Insights into the genome and secretome of *Fusarium metavorans* DSM105788 by cultivation on agro-residual biomass and synthetic nutrient sources

Sophie C. Brandt1†, Hévila Brognaro2,3†, Arslan Ali3,4,5, Bernhard Ellinger6, Katharina Maibach7, Martin Rühl8, Carsten Wrenger2,3, Hartmut Schlüter3,5, Wilhelm Schäfer1, Christian Betzel3, Stefan Janssen7 and Martin Gand1,8*

Abstract

**Background:** The transition to a biobased economy involving the depolymerization and fermentation of renewable agro-industrial sources is a challenge that can only be met by achieving the efficient hydrolysis of biomass to monosaccharides. In nature, lignocellulosic biomass is mainly decomposed by fungi. We recently identified six efficient cellulose degraders by screening fungi from Vietnam.

**Results:** We characterized a high-performance cellulase-producing strain, with an activity of 0.06 U/mg, which was identified as a member of the *Fusarium solani* species complex linkage 6 (*Fusarium metavorans*), isolated from mangrove wood (FW16.1, deposited as DSM105788). The genome, representing nine potential chromosomes, was sequenced using PacBio and Illumina technology. In-depth secretome analysis using six different synthetic and artificial cellulose substrates and two agro-industrial waste products identified 500 proteins, including 135 enzymes assigned to five different carbohydrate-active enzyme (CAZyme) classes. The *F. metavorans* enzyme cocktail was tested for saccharification activity on pre-treated sugarcane bagasse, as well as untreated sugarcane bagasse and maize leaves, where it was complemented with the commercial enzyme mixture Accellerase 1500. In the untreated sugarcane bagasse and maize leaves, initial cell wall degradation was observed in the presence of at least 196 µg/mL of the in-house cocktail. Increasing the dose to 336 µg/mL facilitated the saccharification of untreated sugarcane biomass, but had no further effect on the pre-treated biomass.

**Conclusion:** Our results show that *F. metavorans* DSM105788 is a promising alternative pre-treatment for the degradation of agro-industrial lignocellulosic materials. The enzyme cocktail promotes the debranching of biopolymers surrounding the cellulose fibers and releases reduced sugars without process disadvantages or loss of carbohydrates.

**Keywords:** *Fusarium solani* species complex, *Fusarium metavorans*, Genome analysis, Secretome, Mass spectrometry, Proteomics, CAZyme analysis, Cellulose degradation, Residual biomass treatment

© 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.
be considered in the context of sustainable agriculture to avoid competition with food and feed production [2]. Biotechnological approaches are therefore required to valorize non-edible biomass, focusing on abundant sources such as forestry and agricultural wastes [3]. Sugarcane is the dominant crop in tropical areas such as South America and South Asia [4], whereas maize dominates in sub-tropical and temperate regions such as North America and Northern Europe [5]. The widespread agricultural use of these two C4 crops generates large quantities of lignocellulosic biomass that can be valorized without compromising food/feed production.

Lignocellulosic biomass has a heterogeneous structure and composition dependent on the plant species [6–8]. The main component is cellulose, the most abundant polymer on earth, consisting of linear chains of several hundred to many thousand β-(1,4)-d-glucose units. The other components are hemicellulose, pectin, lignin and extractives, the latter comprising a diverse range of substances that impede the enzymatic hydrolysis of biomass [9, 10]. Hemicellulose, the second most abundant polymer in plant cell walls [11], features at least six different macromolecules with varying ratios of pentose (xylose and arabinose) and hexose (mostly mannose and glucose) residues [12]. Xylans have a linear backbone of β-(1,4)-linked β-d-xylopyranosyl residues, whereas glucuronoxylans feature substituted 4-O-methyl-α-d-glucuronopyranosyl units and acetyl groups, and arabinoxylans contain xylose substituted with α-L-arabinofuranosyl units [11]. Xyloglucans have a cellulose-like linear backbone of β-(1,4)-d-glucose with additional β-(1,6)-linked xylose sidechains, often capped with galactose and fucose [13]. Glucomannans have backbones of β-(1,4)-linked d-mannose and d-glucose, sometimes with branching β-(1,6)-glucosyl residues [14], but if α-(1,6)-linked galactose units are present the polymers are known as galactoglucomanans [15]. Pectin is a complex heteropolymer of covalently linked d-galacturonic acid and other residues, and is also a significant component of sugarcane and maize bagasse [16]. The main constituents are (1) homogalacturonan, comprising linear α-(1,4)-d-galacturonic acid chains with some esterified or O-acetylated modifications; (2) rhamnogalacturonan-I, comprising repeated disaccharides of galacturonic acid and C-3 or C-2 O-acetylated rhamnosyl residues, with linear or branched α-L-arabinofuranosyl and/or galactopyranosyl side chains on C-4; and (3) substituted galacturonans as linear and side chain residues (rhamnogalacturonan-II), resulting in 12 types of glycosyl units that form at least 22 types of glycosidic bonds [17].

The recalcitrance of lignocellulosic biomass in part reflects the complexity of the substrate, with complete hydrolysis requiring efficient enzymes for the digestion of cellulose as well as palettes of enzymes that can digest the components of hemicellulose [18] and pectin [19]. However, enzymatic hydrolysis is also impeded by the inaccessibility of the substrates, which can be addressed by physical and/or chemical pre-treatment. Such processes can generate inhibitors that limit the activity of cellulases and other enzymes, as well as toxic molecules such as furfural, acetic acid, formic acid and lignin-phenolic compounds that interfere with fermentation [20]. The effect of biomass pre-treatment [21, 22] can therefore be improved by optimizing the enzymatic cocktails used to hydrolyze lignocellulosic biomass, tailoring them for the type of biomass and for the ability to tolerate inhibitors [1, 9, 10, 23]. Although the polysaccharide content of maize leaf and sugarcane culm cell walls is similar [24, 25], the cross-linking of polysaccharides and the interactions between polysaccharide and lignin/phenolic compounds differ, resulting in unique cell wall architectures. The physical and chemical characteristics of the biomass therefore reflect variations in the degree of cellulose polymerization, crystallinity, and lignin content, the hemicellulose and pectin content, and cell wall thickness [26].

Lignocellulosic biomass in nature is mainly decomposed by fungi, which are therefore promising candidates for the discovery of enzymes or enzyme cocktails for biomass degradation [27]. More than 5 million species of fungi have been described, and the number is likely to increase given that only 5% of species are formally classified [28, 29]. The subkingdom Dikarya consists of two phyla: Ascomycota, the largest phylum, commonly known as sac fungi [30], and Basidiomycota, the second largest phylum, commonly known as higher mushrooms or pillar fungi. The filamentous ascomycetes are ubiquitous and Fusarium is one of the most abundant genera in that phylum [31]. Fusarium species are frequently isolated from tropical, sub-tropical, and temperate environments, and less frequently from alpine habitats [32]. The genus Fusarium was first described at the beginning of the nineteenth century [33, 34]. Nine species have been described, including the easily recognized Fusarium solani, based on its striking morphology [35]. However, the current concept of F. solani is a species complex (FSSC) within the class Sordariomycetes, order Hypocreales, and the family Nectriaceae. The FSSC is thought to contain at least 60 phylogenetically distinct but closely related and morphologically similar species [36], and is allied with the sexual species Nectria haematococca. Robust classification within the FSSC and the genus Fusarium is achieved by analyzing polymorphisms in the genes encoding translation elongation factor 1α (TEF1) and the second largest subunit of RNA polymerase II (RPB2) as well as the internal transcribed spacer.
(ITS) together with 28S ribosomal RNA (ITS + 28S) [36–38]. Members of the FSSC collectively have a broad host range and can be found as soil-dwelling saprophytes, rhizosphere colonizers, or pathogens of pea, bean, potato, soybean, maize and many cucurbit plants, as well as animals including humans [39]. *Fusarium* sp. of the FSSC has 5–17 chromosomes, with a genome size of 40–54 Mbp and a GC content of ~50% [35, 40–42].

Our previously reported analysis of 295 fungal isolates, collected from different substrates and various environments in Vietnam, revealed their ability to degrade lipids, chitin, cellulose and xylan [43]. Six isolates were able to digest carboxymethylcellulose (CMC) with remarkable efficiency, two of which were *Fusarium* strains. We selected the most active member of FSSC linkage 6, isolated from dead mangrove wood, for further analysis. We characterized this strain as *F. metavorans* FW16.1 by analyzing its genome and secretome, leading to the identification of undiscovered lignocellulose degrading enzymes with the ability to convert sugarcane bagasse and maize leaves into fermentable sugars.

### Results

**Characterization, genomic analysis and phylogenetics of *F. metavorans* FW16.1**

We tested the carboxymethylcellulase (CMCase) activity of *F. metavorans* FW16.1 on media containing 1% CMC 3 days after inoculation, revealing a value of 0.055 ± 0.001 U/mg (Additional file 1: Table S1). Genomic DNA was isolated and analyzed by agarose gel electrophoresis (Additional file 1: Fig. S1) and the ITS region was amplified and sequenced (Additional file 1: Supplementary Data). Sequencing identified the isolate as a *F. solani* strain in the FSSC. The strain is preserved at the German Collection of Microorganisms and Cell Cultures (DSMZ) under the identifier DSM105788. The assembled FW16.1 genome was 48.28 Mbp in length, distributed over nine scaffolds with a GC content of 50.83% and an N50 scaffold length (weighted median of a contig in genomic sequence) following assembly with SOAPdenovo was k = 15 bp, with a pkdepth (peak depth estimated from k-mer distribution) of 30. Gene prediction revealed the presence of 15,626 putative open reading frames (ORFs) with an average of 1618.9 bp per gene or 1459.85 bp per coding sequence. The whole genome is available as a biosample from the National Center for Biotechnology Information (NCBI) under the bioproject PRJN413482, accession number JADNRB000000000. Phylogenetic analysis assigned FW16.1 to the FSSC 6 linkage, with highest similarity to *F. metavorans* NRRL 43489 (Fig. 1). Growth on six different media resulted in the formation of pale mycelia (Fig. 2).

---

**Carbohydrate-active enzyme analysis**

The FW16.1 genomic regions marked as protein coding sequences (CDS) in our de novo assembly were searched for homologs of families (and subfamilies) in the CAZyme database representing enzymes involved in cellulose and sugar metabolism, revealing 694 putative genes (Fig. 3; Table 1). The candidates were assigned to five different carbohydrate-active enzyme (CAZyme) classes, which were divided into their families (Additional file 1: Table S2).

**Evaluation of enzymatic activity**

FW16.1 was cultivated in liquid yeast extract peptone dextrose (YPD) medium, and the enzymatic activity of the supernatant was tested. We observed CMCase activity that increased over the first 2 days, reaching a plateau of ~19.5 ± 0.3 U/mg that lasted until day 5. A further increase in activity on days 6 and 7 led to a new plateau at ~30 U/mg (Additional file 1: Fig. S2). We then measured enzyme activity induced by cultivation in a range of liquid media containing synthetic and artificial cellulose substrates for 72 h. The activity of the FW16.1 supernatant was 0.039 ± 0.001 U/mg against the crystalline cellulose Avicel PH-101 (Additional file 1: Fig. S3), increasing to 0.07 ± 0.01 U/mg against α-cellulose, and 0.18 ± 0.06 U/mg against hydroxyethylcellulose (HEC). The specific activity against high, medium and low-viscosity forms of CMC, described hereafter as H-CMC, M-CMC and L-CMC for simplicity, was comparable (ranging from 0.07 ± 0.01 to 0.1 ± 0.01 U/mg). We also tested the activity of FW16.1 against agro-residual biomass (sugarcane bagasse and maize leaves) focusing on the properties of the crude secretome. We therefore prepared lyophilized secretome fractions from both biomass types and resuspended them at a 1:1 ratio. The highest polygalacturonase and laminarinase activity was observed after 24 h, whereas the highest CMCase and xylanase activity was observed after 96 h (Additional file 1: Fig. S4). We observed little activity against arabinan, arabinoxylin, galactan, pectin and starch, either due to low enzymatic specificity for these substrates or the low sensitivity of 3,5-dinitrosalicylic acid (DNS) assay.

**Secretome profiling of *F. metavorans* on synthetic substrates and agro-residual biomass**

Tandem mass spectrometric proteomics was used to analyze the FW16.1 secretome fractions, revealing the presence of 500 proteins (Additional file 1: Table S3). Different numbers of proteins were identified on each substrate, ranging from 122 for α-cellulose to 235 for...
H-CMC. We identified 124 proteins on Avicel PH-101, 144 on M-CMC, 160 on HEC, 174 on sugarcane bagasse, 176 on maize leaves and 202 on L-CMC. We identified 284 proteins on synthetic or artificial cellulose alone, with the number of unique proteins ranging from six on α-cellulose and Avicel PH-101 to 65 on H-CMC. We identified 13 unique proteins on M-CMC, 26 on HEC, and 31 on L-CMC. We identified 78 proteins solely in the sugarcane bagasse and maize leaf secretome fractions, 23 unique to sugarcane and 31 unique to maize. The largest number of proteins was co-expressed when FW16.1 was grown on the agro-residual biomass, suggesting some of the proteins may be involved in processes not related to energy metabolism (Fig. 4). The second largest number of proteins was co-expressed when FW16.1 was grown on synthetic and artificial cellulose substrates, reflecting the subset of genes required to metabolize these polymers. The third largest number of proteins was common...
to all conditions, including general sugar conversion and homeostasis genes. Interestingly, the fourth largest group of proteins found on more than one substrate was identified on the CMC media, representing genes specifically required for this artificial substrate. These findings indicate that FW16.1 can fine-tune the expression of relevant genes enabling its survival in different habitats.

The theoretical protein distribution was plotted as a function of isoelectric point (pI) (Fig. 5a) and molecular weight (MW) (Fig. 5b), revealing that 90% of the secretome proteins fell within the MW range 6.5–263.4 kDa (median = 40.8 kDa) and the pI range 2.9–11.8 (median = 5.4). On the six synthetic and artificial cellulose substrates, the median size of the secretome was 38.5–39.5 kDa, but this shifted to 42.5 and 45.1 kDa on the two biomass substrates. Similarly, the median pI was 5.3–5.6 on the synthetic and artificial cellulose substrates, but shifted to 5.0 and 5.1 on maize and sugarcane bagasse, respectively. This effect appears small, but the pI has a logarithmic scale and more than 135 proteins were analyzed for both parameters, resulting in significant deviations (\( p < 0.0001 \)) based on an unpaired \( t \)-test assuming Gaussian distribution (Fig. 5).

To gain insight into the metabolic diversity of the secretome on each substrate, the identified proteins were classified according to biological function (Fig. 5c) based on the sequences listed in Additional file 1: Table S3. Several molecular functions were identified, including carbohydrate, lipid, RNA and amino acid metabolism, protein synthesis, redox processes, proteolysis, and proteins with unknown functions. The proteins identified on the synthetic and artificial cellulose substrates were distributed similarly according their molecular functions, whereas the relative frequency of proteins related to carbohydrate metabolism was higher on the biomass substrates. The substrate-dependent profiles of the 135 CAZymes are shown in Fig. 5d; a complete list of identified CAZymes with associated modules is provided in Table 2. Predictions based on putative molecular functions for all proteins are summarized in Additional file 1: Table S3. The 135 CAZymes were assigned to five different classes (Table 2): 93 glycoside hydrolases (GHs), 17 auxiliary activities (AAs), 12 carbohydrate esterases (CEs), 12 polysaccharide lyases (PLs), and one glycosyltransferase (GT), as well as three non-catalytic carbohydrate-binding modules (CBMs). The distribution over the scaffolds is presented in Fig. 6.

The most abundant CAZyme class was the GHs (36–58%, or 62–73% when including GHs with CBMs), confirming their general role in cellulose degradation. The lowest proportion of GHs (36%) was identified on the substrate HEC, which also featured the highest proportion of CE\( s \) (16%) compared to the median 11%. Another unusual profile was the overrepresentation of PLs on
natural substrates, with 7–8 enzymes (10%) compared to 1–3 (2–6%) on the synthetic and artificial cellulose substrates. The natural substrates also featured more GHs (57–58%), but fewer GH-CBMs (12–14%) than the synthetic and artificial cellulose substrates, which featured 47% GHs and 22% GH-CBMs on average.

Some CAZymes were produced on all substrates, whereas others were more specific. When the "core" GH family proteins (produced on at least five of the six synthetic and artificial cellulosics) were evaluated, few differences were found: two GH5, one GH6, two GH7 (all five with CBM1) one GH10, one GH11, one GH16, one GH43, one GH55, one GH71-CBM24-CBM24, one GH72-CBM43, one GH74 and one GH75. Predicted cellulase activities were confirmed in several cases: endo-β-(1,4)-glucanase activity for FW16 GLEAN_10000416 (GH5-CBM1), celllobiohydrolase activity for FW16_GLEAN_10006835 (GH6-CBM1), reducing-end celllobiohydrolase activity for FW16_GLEAN_10001888 and FW16_GLEAN_10005918 (both GH7-CBM1), and potentially xyloglucanase activity for FW16_GLEAN_10000631 (GH74). Remarkably, no GH with predicted β-glucosidase activity was found on the cellulose and cellulose-like substrates, whereas FW16_GLEAN_10003711 (GH1) and FW16_GLEAN_10003498 (GH3) were found on four of the six substrates. Furthermore, enzymes with predicted β-(1,3)-glucanase activity representing GH families 16, 17, 55, 72, 81, 128 and 132 were found mostly on the artificial cellulose substrates and especially on the CMCs.
suggesting the broader substrate specificity of those enzymes or a weak catalytic promiscuity. More diverse cellulose-degrading enzymes were identified on the biomass: FW16_GLEAN_10000066 (GH2), FW16_GLEAN_10001573 (GH10) and FW16_GLEAN_10013304 (GH11), FW16_GLEAN_10003286, FW16_GLEAN_10006822 and FW16_GLEAN_10010955 (GH43). However, a similar distribution was found under both conditions for β-galactosidase (FW16_GLEAN_10000066, GH2); a potential xylan β-(1,4)-xylosidase (FW16_GLEAN_10011639, GH3), β-(1,3)-glucosidase (FW16_GLEAN_10010890, GH17), exo-polygalacturonase (FW16_GLEAN_10001538 and FW16_GLEAN_10005091, both GH28), and potential disaccharide hydrolases such as FW16_GLEAN_10003104 (GH9), FW16_GLEAN_10001518, FW16_GLEAN_10009840 and FW16_GLEAN_10011918 (all GH43) and two exo-α-L-(1,5)-arabinanases (FW16_GLEAN_10001805 and FW16_GLEAN_10011917, both GH93). GH proteins identified solely on sugarcane bagasse were related to xylan, amylase and dextran degradation (GH10, GH13, and four of the 11 GH43 and GH49 proteins). In contrast, those identified solely on maize leaves were primarily related to disaccharide hydrolysis, including FW16_GLEAN_10006734 (GH1), FW16_GLEAN_10003498 and FW16_GLEAN_10008834 (both GH3, β-glucosidase), FW16_GLEAN_1000618 (GH35, β-galactosidase or β-(1,3)-galactosidase) and two of the 11 GH43 proteins (FW16_GLEAN_10001775 β-p-galactofuranosidase, and FW16_GLEAN_10001821, predicted arabinanase or xylosidase).

A clearer picture emerged for the AAs. The synthetic and artificial cellulose substrates mainly featured AA9 proteins with lytic cellulose monooxygenase activity, to the biomass: FW16_GLEAN_10000066 (GH2), FW16_GLEAN_10001573 (GH10) and FW16_GLEAN_10013304 (GH11), FW16_GLEAN_10003286, FW16_GLEAN_10006822 and FW16_GLEAN_10010955 (GH43). However, a similar distribution was found under both conditions for β-galactosidase (FW16_GLEAN_10000066, GH2); a potential xylan β-(1,4)-xylosidase (FW16_GLEAN_10011639, GH3), β-(1,3)-glucosidase (FW16_GLEAN_10010890, GH17), exo-polygalacturonase (FW16_GLEAN_10001538 and FW16_GLEAN_10005091, both GH28), and potential disaccharide hydrolases such as FW16_GLEAN_10003104 (GH9), FW16_GLEAN_10001518, FW16_GLEAN_10009840 and FW16_GLEAN_10011918 (all GH43) and two exo-α-L-(1,5)-arabinanases (FW16_GLEAN_10001805 and FW16_GLEAN_10011917, both GH93). GH proteins identified solely on sugarcane bagasse were related to xylan, amylase and dextran degradation (GH10, GH13, and four of the 11 GH43 and GH49 proteins). In contrast, those identified solely on maize leaves were primarily related to disaccharide hydrolysis, including FW16_GLEAN_10006734 (GH1), FW16_GLEAN_10003498 and FW16_GLEAN_10008834 (both GH3, β-glucosidase), FW16_GLEAN_1000618 (GH35, β-galactosidase or β-(1,3)-galactosidase) and two of the 11 GH43 proteins (FW16_GLEAN_10001775 β-p-galactofuranosidase, and FW16_GLEAN_10001821, predicted arabinanase or xylosidase).

A clearer picture emerged for the AAs. The synthetic and artificial cellulose substrates mainly featured AA9 proteins with lytic cellulose monooxygenase activity, to the biomass: FW16_GLEAN_10000066 (GH2), FW16_GLEAN_10001573 (GH10) and FW16_GLEAN_10013304 (GH11), FW16_GLEAN_10003286, FW16_GLEAN_10006822 and FW16_GLEAN_10010955 (GH43). However, a similar distribution was found under both conditions for β-galactosidase (FW16_GLEAN_10000066, GH2); a potential xylan β-(1,4)-xylosidase (FW16_GLEAN_10011639, GH3), β-(1,3)-glucosidase (FW16_GLEAN_10010890, GH17), exo-polygalacturonase (FW16_GLEAN_10001538 and FW16_GLEAN_10005091, both GH28), and potential disaccharide hydrolases such as FW16_GLEAN_10003104 (GH9), FW16_GLEAN_10001518, FW16_GLEAN_10009840 and FW16_GLEAN_10011918 (all GH43) and two exo-α-L-(1,5)-arabinanases (FW16_GLEAN_10001805 and FW16_GLEAN_10011917, both GH93). GH proteins identified solely on sugarcane bagasse were related to xylan, amylase and dextran degradation (GH10, GH13, and four of the 11 GH43 and GH49 proteins). In contrast, those identified solely on maize leaves were primarily related to disaccharide hydrolysis, including FW16_GLEAN_10006734 (GH1), FW16_GLEAN_10003498 and FW16_GLEAN_10008834 (both GH3, β-glucosidase), FW16_GLEAN_1000618 (GH35, β-galactosidase or β-(1,3)-galactosidase) and two of the 11 GH43 proteins (FW16_GLEAN_10001775 β-p-galactofuranosidase, and FW16_GLEAN_10001821, predicted arabinanase or xylosidase).

A clearer picture emerged for the AAs. The synthetic and artificial cellulose substrates mainly featured AA9 proteins with lytic cellulose monooxygenase activity,

Table 1 Identified CAZyme classes based on the *F. metavorans* FW16.1 genome annotation using RAPSearch2 search and HMMER scanning

| CAZyme classes                                      | Number of detected genes |
|----------------------------------------------------|--------------------------|
| Glycoside hydrolases (GH)                          | 322                      |
| Glycoside hydrolases (GH) with carbohydrate-binding modules (CBM) | 32                       |
| Carbohydrate esterases (CE)                        | 54                       |
| Carbohydrate esterases (CE) with CBM               | 2                        |
| Polysaccharide lyases (PL)                         | 31                       |
| Polysaccharide lyases (PL) with CBM                | 3                        |
| Glycosyl transferases (GT)                         | 103                      |
| Glycosyl transferases (GT) with CBM                | 0                        |
| Auxiliary activities (AA)                          | 142                      |
| Auxiliary activities (AA) with CBM                 | 5                        |

Fig. 4 Co-expression of proteins found by mass spectrometric proteomics in different *F. metavorans* secretomes. The connected black dots in the lower part of the figure indicate growth conditions resulting in the expression of a shared set of proteins. The number of proteins found under the specific condition is shown by the size of the black bars. The combinations are sorted to first show the unique proteins, which are specific for certain growth media, and the set of proteins found in all conditions last. From all the possible combinations, only those with more than two co-expressed proteins are shown.
whereas the biomass substrates showed a greater diversity of AA families. Some were predicted to modify lignin, such as the laccases FW16_GLEAN_10001275 and FW16_GLEAN_10013360 (both AA1), the alcohol oxidase FW16_GLEAN_10000205 (AA3), the cellobiose dehydrogenase FW16_GLEAN_10000721 (AA3), and glyoxal oxidase FW16_GLEAN_10000164 (AA5). Interestingly, no AA9 proteins were found on maize leaves, but two of the four identified AA9 proteins were found on sugarcane bagasse.

Among the 12 identified PLs, six were found on sugarcane bagasse and 10 were found on maize, highlighting their role in pectin degradation. Only 1–3 PLs were found on the synthetic and artificial cellulose substrates, with FW16_GLEAN_10000207 (PL20, predicted endo-β-(1,4)-glucuronan lyase) present on five of the six cellulase substrates but not on the biomass substrates.

We identified 4–5 CEs restricted to the synthetic and artificial cellulose substrates, five produced on sugarcane, and seven produced on maize. In the latter case, roles in hemicellulose and pectin degradation are likely, such as FW16_GLEAN_10004777 (CE1) and FW16_GLEAN_10007169 (CE5), both with predicted (acetyl)xylan esterase activity, FW16_GLEAN_10001547 and FW16_GLEAN_10001601 (both CE8), FW16_GLEAN_10012229 and FW16_GLEAN_10013316 (both CE12), all four with predicted pectinase activity. Sugarcane bagasse contained both CE12 enzymes also found on maize leaves, as well as one common CE8 and CE4 protein, and the CE1 protein FW16_GLEAN_10014832 with predicted feruloyl esterase activity. CEs solely present on the synthetic and artificial cellulose substrates included FW16_GLEAN_10001089 (CE2, acetylxylan esterase), FW16_GLEAN_10006900 (CE5, pectin esterase), FW16_GLEAN_10011996 (CE8, cutinase).
| No | Accession IDs         | Family | Molecular function prediction | Possible substrate(s) | Carbon Source |
|----|-----------------------|--------|-------------------------------|-----------------------|--------------|
|    |                       |        |                               |                       |              |
| 1  | FW16_GLEAN_10001275   | AA1    | Laccase / Laccase-like        | found                 | L- CMC       |
|    |                       |        | multicomponent oxidase        | found                 | M- CMC       |
| 2  | FW16_GLEAN_1001360    | AA1    | Laccase / Laccase-like        | found                 | H- CMC       |
|    |                       |        | multicomponent oxidase        | found                 | HEC          |
| 3  | FW16_GLEAN_10001928   | AA2    | Manganese peroxidase /        | found                 | α- cellulose  |
|    |                       |        | versatile peroxidase /        | found                 | Avicel PJH-101 |
|    |                       |        | lignin peroxidase             | found                 | MZ           |
| 4  | FW16_GLEAN_10000205   | AA3    | Alcohol oxidase               | found                 |              |
| 5  | FW16_GLEAN_10000721   | AA3    | Cellulose dehydrogenase       | found                 |              |
| 6  | FW16_GLEAN_10000164   | AA5    | Glycerol oxidase              | found                 |              |
| 7  | FW16_GLEAN_10006136   | AA6    | 1,4-Semiquinone reductase     | found                 |              |
| 8  | FW16_GLEAN_10008842   | AA7    | Glucosylsaccharide oxidase    | found                 |              |
|    |                       |        | / chitosylsaccharide oxidase  | found                 |              |
| 9  | FW16_GLEAN_10004754   | AA7    | Glucosylsaccharide oxidase    | found                 |              |
|    |                       |        | / chitosylsaccharide oxidase  | found                 |              |
| 10 | FW16_GLEAN_10014336   | AA7    | Glucosylsaccharide oxidase    | found                 |              |
|    |                       |        | / chitosylsaccharide oxidase  | found                 |              |
| 11 | FW16_GLEAN_10012260   | AA7    | Glucosylsaccharide oxidase    | found                 |              |
|    |                       |        | / chitosylsaccharide oxidase  | found                 |              |
| 12 | FW16_GLEAN_10001346   | AA9    | Copper-dependent lytic        | found                 |              |
|    |                       |        | polysaccharide monoxygenases  | found                 |              |
|    |                       |        | (LPMO)                        | found                 |              |
| 13 | FW16_GLEAN_10014384   | AA9    | Copper-dependent lytic        | found                 |              |
|    |                       |        | polysaccharide monoxygenases  | found                 |              |
|    |                       |        | (LPMO)                        | found                 |              |
| 14 | FW16_GLEAN_10007711   | AA9    | Copper-dependent lytic        | found                 |              |
|    |                       |        | polysaccharide monoxygenases  | found                 |              |
|    |                       |        | (LPMO)                        | found                 |              |
| 15 | FW16_GLEAN_10005570   | AA11   | Copper-dependent lytic        | found                 |              |
|    |                       |        | polysaccharide monoxygenases  | found                 |              |
|    |                       |        | (LPMO)                        | found                 |              |
| 16 | FW16_GLEAN_10004602   | AA9-CBM1| Copper-dependent lytic        | found                 |              |
|    |                       |        | polysaccharide monoxygenases  | found                 |              |
|    |                       |        | (LPMO)                        | found                 |              |
| 17 | FW16_GLEAN_10002187   | AA13-CBM20| Copper-dependent lytic       | found                 |              |
|    |                       |        | polysaccharide monoxygenases  | found                 |              |
|    |                       |        | (LPMO)                        | found                 |              |
|    |                       |        | modules binding xylan-        | found                 |              |
|    |                       |        | cellulose-binding function    | found                 |              |
|    |                       |        | demonstrated in one case.     | found                 |              |
| 18 | FW16_GLEAN_10015530   | CBM9   | Modules binding xylan-        | found                 |              |
|    |                       |        | cellulose-binding function    | found                 |              |
|    |                       |        | demonstrated in one case.     | found                 |              |
| 19 | FW16_GLEAN_10000334   | CBM13  | Multivalent sugar-binding sites, | found                 |              |
|    |                       |        | as demonstrated for their     | found                 |              |
|    |                       |        | interaction with xylan and    | found                 |              |
|    |                       |        | arabinosylxylan               | found                 |              |
| 20 | FW16_GLEAN_10007143   | CBM63  | Bind cellulose                | found                 |              |
| 21 | FW16_GLEAN_10004777   | CE1    | Xylan esterase                | found                 |              |
|    |                       |        | Xylan esterase                | found                 |              |
| 22 | FW16_GLEAN_10014832   | CE1    | Furfural esterase             | found                 |              |
|    |                       |        | found                        | found                 |              |
| 23 | FW16_GLEAN_10001089   | CE2    | Acetyl xylan esterase         | found                 |              |
|    |                       |        | found                        | found                 |              |
| 24 | FW16_GLEAN_10013305   | CE4    | Acetyl xylan esterase /       | found                 |              |
|    |                       |        | chitin deacetylase            | found                 |              |
| 25 | FW16_GLEAN_10006908   | CE5    | Cutinase                      | found                 |              |
| 26 | FW16_GLEAN_10007169   | CE5    | Acetyl xylan esterase         | found                 |              |
| 27 | FW16_GLEAN_10001547   | CE8    | Pectinesterase                | found                 |              |
| 28 | FW16_GLEAN_10001061   | CE8    | Pectinesterase                | found                 |              |
| 29 | FW16_GLEAN_10011996   | CE8    | Pectinesterase                | found                 |              |
|    |                       |        | found                        | found                 |              |
| 30 | FW16_GLEAN_10012229   | CE12   | Pectin acetylecsterase /      | found                 |              |
|    |                       |        | rhamnogalacturonan acetyl      | found                 |              |
|    |                       |        | xylan esterase                | found                 |              |
|    |                       |        | rhamnogalacturonan acetyl      | found                 |              |
|    |                       |        | xylan esterase                | found                 |              |
| 31 | FW16_GLEAN_10013316   | CE12   | Pectin acetylecsterase /      | found                 |              |
|    |                       |        | rhamnogalacturonan acetyl      | found                 |              |
|    |                       |        | xylan esterase                | found                 |              |
|    |                       |        | rhamnogalacturonan acetyl      | found                 |              |
|    |                       |        | xylan esterase                | found                 |              |
| 32 | FW16_GLEAN_10015498   | CE16   | Acetylecsterase               | found                 |              |
| 33 | FW16_GLEAN_10003717   | GH1    | B-Glucosidase                 | found                 |              |
| 34 | FW16_GLEAN_10006734   | GH1    | B-Glucosidase                 | found                 |              |
| 35 | FW16_GLEAN_10010580   | GH2    | B-Mannosidase                 | found                 |              |
| 36 | FW16_GLEAN_10000866   | GH2    | B-Galactosidase               | found                 |              |
| 37 | FW16_GLEAN_10003498   | GH3    | B-Glucosidase                 | found                 |              |
| 38 | FW16_GLEAN_10008834   | GH3    | B-Glucosidase                 | found                 |              |
| 39 | FW16_GLEAN_10011639   | GH3    | β-Glucosidase                 | found                 |              |
| 40 | FW16_GLEAN_10011383   | GH3    | β-Glucosidase                 | found                 |              |
| No. | Accession IDs | Family | Molecular function prediction / Possible substrate(s) | Carbon Source |
|-----|--------------|--------|------------------------------------------------------|--------------|
| 41  | FW16_GLEAN_10004843 | GH3 | β-Glucosidase | found found |
| 42  | FW16_GLEAN_10001962 | GH5 | Endo-β-(1,4)-glucanase / cellulase / other | found found |
| 43  | FW16_GLEAN_10004691 | GH5 | Glucan endo-β-(1,6)-glucosidase | found found |
| 44  | FW16_GLEAN_10015175 | GH5 | Endo-β-(1,6)-galactanase | found found |
| 45  | FW16_GLEAN_10015497 | GH5 | Mannan endo-β-(1,4)-mannosidase | found found |
| 46  | FW16_GLEAN_10007085 | GH7 | β-(1,4)-Glucanase | found found |
| 47  | FW16_GLEAN_10003305 | GH10 | Endo-β-(1,4)-xylanase | found found |
| 48  | FW16_GLEAN_10001573 | GH10 | Endo-β-(1,4)-xylanase | found found |
| 49  | FW16_GLEAN_10013894 | GH11 | Endo-β-(1,4)-xylanase | found found |
| 50  | FW16_GLEAN_10013895 | GH11 | Endo-β-(1,4)-xylanase | found found |
| 51  | FW16_GLEAN_10008963 | GH12 | Xyloglucan hydrolase | found found |
| 52  | FW16_GLEAN_10001440 | GH12 | α-Amylase | found found |
| 53  | FW16_GLEAN_10009774 | GH16 | Endo-β-(1,4)-galactosidase / chitin β-(1,6)-glucanase / licheninase | found found |
| 54  | FW16_GLEAN_10010201 | GH16 | Chitin β-(1,6)-glucanose transferase / β-glucosidase / β-(1,3)-glucosyltransferase / β-(1,3)-glucanase / laminarinase | found found |
| 55  | FW16_GLEAN_10003917 | GH17 | Endo-β-(1,4)-galactosidase / chitin β-(1,6)-glucanose transferase / β-glucosidase / β-(1,3)-glucosyltransferase / β-(1,3)-glucanase / laminarinase | found found |
| 56  | FW16_GLEAN_10009225 | GH17 | Chitin β-(1,6)-glucanose transferase / β-glucosidase / β-(1,3)-glucosyltransferase / β-(1,3)-glucanase / laminarinase | found found |
| 57  | FW16_GLEAN_1001346 | GH17 | Endo-β-(1,3)-galactanase | found found |
| 58  | FW16_GLEAN_10004522 | GH17 | Glucan endo-β-(1,3)-glucosidase / licheninase / Aβ-1 specific β-glucosidase / β-(1,3)-glucosyltransferase / β-(1,3)-glucanase | found found |
| 59  | FW16_GLEAN_10010890 | GH17 | β-(1,3)-Glucosyltransferase / β-(1,3)-glucosidase | found found |
| 60  | FW16_GLEAN_10012536 | GH17 | β-(1,3)-Glucosyltransferase / β-(1,3)-glucosidase | found found |
| 61  | FW16_GLEAN_10013337 | GH17 | β-(1,3)-Glucosyltransferase / β-(1,3)-glucosidase | found found |
| 62  | FW16_GLEAN_10014742 | GH17 | β-(1,3)-Glucosidase | found found |
| 63  | FW16_GLEAN_10011231 | GH18 | Chitinase | found found |
| 64  | FW16_GLEAN_10006294 | GH20 | β-Hexosaminidase | found found |
| 65  | FW16_GLEAN_10007529 | GH24 | Leesoyzme | found found |
| 66  | FW16_GLEAN_10000319 | GH28 | Exo-polygalacturonase / xylogalacturonan hydrolase | found found |
| 67  | FW16_GLEAN_10005091 | GH28 | Exo-polygalacturonase | found found |
| 68  | FW16_GLEAN_10000723 | GH28 | Exo-polygalacturonase | found found |
| 69  | FW16_GLEAN_10008243 | GH28 | Exo-polygalacturonase | found found |
| 70  | FW16_GLEAN_10014771 | GH31 | α-Glucosidase | found found |
| 71  | FW16_GLEAN_10011615 | GH32 | Fructan:fructan 1-fructosyltransferase | found found |
| 72  | FW16_GLEAN_1000618 | GH35 | β-Galactosidase / β-(1,3)-galactosidase | found found |
| 73  | FW16_GLEAN_10002565 | GH37 | α-L-Arabinofuranosidase | found found |
| 74  | FW16_GLEAN_10003104 | GH39 | β-Xylosidase / β-1,4-glucanase / β-xylosidase / α-L- arabinoarabinofuranosidase | found found |
| 75  | FW16_GLEAN_10001518 | GH43 | β-Xylosidase / α-L-arabinofuranosidase | found found |
| No. | Accession IDs       | Family | Molecular function prediction (Possible substrate(s)) | Carbon Source |
|-----|---------------------|--------|----------------------------------------------------------|---------------|
| 76  | FW16_GLEAN_10000029 | GH43   | α-(1,2)-l-Arabinofuranosidase / exo-α-(1,5)-l-| found        |
|     |                     |        | arabinofuranosidase / exo-α-(1,5)-l-arabinanase / β-(1,3)- |               |
|     |                     |        | xylosidase / β-(1,3)-xylosidase / exo-α-(1,5)-l-arabinanase / |               |
|     |                     |        | β-(1,5)-xylosidase / exo-α-(1,5)-l-arabinanase /β-(1,3)- |               |
| 77  | FW16_GLEAN_10003286 | GH43   | ENDO-α-(1,5)-l-arabinanase                                | found        |
| 78  | FW16_GLEAN_10006922 | GH43   | β-Xylosidase / α-l-arabinofuranosidase                   | found        |
| 79  | FW16_GLEAN_10007175 | GH43   | α-Galactofuranosidase                                   | found        |
| 80  | FW16_GLEAN_10009940 | GH43   | α-L-Arabinofuranosidase                                 | found        |
| 81  | FW16_GLEAN_10010955 | GH43   | β-Xylosidase                                            | found        |
| 82  | FW16_GLEAN_10001821 | GH43   | α-(1,2)-l-Arabinofuranosidase / exo-α-(1,5)-l- | found        |
|     |                     |        | arabinofuranosidase / exo-α-(1,5)-l-arabinanase / β-(1,3)- |               |
|     |                     |        | xylosidase / β-(1,3)-xylosidase / exo-α-(1,5)-l-arabinanase / |               |
|     |                     |        | α-L-arabinofuranosidase                                 | found        |
| 83  | FW16_GLEAN_10011918 | GH43   | α-(1,2)-l-Arabinofuranosidase / exo-α-(1,5)-l- | found        |
|     |                     |        | arabinofuranosidase / exo-α-(1,5)-l-arabinanase / β-(1,3)- |               |
|     |                     |        | xylosidase / β-(1,3)-xylosidase / exo-α-(1,5)-l-arabinanase / |               |
|     |                     |        | α-L-arabinofuranosidase                                 | found        |
| 84  | FW16_GLEAN_10012038 | GH43   | β-Xylosidase / α-l-arabinofuranosidase                   | found        |
| 85  | FW16_GLEAN_10012117 | GH49   | Dextranase                                               | found        |
| 86  | FW16_GLEAN_10003129 | GH51   | α-L-Arabinofuranosidase                                 | found        |
| 87  | FW16_GLEAN_10014327 | GH53   | ENDO-β-(1,4)-galactanase                                | found        |
| 88  | FW16_GLEAN_10002430 | GH54   | α-L-Arabinofuranosidase                                 | found        |
| 89  | FW16_GLEAN_10003512 | GH55   | EXO-β-(1,3)-glucanase                                   | found        |
| 90  | FW16_GLEAN_10000958 | GH71   | α-(1,3)-glucanase                                       | found        |
| 91  | FW16_GLEAN_10004186 | GH72   | β-(1,3)-Glucanoyltriglycosylase                          | found        |
| 92  | FW16_GLEAN_10013550 | GH72   | β-(1,3)-Glucanoyltriglycosylase                          | found        |
| 93  | FW16_GLEAN_10006031 | GH74   | Xyloglucanase                                           | found        |
| 94  | FW16_GLEAN_10007311 | GH75   | Chitosanase                                              | found        |
| 95  | FW16_GLEAN_10004573 | GH79   | β-Glucuronidase / β-4-O-methylglucuronidase             | found        |
| 96  | FW16_GLEAN_10011474 | GH81   | ENDO-β-(1,3)-glucanase                                  | found        |
| 97  | FW16_GLEAN_10001005 | GH93   | EXO-α-(1,5)-l-arabinanase                               | found        |
| 98  | FW16_GLEAN_10011917 | GH93   | EXO-α-(1,5)-l-arabinanase                               | found        |
| 99  | FW16_GLEAN_10012182 | GH99   | Glycosprotein endo-α-(1,2)-mannosidase                   | found        |
| 100 | FW16_GLEAN_10003277 | GH115  | Xylan α-(1,2)-glucuronidase                             | found        |
| 101 | FW16_GLEAN_10003558 | GH115  | Xylan α-(1,2)-glucuronidase                             | found        |
| 102 | FW16_GLEAN_10005637 | GH128  | β-(1,3)-Glucanase / β-(1,3)-glucosidase                 | found        |
| 103 | FW16_GLEAN_10006366 | GH132  | Glycosidases, potential activity on β-(1,3)-glucan      | found        |
| 104 | FW16_GLEAN_10010582 | GH134  | ENDO-β-(1,4)-mannanase                                  | found        |
| 105 | FW16_GLEAN_1000830  | GH146  | β-L-Arabinofuranosidase                                 | found        |
| 106 | FW16_GLEAN_10004046 | GH5-CBM1| ENDO-β-(1,4)-glucanase / cellulase                     | found        |
| 107 | FW16_GLEAN_10004089 | GH5-CBM1| Mannan endo-β-(1,4)-mannosidase                        | found        |
| 108 | FW16_GLEAN_10006035 | GH6-CBM1| Celllobiohydrolase                                     | found        |
| 109 | FW16_GLEAN_10001888 | GH7-CBM1| Reducing end-acting celllobiohydrolases                 | found        |
| 110 | FW16_GLEAN_10005918 | GH7-CBM1| Reducing end-acting celllobiohydrolases                 | found        |
| 111 | FW16_GLEAN_10010356 | GH11-CBM1| Endo-β-(1,4)-xylanase                                  | found        |
| 112 | FW16_GLEAN_10002366 | GH134-BM48| α-(1,4)-Glucan branching enzyme                      | found        |
| 113 | FW16_GLEAN_10004332 | GH15-CBM20| Glucoamylase                                         | found        |
| 114 | FW16_GLEAN_10007569 | GH16-CBM18| Chitin β-(1,3)-glucanoyltriglycosylase / β- | found        |
|     |                     |        | glucanoyltriglycosylase / β-glycosidase / β- |               |
|     |                     |        | carrageenase / endo-β-(1,3)-glucanase / laminarinase   |               |
| 115 | FW16_GLEAN_10008961 | GH18-3BM18| Chitinase                                              | found        |
| 116 | FW16_GLEAN_10002083 | GH18-CBM18-CBM50| Chitinase                               | found        |
| 117 | FW16_GLEAN_10008188 | GH18-CBM18-CBM50| Chitinase                               | found        |
Table 2 (continued)

| No. | Accession IDs | Family | Molecular function prediction/Possible substrate(s) | Carbon Source |
|-----|---------------|--------|----------------------------------------------------|---------------|
| T18 | FW16_GLEAN_10001593 | GH27-CBM35 | α-galactosidase / β-L-arabinofuranosidase          | Found         |
| T19 | FW16_GLEAN_1000952  | GH43-CBM35 | β-Xylosidase (subfamily 24)                         | Found         |
| T20 | FW16_GLEAN_10007861 | GH45-CBM1 | Endoglucanase                                      | Found         |
| T21 | FW16_GLEAN_10014717 | GH71-3.M24-CBM24 | α-(1,3)-Glucanase                              | Found         |
| T22 | FW16_GLEAN_10012817 | GH72-CBM43 | β-(1,3)-Glucanoyltranseglycosylase                 | Found         |
| T23 | FW16_GLEAN_10004549 | GT20  | Trehalose 6-phosphatase                            | Found         |
| T24 | FW16_GLEAN_10001028 | PL1  | Pectate lyase                                      | Found         |
| T25 | FW16_GLEAN_10001932 | PL1  | Pectate lyase                                      | Found         |
| T26 | FW16_GLEAN_10012037 | PL2  | Pectate lyase                                      | Found         |
| T27 | FW16_GLEAN_10001411 | PL3  | Pectate lyase                                      | Found         |
| T28 | FW16_GLEAN_10001949 | PL3  | Pectate lyase                                      | Found         |
| T29 | FW16_GLEAN_10008463 | PL3  | Pectate lyase                                      | Found         |
| T30 | FW16_GLEAN_10012035 | PL3  | Pectate lyase                                      | Found         |
| T31 | FW16_GLEAN_10008083 | PL4  | Pectate lyase / exopolysaccharide lyase            | Found         |
| T32 | FW16_GLEAN_10015384 | PL4  | Pectate lyase / exopolysaccharide lyase            | Found         |
| T33 | FW16_GLEAN_10001628 | PL9  | Pectate lyase / exopolysaccharide lyase            | Found         |
| T34 | FW16_GLEAN_1000207 | PL20 | Endo-β-(1,4)-glucuronan lyase                     | Found         |
| T35 | FW16_GLEAN_10013878 | PL3-CBM1 | Pectate lyase                                    | Found         |

Fig. 6 Mapping of 135 CAZymes found by mass spectrometric proteomics in different F. metavorans secretomes. *F. metavorans* FW16.1 was grown on eight different substrates (y-axis) differing in complexity. The CAZymes identified by MS were mapped back to protein coding regions (CDS) in our de novo genome assembly, which consists of nine scaffolds. A circle indicates the genomic location of a CDS with a positive proteomics mapping. The circle color corresponds to the growth substrate. The CDS is annotated with its CAZyme family name on the x-axis. For example, the same CDS on scaffold 2, annotated as GH16, is expressed/secreted in all growth substrates except sugarcane bagasse, and is located at position 389,635–390,517 bp (exact position not shown for clarity)
and FW16_GLEAN_10015496 (CE16, acetyl esterase). We identified only one GT protein (FW16_GLEAN_10004549, GT20) and this was found on the L-CMC substrate. Finally, we identified proteins representing three CMB families present solely on the synthetic and artificial cellulose substrates: FW16_GLEAN_10015530 (CBM9), FW16_GLEAN_10000334 (CBM13) and FW16_GLEAN_10007143 (CBM63). Interestingly, CBM9 and CBM63 are predicted to bind cellulose but CBM13 is not. Another eight CBM families were represented in the modular proteins described above, combined with GH, AA or PL domains, and these were distributed similarly between the synthetic cellulose and biomass substrates. CBM1 was the most abundant module (nine identified in total), and was associated with GH, AA and PL proteins, whereas the other CBMs were found only 1–3 times each.

Conversion of biomass with the in-house F. metavorans cocktail
The overall enzymatic activity of the crude secretome preparations was low. We therefore lyophilized the enzymes secreted on both biomass substrates, resuspended them in 50 mM citrate buffer (pH 4.8), and combined them at a 1:1 ratio with a final protein concentration of 312 ± 2.7 µg/mL. We then prepared saturation curves (Additional file 1: Table S5). Hydrolysis assays were evaluated against three different substrates: steam-exploded sugarcane bagasse (XSCB), untreated (in nature) sugarcane bagasse (NSCB) and untreated maize leaves (MZ), each present at a concentration of 5% (w/v) for 24 h. Control assays without in-house enzymes (A1) were also prepared. All assays were supplemented with the commercial Accellerase 1500 enzyme mixture containing exoglucanase, endoglucanase, hemi-cellulase and β-glucosidase at a concentration of 5 FPU/mL (filter paper unit). Under control conditions (A1), XSCB was converted to glucose 1.6-fold more efficiently than the other substrates (Fig. 7). To test the activity of the secretome preparation, we supplemented the assay with the F. metavorans in-house cocktail at concentrations ranging from 10% (v/v) in assay A2 to 70% (v/v) in assay A6 (Additional file 1: Table S5).

Figure 7 shows the glucose profile following biomass hydrolysis in all assays (A1–A6). XSCB was easily converted to glucose by the commercial Accellerase 1500 enzyme mix, but the in-house cocktail did not facilitate further saccharification. In contrast, the in-house cocktail enhanced the release of sugars from the NSCB and MZ substrates starting at concentrations of 25% (v/v), corresponding to 196 µg/mL. When the concentration of the in-house cocktail reached 55% (v/v), corresponding to 0.289 µg/mL, the efficiency of saccharification became equivalent to that of the pre-treated (XSCB) substrate. An in-house enzyme cocktail with a protein load of 35–36 mg/g biomass therefore facilitated synergistic depolymerization without pre-treatment, achieving a statistically significant improvement in glucose yields (p < 0.05, 95% confidence).

Discussion
We set out to characterize an active fungal isolate by identifying enzymes that facilitate the utilization of plant biomass, particularly those involved in cellulose
degradation. We compared the enzymes induced by different synthetic cellulose substrates, and analyzed secretome components on two different types of agro-residual biomass representing the C4 crops sugarcane and maize [25, 44]. We also assigned the fungal isolate to the correct FSSC linkage. To the best of our knowledge, this is the first comparative analysis of the F. metavorans as a strain of the FSSC secretome on different substrates.

Analysis of the 62 proteins produced on all six artificial cellulose substrates revealed only 16 CAZymes, five of which were predicted to degrade cellulose. The enzymes were assigned to CAZy families GH5, GH7 and AA9. The corresponding genes were distributed over four different scaffolds, but there was no clear evidence for clusters of colocalized or coregulated genes. The hydrolytic degradation of cellulose by fungi involves at least three steps: (1) internal cellulose bonds are cleaved by endo-β-(1,4)-glucanases (GH5) [45–47] to create shorter polymers; (2) these are digested by exo-β-(1,4)-glucanases and/or cellobiohydrolases (GH7 and GH6) ultimately to produce cellobiose, which is (3) finally converted into two glucose molecules by β-glucosidases (mainly GH1 or GH3, and some others such as GH39) [48, 49]. At least the first two steps were recapitulated in the F. metavorans FW16.1 secretome fractions. For the first step, one predicted GH5 protein with cellulase activity (FW16_GLEAN_10000416) was found on all cellulose substrates, whereas another (FW16_GLEAN_10001962) was found on the biomass substrates. For the second step, one GH7 with a CBM1 domain (FW16_GLEAN_10005918, predicted cellobiohydrolase) was found on all substrates, another (FW16_GLEAN_10001888) was found on the artificial cellulose substrates, and a third without a CBM (FW16_GLEAN_10007085) was found on six of the eight substrates. Some proteins with predicted β-glucosidase activity (GH1, GH3 maybe GH39) were also found, but none of them were present on all substrates.

We also identified an AA9 lytic polysaccharide monooxygenase (LPMO) that can oxidize the C-1 or C-4 (and perhaps C-6) positions of the glycosidic bond in cellulose and disrupt its structure, as shown for the fungi Podospora anserina and Neurospora crassa [50, 51]. An interesting combination of AA9 and PL20 was observed, where glycosidic bonds of glucuronic acid-containing cello-oligosaccharides produced by AA9 proteins may be cleaved at the C4-position by the PL20 family via β-elimination to produce a reducing end [52].

This mechanism could also be involved in cellulose degradation, as already postulated for the fungus Humicola insolens [53]. A clear difference in cellulose degradation was identified between the biomass substrates, with more GHs found on maize leaves contrasting with more AAs catalyzing oxidative cellulose degradation on sugarcane bagasse, the latter indicating a more complex cellulose architecture [54]. The GH74 family, with predicted xyloglucanase activity, was also found on all substrates, and may therefore contribute to cellulose degradation. This is supported by the identification of a GH74 xylglucanase from the bacterium Cellvibrio japonicus with a strong preference for xylglucans but some activity (24–165-fold lower) against artificial substrates such as CMC and HEC [55]. Another protein (AA13) fused to the starch-binding module CBM20 [56] was found on four of the six synthetic cellulose substrates, perhaps indicating promiscuous activity against artificial cellulosic substrates. However, no cellulase activity was previously reported for AA13 enzymes isolated from the fungi Neurospora crassa and Aspergillus nidulans [57, 58].

The distribution of the CAZymes on the two biomass substrates was more complex, mirroring the complexity of the substrates, including the presence of hemicellulose, pectin and lignin. The secretome fractions thus included a lignocellulolytic enzyme cocktail with the ability to degrade all cell wall polymers and stored starch granules, including cleavage by lyses and oxidation.

The degradation of hemicellulose requires enzymes specific for β-(1,4)-linked xyloses or xyloglucan and arabinoxylan acetylated at the C2 and/or C3 positions as well as β-(1,3), β-(1,4) and β-(1,6) glucan branches [59] that connect pectin to cellulose [17, 60]. These include endo-β-(1,4)-xylanases (GH10, GH11), α-l-arabinofuranosidases and exo-α-l-(1,5)-arabinoxylanases (GH3, GH43, GH51, GH54, GH62 and GH93), β-xylosidases (GH43 and GH3), acetylxylan esterases (CE1–CE7), and ferulic acid esterases (CE1) and acetyl-esterases (CE16) [48]. We found three GH10 and GH11 proteins on maize and four on sugarcane. We also found GH27-CBM35, GH31 and GH35 proteins (the latter two restricted to maize) and GH93 proteins and one GH115 protein, all probably responsible for hemicellulose or rhamnogalacturonan I (pectin) degradation [61]. The GH43 family, which converts xylo-oligosaccharides containing arabinose and galactose to monomers, was also found on both substrates. We identified 11 GH43 enzymes in total, three exclusively on sugarcane and two on maize. Previous secretome analysis of Trichoderma reesei and Aspergillus niger on sugarcane bagasse [62], A. nidulans on sorghum [60], Myceliophthora thermophila on sugarcane bagasse [63] and N. haematococca on maize bran [64] identified one, five, eight, four and four GH43 proteins, respectively. We also identified multiple xylan esterases (CE1, CE4 and CE5): three on maize and one on sugarcane, similar to the T. reesei secretome on sugarcane bagasse which features two CE5 proteins [62]. In contrast, no CEs were found in the secretome fractions of N. haematococca on maize bran [64].
A combination of GHs, CEs and PLs was needed to break down pectin in our biomass substrates [19]. The GHs we identified represented families GH28 (four in total, one only found on sugarcane), GH43 and GH79 [19], perhaps also including GH35, GH51 and GH93 (which can digest rhamnogalacturonan I) [65]. We identified three CE8 proteins (two found only on sugarcane) and two CE12 proteins (required to remove branches from non-sugar components containing methyl and acetyl groups). Finally, we identified six PLs from families PL1, PL3 and PL9 on sugarcane, and 10 PLs from families PL1, PL3, PL4 and PL9 on maize. These are necessary for the efficient utilization of homogalacturonan and rhamnogalacturonan. In contrast, no pectin-digesting GHs, CEs and PLs were identified in the secretome of *N. haematococca* on maize bran, whereas the *A. niger* BRFM442 secretome contained six GH28, two CE8 and one PL proteins on the same substrate [64].

The AA superfamily of lignolytic enzymes and monooxygenases [66] was also found in the secretome induced on our maize and sugarcane substrates. Laccases (AA1) oxidize a wide range of aromatic compounds including polyphenols, methoxy-substituted monophenols and aromatic amines [67] and these were found on both substrates. When other *F. solani* strains were cultured on substrates such as oak combined with millet and wheat bran or corn, wheat, rye and oat, the secretome fractions contained laccases as well as manganese-dependent peroxidases (MnP) and lignin peroxidases (LiP), both of which represent family AA2 [68, 69]. We did not find any AA2 proteins, perhaps because we investigated only a limited set of time points, thus providing an incomplete picture of oxidative lignin degradation. However, we identified AA3 flavoenzymes on both substrates, and this family includes glucose oxidases and aryl alcohol oxidases that act on the anomic carbon of β-d-glucose and alcohols using molecular oxygen as an electron acceptor, releasing hydrogen peroxide [70]. It is interesting to note that feruloyl and p-coumaroyl esterases were not found on the maize substrate, whereas one CE1 protein with that predicted function was found on sugarcane and all the cellulose substrates. These esterases normally remove the crosslinks between polysaccharides and lignin to increase enzymatic access to the cell wall [62, 63]. The analysis of an *A. nidulans* strain on sorghum stover revealed only two esterases in the secretome [54].

Several of the enzymes discussed above overcome the inaccessibility of insoluble substrates by using one or more non-catalytic CBMs [71]. Examples include the GH5, GH7, GH11, GH45 and PL3 families, which are frequently associated with CBM1 (which typically binds cellulose) [72]. Three of the five GH18 family members we identified were associated with the chitin-binding modules CBM18 and CBM50. Similarly, the *T. reesei* genome encodes at least 18 GH18 proteins, four with additional CBMs [73]. A glucoamylase (GH15) associated with the starch-binding module CBM20 was found in *Penicillium echinulatum* [74], and we identified α-(1,4)-glucan branching enzymes (GH13) associated with the glycogen-binding module CBM48, which has been found in several other species [75]. We also identified an α-(1,3)-glucanase (GH71) associated with the starch-binding module CBM24, and an α-galactosidase (GH27) associated with CBM35, which was shown to bind β-galactans in *Phanerochaete chrysosporium* [66].

Our comparative approach revealed 500 secretome proteins, including 93 GH proteins representing 40 different families. A similar range was reported *F. solani* ATCC MYA 4552 cultivated on a mixture of oak, millet and wheat, where 398 proteins were identified, including 48 GH proteins representing 28 families [69]. We compared the secretome proteins of our *F. metavorans* FW16.1 isolate on natural substrates with nine other fungal secretome fractions [60, 62, 64, 76–78]. In most cases, our isolate produced a larger number of secreted CAZymes, with only *A. nidulans* strain A78 grown on sorghum stover and *A. niger* BRFM442 grown on maize bran producing more (Table 3). The cultivation of *N. haematococca* on maize bran produced four GH43 proteins but no members of the families GH5, GH6, GH7 or AA9, arguing that maize bran induces the secretion of hemicellulases [64]. We found that *Fusarium* sp. of the FSSC uses their diverse arsenal of depolymerizing and accessory enzymes as destruents to break down complex substrates, supported by their adaptation to different environments, their metabolic plasticity, and their ability to degrade different lignocellulosic materials [69, 79], as well as other compounds such as the pesticide dichlorodiphenyltrichloroethane [80].

The *F. metavorans* in-house enzymatic cocktail proved a suitable alternative to the chemical pre-treatment of agro-residual lignocellulosic biomass, clearly allowing the debranching of polymers surrounding the cellulose fibers and releasing reducing sugars (Fig. 7). Pre-treatment methods are often needed for recalcitrant biomass such as hemicellulose, lignin and crystalline cellulose, to open up the fibers and improve accessibility to the polymers [44, 81]. Accordingly, the *F. metavorans* in-house cocktail did not enhance the production of sugars from sugarcane biomass subjected to steam explosion, because pre-treatment had already rendered the polymers fully accessible to the Accellerase 15,000 cocktail. However the in-house cocktail had a strong impact on the saccharification of untreated maize and sugarcane biomass, with additional advantages over chemical pre-treatment such as selectivity, mass efficiency (the released carbohydrates...
are retained and utilized), and the avoidance of inhibitory by-products. Furthermore, no toxic compounds are dispersed into the environment, avoiding the need to recycle or remove them. The *F. metavorans* enzyme cocktail therefore provides a sustainable, low-energy process to enhance the efficiency of enzymatic saccharification [82–84].

**Conclusion**

The CAZymes identified in this study can be used to enhance the enzymatic saccharification of agro-residual biomass. Our workflow involved strain isolation, genome sequencing, CAZyme analysis and secretome analysis by mass spectrometric proteomics, revealing 135 relevant enzymes. The *F. metavorans* in-house cocktail was used to increase the amount of glucose generated from maize leaves and untreated sugarcane bagasse by selective pretreatment, improving the turnover of the hemicellulose fraction without carbohydrate loss or the formation of inhibitory by-products.

**Materials and methods**

**Fungus isolation and growth conditions**

The fungal isolate *F. metavorans* FW16.1, was obtained from mangrove wood [43] in Vietnam (longitude 10°36′015″N, latitude 106°56′045″E) and prepared as a conidial suspension. Mycelium pieces (5 mm diameter) on potato dextrose agar (PDA) were transferred to a fresh PDA plate and grown in the dark for 5–7 days at 28 °C. The conidia were scraped with a Drigalski cell spreader in sterile water and centrifuged at 2693 × g for 15 min at 4 °C. The pellet was washed in sterile water, filtered through a 40-µm mesh sieve and centrifuged as above. The pellet was resuspended in sterile water, aliquoted and stored at –70 °C. To investigate mycelial growth and color formation, fungal growth was assessed on PDA, YPD [85], complete medium (CM) [86], malt extract agar (MEA) [87], starch casein agar (SCA) [88] and Mandels’ salt medium (MS) [89] for 15 days (Fig. 2).

**Phylogenetic analysis and de novo sequencing**

Submerged cultures of *F. metavorans* FW16.1 were established in potato dextrose broth (PDB) and incubated at 28 °C, shaking at 150 rpm. DNA was isolated according to the CTAB method [90, 91] and purity and quality were confirmed by gel electrophoresis and spectrophotometry. We used 11 µg of pure high-molecular-weight genomic DNA (gDNA) for the de novo preparation of 270-bp short HiSeq and PACBIO RSII 20 K sequencing libraries. Following gene prediction, ORFs were identified and annotated according to Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Clusters of Orthologous Groups (COGs) using BGI (Beijing Genomics Institute, China) to create a...
fungus-specific database (Additional file 2: FW16.IntegrationTable.lxs). The genome sequence of *F. metavorans* FW16.1 was deposited in GenBank (bioproject number PRJNA413482, biosample number SAMN07749916) with the accession number JADNBR00000000.

The ITS-1/8S rRNA/ITS-2 region was amplified and sequenced using primers ITS1_fw (5'-TCC GTA GGT GAA CCT GCG G-3') and ITS4_rv (5'-TCC TCC GCT TAT TGA TAT GC-3') [92] and the ITS sequence was deposited in GenBank (accession number MG098676). Multiple sequence alignments for marker genes *TEF1*, *RPB2* and ITS + 28S for 79 *Fusarium* taxa were kindly provided by Kerry O'Donnell (personal communication). We built three independent covariance models using cmbuild v1.1.3 in the Infernal package (https://doi.org/10.1093/bioinformatics/btt509) from the sequence alignments without consensus structure information (parameter -noss). The bit scores depend on multiple sequence alignment length (more precisely, the covariance model length), so we ran the ungapped alignment sequences against their covariance model (cmalign -noss -g) and obtained 639, 1668 and 981 bits as average scores for *TEF1*, *RPB2* and ITS + 28S, respectively. Given that a covariance model without a consensus structure is basically a hidden Markov model (HMM), we initially used hmmbuild and hmmsearch (www.hmmer.org) instead, but this did not yield hits with sufficient scores, most likely due to high penalties for the insertion of introns.

Using the covariance model for *TEF1*, we found a hit in scaffold2 at position 6,427,210–6,427,837 with 643 bits (slightly above average). The model for *RPB2* returned two partial hits in close proximity on the reverse strand of scaffold 3. Manual inspection revealed overlapping full models for those hits, but a 130-bp region (probably an intron) divided the region in half. Enforcing global alignment of the combined region 2,964,591–2,966,345 (cmalign -mxsize 100,000 -noss) resulted in a score of 1671 bits, which was above the expected average score.

The covariance model for the ITS + 28S region did not return significant hits, probably due to the omission of this region in the assembly, reflecting multiple gene copies and repetitive regions that complicated the coverage information [93]. We therefore used the covariance model to identify 5737 of the raw 465,771 PacBio reads with significant scores. We next used proovread v2.14.1 (https://doi.org/10.1093/bioinformatics/btu392) to polish the frequent insertion or deletion of bases (indels) in the PacBio reads with short Illumina reads and mapped the results against the FW16.1 scaffold using bowtie2 v2.3.4.1 (https://doi.org/10.1038/nmeth.1923) with the settings-p 20-very-sensitive-local-f, yielding a clear pile up on scaffold 1 from 2,806,169 to 2,812,396. The local alignment of 2245 polished reads against the model (cmalign -mxsize 100,000 -noss) resulted in 456 high-scoring identical alignment rows of 1227 bp. Finally, we added the three identified marker gene regions to the initial sequence alignment (Additional file 3: 79-FSSC-3-locus.nex) and used IQTree v1.6.12 (https://doi.org/10.1093/molbev/msaa015) with the settings -nt AUTO -bb 5000, and partitions *TEF1* = 1–665, ITS + 28S = 666–1621 and *RPB2* = 162–3209 to construct the phylogenetic tree with partitioned maximum likelihood bootstrapping. The resulting newick tree file (Additional file 4, tree_fw16+79.figtree) was rooted at NRRL 22,090 *F. iludens* and NRRL 22,632 *F. plagiophanii* and colored using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

### CAZyme analysis

All genomic regions marked as CDS in our de novo assembly were screened for homologs to families and subfamilies in the CAZyme database [66] using a combination of RAPSearch2 [94, 95] and hmmsearch from the HMMER package [96] as previously described [97]. The CAZyme families/subfamilies were represented by sequence members with different enzymatic activities, annotated as different EC numbers, thus a single homolog CDS can yield multiple EC annotations. To reduce EC number ambiguity, we used BLASTP (v2.9.0+) to score the CDS identified by LC–MS/MS against all sequences of the homologous CAZyme family obtained from dbCAN2 (http://csbl.bmb.uga.edu/dbCAN/index.php) [98]. The CDS was only annotated with EC numbers of the top BLASTP hits for each protein. The corresponding descriptors of EC numbers were used as possible functional annotations (Additional file 1: Table S4). CAZymes identified by LC–MS/MS were mapped to the genome.

### Secretome analysis and SDS-PAGE

The *F. metavorans* secretomes were induced by fermentation in 100-mL Erlenmeyer flasks at 28 °C for up to 96 h, shaking at 150 rpm. Each liquid fermentation was carried out in duplicate (agro-residual biomass) or triplicate (synthetic substrates). Mycelia were pre-cultivated inYPD medium at 28 °C for 3 days, shaking at 150 rpm, then washed briefly and dried between sheets of filter paper (Whatman, Dassel, Germany). We then incubated 0.1 g of the semi-dried mycelia with 50 mL inductive medium at 28 °C for 72 h, shaking at 150 rpm. The inductive medium comprised mineral salts (0.35% NaNO₂, 0.15% K₂HPO₄, 0.05% MgSO₄ × 7H₂O, 0.05% KCl, 0.001% FeSO₄ × 7H₂O) supplemented with 1% (w/v) synthetic or artificial cellulose (Avicel, α-cellulose, HEC, H-CMC, M-CMC or L-CMC, all from Sigma-Aldrich, Steinheim, Germany). The agro-residual biomass was prepared at a final concentration of 1% in Mandels and Weber medium.
[99], with additional yeast extract and peptone (0.03%). The sugarcane bagasse was milled to 1 mm and the maize leaves to 1.5 mm as untreated substrates. After 96 h, the fungal biomass was removed by centrifugation (3250 × g for 30 min) and the supernatant was harvested for secretome analysis, followed by lyophilization and resuspension in 50 mM citrate buffer (pH 4.5). The secretome samples were separated by SDS-PAGE on 12% polyacrylamide gels [100]. The gels were stained with 0.1% Coomassie Brilliant Blue R250 and destained with 45% methanol and 10% acetic acid. The gels were set aside for analysis by mass spectrometric proteomics and remaining samples were retained for enzymatic assays.

**Proteomics**

**Sample preparation**

In-gel tryptic digestion [101] was carried out by dividing each gel lane into 4–5 equal parts and dicing them, followed by reduction with 10 mM dithiothreitol in 100 mM ammonium bicarbonate, alkylation with 55 mM iodoacetamide in 100 mM ammonium bicarbonate and digestion with 13 ng/µL trypsin in 10 mM ammonium bicarbonate containing 10% (v/v) acetonitrile (Promega, Mannheim, Germany). Tryptic peptides were extracted with a 1:1 mixture of 5% formic acid and acetonitrile and were completely lyophilized. The peptides were resuspended in 40 µL 0.1% formic acid before LC–MS/MS analysis.

**LC–MS/MS analysis of the tryptic peptides**

We injected 2-µL samples onto an Acclaim PepMap C-18 nanoViper trapping column (Thermo Fisher Scientific, Waltham, MA, USA; 100 µm × 20 mm, 5 µm, 100 Å) at a flow rate of 3 µL/min and washed for 5 min with 2% buffer B (0.1% formic acid in acetonitrile). The peptides were separated on an Acclaim PepMap C-18 nanoViper reversed-phase capillary column (Thermo Fisher Scientific; 75 µm × 50 cm, 2 µm, 100 Å) at 45 °C using a Dionex Ultimate 3000 nano-UPLC system (Thermo Fisher Scientific) connected to a Fusion trilbd (quadrupole/Orbitrap/linear ion-trap) mass spectrometer (Thermo Fisher Scientific). The gradient system consisted of buffer A (0.1% formic acid in MS-grade water) and buffer B at a constant flow rate of 300 nL/min for 70 min. The profile was held at 3% B for 5 min followed by a gradient to 28% B, at 35 min, then 35% B at 40 min, and 90% B at 40 min 6 s. After a hold at 90% B for 9 min 54 s, the column was equilibrated at 3% B for 19 min 54 s. Eluted peptides were ionized in positive ion mode using a nanospray Flex with an electrospray ionization source (Thermo Fisher Scientific) and a fused-silica nano-bore emitter with an internal diameter of 10 µm (New Objective, Woburn, MA, USA) at a capillary voltage of 1800 V. The ion transfer tube temperature was set to 300 °C. Parent ion scans were carried out in the range 400–1300 m/z in the Orbitrap mass analyzer at 120 K resolution with a maximum injection time of 120 ms and an AGC target value of 2 × 10^6. Data-dependent acquisition mode was set to top speed mode for precursor ion selection. The most intense peaks with (intensity threshold of 5 × 10^5) were isolated with a quadrupole isolation width of 1.6 m/z, fragmented by high-energy collisional dissociation (collision energy 30%) and detected in the ion-trap mass analyzer. A dynamic exclusion filter was applied for 30 s and excluded after one time. For ion-trap detection, the scan rate was set to a rapid scan range 400–1300 m/z. The maximum injection time was 60 ms, and the AGC target value was 1 × 10^4.

**Protein identification by database matching**

The LC–MS/MS data files were used to search the translated database of *F. metavorans* DSM105788 sequences (Additional file 2: FW16.IntegrationTable.lxs) with Proteome Discoverer v2.0 (Thermo Fisher Scientific) including the search engine Sequest HT. The search parameters included precursor and product ion mass tolerances of 10 ppm and 0.5 Da, respectively, two missed cleavages allowed, cysteine carbamidomethylation as a fixed modification, and methionine oxidation as a variable modification. Proteins found with at least one unique peptide and a false discovery rate (FDR) of 1% (determined by percolator) were accepted [101].

**Enzymatic activity**

Enzymatic hydrolysis was measured using the DNS method [102] after liquid fermentation at 50 °C for 24 or 96 h with the substrates arabamin, arabinoxylan, galactan, xylan, starch, CMC and polygalacturonic acid (all at 0.5%) or pectin citrus and laminarin (at 0.2%). We mixed 10 µL of the *F. metavorans* extract with 50 µL of each substrate and 40 µL 50 mM citrate buffer (pH 4.8). Xylan was assayed for 10 min and the remaining substrates for 3 h. *F. metavorans* was grown in YPD medium, we also measured CMCase activity against CMC every 24 h for up to 7 days. Furthermore, if the fungus was cultivated in Mandels’ mineral salts medium supplemented with 1% (w/v) cellulose and artificial cellulose substrates Avicel PH-101, α-cellulose, HEC, H-CMC, M-CMC or L-CMC, we also measured the CMCase activity on day 3. The protein concentration was determined using the ROTI Nanoquant protein detection kit (Carl Roth, Karlsruhe, Germany) by adding 50 µL of the supernatant to 200 µL of the detection solution. Measurements were collected from at last three experimental replicates.
Saccharification of sugarcane bagasse and maize leaves

The conversion of 5% (w/v) NSCB, XSCB [44] and MZ into glucose, was tested in saturation curve assays supplemented with increasing amounts of the *F. metarvorum* in-house crude enzymatic cocktail to a fixed amount of Accellerase 1500 (Genecor, Rochester, NY, USA) at final total cellulase activity of 5 FPU/g biomass, corresponding to 118 µg/mL. For the in-house enzymatic cocktail, the lyophilized secretome fractions from both biomass substrates were resuspended in 50 mM citrate buffer (pH 4.8) and combined at a 1:1 ratio (NSCB:MZ) before saturation curve experiments, such that the final protein concentration of 312 ± 2.7 µg/mL represented 100%. Saccharification was carried out in 2-mL Eppendorf tubes containing 50 mM citrate buffer (pH 4.5) and up to 70% (v/v) of the in-house enzymatic cocktail from *F. metarvorum* at 50 °C for 24 h in a thermomixer (Eppendorf, Hamburg, Germany) at an agitation rate of 1000 rpm. The amount of protein applied for the saturation curve experiments can be found in Additional file 1: Table S5. Each experiment was replicated and the reducing sugars were measured in triplicate using the DNS assay [102]. Glucose standards were used to calibrate the glucose released under each condition. The statistical significance (threshold p < 0.05) was determined using Perseus (www.ezbioinformatics.org/doi.php).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13068-021-01927-9.

### Additional file 1: Figure S1

Genomic DNA from *Fusarium metarvorum* FW16.1 (DSM105788) was isolated using the CTAB method and 5 µL was mixed with 6 x loading buffer (0.25% (v/v) xylene cyanol, 0.25% (v/v) bromophenol blue, 30% (v/v) glycerol and separated by 0.8% (w/v) agarose gel electrophoresis in Tris–borate EDTA (TBE) buffer at 80 V for 60 min, with the GeneRuler 1 kb Plus DNA Ladder (Thermo Fisher Scientific) as a marker. The DNA was stained with 1% ethidium bromide for 15 min and observed on a uv transilluminator (SynGene Genius, BioImaging System).

### Figure S2

Specific CMCase activity of the supernatants against high-viscosity CMC over time in YPD medium. *Figure S3* Specific CMCase activity of the supernatants using different synthetic nutrient sources.

### Figure S4

Enzymatic activities for polygalacturonase (A), laminarinase (B), CMCase (C) and xylanase (D). *Table S1* CMCase activity of 48 fungal strains.

### Table S2

CAZyme analysis of fungal isolate FW16.1 and other fungal species. The coding regions were compared with the CAZyme database (Cantarel et al. 2009; Lombard et al. 2014). *Table S3* Proteins of the fungal isolate FW16.1 induced on different synthetic and artificial cellulose and biomass substrates (maize leaves (MZ) or sugar cane bagasse (SCB)). The proteins were separated by SDS-PAGE followed by in-gel tryptic digestion and LC–MS/MS. The accession number, description, coverage (%), number of peptides (# peptides), peptide-to-spectrum matches (# PSMs), molecular weight in KDa (MW [KDa]), the calculated isoelectric point (calc. pI), Score Sequest HT and number of Peptides Sequest HT (# Peptides Sequest HT) were compared with the automated translation of the genome of the fungal isolate *Fusarium metarvorum* DSM105788. BLASTP annotations and cellular functions are also shown. *Table S4* Functional prediction of the CAZymes found on different synthetic and artificial cellulose and biomass substrates. *Table S5* Quantity of protein applied in the saturation curve experiments.

Acknowledgements

We thank Dr. Richard L. Hahnke, Dr. Tobias Stacke, Kerry O'Donnell and David M. Geiser for valuable discussion and support. We thank Frank Förster, who helped assessing assembly quality, construction of trees and "polishing" of PaclBio reads. We thank Richard M. Twyman for proofreading this manuscript.

Authors’ contributions

SCB isolated the fungal DNA and carried out ITS-PCR, secretome analysis and SDS-PAGE. HB cultivated the fungi and conducted the BLAST analysis of the proteins. BE measured the activities of the supernatants. SJ and KM analyzed the ITS-28S, TEF1 and RPB2 and genomic data and generated the phylogenetic tree. AA and HS carried out mass spectrometric proteomics. MR analyzed the FW16.1 genome. CB, CW and WS supervised the project. MG organized the laboratory work, coordinated the participating institutions and evaluated the data. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This work was funded by the German federal ministry BMFB under contract no. 03A0150B within the framework of Bioeconomy International 2014. Dr. Arslan Ali acknowledges a fellowship under the faculty development program of the International Center for the Chemical and Biological Sciences, University of Karachi. Dr. Héliva Brognaro acknowledges financial support from the Fundação Faculdade de Medicina (FFM, São Paulo/Brazil).

Availability of data and materials

All data generated or analyzed during this study are either included in this published article or can be found in the Supplementary Material.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

### Author details

1 Faculty of Mathematics, Computer Science and Natural Science, Department of Biology, Biozentrum Klein Flottbek, Molecular Phytopathology, University of Hamburg, Ohnhorststr. 18, 22609 Hamburg, Germany. 2 Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, Av. Prof. Lineu Prestes, 1374, São Paulo, CEP 05508-000, Brazil. 3 Institute of Biochemistry and Molecular Biology, University of Hamburg, Martin Luther King Platz 6, 20146 Hamburg, Germany. 4 Dr. Panyiyan Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, University Road, Karachi 75270, Pakistan. 5 Institute of Clinical Chemistry and Laboratory Medicine Diagnostic Center, Campus Research. Martinstr. 52, N27, 20246 Hamburg, Germany. 6 Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Department Screening-Port, Schnackenburgallee 114, 22525 Hamburg, Germany. 7 Department Biology and Chemistry, Algorithmic Bioinformatics, Justus Liebig University Giessen, Heinrich-Buff-Ring 58, 35392 Giessen, Germany. 8 Department Biology and Chemistry, Institute of Food Chemistry and Food Biotechnology, Justus Liebig University Giessen, Heinrich-Buff-Ring 17, 35392 Giessen, Germany.

Received: 17 November 2020 Accepted: 11 March 2021

Published online: 20 March 2021
References

1. Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick-WJ, Halliott JP, Leak DJ, Liotta CL, et al. The path forward for biofuels and biomaterials. Science. 2006;311(5764):489–494.

2. Serrano-Ruiz JC, Luque R, Sepulveda-Escribano A. Transformations of biomass-derived platform molecules: from high added-value chemicals to fuels via aqueous-phase processing. Chem Soc Rev. 2011;40(5):266–81.

3. Ahorsu R, Medina F, Constantí M. Significance and challenges of biomass as a suitable feedstock for bioenergy and biochemical production. Energies. 2018;11(12):3426–38.

4. Leff B, Ramankutty N, Foley JA. Geographic distribution of major crops with an appropriate use of the genetic variability of flowering time. PNAS. 2018;115(2):6942–7.

5. Parent B, Leclere M, Lacube S, Semenov MA, Welcker C, Martre P, Tardieu F. Maize yields over Europe may increase in spite of climate change, with an appropriate use of the genetic variability of flowering time. PNAS. 2018;115(2):6942–7.

6. Nascimento MS, Santana ALBD, Maranhão CA, Oliveira LS, Bieber L. Phenolic extracts and natural resistance of wood. 2013.

7. Gilbert HJ. The biochemistry and structural biology of plant cell wall deconstruction. Plant Physiol. 2010;153(2):444–55.

8. Cantrell SA, Dianese JC, Fell J, Gunde-Cimerman N, Zalar P. Unusual fungal niches. Mycologia. 2011;103(6):1161–74.

9. Chundawat SPS, Beckham GT, Himmel ME, Dale BE. Deconstruction of lignocellulosic biomass to fuels and chemicals. Annu Rev Chem Biomol. 2011;12:1–45.

10. Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Fouad TT. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science. 2007;315(5813):804–7.

11. Bajpai P. Chapter 2-Xylan: occurrence and structure. In: Xylanolytic enzymes. Amsterdam: Academic Press; 2014. pp. 9–18.

12. Holtzapple MT. Hemicelluloses. In: Caballero B, Finglas P, Trugo L, editors. Encyclopedia of food sciences and nutrition (Second Edition). Oxford: Academic Press; 2003. p. 3060–71.

13. York WS, Vansalbeek H, Davill AG, Albersheim P. Structural analysis of fungal cell walls. In: Donev R, editor. Advances in protein chemistry and structural biology. vol. 98. Academic Press; 2015. pp. 223–261.

14. dos Santos MA, Grenha A. Chapter Seven—polysaccharide nanoparticles for protein and peptide delivery: exploring less-known materials. In: Donev R, editor. Advances in protein chemistry and structural biology. vol. 98. Academic Press; 2015. pp. 223–261.

15. Willför S, Sundberg K, Tenkanen M, Holmbom B. Spruce-derived mannan—a potential raw material for hydrocolloids and novel advanced natural materials. CarbohydrPolym. 2008;72(2):197–210.

16. Mohnen D. Pectin structure and biosynthesis. Curr Opin Plant Biol. 2008;11(3):266–77.

17. Kashyap DR, Vohra PK, Chopra S, Tewari R. Applications of pectin. In: Donev R, editor. Advances in protein chemistry and structural biology. vol. 98. Academic Press; 2015. pp. 223–261.

18. Wilflor S, Sundberg K, Tenkanen M, Holmbom B. Spruce-derived mannan—a potential raw material for hydrocolloids and novel advanced natural materials. CarbohydrPolym. 2008;72(2):197–210.

19. Willför S, Sundberg K, Tenkanen M, Holmbom B. Spruce-derived mannan—a potential raw material for hydrocolloids and novel advanced natural materials. CarbohydrPolym. 2008;72(2):197–210.

20. Mohnen D. Pectin structure and biosynthesis. Curr Opin Plant Biol. 2008;11(3):266–77.

21. Kashyap DR, Vohra PK, Chopra S, Tewari R. Applications of pectin. In: Donev R, editor. Advances in protein chemistry and structural biology. vol. 98. Academic Press; 2015. pp. 223–261.

22. Nelson PE, Dignani MC, Anaisse EJ. Taxonomy, biology, and clinical aspects of Fusarium species. Clin Microbiol Rev. 1994;7(4):479–80.

23. Wellenweber H, Reinköig O. Die Fusarien: Ihre Beschreibung, Schadwirkung und Bekämpfung. Berlin: P. Parey; 1935.

24. Snyder WC, Hansen HN. The species concept in Fusarium with reference to section martiella. Am J Bot. 1941;28(9):738–42.

25. Summerell BA, Laurence MH, Liew ECY, Leslie JL. Biogeography and phylogeny of Fusarium: a review. Fungal Divers. 2010;44(1):1–3.

26. Coleman JJ. The Fusarium solani species complex: ubiquitous pathogenicity and host specificity for crop disease. Annals of Applied Biology. 2001;138(4):397–409.

27. Lopes AM, Ferreira EX, Moreira LRS. An update on enzymatic cocktails for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2016;44(8):e80.

28. Mueller GM, Bills GF. Introduction. In: Biodiversity of fungi: classification and identification. Burlington: Academic Press; 2014. pp. 9–18.
50. Bey M, Zhou SM, Poidevin L, Henriiss R, Coutinho PM, Berrin JS, Sigol- lot JC. Cello-oligosaccharide oxidation reveals differences between two lytic polysaccharide monooxygenases (Family GH61) from Podospora anserina. Appl Environ Microb. 2013;79(2):488–96.

51. Phillips CM, Beeson WT, Gate JH, Marletta MA. Cellulose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by Neurospora crassa. Acs Chem Biol. 2011;6(12):1399–406.

52. Konno N, Ishida T, Igarashi K, Fushinobu S, Habu N, Samejima M, Isogai A. Crystal structure of polysaccharide lyase family 20 endo-β-1,4-glucuronanase from the filamentous fungus Trichoderma reesei. Febs Lett. 2009;583(8):1323–6.

53. Chen Y, Guo XN, Zhu M, Chen C, Li DC. Polysaccharide monooxygenase-catalyzed oxidation of cellulose to gluconic acid-containing cello-oligosaccharides. Biotechnol Biofuels 2019;12.

54. Prado-Martinez M, Anzaldo-Hernández J, Becerra-Aguilar B, Palacios-Juarez H, Vargas-Radillo JD, Renteria-Urquiza M. Characterization of maize leaves and of sugarcane bagasse to elaborate of a mixed cell- lulose pulp. Madera Bosques. 2012;18(3):37–51.

55. Attia M, Stepper J, Davies GJ, Brumer H. Functional and structural characterization of a putative GH74 endo-xylulcanase from the soil saprophyte Cellulopha japonica unravels the first step of xylulcan degradation. Febs J. 2016;283(8):1701–19.

56. Foumani M, Vuong TV, Master ER. Altered substrate specificity of the gluco-oligosaccharide oxidase from Acremonium strictum. Biotechnol Bioeng. 2011;108(10):2261–9.

57. Lo Leggio L, Simmons TJ, Poulsen JCN, Frandsen KEH, Hemsworth GR, Stringer MA, von Freeseleben P, Tovborg M, Johansen KS, De Maria L, et al. Structure and boosting activity of a stalk-degrading lytic polysaccharide monooxygenase. Nat Commun. 2015;6.

58. Vu VV, Beeson WT, Span EA, Farquhar ER, Marletta MA. A family of stalk-active polysaccharide monooxygenases. PNAS. 2014;111(138):13822–7.

59. Poutanen K, Sundberg M, Korte H, Puls J. Deacetylation of xylans by acetyl esterases of Trichoderma reesei. Appl Microbiol Biot. 1990;33(3):506–10.

60. Saykhedkar S, Ray A, Ayobi-Canaan P, Hartson SD, Prade R, Mort AJ. Comparative analysis of fungal polysaccharide lytic polysaccharide monooxygenases (Family GH61) from Penicillium echinulatum. Acs Chem Biol. 2013;8(12):3277–86.

61. Zhao ZT, Liu HQ, Wang CF, Xu JR. Comparative secretome analysis of Trichoderma asperellum S4F8 and T. harzianum rut C30 during solid-state fermentation on sugarcane bagasse. Biotechnol Biofuels. 2013;6(1):172.

62. Delabona PD, Cota J, Hoffmann ZB, Farias CA, Ricardo FS. Comparative secretome analysis during growth on sugarcane bagasse of Trichoderma reesei by acetyl esterases of Trichoderma reesei growing on sorghum stover. Biotechnol Biofuels 2012;5.

63. Couturier M, Navarro D, Favel A, Haon M, Lesage-Meessen A. Crystal structure of polysaccharide lyase family 20 endo-β-1,4-lulose pulp. Madera Bosques. 2012;18(3):37–51.

64. Borin GP, Sanchez CC, de Souza AP, de Santana ES, de Souza AT, Leme JPLF, Lima DJ, Squina FM, Ruller R, Pradella JGD. Understanding the cellulolytic system of Trichoderma harzianum P49P11 and enhancing saccharification of pretreated sugarcane bagasse by supplementation with pectinase and α1-arabinofuranosidase. Bioreourc Technol. 2013;131:500–7.

65. Rodriguez A, Perestelo F, Carriero A, Regalado V, Perez R, de la Fuente G, Falcon MA. Degradation of natural lignins and lignocellulosic substrates by soil-inhabiting fungi imperfecti. Fems Microbiol Ecol. 1996;21(3):213–9.

66. Mitra J, Mukherjee PK, Kale SP, Murthy N. Bioremediation of DDT in soil by genetically improved strains of soil fungus Fusarium solani. Biodegrad. 2001;12(4):235–45.

67. Wang K, Chen J, Sun S-N, Sun R-C. Chapter 6 - Steam Explosion. In: Pandey A, Negi S, Binod P, Larroche C, editors. Pretreatment of biomass. Pandey A, Binod P, Negi S, editors. Pretreatment of lignocellulosic biomass—an overview. Bioresource Technol. 2016;199:49–58.

68. Raper KB. A manual of the penicillia. Baltimore: Williams and Wilkins; 1949.

69. Mandels M, Reese ET. Induction of cellulase in fungi by cellobiose. J Bacteriol. 1982;128:1719–29.

70. Scully ED, Hooper K, Carlson J, Tien M, Geib SM. Analysis of Fusarium solani isolated from the longhorned beetle, Anoplophora glabripennis. PLoS ONE. 2012;7(4):e32990.
93. Brankovics B, Zhang H, van Diepeningen AD, van der Lee TAJ, Waalwijk C, de Hoog GS. GRABB: selective assembly of genomic regions, a new niche for genomic research. PloS Comput Biol. 2016;12(6):e1004753.
94. Ye YZ, Choi JH, TangHX. RAPSearch: a fast protein similarity search tool for short reads. BMC Bioinform. 2011;12.
95. Zhao YA, TangHX, Ye YZ. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. Bioinformatics. 2012;28(1):125–6.
96. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42(D1):D222–30.
97. Hahnke RL, Stackebrandt E, Meier-Kolthoff JP, Tindall BJ, Huang SX, Rohde M, Lapidus A, Han J, Trong S, Haynes M et al. High quality draft genome sequence of Flavobacterium rivuli type strain WB 3.3-2(T) (DSM 21788(T)), a valuable source of polysaccharide decomposing enzymes. Stand Genomic Sci 2015;10.
98. Zhang H, Yohe T, Huang L, Entwistle S, Wu PZ, Yang ZL, Busk PK, Xu Y, Yin YB. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018;46(W1):W95–101.
99. Mandels M, Weber J. The production of cellulases. Am Chem Soc. 1969;95:391.
100. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970;227:680.
101. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc. 2007;1:2856–60.
102. Miller GL. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem. 1959;31:426.

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