , from dead caterpillar of <i>Carpocapsa pomonella</i> , Czechia	Filo_depa_CBS7855_V2	15,656,276	41	882,956	44.73	6,232	6,106	126	92.4	0.7
<i>Kwoniella heveanensis</i> BCC8398	Insect frass, Thailand (CBS 12232)	Cryp_heve_BCC8398_V1	25,469,355	67	709,378	51.85	8,039	7,957	82	91.7	0.3

(Continued)

TABLE 1 | Continued

Strains	Isolation source/locality	Assembly name	Assembly Size	# Scaffolds	Scaffold N50	% GC	# Genes	# Proteins	# tRNA	Com*	Dup ^{3%}
<i>Kwoniella heveanensis</i> CBS 569^T	Sheet rubber, Indonesia	Cryp_heve_CBS569_V2	25,253,281	242	174,698	51.78	7,939	7,853	86	92.1	0.7
<i>Kwoniella dejecticola</i> CBS 10117^T	Frass of the litchi fruit borer <i>Conopomorpha sinensis</i> , Vietnam	Cryp_deje_CBS10117_V1	23,862,392	13	2,116,304	48.38	8,519	8,445	74	94.1	0.3
<i>Kwoniella pini</i> CBS 10737^T	Dead needles of <i>Pinus sylvestris</i> in a mixed forest, Russia	Cryp_pinu_CBS10737_V1	20,828,584	18	1,626,097	40.20	7,714	7,639	75	94.8	0.7
<i>Kwoniella bestiolae</i> CBS 10118^T	Frass of the litchi fruit borer <i>Conopomorpha sinensis</i> , Vietnam	Cryp_best_CBS10118_V1	24,360,772	12	3,422,270	47.22	9,068	8,992	76	96.2	0.3
<i>Kwoniella mangroviensis</i> CBS 10435	Seawater, Bahamas	Kwon_mang_CBS10435_V2	22,653,759	37	1,966,685	44.78	8,324	8,243	81	94.1	0.7
<i>Kwoniella mangroviensis</i> CBS 8507^T	Seawater	Kwon_mang_CBS8507_V2	22,654,511	62	1,048,156	44.76	8,326	8,249	77	95.5	0.0
<i>Kwoniella mangroviensis</i> CBS 8886	Seawater (mangrove), Bahamas	Kwon_mang_CBS8886_V1	22,872,713	41	2,035,775	44.53	8,342	8,264	78	95.2	0.0
<i>Saitozyma podzolica</i> DSM 27192	Peat bog soil, Germany	ASM394221v1	29,888,215	46	1,066,819	58.79	10,312	10,224	88	91.4	0.0
<i>Cryptococcus</i> sp. JCM 24511	Soil, Japan	JCM_24511_assembly_v001	28,180,937	55	1,281,587	58.53	9,890	9,835	55	94.8	0.0
<i>Dioszegia crocea</i> JCM 2961^T	Strawberries, United Kingdom	JCM_2961_assembly_v001	20,595,339	26	1,954,963	53.18	8,685	8,639	46	94.8	1.0
<i>Dioszegia aurantiaca</i> JCM 2956^T	Over-wintered nettle stems of <i>Urtica</i> sp., Canada	JCM_2956_assembly_v001	19,344,119	52	1,282,637	53.56	7,883	7,835	48	93.1	0.7
<i>Tremella mesenterica</i> ATCC 28783	<i>Alnus rubra</i> , Canada	Trem_mese_ATCC28783_V1	27,110,854	295	180,988	41.28	7,348	7,309	39	93.8	1.0
<i>Tremella mesenterica</i> DSM 1558	<i>Alnus rubra</i>	Treme1	28,639,919	45	1,622,698	45.67	8,548	8,513	35	87.6	0.7
<i>Tremella fuciformis</i> tr26	(derived from <i>T. fuciformis</i> T0040) <i>Annulohypoxyton stygium</i> , China	ASM98790v1	23,635,607	3,502	18,448	57.03	10,769	10,713	56	88.6	0.7
<i>Fibulobasidium inconspicuum</i> Phaff 89-39	Unknown (https://phaffcollection.ucdavis.edu/search-online-ordering)	Phaff89-39v1.0	20,291,509	31	1,209,751	59.08	7,944	7,900	44	92.4	1.0

(Continued)

TABLE 1 | Continued

Strains	Isolation source/locality	Assembly name	Assembly Size	# Scaffolds	Scaffold N50	% GC	# Genes	# Proteins	# tRNA	Com*	Dup [‡]
<i>Naematelia encephala</i> UCDFST 68-887.2	Isolated from cracked bark of <i>Salix</i> sp., Canada (https://phaffcollection.ucdavis.edu/search-online-ordering)	Treen1	19,786,307	151	209,500	49.30	7,674	7,631	43	87.3	0.7
<i>Fellomyces penicillatus</i> Phaff 54-35	Isolated from Hawaiian pineapple (https://phaffcollection.ucdavis.edu/search-online-ordering)	Phaff54-35v1.0	21,049,006	22	2,827,852	47.59	8,094	8,055	39	96.6	0.7
<i>Kockovaella imperatae</i> NRRL Y 17943	Leaf of cogon, <i>Imperata cylindrica</i> Thailand (https://nrrl.ncaur.usda.gov/cgi-bin/usda/yeast/report.html?nrrlcodes=Y%2d17943)	Kocim1	17,465,713	38	1,071,374	52.20	7,122	7,091	31	87.2	0.3
<i>Bullera alba</i> JCM 2954^T	Air in a dairy, United States	JCM_2954_assembly_v001	19,417,308	8	2,864,929	54.31	7,835	7,794	41	96.9	0.3
<i>Sirobasidium intermedium</i> CBS 7805	<i>Eutypella leprosa</i> on twig of <i>Ulmus</i> sp. (elm tree), United Kingdom	CBS7805v1.0	21,959,807	102	832,995	56.14	9,016	8,966	50	92.0	1.0
<i>Papiliotrema flavescens</i> NRRL Y 50378	Derived from OH182.9 (https://www.ars.usda.gov/research/publications/publication/?seqNo115=237942) Wheat anther, United States (https://www.sciencedirect.com/science/article/pii/S2352407316300348#bb0055)	Cf_30_300r_Split10plusN	22,790,521	712	71,416	58.46	9,328	9,262	66	91.4	0.7
<i>Papiliotrema laurentii</i> RY1	Food (Kombucha tea); India	ASM73882v1	19,145,913	1,152	32,353	56.13	7,687	7,584	103	91.3	0.3
<i>Phaeotremella skinneri</i> JCM 9039^T	Insect frass beneath the bark of hemlock <i>Tsuga heterophylla</i> , United States	JCM_9039_assembly_v001	20,788,256	58	1,003,985	50.23	7,471	7,423	48	95.9	0.7
<i>Phaeotremella fagi</i> JCM 13614	Rotten beech, Netherlands	JCM_13614_assembly_v001	22,649,946	30	1,602,030	42.43	10,146	10,089	57	94.5	0.7
<i>Cryptococcus</i> sp. 05/00	Rainbow hydrothermal site on the Mid-Atlantic Ridge	Mo29	23,801,231	687	165,604	53.09	8,179	7,984	195	91.4	0.7
<i>Cutaneotrichosporon oleaginosum</i> ATCC 20509^T	Dairy plant	ASM171244v1	19,908,169	16	250,97,47	60.59	8,302	8,036	266	92.0	0.3

*Completeness and [‡]duplication determined using BUSCO based on *fungi_odb9*.

DSMZ. *Cutaneotrichosporon oleaginosum* ATCC 20509 was acquired from ATCC Culture Collection, United States. *Dioszegia aurantiaca* CBS 6980, *Kwoniella mangrovensis* CBS 8507, *Sirobasidium intermedium* CBS 7805 and *Cryptococcus amylolentus* CBS 6039 were purchased from CBS, Netherlands. *Cryptococcus* sp. JCM 24511 and *Cryptococcus fagi* JCM 13614 were ordered from JCM, Japan.

The ability of isolates to metabolize certain carbohydrates was characterized with a standardized API 50 CHL system (BioMérieux, Nürtingen, Germany) consisting of 50 biochemical tests. The strains were activated in liquid culture containing YM medium (3 g/L yeast extract, 3 g/L malt extract, 5 g/L peptone, pH 7, sterile supplemented with 10 g/L glucose after autoclaving). After 24 h of growth OD_{600 nm} was determined. The cultures were washed twice with sterile saline (0.9% w/v, NaCl) and the pellets were resuspended in sterile distilled water containing 0.17 g/L bromocresol purple to an OD_{600 nm} = 1. Each isolate suspension was applied into the pockets of the API 50 CH. Strips were moistened and covered as recommended by the manufacturer and incubated at optimal growth temperature of the respective strain (20, 25, or 30°C). Colorimetric changes were recorded after the first 3 days and verified after 7 and 10 days.

For the plate-based tests, YNB medium (HP26.1; Carl Roth) was prepared according to manufacturer's protocol with 15 g/L agar. Xylan from beechwood (4414.4 Carl Roth), xylan from corncobs (8659.3 Carl Roth), inulin (I2255; Sigma-Aldrich), cellulose (Avicel PH-101; 11365; Sigma-Aldrich), starch (S4126; Sigma-Aldrich), carboxymethylcellulose (CMC; C5013; Sigma-Aldrich), pectin (93854; Sigma-Aldrich), chitin (8845.1; Sigma-Aldrich), N-acetyl-D-glucosamine (8993.2; Carl Roth), D(+)-Glucosamine (3769.1 Carl Roth) and D-Glucose monohydrate as positive control, were dissolved in distilled water, autoclaved separately and supplemented to the medium to a final concentration of 10 g/L. As negative control agar without carbon source was prepared. All nine strains were activated in YM medium, as described above, and washed twice with sterile 0.9% NaCl. The pellets were then subsequently resuspended in saline to an OD_{600 nm} = 1. Each agar plate was inoculated with three 10 µL of strain suspension. The plates were incubated at the optimal growth temperature (20, 25, and 30°C) and monitored for 10 days.

Principal component analysis (PCA) plots and heatmaps are generated using Clustvis web tool (Metsalu and Vilo, 2015) while correlation analyses and visualization were conducted in R 3.5.2 using packages corrplot (Wei et al., 2017) and PerformanceAnalytics (Peterson et al., 2018).

RESULTS

Genome Characteristics and Niche Specialization Among the Tremellales

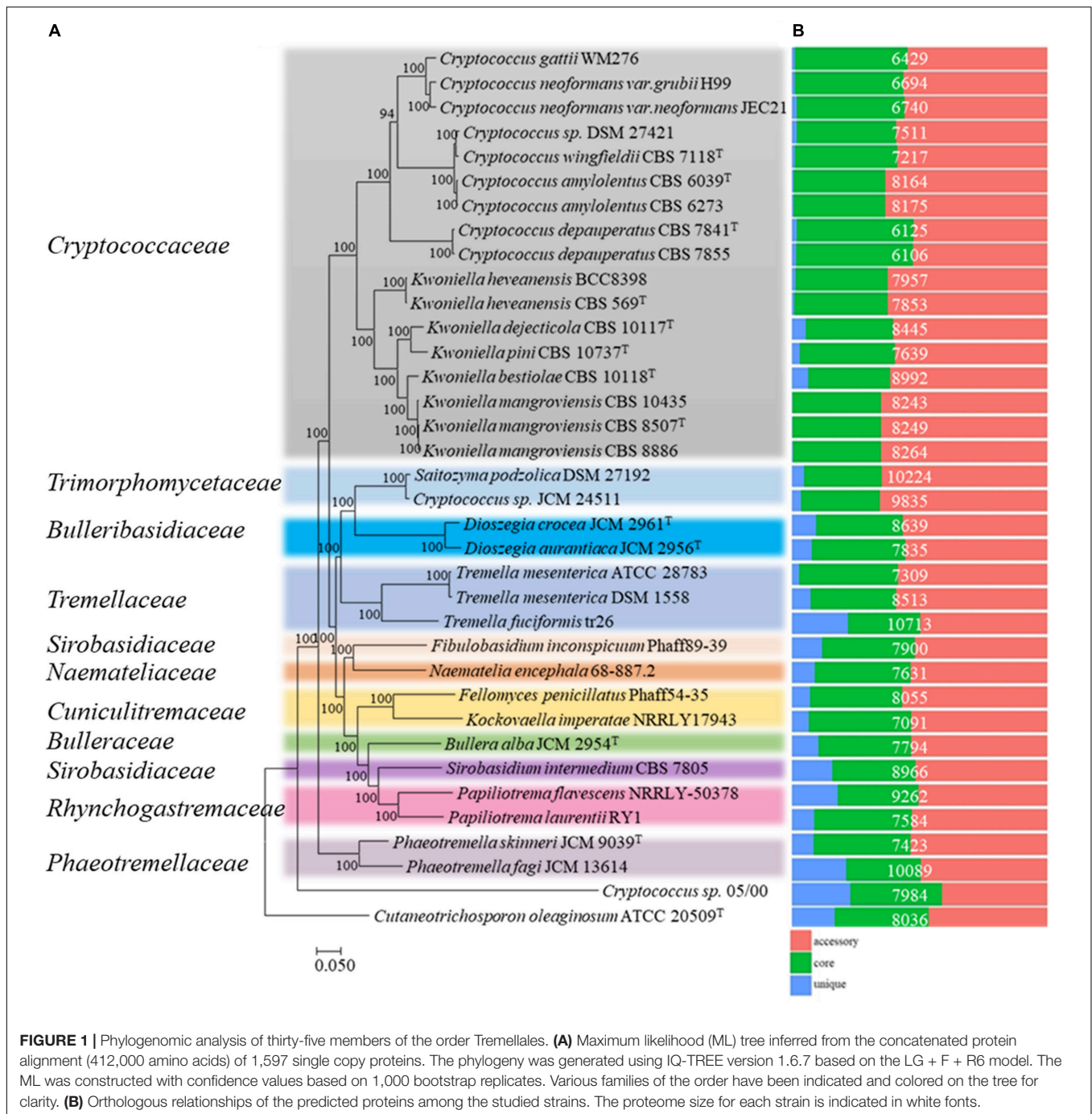
A survey of the NCBI database showed that as at October, 2019, 106 genome assemblies of strains affiliated with Tremellales are public. Majority of these assemblies are from the family *Cryptococcaceae* (93; ~ 88%), with isolates of *Cryptococcus neoformans* and *C. gattii* comprising 54 (~51%) and 20 (~19%),

respectively, of the total available sequences. The present study, however, focuses on genome sequences of thirty-five isolates from eleven families of Tremellales, including one genome sequence each of *C. gattii* WM276, *C. neoformans* var. *grubii* H99 and *C. neoformans* var. *neoformans* JEC21 selected as proxies for the various species complexes to which they belong. The genome of *Cutaneotrichosporon oleaginosum* ATCC 20509^T (order Trichosporonales) was included as an outgroup (**Table 1**). The sizes of the thirty-five studied genomes range between 15.66 Mb (*C. depauperatus* CBS 7855) and 29.88 Mb (*Saitozyma podzolica* DSM 27192) base pairs and code between 6,232 (*C. depauperatus* CBS 7855) and 10,769 (*Tremella fuciformis* tr26) genes including between 31 (*Kockovaella imperatae* NRRL Y 17943) and 195 (*C. wingfieldii* CBS 7118^T) tRNA genes. These genome sizes fall within the lower end of reported fungal genomes (8.97–177.57 Mb) and well below the average of those of Basidiomycota which has been estimated at 46.48 Mb (Mohanta and Bae, 2015). The number of protein coding genes ranged between 6,106 in *Cryptococcus depauperatus* CBS 7855 and 10,713 in *Tremella fuciformis* tr26 both which are lower than the average of ~15,432 protein models reported for Basidiomycota. However, tr26 genome shows the highest contigs number of 3,502, which may be indicative of a highly fragmented genome sequence. Strain tr26 is also characterized by low level of genome completeness (~88.6%). However, based on BUSCO (Waterhouse et al., 2018) fungi_odb9 single copy orthologs, genome completeness among the studied strains varied between 87.2% in *Kockovaella imperatae* NRRL Y 17943 (38 contigs) and 97.6% in *C. gattii* WM276 (14 contigs). The latter isolate along with *S. podzolica* DSM 27192, *Cryptococcus* sp. JCM 24511, *Kwoniella mangrovensis* CBS 8886 and *K. mangrovensis* CBS 8507 showed zero evidence of genome duplication while the rest of the studied strains contain low level of duplications, ranging between 0.3 and 1%. Its noteworthy, however, that this work reports only the genome properties of the monokaryotic (asexual) yeast state of these fungi.

With reference to their sources of origin (**Table 1**), the studied isolates show wide distribution with strains isolated from dairy (1), fungi (3), human (2), insect frass (7), plant (11), soil (2), sea (4) and one isolate each from dead spider, sheet rubber, rotten beech and kombucha tea. The origin of *Fibulobasidium inconspicuum* Phaff 89-39 is unknown. Thus, considering their preferred habitat and views from various literature (Findley et al., 2009; Millanes et al., 2011; Yurkov et al., 2015; May et al., 2016), the strains have been classified into eight groups, namely mycoparasitic, pathogenic, saprobic (arthropod frass, dead arthropods, plants, and others), sea- and soil -inhabiting.

Phylogenomics of the Tremellales

To place the studied strains into phylogenomic perspective, the orthologous relationships among the predicted proteome (289,686 proteins) of the thirty-five Tremellales and the outgroup strain were determined using OrthoFinder (Emms and Kelly, 2015). Analysis of the protein families (orthogroups) revealed 269,580 (93.1%) proteins are assigned in 13,504 orthogroups of which 122 (614; 0.2% proteins) are strain specific. Evaluation of the orthologous proteome identified 2,634 orthogroups (104,



039; ~36% proteins) are core to all the studied strains. Of these, 1,597 comprise the single copy orthogroups (SCO), representing between 15 and 25% of the individual proteomes of the studied fungi. To infer phylogenetic relationships among the studied Tremellales, a maximum likelihood (ML) tree (Figure 1A) was generated from a concatenated and trimmed alignment of the SCO proteins comprising 412,000 amino acids using IQ-TREE (Schmidt et al., 2014; Chernomor et al., 2016). Based on this tree, *Cryptococcus* sp. 05/00 clusters with the outgroup strain suggesting a distant relationship with the strains in the

order Tremellales. This strain also harbors greater proportion of unique proteins (Figure 1B) relative to the other strain, indicating disparate evolutionary history. Aside this, members of the family *Phaeotremellaceae* (*Phaeotremella* spp.) form a monophyletic clade distinct from the rest of the Tremellales which are placed in two separate major clusters. The first cluster comprises members of the genera *Cryptococcus* and *Kwoniella* while the second comprises ten genera from different families, including two *Saitozyma* spp. The latter cluster includes strains harboring greater number of strains specific proteins (Figure 1B).

To highlight the relationships at various taxa levels, average amino acid identity (AAI) values were determined among the thirty-five Tremellales (**Supplementary Figure S1A**). Excluding *Cryptococcus* sp. 05/00, the AAI between the compared strains ranged between 60.7 (*Kockovaella imperatae* NRRL Y 17943 and *Dioszegia aurantiaca* JCM 2956^T) and 98.72 (*Cryptococcus amylolentus* CBS 6273 and *C. amylolentus* CBS 6039^T). The *Cryptococcus* clade incorporates strains whose AAI ranged between 61.36 and 98.71% while isolates in the *Saitozyma* clade share AAI ranging between 56.15 and 92.67%, further confirming the greater diversity in the latter clade. Within the subclades, the *Cryptococcus* spp. share an AAI range of 64.61–98.71% with *C. gattii* WM276 and *C. neoformans* sharing an AAI between 86.28 and 90.02% and *C. amylolentus*, *C. wingfieldii* and *C. floricola* sharing between 91.62 and 98.71% AAI. Furthermore, the *Kwoniella* spp. show an AAI ranging between 67.28 and 97.98%. On the other hand, majority of the strains in *Saitozyma* clade form monophyletic branches with exception of *Saitozyma*, *Dioszegia*, and *Tremella* spp. for which the strains within each branch share an AAI of 88.18, 77.35, and 92.67%, respectively. To further ascertain the genetic relationship of strains at species level, average nucleotide identity (ANI) values among closely related isolates were computed (**Supplementary Figure S1B**). The ANI among *C. gattii* WM276 and *C. neoformans* ranged between 83.60 and 88.31% while *C. amylolentus*, *C. wingfieldii* and *C. floricola* shared ANI that range of 93.24–99.59%. In contrast, ANI of 86.52% is shared between *Saitozyma podzolica* DSM 27192 and *Cryptococcus* sp. JCM 24511 both of which have been affiliated with *Saitozyma podzolica*.

Genome Wide Comparisons of Carbohydrate-Active Enzymes and Proteolytic Enzymes

Proteolytic Enzymes

Evaluation of the predicted proteomes of thirty-five Tremellales revealed 7,079 (average: ~197) peptidases and peptidase inhibitors (**Supplementary Table S1**). Genomes of *C. depauperatus* CBS 7841^T isolated from a dead arthropod and *Tremella mesenterica* DSM 1558 isolated from plant code for the least peptidases of ~85% of the average peptidase number and those of *S. podzolica* DSM 27192 isolated from soil and *Phaeotremella fagi* JCM 13614 isolated from rotten beech harbor the highest number of peptidases; ~120% of the average. Clustering of the predicted peptidases, including information on isolation sources among the studied fungi revealed two major clusters in which the sea-, soil-dwelling and dead arthropods isolates group distinctly from the arthropod frass isolates while strains of fungal and plant sources are distributed across both clusters. However, the plant associated strains within soil isolates clade show distinct clustering (**Figures 2A,B**).

Predictions for possible explanation for the observed differences in peptidase abundance among the Tremellales revealed a significant ($P < 0.01$) and strong ($r = 0.71$) correlation between proteome size and protease abundance (**Figure 3** and **Supplementary Figure S2**). Despite the significant ($P < 0.01$) correlation between genome and proteome sizes, the former show

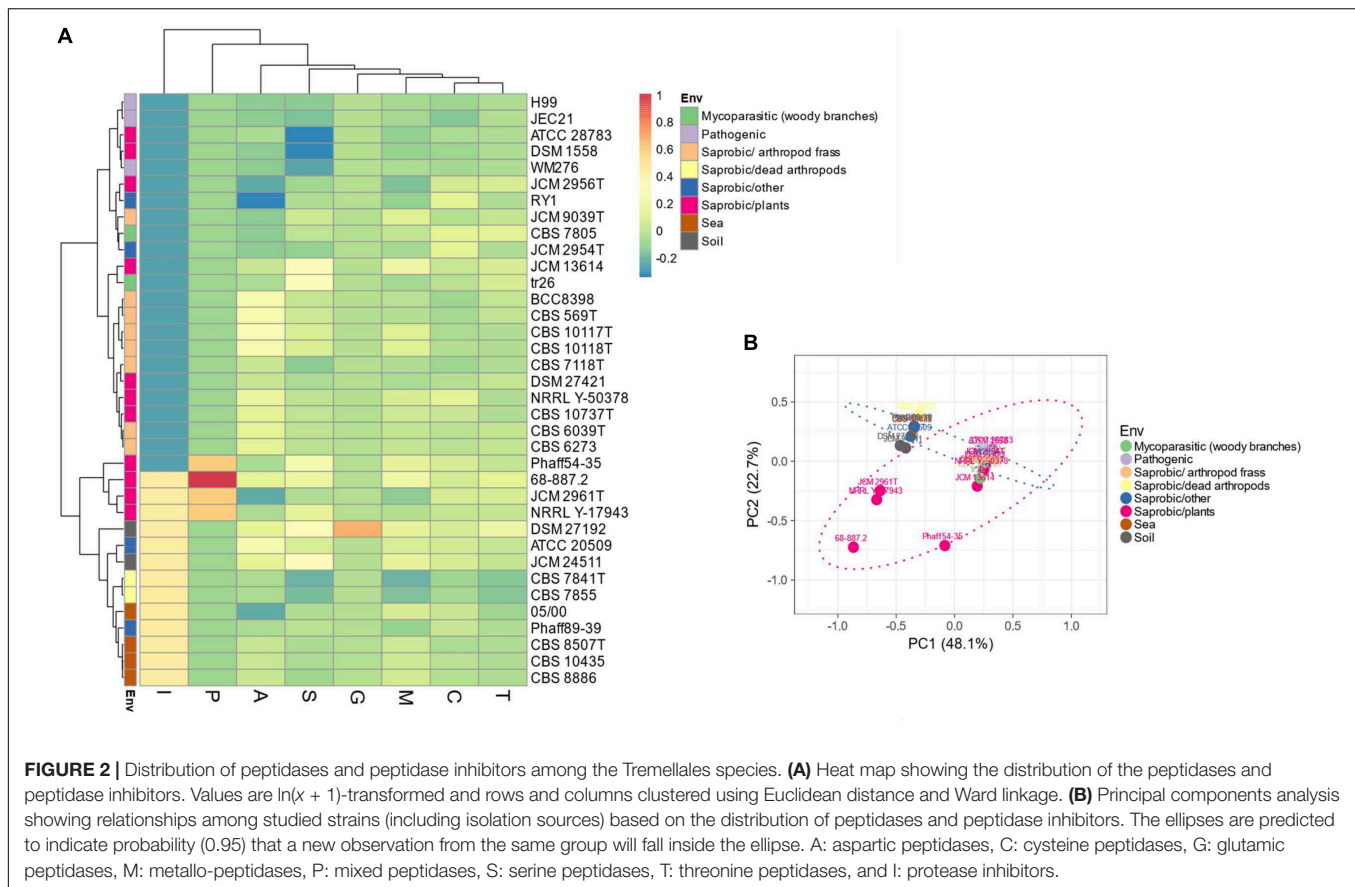
no significant association with peptidases number. Similarly, habitat type appears to contribute to peptidase abundance with the soil and certain saprobic plant associated strains clearly harboring higher numbers of peptidases. Further, the proteomes were queried for two families of amino acid transporters, namely amino acid permease (PF00324) and transmembrane amino acid transporter protein (PF01490). Evaluation of these transporters revealed that only the former shows significant ($P < 0.01$) correlation with protease abundance (**Supplementary Figure S2**). The predicted inhibitors of peptidases (I) and two peptidase families (G and P) are uniquely associated with specific strains in the soil isolates cluster (**Figures 2A,B** and **Supplementary Table S1**). A putative aspergilloglutamic peptidase (G01) is uniquely present in *S. podzolica* DSM 27192 while the mixed peptidase, P01 (putative β -aminopeptidase), is exclusive found among the plant associated isolates of the group; *Naematelia encephala* UCDFST 68-887.2, *Dioszegia crocea* JCM 2961^T, *K. imperatae* NRRL Y 17943 and *Fellomyces penicillatus* Phaff 54-35. The genome of UCDFST 68-887.2 encodes homologs of both BapF peptidase and DmpA aminopeptidase, JCM 2961^T and Phaff54-35 genomes encode only putative BapF peptidase and NRRL Y 17943 genome harbors only DmpA aminopeptidase. Both enzymes have been suggested to have similar substrate specificity (John-White et al., 2017). The predicted peptidase inhibitors include I32 (survivin) present in the two soil isolates (*S. podzolica* DSM 27192 and *Cryptococcus* sp. JCM 24511), the hydrothermal isolate (05/00) and Phaff89-39 which is of unknown origin while I51 (serine carboxypeptidase Y inhibitor) was identified in eight strains including three each isolated from plants and sea, and one each from a dead spider and fungi.

Aspartic peptidases (APs)

Evaluation of the peptidase sets (**Supplementary Table S1**) predicted from the proteomes of the studied isolates revealed only three putative APs belonging to the families A01A (pepsin A), A22B (impas 1 peptidase) and A28A (DNA-damage inducible protein 1). The latter two families occur in single copies in all studied strains while the number of A01A ranged between 6 and 14. Although, the distribution of A1 differs among isolates of similar habitat types, this peptidase is more prevalent among isolates associated with arthropods frass relative to the pathogenic strains (**Figures 2A,B** and **Supplementary Table S1**).

Cysteine peptidases (CPs)

Twenty putative CPs families have been identified in the proteomes of the studied isolates with members of four families, namely C13 (putative glycosylphosphatidylinositol: protein transamidase), C54 (autophagin-1), C85 (OTUD5 peptidase), C86 (ataxin-3) being shared by all strains (**Figures 2A,B** and **Supplementary Table S1**). Further three families, C50 (separase), C65 (otubain), and C78B (UfSP1 peptidase) are common to most of the strains while orthologs of C45 (acyl-coenzyme A:6-aminopenicillanic acid acyl-transferase precursor) and C110 (kyphoscoliosis peptidase) are rare, occurring in one strain each. On the other hand, C19 (ubiquitin-specific peptidase 14) and C26 (gamma-glutamyl hydrolase) are the most abundant



C-peptidases represented at an average of 14 and 9 proteins, respectively, in each strain. While there appears to be no pattern of association of specific C-peptidase families with fungal lifestyles (**Supplementary Figure S3A**), certain plant associated saprobes contain relatively higher numbers of C26 compared to the pathogenic strains (**Supplementary Figure S3A** and **Supplementary Table S1**). Conspicuously, three pathogenic, two soil, four arthropods frass (CBS 10117^T, CBS 6039^T, CBS 6273, and CBS 7118^T), two dead arthropods and two plant (CBS10737^T, DSM27421) associated isolates lack homologs of C01B (bleomycin hydrolase). Similarly, C115 (MINDY-1 protein) was restricted to 05/00, JCM 9039T and JCM 13614 while, DSM 27192, JCM 24511 (soil), JCM 2961T, 68-887.2 and Phaff54-35 harbor homologs of C15 (pyroglutamyl-peptidase I). Evaluation of the association between C-peptidases and genome features revealed no association with genome size (**Supplementary Figures S4A,B**). However, there was significant but weak correlation between number C-peptidases and proteome sizes with the smaller proteomes harboring fewer C-peptidases.

Metallopeptidases (MPs)

The MPs represent the most abundant and diverse proteases harboured by the studied Tremellales (**Figures 2A,B** and **Supplementary Table S1**). Analysis of the predicted proteases identified twenty-three distinct families of M-peptidases including M13 (nepriylisin) and M14A (carboxypeptidase A1)

occurring in single copies and M41 (FtsH peptidase) in duplicate copies in all the studied strains. On the average, 64 (range: 52–78) M-peptidases are encoded on the genomes of the studied isolates with M20 (putative glutamate carboxypeptidase) and M24 (methionyl aminopeptidase) being the most prevalent families with an average of 9 (range: 3–17) and 9 (range: 9–11) proteins, respectively. The distribution of M19 (membrane dipeptidase), M20 and M38 (isoaspartyl dipeptidase) seems to distinguish the soil, majority of the sea and arthropods frass from the pathogenic and mycoparasitic isolates (**Supplementary Figure S3b**). Strains JCM 2956T and Phaff89-39 uniquely harbor homologs of M06 (myroilysin) and M10B (membrane-type matrix metallopeptidase-6), respectively. Interestingly, only proteomes of three plant isolates, 68-887.2, NRRL Y 17943 and Phaff54-35 contain the putative M81 (microcystinase MlrC) while 05/00 harbors large numbers of M03A (thimet oligopeptidase).

Serine peptidases (SPs)

Prediction of the putative peptidases revealed 12 families of serine peptidases are encoded in the various genomes of Tremellales at an average of 55 (range: 38–77) proteins per genome. Members of the families S09 (prolyl oligopeptidase) and S33 (prolyl aminopeptidase) are the most prevalent S-proteases with an average of 20 (range: 12–35) and 14 (range: 10–24) proteins, respectively, while S14 (peptidase Clp), S28

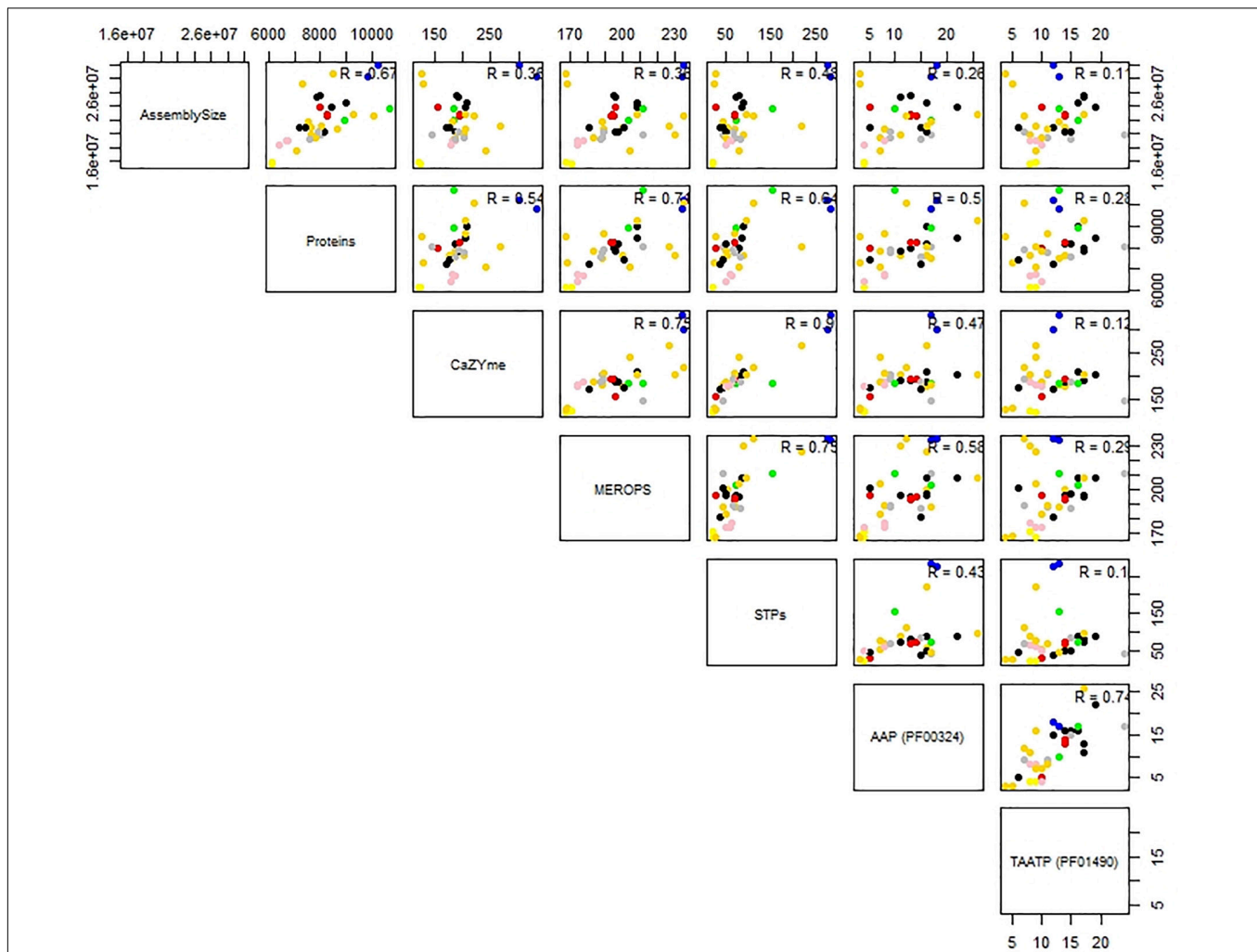


FIGURE 3 | Correlation analysis of genome and proteome sizes, and predicted CAZymes, peptidases, sugar, and amino acid transporters of thirty-five Tremellales. Coefficient of correlation R values indicates the strength of the association and niche specialization of the strains are color coded; black, blue, cyan, gold, green, gray, pink, red representing saprobic (arthropods frass), soil, saprobic (dead arthropods), saprobic (plants), mycoparasitic, saprobic (others), pathogenic, sea isolates, respectively.

(lysosomal Pro-Xaa carboxypeptidase), and S59 (nucleoporin 145) show similar distribution among all the studied strains (**Supplementary Table S1**).

Threonine peptidases (TPs)

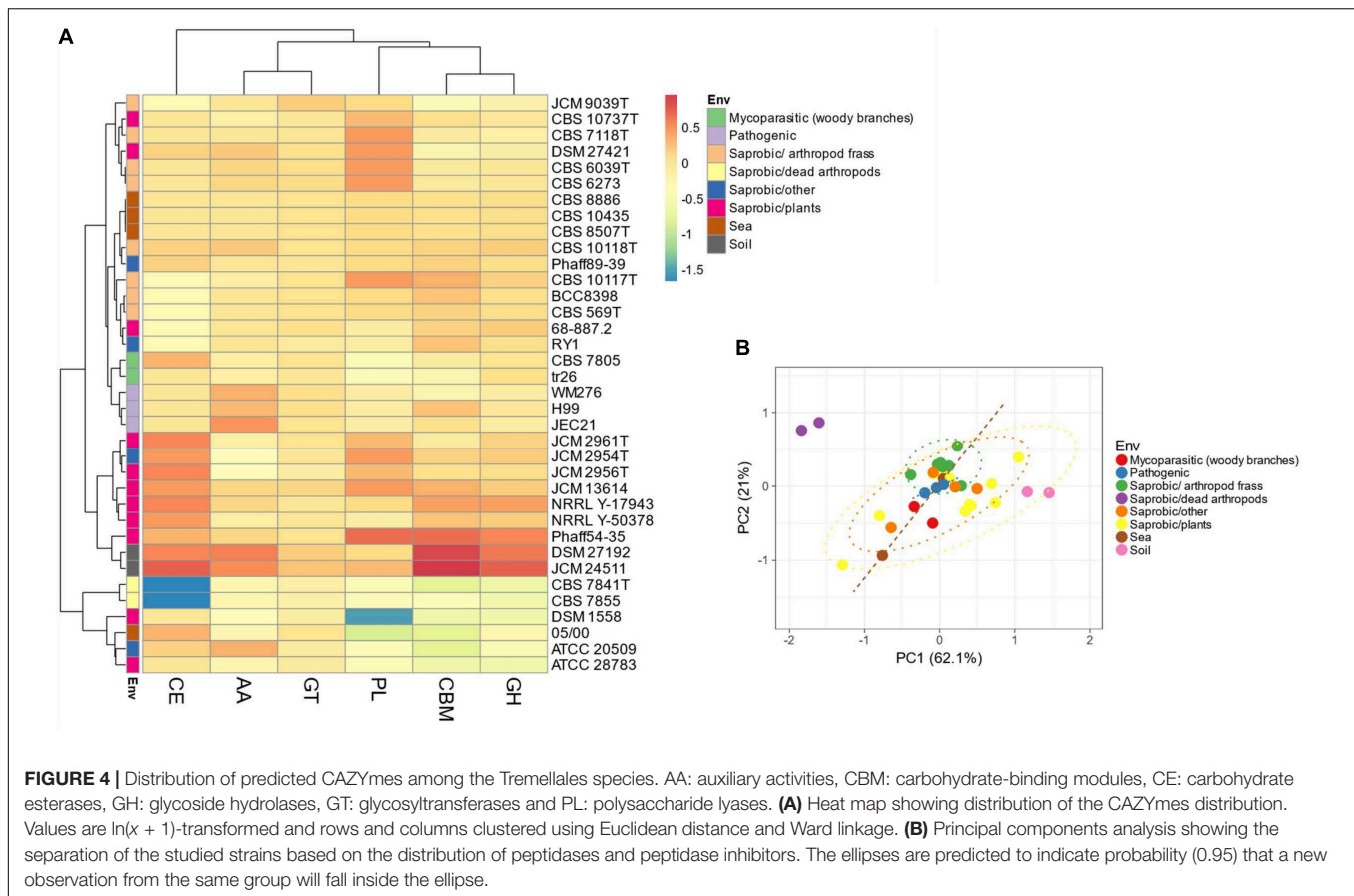
Of the 6 known T-peptidases, three have been identified in the genomes of the studied Tremellales with an overall average of 19 (range: 16–23, **Supplementary Table S1**) proteins. T05 (ornithine acetyltransferase precursor) occur in one copy in all but one strain while T01 (proteasome, beta component) family is overrepresented with 14 putative orthologs in all strains except DSM 27192 which contain 16 T01 proteins (**Supplementary Figure S3D**).

Carbohydrate-Active Enzymes

The thirty-five Tremellales and the outgroup strains harbor 6,918 carbohydrate-active enzymes (CAZymes) at an average of 192

proteins per proteome (**Supplementary Table S2**). The two soil isolates harbor the highest number of CAZymes with 301 and 333 proteins in *S. podzolica* DSM 27192 and *Cryptococcus* sp. JCM 24511, respectively, while the genomes of the two dead arthropods associated *C. depauperatus* strains, CBS 7841T and CBS 7855 code for the least number of CAZymes of 121 and 125 proteins, respectively. Consequently, the soil isolates alongside majority of the plant associated saprobes cluster separately from the rest of the strains based on the CAZymes distribution (**Figures 4A,B**). Furthermore, the mycoparasitic and pathogenic isolates appear to share similar CAZymes distribution relative to the sea and the arthropods frass isolates (**Figures 4A,B**).

Like the proteolytic enzymes, correlation analyses revealed significant ($P < 0.01$; $r = 0.54$) association between CAZymes and proteome size and lack of association with genome size (**Figure 3** and **Supplementary Figure S2**). Similarly, there was highly significant ($P < 0.01$) and strong association ($r = 0.90$)



between CAZymes abundance and predicted sugar transporters STs (**Figure 3**). The proteomes of the soil isolates, which harbor the highest number of CAZymes, also contain more STs, comprising 275 and 284 proteins in DSM 27192 and JCM 24511, respectively. By contrast the rest of the strains harbor between 20 and 220 proteins with the proteomes of the dead arthropod associated strains showing the least STs of 20–22 proteins (**Supplementary Figure S2**).

Auxiliary activities (AA)

Evaluation of the CAZymes datasets revealed 425 putative enzymes linked to ten auxiliary activities (AA) family (**Supplementary Table S2**). The soil isolates harbor relatively higher numbers of AA ranging between 19 and 20 proteins and distributed in seven AA classes. Although more diverse with 8 AA classes, the pathogenic strains follow closely with 15–18 AA proteins. However, the distribution of AA among the plant saprobes ranged between 8 and 14 proteins indicating potential functional diversity among the various saprobes. The genomes of all the studied Tremellales code for orthologs of AA families AA1 (laccase-like multicopper oxidases), AA3 (glucose-methanolcholine (GMC) oxidoreductases family) and AA5_1 (glyoxal oxidase). The genomes of DSM 27192 and Phaff54-35 uniquely code AA2 (class II lignin-modifying peroxidases) in single copies while orthologs of AA9 (formerly GH61; copper-dependent lytic polysaccharide monoxygenases) are present in single copies

uniquely among the three pathogenic alongside the two dead arthropods and one mycoparasitic (CBS 7805) strains.

Carbohydrate-binding modules (CBMs)

The genomes of the studied strains code for 341 proteins that contain CBMs with three, namely, CBM13 (associated with xylanase in fungi), CBM43 (usually associated with GH17 or GH72) and CBM48 (usually appended to GH13 or β subunit of AMPK) occurring in all strains (**Supplementary Table S2**). One protein each from JCM 2954^T, NRRL Y-17943 and Phaff54-35 contain CBM20 (associated with glucoamylase in fungi), CBM35 (usually associated with xylanases) and CBM66 (associated with β -fructosidase reported in *Bacillus subtilis*), respectively. However, CBM67 (linked to α -L-rhamnosidase) is the most abundant module observed in the strains occurring nearly in 2-folds (12–14 proteins) among the two soil strains compared to a range of 0–7 proteins among the thirty-three other strains. However, of the 341 CBMs, 223 (65.4%) occur in CAZymes, predominantly (115 CBMs) CBM67 in proteins with GH78 signatures while 118 CBMs occur in proteins without signatures of hydrolysis activity (**Supplementary Table S3**).

Carbohydrate esterases (CEs)

Genomes of the strains included in this study code for seven CEs with only CE4 being common to all strains (**Supplementary Table S2**). CE10 (arylesterase and others) was identified

only in strain JCM 2954^T, CE15 (4-O-methyl-glucuronoyl methyltransferase) occur in two strains; JCM 24511 and Phaff54-35 while CE8 (pectin methyltransferase) is unique to the two soil isolates DSM 27192 and JCM 24511, 05/00 (sea) and Phaff89-39 (unknown origin). The three *K. mangroviensis* (isolated from sea) and two saprobic isolates uniquely contain CE1 (wide range of esterases) while CE5 (acetyl xylan esterase and cutinases) occurs uniquely among five saprobic strains and one sea isolate. Furthermore, the sea isolates alongside two isolates each of arthropods frass, dead arthropods and plants lack orthologs of CE9 (N-acetylglucosamine 6-phosphate deacetylase).

Glycoside hydrolases (GHs)

Putative GHs proteins identified in the studied proteomes include 3,567 proteins distributed in 57 families, thus constituting the most abundant (~51.2%) and diverse putative CAZymes of the studied strains (**Supplementary Table S2**). Of the 57 GH families, 13 (GH105, GH13, GH133, GH16, GH17, GH18, GH3, GH37, GH47, GH5, GH71, GH72, and GH9) occur in at least one copy in all strains while 11 others (GH11, GH125, GH130, GH141, GH151, GH30_7, GH33, GH39, GH49, GH76, and GH97) show restricted distribution occurring only in between one to four strains. Consistent with overall CAZymes profile, the proteomes of the two soil, strains DSM 27192 and JCM 24511 contain the highest number of 175 and 198 GHs, respectively, represented each in 45 families. However, relative to the rest of the studied strains, two plant associated isolates NRRL Y-17943 (143 GHs) and Phaff54-35 (160 GHs) show greater GHs diversity with 52 and 49 families, respectively. Despite the difference between the four strains above, they uniquely harbor GH125 (exo- α -1,6-mannosidase), GH30_7 (endo- β -1,4-xylanase and several other enzymes), and GH67 (α -glucuronidase or xylan α -1,2-glucuronidase) while the latter strains uniquely contain single copies of GH55 (exo- β -1,3-glucanase or endo- β -1,3-glucanase) and GH76 (α -1,6-mannanase or α -glucosidase). By contrast the soil associated isolates show overrepresentation of GH1 (β -glucosidase and several other enzymes), GH106 (α -L-rhamnosidase or rhamnogalacturonan α -L-rhamnohydrolase), GH3 (β -glucosidase and several other enzymes) GH43 (β -xylosidase and several other enzymes), GH5 (endo- β -1,4-glucanase/cellulase and several other enzymes) and GH78 (α -L-rhamnosidase, rhamnogalacturonan α -L-rhamnohydrolase and L-Rhap- α -1,3-D-Apif-specific α -1,3-L-rhamnosidase) by at least 2-folds relative to the rest of the strains.

Glycosyl transferase (GTs)

A survey of the predicted CAZymes of the studied strains revealed 2,198 GTs grouped in 30 distinct families (**Supplementary Table S2**). Twenty-two GTs are represented in all studied proteomes with GT20, GT21, GT24, GT33, GT39, GT48, GT66 showing similar distributions. The soil isolate, JCM 24511 harbors the highest number of GTs of 76 distributed in 26 families but in terms of diversity, CBS 7118T, 05/00 and DSM 27192 each harbor 27 distinct GT families. GT2 (cellulose synthase, chitin synthase and numerous others) and GT90 (UDP-Xyl:

(mannosyl) glucuronoxylomannan/galactoxylomannan β -1,2-xylosyltransferase, UDP-Glc: protein O- β -glucosyltransferase and UDP-Xyl: protein O- β -xylosyltransferase) with 434 and 269 proteins, respectively constitute 31% of the identified GTs while GT17 (β -1,4-mannosyl-glycoprotein or β -1,4-N-acetylglucosaminyltransferase) and GT49 (β -1,3-N-acetylglucosaminyltransferase) constitute the rarest GTs occurring only in two strains each. Notably GT57 (Dol-P-Glc: α -1,3-glucosyltransferase) and GT59 (Dol-P-Glc: Glc2Man9GlcNAc2-PP-Dol α -1,2-glucosyltransferase) are shared by eight strains including two soil isolates while the pathogenic strains along with four arthropods frass, two plants and one sea isolates harbor GT76 (Dol-P-Man: α -1,6-mannosyltransferase), which is missing in the latter group. This group, however, harbors GT57 and GT59. Alpha-1,6-mannosyltransferase, involved in cell wall α -1,6-mannose backbone extension, has been implicated in pathogenicity of *Candida albicans* (Zhang et al., 2016) and may play similar role among the studied pathogenic strains.

Polysaccharide lyases (PLs)

Analysis of the putative CAZymes of the studied Tremellales reveals 177 PLs belonging to seven families including PL0 (non-classified) present in all strains (**Supplementary Table S2**). Surprisingly, two families PL1_4 (exo-pectate or pectate lyase) and PL3_2 (pectate lyase) are exclusively present in four arthropods frass isolates. With exception of CBS 10118T and CBS 10117T, all the arthropods frass, dead arthropods and the pathogenic isolates lack PL22_2 (oligogalacturonate lyase/oligogalacturonide lyase). The pathogenic, dead arthropods and the sea isolates lack PL8_4 (hyaluronate lyase and others) and the latter two groups are missing the orthologs of PL4 (rhamnogalacturonan endolyase). The implication of various combinations of PLs and other CAZymes (**Figure 5**) predicted to partake in complex biomass degradation is discussed below.

Utilization of Carbon Substrates

To assess the ability to assimilate various carbon sources and to deconstruct complex substrates, nine strains selected from distinct clades in the presented phylogeny (**Figure 1**), representing various habitat preferences were tested on API 50 CH strips and grown on agar plates containing eight polysaccharides and two monosaccharides. Results on 47 substrates for which at least one strain tested positive is summarized in **Figure 6**. All tested strains showed activity on eleven API substrates; D-cellobiose, D-fructose, D-glucose, D-lyxose, D-mannose, D-trehalose, D-xylose, esculin ferric citrate, inulin, and L-arabinose and grew on two of the eight polysaccharides; xylan (beech tree) and xylan (corncoobs). The *K. mangroviensis* CBS 8507^T and *S. podzolica* DSM 27192 isolated from sea and soil, respectively, each grew on 42 substrates while *T. mesenterica* DSM 1558 (plant isolate) metabolized only 26 substrates. *S. podzolica* DSM 27192 utilized seven carbon sources (methyl- α -D-glucopyranoside, D-ribose, L-arabitol, erythritol, L-xylose, D-mannitol, and xylitol) on which no activity was observed for its closest relative *Cryptococcus* sp. JCM 24511. Utilization of these seven substrates by DSM 27192 and the

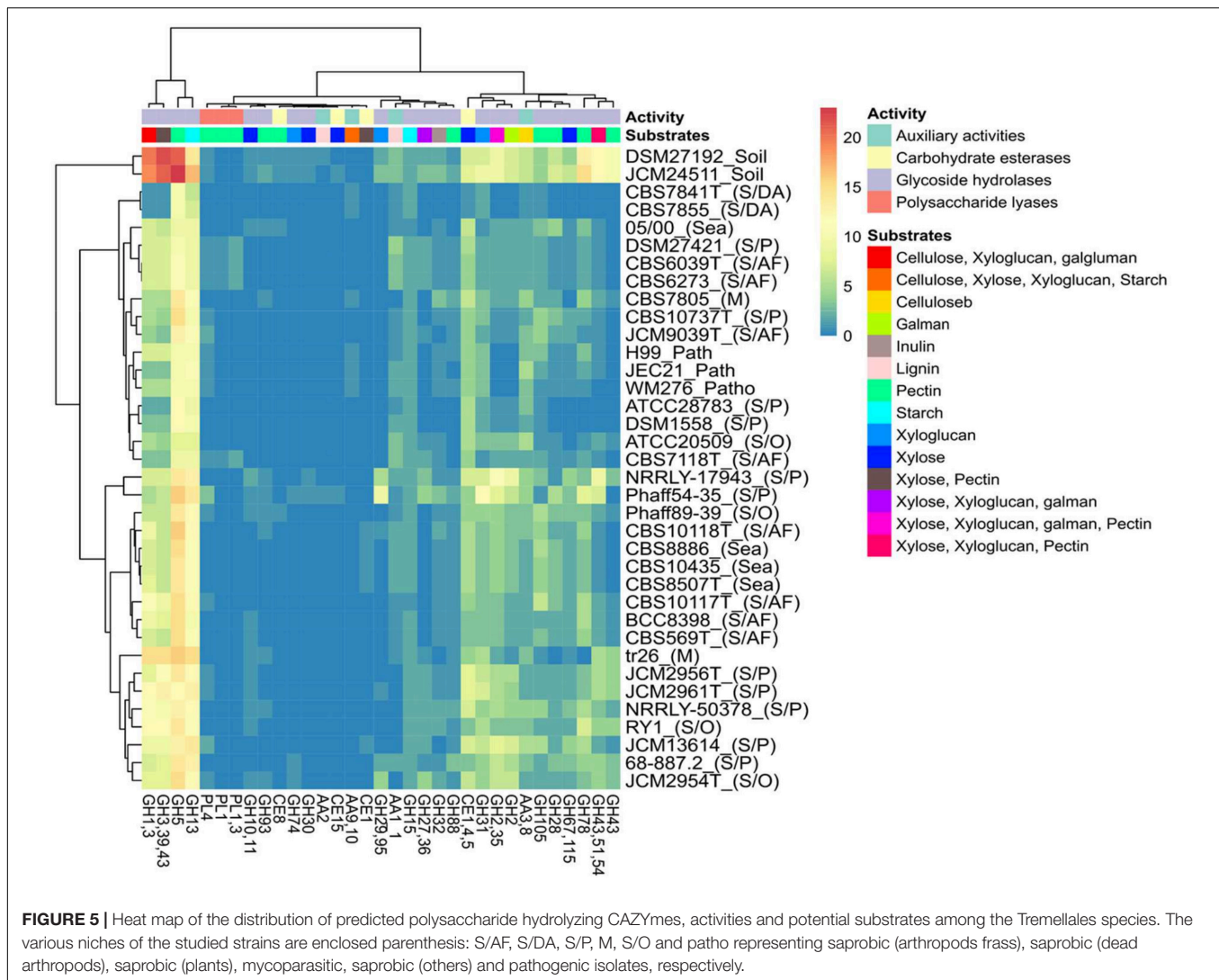


FIGURE 5 | Heat map of the distribution of predicted polysaccharide hydrolyzing CAZymes, activities and potential substrates among the Tremellales species. The various niches of the studied strains are enclosed parenthesis: S/AF, S/DA, S/P, M, S/O and patho representing saprobic (arthropods frass), saprobic (dead arthropods), saprobic (plants), mycoparasitic, saprobic (others) and pathogenic isolates, respectively.

lack thereof by JCM 24511 further supports the evolutionary divergence of the two strains as predicted via phylogenomic analyses. Aside these, the two soil isolates utilize more substrates in common and uniquely use D-Fucose compared to the rest of the strains tested (Figure 6). The studied strains grew on five of the eight polysaccharides included in this study and no visible growth was observed on cellulose, CMC, and chitin (crab shell). Five strains, including the two soil isolates showed growth on pectin and only three strains *K. mangroviensis* CBS 8507T, *D. aurantiaca* CBS 6980^T (JCM 2956^T) and *C. amyolentus* CBS 6039^T isolated from sea, plant, arthropods frass, respectively, grew on plates containing starch (corn).

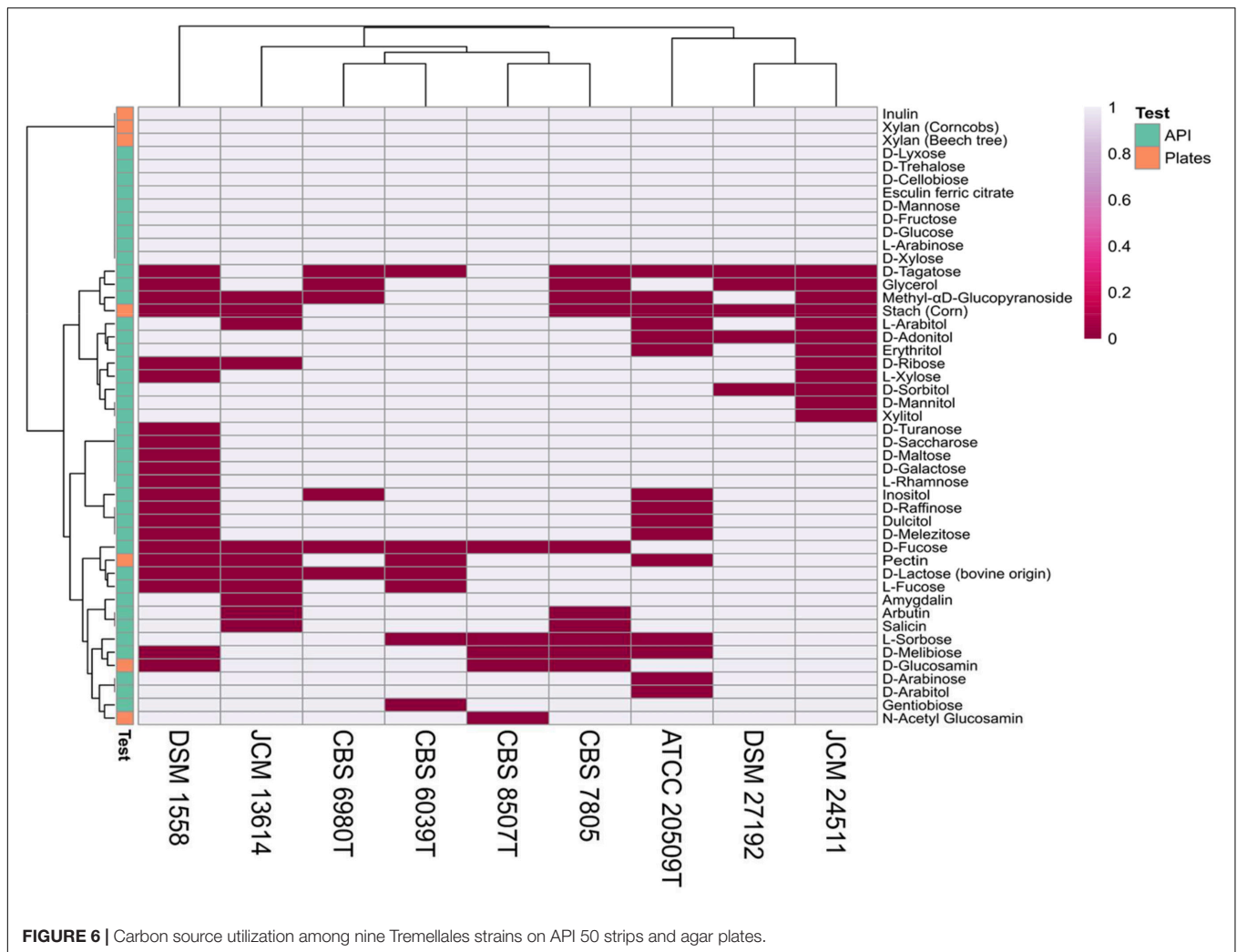
DISCUSSION

This study provides an in-depth evaluation of the genetic relationship, diversification and biomass biodegrading potential of the Tremellales. We previously described the genome sequence of the oleaginous strain DSM 27192 (Aliyu et al., 2019) and in

the current extended study, thirty-four other Tremellales have been incorporated with a view to further explore their genetic potentials. These isolates originated from a variety of habitats, potentially representing diverse niche specializations which could be explored for the identification of novel enzymes with wide ranges of biotechnological applications. Aside the diversity of isolation sources, the taxa included in this study (Table 1) are evolutionary diverse and based on a combination of genomics metrics the data provides additional insight into the application of phylogenomics for the delineation of these taxa.

Phylogenomics Provides Insight Into the Evolutionary Relationships of the Tremellales

The presented phylogeny generated from 1,597 single copy orthologous proteins (Figure 1) largely agrees with the delineation of the Tremellales within the *Tremellomyces* phylogenomic framework reported by Liu et al. (2015). Using taxa of known taxonomic standing, inferences were derived on the



interrelationships between closely related strains. For instance, based on phylogeny and percent “nucleotide” identity values, Passer et al. (2019) proposed that CBS 6039^T and CBS 6273 are conspecific (*C. amyloletus*) while *C. wingfieldii* CBS 7118^T and *C. floricola* DSM 27421 represent distinct species. Similarly, previous phylogenetic analysis of the *C. gattii/neoformans* species complex has proposed seven species with the three strains included in this study belonging to three distinct species (Hagen et al., 2015). The AAI value of 88.8% shared between the two closely related soil and oleagenic isolates (Schulze et al., 2014; Tanimura et al., 2014) DSM 27192 and JCM 24511, affiliated with *S. podzolica* based on D1/D2 region of LSU rRNA gene compares well with the interspecific AAI range of *C. gattii* WM276, *C. neoformans* var. *grubii* H99 and *neoformans* JEC21 (86.3–90.0%), as well as that of *C. amyloletus* CBS 6039^T, *C. wingfieldii* CBS 7118^T and *C. floricola* DSM 27421 (91.6–92.7%). This value is also lower than the intraspecific AAI of *C. amyloletus* species, CBS 6039^T and CBS 6273 (98.7%), *C. depauperatus* strains, CBS 7841^T, CBS 7855 (96.2%) and *Kwoniella heveanensis* CBS 569^T and BCC8398 (96.7%) and the three *K. mangrovensis* strains, CBS 8507^T, CBS 10435 and CBS 8886 (97.6%). Consequently,

DSM 27192 and JCM 24511 may represent two distinct species. In the same vein, ATCC 28783 and DSM 1558 affiliated to *Tremella mesenterica* (AAI: 92.7%) apparently constitute two distinct species. However, the *T. mesenterica* strains share higher average nucleotide identity (ANI) of 96.73% compared to 86.52% for the *Saitozyma podzolica* strains (**Supplementary Figure S1B**). The latter strains also show greater genetic dissimilarity relative to *Cryptococcus neoformans* complex (ANI: 88.31%) or *C. amyloletus* and *C. wingfieldii* (ANI: 92.26), further supporting the possibility that they represent two distinct species. In addition to phylogeny, AAI and ANI values, these strains differ in genome sizes, predicted proteomes and numbers of strain specific proteins. However, the absence of the genomes of the type strains of these taxa may exclude any further speculations regarding genome-based taxonomy for the strains. While genomic metrics like AAI (Konstantinidis and Tiedje, 2005b), ANI (Konstantinidis and Tiedje, 2005a; Jain et al., 2018) and *in silico* DNA–DNA (Meier-Kolthoff et al., 2013, 2014) show remarkable power in delineating microorganisms and correlate well with the conventional lab-based DNA–DNA hybridization, often used in fungal taxonomy (Lachance, 2018), their application is yet to gain

full traction in fungal taxonomy partly due to lack of baseline data that validate DNA–DNA reassociation used in fungal taxonomy (Takashima and Sugita, 2019).

Tremellales Harbor Abundant and Diverse Carbohydrate-Active and Proteolytic Enzymes

To leverage functional capabilities associated with the diversity and niche specializations of the Tremellales for the identification of enzymes of potential industrial and biotechnological applications, proteomes of the studied strains were scanned for carbohydrate-active and proteolytic enzymes using MEROPS and dbCAN database, respectively. Several studies have highlighted the potential applications of carbohydrate-active and proteolytic enzymes and in some instances, the two groups of enzymes have been reported to function complementarily, for example, in the biodegradation of plant and enhanced plant fiber digestibility in the diet of ruminant (Cabaleiro et al., 2002; da Silva, 2017).

Metallopeptidases and Cysteine Proteases Are the Most Abundant and Diverse Proteolytic Enzymes of Tremellales

The present data showed a strong correlation between peptidases and amino acid transporters both of which have been implicated in the regulation and metabolism of nitrogen compounds (Uemura et al., 2004; Abdel-Sater et al., 2011; Kohl et al., 2012; Shah et al., 2013). As previously suggested about nitrogen partitioning for grain filling (Kohl et al., 2012), elucidation of the functional roles of both the transporters and specific peptidases in fungi could have practical implication toward further development of nitrogen limited function like oil accumulation.

Inhibitors of peptidases are of interest because of their potential applications in various fields, including but not limited to medicine, crop protection and biotechnology (Sah et al., 2006; Sabotìè and Kos, 2012). They are widely distributed protein families that control the activity of peptidases. However, inhibitors of cysteine peptidases show restricted distribution in fungi (Sabotìè and Kos, 2012). The caspase (C14) inhibitor I32 belongs to a group of inhibitor of apoptosis proteins (IAP) which regulate a wide range of cellular activities (Dubin, 2005; Sah et al., 2006; Dunaevsky et al., 2014). Although, the antiapoptotic activity of I32 homologs has been demonstrated in *S. cerevisiae*, the inhibitory effects on C14 has not been proven (Walter et al., 2006; Owsianowski et al., 2008). The presence of C14 in the proteomes of all studied strains and restriction of I32 to four (mainly free living) strains may support the lack of association of the two proteins in fungi but instead, suggest a probable role in specific survival strategy in dynamic environments such as soil. Similarly, the precise function of the serine carboxypeptidase Y inhibitor (I51) remains largely obscure (Dunaevsky et al., 2014) but linked to various environmental stress responses such as heat shock and oxidative stress (Sabotìè and Kos, 2017). Unlike I32 and although, the homologs of the putative target S10 (serine carboxypeptidase) are prevalent in all the studied genomes, only 16 genomes, including those of the three sea

(*K. mangroviensis* CBS 10435, CBS 8507^T and CBS 8886) and one plant (*N. encephala* UCDFST 68-887.2) isolates, harbor the specific serine carboxypeptidase Y (S10.001: MER0002010) homologs inhibited by I51. Thus, I51 may play multiple roles including responses to environmental stress and specifically nutrient limitation (Parzych et al., 2018) among the strains harboring serine carboxypeptidase Y.

Among the predicted peptidases, two families, glutamic (G) and mixed (P) peptidases show restricted distribution with the former encoded on the genome of one soil isolate and the latter present in only four plant related isolates. Glutamic peptidases are among the well-studied fungal proteolytic enzymes to be identified and show restricted distribution among fungal taxa (da Silva, 2017). Specifically, the aspergilloglutamic peptidase (G01) first discovered in *A. niger* var. *macrosporus* and named eqolisins (Kataoka et al., 2005) has found various industrial applications, including a role as hydrolyzing agent against various inhibitors present in numerous components of animal feed (Bruins et al., 2018) and dietary supplement (Bruins et al., 2019). Thus, the prediction of a putative G01 protein in DSM 27192 highlights the biotechnological potential of this strain that requires further investigation. P1 or β -aminopeptidases include aminopeptidases and self-processing peptides that are of biotechnological and pharmaceutical interest for cleavage of β - and mixed α,β -peptides and amides. Orthologs of P1 has been identified in a wide range of prokaryotes and eukaryotes, including fungi (Heck et al., 2006; Hiraishi, 2016; John-White et al., 2017). However, the natural substrates for these peptidases are not known and therefore their restriction to only a few plant isolates could be associated with their potential capacity to use synthetic β -peptidase (John-White et al., 2019) and hence characterizing these proteins will expand the limited knowledge regarding their function.

Non-animal aspartic peptidases sources such as fungi are of special biotechnological interest as cheap alternative to chymosin obtained from the stomach of calves (Kumar et al., 2010; da Silva, 2017). They are mainly applied in food industry, for instance, in cheese processing (Kumar et al., 2010; Yegin et al., 2011). Prevalence of the A01A family, which is secreted aspartic proteases, among the studied isolates, especially those from arthropods frass may underpin the role of these enzymes in proteolytic activities associated with both general nutrients acquisition and perhaps pathogenesis among the pathogenic strains (Monod et al., 2002; Mandujano-Gonzalez et al., 2016).

Scanty information is available about the functional role of C-peptidases in fungi (da Silva, 2017). However, these enzymes have been associated with parasitism in several organisms, including fungi and hence considered to be of clinical importance (Barrett and Rawlings, 1996; Atkinson et al., 2009). For example, expression of C14 (caspase) has been reported in parasitic microorganisms including fungi (Atkinson et al., 2009; McLuskey and Mottram, 2015). However, the genomic data presented here showed similar distribution of these enzymes among all the studied strains.

M-peptidases such as matrix metallopeptidases (MMPs) have potential applications in cancer and arthritis therapy because of their ability to degrade extracellular matrix (Rao et al., 1998).

However, orthologs of MMPs have been reported to occur rarely in fungi, strictly phylum Ascomycota and no activity data are available (Cerdà-Costa and Gomis-Rüth, 2014; Marino-Puertas et al., 2017). M-peptidases have also been implicated as virulence factors in various organisms (Gravi et al., 2012). However, this study could not establish enrichment of any of these peptidases in either the pathogenic or the parasitic relative to the saprobic isolates.

Although shown to be among the most prevalent peptidases (da Silva, 2017), this study revealed that serine peptidases are less abundant and less diverse compared to the metallopeptidases and cysteine proteases. Consistent with previous studies (Muszewska et al., 2017), serine peptidase abundance appears to be significantly associated with proteome size (**Supplementary Figures S5A,B**). However, contrary to previous reports where plant associated taxa have been shown to harbor more orthologs of S9, S10, S12, S33, and S53 (Muszewska et al., 2017), this study revealed predominance of orthologs of S9 among the isolates obtained from soil (JCM 24511 and DSM 27192), rotten beech (JCM 13614) and a fungus (tr26) relative to isolates from plants and insect frass (**Supplementary Figure S3C**). Furthermore, in contrast to previous reports (Muszewska et al., 2017; Palmer et al., 2018) the distribution of S8 (subtilisin) does not appear to be shaped by niche specialization among the studied Tremellales.

Previous reports have indicated that T-peptidases are more abundant among aquatic and soil inhabiting Bacteroidetes (Nguyen et al., 2019). The fungi studied here, however, appear to show similar distribution of T-peptidases.

Carbohydrate-Active Enzymes Enzyme (CAZyme) Repertoire and Substrate Utilization Among Thirty-Five Tremellales

Evaluation of the predicted proteomes of the thirty-five Tremellales isolated from different environments revealed greater CAZymes abundance among the soil isolates similar to numbers reported from previous studies of representative fungi, where more than 50% of the strains harbor >300 CAZymes (Zhao et al., 2014). However, CAZymes distribution among the majority (thirty-three) of the studied strains falls within the lower 50% range and as observed previously, modes of life do not appear to be associated with abundance or diversity of CAZymes (Zhao et al., 2014). Further reflecting their versatility, the soil isolates also harbor greater number of sugar transporters (STs). STs distribution have been correlated with fungal capability to utilize carbon sources with greater numbers of STs typically observed among metabolically versatile fungi capable of using broader range of substrates (Cornell et al., 2007). In addition to sugar transduction, certain STs also regulate the expression of CAZymes such as cellulase and xylanases (Zhang et al., 2013; Huang et al., 2015) and have therefore been considered veritable targets for strain improvement in various biotechnological applications (Borin et al., 2017).

Although CBMs are most abundant among the soil isolates, majority of the proteins containing these modules do not have CAZymes modules. CAZymes without CBMs have been

reported to show poor catalytic activities relative to their CBMs associated counterparts (Pasari et al., 2017). To enhance catalytic activity of certain CAZymes lacking CBMs, the latter modules could be engineered in the enzymes. For instance, improved hydrolytic activities have been detected with CBM augmented *M. albomyces* cellulases (Szijártó et al., 2008). Non-hydrolytic proteins containing CBMs, however, enhance substrate degradation by participating in a multienzyme complex called cellulosome whose catalytic efficiency is greatly reduced without these proteins (Shoseyov et al., 2006). CBMs are important targets of enzyme engineering for enhanced hydrolytic activities via the construction of hybrid enzymes (Kim et al., 1998; Limón et al., 2001; Shoseyov et al., 2006) or hydrolytic scaffolds such as the cellulosome. Detection of CBMs in proteins with no known CAZyme modules could potentially signify unknown hydrolytic proteins.

The two *Saitozyma* isolates from soil also harbor more proteins linked to auxiliary activities, probably indicating a greater lignin-degrading capacity (Whittaker et al., 1996). However, some of the specific AA families such as AA2 which is associated with lignin modification (Lundell et al., 2010), occur uniquely, albeit in single copies, in two of the studied strains, signifying their putative specialty in plant biomass degradation. In contrast, AA9 associated with cellulose and hemicellulose utilization (Hemsworth et al., 2013; Riley et al., 2014), has been identified across pathogenic, symbiotic and saprobic fungi. However, involvement of AA9 in host invasion has been suggested in *Pyrenochaeta lycopersici* (Valente et al., 2011) and hence a link to this trait among these strains could be speculated. Similarly, AA8 (iron reductases) is exclusively present in the pathogenic, soil and four saprobic isolates. The pathogenic isolates also have two copies of AA6 (1,4-benzoquinone reductase) along with one arthropods frass strain while six other saprobic strains have only single copies of AA6. AA8 is suggested to be involved in the production of reactive hydroxyl radical potentially associated with non-enzymatic breakages of cellulose and hence a component of plant biomass degradation proteins while AA6 is involved in degradation of aromatic compounds intracellularly and serve as protection against reactive quinone compounds in fungi (Levasseur et al., 2013).

Aside CE4, the largest family of CEs for which known activities include acetyl xylan esterases, chitin deacetylase, peptidoglycan GlcNAc deacetylase, among others (Nakamura et al., 2017), that is shared by all the studied strains, the studied proteomes showed variation in terms of CEs distribution and potential biotechnological potentials. For instance, the recently discovered CE15 occurs in only two strains and has the unique ability to cleave the ester bond between lignin and glucuronic acid residues of glucuronoxylans (Arnling Bååth et al., 2016; Mosbech et al., 2018). However, most of the CEs identified in this study have been reported to show wide range of substrate specificity. For example, most CE10 CAZymes do not act on carbohydrates substrates (Zhao et al., 2014). Despite this, the diversity of the CEs represents enormous potential in various fields, for instance, CE1, present in four strains, includes enzymes of potential biomedical application as drug design targets (Nakamura et al., 2017).

Because of their important role in breaking down plant cell walls, GHs are among the most characterized enzymes of fungal origin (Murphy et al., 2011). Majority of GHs shared by all the studied Tremellales are among those reported to be widely distributed in fungi and the overrepresentation of GH2, GH13, and GH16 is also consistent with reports on the top most abundant fungal GHs (Zhao et al., 2014). The current study showed varied distribution and abundance of GH among the thirty-five studied Tremellales and different life styles with the two soil isolates harboring greater numbers but less diverse GHs relative to two plant associated strains, thus indicating that the latter group may have greater substrate range. Interestingly, GH30, GH67, and GH125 shared by the four (soil and plant) isolates occur rarely among previously studied *Basidiomycota* with GH67 only found among the Ascomycetes (Zhao et al., 2014).

Despite prediction of CAZYmes repertoire that could potentially act on numerous substrates (Figure 5), the nine tested isolates could not utilize cellulose (carboxymethylcellulose; CMC) and chitin. GH18, GH19, and GH85 and AA10 are chitinases involved in the degradation of chitin (Lombard et al., 2014; Berlemont, 2017). The inability of these strains to utilize chitin may be explained by the absence of the complete set of chitinases as only GH18 orthologs are present in the proteome of all tested strains. Of the three cellulose active GHs sets (Coutinho et al., 2009; van den Brink and de Vries, 2011), all nine strains lacked β -1,4-endoglucanase (GH5, GH7, GH12 and GH45) and cellobiohydrolase (GH6 and GH7) but harbor the orthologs of β -1,4-glucosidase (GH1 and GH3). In addition, *S. intermedium* CBS 7805 harbors AA9, a module known for its cellulolytic activity (Langston et al., 2011). It is not surprising, however, that these strains could not utilize this substrate since GH1, GH3, and AA9 alone may not be adequate for cellulose deconstruction as a combined action of several cellulases is essential for the complete breakdown of this polysaccharide (Singh et al., 2017). Several GH and AA proteins identified in plant biomass-degrading fungi, including AA9 have been reported to contain CBMs (Várnai et al., 2014; Berlemont, 2017). For instance, CBM1, which is cellulose targeting module has been reported occur at least once in ~79% of 1,425 multi-domain CAZYmes (Berlemont, 2017). However, majority of AA9 proteins of several plant biomass-degrading fungi do not harbor CBM1 (Várnai et al., 2014). Regardless, the observed lack cellulolytic activity in all nine strains, including CBS 7805 may also be associated with the absence of these modules.

On the other hand, the proteomes of the nine strains contain numerous CAZYmes (Figure 5) associated with xylan decomposition (Coutinho et al., 2009; van den Brink and de Vries, 2011; Nagy et al., 2015), including acetyl xylan esterase (CE1,4,5), α -L-arabinofuranosidase (GH43,51,54) and β -1,4-D-xylosidase (GH3,39,43) that are present in all the isolates. By contrast, only three of the nine tested strains harbor β -1,4-D-Endoxylanase (GH10,11) (Coutinho et al., 2009; van den Brink and de Vries, 2011; Nagy et al., 2015). Despite these differences, all strains utilized the two xylan sources (beech tree and corncobs) as only carbon source. All strains were also able to grow on inulin.

However, the current analysis did not predict inulinase (GH32) in two strains, DSM 1558 and JCM 13614.

Contrary to the above two scenarios, the proteomes of all tested strains (Figure 5) include orthologs of unsaturated rhamnogalacturonan hydrolase (GH105), β -1,4-D-galactosidase (GH2,35), β -1,4-D-xylosidase (GH3,39,43) and β -1,6-endogalactanase (GH5) and except DSM 1558 (no growth on pectin) orthologs of endopolygalacturonase or exopolygalacturonase or rhamnogalacturonan galacturonohydrolase or rhamnogalacturonan hydrolase (GH28), α -L-arabinofuranosidase (GH43,51,54) and α -rhamnosidase (GH78), all of which partake in the disintegration of pectin (van den Brink and de Vries, 2011; Nagy et al., 2015). Despite the range of pectinolytic enzymes, only five of the nine strains showed growth on pectin. Although, CE1 and pectin methyl esterase (CE8) and PLs are known to participate in degradation of pectin components (van den Brink and de Vries, 2011), strains lacking one or both enzymes grew on pectin. Similarly, all tested strains harbour α -amylase (GH13), α -1,4-D-glucosidase/ α -D-xylosidase (GH31) and glucoamylase (GH15), a set of CAZYmes known to hydrolyze starch (van den Brink and de Vries, 2011; Nagy et al., 2015). However, majority of the strains sharing similar starch-degrading enzymes to those in isolates capable of hydrolyzing starch could not grow on it. The inability of strains harboring sets of polysaccharides hydrolyzing enzymes to breakdown the substrate may be associated with poor substrate binding capacity of the enzymes or low rate of enzyme activity which could also be shaped by numerous factors including growth conditions. Subsequent work will therefore focus on characterization of a selection of the predicted proteolytic and carbohydrate active enzymes from the soil inhabiting *Saitozyma* strains with emphasis on enzymes of potential novelty both in terms of structure and perhaps function.

CONCLUSION

Considering pending economical, societal, and environmental challenges regarding the reduction of greenhouse gas emissions and the highly controversial “food-or-fuel” debate, it is highly important to enable competitive processing of renewable alternatives over fossil oil based or non-sustainable products. This can be achieved via identification and deployment of biomass deconstruction enzymes to generate the raw materials to produce biofuel and other industrial chemicals. This study aimed to provide an insight into the evolutionary relationships and functional diversification, with emphasis on biomass deconstruction capabilities, among thirty-five members of the order Tremellales. Contributing to the debate on genome-based circumscription of fungal isolates, our work revealed evolutionary distinction of two closely related, soil and oleagenic strains (*Saitozyma* species) both at genomic and phenotypic levels. We also identified 6,918 putative CAZYmes and 7,066 peptidases belonging to various families of the enzymes. The *Saitozyma* species, which have been isolated from soil harbor the largest numbers of both CAZYmes and peptidases. Although, the soil isolates show greater abundance of these feature, the limited

number of soil isolates, and the lack of obvious trend among taxa of plant and arthropods origin, indicates that the distribution of these enzymes may not be associated with specific habitat type. Our future work includes isolation and characterization of the predicted biomass degrading enzymes from available fungal isolates using a range of analytics with a view to develop a robust enzyme cocktails with potential biotechnological applications.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found in the NCBI or JGI databases: ASM18594v1, CNA3, ASM9104v1, ASM635230v1, ASM614915 v1, Cryp_amyl_CBS6039_V3, Cryp_amyl_CBS6273_V2, Filo_depa_CBS7841_V1, Filo_depa_CBS7855_V2, Cryp_heve_BC C8398_V1, Cryp_heve_CBS569_V2, Cryp_deje_CBS10117_V1, Cryp_pinu_CBS10737_V1, Cryp_best_CBS10118_V1, Kwon_mang_CBS10435_V2, Kwon_mang_CBS8507_V2, Kwon_mang_CBS8886_V1, ASM394221v1, JCM_24511_assembly_v001, JCM_2961_assembly_v001, JCM_2956_assembly_v001, Trem_mese_ATCC28783_V1, Tremel, ASM98790v1, Phaff89-39v1.0, Treen1, Phaff54-35v1.0, Kocim1, JCM_2954_assembly_v001, CBS7805v1.0, Cf_30_300r_Split10plusN, ASM73882v1, JCM_90_39_assembly_v001, JCM_13614_assembly_v001, Mo29, ASM17_1244v1.

REFERENCES

- Abdel-Sater, F., Jean, C., Merhi, A., Vissers, S., and André, B. (2011). Amino acid signaling in yeast: activation of Ssy5 protease is associated with its phosphorylation-induced ubiquitylation. *J. Biol. Chem.* 286, 12006–12015. doi: 10.1074/jbc.M110.200592
- Aliyu, H., Gorte, O., Neumann, A., and Ochsenreither, K. (2019). Draft genome sequence of the oleaginous yeast *Saitozyma podzolica* (syn. *Cryptococcus podzolicus*. *Microbiol. Resour. Announc.* 8:e01676-18.
- Arnling Bääth, J., Giummarella, N., Klaubauf, S., Lawoko, M., and Olsson, L. (2016). A glucuronoyl esterase from *Acremonium alcalophilum* cleaves native lignin-carbohydrate ester bonds. *FEBS Lett.* 590, 2611–2618. doi: 10.1002/1873-3468.12290
- Atkinson, H. J., Babbitt, P. C., and Sajid, M. (2009). The global cysteine peptidase landscape in parasites. *Trends Parasitol.* 25, 573–581. doi: 10.1016/j.pt.2009.09.006
- Balat, M., and Ayar, G. (2005). Biomass energy in the world, use of biomass and potential trends. *Energy Sour.* 27, 931–940. doi: 10.1080/00908310490449045
- Barrett, A. J., and Rawlings, N. D. (1996). Families and clans of cysteine peptidases. *Perspect. Drug Discov. Design* 6, 1–11. doi: 10.1007/bf02174042
- Berlemont, R. (2017). Distribution and diversity of enzymes for polysaccharide degradation in fungi. *Sci. Rep.* 7:222. doi: 10.1038/s41598-017-00258-w
- Bilal, M., Asgher, M., Iqbal, H. M. N., Hu, H., and Zhang, X. (2017). Biotransformation of lignocellulosic materials into value-added products—a review. *Intern. J. Biol. Macromol.* 98, 447–458. doi: 10.1016/j.ijbiomac.2017.01.133
- Borin, G. P., Sanchez, C. C., de Santana, E. S., Zanini, G. K., dos Santos, R. A. C., de Oliveira Pontes, A., et al. (2017). Comparative transcriptome analysis reveals different strategies for degradation of steam-exploded sugarcane bagasse by *Aspergillus niger* and *Trichoderma reesei*. *BMC Genom.* 18:501. doi: 10.1186/s12864-017-3857-5
- Bruins, M. J., Edens, L., and Nan, H. M. (2018). *Use of Aspergillus niger Aspergilloglutamyl Peptidase To Improve Animal Performance*. Denmark: Novozymes.

AUTHOR CONTRIBUTIONS

HA, OG, AN, and KO co-conceived the study and designed the experiments. HA, OG, XZ, and KO, performed the experiments and analyzed the data. HA, OG, and KO drafted the manuscript. All authors contributed to the final manuscript.

FUNDING

HA acknowledged the funding from Alexander von Humboldt Foundation. Bioeconomy International BMBF (grant #031B0452) supported OG.

ACKNOWLEDGMENTS

We acknowledge support by Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of Karlsruhe Institute of Technology.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2020.00226/full#supplementary-material>

- Bruins, M. J., Edens, L., and Nan, H. M. (2019). *Medicament and Method For Treating Innate Immune Response Diseases*. Netherlands: DSM IP Assets BV.
- Cabaleiro, D. R., Rodriguez-Couto, S., Sanromán, A., and Longo, M. A. (2002). Comparison between the protease production ability of ligninolytic fungi cultivated in solid state media. *Process Biochem.* 37, 1017–1023. doi: 10.1016/S0032-9592(01)00307-7
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552.
- Cerdà-Costa, N., and Gomis-Rüth, F. X. (2014). Architecture and function of metallopeptidase catalytic domains. *Protein Sci.* 23, 123–144. doi: 10.1002/pro.2400
- Chapman, J., Ismail, A. E., and Dinu, C. Z. (2018). Industrial applications of enzymes: recent advances, techniques, and outlooks. *Catalysts* 8:238.
- Chernomor, O., von Haeseler, A., and Minh, B. Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* 65, 997–1008. doi: 10.1093/sysbio/syw037
- Cornell, M. J., Alam, I., Soanes, D. M., Wong, H. M., Hedeler, C., Paton, N. W., et al. (2007). Comparative genome analysis across a kingdom of eukaryotic organisms: specialization and diversification in the fungi. *Genome Res.* 17, 1809–1822.
- Coutinho, P. M., Andersen, M. R., Kolenova, K., Benoit, I., Gruben, B. S., Trejo-Aguilar, B., et al. (2009). Post-genomic insights into the plant polysaccharide degradation potential of *Aspergillus nidulans* and comparison to *Aspergillus niger* and *Aspergillus oryzae*. *Fungal Genet. Biol.* 46, S161–S169.
- da Silva, R. R. (2017). Bacterial and fungal proteolytic enzymes: production, catalysis and potential applications. *Appl. Biochem. Biotechnol.* 183, 1–19. doi: 10.1007/s12010-017-2427-2
- de Paula, R. G., Antoniêto, A. C. C., Ribeiro, L. F. C., Srivastava, N., O'Donovan, A., Mishra, P. K., et al. (2019). Engineered microbial host selection for value-added bioproducts from lignocellulose. *Biotechnol. Adv.* 37:107347. doi: 10.1016/j.biotechadv.2019.02.003
- De Schouwer, F., Claes, L., Vandekerckhove, A., Verduyck, J., and De Vos, D. E. (2019). Protein-rich biomass waste as a resource for future biorefineries: state

- of the art, challenges, and opportunities. *Chemoschem* 12, 1272–1303. doi: 10.1002/cssc.201802418
- Dubin, G. (2005). Proteinaceous cysteine protease inhibitors. *Cell. Mol. Life Sci.* 62:653. doi: 10.1007/s00018-004-4445-9
- Dunaevsky, Y. E., Popova, V. V., Semenova, T. A., Beliakova, G. A., and Belozersky, M. A. (2014). Fungal inhibitors of proteolytic enzymes: classification, properties, possible biological roles, and perspectives for practical use. *Biochimie* 101, 10–20. doi: 10.1016/j.biochi.2013.12.007
- Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157. doi: 10.1186/s13059-015-0721-2
- Findley, K., Rodriguez-Carres, M., Metin, B., Kroiss, J., Fonseca, Á, Vilgalys, R., et al. (2009). Phylogeny and phenotypic characterization of pathogenic *Cryptococcus* species and closely related saprobic taxa in the *Tremellales*. *Eukaryot. Cell* 8, 353–361.
- Gravi, E. T., Paschoalin, T., Dias, B. R., Moreira, D. F., Belizario, J. E., Oliveira, V., et al. (2012). Identification of a metallopeptidase with TOP-like activity in *Paracoccidioides brasiliensis*, with increased expression in a virulent strain. *Med. Mycol.* 50, 81–90. doi: 10.3109/13693786.2011.590825
- Hagen, F., Khayhan, K., Theelen, B., Kolecka, A., Polacheck, I., Sionov, E., et al. (2015). Recognition of seven species in the *Cryptococcus gattii/Cryptococcus neoformans* species complex. *Fungal Genet. Biol.* 78, 16–48. doi: 10.1016/j.fgb.2015.02.009
- Harris, P. J., and Stone, B. A. (2009). “Chemistry and molecular organization of plant cell walls,” in *Biomass Recalcitrance: Deconstructing The Plant Cell Wall For Bioenergy*, ed. M. E. Himmel (Hoboken, NJ: Wiley), 61–93.
- Heck, T., Limbach, M., Geueke, B., Zacharias, M., Gardiner, J., Kohler, H.-P. E., et al. (2006). Enzymatic degradation of β - and Mixed α,β -Oligopeptides. *Chem. Biodiver.* 3, 1325–1348. doi: 10.1002/cbdv.200690136
- Hemsworth, G. R., Davies, G. J., and Walton, P. H. (2013). Recent insights into copper-containing lytic polysaccharide mono-oxygenases. *Curr. Opin. Struct. Biol.* 23, 660–668.
- Hiraishi, T. (2016). Poly(aspartic acid) (PAA) hydrolases and PAA biodegradation: current knowledge and impact on applications. *Appl. Microbiol. Biotechnol.* 100, 1623–1630. doi: 10.1007/s00253-015-7216-7
- Huang, Z.-B., Chen, X.-Z., Qin, L.-N., Wu, H.-Q., Su, X.-Y., and Dong, Z.-Y. (2015). A novel major facilitator transporter TrSTR1 is essential for pentose utilization and involved in xylanase induction in *Trichoderma reesei*. *Biochem. Biophys. Res. Commun.* 460, 663–669.
- Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9
- John-White, M., Dumsday, G. J., Johanesen, P., Lyras, D., Drinkwater, N., and McGowan, S. (2017). Crystal structure of a β -aminopeptidase from an Australian *Burkholderia* sp. *Acta Crystallographica. Sec. F Struct. Biol. Commun.* 73(Pt 7), 386–392. doi: 10.1107/S2053230X17007737
- John-White, M., Gardiner, J., Johanesen, P., Lyras, D., and Dumsday, G. (2019). β -Aminopeptidases: insight into enzymes without a known natural substrate. *Appl. Environ. Microbiol.* 85, e00318–e00319. doi: 10.1128/aem.00318-19
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Kai, D., Chow, L. P., and Loh, X. J. (2018). Lignin and its properties. *Funct. Mater. Lignin* 3:1.
- Kataoka, Y., Takada, K., Oyama, H., Tsunemi, M., James, M. N. G., and Oda, K. (2005). Catalytic residues and substrate specificity of scytalidoglutamic peptidase, the first member of the eqolisin in family (G1) of peptidases. *FEBS Lett.* 579, 2991–2994. doi: 10.1016/j.febslet.2005.04.050
- Kim, H., Goto, M., Jeong, H.-J., Jung, K. H., Kwon, H., and Furukawa, K. (1998). Functional analysis of a hybrid endoglucanase of bacterial origin having a cellulose binding domain from a fungal exoglucanase. *Appl. Biochem. Biotechnol.* 75:193. doi: 10.1007/BF02787774
- Kohl, S., Hollmann, J., Blattner, F. R., Radchuk, V., Andersch, F., Steuernagel, B., et al. (2012). A putative role for amino acid permeases in sink-source communication of barley tissues uncovered by RNA-seq. *BMC Plant Biol.* 12:154. doi: 10.1186/1471-2229-12-154
- Konstantinidis, K. T., and Tiedje, J. M. (2005a). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409272102
- Konstantinidis, K. T., and Tiedje, J. M. (2005b). Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* 187, 6258–6264. doi: 10.1128/JB.187.18.6258-6264.2005
- Kumar, A., Grover, S., Sharma, J., and Batish, V. K. (2010). Chymosin and other milk coagulants: sources and biotechnological interventions. *Crit. Rev. Biotechnol.* 30, 243–258. doi: 10.3109/07388551.2010.483459
- Lachance, M.-A. (2018). C. P. Kurtzman’s evolving concepts of species, genus and higher categories. *FEMS Yeast Res.* 18:foy103. doi: 10.1093/femsyr/foy103
- Langston, J. A., Shaghasi, T., Abbate, E., Xu, F., Vlasenko, E., and Sweeney, M. D. (2011). Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Appl. Environ. Microbiol.* 77, 7007–7015. doi: 10.1128/aem.05815-11
- Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., and Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* 6, 41–41. doi: 10.1186/1754-6834-6-41
- Limón, M. C., Margolles-Clark, E., Benítez, T. A., and Penttilä, M. (2001). Addition of substrate-binding domains increases substrate-binding capacity and specific activity of a chitinase from *Trichoderma harzianum*. *FEMS Microbiol. Lett.* 198, 57–63. doi: 10.1111/j.1574-6968.2001.tb10619.x
- Liu, X. Z., Wang, Q. M., Theelen, B., Groenewald, M., Bai, F. Y., and Boekhout, T. (2015). Phylogeny of tremellomycetous yeasts and related dimorphic and filamentous basidiomycetes reconstructed from multiple gene sequence analyses. *Stud. Mycol.* 81, 1–26. doi: 10.1016/j.simyco.2015.08.001
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178
- Love, J., Palmer, J., Stajich, J., Esser, T., Kastman, E., and Winter, D. (2018). nextgenusfun/nextgenusfun: funannotate v1.5.0 (Version 1.5.0). *Zenodo*. doi: 10.5281/zenodo.1342272
- Lundell, T. K., Mäkelä, M. R., and Hildén, K. (2010). Lignin-modifying enzymes in filamentous basidiomycetes – ecological, functional and phylogenetic review. *J. Basic Microbiol.* 50, 5–20. doi: 10.1002/jobm.200900338
- Magis, C., Taly, J.-F., Bussotti, G., Chang, J.-M., Di Tommaso, P., Erb, I., et al. (2014). T-coffee: tree-based consistency objective function for alignment evaluation. *Multip. Sequen. Alignm. Methods* 1079, 117–129.
- Mandujano-Gonzalez, V., Villa-Tanaca, L., Anducho-Reyes, M. A., and Mercado-Flores, Y. (2016). Secreted fungal aspartic proteases: a review. *Revist. Iberoam. Micol.* 33, 76–82.
- Marino-Puertas, L., Goulas, T., and Gomis-Rüth, F. X. (2017). Matrix metalloproteinases outside vertebrates. *Biochim. Biophys. Acta Mol. Cell Res.* 1864, 2026–2035. doi: 10.1016/j.bbamcr.2017.04.003
- May, R. C., Stone, N. R. H., Wiesner, D. L., Bicanic, T., and Nielsen, K. (2016). *Cryptococcus*: from environmental saprophyte to global pathogen. *Nat. Rev. Microbiol.* 14, 106–117. doi: 10.1038/nrmicro.2015.6
- McLuskey, K., and Mottram, J. C. (2015). Comparative structural analysis of the caspase family with other clan CD cysteine peptidases. *Biochem. J.* 466, 219–232. doi: 10.1042/BJ20141324
- Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60
- Meier-Kolthoff, J. P., Klenk, H.-P., and Göker, M. (2014). Taxonomic use of DNA G+C content and DNA–DNA hybridization in the genomic age. *Intern. J. Syst. Evolut. Microbiol.* 64, 352–356.
- Metsalu, T., and Vilo, J. (2015). ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Res.* 43, W566–W570. doi: 10.1093/nar/gkv468
- Millanes, A. M., Diederich, P., Ekman, S., and Wedin, M. (2011). Phylogeny and character evolution in the jelly fungi (*Tremellomycetes*, *Basidiomycota*, *Fungi*). *Mol. Phylogenet. Evolut.* 61, 12–28. doi: 10.1016/j.ympev.2011.05.014
- Mohanta, T. K., and Bae, H. (2015). The diversity of fungal genome. *Biol. Proced. Online* 17:8. doi: 10.1186/s12575-015-0020-z

- Monod, M., Capoccia, S., Léchenne, B., Zaugg, C., Holdom, M., and Jousson, O. (2002). Secreted proteases from pathogenic fungi. *Intern. J. Med. Microbiol.* 292, 405–419. doi: 10.1078/1438-4221-00223
- Mosbach, C., Holck, J., Meyer, A. S., and Agger, J. W. (2018). The natural catalytic function of CuGE glucuronoyl esterase in hydrolysis of genuine lignin-carbohydrate complexes from birch. *Biotechnol. Biofuels* 11:71. doi: 10.1186/s13068-018-1075-2
- Murphy, C., Powlowski, J., Wu, M., Butler, G., and Tsang, A. (2011). Curation of characterized glycoside hydrolases of fungal origin. *Database* 2011:bar020. doi: 10.1093/database/bar020
- Muszewska, A., Stepniewska-Dziubinska, M. M., Steczkiewicz, K., Pawlowska, J., Dziedzic, A., and Ginalski, K. (2017). Fungal lifestyle reflected in serine protease repertoire. *Sci. Rep.* 7:9147. doi: 10.1038/s41598-017-09644-w
- Nagy, L. G., Riley, R., Tritt, A., Adam, C., Daum, C., Floudas, D., et al. (2015). Comparative genomics of early-diverging mushroom-forming fungi provides insights into the origins of lignocellulose decay capabilities. *Mol. Biol. Evol.* 33, 959–970. doi: 10.1093/molbev/msv337
- Nakamura, A. M., Nascimento, A. S., and Polikarpov, I. (2017). Structural diversity of carbohydrate esterases. *Biotechnol. Res. Innov.* 1, 35–51. doi: 10.1016/j.biori.2017.02.001
- Nguyen, T. T. H., Myrold, D. D., and Mueller, R. S. (2019). Distributions of extracellular peptidases across prokaryotic genomes reflect phylogeny and habitat. *Front. Microbiol.* 10:413. doi: 10.3389/fmicb.2019.00413
- Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217.
- Owsianowski, E., Walter, D., and Fahrenkrog, B. (2008). Negative regulation of apoptosis in yeast. *Biochim. Biophys. Acta Mol. Cell Res.* 1783, 1303–1310.
- Palmer, J. M., Drees, K. P., Foster, J. T., and Lindner, D. L. (2018). Extreme sensitivity to ultraviolet light in the fungal pathogen causing white-nose syndrome of bats. *Nat. Commun.* 9:35. doi: 10.1038/s41467-017-02441-z
- Parzych, K. R., Ariosa, A., Mari, M., and Klionsky, D. J. (2018). A newly characterized vacuolar serine carboxypeptidase, Atg42/Ybr139w, is required for normal vacuole function and the terminal steps of autophagy in the yeast *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 29, 1089–1099.
- Pasari, N., Adlakha, N., Gupta, M., Bashir, Z., Rajacharya, G. H., Verma, G., et al. (2017). Impact of Module-X2 and carbohydrate binding module-3 on the catalytic activity of associated glycoside hydrolases towards plant biomass. *Sci. Rep.* 7:3700. doi: 10.1038/s41598-017-03927-y
- Passer, A. R., Coelho, M. A., Billymyre, R. B., Nowrousian, M., Mittelbach, M., Yurkov, A. M., et al. (2019). Genetic and genomic analyses reveal boundaries between species closely related to *Cryptococcus* pathogens. *mBio* 10:e00764. doi: 10.1128/mBio.00764-19
- Peterson, B. G., Carl, P., Boudt, K., Bennett, R., Ulrich, J., Zivot, E., et al. (2018). *Package 'PerformanceAnalytics'*. Vienna: R Team Cooperation.
- Rao, M. B., Tanksale, A. M., Ghatge, M. S., and Deshpande, V. V. (1998). Molecular and biotechnological aspects of microbial proteases. *Microbiol. Mol. Biol. Rev.* 62, 597–635.
- Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., and Finn, R. D. (2017). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 46, D624–D632. doi: 10.1093/nar/gkx1134
- Riley, R., Salamov, A. A., Brown, D. W., Nagy, L. G., Floudas, D., Held, B. W., et al. (2014). Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9923–9928.
- Rosnow, J. J., Anderson, L. N., Nair, R. N., Baker, E. S., and Wright, A. T. (2017). Profiling microbial lignocellulose degradation and utilization by emergent omics technologies. *Crit. Rev. Biotechnol.* 37, 626–640.
- Rytioja, J., Hildén, K., Yuzon, J., Hatakka, A., de Vries, R. P., and Mäkelä, M. R. (2014). Plant-polysaccharide-degrading enzymes from basidiomycetes. *Microbiol. Mol. Biol. Rev.* 78, 614–649. doi: 10.1128/mmlbr.00035-14
- Sabotie, J., and Kos, J. (2012). Microbial and fungal protease inhibitors—current and potential applications. *Appl. Microbiol. Biotechnol.* 93, 1351–1375. doi: 10.1007/s00253-011-3834-x
- Sabotie, J., and Kos, J. (2017). “Fungal protease inhibitors,” in *Fungal Metabolites Reference Series in Phytochemistry*, eds J. M. Mérillon, and K. Ramawat (Cham: Springer), 853–885.
- Sah, N. K., Khan, Z., Khan, G. J., and Bisen, P. S. (2006). Structural, functional and therapeutic biology of survivin. *Cancer Lett.* 244, 164–171. doi: 10.1016/j.canlet.2006.03.007
- Schmidt, H. A., Minh, B. Q., von Haeseler, A., and Nguyen, L.-T. (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Schulze, I., Hansen, S., Großhans, S., Rudsuck, T., Ochsenreither, K., Syltatk, C., et al. (2014). Characterization of newly isolated oleaginous yeasts - *Cryptococcus podzolicus*, *Trichosporon porosum* and *Pichia segobiensis*. *AMB Express* 4:24. doi: 10.1186/s13568-014-0024-0
- Shah, F., Rineau, F., Canbäck, B., Johansson, T., and Tunlid, A. (2013). The molecular components of the extracellular protein-degradation pathways of the ectomycorrhizal fungus *Paxillus involutus*. *New Phytol.* 200, 875–887. doi: 10.1111/nph.12425
- Sheridan, C. (2013). *Big Oil Turns On Biofuels*. Berlin: Nature Publishing Group.
- Shoseyov, O., Shani, Z., and Levy, I. (2006). Carbohydrate binding modules: biochemical properties and novel applications. *Microbiol. Mol. Biol. Rev.* 70, 283–295. doi: 10.1128/MMBR.00028-05
- Silveira, M. H. L., Morais, A. R. C., da Costa Lopes, A. M., Oleksyzszen, D. N., Bogel-Lukasik, R., Andraus, J., et al. (2015). Current pretreatment technologies for the development of cellulosic ethanol and biorefineries. *ChemSuschem* 8, 3366–3390. doi: 10.1002/cssc.201500282
- Singh, A., Patel, A. K., Adsul, M., Mathur, A., and Singhania, R. R. (2017). Genetic modification: a tool for enhancing cellulase secretion. *Biofuel Res. J.* 4, 600–610. doi: 10.18331/BRJ2017.4.2.5
- Szjártó, N., Siika-aho, M., Tenkanen, M., Alapuranen, M., Vehmaanperä, J., Réczey, K., et al. (2008). Hydrolysis of amorphous and crystalline cellulose by heterologously produced cellulases of *Melanocarpus albomyces*. *J. Biotechnol.* 136, 140–147. doi: 10.1016/j.jbiotec.2008.05.010
- Takashima, M., and Sugita, T. (2019). Draft genome analysis of *Trichosporonales* species that contribute to the taxonomy of the genus *trichosporon* and related taxa. *Med. Mycol. J.* 60, 51–57.
- Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577.
- Tanimura, A., Takashima, M., Sugita, T., Endoh, R., Kikukawa, M., Yamaguchi, S., et al. (2014). Selection of oleaginous yeasts with high lipid productivity for practical biodiesel production. *Bioresour. Technol.* 153, 230–235. doi: 10.1016/j.biortech.2013.11.086
- Uemura, T., Tomonari, Y., Kashiwagi, K., and Igarashi, K. (2004). Uptake of GABA and putrescine by UGA4 on the vacuolar membrane in *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* 315, 1082–1087. doi: 10.1016/j.bbrc.2004.01.162
- Valente, M. T., Infantino, A., and Aragona, M. (2011). Molecular and functional characterization of an endoglucanase in the phytopathogenic fungus *Pyrenochaeta lycopersici*. *Curr. Genet.* 57, 241–251. doi: 10.1007/s00294-011-0343-5
- van den Brink, J., and de Vries, R. P. (2011). Fungal enzyme sets for plant polysaccharide degradation. *Appl. Microbiol. Biotechnol.* 91:1477. doi: 10.1007/s00253-011-3473-2
- Várnai, A., Mäkelä, M. R., Djajadi, D. T., Rahikainen, J., Hatakka, A., and Viikari, L. (2014). “Chapter four - carbohydrate-binding modules of fungal cellulases: occurrence in nature, function, and relevance in industrial biomass conversion,” in *Advances in Applied Microbiology*, eds S. Sariaslani and G. M. Gadd (Cambridge, MA: Academic Press), 103–165.
- Walter, D., Wissing, S., Madeo, F., and Fahrenkrog, B. (2006). The inhibitor-of-apoptosis protein Bir1p protects against apoptosis in *S. cerevisiae* and is a substrate for the yeast homologue of Omi/HtrA2. *J. Cell Sci.* 119, 1843–1851.
- Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548. doi: 10.1093/molbev/msx319
- Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., and Zemla, J. (2017). Package ‘corrplot’. *Statistician* 56, 316–324.
- Whittaker, M. M., Kersten, P. J., Nakamura, N., Sanders-Loehr, J., Schweizer, E. S., and Whittaker, J. W. (1996). Glyoxal oxidase from *phanerochaete chrysosporium* is a new radical-copper oxidase. *J. Biol. Chem.* 271, 681–687. doi: 10.1074/jbc.271.2.681

- Yegin, S., Fernandez-Lahore, M., Salgado, A., Guvenc, U., Goksungur, Y., and Tari, C. (2011). Aspartic proteinases from *Mucor* spp. in cheese manufacturing. *Appl. Microbiol. Biotechnol.* 89, 949–960. doi: 10.1007/s00253-010-3020-6
- Yoon, S.-H., Ha, S.-M., Lim, J., Kwon, S., and Chun, J. (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 110, 1281–1286. doi: 10.1007/s10482-017-0844-4
- Yurkov, A., Guerreiro, M. A., Sharma, L., Carvalho, C., and Fonseca, A. (2015). Multigene assessment of the species boundaries and sexual status of the basidiomycetous yeasts *Cryptococcus flavescens* and *C. terrestris* (Tremellales). *PLoS One* 10:e0120400. doi: 10.1371/journal.pone.0120400
- Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, W95–W101. doi: 10.1093/nar/gky418
- Zhang, S. Q., Zou, Z., Shen, H., Shen, S. S., Miao, Q., Huang, X., et al. (2016). Mnn10 maintains pathogenicity in candida albicans by extending α -1,6-mannose backbone to evade host dectin-1 mediated antifungal immunity. *PLoS Pathog.* 12:e1005617. doi: 10.1371/journal.ppat.1005617
- Zhang, W., Kou, Y., Xu, J., Cao, Y., Zhao, G., Shao, J., et al. (2013). Two major facilitator superfamily sugar transporters from *Trichoderma reesei* and their roles in induction of cellulase biosynthesis. *J. Biol. Chem.* 288, 32861–32872.
- Zhao, Z., Liu, H., Wang, C., and Xu, J.-R. (2014). Erratum to: comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* 15:6. doi: 10.1186/1471-2164-15-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Aliyu, Gorte, Zhou, Neumann and Ochsenreither. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.