1 Whole Genome Sequence and CAZyme distribution of the cellulase hyper producing
2 filamentous fungus *Penicillium janthinellum* NCIM 1366

3 Meera Christopher ^{1,2}, Athiraraj Sreeja-Raju ^{1,2}, Prajeesh Kooloth-Valappil ^{1,2}, Amith
 4 Abraham ³, Digambar Vitthal Gokhale ⁴, Rajeev K. Sukumaran ^{1*}

5 ¹ *Biofuels and Biorefineries Section, Microbial Processes and Technology Division, CSIR-*
 6 *National Institute for Interdisciplinary Science and Technology, Industrial Estate P.O.,*
 7 *Pappanamcode, Thiruvananthapuram- 695019, Kerala, India*

8 ² *Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, Uttar Pradesh-*
 9 *201002, India*

10 ³ *Department of Chemical Engineering, Hanyang University, Seoul, Republic of Korea*

11 ⁴ *National Chemical Laboratory, Pune, India*

12 *** Corresponding Author**

13 *Rajeev K. Sukumaran*

14 *Biofuels and Biorefineries Section, Microbial Processes and Technology Division*

15 *CSIR-National Institute for Interdisciplinary Science and Technology*

16 *Industrial Estate P.O., Pappanamcode, Thiruvananthapuram 695019, Kerala, India*

17 *Tel : +91 471 2515368 (O)*

18 *Fax : +91 471 2491712*

19 *email : rajeevs@niist.res.in*

1   **Abstract**

2   *Penicillium janthinellum* NCIM 1366, capable of secreting cellulases that are highly efficient
3   in the hydrolysis of lignocellulosic biomass, was sequenced to understand its cellulolytic
4   machinery. *De novo* sequencing and assembly revealed a 37.6 Mb genome encoding 11,848
5   putative proteins, 93% of which had significant *BLAST-P* hits. The majority of the top hits
6   (those with over 60% UniProt identity) belonged to *P. brasilianum*. Carbohydrate active
7   enzymes (CAZymes) and other enzymes involved in lignocellulose degradation were also
8   predicted from this strain and compared with those of the industrial workhorse of cellulase
9   production- *Trichoderma reesei* RUT-C30. The comparison showed that the fungus encodes
10   a far higher number of CAZymes (422) as compared to *T. reesei* RUT-C30 (244), which
11   gives a plausible explanation for its overall effectiveness in biomass hydrolysis. An analysis
12   of the secreted CAZymes and annotated ligninases identified 216 predicted proteins which
13   may be directly involved in the breakdown of lignocellulose.

14   **Keywords**

15   *Penicillium*, genome, CAZymes, cellulase, biofuel, lignocellulose

1 1. Introduction

2 The continuous cycling of carbon compounds is one of the fundamental activities in nature.
3 Together with bacteria, fungi are responsible for recycling the bulk of recalcitrant polymers
4 such as lignocellulose [1]. Cellulolytic fungi, i.e., fungi capable of utilizing cellulose-
5 containing material, are widespread in nature and belong to diverse subdivisions like
6 *Ascomycetes*, *Basidiomycetes*, and *Deuteromycetes*. Amongst them, the most studied aerobic
7 fungal genera are *Chaetomium*, *Coriolus*, *Phanerochaete*, *Poria*, *Schizophyllum*, *Serpula*,
8 *Aspergillus*, *Fusarium*, *Geotrichum*, *Paecilomyces*, *Penicillium* and *Trichoderma* [2,3].

9 The efficient degradation of lignocellulose requires a mixture of different enzymes-
10 cellulases, hemicellulases and ligninases- acting sequentially and in tandem. Cellulases have
11 been traditionally categorized into three different classes by the International Union of
12 Biochemistry (IUB) based on their mode of catalytic action- Endo-1,4-glucanases (EGs) (EC
13 3.2.1.4) that cleave the cellulose chains internally, Exoglucanases (cellobiohydrolases-
14 CBHs) that include cellodextrinases (EC 3.2.1.74), non-reducing end acting CBHs (EC
15 3.2.1.91) and reducing end acting CBHs (EC 3.2.1.176) that collectively degrade cellulose
16 from free chain ends; and 1,4- β -glucosidases (EC 3.2.1.21) that hydrolyze soluble
17 cellodextrins and cellobiose into glucose [4]. More recently, carbohydrate-active enzymes
18 (CAZymes), which includes lignocellulases, have been classified into different families based
19 on their structural similarities. These include glycoside hydrolases (GH), glycosyl
20 transferases (GT), carbohydrate esterases (CE), polysaccharide lyases (PL) and enzymes with
21 auxiliary activities (AA) such as the oxidoreduction of phenolic carbohydrates [5].

22 One of the most efficient and best characterized producers of cellulases is *Trichoderma reesei*
23 (teleomorph: *Hypocrea jecorina*) and its hypersecreting mutants RUT C-30 and CL847 [6,7].
24 *T. reesei* secretes large amounts of all three classes of cellulases required for the degradation

1 of cellulose. To date, 9 secretory cellulase encoding genes have been discovered and from *T.*
 2 *reesei*- two CBHs, five EGs, and two BGLs [8–10]. One of the inherent drawbacks of the *T.*
 3 *reesei* cellulase system is its low content of BGLs (~0.5% of the total secreted cellulases),
 4 which leads to insufficient degradation of cellobiose and consequently causes feedback
 5 inhibition of the endo- and exo-glucanases. To overcome this, commercially available
 6 cellulase cocktails, such as those from Novozymes (C-Tec™) and Genencor International
 7 Inc. (Accellerase™) rely on the supplementation of heterologous BGLs to improve
 8 saccharification efficiency [11,12].

9 In the last 20 years, the ascomycete *Penicillium* has emerged as a viable producer of
 10 cellulases [13,14] and sometimes bettering the established industrial work horse for cellulase
 11 production *Trichoderma reesei* RUT C30 [15]. Although this genus is traditionally known for
 12 industrial drug production, many species, such as *P. pinophilum* [16], *P. brasilianum* [17], *P.*
 13 *citrinum* MTCC 6489 [18], *P. decumbens* [19] and *P. echinulatum* [20,21], are now known
 14 to have extracellular enzyme systems with good lignocellulose hydrolysis potential. Most of
 15 these species possess an extensive array of cellulases and hemicellulases that have good
 16 stability at 50 °C, a feature that is essential for industrial applications in cellulose hydrolysis.
 17 Additionally, in comparison to the enzyme complex of *T. reesei*, some species, such as *P.*
 18 *echinulatum*, have been reported to show more BGL activity in relation to the overall
 19 cellulase activity [22].

20 Previous studies using *P. janthinellum* NCIM 1366 (hereafter referred to as PJ-1366) and its
 21 mutants have reported that its cellulases are very efficient in hydrolyzing biomass, including
 22 highly recalcitrant lignocelluloses like cotton stalks, at low enzyme loadings [23]. Profiling of
 23 the enzyme activities also revealed higher amounts of endoglucanases and β -glucosidases
 24 (BGL) per milliliter of the enzyme as compared to *T. reesei* cellulase. Supplementation of the
 25 crude *P. janthinellum* cellulases with BGL from *Aspergillus* species has led to enhancements

1 of its biomass hydrolyzing potential; and in some combinations, it has performed at par or
2 better than the world's best commercial enzymes for biomass hydrolysis [24]. A recent
3 comparison of secretomes by Sreeja-Raju *et al.* [25] has also revealed that PJ-1366 secretes a
4 higher number of CAZymes on induction, as compared with *T. reesei* RUT-C30 (henceforth
5 given as TR-RC30).

6 With the implementation of a bio-based economy becoming a logical and necessary option
7 due to environmental and economic concerns, there has been a rising focus on the conversion
8 of cheap and abundantly-available lignocellulose-rich crop residues into a variety of
9 commodity and specialty chemicals and materials, in addition to potential liquid
10 transportation fuels like bioethanol. However, the first step in any of these processes is the
11 biochemical conversion of cellulosic biomass into fermentable sugars by using enzymes,
12 micro-organisms, or other catalysts. Enzymatic transformation of cellulose and hemicellulose
13 into monosaccharides is an effective and sustainable choice, however, worldwide the
14 commercialization of biomass saccharification technologies have proved to be difficult due to
15 the prohibitively high cost of the hydrolases used.

16 In this scenario, the genome of PJ-1366 has been sequenced, proteins have been annotated
17 using the UniProt Fungi database, and available databases on CAZymes have been used to
18 identify the putative lignocellulose-degrading enzymes of this fungus. This knowledge base
19 will serve as a platform for characterizing the potent cellulase system of this strain, and using
20 it for the development of better biomass hydrolyzing enzyme mixtures.

21 **2. Results**

22 2.1. Strain confirmation and phylogenetic analysis

23 *BLAST-N* analysis of the ITS region identified the isolate as *Penicillium janthinellum*. The
24 highest similarity (query coverage- 98%, percent identity- 99%) was with the ITS region of

1 *Penicillium janthinellum* strain NZD-mf104 (Accession No. **KM278042.1**). The slanted
2 cladogram generated for the phylogenetic analysis showed that PJ-1366 is developed from a
3 basal taxon that also gave rise to *P. brasilianum*, and evolved separately from *T. reesei*. As
4 per the analysis, PJ-1366 is part of a sister taxa to currently reported *P. janthinellum* strains
5 (Fig. 1).

6 2.2. Whole genome sequencing by NGS, de novo assembly and annotation

7 A total of 30,693,398 raw reads corresponding to 7673.34 megabases, with approximately
8 100-fold genome coverage were generated by paired-end sequencing. The average base
9 quality was above Q30 (error-probability ≤ 0.0001) for 93.4% of bases. After data pre-
10 processing, the final data set used for assembly contained 30,055,390 reads covering 7080.7
11 megabases.

12 The final assembly consisted of 1773 contigs, with an N50 of 71,572 bp. Based on the
13 assembly, the genome size was estimated to be 37.62 Mb, with a GC content of 50.72%. The
14 genome assembly was deposited in NCBI as a whole genome shotgun sequencing project
15 with accession no. **NPFE00000000.1**

16 Gene prediction servers predicted 11,848 genes from the genome of PJ-1366. The total length
17 of the coding sequences was 18,352,632 bp, which corresponds to 48.78% of the genome.
18 Out of this, 11,036 predicted proteins (93.14% of total predicted proteins) had significant
19 *BLAST-P* matches with protein sequences in the Fungal UniProt Knowledge Base. Around
20 94% of the proteins found using *BLAST-P* had a confidence level of at least $1e^{-50}$, which
21 indicates a high level of protein conservation. Nearly 89% of the predicted proteins had a
22 BLAST similarity score of more than 60% at the protein level with existing proteins in
23 UniProt Fungi. The top *BLAST-P* hit of each protein was studied and it was observed that the
24 majority of the top hits belonged to *Penicillium brasilianum* sp. (Fig. 2)

1 The gene ontology (GO) terms extracted for the predicted proteins showed that 2721 proteins
2 had predicted molecular functions, 2866 were involved in biological processes and 1992 were
3 cellular components. Of the proteins having molecular function, the majority of them were
4 involved in DNA binding [GO:0003677]. Among the proteins involved in biological
5 processes, 344 were predicted to have a role in transmembrane transport [GO:0055085]. 1364
6 of the predicted proteins were annotated as being integral component of membrane
7 [GO:0016021] (Fig. 3). Additionally, 206 putative tRNA genes were also extracted by
8 tRNAscan-SE.

9 2.3. CAZymes of PJ-1366 and TR-RC30

10 Prediction of CAZymes using the dbCAN server revealed that the genome of PJ-1366
11 encoded 422 putative CAZymes, as compared to 244 from the genome of TR-RC30. While
12 the number of predicted glycosyl transferases (enzymes that participate in glycosidic bond
13 formation by catalyzing the transfer of sugar moieties from activated donor molecules to
14 specific acceptor molecules) was only slightly higher for PJ-1366, there was a significant
15 difference in the number of enzymes belonging to the other CAZy classes: 1.8x times more
16 glycoside hydrolases (enzymes which hydrolyze the glycosidic bond between two or more
17 carbohydrates or between a carbohydrate and a non-carbohydrate moiety); 12x more
18 polysaccharide lyases (enzymes that cleave uronic acid-containing polysaccharide chains),
19 >2x more carbohydrate esterases (enzymes that catalyze the de-O or de-N-acylation of
20 substituted saccharides) and 1.7x more enzymes with auxiliary activities (Fig. 4).

21 2.4. Extracellular biomass hydrolyzing enzymes of PJ-1366

22 Further, SignalP analysis of PJ-1366 CAZymes showed that 216 of the proteins were
23 extracellular, and could thus be directly involved in biomass hydrolysis. This comprised of
24 165 GH proteins, 17 proteins belonging to AA class, 10 PLs, 14 CEs and 10 GT proteins. Of

1 the extracellular proteins, 81 of them had predicted CBMs (carbohydrate binding module)
2 that can bind the carbohydrate ligand and direct the catalytic machinery onto its substrate,
3 thus capable of enhancing the catalytic efficiency of the multimodular CAZyme.

4 Comparison of the putative secreted CAZymes of PJ-1366 with those of TR-RC30 (131
5 secreted CAZymes) revealed interesting differences- there is an appreciable difference in the
6 distribution of certain CAZy families. PJ-1366 encodes a far higher no. of AA3 (cellobiose
7 dehydrogenase), AA9 (LPMO), CE2 (acetyl xylan esterase), CE8 (pectin methylesterase),
8 GH13 (alpha-amylase), GH19 (chitinase), GH28 (polygalacturonase), GH3 (beta-
9 glucosidase), GH32 (invertase), GH35 (beta-galactosidase), GH43 (beta-xylosidase), GH5
10 (endo-beta-1,4-glucanase), GT5 (UDP-Glc: glycogen glucosyltransferase), PL1 (pectate
11 lyase) and PL4 (rhamnogalacturonan endolyase) family genes, while TR-RC30 has a higher
12 no. of GH18 (chitinase), GH64 (beta-1,3-glucanase), GH65 (α,α -trehalase), GH67 (alpha-
13 glucuronidase), GH74 (endoglucanase), GT15 (glycolipid 2-alpha-mannosyltransferase),
14 GT17 (β -1,4-mannosyl-glycoprotein β -1,4-N-acetylglucosaminyltransferase) and PL20
15 (endo- β -1,4-glucuronan lyase) proteins (Fig. 5).

16 Among the secreted PJ-1366 CAZymes, the most prevalent were GH28 (13 genes), GH3 (13
17 genes), GH43 (11 genes), GH18 (10 genes), GH5 (10 genes), AA3 (8 genes), and CE8,
18 GH30, GH35, GH76 and PL1 (6 genes each). Six families of auxiliary activity enzymes were
19 represented- AA1 (laccase), AA3, AA7 (glucooligosaccharide oxidase) and 3 families of
20 LPMOs (AA11, AA13 and AA9). Six families of CEs were present, amongst which CE8 was
21 the highest in number. 50 families of GHs were present; the number of GT families
22 represented was 7. As compared to the sole PL of TR-RC30, which is an endo-beta-1,4-
23 glucuronan lyase, PJ-1366 encoded 4 families of PLs- 6 genes of PL1, one PL26
24 (rhamnogalacturonan exolyase) and 3 rhamnogalacturonan endolyases (PL11 and PL4) (Fig.

6). Totally, PJ-1366 encoded CAZymes belonging to 73 unique families, while the CAZymes of TR-RC30 belonged to 59 different families.

From the 165 GHs, 40 sequences were identified which belonged to putative cellulase families GH12, GH131, GH2, GH3, GH30, GH5, GH6 and GH7. *BLAST-P* analysis of these sequences revealed that 19 of them are potential cellulases- 4 cellobiohydrolases, 6 endoglucanases and 9 betaglucosidases (Table 1). Out of the remaining 146 sequences, eight sequences were predicted as alpha/glucoamylases, nine sequences were chitinases/chitin synthases, five sequences were GH72 transglycosylases and two sequences were annotated as belonging to the six-hairpin glycosidase superfamily (Supplementary File 1).

Further, using the available UniProt annotation, 71 sequences were flagged as probable hemicellulases. The most common were xylanases/xylosidases (18 sequences) which cleave the main chain and subsequent oligosaccharides, followed by galactanases/galactosidases (16 sequences). Along with mannosidases (14 sequences), arabinases (9 sequences), rhamnosidases (2 sequences) and galacturonidases (2 sequences), these enzymes are involved in cleaving side chains that bind hemicellulose to cellulose and lignin [26]. Ten sequences encoding polygalacturonases were also present, which, together with pectin lyases (PL1), are involved in the breakdown of pectin [27] (Supplementary File 1)

2.5. *Ligninolytic enzymes of PJ-1366*

Interestingly, the only extracellular ligninolytic enzymes predicted by the dBCAN server were two AA1 laccases (ctg7180000015091.g77 and ctg7180000015373.g58). While laccases (EC 1.10.3.2) facilitate the initial depolymerization of lignin by decarboxylation, demethylation and demethoxylation of its constituent phenolic acid units [28], further degradation requires lignin peroxidases (LiP, EC 1.11.1.14), manganese peroxidases (MnP,

1 EC 1.11.1.13) and versatile peroxidases (VP, EC 1.11.1.16), which are usually grouped in the
2 AA2 family.

3 There were also no AA2 proteins predicted amongst intracellular CAZymes, though there
4 were a further 2 laccases and one AA5 (oxidase with oxygen as acceptor (EC 1.1.3.-)) protein
5 (Supplementary File 2). An analysis of the annotated proteins of PJ-1366 revealed that there
6 are 5 secretory sequences that are annotated as phenol oxidases and have significant
7 similarity (>60% identity) to fungal laccases in NCBI Protein databases (Table 2). A search
8 for lignin/manganese peroxidases did not give any significant results.

9 **3. Discussion**

10 In nature, different organisms are capable of degrading lignocellulose; however, the most
11 effective ones are the fungi belonging to *Ascomycota*, like *Aspergillus*, *Trichoderma* and
12 *Penicillium*. An analysis of the sequenced genomes of some cellulolytic fungi reveals that the
13 size varies from ~ 26 Mb for *P. digitatum* and *P. marneffei* strains [29,30] to nearly 40 Mb
14 for *T. harzianum* CBS 226.95 [31]. The genome of PJ-1366 is on the larger side, as compared
15 to other sequenced *Penicillium* sp. like *P. chrysogenum* Wisconsin 54-1255 (32.19 Mb) [32]
16 and *P. aurentiogriseum* NRRL 62431 (32.7 Mb) [33]. However, the number of predicted
17 proteins of PJ-1366 (11,848) is comparable to that of *P. aurentiogriseum* (11,476), but higher
18 than that of *P. digitatum* strains (~ 9000). Both the genome size and the number of putative
19 proteins of PJ-1366 is appreciably larger than that of the industrial standard for cellulase
20 production- TR-RC30 (32.7 MB genome encoding 9852 proteins)

21 Cellulases are multicomponent enzyme mixtures that work in tandem to degrade celluloses
22 and hemicelluloses. In addition to hydrolytic enzymes, an oxidative enzyme system
23 comprised of monooxygenases, dehydrogenases and peroxidases is also involved in boosting

1 the breakdown of lignocellulose [34,35]. The dbCAN server predicted 422 CAZymes for PJ-
2 1366, a similar number to that reported for *P. decumbens* [19] and *P. verruculosum* [36], but
3 nearly 1.7 times more than the 244 CAZymes reported for TR-RC30

4 SignalP analysis of the putative CAZymes revealed that more than half of them are
5 extracellular, and thus may be directly involved in the degradation of lignocellulose. Of these,
6 19 sequences were annotated (using UniProt Fungi and CAZy databases) as cellulases- four
7 CBHs, six EGs and nine BGLs. The CBHs included two each of GH6 and GH7 family
8 proteins. Cellobiohydrolases are considered as the main degraders of crystalline cellulose,
9 with GH6 CBHs being specific towards non-reducing ends of the cellulose chain, and GH7
10 CBHs attacking the reducing ends [37]. A mutant strain of *P. verruculosum* is known to
11 secrete five different CBHs, but TR-RC30 secretes only two CBHs (one each of GH6 and
12 GH7), which together account for ~80% of its total cellulolytic activity [8,9]. CBHs are
13 crucial in hydrolyzing pretreated lignocelluloses since the pretreatments carried out to
14 remove the amorphous hemicelluloses and lignin generally result in increased contents of
15 crystalline cellulose [38,39].

16 Endoglucanases, on the other hand, cleave the amorphous regions in cellulose chains and are
17 widespread among GH families [40]. Six putative extracellular EGs were identified from PJ-
18 1366, all of which were associated with CBM domains. The presence of a CBM increases
19 substrate binding and generally leads to more thermostable enzymes [41]. Also, the number
20 of predicted EGs in PJ-1366 is higher than both *P. verruculosum* (5) and TR-RC30 (5), which
21 is renowned for its high endoglucanase activity [2]. Even trace levels of EG activity can
22 stimulate CBHs [42], and this might be a reason why, in spite of having similar numbers of
23 EG genes, there is a marked difference in the endoglucanase and FPase activities of cellulases
24 from PJ-1366 and TR-RC30 [25]. Additionally, proteins in GH families are reported to have

1 overlapping specificities, so the function of EGs may also be contributed by the other
2 hydrolases.

3 PJ-1366 encodes nine putative extracellular GH3 family BGLs, in contrast to the two secreted
4 GH1 and GH3 family BGLs of TR-RC30. GH3 BGLs are known to be functionally diverse,
5 having different pH optima, glucose tolerance abilities, and substrate specificities [43]. Many
6 *Penicillium* species have been reported to have a higher BGL activity than TR-RC30 [44–46],
7 which is advantageous since BGL catalyzes the rate-limiting step in cellulolysis; and
8 therefore the enzyme, if sourced from this fungus may have a bearing on the operational cost
9 of biorefineries [47].

10 In addition to encoding more cellulases, the number of predicted hemicellulase genes of PJ-
11 1366 (71) also far exceeds the 16 characterized hemicellulases of TR-RC30 [48]. The
12 hydrolysis of hemicellulose is a vital aspect of lignocellulose degradation since hemicellulose
13 typically constitutes 20-40% of cellulose enriched biomass [49]. TR-RC30 has been reported
14 to have a lower secretion of xylan-degrading enzymes, while *Penicillium* species has been
15 reported to secrete a more diversified set of enzymes [6]. The repertoire of PJ-1366
16 hemicellulases include xylanases and mannanases- that cleave the xylan and mannan main
17 chains of hemicellulose; galacturonases, furanosidases and endo-1,3-glucanases- that digest
18 side chain moieties; and xylosidases, mannosidases and galactosidases- that complete the
19 digestion of oligosaccharides derived from hemicellulose [50]. Together with the BGLs, the
20 GH16/GH72 family proteins (which have probable β -transglycosylation activity), and the ten
21 secreted glycosyltransferases, these enzymes may also be involved in generating soluble
22 inducers for genome-wide induction of cellulases [51,52]. Studies on bacterial cellulase
23 systems have suggested that while not directly involved in biomass degradation, GT proteins
24 are also invaluable for functionalization (via glycosylation) of the effectors of biomass
25 hydrolysis [53]

1 Sequences for AA class proteins, which cleave cellulose by oxidative mechanisms, have also
 2 been detected, including lytic polysaccharide monooxygenases (LPMOs), which are now
 3 understood to create access points for cellulases and hemicellulases [34,54,55]. Surprisingly,
 4 only some of the enzymes required for lignin degradation were present among the predicted
 5 CAZymes of PJ-1366. While a number of lignases have been predicted for TR-RC30 and
 6 other fungi, and metal-containing oxidases potentially linked to lignin degradation have been
 7 detected in the secretome of TR-RC30 [56], no concrete data exists on their expression
 8 patterns. A shortage of information on *Penicillium* peroxidases (550 sequences in NCBI
 9 Protein database, as compared to 26881 sequences for *Aspergillus* peroxidases) may also be a
 10 contributing factor for the non-availability of structure/sequence data on its lignases.
 11 Nevertheless, saprophytic fungi like *Penicillium*, which are involved in recycling
 12 lignocellulose residues in nature, must unquestionably be able to efficiently degrade lignins;
 13 therefore, there is a need to identify and characterize these proteins by studying the
 14 secretomes of lignin-induced fungal cultures.

15 In conclusion, a cursory analysis of the genome of PJ-1366 and its proteins has revealed that
 16 this isolate possesses a complete set of lignocellulose degrading enzymes, which explains its
 17 previously reported ability to hydrolyze recalcitrant biomass without the addition of
 18 accessory enzymes [23,25]. The fact that this fungus possesses a more diversified set of
 19 lignocellulases than TR-RC30 may be partly due to the fact that *P. janthinellum* is a white rot
 20 fungus, unlike *T. reesei* which is generally classified as a soft-rot fungus. By nature of their
 21 adaptations to natural habitats, white-rot fungi are capable of degrading all major wood
 22 components while soft-rot fungi are only capable of surface degradation of cellulose and
 23 hemicellulose. On the whole, the genome assembly and annotation provide a framework for
 24 identifying the cellulase machinery of this fungus. In order to fully characterize its
 25 lignocellulases so as to use it as a base for strain improvement and development of

1 commercial cellulase cocktails, further studies, including secretome analysis of induced
2 fungal cultures coupled with cDNA profiling, is indispensable.

3 **4. Materials and methods**

4 4.1. Strain confirmation and phylogenetic analysis

5 The identity of the strain used for analysis was first reconfirmed by analysis of its ribosomal
6 internal transcribed spacer (ITS) region. 1×10^6 spores were inoculated into potato dextrose
7 broth, and genomic DNA was isolated from the mycelia of 120h old culture by the phenol-
8 chloroform method [57]. Standard protocols were used for the amplification of the ITS region
9 using the conserved eubacterial primers ITS1 (5' TCCGTAGGTGAACCTGCGG 3') and
10 ITS4 (5' TCCTCCGCTTATTGATATGC 3'). The purified PCR product (≈ 500 bp) was then
11 sequenced and the resulting sequence was compared with the NCBI Nucleotide database
12 using *BLAST-N* [58].

13 The highest scoring sequences were retrieved from the database as reference sequences and
14 aligned with the ITS sequence of *P. janthinellum*. To assess the relationship between the
15 different strains/species, the taxonomic similarity of isolates was compared using a
16 phylogenetic tree constructed based on the Fast Minimum Evolution method [59], with the
17 maximum sequence difference set at 0.45. The sequences of the ITS regions of *P.*
18 *brasiliense* (Accession no. **KY701767.1**), *T. reesei* (Accession no. **NR_120297.1**) and the
19 basidiomycete *Ustilago maydis* (Accession no. **AF135431.1**) were used as the outgroups.

20 4.2. Whole genome sequencing by NGS, de novo assembly and annotation

21 Genomic DNA was extracted as previously described, and treated with RNaseA to remove
22 RNA contamination. The purity and concentration of the DNA were checked using a

1 NanoDrop™ ND-1000 UV-Vis spectrophotometer, and the integrity of the DNA was
2 checked by agarose gel electrophoresis.

3 Next Generation Sequencing (NGS) using the Illumina HiSeq 2500 platform was carried out
4 at AgriGenome Labs Private Limited, Cochin, India. Following paired-end sequencing, the
5 quality of the sequences was checked by assessing various parameters like base quality score
6 distributions, average base content per read and the GC distribution in the reads

7 All the raw reads were processed before genome assembly. The Illumina adapter sequences
8 were trimmed using Cutadapt v1.8 [60]. All low quality bases (Phred value score <Q30) were
9 eliminated using Sickle v1.33 [61] and duplicate reads were removed using FastUniq [62].

10 The cleaned reads were subjected to Kmergenie [63] to predict the optimal k-value and
11 assembly size. Contigs were then assembled *de novo* using MaSuRCA v4.0.2 [64]. After the
12 assembly generation, the presence of conserved genes in the assembled contigs was checked
13 using BUSCO v2 [65].

14 The gene prediction was carried out using an Augustus 3.1 [66] and the predicted genes were
15 annotated using AgriGenome's in-house pipeline CANoPI (Contig Annotator Pipeline).
16 Briefly it consisted of comparison of the predicted sequences with UniProt Fungi database
17 using BLAST-P program (matches with E-value $\leq 10^{-5}$ and similarity score $\geq 40\%$ were
18 retained for further annotation), organism annotation, gene and protein annotation to the
19 matched genes, gene ontology annotation and pathway annotation.

20 Finally, tRNA genes were predicted from the contigs using the tRNAscan-SE program [67]

21 4.3. Prediction of CAZymes and identification of lignocellulases

22 The full repertoire of CAZymes from PJ-1366 were predicted using the web-based dbCAN2
23 meta server [68]. For comparison purposes, the list of predicted proteins from the genome of

1 *T. reesei* (*Trichoderma reesei* RUT C-30 v1.0: Project ID- 403119) [69] were downloaded
2 from the JGI Genome portal and CAZymes were extracted from it. Only proteins identified
3 as CAZymes using all three tools- HMMER, DIAMOND and Hotpep, were used for further
4 analysis.

5 The secretory nature of the proteins was predicted using SignalP 5.0 [70]. By comparing the
6 list of secretory CAZymes with the annotated proteins from PJ-1366, a comprehensive list of
7 all lignocellulases, viz., cellulases, hemicellulases, lignases and lytic polysaccharide
8 monooxygenases, was prepared.

9 **5. Data Availability**

10 The genome sequence is available in NCBI under the BioProject accession number
11 **PRJNA389228**. The contig level assembly is present under accession number
12 **ASM236980v1**. The ITS sequence has been submitted at GenBank with accession no.
13 **MT275691.1**.

14 **6. Acknowledgments**

15 MC, ASR and PKV are thankful to the Council of Scientific and Industrial Research (CSIR),
16 Govt. of India for the Junior /Senior research fellowships for Ph.D. studies. This study was
17 partially funded by the Department of Biotechnology, Govt. of India
18 (BT/PR20695/BBE/117/211/2016) and CSIR (MLP 0035: 33/2018/MD-FTT&FTC-ANB).
19 AA is thankful to Kerala State Council for Science, Technology and Environment for a Post-
20 Doctoral Fellowship (001/PDF/2015/KSCSTE). We are thankful to Dr. N. Ramesh Kumar at
21 MPTD, CSIR-NIIST for ITS sequencing service.

References

- [1]. H. Nevalainen, M. Penttilä, Molecular Biology of Cellulolytic Fungi, in: Genet. Biotechnol., 1995. https://doi.org/10.1007/978-3-662-10364-7_18.
- [2]. W. Chulalaksananukul, Beta-Glucosidase from *Trichoderma* to Improve the Activity of Cellulase Cocktails, in: Biotechnol. Biol. *Trichoderma*, 2014. <https://doi.org/10.1016/B978-0-444-59576-8.00019-9>.
- [3]. L. Viikari, A. Suurnäkki, S. Grönqvist, L. Raaska, A. Ragauskas, Forest Products: Biotechnology in Pulp and Paper Processing, in: Encycl. Microbiol., 2009. <https://doi.org/10.1016/B978-012373944-5.00123-1>.
- [4]. L.R. Lynd, Y. Zhang, Quantitative determination of cellulase concentration as distinct from cell concentration in studies of microbial cellulose utilization: Analytical framework and methodological approach, Biotechnol. Bioeng. (2002). <https://doi.org/10.1002/bit.10142>.
- [5]. V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Res. (2014). <https://doi.org/10.1093/nar/gkt1178>.
- [6]. I. Herpoël-Gimbert, A. Margeot, A. Dolla, G. Jan, D. Mollé, S. Lignon, H. Mathis, J.C. Sigoillot, F. Monot, M. Asther, Comparative secretome analyses of two *Trichoderma reesei* RUT-C30 and CL847 hypersecretory strains, Biotechnol. Biofuels. (2008). <https://doi.org/10.1186/1754-6834-1-18>.
- [7]. R. Peterson, H. Nevalainen, *Trichoderma reesei* RUT-C30 - Thirty years of strain improvement, Microbiology. (2012). <https://doi.org/10.1099/mic.0.054031-0>.
- [8]. A.V. Gusakov, Alternatives to *Trichoderma reesei* in biofuel production, Trends Biotechnol. (2011). <https://doi.org/10.1016/j.tibtech.2011.04.004>.
- [9]. L. Rosgaard, S. Pedersen, J. Langston, D. Akerhielm, J.R. Cherry, A.S. Meyer, Evaluation of minimal *Trichoderma reesei* cellulase mixtures on differently pretreated Barley straw substrates, Biotechnol. Prog. (2007). <https://doi.org/10.1021/bp070329p>.
- [10]. R.K. Sukumaran, R.R. Singhanian, A. Pandey, Microbial cellulases - Production, applications and challenges, J. Sci. Ind. Res. (India). 64 (2005) 832–844. <https://doi.org/http://nopr.niscair.res.in/handle/123456789/5375>.
- [11]. R.P. de Vries, J. Visser, Aspergillus Enzymes Involved in Degradation of Plant Cell Wall Polysaccharides, Microbiol. Mol. Biol. Rev. 65 (2001) 497–522. <https://doi.org/10.1128/mmbr.65.4.497-522.2001>.

- 1 [12]. A. Berlin, N. Gilkes, D. Kilburn, R. Bura, A. Markov, A. Skomarovsky, O. Okunev, A.
2 Gusakov, V. Maximenko, D. Gregg, A. Sinitsyn, J. Saddler, Evaluation of novel fungal
3 cellulase preparations for ability to hydrolyze softwood substrates - Evidence for the
4 role of accessory enzymes, *Enzyme Microb. Technol.* 37 (2005) 175–184.
5 <https://doi.org/10.1016/j.enzmictec.2005.01.039>.
- 6 [13]. M.G. Adsul, K.B. Bastawde, A.J. Varma, D. V Gokhale, Strain improvement of
7 *Penicillium janthinellum* NCIM 1171 for increased cellulase production, *Bioresour.*
8 *Technol.* 98 (2007) 1467–1473. <https://doi.org/10.1016/j.biortech.2006.02.036>.
- 9 [14]. H. Jorgensen, A. Morkeberg, K.B.R. Krogh, L. Olsson, Production of cellulases and
10 hemicellulases by three *Penicillium* species: Effect of substrate and evaluation of
11 cellulase adsorption by capillary electrophoresis, *Enzyme Microb. Technol.* 36 (2005)
12 42–48. <https://doi.org/10.1016/j.enzmictec.2004.03.023>.
- 13 [15]. K. Marjamaa, K. Toth, P.A. Bromann, G. Szakacs, K. Kruus, Novel *Penicillium*
14 cellulases for total hydrolysis of lignocellulosics, *Enzyme Microb. Technol.* 52 (2013)
15 358–369. <https://doi.org/10.1016/j.enzmictec.2013.03.003>.
- 16 [16]. E.M. Visser, T.F. Leal, M.N. de Almeida, V.M. Guimarães, Increased enzymatic
17 hydrolysis of sugarcane bagasse from enzyme recycling., *Biotechnol. Biofuels.* 8 (2015)
18 5. <https://doi.org/10.1186/s13068-014-0185-8>.
- 19 [17]. M.N. Anwar, M. Suto, F. Tomita, Isolation of mutants of *Penicillium purpurogenum*
20 resistant to catabolite repression, *Appl. Microbiol. Biotechnol.* (1996).
21 <https://doi.org/10.1007/s002530050748>.
- 22 [18]. T. Dutta, R. Sengupta, R. Sahoo, S. Sinha Ray, A. Bhattacharjee, S. Ghosh, A novel
23 cellulase free alkaliphilic xylanase from alkali tolerant *Penicillium citrinum*:
24 Production, purification and characterization, *Lett. Appl. Microbiol.* (2007).
25 <https://doi.org/10.1111/j.1472-765X.2006.02042.x>.
- 26 [19]. G. Liu, L. Zhang, X. Wei, G. Zou, Y. Qin, L. Ma, J. Li, H. Zheng, S. Wang, C. Wang,
27 L. Xun, G.P. Zhao, Z. Zhou, Y. Qu, Genomic and Secretomic Analyses Reveal Unique
28 Features of the Lignocellulolytic Enzyme System of *Penicillium decumbens*, *PLoS One.*
29 (2013). <https://doi.org/10.1371/journal.pone.0055185>.
- 30 [20]. K. Fujii, Y. Kai, S. Matsunobu, H. Sato, A. Mikami, Isolation of digested sludge-
31 assimilating fungal strains and their potential applications, *J. Appl. Microbiol.* (2013).
32 <https://doi.org/10.1111/jam.12266>.
- 33 [21]. J. Sun, N.L. Glass, Identification of the CRE-1 cellulolytic regulon in *Neurospora*
34 *crassa*, *PLoS One.* 6 (2011). <https://doi.org/10.1371/journal.pone.0025654>.

- 1 [22]. M. Nitta, T. Furukawa, Y. Shida, K. Mori, S. Kuhara, Y. Morikawa, W. Ogasawara, A
2 new Zn(II) 2Cys 6-type transcription factor BglR regulates β -glucosidase expression in
3 *Trichoderma reesei*, Fungal Genet. Biol. 49 (2012) 388–397.
4 <https://doi.org/10.1016/j.fgb.2012.02.009>.
- 5 [23]. M. Christopher, A.K. Mathew, M. Kiran Kumar, A. Pandey, R.K. Sukumaran, A
6 biorefinery-based approach for the production of ethanol from enzymatically
7 hydrolysed cotton stalks, Bioresour. Technol. 242 (2017).
8 <https://doi.org/10.1016/j.biortech.2017.03.190>.
- 9 [24]. K.P. Rajasree, G.M. Mathew, A. Pandey, R.K. Sukumaran, Highly glucose tolerant
10 Beta-glucosidase from *Aspergillus unguis*: NII 08123 for enhanced hydrolysis of
11 biomass, J. Ind. Microbiol. Biotechnol. 40 (2013) 967–975.
12 <https://doi.org/10.1007/s10295-013-1291-5>.
- 13 [25]. A. Sreeja-Raju, M. Christopher, P. Kooloth-Valappil, R. Kuni-Parambil, D.V. Gokhale,
14 M. Sankar, A. Abraham, A. Pandey, R.K. Sukumaran, *Penicillium janthinellum*
15 NCIM1366 shows improved biomass hydrolysis and a larger number of CAZymes with
16 higher induction levels over *Trichoderma reesei* RUT-C30, Biotechnol. Biofuels. 13
17 (2020). <https://doi.org/10.1186/s13068-020-01830-9>.
- 18 [26]. M. Andlar, T. Rezić, N. Marđetko, D. Kracher, R. Ludwig, B. Šantek, Lignocellulose
19 degradation: An overview of fungi and fungal enzymes involved in lignocellulose
20 degradation, Eng. Life Sci. 18 (2018) 768–778. <https://doi.org/10.1002/elsc.201800039>.
- 21 [27]. Y. Yang, Y. Yu, Y. Liang, C.T. Anderson, J. Cao, A profusion of molecular scissors for
22 pectins: classification, expression, and functions of plant polygalacturonases, Front.
23 Plant Sci. 9 (2018). <https://doi.org/10.3389/fpls.2018.01208>.
- 24 [28]. L.P. Christopher, B. Yao, Y. Ji, Lignin biodegradation with laccase-mediator systems,
25 Front. Energy Res. 2 (2014) 12. <https://doi.org/10.3389/fenrg.2014.00012>.
- 26 [29]. M. Marcet-Houben, A.R. Ballester, B. de la Fuente, E. Harries, J.F. Marcos, L.
27 González-Candelas, T. Gabaldón, Genome sequence of the necrotrophic fungus
28 *Penicillium digitatum*, the main postharvest pathogen of citrus, BMC Genomics.
29 (2012). <https://doi.org/10.1186/1471-2164-13-646>.
- 30 [30]. K.Y. Yuen, G. Pascal, S.S.Y. Wong, P. Glaser, P.C.Y. Woo, F. Kunst, J.J. Cai, E.Y.L.
31 Cheung, C. Médigue, A. Danchin, Exploring the *Penicillium marneffe* genome, Arch.
32 Microbiol. (2003). <https://doi.org/10.1007/s00203-003-0533-8>.
- 33 [31]. I.S. Druzhinina, C.P. Kubicek, M. Komo-Zelazowska, T.B. Mulaw, J. Bissett, The
34 *Trichoderma harzianum* demon: Complex speciation history resulting in coexistence of

- 1   hypothetical biological species, recent agamospecies and numerous relict lineages,
2   BMC Evol. Biol. (2010). <https://doi.org/10.1186/1471-2148-10-94>.
- 3   [32]. M.A. Van Den Berg, R. Albang, K. Albermann, J.H. Badger, J.M. Daran, A.J. M
4   Driessen, C. Garcia-Estrada, N.D. Fedorova, D.M. Harris, W.H.M. Heijne, V. Joardar,
5   J.A.K. W Kiel, A. Kovalchuk, J.F. Martín, W.C. Nierman, J.G. Nijland, J.T. Pronk, J.A.
6   Roubos, I.J. Van Der Klei, N.N.M.E. Van Peij, M. Veenhuis, H. Von Döhren, C.
7   Wagner, J. Wortman, R.A.L. Bovenberg, Genome sequencing and analysis of the
8   filamentous fungus *Penicillium chrysogenum*, Nat. Biotechnol. (2008).
9   <https://doi.org/10.1038/nbt.1498>.
- 10   [33]. Y. Yang, H. Zhao, R.A. Barrero, B. Zhang, G. Sun, I.W. Wilson, F. Xie, K.D. Walker,
11   J.W. Parks, R. Bruce, G. Guo, L. Chen, Y. Zhang, X. Huang, Q. Tang, H. Liu, M.I.
12   Bellgard, D. Qiu, J. Lai, A. Hoffman, Genome sequencing and analysis of the
13   paclitaxel-producing endophytic fungus *Penicillium aurantiogriseum* NRRL 62431,
14   BMC Genomics. (2014). <https://doi.org/10.1186/1471-2164-15-69>.
- 15   [34]. J.W. Agger, T. Isaksen, A. Várnai, S. Vidal-Melgosa, W.G.T. Willats, R. Ludwig, S.J.
16   Horn, V.G.H. Eijsink, B. Westereng, Discovery of LPMO activity on hemicelluloses
17   shows the importance of oxidative processes in plant cell wall degradation, Proc. Natl.
18   Acad. Sci. U. S. A. 111 (2014) 6287–6292. <https://doi.org/10.1073/pnas.1323629111>.
- 19   [35]. J.A. Langston, T. Shaghasi, E. Abbate, F. Xu, E. Vlasenko, M.D. Sweeney,
20   Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase
21   and glycoside hydrolase 61, Appl. Environ. Microbiol. (2011).
22   <https://doi.org/10.1128/AEM.05815-11>.
- 23   [36]. V.V. Morozova, A. V. Gusakov, R.M. Andrianov, A.G. Pravilnikov, D.O. Osipov, A.P.
24   Sinitsyn, Cellulases of *Penicillium verruculosum*, Biotechnol. J. (2010).
25   <https://doi.org/10.1002/biot.201000050>.
- 26   [37]. S.A. Teter, K.B. Sutton, B. Emme, Enzymatic processes and enzyme development in
27   biorefining, in: Adv. Biorefineries Biomass Waste Supply Chain Exploit., 2014.
28   <https://doi.org/10.1533/9780857097385.1.199>.
- 29   [38]. M. Christopher, A.K. Mathew, M. Kiran Kumar, A. Pandey, R.K. Sukumaran, A
30   biorefinery-based approach for the production of ethanol from enzymatically
31   hydrolysed cotton stalks, Bioresour. Technol. (2017).
32   <https://doi.org/10.1016/j.biortech.2017.03.190>.

- 1 [39]. L. Zhu, J.P. O'Dwyer, V.S. Chang, C.B. Granda, M.T. Holtzaple, Structural features
2 affecting biomass enzymatic digestibility, *Bioresour. Technol.* (2008).
3 <https://doi.org/10.1016/j.biortech.2007.07.033>.
- 4 [40]. M.D. Sweeney, F. Xu, Biomass converting enzymes as industrial biocatalysts for fuels
5 and chemicals: Recent developments, *Catalysts*. 2 (2012) 244–263.
6 <https://doi.org/10.3390/catal2020244>.
- 7 [41]. A. Várnai, M. Siika-Aho, L. Viikari, Carbohydrate-binding modules (CBMs) revisited:
8 Reduced amount of water counterbalances the need for CBMs, *Biotechnol. Biofuels*.
9 (2013). <https://doi.org/10.1186/1754-6834-6-30>.
- 10 [42]. T.M. Wood, S.I. McCrae, K.M. Bhat, The mechanism of fungal cellulase action.
11 Synergism between enzyme components of *Penicillium pinophilum* cellulase in
12 solubilizing hydrogen bound-ordered cellulose, *Biochem. J.* (1989).
13 <https://doi.org/10.1042/bj2600037>.
- 14 [43]. W. Xia, Y. Bai, Y. Cui, X. Xu, L. Qian, P. Shi, W. Zhang, H. Luo, X. Zhan, B. Yao,
15 Functional diversity of family 3 β -glucosidases from thermophilic cellulolytic fungus
16 *Humicola insolens* Y1, *Sci. Rep.* 6 (2016). <https://doi.org/10.1038/srep27062>.
- 17 [44]. L.F. Martins, D. Kolling, M. Camassola, A.J.P. Dillon, L.P. Ramos, Comparison of
18 *Penicillium echinulatum* and *Trichoderma reesei* cellulases in relation to their activity
19 against various cellulosic substrates, *Bioresour. Technol.* (2008).
20 <https://doi.org/10.1016/j.biortech.2007.01.060>.
- 21 [45]. A.J.P. Dillon, M. Camassola, J.A.P. Henriques, M.H.P. Fungaro, A.C.S. Azevedo,
22 T.A.F. Velho, S.E. Laguna, Generation of recombinants strains to cellulases production
23 by protoplast fusion between *Penicillium echinulatum* and *Trichoderma harzianum*,
24 *Enzyme Microb. Technol.* (2008). <https://doi.org/10.1016/j.enzmictec.2008.07.009>.
- 25 [46]. M. Camassola, A.J.P. Dillon, Cellulases and xylanases production by *Penicillium*
26 *echinulatum* grown on sugar cane bagasse in solid-state fermentation, *Appl. Biochem.*
27 *Biotechnol.* (2010). <https://doi.org/10.1007/s12010-010-8967-3>.
- 28 [47]. N. Srivastava, R. Rathour, S. Jha, K. Pandey, M. Srivastava, V.K. Thakur, R.S. Sengar,
29 V.K. Gupta, P.B. Mazumder, A.F. Khan, P.K. Mishra, Microbial beta glucosidase
30 enzymes: Recent advances in biomass conversion for biofuels application,
31 *Biomolecules*. (2019). <https://doi.org/10.3390/biom9060220>.
- 32 [48]. D. Martinez, R.M. Berka, B. Henrissat, M. Saloheimo, M. Arvas, S.E. Baker, J.
33 Chapman, O. Chertkov, P.M. Coutinho, D. Cullen, E.G.J. Danchin, I. V. Grigoriev, P.
34 Harris, M. Jackson, C.P. Kubicek, C.S. Han, I. Ho, L.F. Larrondo, A.L. De Leon, J.K.

- 1 Magnuson, S. Merino, M. Misra, B. Nelson, N. Putnam, B. Robbertse, A.A. Salamov,
2 M. Schmoll, A. Terry, N. Thayer, A. Westerholm-Parvinen, C.L. Schoch, J. Yao, R.
3 Barbote, M.A. Nelson, C. Detter, D. Bruce, C.R. Kuske, G. Xie, P. Richardson, D.S.
4 Rokhsar, S.M. Lucas, E.M. Rubin, N. Dunn-Coleman, M. Ward, T.S. Brettin, Genome
5 sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn.
6 *Hypocrea jecorina*), Nat. Biotechnol. (2008). <https://doi.org/10.1038/nbt1403>.
- 7 [49]. K. Kucharska, P. Rybarczyk, I. Hołowacz, R. Łukajtis, M. Glinka, M. Kamiński,
8 Pretreatment of lignocellulosic materials as substrates for fermentation processes,
9 Molecules. (2018). <https://doi.org/10.3390/molecules23112937>.
- 10 [50]. P.K. Foreman, D. Brown, L. Dankmeyer, R. Dean, S. Diener, N.S. Dunn-Coleman, F.
11 Goedegebuur, T.D. Houfek, G.J. England, A.S. Kelley, H.J. Meerman, T. Mitchell, C.
12 Mitchinson, H.A. Olivares, P.J.M. Teunissen, J. Yao, M. Ward, Transcriptional
13 regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma*
14 *reesei*, J. Biol. Chem. (2003). <https://doi.org/10.1074/jbc.M304750200>.
- 15 [51]. J. Gao, Y. Qian, Y. Wang, Y. Qu, Y. Zhong, Production of the versatile cellulase for
16 cellulose bioconversion and cellulase inducer synthesis by genetic improvement of
17 *Trichoderma reesei*, Biotechnol. Biofuels. 10 (2017). [https://doi.org/10.1186/s13068-](https://doi.org/10.1186/s13068-017-0963-1)
18 [017-0963-1](https://doi.org/10.1186/s13068-017-0963-1).
- 19 [52]. J.M. Rini, J.D. Esko, A. Varki, Glycosyltransferases and Glycan-Processing Enzymes,
20 in: V. A, C. R, E. J (Eds.), Essentials Glycobiol., Cold Spring Harbor Laboratory Press,
21 2015. <https://www.ncbi.nlm.nih.gov/books/NBK20718/> (accessed January 4, 2021).
- 22 [53]. J. Russell, S.-K. Kim, J. Duma, H. Nothaft, M.E. Himmel, Y.J. Bomble, C.M.
23 Szymanski, J. Westpheling, Deletion of a single glycosyltransferase in
24 *Caldicellulosiruptor bescii* eliminates protein glycosylation and growth on crystalline
25 cellulose, Biotechnol. Biofuels. (2018). <https://doi.org/10.1186/s13068-018-1266-x>.
- 26 [54]. W.T. Beeson, V. V. Vu, E.A. Span, C.M. Phillips, M.A. Marletta, Cellulose
27 Degradation by Polysaccharide Monooxygenases, Annu. Rev. Biochem. 84 (2015) 923–
28 946. <https://doi.org/10.1146/annurev-biochem-060614-034439>.
- 29 [55]. A. Villares, C. Moreau, C. Bennati-Granier, S. Garajova, L. Foucat, X. Falourd, B.
30 Saake, J.G. Berrin, B. Cathala, Lytic polysaccharide monooxygenases disrupt the
31 cellulose fibers structure, Sci. Rep. 7 (2017) 1–9. <https://doi.org/10.1038/srep40262>.
- 32 [56]. I.J. Marx, N. Van Wyk, S. Smit, D. Jacobson, M. Viljoen-Bloom, H. Volschenk,
33 Comparative secretome analysis of *Trichoderma asperellum* S4F8 and *Trichoderma*

- 1   *reesei* Rut C30 during solid-state fermentation on sugarcane bagasse, Biotechnol.
- 2   Biofuels. (2013). <https://doi.org/10.1186/1754-6834-6-172>.
- 3   [57]. M.R. Green, J. Sambrook, Molecular Cloning: A Laboratory Manual, 4th ed., Cold
- 4   Spring Harbor Laboratory Press, 2012.
- 5   [58]. S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment
- 6   search tool, J. Mol. Biol. (1990). [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- 7   [59]. R. Desper, O. Gascuel, Fast and accurate phylogeny reconstruction algorithms based on
- 8   the Minimum-Evolution principle, J. Comput. Biol. (2002).
- 9   <https://doi.org/10.1089/106652702761034136>.
- 10   [60]. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing
- 11   reads, EMBnet.Journal. (2011). <https://doi.org/10.14806/ej.17.1.200>.
- 12   [61]. N. Joshi, J. Fass, Sickel: A sliding-window, adaptive, quality-based trimming tool for
- 13   FastQ files (Version 1.33) [Software], Available at <https://github.com/Najoshi/Sickle>.
- 14   (2011).
- 15   [62]. H. Xu, X. Luo, J. Qian, X. Pang, J. Song, G. Qian, J. Chen, S. Chen, FastUniq: A Fast
- 16   De Novo Duplicates Removal Tool for Paired Short Reads, PLoS One. (2012).
- 17   <https://doi.org/10.1371/journal.pone.0052249>.
- 18   [63]. R. Chikhi, P. Medvedev, Informed and automated k-mer size selection for genome
- 19   assembly, Bioinformatics. (2014). <https://doi.org/10.1093/bioinformatics/btt310>.
- 20   [64]. A.V. Zimin, G. Marçais, D. Puiu, M. Roberts, S.L. Salzberg, J.A. Yorke, The
- 21   MaSuRCA genome assembler, Bioinformatics. (2013).
- 22   <https://doi.org/10.1093/bioinformatics/btt476>.
- 23   [65]. F.A. Simão, R.M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E.M. Zdobnov,
- 24   BUSCO: Assessing genome assembly and annotation completeness with single-copy
- 25   orthologs, Bioinformatics. (2015). <https://doi.org/10.1093/bioinformatics/btv351>.
- 26   [66]. M. Stanke, R. Steinkamp, S. Waack, B. Morgenstern, AUGUSTUS: A web server for
- 27   gene finding in eukaryotes, Nucleic Acids Res. (2004).
- 28   <https://doi.org/10.1093/nar/gkh379>.
- 29   [67]. T.M. Lowe, P.P. Chan, tRNAscan-SE On-line: integrating search and context for
- 30   analysis of transfer RNA genes, Nucleic Acids Res. (2016).
- 31   <https://doi.org/10.1093/nar/gkw413>.
- 32   [68]. Y. Yin, X. Mao, J. Yang, X. Chen, F. Mao, Y. Xu, DbCAN: A web resource for
- 33   automated carbohydrate-active enzyme annotation, Nucleic Acids Res. (2012).
- 34   <https://doi.org/10.1093/nar/gks479>.

- 1 [69]. E. Jourdier, L. Baudry, D. Poggi-Parodi, Y. Vicq, R. Koszul, A. Margeot, M. Marbouty,
2 F. Bidard, Proximity ligation scaffolding and comparison of two *Trichoderma reesei*
3 strains genomes, Biotechnol. Biofuels. (2017). [https://doi.org/10.1186/s13068-017-](https://doi.org/10.1186/s13068-017-0837-6)
4 [0837-6](https://doi.org/10.1186/s13068-017-0837-6).
5 [70]. J.J. Almagro Armenteros, K.D. Tsirigos, C.K. Sønderby, T.N. Petersen, O. Winther, S.
6 Brunak, G. von Heijne, H. Nielsen, SignalP 5.0 improves signal peptide predictions
7 using deep neural networks, Nat. Biotechnol. (2019). [https://doi.org/10.1038/s41587-](https://doi.org/10.1038/s41587-019-0036-z)
8 [019-0036-z](https://doi.org/10.1038/s41587-019-0036-z).

Fig. 1. Phylogenetic analysis showing relation of PJ-1366 ITS region with reference sequences from NCBI GenBank. *T. reesei* (ascomycete) and *U. maydis* (basidiomycete) were used as the outgroups

Fig. 2. Top 10 organism distribution of proteins of PJ-1366, as obtained by BLAST-P analysis

Fig. 3. No. of proteins (x-axis) predicted for top 10 GO terms (y-axis) in (a) molecular function, (b) biological process and (c) cellular component

Fig. 4. Comparison of putative CAZymes of PJ-1366 and TR-RC30, predicted using dbCAN (No. of tools = 3)

Fig. 5. Comparison of extracellular CAZymes of PJ-1366 and TR-RC30

Fig. 6. Classification and distribution of extracellular CAZymes of PJ-1366 (Diagram not to scale). Number of putative genes are given within brackets.

Fig. 1.

Fig. 2.

Fig. 3.

(a)

(b)

(c)

Fig. 4.

Fig. 5.

Fig. 6.

Table 1. Putative secreted cellulases of PJ-1366

Gene ID	Accession	Description	% Identity	CAZy Family
ctg7180000009921.g242	AGW24292.1	cellobiohydrolase 2 [<i>P. oxalicum</i>]	80.622	GH7
ctg7180000014196.g218	CRL19618.1	1, 4- β cellobiohydrolase [<i>P. camemberti</i>]	81.704	GH6
ctg7180000014411.g234	ADX86895.1	cellobiohydrolase II [<i>P. decumbens</i>]	78.891	GH6
ctg7180000015101.g188	AGW24291.1	cellobiohydrolase 1 [<i>P. oxalicum</i>]	83.664	GH7
ctg7180000014196.g220	KAF3385282.1	putative endo- β -1,4-glucanase celB [<i>P. rolfisii</i>]	78.759	GH7
ctg7180000014395.g40	AFG25592.1	endoglucanase Cel5C [<i>P. decumbens</i>]	80.183	GH5
ctg7180000014428.g38	GAQ03870.1	endoglucanase 7a [<i>A. lentulus</i>]	77.755	GH7
ctg7180000014467.g196	CEJ59332.1	Putative Endoglucanase [<i>P. brasilianum</i>]	76.411	GH5
ctg7180000014484.g138	ABY28340.1	endoglucanase II [<i>P. decumbens</i>]	78.832	GH5
ctg7180000015246.g308	ACB06750.1	endo-1,4- β -D-glucanase [<i>P. brasilianum</i>]	82.324	GH5
ctg7180000014229.g247	ABP88968.1	β -glucosidase [<i>P. brasilianum</i>]	91.002	GH3
ctg7180000014236.g84	CEJ60129.1	Putative B-glucosidase (Fragment) [<i>P. brasilianum</i>]	91.388	GH3
ctg7180000014474.g297	ACD86466.1	β -glucosidase [<i>P. decumbens</i>]	88.167	GH3
ctg7180000014493.g241	GCB27413.1	probable β -glucosidase M [<i>A. awamori</i>]	52.747	GH3
ctg7180000014502.g17	XP_025470619.1	β -glucosidase [<i>A. sclerotium</i> CBS 115572]	52.586	GH3
ctg7180000015106.g13	CDM32885.1	Probable β -glucosidase L [<i>P. roqueforti</i> FM164]	79.592	GH3
ctg7180000015125.g119	EPS27960.1	putative β -glucosidase [<i>P. oxalicum</i> 114-2]	81.454	GH3
ctg7180000015232.g357	CEO61051.1	Putative B-1,4-glucosidase [<i>P. brasilianum</i>]	91.409	GH3
ctg7180000015283.g319	GAQ05079.1	probable β -glucosidase G [<i>A. lentulus</i>]	88.821	GH3

Table 2. BLAST-P analysis of secretory PJ-1366 proteins annotated as phenol oxidases

Gene ID	Accession	<i>blastp</i> Hit	Query Cover	% Identity
ctg7180000015091.g77	CEJ58437.1	Putative ferro-O ₂ -oxidoreductase [<i>Penicillium brasilianum</i>]	100%	90.57%
ctg7180000015373.g58	KAF3396287.1	multicopper oxidase abr1 [<i>Penicillium rolsii</i>]	99%	77.20%
ctg7180000015113.g180	KDB16999.1	laccase [<i>Ustilaginoidea virens</i>]	92%	63.59%
jtg7180000014473f_7180000014367f.g266	OKP09932.1	laccase-1 [<i>Penicillium subrubescens</i>]	87%	89.12%
ctg7180000009992.g230	OOQ87800.1	putative laccase TILA [<i>Penicillium brasilianum</i>]	97%	86.47%