



## **CAZyme prediction in ascomycetous yeast genomes guides discovery of novel xylanolytic species with diverse capacities for hemicellulose hydrolysis**

Downloaded from: <https://research.chalmers.se>, 2025-12-04 23:21 UTC

Citation for the original published paper (version of record):

Ravn, J., Engqvist, M., Larsbrink, J. et al (2021). CAZyme prediction in ascomycetous yeast genomes guides discovery of novel xylanolytic species with diverse capacities for hemicellulose hydrolysis. *Biotechnology for Biofuels*, 14(1). <http://dx.doi.org/10.1186/s13068-021-01995-x>

N.B. When citing this work, cite the original published paper.

RESEARCH

Open Access



# CAZyme prediction in ascomycetous yeast genomes guides discovery of novel xylanolytic species with diverse capacities for hemicellulose hydrolysis

Jonas L. Ravn<sup>1</sup>, Martin K. M. Engqvist<sup>1</sup>, Johan Larsbrink<sup>1,2</sup> and Cecilia Geijer<sup>1\*</sup> 

## Abstract

**Background:** Ascomycetous yeasts from the kingdom fungi inhabit every biome in nature. While filamentous fungi have been studied extensively regarding their enzymatic degradation of the complex polymers comprising lignocellulose, yeasts have been largely overlooked. As yeasts are key organisms used in industry, understanding their enzymatic strategies for biomass conversion is an important factor in developing new and more efficient cell factories. The aim of this study was to identify polysaccharide-degrading yeasts by mining CAZymes in 332 yeast genomes from the phylum Ascomycota. Selected CAZyme-rich yeasts were then characterized in more detail through growth and enzymatic activity assays.

**Results:** The CAZyme analysis revealed a large spread in the number of CAZyme-encoding genes in the ascomycetous yeast genomes. We identified a total of 217 predicted CAZyme families, including several CAZymes likely involved in degradation of plant polysaccharides. Growth characterization of 40 CAZyme-rich yeasts revealed no cellulolytic yeasts, but several species from the *Trichomonascaceae* and CUG-Ser1 clades were able to grow on xylan, mixed-linkage  $\beta$ -glucan and xyloglucan. *Blastobotrys mokoensis*, *Sugiyamaella lignohabitans*, *Spencermartinsiella europaea* and several *Scheffersomyces* species displayed superior growth on xylan and well as high enzymatic activities. These species possess genes for several putative xylanolytic enzymes, including ones from the well-studied xylanase-containing glycoside hydrolase families GH10 and GH30, which appear to be attached to the cell surface. *B. mokoensis* was the only species containing a GH11 xylanase, which was shown to be secreted. Surprisingly, no known xylanases were predicted in the xylanolytic species *Wickerhamomyces canadensis*, suggesting that this yeast possesses novel xylanases. In addition, by examining non-sequenced yeasts closely related to the xylanolytic yeasts, we were able to identify novel species with high xylanolytic capacities.

**Conclusions:** Our approach of combining high-throughput bioinformatic CAZyme-prediction with growth and enzyme characterization proved to be a powerful pipeline for discovery of novel xylan-degrading yeasts and enzymes. The identified yeasts display diverse profiles in terms of growth, enzymatic activities and xylan substrate preferences, pointing towards different strategies for degradation and utilization of xylan. Together, the results provide

\*Correspondence: [cecilia.geijer@chalmers.se](mailto:cecilia.geijer@chalmers.se)

<sup>1</sup> Department of Biology and Biological Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden  
Full list of author information is available at the end of the article



© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

novel insights into how yeast degrade xylan, which can be used to improve cell factory design and industrial bioconversion processes.

**Keywords:** Ascomycota, Non-conventional yeasts, CAZymes, Xylanase, Xylan, Xylanolytic yeasts

## Background

Revolutionizing the use of biomass is one of the most promising pathways to a more sustainable production of liquid fuels, chemicals and materials and a reduced fossil fuel dependence. The global benefits of a 'green shift' towards a circular, biobased economy are numerous and include lower CO<sub>2</sub> emissions, resilient product and food chains and creation of stimulating high-skilled jobs [1]. However, for it to be realized, many technological hurdles and biochemical challenges in waste minimization and resource conversion efficiency must be overcome [1, 2].

Lignocellulosic biomass is mainly composed of the homopolysaccharide cellulose (40–60% of dry weight), various hemicellulosic heteropolysaccharides (20–35% of dry weight), and the aromatic polymer lignin (15–40% of dry weight) [3]. Cellulose is a linear polysaccharide consisting of  $\beta$ -1,4-linked D-glucose units that form crystalline and insoluble microfibrils [4]. Hemicelluloses coat the cellulose fibrils and their proportions and abundances differ between plant species. In industrially important grasses and hardwoods, xylans are the most abundant hemicellulose type, while in other species galacto-glucomannans, xyloglucans and mixed-linkage  $\beta$ -glucans are more abundant [5–7]. Xylans comprise a backbone of  $\beta$ -1,4-linked D-xylose residues which are commonly O-acylated and further substituted by  $\alpha$ -1,2- or  $\alpha$ -1,3-linked arabinosyl units and  $\alpha$ -1,2-linked (methyl)-glucuronic acid moieties, and these carbohydrate decorations can in turn be further substituted in various patterns. The xylans are typically grouped into arabinoxylan (AX), glucuronoxylan (GX) and glucuronoarabinoxylan (GAX) [8]. The arabinosyl substitutions found on xylans can be esterified with ferulic acid that can in turn form phenolic crosslinks to other feruloylated xylans or the hydrophobic lignin polymers in the plant cell wall [9], thereby exerting biomechanical contributions to cellulose fibrillar networks [10]. For biorefining purposes, the complex and heterogenous carbohydrate matrix in the plant cell wall represents one of the main challenges in efficient and rapid conversion of biomass and biowastes to value-added chemicals and fuels [11, 12].

Microbes and their carbohydrate-active enzymes (CAZymes) are central for depolymerization of the complex lignocellulosic polysaccharides in the global carbon cycle as well as in industrial bioconversion processes [13]–[15]. Complete or semi-complete enzymatic

breakdown of biomass requires multiple exo-, endo- and auxiliary CAZymes to hydrolyze the diversity of polysaccharide backbones and side chains [16, 17]. CAZymes are divided into classes and families in the carbohydrate-active enzymes database (CAZy, [www.cazy.org](http://www.cazy.org); [18]) based on their sequence similarities, which in turn determine their structures and functions, e.g., catalysis reactions [19]. The enzyme classes in CAZy comprise glycoside hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs) and auxiliary activities (AAs). The database also comprises the non-catalytic carbohydrate-binding modules (CBMs), which are often found linked to degradative CAZymes, where their main function is to provide additional substrate-binding capabilities and improve overall enzyme efficiency [20].

Knowledge on CAZymes targeting the plant cell wall has mainly been generated from research on filamentous fungi and bacteria [21–23]. Various yeast species have been shown to grow on diverse and complex substrates, but their contribution to biomass degradation and as a source of CAZymes has been largely overlooked [21, 24]. Thus, yeast species represent an untapped source of CAZymes of potential industrial relevance. Moreover, yeasts capable of growing on diverse and complex substrates in challenging environments combined with a unicellular growth pattern, ease of cultivation and genetic manipulation make them attractive candidates as future biorefinery cell factories for consolidated bioprocessing (CBP) [25, 26].

Based on known CAZyme protein domains, it has recently become possible to annotate and predict CAZymes in whole genomes in a high-throughput manner using the automated online meta server dbCAN2 [27]. Moreover, advances in next-generation sequencing and bioinformatical tools have considerably increased knowledge of yeast genetics and evolution [28] and about a quarter of the approx. 1500 yeast species described to date have been sequenced [26, 29]. Together, these technical advances provide an opportunity to identify polysaccharide-degrading yeast species through bioinformatic mining, complementing time-consuming and labor-intensive bioprospecting approaches. The aim of this study was to identify polysaccharide-degrading yeasts by mining 332 yeast genomes from the Ascomycota phylum [26]. We used the results of the initial prediction of the species' CAZyme repertoires to select a

subgroup of CAZyme-rich yeasts for more in-depth characterization of polysaccharide metabolism and enzymatic activities. This bioinformatic-based approach allowed us to map phylogenetic clades rich in xylanolytic yeast species and identify additional highly xylanolytic non-sequenced yeast species.

## Methods

### Prediction of CAZymes by dbCAN2 in 332 ascomycetous yeasts

A bioinformatic analysis was carried out to identify CAZymes in the 332 ascomycetous yeasts [26]. Fasta files containing protein sequences were downloaded from Figshare (<https://doi.org/10.6084/m9.figshare.5854692>) in November 2019. The protein sequences in each fasta file were de-duplicated by clustering at 98% identity using CD-HIT [30] and cluster representatives were carried forward for further analysis. Hidden Markov Models (HMMs) for CAZymes were downloaded from dbCAN (<http://bcb.unl.edu/dbCAN2/>, version 8) [27]. Each sequence in the fasta files was matched against these HMMs using HMMER3 [31] with the *-E* flag set to filter hits with *e*-values below  $10^{-15}$  as well as with the *-domtblout* flag to obtain an easily parsable output file. Hits covering less than 35% of the corresponding HMM model were removed. Additionally, if two domains showed more than 20% overlap on a single protein, only the domain with a better *e*-value score was retained. For each of the enzymes in the fasta files, potential signal peptides, indicating secretion, were also predicted using SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) [32]. In these runs the *-org* flag was set to "euk" (for eukaryote), and the *-format* flag to "short" to obtain easily parsable output files. The final data was visualized, using the ETE toolkit, on phylogenetic trees comprising the 332 yeast species, likewise retrieved from the Figshare record (<https://doi.org/10.6084/m9.figshare.5854692>).

The entire bioinformatic analysis was carried out in an Anaconda environment (<https://www.anaconda.com/>; version 2018.12) using the Python programming language (version 3.7.1) on a Unix system. Python libraries used include Pandas version 0.25.1, re version 2.2.1, BioPython version 1.72 [33], ETE3 version 3.1.2 [34] and Matplotlib version: 3.3.2 [35]. Other dependencies include CD-HIT version 4.8.1, HMMER version 3.2.1, and SignalP version 5.0b.

### Yeast selection

Yeasts were selected based on their total number of predicted CAZymes and CAZyme functional activity clustering in polysaccharide degradation (Additional file 1: Table S1). In total, 40 sequenced yeasts and six non-sequenced yeasts were ordered from the ARS

Culture Collection, USA (NRRL; <https://nrrl.ncaur.usda.gov/>). The selected sequenced species that we managed to cultivate in the lab are listed in Table 1. The six non-sequenced species were: *Sugiyamaella novakii* (CBS 8402), *Sugiyamaella smithiae* (CBS 5657), *Blastobotrys malaysiensis* (CBS 10336), *Blastobotrys illinoisensis* (CBS 10339), *Blastobotrys parvus* (CBS 6147) and *Scheffersomyces shehatae* (CBS 5813). All strains were either received freeze-dried in ampules which were re-grown in liquid yeast extract–peptone–dextrose (YPD) at room temperature, or as agar slants which were re-streaked on YPD agar plates and grown at room temperature. YPD contained 10 g L<sup>-1</sup> yeast extract, 20 g L<sup>-1</sup> peptone and 20 g L<sup>-1</sup> glucose.

### Yeast growth characterization

Growth on polysaccharides was measured in both semi-solid and liquid media. Polysaccharides included wheat arabinoxylan (Megazyme, Ireland), birchwood glucuronoxylan (Sigma-Aldrich, Germany), xyloglucan (tamarind, Megazyme, Ireland), mixed-linkage  $\beta$ -1,3/1,4-glucan (barley, Megazyme, Ireland), galactomannan (guar/locust bean gum, Sigma-Aldrich, Germany), glucomannan (konjac, Sigma-Aldrich, Germany), curdlan (Merck, USA), poly-methylgalacturonan (Sigma-Aldrich, Germany), pectin (citrus, Sigma-Aldrich, Germany), carboxymethyl cellulose (Sigma-Aldrich, Germany), Avicel (Sigma-Aldrich, Germany) and potato starch (Sigma-Aldrich, Germany). For semi-solid growth, agar plates were prepared using autoclaved Delft minimal medium with different polysaccharides 0.2% (w/v) and 2% agar (w/v). The Delft media contained 5 g L<sup>-1</sup> ammonium sulfate, 3 g L<sup>-1</sup> potassium phosphate, 1 g L<sup>-1</sup> magnesium sulfate, vitamins and trace metals as described previously [36], and pH was adjusted to 5 using 2 M KOH. Yeasts were inoculated in Delft medium 2% glucose (w/v) and grown at 30 °C, 150 rpm for 24 h before harvested, washed, and resuspended in water to a cell density of OD<sub>600</sub> = 5.10  $\mu$ l of the cell suspensions were spotted on plates that were then sealed with parafilm and kept at room temperature for 10 days before scoring growth. All strains were also spotted on agar plates either without any carbon source (where no strains were expected to grow) or with 2% glucose (where all strains were expected to grow). The *Saccharomyces cerevisiae* strain CEN.PK 113-7D that is unable to grow on polysaccharides was also included as a negative control. Growth was scored by visual inspection of colony thickness and size (including hyphae) in comparison to cell droplets on plates without carbon source. For growth in liquid cultures, yeasts were inoculated with a starting OD<sub>600</sub> = 0.05 in Delft minimal media containing 10 g

**Table 1** Overview of budding yeast growth assessment on agar plates and liquid cultures using different polysaccharides

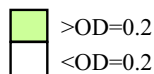
Clade	Species	NRRL	Wheat AX	Birch GX	Xyloglucan	$\beta$ -glucan	GluMan	GalMan	Pectin	Poly-MeGal	Curdlan	Avicel	CMC
Saccharom.	<i>Saccharomyces cerevisiae</i> (CENPK)		-	-	-	-	+	+	-	W	-	-	-
Lipomyc.	<i>Lipomyces arxii</i> , CBS 7333	Y-17921	-	-	-	-	-	-	-	+	-	-	-
Trichomon asceae	<i>Spenceriia europaea</i> , CBS 11730	Y-48265	++	++	+	+	W	-	-	+	-	-	-
	<i>Sagiyamaella lignohabitus</i> , CBS 10342	YB-1473	+	++	-	W	W	-	++	++	-	-	-
	<i>Diddensella caesiifluorescens</i> , CBS 12613	Y-48781	W	W	-	-	-	-	-	W	-	-	-
	<i>Blastobotrys peariensis</i> , CBS 10340	YB-2290	++	++	-	++	-	-	-	+	-	-	-
	<i>Blastobotrys muscicola</i> , CBS 10338	Y-7993	W	W	-	W	-	W	-	++	-	-	-
	<i>Blastobotrys mokenaii</i> , CBS 8435	Y-27120	+++	++	+++	W	+	+++	-	+	W	-	-
	<i>Blastobotrys americana</i> , CBS 10337	Y-6844	W	W	-	-	-	W	-	+	W	-	-
	<i>Blastobotrys adenivorans</i> , CBS 7370	Y-17693	+	+	-	++	-	-	-	+	-	-	-
	<i>Blastobotrys proliferans</i> , CBS 522.75	Y-17577	++	W	W	++	+	++	-	++	W	-	-
	<i>Blastobotrys serpentis</i> , CBS 10541	Y-48249	+	+	+	W	W	W	-	+	-	-	-
	<i>Blastobotrys raffinosefermentans</i> , CBS 6800	Y-27150	+	+	+	+	+	-	-	+	+	-	-
Pichiaceae	<i>Ambrosiozyma philentoma</i> , CBS 6276	Y-7523	W	W	-	+	-	-	-	+	W	-	-
	<i>Ambrosiozyma monospora</i> , CBS 6392	Y-7403	W	-	-	-	-	-	W	+	-	-	-
	<i>Ambrosiozyma oregonensis</i> , CBS 5560	Y-6106	W	W	-	++	-	-	W	+	-	-	-
CUG-Ser1	<i>Spathaspora passalidarum</i> , CBS 10155	Y-27907	+	++	-	-	W	-	-	++	-	-	-
	<i>Debaryomyces subglobosus</i> , CBS 792	Y-6666	-	W	-	-	-	-	-	+	-	-	-
	<i>Debaryomyces fabryi</i> , CBS 5948	Y-1453	-	W	-	-	-	-	-	+	-	-	-
	<i>Debaryomyces prosopidis</i> , CBS 8450	Y-27369	-	W	-	-	-	-	-	+	-	-	-
	<i>Debaryomyces nepalensis</i> , CBS 5921	Y-7108	+	W	-	-	-	-	-	+	-	W	-
	<i>Candida gorgasii</i> , CBS 9880	Y-27707	W	+	-	-	-	-	-	+	-	-	-
	<i>Debaryomyces marasmius</i> , CBS 1958	Y-2171	W	W	-	-	-	-	-	+	-	W	-
	<i>Scheffersomyces lignosus</i> , CBS 4705	Y-12856	++	++	-	-	-	-	-	+	-	-	-
	<i>Scheffersomyces stipitidis</i> , CBS 6054	Y-7124	+	+	-	-	-	-	-	+	-	-	-
	<i>Wickerhamomyces anomalus</i> , CBS 5759	Y-17698	W	-	W	-	-	-	-	+	-	-	-
Phaffomyce taceae	<i>Wickerhamomyces ciferrii</i> , CBS 111	Y-1031	W	-	-	W	+	-	W	++	-	-	-
	<i>Wickerhamomyces canadensis</i> , CBS 1992	Y-1888	+	+	-	-	-	-	-	+	-	-	-
	<i>Wickerhamomyces hamphshirensis</i> , CBS 7208	YB-4128	W	+	-	-	-	W	W	W	W	-	-
	<i>Saccharomycopsis capsularis</i> , CBS 2519	Y-17639	+	+	W	W	-	-	-	W	W	-	-
CUG-Ser2	<i>Saccharomycopsis malanga</i> , CBS 6267	Y-7175	-	-	+	+	-	-	-	-	-	-	-

Growth was scored by visual comparison to a negative control plate not containing a carbon source and by the difference in colony thickness and size (including hyphae, if present)

Growth was ranked from + to +++, where + was regular growth and +++ extensive growth, while W indicates weak growth and - no growth. Growth after 72 h in liquid cultures > OD = 0.2 is indicated by a green color

AX, arabinosyl; GX, glucuronosyl; GluMan, glucomannan; GalMan, galactomannan; Poly-MeGal, poly-methylgalacturonan; CMC, carboxymethyl cellulose; Saccharom., Saccharomycetaceae; Lipomyc., Lipomycetaceae

Liquid culture 72 h



L<sup>-1</sup> (w/v) of the different polysaccharides except curd-  
lan, CMC and Avicel, and cultivated at 30 °C, 150 rpm  
for 72 h before determining growth through optical  
density (OD<sub>600</sub>) measurements. Yeast cultures that dis-  
played optical densities of OD<sub>600</sub> ≥ 0.2 were considered  
as growing on the respective polysaccharide.

To follow growth on xylan substrates over time,  
selected species were precultured at 30 °C, 150 rpm  
for 24 h in Delft medium containing 2% xylose (w/v).  
Here, xylose was selected as carbon source as it has  
previously been shown to induce expression of xyla-  
nases in other xylanolytic yeasts [24, 37]. Precultured  
cells were then inoculated in 250 µl Delft medium sup-  
plemented with 10 g L<sup>-1</sup> xylan (either wheat AX or  
birchwood GX) to a starting OD<sub>600</sub> = 0.2. While wheat  
AX was soluble in Delft medium, birchwood GX was  
not fully soluble. All yeast strains were grown in bio-  
logical triplicates in a 96-well plate setup in a Growth-  
Profiler 960 (EnzyScreen, Netherlands). ‘Green Values’  
(GV) measured by the GrowthProfiler correspond to  
growth based on pixel counts, and GV changes were  
recorded every 20 min for 72 h at 30 °C and 150 rpm.

### Xylanolytic activity determination

To quantify the xylanolytic yeasts’ secretome and cell-  
associated xylanase activities, the final cultures from  
the GrowthProfiler experiment were collected by cen-  
trifugation (2000×g 15 min) and xylanase activity was  
assayed in the cell-free supernatant or the intact cell  
pellets, respectively. The assay mixture consisted of a  
175 µl xylan suspension of 10 g L<sup>-1</sup> wheat AX or birch-  
wood GX and 50 mM sodium acetate buffer (pH 5.5)  
added to cell pellets or 25 µl cell-free supernatants  
mixed in a 96-well plate. The mixture was incubated at  
30 °C for 30 min followed by immediate chilling on ice.  
Reducing sugar ends released by xylanases was deter-  
mined by the dinitrosalicylic acid (DNS) method [38]  
as end point assay. All enzymatic measurements were  
performed in triplicates. One unit of enzyme activ-  
ity was defined as the amount of enzyme required to  
release 1 µmol of reducing sugars in 1 min under the  
assay conditions. Volumetric activity (U mL<sup>-1</sup>) was cal-  
culated by converting mM reduced sugar to Units by  
multiplying with total assay volume (L), dividing with



assay time (min) and then dividing with sample volume (L) as described previously [39].

### Phylogenetic analysis

Phylogenetic trees of GH10 and 11 xylanases were constructed using the identified yeast enzymes as well as sequences from 259 characterized GH10 members and 208 characterized GH11 members retrieved from the CAZy database ([www.cazy.org](http://www.cazy.org)), respectively. The sequences were aligned using MUSCLE (<https://www.ebi.ac.uk/Tools/msa/>) [40], and then submitted for tree building using the online Iqtree tool <http://iqtree.cibiv.univie.ac.at/> with 1000 bootstrap alignments and viewed in MEGA-X as Newick trees [41]. For species phylogenetic analysis, Internal Transcribed Spacer (ITS) nucleotide sequences from xylanolytic yeasts, their closely related species and *Schizosaccharomyces pombe* as out-group were aligned using ClustalW, and a maximum likelihood (ML) phylogenetic tree with bootstrap value 1000 was constructed using MEGA-X.

### In-gel proteomics of the GH11 enzyme from *Blastobotrys mokoensis*

Supernatants from yeast cultures of *Blastobotrys mokoensis* grown in Delft minimal medium with 10 g L<sup>-1</sup> birchwood GX or wheat AX (72 h, 30 °C, 150 rpm) were concentrated using 10 kDa ultra centrifugal filters (Amicon, Merck, Germany) by centrifugation (2000×g, 10 min, repeated 3×). Secreted proteins were identified by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE). A protein at ~24 kDa was cut out from the gel using a scalpel and kept at –20 °C before sent for proteomic analysis. The protein identity was confirmed by MS/MS analysis as a 23.19-kDa GH11 xylanase with the following predicted protein sequence (217 residues):

MKLSNAITAICAAVLAAPLEEEVAKRSVTPSSTGTNNGYYYFSWSDGGGDVITYTNGNGGSSVSEWTNCGNFVGGKGWNPAAAREINFSGSFNPSGNGYLSVYGWTTNPLVEYYIVESYGDYNPGTAGTFLGTVDSDGSTYDIYKAVRTNAPSIEGTATFDQYWSIRRNHRTSGTVNTGNHFNAWAQHGLQLGTHNYQIVATEGYQSSGSSSITVS.

## Results

### CAZyme abundance and distribution in ascomycetous yeasts

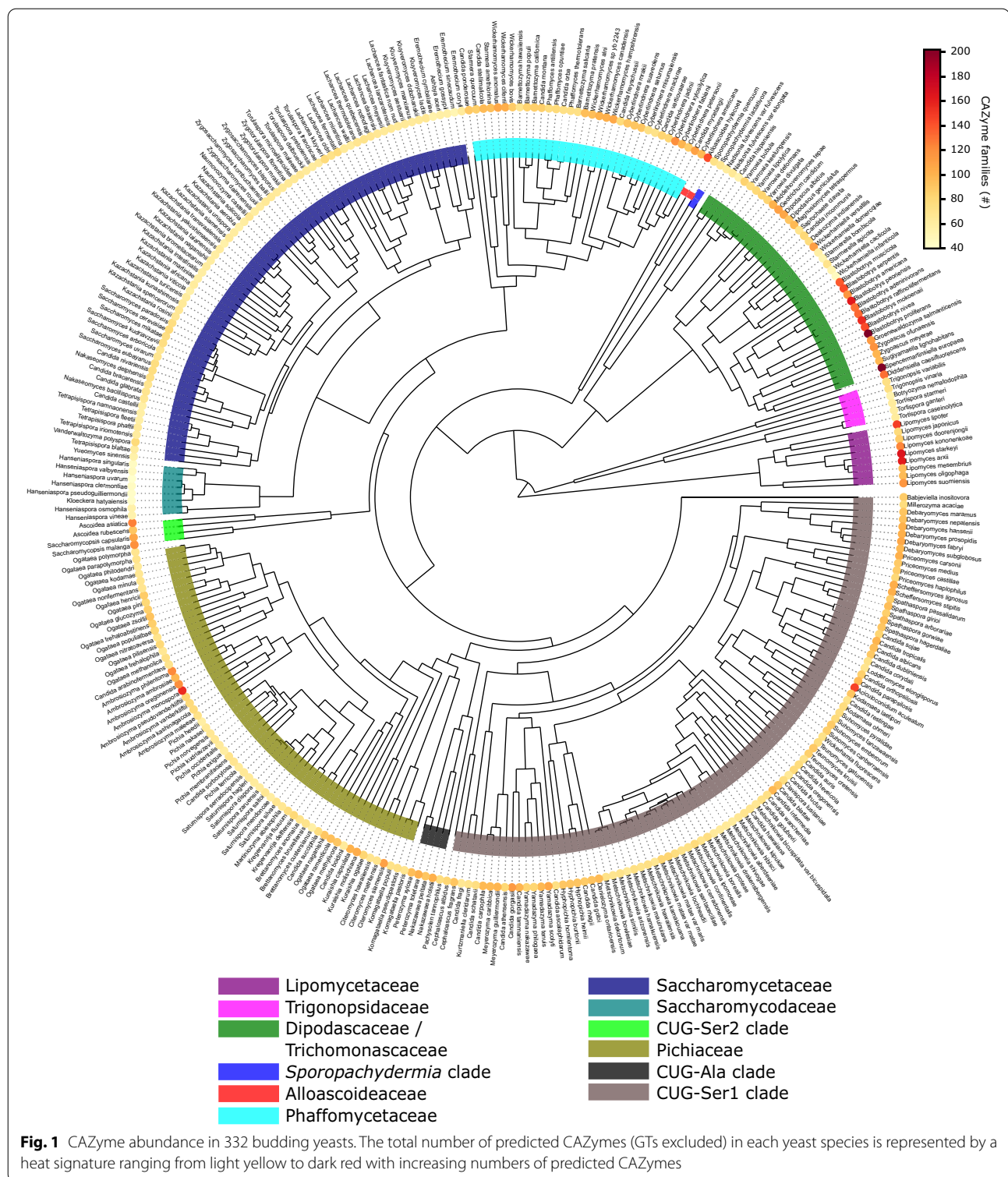
To identify polysaccharide-degrading yeasts, the dbCAN2 meta server was used to scan the genomes of 332 yeast species within the Ascomycota phylum to predict and compile their encoded CAZymes [26, 27]. It is important to note that the 332 genomes included in the dataset are not all complete [26], and therefore, the list

of CAZymes is likely incomplete. Nonetheless, we identified a total 217 different CAZyme families, with GT and GH as the most predominant classes. The full genetic CAZyme prediction and associated protein sequences can be downloaded from Zenodo (<https://zenodo.org/record/4548336/export/hx#.YLSzGqgzaUk>, <https://doi.org/10.5281/zenodo.4548335>).

The 332 yeasts encoded on average 152 CAZymes with 20 yeasts having more than 200 CAZymes. Yeasts containing the highest number of CAZymes primarily belong to the Trichomonascaceae clade, followed by species in the Lipomycetaceae and the Pichiaceae clades. Species with low amounts of CAZymes include *Hanseniaspora* and *Eremothecium* species. With our focus being on potential polysaccharide-degrading yeasts and their respective CAZymes, GTs involved in biosynthesis of disaccharides were not included in subsequent analyses. An overview of the abundance of CAZymes (except GTs) in individual yeast species throughout the phylogenetic tree can be viewed in Fig. 1, with increasing CAZyme numbers represented by yellow to dark red color. From this initial analysis, we can conclude that some clades more than others appear to be hotspots for identification and characterization of yeast CAZymes.

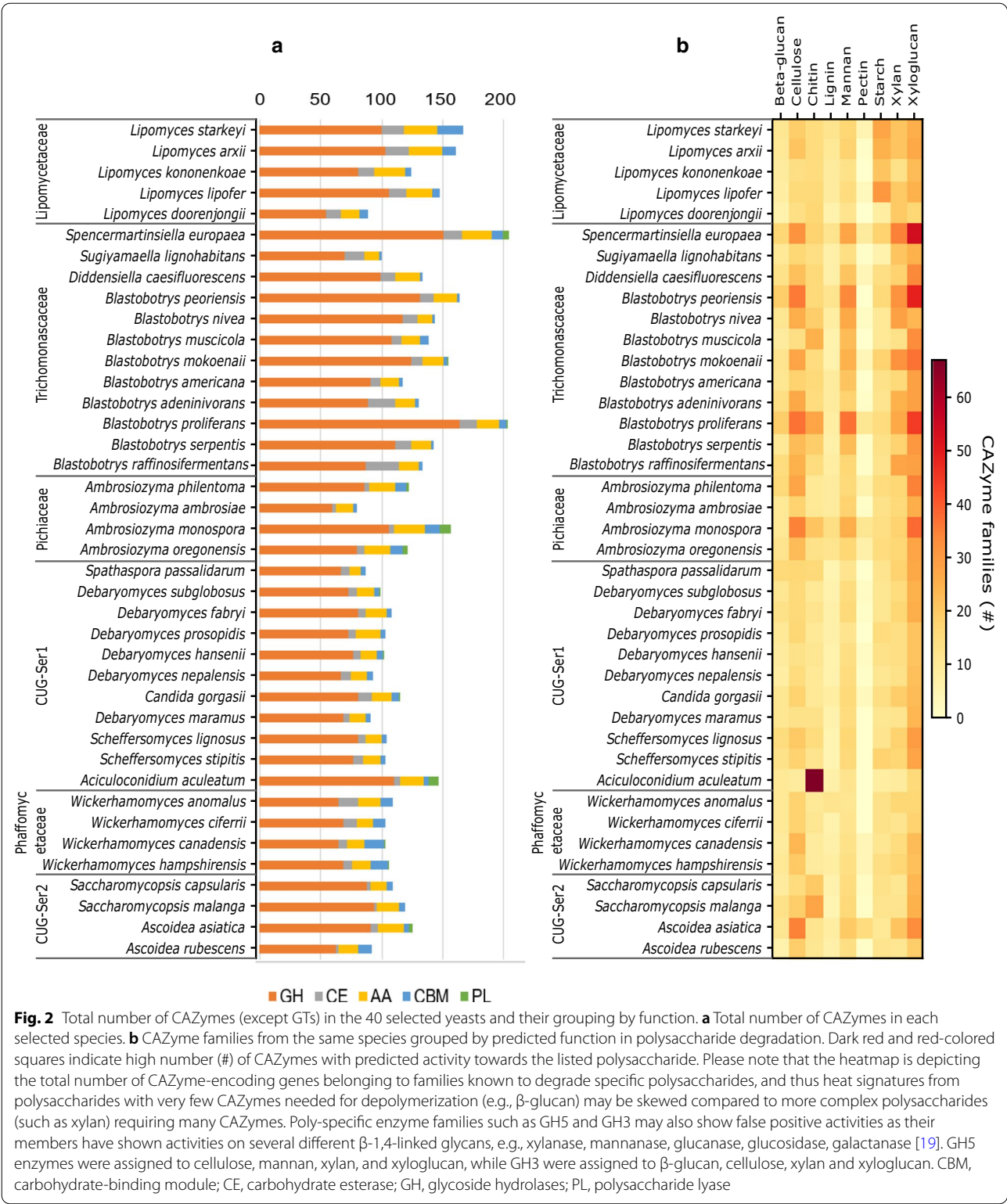
### CAZyme annotation in putative polysaccharide-degrading yeasts

To relate and confirm genetic CAZyme predictions to real capacity of polysaccharide utilization, 40 yeasts from six different phylogenetic clades were selected for further characterization. The species were chosen based on their total number of predicted CAZymes and the clustering of their CAZymes by functional activity involved in polysaccharide degradation. The distribution of the different enzyme classes (excluding GTs) and CBMs in the selected yeasts is shown in Fig. 2a. The GHs showed the highest variation in number among species while relatively few CBM and PL families were predicted. The yeast with highest number of CAZymes (excluding GTs) was *Spencermartinsiella europaea* with 204 predicted CAZymes followed by *Blastobotrys proliferans* (203) from the same clade, then *Lipomyces starkeyi* (167) from the Lipomycetaceae clade. These numbers are around twice the number of CAZymes found in the more commonly studied yeasts such as *Saccharomyces cerevisiae* (79) and *Schizosaccharomyces pombe* [22]. Notably, another five *Blastobotrys* species from the Trichomonascaceae clade also ranked among the top 25 yeasts in terms of absolute CAZyme numbers. In addition, we grouped the yeasts' CAZymes by predicting functional polysaccharide degradation activities (Additional file 1: Table S1), e.g.,  $\beta$ -glucanases, cellulases, chitinases, lignin-degrading enzymes, mannanases, pectinases, starch



degrading enzymes, xylanases and xyloglucanases [18] and created a heatmap based on the resulting number of enzymes (Fig. 2b). The analysis suggests that the yeasts from the Trichomonascaceae clade have diverse enzyme

portfolios and with a particular enrichment of mannan-, xylan-, xyloglucan-, and cellulose-degrading CAZymes. The Lipomycetaceae clade appears rich in starch degrading CAZymes, while *Aciculoconidium aculeatum* in



the CUG-Ser1 clade contains multiple enzymes for chitin degradation with a total of 57 predicted GH18 chitinases. In general, relatively few CAZymes associated with pectin and lignin degradation were predicted in the ascomycetous yeast genomes (Fig. 2b). Collectively, the results suggest that the assessed yeasts are equipped with a range of different polysaccharide-degrading enzymes, where some species seem specialized to degrade specific



polysaccharides while others appears to be polysaccharide generalists.

### Growth characterization on different polysaccharides

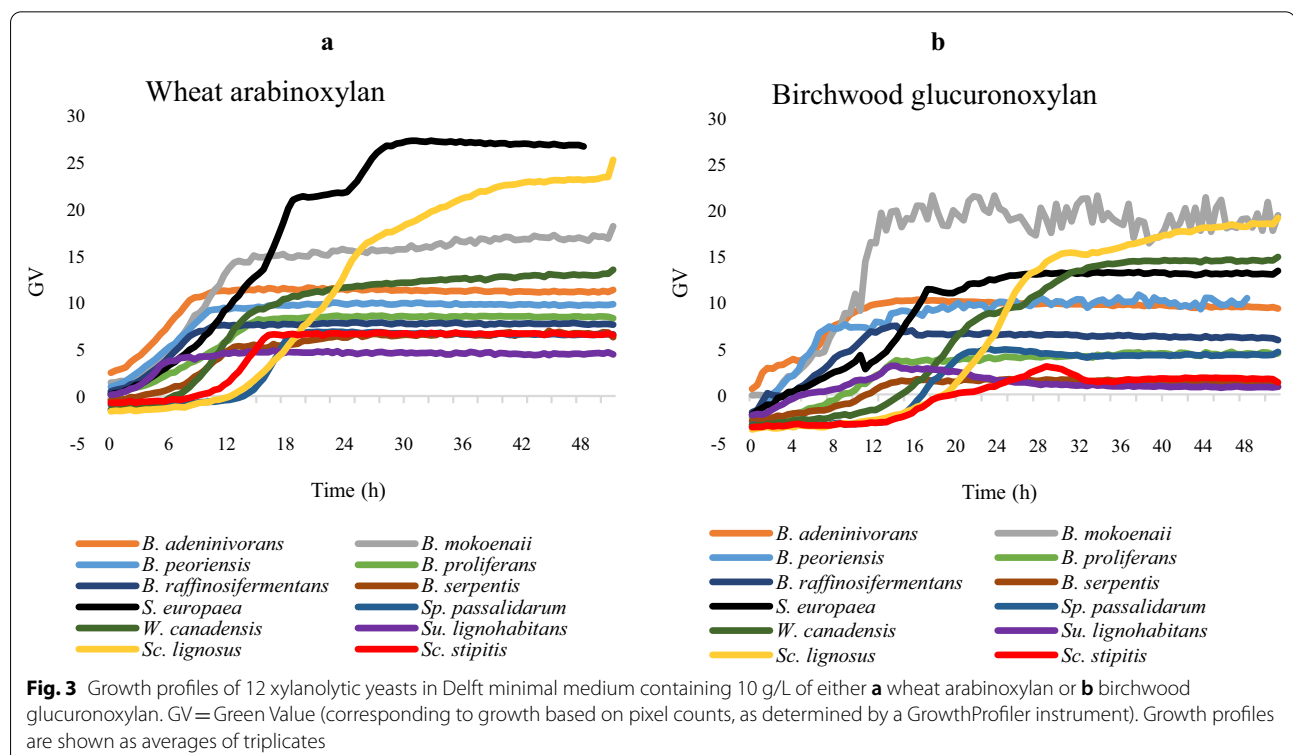
To determine if polysaccharides could support growth for the 40 selected ascomycetous yeasts, the yeasts were cultivated on agar plates with semisolid minimal media (Delft) supplemented with different polysaccharides as the sole carbon source. Growth on xylan, xyloglucan,  $\beta$ -glucan, galactomannan, glucomannan, pectin and poly-methylgalacturonan polysaccharides was also confirmed in liquid cultures and the accumulated growth results are shown in Table 1. Several species—*Lipomyces doorenjongii*, *Lipomyces kononenkoae*, *Lipomyces lipofer*, *Lipomyces starkeyi*, *Aciculoconidium aculeatum*, *Ambrosiozyma ambrosiae*, *Ascoidea rubescens* and *Blastobotrys nivea*—did not grow in the pre-cultures and were therefore discarded from further analysis. In general, growth on agar plates corresponded well with the increased optical density ( $OD_{600} > 0.2$ ) observed in liquid cultures, though some species from the CUG-Ser1 clade, particularly *Scheffersomyces* species, showed better growth in liquid culture than on agar plates with mannan-based, pectin and xyloglucan polysaccharides (Table 1). In accordance with the CAZyme heatmap (Fig. 2b), species from the Trichomonascaceae clade showed substantial growth on hemicellulosic substrates, particularly xylans,

$\beta$ -glucan, glucomannan and galactomannan. Also yeasts from the CUG-Ser1 and Phaffomycetaceae clades showed growth on xylan, whereas those from the Pichiaceae clade did not. Some of the herein characterized species have been identified as xylan-growers also in other screens, for example *Scheffersomyces stipitis*, *Sugiyamaella lignohabitans* and *Spenceriartinsella* sp. [42] while, to the best of our knowledge, other species such as *Blastobotrys serpensis*, *Blastobotrys peoriensis* and *Scheffersomyces lignosus* have so far escaped attention in this regard [43, 44].

In opposite to hemicellulosic substrates, the assessed yeasts did not grow well on cellulose despite predictions of cellulase activities (Table 1, Fig. 2b). In line with these results, a large-scale screen to identify wild cellulolytic yeasts showed that only 16 of 390 strains grew on cellulose and just 5 had significant enzyme activity levels [45], indicating that most yeasts are unable to utilize crystalline cellulose [24]. Overall, we can conclude that the polysaccharide-degrading ascomycetous yeasts identified in this study display better growth on hemicellulosic substrates compared to cellulosic substrates in accordance with previous studies.

### Growth and enzymatic activities of xylan-utilizing yeasts

To further characterize the top xylan-utilizing yeast species, we determined their growth profiles over time in both wheat AX and birchwood GX (Fig. 3). The



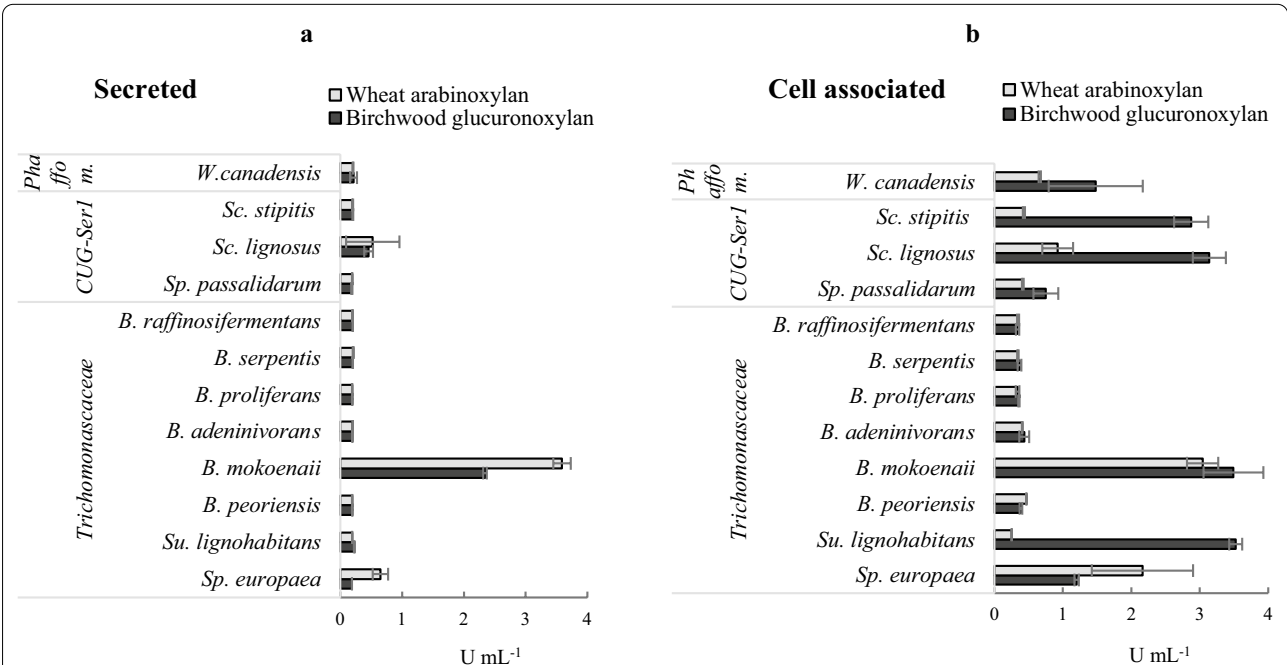
xylanolytic yeasts showed different growth profiles and reached stationary phase between 8–28 h, with some species showing biphasic growth curves. Notably, *B. mokoensis*, *Sp. europaea*, *Sc. lignosus* and *Wickerhamomyces canadensis* reached the highest optical densities in both xylans, although the yeasts' growth profiles differed somewhat between the two substrates (Fig. 3a, b).

Next, the xylanolytic yeasts were characterized in terms of xylanase activities. Both the secretome and cell-associated enzymatic activities were assayed to gain deeper insight into the xylanolytic strategies used by these species. Xylanase activity of the secretome was particularly high in *B. mokoensis* for both types of xylans, with a higher activity on wheat AX compared to birchwood GX (3.6 and 2.3 U mL<sup>-1</sup>, respectively) (Fig. 4a). These values were 7.2-fold higher than those of *Sc. lignosus* that had the second highest secretome activity values, and also higher than what has been reported previously on yeasts that secrete xylanases [37]. This indicates that *B. mokoensis* possesses a unique xylanolytic strategy among the studied species. *B. mokoensis* also had a high cell-associated xylanase activity on both wheat AX and birchwood GX, a feature shared with several other species. These included the other top xylan-growing species *Sp. europaea*, *Sc. lignosus*, and *W. canadensis* (Figs. 3 and 4b), which all showed good correlation between enzyme activity and growth. However, for several other yeasts,

the correlation between measured xylanase activity and growth characteristics was ambiguous. For example, yeasts such as *B. adeninivorans* and *B. peoriensis* with intermediate growth in both xylans showed only modest xylanolytic activities (0.2–0.3 U mL<sup>-1</sup>), whereas *Sc. stipitis* and *Su. lignohabitans* showed high xylanase activities (0.4–2.8 U mL<sup>-1</sup>) but only moderate xylan growth (Figs. 3, 4). Overall, the diverse profiles in terms of growth, enzymatic activities and xylan substrate preferences point towards different yeast strategies for degradation and utilization of xylan.

### CAZyme analysis in xylanolytic yeast species

To connect the experimentally measured xylanolytic activities with the predicted CAZymes, we identified all putative xylanolytic CAZymes for each of the top 12 xylan-growing yeasts (Table 2). Overall, the yeasts, coming from three clades, have similar numbers of genes encoding CEs with expected roles in de-acylation of polysaccharides, and GH3 enzymes predicted to act as exo-β-glycosidases on oligosaccharides. The species from the Trichomonascaceae clade have a more diverse and abundant xylanolytic CAZyme distribution compared to yeasts from other clades. The top-performing xylanolytic yeast *B. mokoensis* encodes a putative GH11 xylanase, which is a unique trait within the whole 332 yeast dataset. We were able to detect the GH11 protein



**Fig. 4** Xylanolytic yeast activities in liquid cultures. Volumetric activities of **a** secretome xylanases and **b** yeast cell-associated xylanases in wheat arabinoxylan (grey) and birchwood glucuronoxylan (black) determined at 30 °C after growth on xylan in liquid medium for 72 h. Phaffom. = Phaffomycetaceae clade

**Table 2** Xylanolytic CAZyme predicted from whole-genome sequenced xylanolytic yeasts

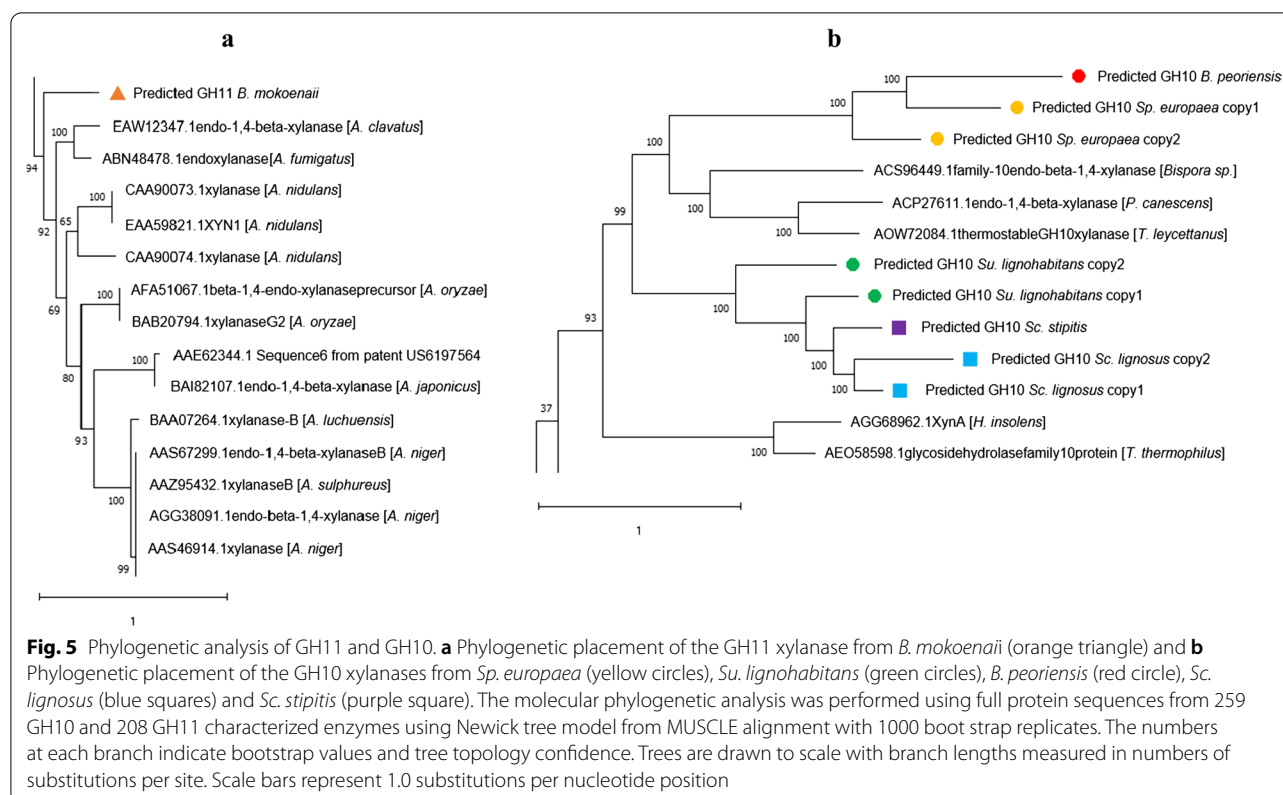
Clade	Species	CE (332/332)	GH3 (263/332)	GH5 (324/332)	GH10(5/332)	GH11 (1/332)	GH30 (11/332)	GH43 (22/332)	GH51 (39/332)	GH62 (1/332)	GH67 (4/332)	GH115 (28/332)	Total
Trichomon- ascaceae	<i>Sp. europaea</i>	CE1, CE4, CE15	GH3(11)	GH5_5(2), GH5_9(2), GH5_12(2), GH5_22(4), GH5_49	GH10(2)		GH30_5, GH30_7	GH43_14, GH43_24	GH51		GH67	GH115(3)	36
	<i>Su. lignohab- itans</i>	CE1(2), CE4(3), CE15	GH3(5)	GH5_9(2), GH5_12(2), GH5_22(2), GH5_49	GH10(2)		GH30_7		GH51			GH115	23
	<i>B. peoriensis</i>	CE1, CE4	GH3(15)	GH5_9(3), GH5_12(2), GH5_22(5), GH5_49	GH10		GH30_3(3)		GH51				33
<i>B. mokoensis</i>		CE1, CE4	GH3(8)	GH5, GH5_5, <b>GH5_7(2)</b> , GH5_9(2), GH5_12, GH5_22(2), GH5_49	<b>GH11</b>		GH30_5, GH30_7	GH43_6, GH43_24	GH51(3)	<b>GH62</b>	GH67	GH115(2)	32
	<i>B. adenini- vorans</i>	CE1, CE4(2)	GH3(8)	GH5_9(3), GH5_12(2), GH5_44, GH5_47			GH30_3	GH43_6	GH51		GH67		22
	<i>B. proliferans</i>	CE1(2), CE4	GH3(12)	GH5_5(3), GH5_9(2), GH5_12, GH5_31, GH5_49			GH30_3		GH51(3)				27
<i>B. serpentina</i>		CE1, CE4	GH3(5)	GH5, GH5_9(3), GH5_12, GH5_22, GH5_49			GH30_3(2)		GH51(2)				18
	<i>B. raffinosifer- mentans</i>	CE1, CE4(2), CE5(6)	GH3(9)	GH5_9(3), GH5_12(2), GH5_49			GH30_3	GH43_6	GH51		GH67		27

**Table 2** (continued)

Clade	Species	CE (332/332)	GH3 (263/332)	GH5 (324/332)	GH10(5/332)	GH11 (1/332)	GH30 (11/332)	GH43 (22/332)	GH51 (39/332)	GH62 (1/332)	GH67 (4/332)	GH115 (28/332)	Total
CUG-Ser1	<i>Spa. passali- darum</i>	CE1, CE4(2)	GH3(9)	GH5, GH5_9(3), GH5_22, GH5_49	GH10(2)							GH115	19
				GH5_5, GH5_9(3), GH5_12, GH5_22(2), GH5_49									
Phaff.	<i>Sc. lignosus</i>	CE1, CE4,	GH3(7)		GH10							GH115	20
Phaff.	<i>Sc. stipitis</i>	CE1, CE4,	GH3(7)	GH5_9(2), GH5_12, GH5_22 (3), GH5_49								GH115	17
Phaff.	<i>W. canadensis</i>	CE1(3), CE4	GH3(5)	GH5_9(2), GH5_12, GH5_22, GH5_49									14

CAZyme families marked in bold in the species within the 332-yeast dataset. Total number of species containing each CAZyme family is marked by (x/332). Copy number of each enzyme is stated in parenthesis next to the enzyme

CE carbohydrate esterase, GH glycoside hydrolase, Phaff. Phaffomycetace



**Fig. 5** Phylogenetic analysis of GH11 and GH10. **a** Phylogenetic placement of the GH11 xylanase from *B. mokoensis* (orange triangle) and **b** Phylogenetic placement of the GH10 xylanases from *Sp. europaea* (yellow circles), *Su. lignohabitanis* (green circles), *B. peoriensis* (red circle), *Sc. lignosus* (blue squares) and *Sc. stipitis* (purple square). The molecular phylogenetic analysis was performed using full protein sequences from 259 GH10 and 208 GH11 characterized enzymes using Newick tree model from MUSCLE alignment with 1000 boot strap replicates. The numbers at each branch indicate bootstrap values and tree topology confidence. Trees are drawn to scale with branch lengths measured in numbers of substitutions per site. Scale bars represent 1.0 substitutions per nucleotide position

with a molecular size of 23.19 kDa in the secretome of *B. mokoensis* grown in medium containing wheat AX or birchwood GX, using in-gel proteomic MS/MS analysis (Additional file 2: Fig. S1). The GH11 gene can be found in the genome position 29298–29948 in the GenBank sequence ID: PPJM02000065.1. *B. mokoensis* is also unique in that it possesses two gene copies for GH5 enzymes from subfamily 7 (GH5\_7; putative endo- $\beta$ -1,4-mannanases) and a GH62  $\alpha$ -L-arabinofuranosidase. Further, *B. mokoensis* encodes a GH30\_7 enzyme (putative exo- $\beta$ -1,4-xylanase or glucuronoxylanase) in common with only two other yeasts that also scored high in our assays: *Su. lignohabitanis* and *Sp. europaea*. Indeed, all eight species in the Trichomonascaceae clade have predicted GH30 enzymes and some species have putative GH67  $\alpha$ -glucuronidases as well as GH43 and GH51 enzymes predicted to be  $\alpha$ -L-arabinofuranosidases), indicating abilities to target complex GAX. A similar setup is not found in the CUG-Ser1 and Phaffomycetaceae clades. However, the CUG-Ser1 clade species possess a putative GH115  $\alpha$ -glucuronidase, potentially enabling them to hydrolyze glucuronic acid side chains present in birchwood GX.

Species displaying good xylanolytic activity (2–4 U mL<sup>-1</sup>) almost all possess predicted GH10 (*Sp. europaea*, *Su. lignohabitanis*, *Sc. stipitis* and *Sc. lignosus*) or GH11

(*B. mokoensis*) xylanases (Fig. 4, Table 2). An interesting exception is *W. canadensis*, which does not appear to encode either GH11, GH10 or GH30 xylanases. However, it possesses putative GH5\_9, GH5\_22 and GH5\_49 CAZymes in common with most of the xylanolytic species listed in Table 2, suggesting that some of these CAZymes may be novel xylanases. No xylanase activities have yet been confirmed in the mentioned GH5 subfamilies, and in fact no GH5\_49 enzymes have to date been biochemically characterized [18]. Although we cannot completely rule out that the lack of genes encoding known xylanases in *W. canadensis* is due to an incomplete genome assembly, this species and its putative xylanases deserve further characterization outside the scope of this study.

#### Phylogenetic analysis of GH10 and GH11 xylanases

To investigate the origin of the genes encoding the identified GH10 and GH11 members in the ascomycetous yeasts, we determined the phylogenetic relationships of these enzymes with 259 characterized enzymes from GH10 and 208 from GH11, listed in the CAZy database [18]. Phylogenetic trees displaying all characterized enzymes can be viewed in Additional file 3: Fig. S2. The GH11 xylanase from *B. mokoensis* shows the highest sequence identity (71.95%) to the xlnB xylanase from



(See figure on next page.)

**Fig. 6** Characterization of non-sequenced xylanolytic yeasts. **a** Phylogenetic analysis of 19 *Blastobotrys*, *Sugiyamaella* and *Scheffersomyces* species as well as *Schizosaccharomyces pombe* serving as outgroup. The molecular phylogenetic analysis was based on ITS sequences using maximum-likelihood model from ClustalW alignment with 1000 bootstrap replicates. The numbers at each branch indicate bootstrap values and tree topology confidence. The tree is drawn to scale, with branch lengths measured in the number (0.2) of substitutions per site. Growth profiles of xylanolytic yeasts grown in Delft medium containing 10 g L<sup>-1</sup> of **b** wheat arabinoxylan and **c** birchwood glucuronoxylan. Yeasts were grown for 48 h at 30 °C. GV = Green Value (corresponding to growth based on pixel counts, as determined by a GrowthProfiler instrument). **d** Secretome and **e** cell-associated volumetric xylanase activities on wheat arabinoxylan (grey) and birchwood glucuronoxylan (black) determined at 30 °C after growth in xylan-containing liquid medium for 72 h. Stars (\*) symbolizes non-sequenced species

*Aspergillus nidulans* FGSCA4 with confirmed ability to hydrolyze oat-spelt xylan [46], suggesting a similar function of the putative *B. mokoensis* enzyme (Fig. 5a). All GH10 copies from *Sp. europaea*, *Su. lignohabitans*, *B. peoriensis*, *Sc. lignosus* and *Sc. stipitis* clustered to the same branch of the phylogenetic tree, together with characterized xylanases from the filamentous fungi *Talaromyces leycettanus*, *Penicillium canescens* and *Bispora sp. MEY-1* (Fig. 5b). Thus, we can conclude that all yeast GH10 and GH11 are of Ascomycota origin and likely have xylanolytic activity. The presence of these genes in the yeast genomes could be a result of horizontal gene transfer within the phylum, or that these genes have been specifically retained by a small number of (ancestral) yeast species after the split between Pezizomycotina (filamentous fungi) and Saccharomycotina (yeasts) in the Ascomycota phylum. In favor for the gene retention explanation model, Morel and co-authors have shown that the genome of the yeast *Geotrichum candidum* within the Trichomonascaceae clade contains a few hundred genes that are orthologous to predicted genes in filamentous fungi rather than other sequenced Saccharomycotina yeasts [47]. Moreover, *B. mokoensis* possesses several other unique CAZymes, including ones from GH62 and GH12 that also show high sequence identity to CAZymes from *Aspergillus* species (77% to *A. pseudonimiae* and 62% to *A. flavus*, respectively), further supporting this model.

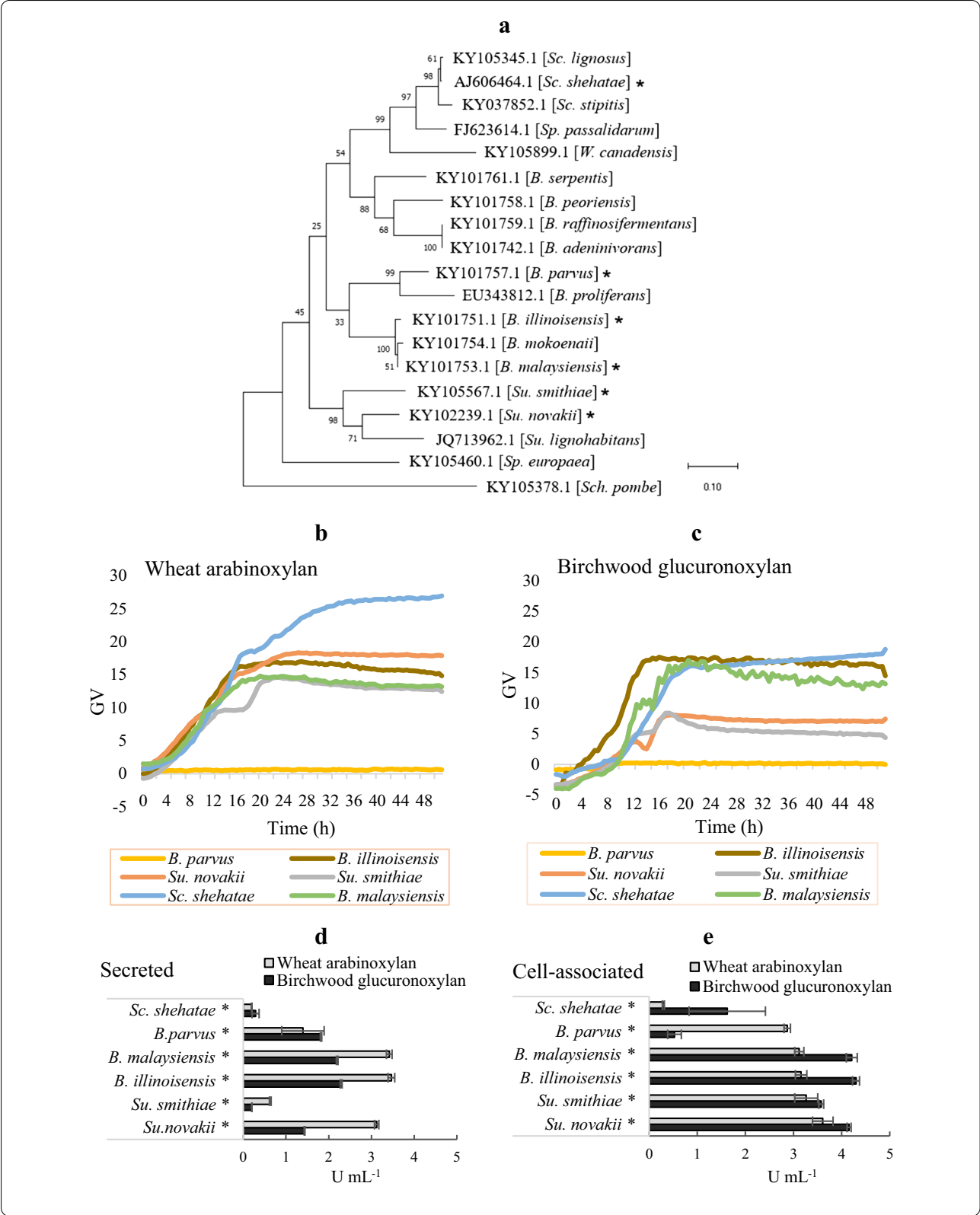
#### Identification of novel xylanolytic species by phylogenetic association

The successful approach of using CAZyme prediction to identify xylan-degrading yeasts in CAZyme-rich clades, prompted us to scout for additional, novel xylanolytic species through phylogenetic association. Six non-sequenced yeast species phylogenetically closely related to the highest scoring xylanolytic species found in this study were therefore included in another round of characterization; *Sugiyamaella novakii* (CBS 8402), *Sugiyamaella smithiae* (CBS 5657), *Blastobotrys malaysiensis* (CBS 10,336), *Blastobotrys illinoisensis* (CBS 10,339), *Blastobotrys parvus* (CBS 6147) and *Scheffersomyces shehatae* (CBS 5813) (Fig. 6a). All species except

*Sc. shehatae* have so far largely escaped scientific attention, and genomic information except for ITS and ribosomal RNA sequences is almost completely missing [37, 42, 44, 48]. The yeast growth profiles in Delft minimal media containing wheat AX and birchwood GX as carbon sources and the secretome and cell-associated xylanolytic activities and can be seen in Fig. 6b–e. All six species showed secreted or cell-associated xylanolytic activities, or both, and all except *B. parvus* grew on both xylan substrates. This species, however, displayed high xylanolytic activity (1.8–2.8 U mL<sup>-1</sup>) on both xylan types (Fig. 6b–e). Interestingly, *Sc. shehatae* reached the highest green values (26.7 GV) in wheat AX out of all the xylanolytic yeasts characterized in this study (Fig. 6b) and *Su. smithiae* and *Su. novakii* showed high activity on both xyans (in contrast to *Su. lignohabitans* which seem to prefer birchwood GX) (Figs. 4, 6d, e). Overall, these results show that CAZyme-rich clades are treasure troves for identifying xylanolytic yeast species, and sequencing and characterization of the new yeasts will most likely lead to additional discoveries of CAZymes with potential industrial value.

#### Discussion

Complete depolymerization of complex lignocellulosic polysaccharides requires a repertoire of enzymes that act together on the different chemical bonds [14]. While CAZyme systems from filamentous fungi and bacteria have been studied for decades, yeast species have received considerably less attention. However, since yeasts are key industrial workhorses, elucidating their plant cell wall-degrading potential may be of great benefit for the development of efficient CBP strains able to both produce the enzymes needed for biomass degradation and convert the released sugars into valuable products [49]. Non-conventional, xylanolytic yeasts can potentially be developed into future CBP cell factories. Alternatively, the strategies these yeasts use may be directly transferable to industrial *Saccharomyces* species in a manner that is not feasible for systems used by filamentous fungi or bacteria. We here present a strategy of high-throughput mining of genomes for putative CAZymes followed by growth studies and enzymatic investigation, through



which we identified several novel yeast species displaying high xylanolytic activities and seemingly diverse strategies for hemicellulose utilization.

### Correlation between predicted CAZymes, growth and enzymatic activities

Identification of CAZyme-encoding genes in a genome is a promising lead in finding polysaccharide-utilizing species, but it does not per se reveal whether the genes are actually expressed into functional enzymes by the organism. Moreover, a series of enzymatic activities is required to degrade complex polysaccharides into oligo- and monosaccharides that can be taken up and metabolized by the microorganism. To find yeasts equipped with all enzymes needed to hydrolyze and grow on plant polysaccharides, we initially selected CAZyme-rich yeasts rather than cherry-picked species containing single GHs of interest. By doing so, our hope was also to increase the chances of capturing species equipped with CAZymes from poly-specific GH families or even GH families not yet associated with the targeted polysaccharides. Overall, this proved to be a highly successful approach, as we identified multiple polysaccharide-degrading species not captured through previous bioprospecting campaigns [37, 42, 48].

Among the CAZyme-rich species characterized in this study, we identified several yeast species from the Trichomonascaceae, CUG-Ser1 and the Phaffomycetaceae clades that were able to grow on xylan substrates. However, not all species enriched in CAZymes associated with xylan degradation grew according to the CAZyme-predictions and some surprisingly displayed relatively high xylanolytic activities but only modest xylan-growth, suggesting an inability to utilize the released oligosaccharides. Moreover, the CAZyme heatmap indicated potential for cellulolytic activity in several species, but this was not observed in the growth studies for the soluble cellulose analog CMC or crystalline and insoluble Avicel. For efficient degradation of cellulose to glucose, endoglucanase, cellobiohydrolase, lytic polysaccharide monooxygenase and  $\beta$ -glucosidase activities are generally regarded as needed [50–53]. The inability of yeasts to grow on cellulose may be due to the absence of one or several of the enzyme families known to encode such enzymes, such as GH6 (totally absent in the 332 yeasts) and GH7 (only present in *B. peoriensis*) and GH12 (only found in *B. mokoensis*). These results are well correlated with the lack of literature describing cellulose-growing yeasts, and only a few studies have reported cellulolytic enzymatic activities in yeasts [45, 55, 56]. Further analysis is needed to decipher the relationships between bioinformatic CAZyme prediction and measured growth and

enzymatic activities, but our results clearly emphasize the need to confirm bioinformatic predictions with wet lab characterization.

### Xylanolytic strategies in ascomycetous yeasts clades

The xylanolytic yeasts identified in this study display volumetric xylanase activities that range between 0.9 and 4.3 U mL<sup>-1</sup> (cell-associated and secreted activities combined), which can be compared to the xylanolytic activities of 28–30 U mL<sup>-1</sup> reported for filamentous fungi *Aspergillus niger*, *Aspergillus flavus* and *Trichoderma viride* [57, 58] and also to previous reports of *Sc. stipitis* (1.6 U mL<sup>-1</sup>) and *Su. lignohabitans* (0.29 U mL<sup>-1</sup>) [48, 59]. Strikingly, the different yeast species characterized here seem to have versatile approaches to degrade xylan. Most of the species that showed high enzymatic activities also possess CAZymes from the well-studied xylan-associated GH families GH5, GH10, GH11 and GH30. GH10 members were found in species in the Trichomonascaceae and CUG-Ser1 clades that preferentially display cell-associated xylanolytic activities. In contrast, *B. mokoensis* was the only species identified with a GH11 enzyme, which seems to be secreted in a soluble form, i.e. not attached to the yeast cell surface. In addition to GH10 and GH11, species in the Trichomonascaceae clade encoded enzymes from a range of additional families potentially containing xylan-acting enzymes that are completely lacking in the other clades. For example, enzymes from subfamily GH30\_5 (putative endo- $\beta$ -1,6-galactanase) were present in *B. mokoensis* and *Sp. europaea*, while a putative exo- $\beta$ -1,4-xylanase or glucuronoxylanase from GH30\_7 was present in *B. mokoensis*, *Su. lignohabitans* and *Sp. europaea* – three of the top scoring xylanolytic species found in our study. Yeasts in this clade were also found to contain genes for different xylanolytic de-branching enzymes, e.g., likely GH43, GH51 and GH62  $\alpha$ -L-arabinofuranosidases and GH67  $\alpha$ -glucuronidases. These could enable de-branching of complex xyans and reduce steric hindrances thus promoting the access to the xylan backbone for main-chain cleaving xylanases [60].

Along with the Trichomonascaceae species, we also identified xylanolytic species such as *W. canadensis* where the expected xylan-depolymerizing CAZyme setup seems to be absent. As mentioned above, the lack of predicted xylanolytic CAZymes in *W. canadensis* could be due to an incomplete genomic sequence [26]. However, other xylanolytic yeasts such as *Sp. passalidarum* and *Candida intermedia* [37, 42] for which complete or almost complete genomes exist, obvious xylanolytic enzyme-candidates are also missing (this study and unpublished results). This supports the hypothesis that these yeasts possess novel xylanases that should be explored more thoroughly. Further studies in terms of

gene deletion studies and/or heterologous expression and characterization of individual enzymes are needed to conclusively assign different enzyme activities of these putative enzymes.

### Future outlook

We here present the identification of several novel ascomycetous yeast species that can grow on polysaccharides and particularly on xylans. Future research includes careful physiological characterization of select species from the Trichomonascaceae, CUG-Ser1 and Phaffomycetaceae clades, to determine their precise growth requirements and full substrate ranges, tolerance-levels to industrial stressors and product portfolios. Additionally, xylanases are in high industrial demand for production of textiles, pulp and paper as well as in modern biotechnology for production of, for example, functional foods and feeds. Characterization of the many putative CAZymes identified in yeasts may provide new enzyme features in terms of fold, stability and specificity with potential to improve current processes and enable new applications. To assign physiological roles and substrate specificities to these enzymes, heterologous expression, purification and biochemical characterization will be needed.

### Conclusions

Yeast biodiversity presents a huge, untapped resource for present and future industrial applications such as CBP in terms of desirable microbial phenotypes and novel CAZyme discovery. In this study, we have developed a bioinformatic pipeline to rapidly process and predict CAZymes in a large number of genome sequenced ascomycetous yeasts. The resulting CAZyme predictions combined with growth and enzymatic activities assays enabled identification of several novel xylanolytic yeasts. Moreover, additional non-sequenced species with xylan-degrading capacity were identified through phylogenetic association. Many species identified and characterized here show equal or better xylanolytic activities compared to described species in literature such as *Scheffersomyces* and *Sugiyamaella* species, highlighting the potential of the approach. Collectively, the results presented expand our current knowledge on polysaccharide-degrading ascomycetous yeasts and opens up for numerous follow-up studies on yeast physiology and CAZyme characterization. The knowledge generated through such studies will be of high importance for the optimization of lignocellulosic biomass conversion processes.

### Abbreviations

AX: Arabinoxylan; CAZyme: Carbohydrate active enzyme; CBM: Carbohydrate-binding module; CE: Carbohydrate esterase; CMC: Carboxymethyl cellulose; CBP: Consolidated bioprocessing; DNS: Dinitro salicylate; GalMan: Galactomannan; GH: Glycoside hydrolase; GluMan: Glucomannan; GT: Glycoside transferases; GX: Glucuronoxylan; PL: Polysaccharide lyase; poly-MeGal: Polymethylgalacturonan; SDS-PAGE: Sodium dodecyl sulphate-polyacrylamide gel electrophoresis.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13068-021-01995-x>.

**Additional file 1: Table S1.** Table S1. CAZyme families grouped by polysaccharide degradation function for heatmap generation.

**Additional file 2: Fig. S1.** SDS-PAGE gel showing proteins secreted by *B. mokoensis* after three days of growth in xylan containing Delft medium.

**Additional file 3: Fig. S2.** Phylogenetic analysis of GH11 and GH10 xylanases.

### Acknowledgements

The authors would like to thank Dr. Marcel Taillefer for his expertise and analysis on MS/MS proteomics and Amanda Sörensen Ristinmaa for her assistance in the agar growth plate studies. The authors would also like to thank the ARS Culture Collection for providing yeast cultures.

### Authors' contributions

All authors conceived the project; JLR, CG and JL designed the experiments; JLR performed the experiments; ME performed bioinformatic work. JLR and CG interpreted the data and wrote the manuscript. JLR, CG, JL and ME revised the manuscript. All authors read and approved the final manuscript.

### Funding

Open access funding provided by Chalmers University of Technology. JLR, CG and JL would like to acknowledge Carl Trygger Fonden (Grant nr. CTS 18:118) for financial support of this research.

### Availability of data and materials

All data generated and analyzed in this study are included in the publication article and its additional information files. All code and data used in this study are made freely available for re-use. The bioinformatic analysis code is available as a GitHub repository ([https://github.com/EngqvistLab/yeast\\_xylanase](https://github.com/EngqvistLab/yeast_xylanase)) and the main output files, with information on predicted CAZyme domains, are downloadable from Zenodo (<https://zenodo.org/deposit/4548336>; <https://doi.org/10.5281/zenodo.4548336>).

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

All authors consented on the publication of this work.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Biology and Biological Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden. <sup>2</sup>Wallenberg Wood Science Center, Chalmers University of Technology, 412 96 Gothenburg, Sweden.

Received: 12 April 2021 Accepted: 14 June 2021

Published online: 02 July 2021

## References

- Fritsche U. Future transitions for the bioeconomy towards sustainable development and a climate-neutral. Publications Office of the European Union, JRC121212. 2020. <https://doi.org/10.2760/667966>
- Sheldon RA. The E factor 25 years on: The rise of green chemistry and sustainability. *Green Chem.* 2017;19(1):18–43. <https://doi.org/10.1039/c6gc02157c>.
- Zoghalmi A, Paës G. Lignocellulosic biomass: understanding recalcitrance and predicting hydrolysis. *Front Chem.* 2019;7:2. <https://doi.org/10.3389/fchem.2019.00874>.
- Heinze T. Cellulose chemistry and properties: fibers, nanocelluloses and advanced materials. *Adv Polym Sci.* 2015;271:47.
- Lundqvist J, et al. Isolation and characterization of galactoglucomannan from spruce (*Picea abies*). *Carbohydr Polym.* 2002;48(1):29–39. [https://doi.org/10.1016/S0144-8617\(01\)00210-7](https://doi.org/10.1016/S0144-8617(01)00210-7).
- Park YB, Cosgrove DJ. Xyloglucan and its interactions with other components of the growing cell wall. *Plant Cell Physiol.* 2015;56(2):180–94. <https://doi.org/10.1093/pcp/pcu204>.
- Wierzbicki MP, Maloney V, Mizrahi E, Myburg AA. Xylan in the middle: understanding xylan biosynthesis and its metabolic dependencies toward improving wood fiber for industrial processing. *Front Plant Sci.* 2019;10(February):1–29. <https://doi.org/10.3389/fpls.2019.00176>.
- Ebringerová A, Heinze T. Xylan and xylan derivatives - Biopolymers with valuable properties. 1: naturally occurring xylans structures, isolation procedures and properties. *Macromol Rapid Commun.* 2000;21(9):542–56.
- Mnich E, et al. Phenolic cross-links: building and de-constructing the plant cell wall. *Nat Prod Rep.* 2020. <https://doi.org/10.1039/c9np00028c>.
- Berglund J, et al. Wood hemicelluloses exert distinct biomechanical contributions to cellulose fibrillar networks. *Nat Commun.* 2020;11(1):1–16. <https://doi.org/10.1038/s41467-020-18390-z>.
- De Buck V, Polanska M, Van Impe J. Modeling biowaste biorefineries: a review. *Front Sust Food Syst.* 2020;4:7. <https://doi.org/10.3389/fsufs.2020.00011>.
- Den W, Sharma VK, Lee M, Nadadur G, Varma RS. Lignocellulosic biomass transformations via greener oxidative pretreatment processes: access to energy and value added chemicals. *Front Chem.* 2018;6:1–23. <https://doi.org/10.3389/fchem.2018.00141>.
- Arntzen M, Bengtsson O, Várnai A, Delogu F, Mathiesen G, Eijssink VGH. Quantitative comparison of the biomass-degrading enzyme repertoires of five filamentous fungi. *Sci Rep.* 2020;10(1):1–17. <https://doi.org/10.1038/s41598-020-75217-z>.
- Chukwuma OB, Rafatullah M, Tajarudin HA, Ismail N. Lignocellulolytic enzymes in biotechnological and industrial processes: a review. *Sustain.* 2020;12(18):1–31. <https://doi.org/10.3390/su12187282>.
- Chettri D, Verma AK, Verma AK. Innovations in CAZyme gene diversity and its modification for biorefinery applications. *Biotechnol Rep.* 2020;28:8. <https://doi.org/10.1016/j.btre.2020.e00525>.
- Minty JJ, Lin XN. Engineering synthetic microbial consortia for consolidated bioprocessing of lignocellulosic biomass into valuable fuels and chemicals. New York: Elsevier; 2015.
- Claes A, Deparis Q, Foulquié-Moreno MR, Thevelein JM. Simultaneous secretion of seven lignocellulolytic enzymes by an industrial second-generation yeast strain enables efficient ethanol production from multiple polymeric substrates. *Metab Eng.* 2020;59:131–41. <https://doi.org/10.1016/j.ymben.2020.02.004>.
- Lombard V, GolacondaRamulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 2014;42:490–5. <https://doi.org/10.1093/nar/gkt1178>.
- Davies G, Gilbert H, Henrissat B, Svensson B, Vocadlo D, Williams S. Ten years of CAZypedia: a living encyclopedia of carbohydrate-active enzymes. *Glycobiology.* 2018;28(1):3–8. <https://doi.org/10.1093/glycob/cwx089>.
- Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J.* 2004;382(3):769–81. <https://doi.org/10.1042/BJ20040892>.
- Sun Y, Cheng J. Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresour Technol.* 2002;83(1):1–11. [https://doi.org/10.1016/S0960-8524\(01\)00212-7](https://doi.org/10.1016/S0960-8524(01)00212-7).
- Zhao Z, Liu H, Wang C, Xu JR. Correction to Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics.* 2014;15:144. <https://doi.org/10.1186/1471-2164-15-6>.
- Despres J, et al. Xylan degradation by the human gut *Bacteroides xylanisolvens* XB1AT involves two distinct gene clusters that are linked at the transcriptional level. *BMC Genomics.* 2016;17(1):1–14. <https://doi.org/10.1186/s12864-016-2680-8>.
- Biely P, Kremnický L. Yeasts and their enzyme systems degrading cellulose, hemicelluloses and pectin. *Food Technology and Biotechnology.* 1998;36(4):305–12.
- Mukherjee V, Radecka D, Aerts G, Verstrepen KJ, Lievens B, Thevelein JM. Phenotypic landscape of non-conventional yeast species for different stress tolerance traits desirable in bioethanol fermentation. *Biotechnol Biofuels.* 2017;10(1):1–19. <https://doi.org/10.1186/s13068-017-0899-5>.
- Shen XX, et al. Tempo and mode of genome evolution in the budding yeast subphylum. *Cell.* 2018;175(6):1533–1545.e20. <https://doi.org/10.1016/j.cell.2018.10.023>.
- Zhang H, et al. DbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 2018;46(W1):W95–101. <https://doi.org/10.1093/nar/gky418>.
- Libkind D, et al. Into the wild: new yeast genomes from natural environments and new tools for their analysis. *FEMS Yeast Res.* 2020;20(2):1–15. <https://doi.org/10.1093/femsyr/foaa008>.
- Kurtzman CP, Fell JW, Boekhout T. Definition, classification and nomenclature of the yeasts, vol. 1. New York: Elsevier; 2011.
- Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 2006;22(13):1658–9. <https://doi.org/10.1093/bioinformatics/btl158>.
- Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol.* 2011;7:10. <https://doi.org/10.1371/journal.pcbi.1002195>.
- AlmagroArmenteros JJ, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* 2019;37(4):420–3. <https://doi.org/10.1038/s41587-019-0036-z>.
- Cock PJA, et al. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics.* 2009;25(11):1422–3. <https://doi.org/10.1093/bioinformatics/btp163>.
- Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol Biol Evol.* 2016;33(6):1635–8. <https://doi.org/10.1093/molbev/msw046>.
- Thiruvathukal EGK, Hunter BJD. M Atplotlib: a 2D G Rraphics E Nvironment. Berlin: Springer; 2007. p. 90–5.
- Hendriks ATWM, van Lier JB, de Kreuk MK. Growth media in anaerobic fermentative processes: The underestimated potential of thermophilic fermentation and anaerobic digestion. *Biotechnol Adv.* 2018;36(1):1–13. <https://doi.org/10.1016/j.biotechadv.2017.08.004>.
- Lara CA, et al. "Identification and characterisation of xylanolytic yeasts isolated from decaying wood and sugarcane bagasse in Brazil", *Antonie van Leeuwenhoek.* Int J Gen Mol Microbiol. 2014;105(6):1107–19. <https://doi.org/10.1007/s10482-014-0172-x>.
- Miller GL. Use of Dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem.* 1959;31(3):426–8. <https://doi.org/10.1021/ac60147a030>.
- Ghose TK, Bisaria VS. Measurement of hemicellulase activities part 1: xylanases. *Pure Appl Chem.* 1987;59(12):1739–51. <https://doi.org/10.1351/pac198759121739>.
- Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7. <https://doi.org/10.1093/nar/gkh340>.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018;35(6):1547–9. <https://doi.org/10.1093/molbev/msy096>.
- Morais CG, Cadete RM, Uetanabaro APT, Rosa LH, Lachance MA, Rosa CA. D-xylose-fermenting and xylanase-producing yeast species from rotting wood of two Atlantic Rainforest habitats in Brazil. *Fungal Genet Biol.* 2013;60:19–28. <https://doi.org/10.1016/j.fgb.2013.07.003>.
- Suh SO, Houseknecht JL, Gujjari P, Zhou JJ. *Scheffersomyces parashehatae* f.a., sp. nov., *Scheffersomyces xylosifermentans* f.a., sp. nov., *Candida broadrunensis* sp. nov. and *Candida manassasensis* sp. Nov., novel yeasts associated with wood-ingesting insects, and their ecological and biofuel implications. *Int J Syst Evol Microbiol.* 2013;63:4330–9. <https://doi.org/10.1099/ijso.0.053009-0>.
- Kurtzman CP. *Blastobotrys americana* sp. nov., *Blastobotrys illinoisensis* sp. nov., *Blastobotrys malaysiensis* sp. nov., *Blastobotrys muscicola* sp. nov., *Blastobotrys peoriensis* sp. nov. and *Blastobotrys raffinosifermentans*



- sp. nov., novel anamorphic yeast species. *Int J Syst Evol Microbiol*. 2007;57(5):1154–62. <https://doi.org/10.1099/ijse.0.64847-0>.
45. Goldbeck R. Screening and identification of cellulase producing yeast-like microorganisms from Brazilian biomes. *African J Biotechnol*. 2012;11(53):11595–603. <https://doi.org/10.5897/ajb12.422>.
  46. Pérez-González JA, De Graaff LH, Visser J, Ramón D. Molecular cloning and expression in *Saccharomyces cerevisiae* of two *Aspergillus nidulans* xylanase genes. *Appl Environ Microbiol*. 1996;62(6):2179–82. <https://doi.org/10.1128/aem.62.6.2179-2182.1996>.
  47. Morel G, et al. Differential gene retention as an evolutionary mechanism to generate biodiversity and adaptation in yeasts. *Sci Rep*. 2015;5:1–18. <https://doi.org/10.1038/srep11571>.
  48. Sena LMF, et al. "d-Xylose fermentation, xylitol production and xylanase activities by seven new species of Sugiyamaella", *Antonie van Leeuwenhoek*. *Int J Gen Mol Microbiol*. 2017;110(1):53–67. <https://doi.org/10.1007/s10482-016-0775-5>.
  49. Olson DG, McBride JE, Joe Shaw A, Lynd LR. Recent progress in consolidated bioprocessing. *Curr Opin Biotechnol*. 2012;23(3):396–405. <https://doi.org/10.1016/j.copbio.2011.11.026>.
  50. Keller MB, Sørensen TH, Krogh KBRM, Wogulis M, Borch K, Westh P. Biotechnology for biofuels activity of fungal  $\beta$ -glucosidases on cellulose. *Biotechnol Biofuels*. 2020;2020(7):1–7. <https://doi.org/10.1186/s13068-020-01762-4>.
  51. Rahikainen J, Ceccherini S, Molinier M, Gro S, Suurna A. Effect of cellulase family and structure on modification of wood fibres at high consistency, vol. 0123456789. Berlin: Springer; 2019. p. 5085–103.
  52. Østby H, Hansen LD, Horn SJ, Eijsink VGH, Várnai A. Enzymatic processing of lignocellulosic biomass: principles, recent advances and perspectives, vol. 47. Berlin: Springer; 2020. p. 9–10.
  53. Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VG. Novel enzymes for the degradation of cellulose. *Biotechnol Biofuels*. 2012;5:45.
  54. Kulon U, Park N. Screening of amylolytic and cellulolytic yeast from *Dendrobium spathilingue* in Bali botanical garden, Indonesia screening of amylolytic and cellulolytic yeast from *Dendrobium spathilingue* in Bali botanical garden, Indonesia. *AIP Conf Proc*. 2020;2242:050013.
  55. Shirnalli G, Ashwini M, Gachhi D. Isolation and evaluation of cellulolytic yeasts for production of ethanol from wheat straw. *PLoS ONE*. 2018;7(12):2215–21.
  56. Kim J, et al. Isolation and characterization of cellulolytic yeast belonging to *Moesziomyces* sp. from the gut of Grasshopper. *Korean J. Microbiol*. 2019;55(3):234–41.
  57. Oyediji O, Iluyomade A, Egbewumi I, Odufuwa A. Isolation and screening of xylanolytic fungi from soil of botanical garden: xylanase production from *Aspergillus flavus* and *Trichoderma viride*. *J Microbiol Res*. 2018;8(1):9–18. <https://doi.org/10.5923/j.microbiology.20180801.02>.
  58. Sridevi A, Sandhya A, Ramanjaneyulu G, Narasimha G, Devi PS. Biocatalytic activity of *Aspergillus niger* xylanase in paper pulp biobleaching. *3 Biotech*. 2016;6(2):1–7. <https://doi.org/10.1007/s13205-016-0480-0>.
  59. Lee H, Biely P, Latta RK. Utilization of xylan by yeasts and its conversion to ethanol by *Pichia stipitis* strains. *Appl Environ Microbiol*. 1986;52(2):320–4. <https://doi.org/10.1128/aem.52.2.320-324.1986>.
  60. Jia L, et al. Synergistic degradation of arabinoxylan by free and immobilized xylanases and arabinofuranosidase. *Biochem Eng J*. 2016;114:268–75. <https://doi.org/10.1016/j.bej.2016.07.013>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

