Supplementary Materials for:

Genome of the anaerobic gut fungal strain *Orpinomyces* sp. C1A reveals the unique evolutionary history of a remarkable plant biomass degrader

Noha H. Youssef, MB Couger, Christopher G. Struchtemeyer, Audra S. Liggenstoffer, Rolf A. Prade, Fares Z. Najar, Hasan K. Atiyeh, Mark R. Wilkins, and Mostafa S. Elshahed

Table of contents

I. Results.

III. Supplementary figures.

IV. Supplementary tables.

I. Results.

1. Metabolism in strain C1A.

A. Glycolysis and gluconeogenesis. The C1A genome encodes all the glycolytic genes involved in conversion of glucose to pyruvate. On the other hand, 2 of the key enzymes of gluconeogenesis are missing from the genome, fructose-1,6-bisphosphatase and glucose-6-phosphatase.

B. Tricarboxylic acid (TCA) cycle. TCA cycle appears to be incomplete in the C1A genome. The genome only encodes citrate synthase, aconitase, isocitrate dehydrogenase, succinyl-CoA synthase, fumarate reductase, and fumarase genes. No gene encoding α-ketoglutarate dehydrogenase was identified. This is in accordance with prior studies that indicated that the TCA cycle enzymes in anaerobic hydrogenosome-bearing organisms have been re-targeted to the cytosol (1) where the cycle is mainly anapleurotic, providing biosynthetic intermediates (e.g α-ketoglutarate, and succinyl-CoA) for biosynthesis (2).

C. Pentose phosphate pathway (PPP). PPP in strain C1A is incomplete. The C1A genome encodes all the enzymes of the reductive branch of the pathway. This branch is essential for the conversion of 6-carbon sugars to both the 5-carbon sugars ribose and deoxyribose (essential for purine and pyrimidine biosynthesis), and to phosphoribosyl-pyrophosphate (essential for histidine biosynthesis). On the other hand, no enzymes for the oxidative branch of the pathway (glucose-6-phosphate dehydrogenase, 6-phosphogluconolactonase, or 6-phosphogluconate dehydrogenase) were identified in the genome.

D. Catabolism of sugars other than glucose.

i. Xylose. Anaerobic fungi have previously been shown to metabolize xylose via the bacterial route (3, 4). Indeed, the C1A genome encodes the enzymes required for xylose metabolism including xylose isomerase, xylulokinase, and ribulose phosphate 3-epimerase.

ii. Galactose. While a galactokinase is encoded by the genome, no homologue of UDP-glucose:hexose-1-phosphate uridylyltransferase, responsible for the conversion of galactose-1-phosphate to UDP-galactose was identified. This could explain the observed inability of strain C1A to grow on galactose.

iii. Nucleotide-sugars. The genome encodes enzymes essential for the synthesis of UDP-glucose, UDP-galactose, UDP-glucuronic acid, and UDP-N-acetylglucosamine. All the enzymes required for the synthesis of GDP-mannose and GDP-fucose (essential for N-glycan metabolism) from fructose-6-phosphate are also encoded by the genome. These include phosphomannomutase, mannose-6-phosphate isomerase, mannose-1-phosphate guanylyltransferase, GDP-mannose 4,6-dehydratase, and GDP-L-fucose synthase. Consistent with the fact that eukaryotes synthesize their storage polysaccharides (e.g. glycogen) from UDP-glucose, as opposed to ADP-glucose in Bacteria (5, 6), no apparent homologues for enzymes responsible for the biosynthesis of ADP-sugars were identified in the genome. Also, CMP-sugars (e.g. CMP-N-acetylneuraminic acid, CMP-N-glycolylneuraminic acid, and CMP-2-keto-3-deoxynononic acid) are apparently not synthesized by C1A due to the absence of all enzymes required for their synthesis from the genome. CMP-sugars are glycosylation precursors common to higher eukaryotes (e.g. vertebrates) (7).

E. Polysaccharides. The C1A genome encodes a rich repertoire of genes involved in breakdown of cellulose, hemicellulose, starch, chitin, pectin, and peptidoglycan. Strain C1A does not appear to be capable of lignin degradation. Details on the genomic and experimental evaluation of such capabilities are presented in a separate section in the manuscript and supplementary materials.

F. Myo-Inositol Synthesis. Myo-inositol is a polyol that constitutes the structural basis of several secondary messengers in eukaryotes, including inositol phosphates and phosphatidylinositol (8). The genome encodes enzymes responsible for conversion of glucose-6-phosphate to myo-inositol.

G. Amino acid metabolism. The genome encodes all the genes responsible for the biosynthesis of all amino acids from glycolytic or TCA cycle intermediates. In addition, the genome encodes enzymes for β-alanine (an essential precursor in pantothenate biosynthesis) biosynthesis from aspartate. Genes for glutathione biosynthesis from glycine, glutamate, and cysteine are also encoded by the genome. On the other hand, the degradation pathways of several amino acids seem to be incomplete. No homologues of glycine dehydrogenase, proline dehydrogenase, and the methionine salvage pathway enzymes were encoded by the genome. The degradation of branched-chain amino acids is also incomplete due to the absence of homologues of α-keto acid oxidoreductase or dehydrogenase, and enoyl-CoA hydratase. Also, genes for the degradation of aromatic amino acids (phe, tyr, trp, his) are completely missing from the genome. Collectively, these results suggest that strain C1A is incapable of utilizing many amino acids such as glycine, proline, branched chain amino acids, and aromatic amino acids as substrates. Previous studies on amino acid utilization by fungi showed similar results where some but not all amino acids supported growth of several Dikarya fungi when utilized as sole N source (9-12). Variation between species with respect to amino acid utilization was also observed (9, 12).

H. Synthesis of N-linked glycans, O-linked glycans, and glycosphosphatidylinositol (GPI) lipid anchor.

i. N-linked glycans. N-glycosylation is one of the most common post-translational protein modifications in eukaryotic cells (13). N-linked glycans are oligosaccharides linked to asparagine residues of proteins, and are essential for proper protein folding and subsequent protein intracellular trafficking (14). Biosynthesis of the N-glycan oligosaccharide starts in the cytoplasm by transferring the sugar moiety of UDP/GDP-sugar to the lipid carrier dolichylphosphate. This is followed by flipping the oligosaccharide ($Man_5GlcNAc_2$) to the lumenal side of the ER membrane, where seven more sugars (4 mannosyl residues and 3 glucosyl residues) are added to the oligosaccharide unit to give the final structure $Glc_3Man_9GlcNAc_2$. Oligosaccharyltransferase enzyme complex then catalyzes the transfer of the oligosaccharide en-bloc to several asparagine residues of growing polypeptide chains, co-translationally. In the ER, the three glucosyl residues, as well as one mannosyl residue are trimmed. The glycoprotein then folds and is transported to the Golgi apparatus where further mannose trimming occurs via α-mannosidase I, and II to produce the core oligosaccharide Manα1–6(Manα1–3)Manβ1–4GlcNAcβ1–4GlcNAcβ1-Asn. With the exception of the 3 glucosyltransferases that add the 3 glucosyl residues to the N-glycan precursor, strain C1A genome encodes all the genes required for the biosynthesis of $Glc_3Man_9GlcNAc_2$. The genome encodes only α-mannosidase I required for trimming of N-glycan to the structure Manα1–6(Manα1–3)Manα1–6(Manα1–3)Manβ1–4GlcNAcβ1–4GlcNAcβ1-Asn. No homologues of α-mannosidase II (required for trimming of Manα1–6(Manα1–3)Manα1–6(Manα1–3)Manβ1–4GlcNAcβ1–4GlcNAcβ1-Asn to the core structureManα1–6(Manα1–3)Manβ1–4GlcNAcβ1–4GlcNAcβ1-Asn) were detected in C1A genome. Following trimming to the core structure in the Golgi, N-glycan maturation takes place to give rise to one of three structures: 1. Oligomannose N-glycans with only mannosyl residues attached to the core, 2. Hybrid N-glycans with only mannose residues attached to the Manα1–6 arm of the core, and one or two branches with other sugar residues attached to the Manα1–3 arm, and 3. Complex N-glycans in which branches with other sugar residues initiated by N-acetylglucosaminyl transferases (GlcNAcTs) are attached to the core (15, 16). Enzymes for N-glycan maturation identified in the genome include β-1,4-mannosyl-glycoprotein β-1,4-N-acetylglucosaminyltransferase, β-1,4-mannosyltransferase, α-1,2-mannosyltransferase, and glycoprotein α-1,3-L-fucosyltransferase to give rise to a hybrid N-glycan structure as the one shown in Fig S3.

ii. O-linked glycosylation. O-linked glycosylation is another form of post-translational protein modification present in eukaryotes. Mucin-type O-proteoglycans are the most abundant form of O-linked glycosylation in higher eukaryotes (17). Extracellularly, mucin-type O-linked glycoproteins (those could either be secreted or attached to the membrane peripherally) serve as protective barriers to the cell, and mediate communication signals between the cell and its surroundings (17). The fundamental enzyme of mucin-type O-linked glycan biosynthesis is the UDP-N-acetyl-D-galactosamine*:* polypeptide *N*-acetyl-α-galactosaminyltransferase (ppGalNAcT) (17). The C1A genome lacks any apparent homologues of ppGalNAcT and hence is thought to be deficient in mucin-type O-linked glycoproteins. In the 1980s, another form of O-linked glycosylation, the O-linked-β-*N*-acetylglucosamine modification on the hydroxyl group of serine or threonine residues, was discovered (18). This form of O-glycosylation modifies cytosolic and nuclear proteins as opposed to surface or membrane-attached proteins (18), and plays an important role in transcription regulation, signaling, and the cell cycle (18). The O-linked modification of proteins by a single N-acetylglucosamine residue is the result of the concerted action of 2 enzymes; O-linked N-acetylglucosaminyl transferase (OGT), and O-N-acetylglucosaminidase (18, 19). The C1A genome encodes both an OGT, as well as a β-*N*-acetylhexosaminidase, which potentially catalyzes the removal of the O-linked N-acetylglucosaminyl residue.

iii. Glycosphosphatidyl inositol lipid anchor. Several integral membrane proteins are linked to the plasma membrane via a glycosphosphatidylinositol (GPI) lipid anchor (20). This comprises yet another form of protein C-terminus post-translational modification. The GPI lipid core is pre-assembled via several enzymatic reactions that start on the cytosolic side of the ER and are completed in the lumen of the ER. Almost all the enzymes required for the synthesis of GPI lipid anchor from UDP-N-acetylglucosamine, phosphatidyl-inositol, phosphatidyl ethanolamine, and mannose β-phosphodolichol are encoded by the genome. Components of the GPI-N-acetylglucosaminyltransferase and the transamidase complexes are both present. No homologues of phosphatidylinositol class G (ethanolamine phosphotransferase), or phosphatidylinositol class W (acyltransferase) were identified in the genome.

I. Purine and pyrimidine metabolism. The genome encodes all the genes necessary for the biosynthesis of purine and pyrimidine nucleotides.

J. Lipid metabolism.

i. Fatty acids biosynthesis and degradation. Genes required for the synthesis of long chain fatty acids (C16-C18) from acetyl CoA were identified in the C1A genome. These include acetyl-CoA carboxylase/biotin carboxylase, fatty acid synthase fungi type α and β subunits, fatty acid synthase animal type, 3-oxoacyl-ACP reductase, enoyl-ACP reductase II, and palmitoyl-protein thioesterase. In spite of the presence of homologues of mitochondrial trans-2-enoyl CoA reductase and palmitoyl-CoA thioesterase in the genome, mitochondrial (or hydrogenosomal) fatty acid elongation is unlikely to occur in strain C1A due to the lack of homologues of mitochondrial fatty acid elongation enzymes acetyl-CoA acyltransferase, enoyl-CoA hydratase, and long chain 3-hydroxyacyl-CoA dehydrogenase. However, very long chain fatty acids (>C18) could possibly be synthesized from long chain fatty acids in the ER. The genome encodes an acyl-CoA synthetase, an elongation of VLCFA protein 3, β-ketoacyl reductase, 3-hydroxyacyl CoA dehydratase, enoyl-CoA reductase, and a predicted acyl-CoA thioesterase. Strain C1A appears capable of catabolizing fatty acids. The genome harbors a long chain fatty acyl-CoA synthase, acyl-CoA dehydrogenase, multifunctional β-oxidation enzyme, 3-hydroxyacyl-CoA dehydrogenase, and acetyl-CoA acetyltransferase. A homologue of an unsaturated fatty acyl-CoA isomerase (dodecenoyl-CoA isomerase) is also encoded by the genome.

ii. Steroids. Sterols are important components of eukaryotic cell membranes contributing to cell membrane fluidity and participating in phagocytosis. Sterols are synthesized from acetyl-CoA in 3 distinct stages; mevalonate biosynthesis, squalene biosynthesis, and squalene oxidation to lanosterol with the subsequent differentiation to cholesterol in mammals, phytosterol in plants, and ergosterol in fungi (21). Only the last stage of squalene oxidation requires molecular oxygen. The C1A genome encodes all the enzymes for the first 2 stages of sterol biosynthesis. Since the subsequent steps of sterol synthesis require molecular O_2, anaerobic eukaryotes were previously found to produce a homologue of prokaryotic hopane, tetrahymanol, as their cell membrane sterol. Tetrahymanol is a triterpenol that does not require molecular O_2 for biosynthesis (22). We identified squalene-tetrahymanol cyclase in the C1A genome. Thus it appears that *Orpinomyces*, similar to other anaerobic eukaryotes, synthesizes tetrahymanol as its membrane sterol as previously suggested for *Piromyces communis* (22, 23).

iii. Glycerolipids and glycerophospholipids biosynthesis. sn-Glycerol-3-phosphate is the precursor of glycerolipids and glycerophospholipids, both of which are constituents of the cell membrane. sn-Glycerol-3-phosphate can be synthesized from glycerone phosphate (an intermediate in glycolysis) via glycerol-3-phosphate dehydrogenase (NADP), which is encoded in the C1A genome. The genome also encodes all the enzymes required for the biosynthesis of mono-, di-, and triacylglycerol. All the enzymes required for the synthesis of glycerophospholipids precursors, CDP-glycerol, CDP-diacylglycerol, CDP-ethanolamine, and CDP-choline, are encoded by the genome. The genome also encodes enzymes required for synthesis and degradation of phosphatidyl-serine, phosphatidyl-ethanolamine, phosphatidyl-choline, phosphatidyl-inositol, and cardiolipin (the phospholipid known to occur in the inner mitochondrial membrane and is most probably present in strain C1A hydrogenosomal membrane).

K. Cofactors and vitamins. The genome encodes all the enzymes responsible for biosynthesis of riboflavin, pyridoxal phosphate, nicotinate and nicotinamide, pantothenate, coenzyme A, and folic acid. Strain C1A might be auxotrophic for thiamine and biotin: Homologues of thiamine biosynthetic enzymes ThiI (essential for the formation of a thiocarboxylate on the carboxylic acid group of the carboxyl-terminal glycine of the ThiS protein), ThiG (catalyzing with ThiH the conversion of 1-deoxy-D-xylulose 5-phosphate to thiazole), and ThiC (catalyzing the conversion of 5-aminoimidazole ribonucleotide (AIR) into 4-amino-2-methyl-5-hydroxymethylpyrimidine (HMP)) are missing from the genome. Similarly, many of the enzymes for biotin biosynthesis e.g. 6-carboxyhexanoate--CoA ligase, 8-amino-7-oxononanoate synthase, adenosylmethionine-8-amino-7-oxononanoate aminotransferase, dethiobiotin synthetase are also missing.

2. Genetic information processing in strain C1A. We analyzed the genetic information processing machinery in strain C1A, and compared it to what has been observed in eukaryotic genomes in general, and Dikarya genomes in particular. Our goal was to determine whether the basal phylogenetic affiliation, or the unique habitat, has influenced genetic information processing in the *Neocallimastigomycota* through reductive evolution or horizontal gene transfer of prokaryotic genes, similar to what was observed in plant polysaccharide degradation machinery, and other metabolic capabilities. Overall, our results indicate that the C1A genome possesses a thoroughly eukaryotic machinery, with no evidence for absence, or modification of salient genes/pathways when compared to other available eukaryotic genomes.

A. DNA replication, transcription, and translation.

DNA Replication. The C1A genome encodes a complete and thoroughly eukaryotic DNA replication machinery composed of DNA replication initiation factors (pre-replication complex, GINS complex), DNA replication elongation factors (primase complex, polymerase ε and δ complexes, replication factor A, and C), and DNA replication termination factors. Strain C1A clearly possesses all the hallmarks for eukaryotic DNA replication (24). These include (i) the presence of different DNA-dependent DNA polymerases for the leading and lagging DNA strands (25), where the C1A genome encodes all 4 subunits of δ DNA-dependent DNA polymerase for the lagging strand replication, and all 4 subunits of ε DNA-dependent DNA polymerase for the leading strand replication, (ii) the presence of histones around which DNA winds to form nucleosomes, where the genome encodes histones H2A, H2B, H3, H4, as well as the linker histones H1/5, and (iii) the presence of telomeres, where the genome encodes RNA-dependent DNA polymerases (telomerase reverse transcriptase), as well as telomere length regulation protein.

Transcription. Similarly, a thoroughly-eukaryotic transcription machinery is identified in the C1A genome (26, 27): the C1A genome harbors all subunits of RNA polymerases I, II, and III and their associated transcription factors. Further, genes encoding enzymes mediating specific eukaryotic transcription processes, e.g. mRNA 5'end cap addition, mRNA 3'end polyA tail addition, and splicing (28) are also encoded by the C1A genome. In addition to RNA polymerase and transcription factors, the initiation of transcription in eukaryotes requires multi-subunit cofactors including the mediator complex, as well as various chromatin modifying or remodeling co-activator complexes (29). The genome encodes a wealth of such upstream-acting elements.

Translation. A complete set of aminoacyl-tRNA synthetases was detected in the genome. All but two of the large and small ribosomal proteins were identified in the C1A genome, as well as various ribonucleoproteins and enzymes involved in ribosome biogenesis. tRNA processing, splicing, transport, modification, and degradation factors are also encoded by the genome.

B. Cytoskeleton. Eukaryotic cytoskeleton is a dynamic structure that combines the activities of several proteins and organelles to help connect the cell to its surroundings, as well as enable movement and division (30). The eukaryotic cytoskeleton is comprised of three main building blocks: actin filaments (microfilaments), tubulin filaments (microtubules), and intermediate filaments. Both microfilaments and microtubules are associated with cytoskeletal motors; myosins for microfilaments, and kinesins and dyneins for microtubules, which mediate directional movement (30, 31). Intermediate filaments are generally absent in Fungi (32). The C1A genome encodes an extensive cytoskeleton machinery, including microfilaments (actins and actin-binding proteins), microtubules (tubulins), and cytoskeletal motors.

C. Intracellular protein trafficking. Detailed analysis of the complex trafficking pathways clearly indicates that strain C1A possess a complete machinery for intracellular protein trafficking. A general view of these pathways in strain C1A is shown in Figure S4. In eukaryotic cells, protein synthesis occurs in the cytosol. Proteins that are not destined for secretion are synthesized on free ribosomes (Fig S4). These proteins will either remain in the cytosol (cytosolic proteins, Fig S4, route 1), be transported to the nucleus (nuclear proteins, Fig S4, route 2), or could be imported into the hydrogenosome (hydrogenosomal proteins, Fig S4, route 3), or peroxisome (peroxisomal proteins, Fig S4, route 4). The C1A genome encodes a near complete list of the nuclear pore complex components as well as importins for protein import into the nucleus, a near complete hydrogenosomal import machinery, almost all of the peroxins involved in protein transport across the peroxisomal membrane, and peroxisomal integral membrane proteins.

On the other hand, proteins synthesized on membrane-bound ribosomes of the rough endoplasmic reticulum (RER) could either be secreted (secretory pathway, Fig S4, blue arrows), or enter the endosomal system and eventually be degraded in the lysosome (endosomal pathway, Fig S4, green arrows). The later 2 pathways will be further discussed below. The general secretory intracellular protein trafficking apparatus in eukaryotes was uncovered almost 40 years ago (33), where it was deciphered that proteins move between compartments via vesicle formation and fusion (33, 34). Proteins are transported into the ER lumen via the sec61/62/63 complex (Fig S4, route 5) (14). All subunits of the sec61 translocon as well as signal recognition particle components are encoded by the genome. In the ER, proteins are modified with N-glycans and properly-folded proteins are transported to the Golgi via COPII vesicles (Fig S4, route 6) (14). All components of the COPII coatamer are encoded by the genome. Misfolded proteins are subjected to the ER-associated protein degradation (ERAD), and are eventually degraded in the proteasome (14). All the components of the ERAD and proteasomal protein degradation are encoded by the genome. In the Golgi, properly-folded proteins are transported between cisternae via retrograde COPI transport vesicles (Fig S4, route 7) (35). COPI-mediated vesicle transport is also responsible for recycling ER-resident proteins back to the ER (Fig S4, route 8) (35). COPI coatamer components are encoded by the genome. From the Golgi, proteins are transported to the trans Golgi network (TGN) again via COPI transport vesicles (Fig S4, route 9). In the TGN, proteins are sorted either to the plasma membrane (exocytosis) (Fig S4, route 10), or to the endosomal system (Fig S4, route 11)(36). Trafficking from the TGN to the plasma membrane requires at least protein kinase D, protein kinase C η, phosphatidyl-inositol 4-kinase III-β, goodpasture antigen binding protein (START domain), and phosphatidylinositol/phosphatidylcholine transfer protein (sec14/PDR16) (37), all of which are encoded by the genome. Exocytosis itself requires AP-1 complex components, and exocyst subunits (38), homologues of all of which were identified in the genome. On the other hand, endosomal transport of proteins occurs via clathrin-coated vesicles (39). Components of the vesicle coat as well as factors essential for membrane curvature, budding and scission are all encoded by the genome. Transport into the lysosome requires multivesicular body (MVB) sorting (40). Endosomal sorting complex required for transport (ESCRTs) are involved in the sorting of proteins into multivesicular bodies (MVBs), generating the intraluminal vesicles (40). All components of the ESCRT complexes have been identified in the genome. Soluble acid hydrolases destined to the lysosome are modified in Golgi by mannose-6-phosphate residues (by the action of UDP-N-acetylglucosamine:lysosomal enzyme N-acetylglucosamine-1-phosphate transferase) to be recognized by mannose-6-phosphate receptors, followed by packaging in clathrin-coated vesicles and transport to the endosomal system (41). While the genome encodes UDP-N-acetylglucosamine:lysosomal enzyme N-acetylglucosamine-1-phosphate transferase, no homologues of mannose-6-phosphate receptors (cation-dependent or cation-independent) were identified. However, the genome encodes a sortilin Vps10-like homologue, which might be involved in lysosomal enzymes transport to the vacuole (42). Retrograde trafficking components (Vps26, Vps29, Vps35, and sorting nexins 1 and 2) implicated in recycling lysosomal hydrolase receptors from the lysosome to the TGN (Fig S4, route 12) (42) are all encoded by the genome. All vesicular transport events require vesicle tethering, and fusion. These 2 events are mediated by vesicle tethering complexes, and SNAREs, respectively (43, 44). Several SNAREs and vesicle tethering complexes were identified in the genome.

3. Salient differences identified between the C1A genome and those of Dikarya Fungi. Detailed comparison of various cellular processes in the C1A genome to available fungal genomes identified multiple processes, in which strain C1A possesses genes/pathways/machineries that are not usually associated with members of the Dikarya. Rather, many of these are more associated with non-fungal Opisthokonta e.g. Choanoflagellate, *Capsaspora owczarzaki*, and metazoa, as well as other multicellular eukaryotes. In some instances, such genes could also be observed in some or all examined basal fungal genomes, while in other cases, these differences appear to be unique to the C1A genome. Examples include members of intramembrane proteases, post-translational modification enzymes, focal adhesion, axoneme and intraflagellar transport, and protease inhibitors. Details are shown below and are summarized in Table 1, Tables S10-11.

3.1. Intra-membrane proteases. Intra-membrane proteases (IMPs) are integral membrane proteolytic enzymes that function in cleaving trans-membrane domains of other integral membrane proteins. Four families of IMPs are recognized: 1. The aspartyl protease presenelin (the active protease of the γ-secretase complex), 2. The site 2 protease family (S2P) metalloprotease, 3. The signal peptidase family serine protease, and 4. The rhomboid family serine protease (45). Out of the four IMP families, both the signal peptidase family serine protease and the rhomboid serine protease are ubiquitous in fungi (46).

We analyzed the occurrence and distribution of intramembrane proteases in the C1A genome, basal and Dikarya fungal genomes, and Opisthokonta and higher eukaryotic genomes. The analysis revealed salient differences between strain C1A intramembrane protease machinery and Dikarya machineries in both the aspartyl protease presenelin (γ-secretase complex), and the site 2 protease family (S2P) as detailed below.

a. The γ-secretase intra-membrane protease complex is an integral multi-subunit protease complex containing at least 4 proteins: 1. Presenelin (the active site of the enzyme complex), 2. Nicastrin, 3. Anterior pharynx defective-1 protein (APH-1), and 4. Presenelin enhancer-2 (47-49). Nicastrin helps recognize the substrate for proteolytic cleavage that is carried out by presenelin. Presenelin enhancer-2 and APH-1 proteins help stabilize the complex (47). The C1A genome encodes 2 homologues of presenelin, 3 homologues of nicastrin, 1 homologue of APH-1, and 1 homologue of presenelin enhancer-2 (Table 1). Comparative analysis indicated that none of the γ-secretase intra-membrane protease complex components are present in any of the Dikarya genomes examined (74 genomes), and few nicastrin, presenelin, and APH-1 proteins could be identified in basal fungal genomes examined (*Spizellomyces punctuates* DAOM BR117, *Batrachochytrium dendorobatidis* JEL423, and *Allomyces macrogynus* ATCC38327, *Rhizopus oryzae* RA 99-880). We further compared C1A γ-secretase proteins to the genomes of 2 Opisthokonts; *Monosiga brevicollis* (a choanoflagellate thought to represent the closest relative to Metazoa (50)), and *Capsaspora owczarzaki* (representing an independent unicellular lineage closely related to animals and choanoflagellates (51)). Multiple C1A nicastrin (2 of the 3 homologues), APH-1, and presenelin were identified in both genomes, while Pen-2 was only observed in the genome of *Capsaspora owczarzaki* (Table 1). An examination of the Merops (52) and Interpro (53) databases shows that in general components of the γ-secretase complex are represented in Kingdom Animalia (mainly Chordata, Arthropoda, and Nematoda), Kingdom Plantae, and few representatives in Amoebozoa, Euglenozoa, and Archaea.

b. The Site 2 protease (S2P) family of intramembrane metalloprotease belongs to the M50 family of proteases (52). The role of S2P proteases are mainly involved in cleaving different membrane-tethered transcription factors, e.g. sterol regulatory element binding protein (54). According to the Merops database, two M50 subfamilies were characterized; M50A and M50B (52). Members of subfamily M50A are mainly represented in kingdom Animalia, specifically phylum Chordata, with few representatives in Kingdom Plantae, and Kingdom Protozoa, and some representatives in Archaea. M50B subfamily members on the other hand are mainly represented in Archaea, Bacteria, with very few examples in Animalia (only 1 example in Chordata), Fungi, Plantae, and Protozoa (55). The C1A genome encodes 1 gene annotated as S2P protease belonging to subfamily M50A that has not been previously identified in Fungi (Table 1). Indeed, analysis of Dikarya fungal genomes failed to identify any members of the M50A protease, while the basal fungal genome *Spizellomyces punctuates* DAOM BR117 encodes an M50 protease. On the other hand, analysis of *Monosiga brevicollis* and *Capsaspora owczarzaki* resulted in the identification of subfamily M50A proteases, similar to C1A homologues. M50A homologues were thus identified in C1A, *Spizellomyces punctuates* DAOM BR117, *Monosiga brevicollis* and *Capsaspora owczarzaki,* but not in Dikarya fungal genomes.

3.2. Post-translational protein modification. Post-translational modifications of eukaryotic proteins include phosphorylation (ser, thr, tyr), acylation (acetylation of lys, N-myristoylation of gly, S-palmitoylation of cys, mono-, and polyubiquitination of lys), alkylation (methylation of lys and arg, prenylation of cys), and glycosylation (N-glycosylation of asn, O-glycosylation of ser) (13). Post-translational modifications of proteins are essential for their function, folding, and/or intracellular fate (13). As detailed above, proteins destined for secretion are co-translationally N-glycosylated in the ER, followed by trimming of the N-glycan oligosaccharidyl residues to the core oligosaccharide in the Golgi. In the Golgi, maturation from the core oligosaccharide takes place to give rise to one of three N-glycan structures; oligomannose, hybrid, or complex N-glycans. Strain C1A appears to be capable of post-translational protein fucosylation, a type of N-glycosylation that occurs on the first N-acetylglucosaminyl residue of the N-glycan oligosaccharide linked to Asn residues of the modified protein. Fucosylation is a post-translational protein modification that has not been previously observed within the Mycota. The genome encodes glycoprotein α-1,3-L-fucosyltransferase activity. This enzyme is known to catalyze the transfer of a fucosyl moiety from GDP-fucose (the C1A genome encodes all the genes required for its biosynthesis) to the first N-acetylglucosaminyl residue of N-glycan to form an α-1,3 linkage. While fucosylation of N-acetylglucosamine residues of N-glycans is known to occur in vertebrates, invertebrates, and plants, the α-1,3 linkage is only known to occur in insects, plants, and invertebrates (e.g., *Schistosoma*, *Haemonchus*, *Lymnaea*) (16). Indeed, the closest relative to C1A fucosyltransferases was *Nicotiana tabacum*. Comparsion to other genomes identified similar fucosyltransferases in *Rhizopus oryzae* RA 99-880. In all Dikarya genomes currently available in Genbank database, only one gene annotated as "hypothetical protein" in *Puccinia graminis* (EFP84060) with low sequence similarity to plant fucosyltransferases has been identified. No similar fucosyltransferases to that observed in C1A were identified in *Monosiga brevicollis* and *Capsaspora owczarzaki* Opisthokonts genomes.

3.3. Focal adhesions. In filamentous fungi within the Dikarya, surface attachment and association with the extracellular matrix (ECM) is known to occur via extracellular "glue" secreted by conidial structures (e.g. adhesin) (56). The C1A genome encodes a putative cell wall adhesin similar to that of *Candida albicans*. On the other hand, focal adhesions (FA) are large multiprotein intracellular assemblies that mediate interaction between the ECM and the cell interior by promoting cell anchorage and mechanical adhesion to the ECM as well as act as signaling milieu where signaling proteins are concentrated at sites of integrin binding and connect the cell's cytoskeleton to the ECM (57). FA appears to be absent from filamentous fungi, and more common in other eukaryotes (Amoebozoa, Metazoa (sponges, placozoans, and cnidarians), and Holozoa (58).

During FA, integrin (α, β) binds to extracellular proteins of ECM via its extracellular domains. Within the cell, integrin connect to the cytoskeleton via adaptor proteins (e.g. α-actinin, talin, paxillin, vinculin) that bind to integrin intracellular domains (57). This integrin-adaptor-ECM protein complex forms the basis of FA with which other proteins, particularly those involved in intracellular signaling, associate. These include the heterotrimer IPP complex (Integrin-linked kinase, PINCH, and α-parvin), and the 2 kinases, FAK (focal adhesion kinase), and c-Src tyrosine kinases known to accumulate at the FA site (58). Analysis of the C1A genome indicates that it encodes multiple integrin-adaptor proteins (e.g. Talin, vinculin, α-parvin, α-actinin, Band 4.1, paxillin), and several FA-associated signaling proteins (e.g. integrin-linked kinase, Nck, Nck-associated protein 1), as well as downstream signaling protein (e.g. ROCK) (Table 1). However, no homologues of integrins, FAK, and c-Src kinase were identified in the C1A genome. This is similar to what was observed in prior studies (58), where it was suggested that basal fungi seemed to have kept some or all of the IPP complex components, and the scaffolding proteins, but lost the integrins and FAK and c-Src kinase during the course of their evolution from their unicellular Opisthokont ancestor. The same study (58) identified all components of FA in the Opisthokont *Capsaspora owczarzaki*, but not *Monosiga brevicollis,* and suggested that the focal adhesion machinery originated before the divergence of *C. owczarzaki* from choanoflagellates+Metazoa, but likely after the Fungi+nucleariid+fonticulid clade had split from Holozoa. Our analysis of the C1A genome fits well with that conclusion (58). Strain C1A seemed to have lost integrins, FAK, and src-kinase, but kept all IPP components as well as the scaffolding proteins. The exact cellular function of FA proteins in the absence of integrins, FAK, and src-kinase in strain C1A is unclear. The process occurring intracellularly during FA and its connection to the actin cytoskeleton is reviewed in (59). Phosphatidylinositol-4,5-bisphosphate (PIP2) in the membrane stimulates the accumulation of FA proteins, and activates the actin filament cross-linker, α-actinin. Following their accumulation, FA proteins interact with the actin cytoskeleton to increase actin polymerization, promote actomyosin assembly, and stabilize the actin filament. Downstream effects on actin cytoskeleton are mediated by Rho kinase. Activation of Rho kinase through the fine-tuning of Rho-GEF (stimulatory) and Rho-GAP (inhibitory) activities leads to (1) activation of Diaphanous (GDIA) that in turn activates profilin causing polymerization of G-actin, (2) activation of ROCK (Rho-associated protein kinase) that has an inhibitory effect on myosin light chain phosphatase (MLCP) keeping myosin light chain (MLC) in the phosphorylated (P) active form (via the action of myosin light chain kinase (MLCK)) to interact with actin, and (3) a stimulatory effect on LIM kinase (LIMK) causing it to phosphorylate and inhibit cofilin, thereby inhibiting its actin-depolymerizing activity.

3.4. Evidence for the presence of an axoneme in the C1A genome. Ciliated and flagellated eukaryotic cells are known to possess a microtubule-based cytoskeletal structure called the axoneme that acts as a scaffold for other protein complexes including motor proteins (e.g kinesin, and dynein) that are essential for intraflagellar transport (IFT) of proteins (60). Neocallimastigomycota, and other basal fungi produce flagellated zoospores. The genome shows evidence for the presence of an axoneme and IFT in *Orpinomyces* strain C1A. Table S10 provides a summary of proteins in the C1A genome that are thought to be associated with the axoneme or involved in IFT. The genome encodes both axonemal structural components including axonemal dynein (heavy, intermediate, and light chains), radial spokes, ribbon, nexin-dynein regulatory complex components, and outer dynein arm docking complex components. When compared to other fungal genomes, all axonemal structural components of the C1A genome were absent from Dikarya Fungal genomes. Several basal body assembly proteins (61, 62) including B9 domain, meckelin, FOP dimerization domain, and Cep290 were also identified in the genome. Due to the absence of ribosomes in the axoneme, proteins are encoded in the cytosol then transported to the axoneme via IFT particles composed of IFT complexes A and B with their accessory components as well as the Bardet Biedl syndrome complex (BBSome) components (63). Homologues of all IFT and BBSome components were identified in the C1A genome (Table S10). When compared to other fungal genomes, IFT complexes A and B homologues were identified in basal but not Dikarya fungal genomes, while BBSome complex components were not identified in any other fungal genome (Table S10).

3.5. *Protease inhibitors.* The C1A genome also encodes several protease inhibitors that are not encountered in other fungal genomes. Table S11 lists protease inhibitors in the C1A genome identified using the MEROPS database (52). Strain C1A protease inhibitors belonged to MEROPS families I4, I32, and I42.

I4 Proetease inhibitors are known serine protease inhibitors (serpins). Serpins of family I4 are present mostly within eukaryotic metazoan phyla (Arthropoda, Chordata, and Nematoda), and have also been identified in Bacteria and Archaea. While abundant in the C1A genome, I4 proteases have not previously been encountered in Fungi. Indeed, no I4 serine proteases were identified in either basal or Dikarya fungi, or in the genomes of *Capsaspora owczarzaki*, and *Monosiga brevicollis* genomes.

It has been previously demonstrated that bacterial commensals of the mammalian gastrointestinal tract produce serpins as defense mechanism against proteolysis by the host elastases (64). A similar function could be attributed to C1A serpins where its habitat in the cow rumen exposes it to a highly proteolytic environment (65). Dockerin or chitin-binding domains were present in some of C1A serpins as previously documented in *Piromyces* sp. E2 (66), further emphasizing serpins suggested extracellular role of protection against proteases. C1A serpins were closely related to *Piromyces* sp. E2 serpins.

Cysteine protease inhibitors of family I42 have previously been identified within the Bacteria and Archaea. In eukaryotes, they appear to be restricted to Protozoa (phyla Apicomplexa, and Sarcomastigophora). Comparison of cysteine protease inhibitors belonging to family I42 in the C1A genome against other fungal genomes identified homologues only in the basal fungal genome of *Allomyces macrogynus.* No homologues of family I42 were identified in Dikarya fungal genomes or in the 2 Opisthokont genomes of *Capsaspora owczarzaki*, and *Monosiga brevicollis*.

Cysteine protease inhibitors of family I32 are eukaryotic proteins that are present within the Metazoa (phyla Arthropoda, Chordata, Nematoda), Fungi, with few representatives in Protozoa, and no representatives in plants. C1A I32 homologues were identified in both basal (*Spizellomyces punctatus* DAOM BR117), as well as most Dikarya Fungal genomes. No homologues of family I32 were identified in the 2 Opisthokont genomes *Capsaspora owczarzaki*, and *Monosiga brevicollis*. It is interesting to note that lysates of rumen protozoa were shown to possess cysteine protease (caspase) activity (67). C1A cysteine protease inhibitors (I32, and I34) could potentially be aimed at protozoal caspases released into the rumen following protozoal cell lysis.

4. Hydrogenosomal structure and function in strain C1A.

<u>4.1. Protein import system.</u> Anaerobic fungi, as well as anaerobic ciliates and flagellates are known to possess double-membrane hydrogenosomes whose main function is ATP production via substrate level phosphorylation and hydrogen production (68-70). We examined the C1A genome to identify genes involved in protein, solutes, and ATP/ADP hydrogenosomal transport, as well as various biosynthetic and metabolic processes occurring within the hydrogenosome (Figure 3). We compared the distribution, function, and phylogenetic affiliation of these identified genes, with those observed in fungal genomes, as well as those observed in the hydrogenosomes of model microeukaryotes e.g. *Trichomonas vaginalis*.

The general eukaryotic mitochondrial protein import system encountered in aerobic fungi and the majority of eukaryotes (71) is composed of: 1. An outer mitochondrial membrane translocase TOM complex (at least 7 proteins), 2. Sorting and assembly machinery of the outer membrane SAM (5 proteins), 3. Mitochondrial intermembrane space import and assembly MIA (3 proteins), 4. Intermembrane space chaperones small TIMs (4 proteins), 5. Translocase of the inner mitochondrial membrane TIM22 (at least 6 proteins), 6. Presequence translocase of the inner mitochondrial membrane TIM23 (4 proteins) and several presequence translocase-associated motor proteins (7 proteins) (72, 73). The C1A genome encodes a near-complete hydrogenosomal protein import system with components of the TOM outer membrane transport system (4 out of 7 genes), the SAM sorting and assembly complex (4 out of 5 genes) for protein insertion in the outer membrane, the MIA intermembrane space import and assembly complex (2 out of 3 genes), small TIMs (2 out of 4 genes), the TIM22 complex for protein insertion in the inner membrane (6 out of 6 genes), and the inner membrane transport system and associated motor (TIM23 complex, 10 out of 11 genes) (Figure 3 and Table S12).

Secondary losses of small TIMs and the TIM22 complex components, as well as simplification of the TOM, SAM, and TIM23 complexes were observed in both hydrogenosome-harboring (e.g. the Parabasalia *Trichomonas vaginalis*), and mitosome-harboring (e.g. the microsporidian *Encephalitozoon cuniculi*, and the amoebozoan *Entamoeba histolytica*) organisms. For example, the hydrogenosomal import machinery of *Trichomonas vaginalis* has been reduced to a few outer membrane proteins (Tom40, Sam50, Hmp35, and Hmp36), a few inner membrane proteins (Tim17/22/23, Tim 44, and PAM16, 18), and one highly modified intermembrane small TIMs (74). The reduced hydrogenosomal import machinery in *Trichomonas vaginalis* is thought to be a specific adaptation to life under anaerobic conditions (74). Such a drastic reduction was not observed in the C1A genome in spite of the anaerobic lifestyle. Further, examining the phylogenetic affiliation of hydrogenosomal import proteins in strain C1A clearly demonstrates their fungal origin (Table S12). This near complete machinery and fungal affiliation of protein import system in C1A hydrogenosome could be explained by the relatively recent adaptation of Neocallimastigomycota to anaerobic gut habitats (75), compared to earlier adaptation dates of Parabasilia and Amoebozoa.

4.2. Identification of hydrogenosomal-bound proteins. Using the approach outlined in supplementary methods, we bioinformatically identified 46 hydrogenosomal proteins in the C1A genome. Table S13 shows the predicted C1A hydrogenosomal proteins. Proteins predicted to be imported into the hydrogenosomal matrix included several hypothetical proteins, Fe-S cluster assembly and maturation proteins, pyruvate metabolism and energy production enzymes, peptidases, as well as intralumenal chaperones and co-chaperones (Table S13).

4.3. Hydrogenosomal Fe-S cluster assembly. Eukaryotic Fe-S clusters are cofactors catalyzing a plethora of biological reactions in the mitochondria, cytosol, as well as the nucleus (76). Mitochondria, and mitochondria-like organelles, e.g. hydrogenosomes (70) are essential for Fe/S protein biogenesis. The C1A genome encodes genes involved in Fe-S cluster assembly, and maturation. Table S14 summarizes C1A proteins involved in Fe/S cluster assembly, export, and cytosolic Fe/S protein maturation. Biogenesis of mitochondrial Fe-S clusters starts by the import of ferrous iron via the inner membrane carriers Mrs3/4 (76, 77). Fe/S clusters are synthesized by the early components of the ISC machinery including cysteine desulfurase, frataxin, and ferredoxin, and are transiently bound by the scaffold protein IscU. The Fe/S cluster is then transferred from IscU to the recipient apoprotein via the late components of the ISC machinery including the chaperone Ssq1, the cochaperone Jac1, the nuceotide exchange factor Mge1, in presence of monothiol glutaredoxin to form the holoprotein. The accessory proteins Isa1 and Isa2 are specifically required for the activation of mitochondrial aconitase. Extramitochondrial assembly of Fe/S proteins require the mitochondrial ISC machinery, the mitochondrial ISC export machinery, and the cytosolic CIA machinery (76-78). First, an unknown compound is transferred via the ABC transporter Atm1 to the cytosol to aid in Fe/S protein maturation. Glutathione and the sulfhydryl oxidase Erv1 of the intermembrane space participate in the export. In the cytosol, the Fe/S clusters are assembled on the scaffold proteins Cfd1, and Nbp35. The Fe/S clusters are then transferred to the cytosolic apoproteins via Nar1 and Cia1.

5. Transporters in the C1A genome.

The genome encodes 339 transporter genes belonging to 38 different transporter families. Table S15-16 summarizes the results of the C1A genome comparison to the transporter classification database (79). The majority of transporters (36% of the total transporters) are secondary transporters. These include sodium symporters (amino acid: Na^+, organic anion: Na^+, and inorganic anion: Na^+), sodium antiporters (Ca^{2+}: Na^+, and drug :Na^+), proton symporters (amino acid: proton, sugar: proton, polyol: proton, nucleoside: proton, and peptide: proton), proton antiporters (Ca^{2+}: proton, heavy metal: proton, drug: proton, Fe: proton, K^+: proton, Na^+: proton, and Cl^-: proton), and antiporters for nucleotide-sugar: nucleotide, ADP:ATP, bicarbonate:Cl^-, and oxoglutarare:malate. The ATP-dependent transporters constitute 32% of the total transporters encoded by the C1A genome. These include: 1. Uptake transporters for sugars, polar amino acids, iron-chelate, and siderophore-Fe^{3+}, 2. Exporters for drugs, drug-conjugates, heavy metal, and acetate, 3. Ca^{2+} transporters in the lysosome, Golgi, and ER, and 4. Phospholipid transporters including Golgi phosphatidylserine, phosphatidylethanolamine translocase, phospholipid uptake flippase, inwardly-directed phospholipid and lyso- phospholipid flippase, and plasma membrane inwardly-directed phosphatidylcholine, phosphatidylserine, lyso-phosphatidylethanolamine flippase, and ATPase pumps (V-type). Finally, the genome encodes ion channels and pores constituting 31% of the total transporters including voltage-gated ion channels for Na^+, Ca^{2+}, and K^+, transient receptor potential Ca^{2+} channels, water transporters, ammonia transporters, mechanosensitive ion channels, Al^{3+}, and Mn^{2+} resistance transporters, as well as voltage-dependent anion channel for Cl^-. This distribution of transporters in the C1A genome with almost one third of transporters belonging to one of the three major groups of transporter families (ATP-dependent, ion channels, and secondary transporters), is in contrast to other fungal genomes that seem to have 70-80% of their transporters belonging to the secondary group of transporters, with only 20-30% belonging to the other 2 groups (http://www.membranetransport.org/).

C1A sugar transporters belong to three major families; the major facilitator superfamily (MFS), the glycoside-pentoside-hexuronide:cation symporter family (GPH), and the ATP-binding cassette superfamily (ABC). C1A MFS transporters import monosaccharides, while the GPH, and the ABC transporters import di- and oligosaccharides (Table S16). As shown, C1A has more transporters for monosaccharides compared to di- or oligosaccharides (33 compared to 13 transporters). Most of the monosaccharide transporters were similar to fungal transporters (19 out of 33), while only a few (9 out of 33) were similar to bacterial transporters. On the other hand, almost all the di- or oligosaccharide transporters were similar to bacterial transporters (12 out of 13).

6. Peptidases in the C1A genome.

The genome encodes proteases belonging to 53 families (MEROPS classification (52)) of aspartyl (19 families), cysteinyl (18 families), metallo- (19 families), seryl (11 families), and threonyl (3 families) proteases. Table S17 summarizes the results of MEROPS database comparison. The distribution of different protease families in the C1A genome is similar to other fungal genomes. Most of the proteases/peptidases are intracellular (Table S17), with the exception of metalloproteases belonging to families M23 (lyse bacterial cell walls), M28, M35, and M36.

7. Identification and analysis of genes containing CBM modules in strain C1A.

Carbohydrate-binding modules (CBM) are non-catalytic domains that are often encountered in lignocellulolytic enzymes and promote the association of the enzyme with the substrate. A total of 103 genes harboring carbohydrate-binding modules (CBM) domains belonging to 6 different CBM families were identified (Table S22). The majority (75.7%) of CBMs were members of the exclusively fungal CBM1 domain. Twenty-six genes with CBM domains were associated with GH enzymes, and 7 were associated with PL enzymes (Table S22). Within GH-CBM dual domain genes, CBM1 domains were associated with several GH10 and GH11 xylanases, CBM18 with GH18 chitinases, and CBM48 with GH13 amylases. No CBM domains were identified in GH genes putatively involved in cellulose metabolism in the C1A genome. No CBM2 or CBM3 domains, the prevalent CBM modules in bacterial plant biomass-degradation genes and in rumen anaerobic cellulosomal bacteria, respectively, were identified in the C1A genome (Table S22).

II. References:

1. **Akhmanova A, Voncken FGJ, Harhangi H, Hosea KM, Vogels GD, and Hackstein JHP**. 1998. Cytosolic enzymes with a mitochondrial ancestry from the anaerobic chytrid *Piromyces* sp. E2. Mol. Microbiol. **30**: 1017-1027.
2. **Boxma B, Voncken F, Jannink S, Van Alen T, Akhmanova A, Van Weelden SWH, Van Hellemond JJ, Ricard G, Huynen M, Tielens AGM, and Hackstein JHP**. 2004. The anaerobic chytridiomycete fungus *Piromyces* sp. E2 produces ethanol via pyruvate:formate lyase and an alcohol dehydrogenase E. Mol. Microbiol. **51**: 1389-1399.
3. **Harhangi HR, Akhmanova AS, Emmens R, van der Drift C, de Laat WTAM, van Dijken JP, Jetten MSM, Pronk JT, and Op den Camp HJM**. 2003. Xylose metabolism in the anaerobic fungus *Piromyces* sp. strain E2 follows the bacterial pathway. Arch. Microbiol. **180**: 134-141.
4. **Madhavan A, Tamalampudi S, Ushida K, Kanai D, Katahira S, Srivastava A, Fukuda H, Bisaria V, and Kondo A**. 2009. Xylose isomerase from polycentric fungus *Orpinomyces* gene sequencing, cloning, and expression in *Saccharomyces cerevisia*e for bioconversion of xylose to ethanol. Appl. Microbiol. Biotechnol. **82**: 1067-1078.
5. **Ball SG and Morell MK**. 2003. From bacterial glycogen to starch: Understanding the biogenesis of the plant starch granule. Ann. Rev. Plant Biol. **54**: 207-233.
6. **Ball S, Colleoni C, Cenci U, Raj JN, and Tirtiaux C**. 2011. The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. J. Exp. Bot. **62**: 1775-1801.
7. **Freeze HH and Elbein AD**. 2009. Chapter 4. Glycosylation precursors. *In* Ajit Varki, Richard D Cummings, Jeffrey D Esko, Hudson H Freeze, Pamela Stanley, Carolyn R Bertozzi, Gerald W Hart, and Etzler M E (ed), Essentials of glycobiology, 2nd. ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
8. **Shears SB**. 1998. The versatility of inositol phosphates as cellular signals. Biochim. Biophys. Acta. **1436**: 49-67.
9. **Abuzinadah RA and Read DJ**. 1988. Amino acids as nitrogen sources for ectomycorrhizal fungi: Utilization of individual amino acids. Trans. Brit. Mycol. Soc. **91**: 473-479.
10. **Oso BA**. 1975. Utilization of amino acids as sole nitrogen sources by thermophilic fungi. Zeitschrift für allgemeine Mikrobiologie. **15**: 39-43.
11. **Jennison MW and Perritt AM**. 1960. Physiology of wood-rotting Basidiomycetes. III. Studies on the utilization of optical isomers of amino acids. Mycologia. **52**: 628-635.
12. **Whittaker SP and Cairney JWG**. 2001. Influence of amino acids on biomass production by ericoid mycorrhizal endophytes from *Woollsia pungens* (Epacridaceae). Mycol. Res. **105**: 105-111.

13. **Walsh CT, Garneau-Tsodikova S, and Gatto GJ**. 2005. Protein posttranslational modifications: The chemistry of proteome diversifications. Angewandte Chemie International Edition. **44**: 7342-7372.

14. **Hebert DN and Molinari M**. 2007. In and out of the ER: protein folding, quality control, degradation, and related human diseases. Physiol. Rev. **87**: 1377-1408.

15. **Helenius A, Aebi, and Markus**. 2001. Intracellular functions of N-Linked glycans. Science. **291**: 2364-2369.

16. **Stanley P, Schachter H, and Taniguchi N**. 2009. Chapter 8. N-glycans. *In* Ajit Varki, Richard D Cummings, Jeffrey D Esko, Hudson H Freeze, Pamela Stanley, Carolyn R Bertozzi, Gerald W Hart, and Etzler M E (ed), Essentials of glycobiology, 2nd. ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

17. **Hang HC and Bertozzi CR**. 2005. The chemistry and biology of mucin-type O-linked glycosylation. Bioorg. Med. Chem. **13**: 5021-5034.

18. **Butkinaree C, Park K, and Hart GW**. 2010. O-linked β-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. Biochim. Biophys. Acta. **1800**: 96-106.

19. **Zeidan Q and Hart GW**. 2010. The intersections between O-GlcNAcylation and phosphorylation: implications for multiple signaling pathways. J. Cell Sci. **123**: 13-22.

20. **Orlean P and Menon AK**. 2007. Thematic review series: Lipid Posttranslational Modifications. GPI anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and love glycophospholipids. J. Lipid Res. **48**: 993-1011.

21. **Espenshade PJ and Hughes AL**. 2007. Regulation of sterol synthesis in eukaryotes. Ann. Rev. Genet. **41**: 401-427.

22. **Takishita K, Chikaraishi Y, Leger M, Kim E, Yabuki A, Ohkouchi N, and Roger A**. 2012. Lateral transfer of tetrahymanol-synthesizing genes has allowed multiple diverse eukaryote lineages to independently adapt to environments without oxygen. Biol. Direct. **7**: 5.

23. **Kemp P, Lander DJ, and Orpin CG**. 1984. The lipids of the rumen fungus *Piromonas communis*. J. Gen. Microbiol. **130**: 27-37.

24. **Chagin VO, Stear JH, and Cardoso MC**. 2010. Organization of DNA replication. Cold Spring Harb. Perspect. Biol. **2**: a000737.

25. **Miyabe I, Kunkel TA, and Carr AM**. 2011. The major roles of DNA polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved. PLoS Genet. **7**: e1002407.

26. **Roeder RG**. 2003. The eukaryotic transcriptional machinery: complexities and mechanisms unforeseen. Nat Med. **9**: 1239-1244.

27. **Woychik NA**. 2001. RNA polymerases and the eukaryotic transcription machinery. eLS. **DOI**: 10.1038/npg.els.0005059.

28. **Kapp LD and Lorsch JR**. 2004. The molecular mechanics of eukaryotic translation. Ann. Rev. Biochem. **73**: 657-704.

29. **Tsonis PA**. 2003. Transcription in Eukaryotes, p. 104-152. (ed), Anatomy of gene regulation. A Three-dimensional Structural Analysis. Cambridge University Press, Cambridge.

30. **Fletcher DA and Mullins RD**. 2010. Cell mechanics and the cytoskeleton. Nature. **463**: 485-492.

31. **Wickstead B and Gull K**. 2011. The evolution of the cytoskeleton. J. Cell Biol. **194**: 513-525.

32. **Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, and Darnell J**. 2000. Chapter 19, Cell motility and shape II: microtubules and intermediate filaments., (ed), Molecular cell biology, 4th ed, W. H. Freeman, New York. http://www.ncbi.nlm.nih.gov/books/NBK21646/.

33. **Palade G**. 1975. Intracellular aspects of the process of protein synthesis. Science. **189**: 347-358.

34. **Rothman JE**. 1994. Mechanisms of intracellular protein transport. Nature. **372**: 55-55.

35. **Hsu VW, Lee SY, and Yang J-S**. 2009. The evolving understanding of COPI vesicle formation. Nat. Rev. Mol. Cell. Biol. **10**: 360-364.

36. **Gu F, Crump CM, and Thomas G**. 2001. Trans-Golgi network sorting. Cell. Mol. Life Sci. **58**: 1067-1084.

37. **Malhotra V and Campelo F**. 2011. PKD regulates membrane fission to generate TGN to cell surface transport carriers. Cold Spring Harb. Perspect. Biol. **3**.

38. **Munson M and Novick P**. 2006. The exocyst defrocked, a framework of rods revealed. Nat. Struct. Mol. Biol. **13**: 577-581.

39. **Doherty GJ and McMahon HT**. 2009. Mechanisms of endocytosis. Ann. Rev. Biochem. **78**: 857-902.

40. **Fader CM and Colombo MI**. 2008. Autophagy and multivesicular bodies: two closely related partners. Cell Death Differ. **16**: 70-78.

41. **Ghosh P, Dahms NM, and Kornfeld S**. 2003. Mannose 6-phosphate receptors: new twists in the tale. Nat. Rev. Mol. Cell. Biol. **4**: 202-213.

42. **Saftig P and Klumperman J**. 2009. Lysosome biogenesis and lysosomal membrane proteins: trafficking meets function. Nat. Rev. Mol. Cell. Biol. **10**: 623-635.

43. **Whyte JRC and Munro S**. 2002. Vesicle tethering complexes in membrane traffic. J. Cell Sci. **115**: 2627-2637.

44. **Schmitt HD and Jahn R**. 2009. A Tethering Complex Recruits SNAREs and Grabs Vesicles. Cell. **139**: 1053-1055.

45. **Ha Y**. 2007. Structural principles of intramembrane proteases. Curr. Opin. Struct. Biol. **17**: 405-411.

46. **Freeman M**. 2008. Rhomboid proteases and their biological functions. Ann. Rev. Genet. **42**: 191-210.

47. **Wolfe MS**. 2006. The γ-secretase complex: a membrane-embedded proteolytic ensemble. Biochemistry. **45**: 7931-7939.

48. **Iwatsubo T**. 2004. The Œ≥-secretase complex: machinery for intramembrane proteolysis. Current Opinion in Neurobiology. **14**: 379-383.

49. **Small D, Klaver D, and Foa L**. 2010. Presenilins and the gamma-secretase: still a complex problem. Mol. Brain. **3**: 7.

50. **King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, Marr M, Pincus D, Putnam N, Rokas A, Wright KJ, Zuzow R, Dirks W, Good M, Goodstein D, Lemons D, Li W, Lyons JB, Morris A, Nichols S, Richter DJ, Salamov A, Sequencing JGI, Bork P, Lim WA, Manning G, Miller WT, McGinnis W, Shapiro H, Tjian R, Grigoriev IV, and Rokhsar D**. 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature. **451**: 783-788.

51. **Ruiz-Trillo I, Lane CE, Archibald JM, and Roger AJ**. 2006. Insights into the evolutionary origin and genome architecture of the unicellular Opisthokonts *Capsaspora owczarzaki* and *Sphaeroforma arctica*. J. Euk. Microbiol. **53**: 379-384.

52. **Rawlings ND, Barrett AJ, and Bateman A**. 2012. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. **40**: D343-D350.

53. **Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, de Castro E, Coggill P, Corbett M, Das U, Daugherty L, Duquenne L, Finn RD, Fraser M, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, McMenamin C, Mi H, Mutowo-Muellenet P, Mulder N, Natale D, Orengo C, Pesseat S, Punta M, Quinn AF, Rivoire C, Sangrador-Vegas A, Selengut JD, Sigrist CJA, Scheremetjew M, Tate J, Thimmajanarthanan M, Thomas PD, Wu CH, Yeats C, and Yong S-Y**. 2012. InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res. **40**: D306-D312.

54. **Kinch LN, Ginalski K, and Grishin NV**. 2006. Site-2 protease regulated intramembrane proteolysis: Sequence homologs suggest an ancient signaling cascade. Protein Sci. **15**: 84-93.

55. **Rawlings ND, Morton FR, Kok CY, Kong J, and Barrett AJ**. 2008. MEROPS: the peptidase database. Nucleic Acids Research. **36**: D320-D325.

56. **Epstein L and Nicholson RL**. 2006. Adhesion and adhesives of Fungi and Oomycetes, p. 41-62. *In* Smith A M and Callow J A (ed), Biological Adhesives. Springer-Verlag, Berlin, Heidelberg.

57. **Wozniak MA, Modzelewska K, Kwong L, and Keely PJ**. 2004. Focal adhesion regulation of cell behavior. Biochim. Biophys. Acta. **1692**: 103-119.

58. **Sebé-Pedrós A, Roger AJ, Lang FB, King N, and Ruiz-Trillo I**. 2010. Ancient origin of the integrin-mediated adhesion and signaling machinery. Proc. Natl. Acad. Sci. **107**: 10142-10147.

59. **Parsons JT, Horwitz AR, and Schwartz MA**. 2010. Cell adhesion: integrating cytoskeletal dynamics and cellular tension. Nat. Rev. Mol. Cell. Biol. **11**: 633-643.

60. **Mitchison TJ and Mitchison HM**. 2010. Cell biology: How cilia beat. Nature. **463**: 308-309.

61. **Fritz-Laylin LK and Cande WZ**. 2010. Ancestral centriole and flagella proteins identified by analysis of *Naegleria* differentiation. J. Cell Sci. **123**: 4024-4031.

62. **Kobayashi T and Dynlacht BD**. 2011. Regulating the transition from centriole to basal body. J. Cell Biol. **193**: 435-444.

63. **Taschner M, Bhogaraju S, and Lorentzen E**. 2012. Architecture and function of IFT complex proteins in ciliogenesis. Differentiation. **83**: S12-S22.

64. **Ivanov D, Emonet C, Foata F, Affolter M, Delley M, Fisseha M, Blum-Sperisen S, Kochhar S, and Arigoni F**. 2006. A serpin from the gut bacterium *Bifidobacterium longum* inhibits eukaryotic elastase-like serine proteases. J. Biol. Chem. **281**: 17246-17252.

65. **Brock FM, Forsberg CW, and Buchanan-Smith JG**. 1982. Proteolytic activity of rumen microorganisms and effects of proteinase inhibitors. Appl. Environ. Microbiol. **44**: 561-569.

66. **Steenbakkers PJM, Irving JA, Harhangi HR, Swinkels WJC, Akhmanova A, Dijkerman R, Jetten MSM, van der Drift C, Whisstock JC, and Op den Camp HJM**. 2008. A serpin in the cellulosome of the anaerobic fungus *Piromyces* sp. strain E2. Mycol. Res. **112**: 999-1006.

67. **Forsberg CW, Lovelock LK, Krumholz L, and Buchanan-Smith JG**. 1984. Protease activities of rumen protozoa. Appl. Environ. Microbiol. **47**: 101-110.

68. **Müller M, Mentel M, van Hellemond JJ, Henze K, Woehle C, Gould SB, Yu R-Y, van der Giezen M, Tielens AGM, and Martin WF**. 2012. Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. Microbiol. Mol. Biol. Rev. **76**: 444-495.

69. **Mentel M and Martin W**. 2008. Energy metabolism among eukaryotic anaerobes in light of Proterozoic ocean chemistry. Phil. Trans. R. Soc. B. **363**: 2717-2729.

70. **Tachezy J and Doležal P**. 2007. Iron-Sulfur proteins and iron-sulfur cluster assembly in organisms with hydrogenosomes and mitosomes., p. 105-134. *In* Martin W F and Müller M (ed), Origin of Mitochondria and Hydrogenosomes. Springer-Verlag, Berlin Heidelberg.

71. **Lithgow T and Schneider A**. 2010. Evolution of macromolecular import pathways in mitochondria, hydrogenosomes and mitosomes. Phil. Trans. R. Soc. B. **365**: 799-817.

72. **Chacinska A, Koehler CM, Milenkovic D, Lithgow T, and Pfanner N**. 2009. Importing mitochondrial proteins: machineries and mechanisms. Cell. **138**: 628-644.

73. **Neupert W and Herrmann JM**. 2007. Translocation of proteins into mitochondria. Ann. Rev. Biochem. **76**: 723-749.

74. **Rada P, Doležal P, Jedelský PL, Bursac D, Perry AJ, Šedinová M, Smíšková K, Novotný M, Beltrán NC, Hrdý I, Lithgow T, and Tachezy J**. 2011. The core components of organelle biogenesis and membrane transport in the hydrogenosomes of *Trichomonas vaginalis*. PLoS ONE. **6**: e24428.

75. **Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PE, Eriksson OE, Huhndorf S, James T, Kirk PM, L¸cking R, Thorsten LH, Lutzoni F, Matheny PB, McLaughlin DJ, Powell MJ, Redhead S, Schoch C, Spatafora JW, Stalpers JA, Vilgalys R, Aime AM, A. Aptroot A, R. Bauer R, D. Begerow D, G. Benny GL, L. A. Castlebury LA, Crous PC, Dai Y-C, Gams W, Geiser D, Griffith GW, Gueidan C, Hawksworth DL, Hestmark G,**

Hosaka K, Humber RA, Hyde KD, Ironside JE, Kıljalg U, Kurtzman CP, Larsson K-H, Lichtwardt R, Longcore J, Miadlikowska J, Miller A, Moncalvo J-M, Mozley-Standridge S, Oberwinkler F, E. Parmasto E, V. Reeb V, Rogers JD, Roux C, Ryvarden L, Sampaio J, Sch AF, J. Sugiyama, Thorn RG, Tibell L, Untereiner WA, Walker C, Wang Z, Weir A, Weiss M, White M, Winka K, Yao Y-J, and Zhang N. 2007. A higher-level phylogenetic classification of the Fungi. Mycol. Res. **111**: 509-547.

76. **Lill R and M√ºhlenhoff U**. 2008. Maturation of iron-sulfur proteins in eukaryotes: mechanisms, connected processes, and diseases. Ann. Rev. Biochem. **77**: 669-700.

77. **Lill R and Mühlenhoff U**. 2006. Iron-sulfur protein biogenesis in eukaryotes: components and mechanisms. Ann. Rev. Cell Develop. Biol. **22**: 457-486.

78. **Sharma AK, Pallesen LJ, Spang RJ, and Walden WE**. 2010. Cytosolic iron-sulfur cluster assembly (CIA) system: factors, mechanism, and relevance to cellular iron regulation. J. Biol. Chem. **285**: 26745-26751.

79. **Saier MH, Yen MR, Noto K, Tamang DG, and Elkan C**. 2009. The Transporter Classification Database: recent advances. Nucleic Acids Res. **37**: D274-D278.

80. **Stamatakis A**. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. **22**: 2688-2690.

81. **Harhangi HR, Akhmanova A, Steenbakkers PJM, Jetten MSM, van der Drift C, and Op den Camp HJM**. 2003. Genomic DNA analysis of genes encoding (hemi-)cellulolytic enzymes of the anaerobic fungus *Piromyces* sp. E2. Gene. **314**: 73-80.

82. **Li X-L, Ljungdahl L, Ximenes E, Chen H, Felix C, Cotta M, and Dien B**. 2004. Properties of a recombinant β-glucosidase from polycentric anaerobic fungus *Orpinomyces* PC-2 and its application for cellulose hydrolysis. Appl. Biochem. Biotechnol. **113**: 233-250.

83. **Harhangi HR, Steenbakkers PJM, Akhmanova A, Jetten MSM, van der Drift C, and Op den Camp HJM**. 2002. A highly expressed family 1 β-glucosidase with transglycosylation capacity from the anaerobic fungus *Piromyces* sp. E2. Biochim. Biophys. Acta. **1574**: 293-303.

84. **Chen H-L, Chen Y-C, Lu M-Y, Chang J-J, Wang H-T, Ke H-M, Wang T-Y, Ruan S-K, Wang T-Y, Hung K-Y, Cho H-Y, Lin W-T, Shih M-C, and Li W-H**. 2012. A highly efficient beta-glucosidase from the buffalo rumen fungus *Neocallimastix patriciarum* W5. Biotechnol. Biofuels. **5**: 24.

85. **Wang T-Y, Chen H-L, Lu M-Y, Chen Y-C, Sung H-M, Mao C-T, Cho H-Y, Ke H-M, Hwa T-Y, Ruan S-K, Hung K-Y, Chen C-K, Li J-Y, Wu Y-C, Chen Y-H, Chou S-P, Tsai Y-W, Chu T-C, Shih C-C, Li W-H, and Shih M-C**. 2011. Functional characterization of cellulases identified from the cow rumen fungus Neocallimastix patriciarum W5 by transcriptomic and secretomic analyses. Biotechnology for Biofuels. **4**: 24.

86. **Steenbakkers PJM, Harhangi HR, Bosscher MW, van der Hooft MMC, Keltjens JT, van der Drift C, Vogels GD, and op den Camp HJM**. 2003. Beta-

glucosidase in cellulosome of the anaerobic fungus *Piromyces* sp. strain E2 is a family 3 glycoside hydrolase. Biochem. J. **370**: 963-970.

87. **Xue G-P, Gobius KS, and Orpin CG**. 1992. A novel polysaccharide hydrolase cDNA (celD) from *Neocallimastix patriciarum* encoding three multi-functional catalytic domains with high endoglucanase, cellobiohydrolase and xylanase activities. J. Gen. Microbiol. **138**: 2397-2403.
88. **Steenbakkers PJ, Li XL, Ximenes EA, Arts JG, Chen H, Ljungdahl LG, and Camp HJOD**. 2001. Noncatalytic docking domains of cellulosomes of anaerobic fungi. J. Bacteriol. **183**: 5325-5333.
89. **Zhou L, Xue GP, Orpin CG, Black GW, Gilbert HJ, and Hazlewood GP**. 1994. Intronless celB from the anaerobic fungus *Neocallimastix patriciarum* encodes a modular family A endoglucanase. Biochem. J. **297**: 359-364.
90. **Qiu X, Selinger B, Yanke LJ, and Cheng KJ**. 2000. Isolation and analysis of two cellulase cDNAs from *Orpinomyces joyonii*. Gene. **245**: 119-126.
91. **Fujino Y, Ogata K, Nagamine T, and Ushida K**. 1998. Cloning, sequencing, and expression of an endoglucanase gene from the rumen anaerobic fungus *Neocallimastix frontalis* MCH3. Biosci. Biotechnol. Biochem. **62**: 1795-1798.
92. **Tsai C-F, Qiu X, and Liu J-H**. 2003. A comparative analysis of two cDNA clones of the cellulase gene family from anaerobic fungus *Piromyces rhizinflata*. Anaerobe. **9**: 131-140.
93. **Chen H, Li X-L, Blum D, Ximenes E, and Ljungdahl L**. 2003. CelF of *Orpinomyces* PC-2 has an intron and encodes a cellulase (CelF) containing a carbohydrate-binding module. Appl. Biochem. Biotechnol. **108**: 775-785.
94. **Harhangi HR, Freelove ACJ, Ubhayasekera W, van Dinther M, Steenbakkers PJM, Akhmanova A, van der Drift C, Jetten MSM, Mowbray SL, Gilbert HJ, and Op den Camp HJM**. 2003. Cel6A, a major exoglucanase from the cellulosome of the anaerobic fungi *Piromyces* sp. E2 and *Piromyces equi*. Biochim. Biophys. Acta. **1628**: 30-39.
95. **Steenbakkers PJM, Ubhayasekera W, Goossen HJAM, van Lierop EMHM, van der Drift C, Vogels GD, Mowbray SL, and Op den Camp HJM**. 2002. An intron-containing glycoside hydrolase family 9 cellulase gene encodes the dominant 90 kDa component of the cellulosome of the anaerobic fungus *Piromyces* sp. strain E2. Biochem. J. **365**: 193-204.
96. **Black GW, Hazlewood GP, Xue GP, Orpin CG, and Gilbert HJ**. 1994. Xylanase B from *Neocallimastix patriciarum* contains a non-catalytic 455-residue linker sequence comprised of 57 repeats of an octapeptide. Biochem. J. **299**: 381-387.
97. **Pai C-K, Wu Z-Y, Chen M-J, Zeng Y-F, Chen J-W, Duan C-H, Li M-L, and Liu J-R**. 2010. Molecular cloning and characterization of a bifunctional xylanolytic enzyme from *Neocallimastix patriciarum*. Appl. Microbiol. Biotechnol. **85**: 1451-1462.
98. **Liu J-H, Selinger BL, Tsai C-F, and Cheng K-J**. 1999. Characterization of a *Neocallimastix patriciarum* xylanase gene and its product. Can. J. Microbiol. **45**: 970-974.
99. **Fanutti C, Ponyi T, Black GW, Hazlewood GP, and Gilbert HJ**. 1995. The conserved noncatalytic 40-residue sequence in cellulases and hemicellulases from

anaerobic fungi functions as a protein docking domain. J. Biol. Chem. **270**: 29314-29322.

100. **Durand R, Rascle C, and Fèvre M**. 1996. Molecular characterization of xyn3, a member of the endoxylanase multigene family of the rumen anaerobic fungus *Neocallimastix frontalis*. Curr. Genet. **30**: 531-540.
101. **Eberhardt RY, Gilbert HJ, and Hazlewood GP**. 2000. Primary sequence and enzymic properties of two modular endoglucanases, Cel5A and Cel45A, from the anaerobic fungus *Piromyces eq*ui. Microbiology. **146**: 1999-2008.
102. **Steenbakkers P, Freelove A, Van Cranenbroek B, Sweegers B, Harhangi H, Vogels G, Hazlewood G, Gilbert H, and Op den Camp H**. 2002. The major component of the cellulosomes of anaerobic fungi from the genus *Piromyces* is a family 48 glycoside hydrolase. DNA seq. **13**: 313-320.
103. **Qi M, Wang P, Selinger LB, Yanke LJ, Forster RJ, and McAllister TA**. 2011. Isolation and characterization of a ferulic acid esterase (Fae1A) from the rumen fungus *Anaeromyces mucronatus*. J. Appl. Microbiol. **110**: 1341-1350.

III. Supplementary Figures:

Figure S1. (**A)** *Orpinomyces* sp. strain C1A on anaerobic agar roll tubes. (**B)** Phase contrast micrograph displaying polycentric growth of strain C1A. (**C)** Strain C1A growing on and colonizing switchgrass (1), compared to uninoculated control (2) (**D)** Distance dendogram based on ITS1-2 region highlighting the phylogenetic affiliation of strain C1A within the Neocallimastigomycota. The tree was constructed using Neighbor Joining algorithm with Jukes-Cantor corrections. Bootstrap values shown are based on 1000 replicates. **(E)** High order phylogenetic relationship between Neocallimastigomycota C1A genome and selected basal and Dikarya fungal genomes. The maximum likelihood tree was constructed in RaxML (80) using a concatenated set of 42 protein-coding genes. The alignment was comprised of 24349 total trimmed amino acid residues. The tree was constructed using BLOSUM62 model with fixed base frequencies, proportion of invariable sites = 0.07, and α parameter =1.27. Bootstrap values (in percentages) are based on 1000 replicates and are shown for branches with more than 50% bootstrap support.

1

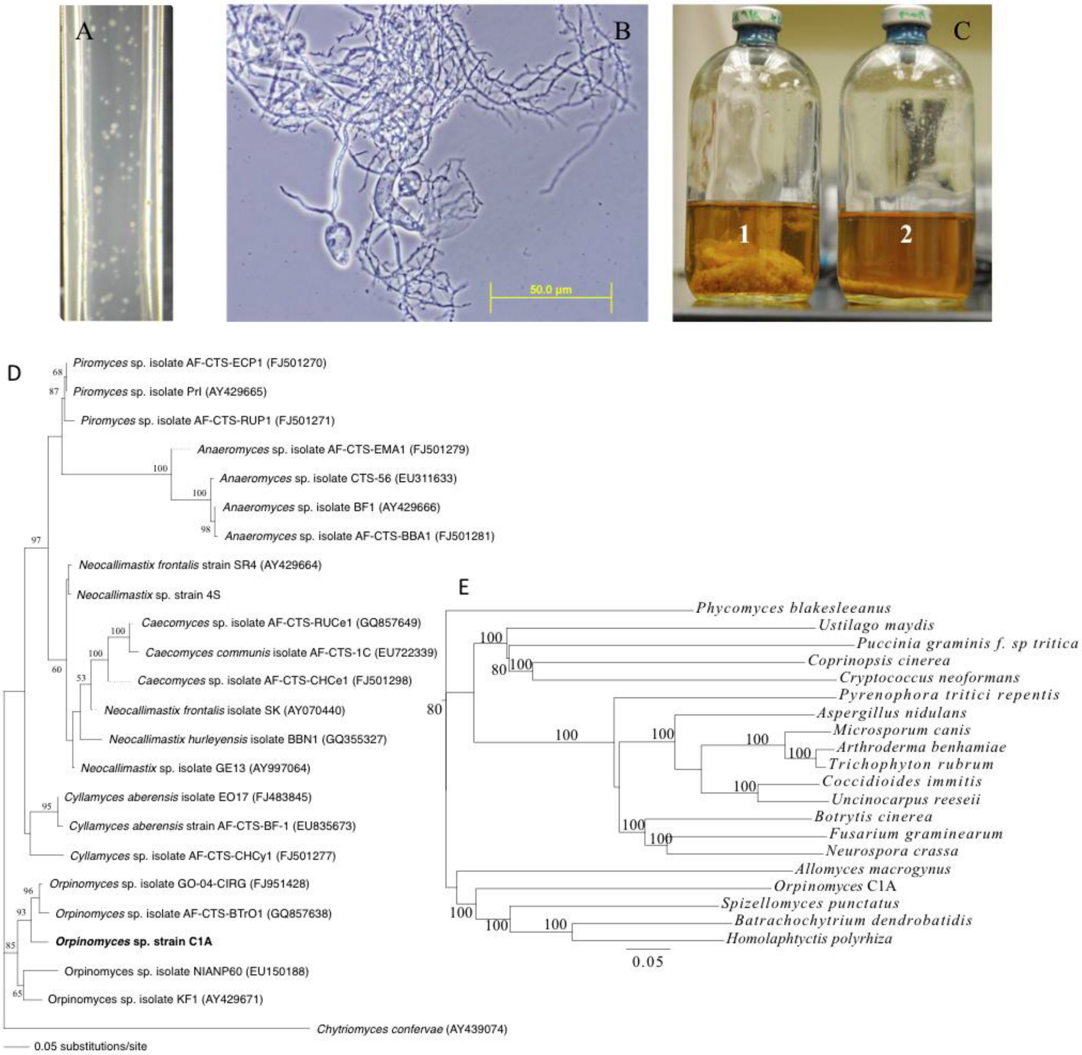

A
B
50.0 µm
C
1
2
D
68
87
97
100
100
98
60
100
100
53
95
96
93
85
65
Piromyces sp. isolate AF-CTS-ECP1 (FJ501270)
Piromyces sp. isolate PrI (AY429665)
Piromyces sp. isolate AF-CTS-RUP1 (FJ501271)
Anaeromyces sp. isolate AF-CTS-EMA1 (FJ501279)
Anaeromyces sp. isolate CTS-56 (EU311633)
Anaeromyces sp. isolate BF1 (AY429666)
Anaeromyces sp. isolate AF-CTS-BBA1 (FJ501281)
Neocallimastix frontalis strain SR4 (AY429664)
Neocallimastix sp. strain 4S
Caecomyces sp. isolate AF-CTS-RUCe1 (GQ857649)
Caecomyces communis isolate AF-CTS-1C (EU722339)
Caecomyces sp. isolate AF-CTS-CHCe1 (FJ501298)
Neocallimastix frontalis isolate SK (AY070440)
Neocallimastix hurleyensis isolate BBN1 (GQ355327)
Neocallimastix sp. isolate GE13 (AY997064)
Cyllamyces aberensis isolate EO17 (FJ483845)
Cyllamyces aberensis strain AF-CTS-BF-1 (EU835673)
Cyllamyces sp. isolate AF-CTS-CHCy1 (FJ501277)
Orpinomyces sp. isolate GO-04-CIRG (FJ951428)
Orpinomyces sp. isolate AF-CTS-BTrO1 (GQ857638)
Orpinomyces sp. strain C1A
Orpinomyces sp. isolate NIANP60 (EU150188)
Orpinomyces sp. isolate KF1 (AY429671)
Chytriomyces confervae (AY439074)
0.05 substitutions/site
E
80
100
80
100
100
100
100
100
100
100
100
100
100
100
Phycomyces blakesleeanus
Ustilago maydis
Puccinia graminis f. sp tritica
Coprinopsis cinerea
Cryptococcus neoformans
Pyrenophora tritici repentis
Aspergillus nidulans
Microsporum canis
Arthroderma benhamiae
Trichophyton rubrum
Coccidioides immitis
Uncinocarpus reeseii
Botrytis cinerea
Fusarium graminearum
Neurospora crassa
Allomyces macrogynus
Orpinomyces C1A
Spizellomyces punctatus
Batrachochytrium dendrobatidis
Homolaphtyctis polyrhiza
0.05

Figure S2. Comparison of total number of introns (A), intron density (B), and introns per gene (C) between the C1A (circled) and 120 different fungal genomes (90 Ascomycota, 16 Basidiomycota, 7 Early-diverging fungal lineages, and 7 Microsporidiae).

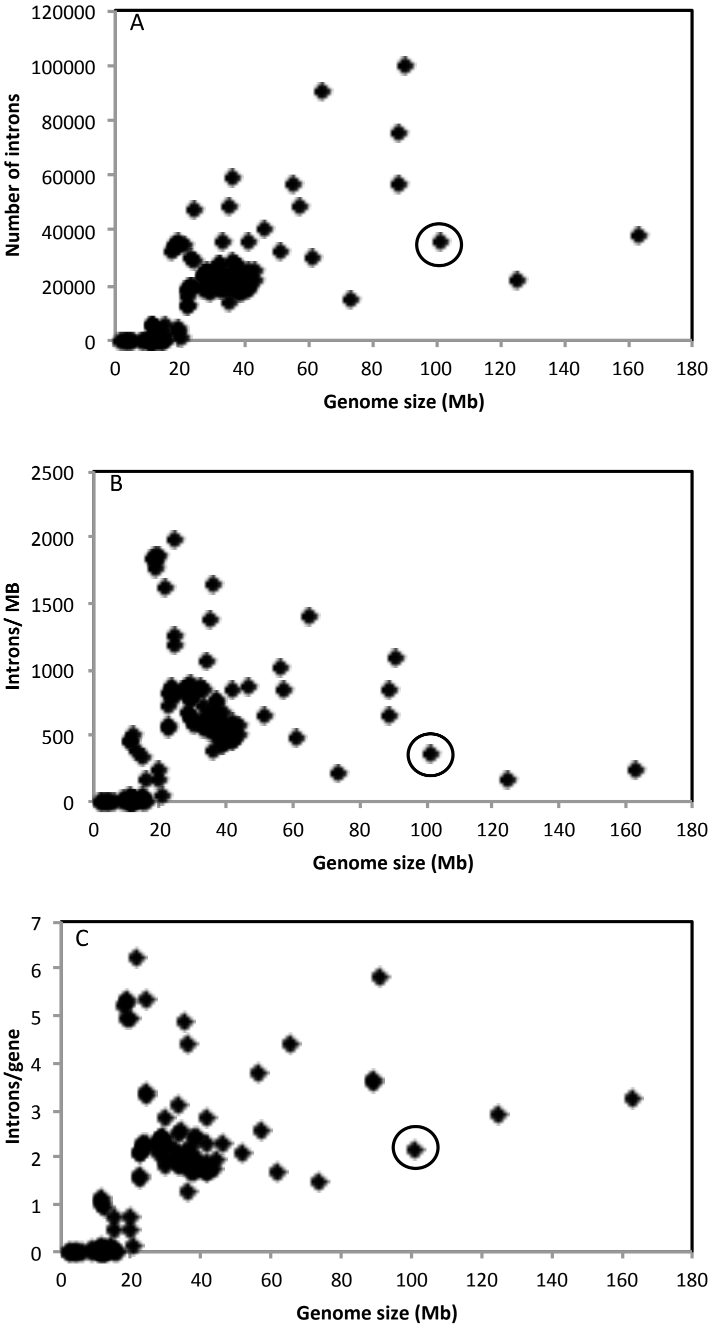


Figure S3. Suggested N-glycan oligosaccharide structure of strain C1A.

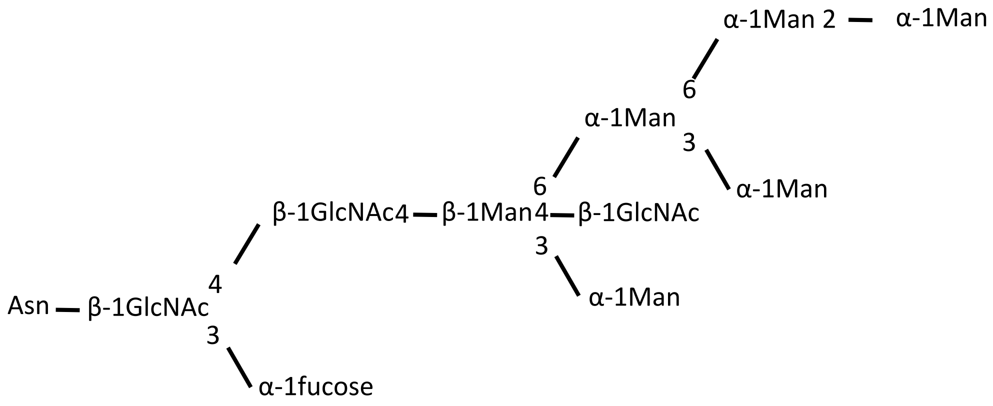


Figure S4. Intracellular protein trafficking in strain C1A. Proteins that are not destined for secretion are synthesized on free ribosomes (brown arrows). Those proteins could either remain in the cytosol (cytosolic proteins, route 1), transport to the nucleus (nuclear proteins, route 2), or could be imported into the hydrogenosome (hydrogenosomal proteins, route 3), or peroxisome (peroxisomal proteins, route 4). On the other hand, proteins synthesized on membrane-bound ribosomes of the rough endoplasmic reticulum (RER) could either be secreted (secretory pathway, blue arrows), or enter the endosomal system and eventually be degraded in the lysosome (endosomal pathway, green arrows). More detailed explanation is presented in the supplementary text. Coat protein components, SNARE components, and vesicle tethering complexes are indicated where appropriate. EE: early endosome, ER: endoplasmic reticulum, H: hydrogenosome, L: lysosome, LE: late endosome, MVB: multivesicular bodies, P: peroxisome, RE: recycling endosome, TGN: trans-Golgi network.

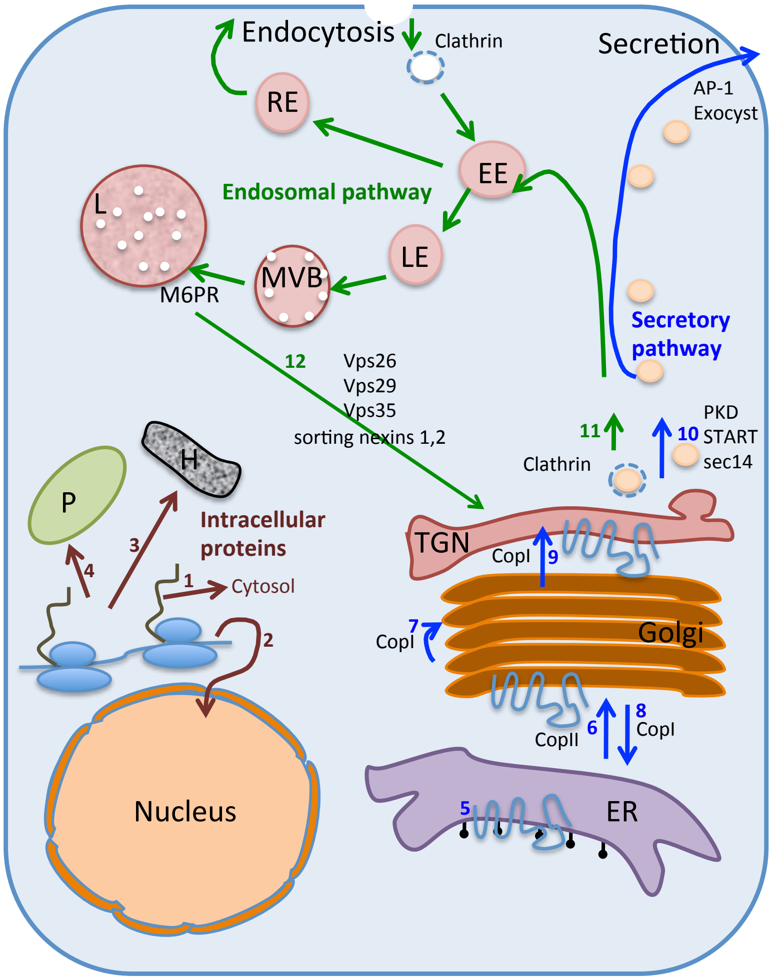
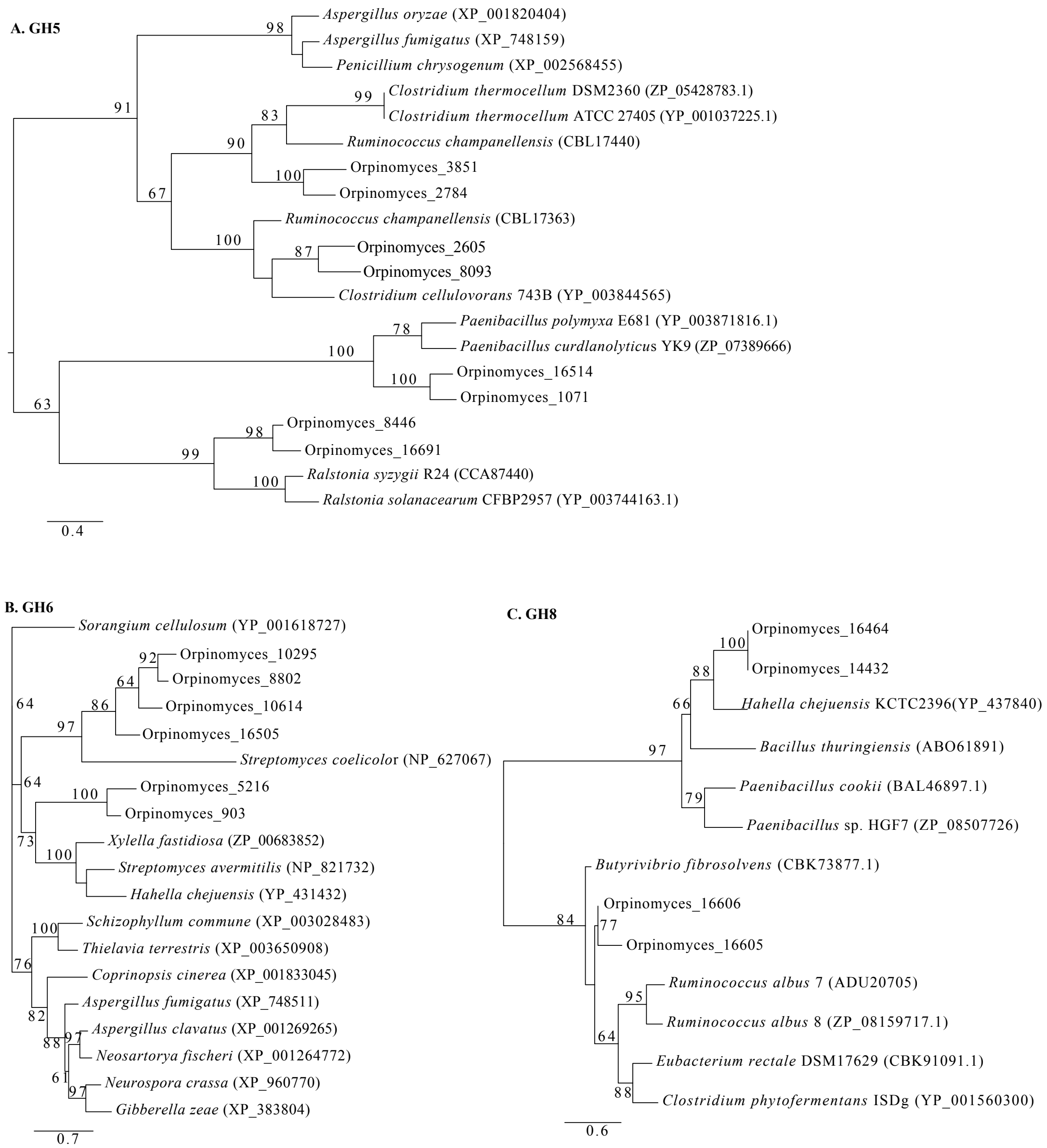
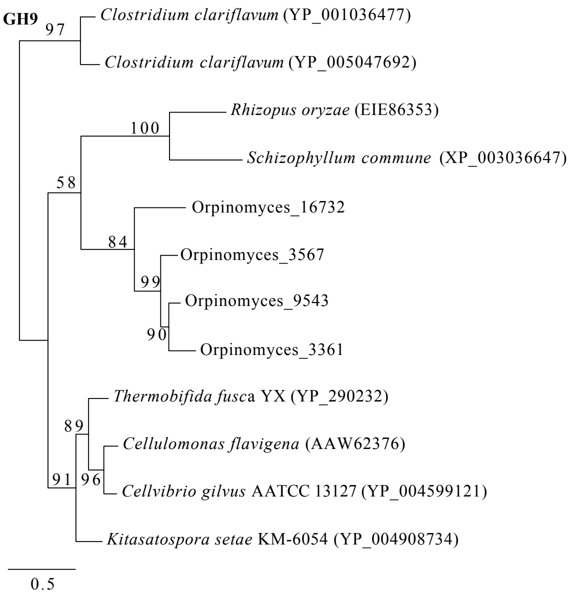


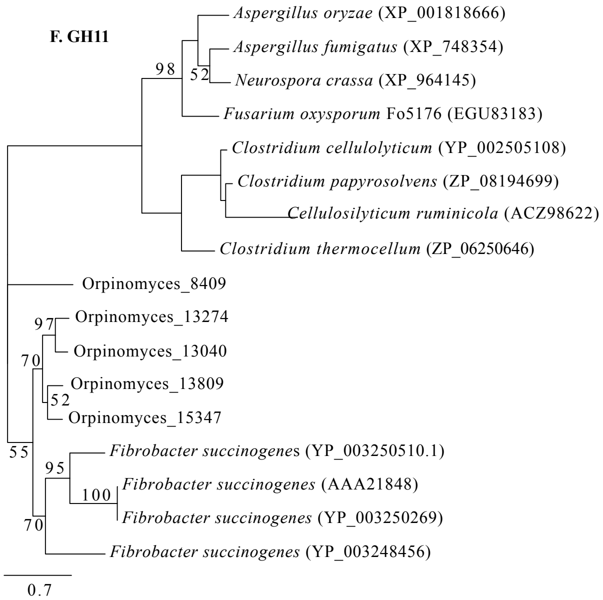
Figure S5. Maximum-likelihood dendograms showing the phylogenetic affiliation of selected C1A plant biomass-degradation enzymes belonging to families (A) GH5, (B) GH6, (C) GH8, (D) GH9, (E) GH10, (F) GH11, (G) GH26, (H) GH43, (I) GH45, (J) GH48, and (K) GH67. The trees were constructed in RaxML (80) using BLOSUM62 model with fixed base frequencies. Bootstrap values (in percentages) are based on 1000 replicates and are shown for branches with more than 50% bootstrap support.



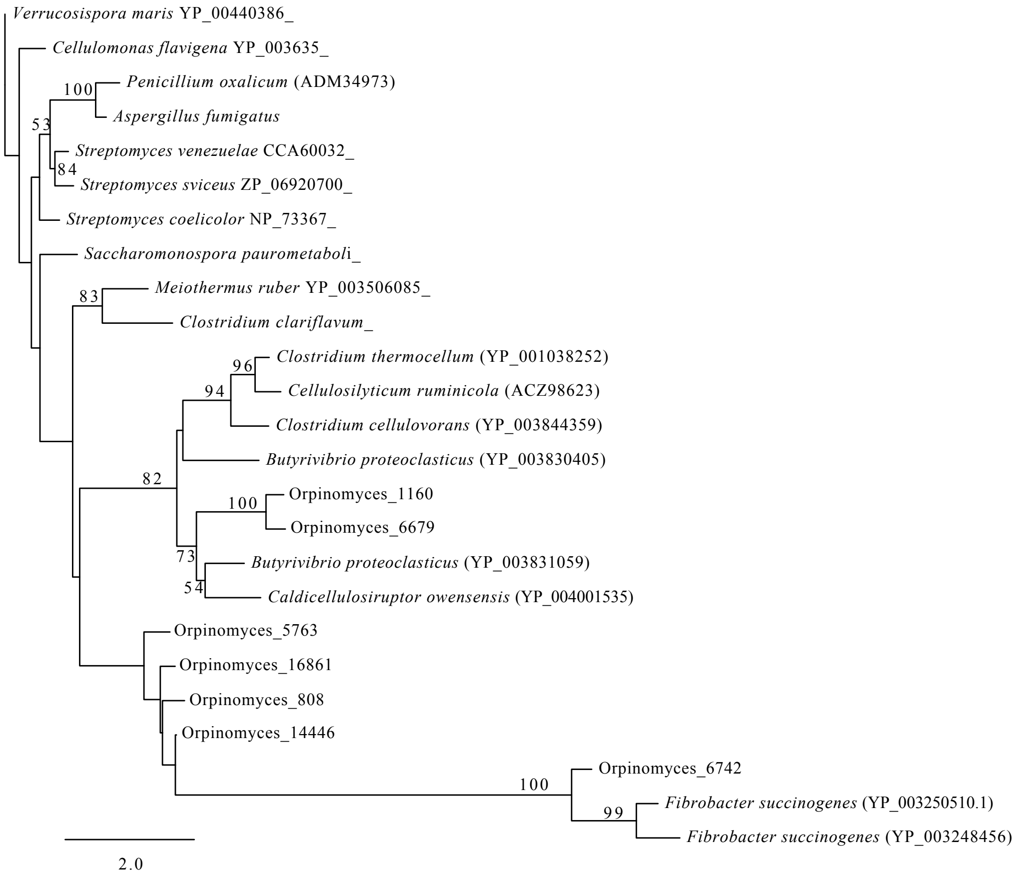
D. GH9



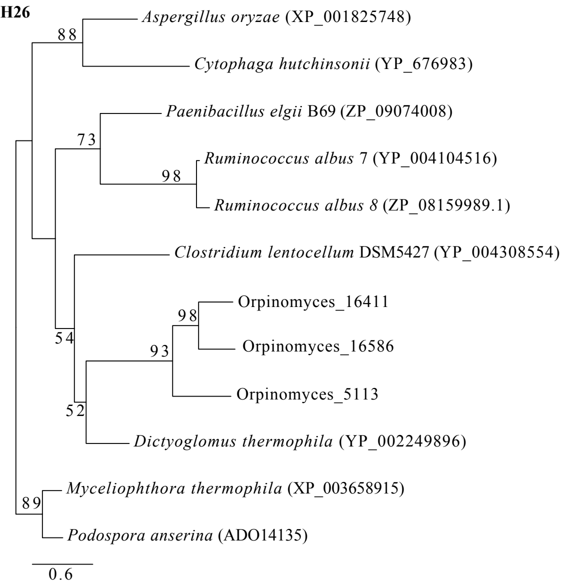
F. GH11



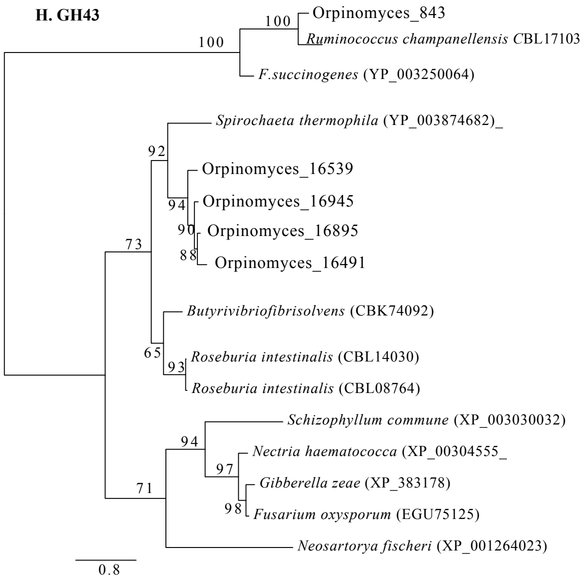
E. GH10



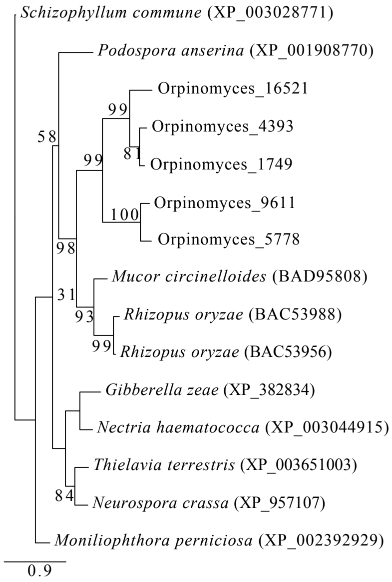
G. GH26



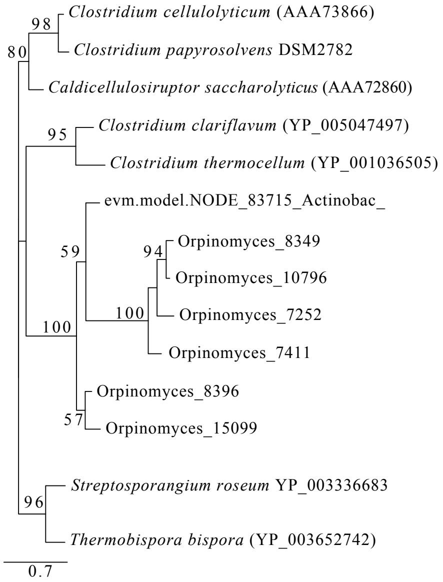
H. GH43



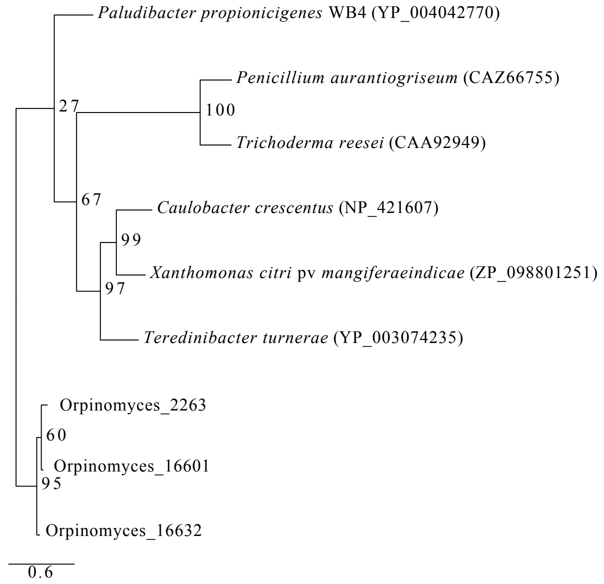
I. GH45



J. GH48



K. GH67



Paludibacter propionicigenes WB4 (YP_004042770)
Penicillium aurantiogriseum (CAZ66755)
100
Trichoderma reesei (CAA92949)
27
67
Caulobacter crescentus (NP_421607)
99
Xanthomonas citri pv *mangiferaeindicae* (ZP_098801251)
97
Teredinibacter turnerae (YP_003074235)
Orpinomyces_2263
60
Orpinomyces_16601
95
Orpinomyces_16632
0.6

Figure S6. Transcription levels of genes belonging to various GH families in cellulose (blue)- versus cellobiose (red)-grown cultures. Transcription levels are expressed as absolute transcripts per million (TPM) in (A), and as normalized TPM relative to a suite of glycolytic genes in (B).

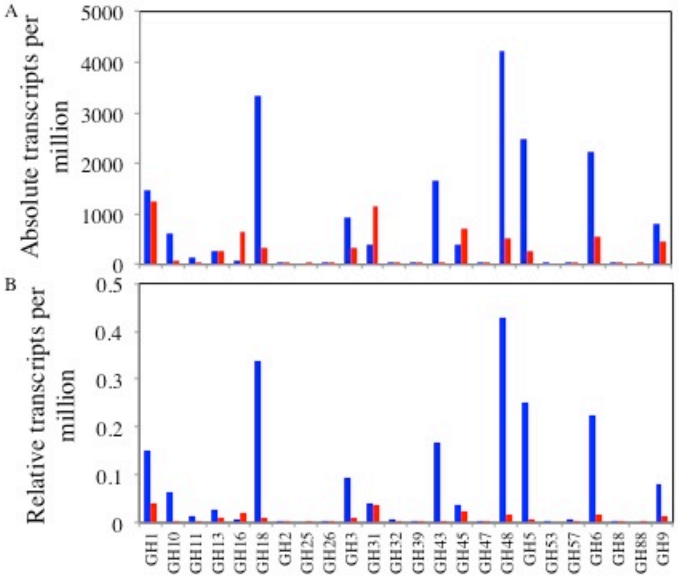


Figure S7. Crystalline cellulose and xylan utilization by *Orpinomyces* strain C1A. (A) Cellulose loss (○) and lactate (◊), formate (□), acetate (△), and ethanol (X) production in microcosms that were amended with avicel and strain C1A. (B) Formate (◊), Lactate (□), acetate (△), and ethanol (X) production in microcosms that were amended with xylan and strain C1A.

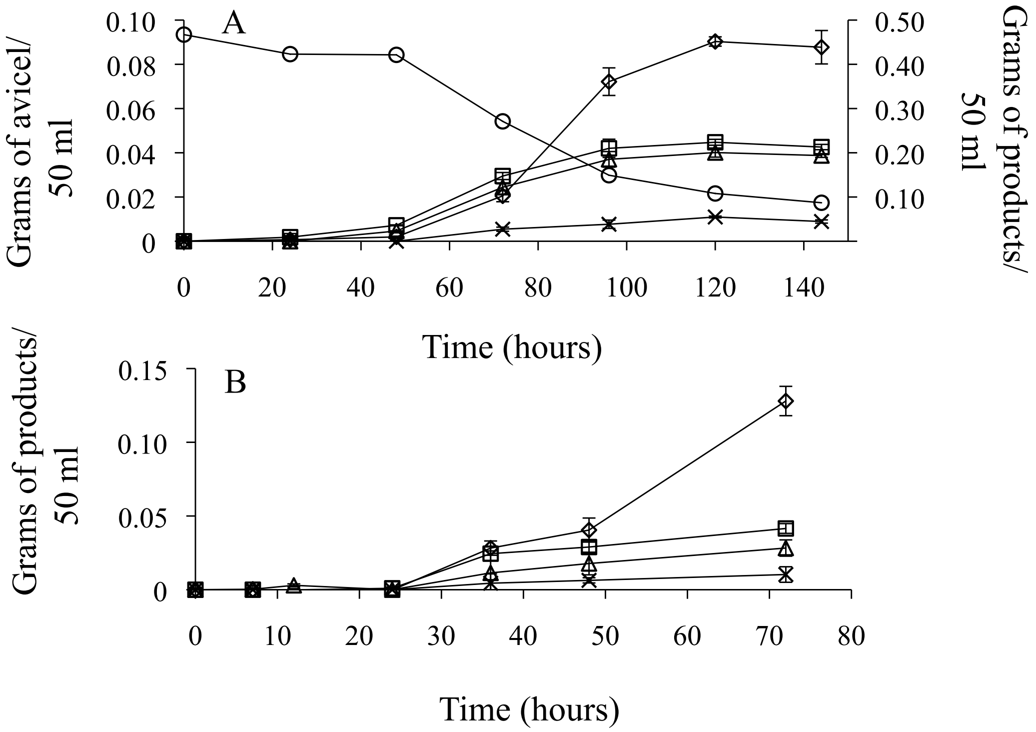


Figure S8. Grams of dry weight (◊), cellulose (□), hemicellulose (△), and lignin (X) lost in microcosms that contained untreated (A), sodium hydroxide-treated (B), acid-treated (C), and hydrothermolysis-treated (D) switchgrass.

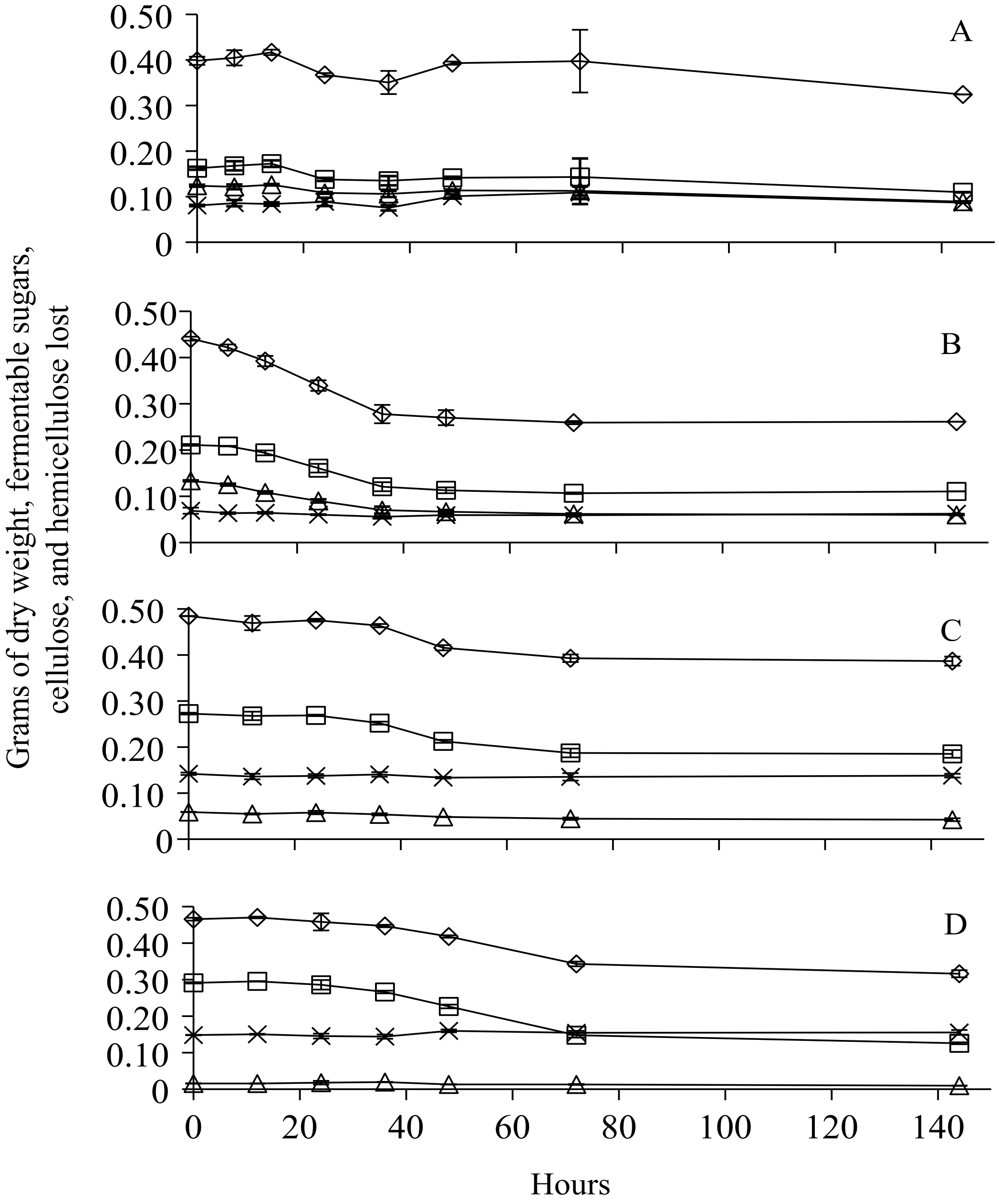
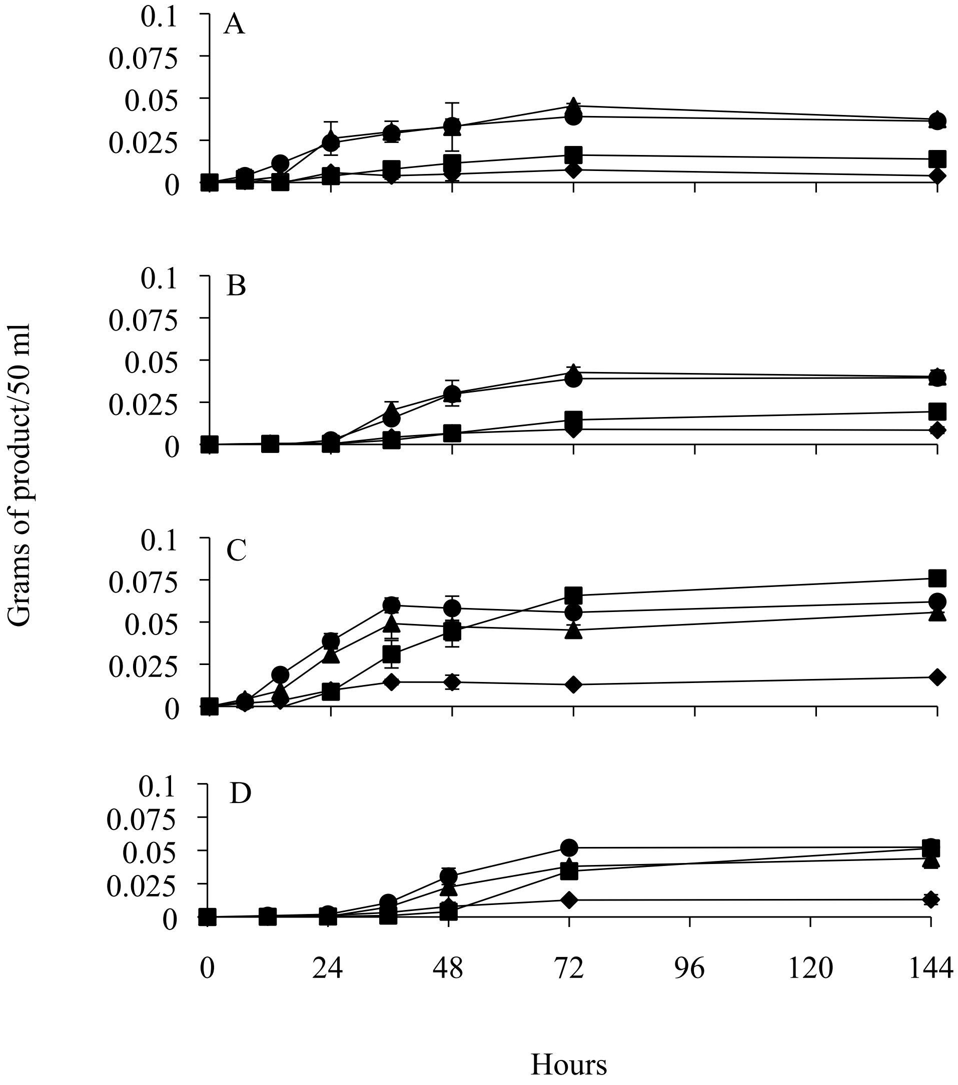


Figure S9. Acids and alcohols produced during switchgrass degradation. Lactate (■), formate (●), acetate (▲), and ethanol (◆) production in microcosms that contained untreated (A), acid-treated (B), sodium hydroxide-treated (C), and hydrothermolysis-treated switchgrass (D).



IV. Supplementary tables.

In all tables, genes with labeled “Orpinomyces_” have been identified in C1A genome assembly. Genes with an IMG number have only been identified in the transcriptome of cellobiose-grown cultures.

Table S1. The C1A genome sequencing efforts summary

Illumina Genomic Passed Filtered Paired Reads	146,385,792
Illumina Genomic Total Bases	29.2 GB
Illumina cellobiose transcriptome Reads	292,771,584
Illumina cellobiose transcriptome Total Bases	35.2 GB
Illumina cellulose transcriptome Reads	26,627,384
Illumina cellulose transcriptome Total Bases	3.9 GB
Pacific Biosciences SMRT Total Reads	480,781
Pacific Bioesceinces SMRT Total Bases	1.01 GB
C-SMRT Total Reads	394,300
C-SMRT Total Bases	570.1 MB

Table S2. Comparison between genome assembly using Illumina only sequence data and SMRT-Illumina hybrid assembly

	Illumina Only	PacBio/Illumina assembly
Genome assembly size	105.1 MB	100.85 MB
In 1kb+ Contigs	73.52 MB	100.95 MB
Number of ambiguous bp	91,688 bp	0
N50 of 1kb+ Contigs	2,226 bp	3,373 bp
N90 of 1kb contigs	1,072 bp	1,829 bp
Average length of gene model	903 bp	1623 bp
Number of introns	2,458	35,697
Number of gene models	14,594	16,437
GC% Content	15.8	17.0
Total PASA/Trinity Assemblies	10,115	14,009

Table S3. Gene model predictions in the C1A genome

Trinity De Novo Assembled Transcripts 300bp+	25,934
N50 of Trinity De Novo Assembled Transcripts 300bp+	1082 bp
GlimmerHMM+Augtsus Models ab inito	60,595
PASA Assemblies	14,009
PASA GMAP RNA Alignments	38,647
Total EVM Consoldated Models	16,347

Table S4. General genomic features of the C1A genome

Genome size	100.95MB
Number of Contigs	32,574
Protein Coding	20.60%
Non- coding intergenic	73.60%
Non-coding introns	5.10%
rRNA	0.67%
5.8S	183 (30,763 bp)
18S	272 (168,110 bp)
28S	366 (457,301 bp)
tRNA	0.06%
	770 (58,292 bp)
Number of Genes	16,347
Number of Genes with transcripts	14,009
Average Gene Length	1623
Number of Intron	35,697
Introns/gene	2.18
Average Intronlength	163
Number of Exons	52,044
GC content	17.00%
Protein Coding	26.80%
Intergenic	14.80%
Intron	8.10%
SSR Repeats	4.90%
TE repeats	3.31%

Table S5. Other genomes used in this study.

Dataset Name	**NCBI taxon ID**	**Kingdom**	**Phylum**	**Genome Size Mb**	**GC Ratio**	**ORF #**	**Intron#**	**coding density**
Ascomycetes								
Pezizomycotina								
Ajellomyces dermatitidis SLH14081	**559298**	**Fungi**	**Ascomycota**	**75.35**	**35.76**	**9555**	**22100**	**17.12**
Arthroderma benhamiae CBS 112371	**663331**	**Fungi**	**Ascomycota**	**22.22**	**48.75**	**7980**	**12672**	**53.24**
Aspergillus clavatus NRRL 1	**344612**	**Fungi**	**Ascomycota**	**27.86**	**49.21**	**9121**	**18838**	**48.57**
Aspergillus flavus NRRL3357	**332952**	**Fungi**	**Ascomycota**	**36.79**	**48.35**	**12604**	**28367**	**47.32**
Aspergillus fumigatus A1163	451804	Fungi	Ascomycota	29.21	49.55	9929	19165	49.00
Aspergillus fumigatus Af293	**330879**	**Fungi**	**Ascomycota**	**29.39**	**49.80**	**9887**	**18277**	**48.83**
Aspergillus nidulans FGSC A4	**227321**	**Fungi**	**Ascomycota**	**30.07**	**50.32**	**10701**	**24824**	**51.45**
Aspergillus niger ATCC 1015	380704	Fungi	Ascomycota	37.20	50.36	11200	23771	45.17
Aspergillus niger CBS 513.88	**425011**	**Fungi**	**Ascomycota**	**33.98**	**50.37**	**14086**	**36285**	**65.42**
Aspergillus oryzae RIB40	**510516**	**Fungi**	**Ascomycota**	**37.12**	**48.24**	**12063**	**23256**	**43.83**
Aspergillus terreus NIH2624	**341663**	**Fungi**	**Ascomycota**	**29.33**	**52.90**	**10406**	**22710**	**53.25**
Chaetomium globosum CBS 148.51	**306901**	**Fungi**	**Ascomycota**	**34.89**	**55.56**	**11124**	**23599**	**56.77**
Coccidioides immitis H538.4	396776	Fungi	Ascomycota	27.73	46.96	10663	23840	40.96
Coccidioides immitis RMSCC 2394	404692	Fungi	Ascomycota	28.82	46.23	10408	24399	43.82
Coccidioides immitis RMSCC 3703	454286	Fungi	Ascomycota	27.65	47.08	10465	23466	39.86
Coccidioides immitis RS	**246410**	**Fungi**	**Ascomycota**	**28.89**	**46.02**	**10457**	**25680**	**44.02**
Coccidioides posadasii RMSCC 3488	454284	Fungi	Ascomycota	28.15	46.33	9964	23520	44.03
Coccidioides posadasii Silveira		**Fungi**	**Ascomycota**	**27.47**	**46.86**	**10125**	**23395**	**45.56**
Cochliobolus heterostrophus ATCC 48331	**665024**	**Fungi**	**Ascomycota**	**32.93**	**50.73**	**12720**		**56.42**
Cochliobolus heterostrophus C5	**701091**	**Fungi**	**Ascomycota**	**34.86**	**49.98**	**9633**	**18374**	**52.63**
Cryphonectria parasitica EP155	660469	Fungi	Ascomycota	43.86	50.83	11184	21906	33.54
Fusarium graminearum GZ3639		Fungi	Ascomycota	15.15	50.47	6694	4998	
Fusarium graminearum PH-1	**229533**	**Fungi**	**Ascomycota**	**36.55**	**48.28**	**13339**	**24232**	**55.63**
Fusarium oxysporum lycopersici FGSC 4287	**426428**	**Fungi**	**Ascomycota**	**61.36**	**48.40**	**17608**	**29443**	**39.69**
Gaeumannomyces graminis var. tritici R3-111a-1	644352	Fungi	Ascomycota	43.26	56.85	14463	25024	42.79
Gibberella moniliformis 7600	334819	Fungi	Ascomycota	41.87	48.67	14199	25309	42.81
Histoplasma (Ajellomyces) capsulatum G186AR	447093	Fungi	Ascomycota	29.91	44.38	7454	17108	41.25

Histoplasma (Ajellomyces) capsulatum G217B	447094	Fungi	Ascomycota	41.29	42.75	8038	18673	
Histoplasma capsulatum H143	544712	Fungi	Ascomycota	38.96	41.71	7365	17799	29.81
Histoplasma capsulatum H88	544711	Fungi	Ascomycota	37.94	42.00	7428	17928	33.28
Histoplasma capsulatum NAm1	**339724**	**Fungi**	**Ascomycota**	**32.99**	**46.16**	**9349**	**23495**	**35.89**
Magnaporthe oryzae 70-15	**242507**	**Fungi**	**Ascomycota**	**41.62**	**51.68**	**12841**	**21348**	**41.10**
Magnaporthe poae ATCC 64411	644358	Fungi	Ascomycota	39.50	56.99	12169	21875	39.54
Microsporum (Arthroderma) canis CBS 113480	**554155**	**Fungi**	**Ascomycota**	**23.26**	**47.50**	**8777**	**19991**	**55.97**
Microsporum (Arthroderma) gypseum CBS 118893	**535722**	**Fungi**	**Ascomycota**	**23.27**	**48.46**	**8876**	**19748**	**55.00**
Myceliophthora thermophila ATCC 42464	573729	Fungi	Ascomycota	38.74	51.40	9110	16657	33.52
Mycosphaerella fijiensis CIRAD86	383855	Fungi	Ascomycota	73.37	45.30	10327	14962	22.91
Mycosphaerella graminicola IPO323	336722	Fungi	Ascomycota	41.86	52.29	11395	19234	34.05
Nectria haematococca MPVI	**660122**	**Fungi**	**Ascomycota**	**51.29**	**50.79**	**15707**	**32680**	**49.33**
Neosartorya fischeri NRRL 181	**331117**	**Fungi**	**Ascomycota**	**32.55**	**49.43**	**10406**	**20578**	**46.87**
Neurospora crassa OR74A	**367110**	**Fungi**	**Ascomycota**	**39.23**	**49.29**	**9842**	**17346**	**45.45**
Paracoccidioides brasiliensis Pb01	**502779**	**Fungi**	**Ascomycota**	**32.97**	**42.80**	**9136**	**28174**	**36.86**
Paracoccidioides brasiliensis Pb03	482561	Fungi	Ascomycota	29.09	44.48	9264	22384	38.84
Paracoccidioides brasiliensis Pb18	502780	Fungi	Ascomycota	30.00	44.33	8741	24498	39.28
Penicillium chrysogenum Wisconsin 54-1255	**500485**	**Fungi**	**Ascomycota**	**32.22**	**48.96**	**12791**	**27650**	**30.58**
Penicillium marneffei ATCC 18224	**441960**	**Fungi**	**Ascomycota**	**28.64**	**46.67**	**10638**	**23668**	**59.52**
Phaeosphaeria nodorum SN15	**321614**	**Fungi**	**Ascomycota**	**37.24**	**50.42**	**16597**	**27420**	**42.17**
Podospora anserina S mat+	**515849**	**Fungi**	**Ascomycota**	**35.71**	**52.04**	**10596**	**13841**	**42.67**
Pyrenophora tritici-repentis Pt-1C-BFP	**426418**	**Fungi**	**Ascomycota**	**38.00**	**50.89**	**12169**	**20548**	**49.33**
Sclerotinia sclerotiorum 1980 UF-70	**665079**	**Fungi**	**Ascomycota**	**38.33**	**41.84**	**14522**	**26101**	**38.59**
Talaromyces stipitatus ATCC 10500	**441959**	**Fungi**	**Ascomycota**	**35.69**	**46.10**	**13356**		**57.30**
Thielavia terrestris NRRL 8126	578455	Fungi	Ascomycota	36.91	54.70	9813	18204	37.00
Trichoderma atroviride IMI 206040	452589	Fungi	Ascomycota	36.09	49.80	11100	21463	44.92
Trichoderma reesei QM6a	431241	Fungi	Ascomycota	33.45	52.81	9129	18762	40.30
Trichoderma virens Gv29-8	431241	Fungi	Ascomycota	38.77	49.44	11643	23030	45.26
Trichophyton equinum CBS 127.97	559882	Fungi	Ascomycota	24.16	47.37	8576	19382	49.32
Trichophyton rubrum CBS 118892	559305	Fungi	Ascomycota	22.50	48.31	8707	18269	53.91
Trichophyton tonsurans CBS 112818	647933	Fungi	Ascomycota	22.96	48.15	8523	18366	52.29
Trichophyton verrucosum HKI 0517	**663202**	**Fungi**	**Ascomycota**	**22.54**	**48.24**	**8024**	**12612**	**52.26**

Tuber melanosporum Mel28	**656061**	**Fungi**	**Ascomycota**	**124.95**	**44.86**	**7496**	**21525**	**9.74**
Uncinocarpus reesii 1704	**336963**	**Fungi**	**Ascomycota**	**22.35**	**48.66**	**7798**	**16296**	**57.24**
Verticillium albo-atrum VaMs.102	**526221**	**Fungi**	**Ascomycota**	**32.86**	**56.03**	**10239**	**18603**	**40.91**
Verticillium dahliae VdLs.17	498257	Fungi	Ascomycota	33.90	55.79	10575	19161	44.29
Saccharomycotina								
Candida dubliniensis CD36	**573826**	**Fungi**	**Ascomycota**	**14.62**	**33.25**	**5972**	**372**	**60.10**
Lachancea thermotolerans CBS 6340	**559295**	**Fungi**	**Ascomycota**	**10.42**	**47.25**	**5102**		**73.05**
Pichia pastoris GS115	**644223**	**Fungi**	**Ascomycota**	**9.22**	**41.10**	**5040**		**78.88**
Ashbya gossypii ATCC 10895	**284811**	**Fungi**	**Ascomycota**	**8.77**	**51.85**	**4717**	**226**	**80.09**
Botryotinia fuckeliana B05.10	332648	Fungi	Ascomycota	42.74	43.04	16584		47.24
Candida albicans SC5314	**237561**	**Fungi**	**Ascomycota**	**14.30**	**33.46**	**6090**	**534**	**62.38**
Candida albicans WO-1	294748	Fungi	Ascomycota	14.47	33.47	6160	235	61.23
Candida glabrata CBS138	**284593**	**Fungi**	**Ascomycota**	**12.28**	**38.61**	**5165**	**84**	**64.36**
Candida lusitaniae ATCC 42720	**306902**	**Fungi**	**Ascomycota**	**12.11**	**44.50**	**5941**	**15**	**65.53**
Candida parapsilosis CDC317	578454	Fungi	Ascomycota	13.09	38.69	5733	0	66.87
Candida tropicalis MYA-3404	**294747**	**Fungi**	**Ascomycota**	**14.63**	**33.15**	**6258**	**34**	**62.13**
Debaryomyces hansenii var hansenii CBS767	**284592**	**Fungi**	**Ascomycota**	**12.22**	**36.28**	**6354**	**356**	**71.39**
Kluyveromyces lactis NRRL Y-1140	**284590**	**Fungi**	**Ascomycota**	**10.69**	**38.76**	**5327**	**130**	**70.57**
Lachancea kluyveri NRRL Y-12651	226302	Fungi	Ascomycota	11.03	41.64	2968	0	
Lachancea waltii NCYC 2644	262981	Fungi	Ascomycota	10.91	44.29	4935	460	70.87
Lodderomyces elongisporus NRRL YB-4239	**379508**	**Fungi**	**Ascomycota**	**15.55**	**36.95**	**5802**	**54**	**56.83**
Meyerozyma guilliermondii ATCC 6260	**294746**	**Fungi**	**Ascomycota**	**10.61**	**43.76**	**5920**	**15**	**77.49**
Nadsonia fulvescens elongata AJ 4281, DSM 6958	**857566**	**Fungi**	**Ascomycota**	**13.75**	**39.25**	**5657**		**61.89**
Naumovozyma castellii CBS 4309	1064592	Fungi	Ascomycota	11.35	36.95	4677	0	
Saccharomyces bayanus 623-6C	226231	Fungi	Ascomycota	11.87	40.20	4966	0	
Saccharomyces cerevisiae RM11-1a	**285006**	**Fungi**	**Ascomycota**	**11.74**	**38.29**	**5696**	**292**	**61.57**
Saccharomyces cerevisiae S288C	**559292**	**Fungi**	**Ascomycota**	**12.16**	**38.15**	**6692**	**350**	**72.58**
Saccharomyces cerevisiae YJM789	**307796**	**Fungi**	**Ascomycota**	**11.99**	**38.31**	**5903**	**250**	**71.60**
Saccharomyces kudriavzevii IFO 1802	226230	Fungi	Ascomycota	11.18	39.85	3768	0	
Saccharomyces mikatae IFO 1815	226126	Fungi	Ascomycota	11.47	37.97	9016	0	
Saccharomyces mikatae IFO 1815	226126	Fungi	Ascomycota	10.83	38.14	3100	0	

Saccharomyces paradoxus NRRL Y-17217	226125	Fungi	Ascomycota	11.87	38.69	8939	0	
Scheffersomyces stipitis CBS 6054	**322104**	**Fungi**	**Ascomycota**	**15.44**	**41.14**	**5839**	**2589**	**61.12**
Vanderwaltozyma polyspora DSM 70294	**436907**	**Fungi**	**Ascomycota**	**14.67**	**33.03**	**5367**	**157**	**57.89**
Yarrowia lipolytica CLIB122	**284591**	**Fungi**	**Ascomycota**	**20.50**	**49.05**	**6524**	**740**	**47.05**
Zygosaccharomyces rouxii CBS 732	**559307**	**Fungi**	**Ascomycota**	**9.77**	**39.20**	**5272**		**76.82**
Taphrinomycotina								
Schizosaccharomyces cryophilus OY26	653667	Fungi	Ascomycota	11.55	37.67	5163	5929	60.56
Schizosaccharomyces japonicus yFS275	**402676**	**Fungi**	**Ascomycota**	**11.30**	**43.67**	**5172**	**5149**	**64.01**
Schizosaccharomyces octosporus yFS286	483514	Fungi	Ascomycota	11.22	37.53	4925	5243	58.97
Schizosaccharomyces pombe 972h-	**284812**	**Fungi**	**Ascomycota**	**12.59**	**36.05**	**5058**	**4811**	**59.89**
Saitoella complicata NRRL Y-17804	**698492**	**Fungi**	**Ascomycota**	**14.14**	**52.60**	**7034**		**67.51**
Basidiomycetes								
Agaricomycetes								
Botryobasidium botryosum	**264124**	**Fungi**	**Basidiomycota**	**46.67**	**52.44**	**16526**		**53.06**
Moniliophthora perniciosa FA553	**554373**	**Fungi**	**Basidiomycota**	**26.67**	**47.74**	**13674**		**35.18**
Coprinopsis cinerea okayama7#130	**240176**	**Fungi**	**Basidiomycota**	**36.25**	**51.64**	**13544**	**59343**	**51.24**
Cryptococcus gattii R265	294750	Fungi	Basidiomycota	17.50	47.79	6210	32178	54.83
Cryptococcus neoformans var neoformans B-3501A	**283643**	**Fungi**	**Basidiomycota**	**18.52**	**48.48**	**6431**	**34511**	**64.92**
Cryptococcus neoformans var neoformans JEC 21	**214684**	**Fungi**	**Basidiomycota**	**19.05**	**48.54**	**6475**	**34336**	**64.77**
Cryptococcus neoformans var. grubii H99	235443	Fungi	Basidiomycota	19.46	48.18	7302	36023	56.81
Cryptococcus neoformans var. neoformans WM276		Fungi	Basidiomycota	19.01	47.84	6870	33719	55.59
Gymnopus luxurians FD-317 M1	**206324**	**Fungi**	**Basidiomycota**	**66.28**	**45.08**	**22057**		**47.16**
Hypholoma sublateritium FD-334 SS-4	**71945**	**Fungi**	**Basidiomycota**	**48.03**	**51.04**	**17911**		**54.50**
Laccaria bicolor S238N-H82	**486041**	**Fungi**	**Basidiomycota**	**64.88**	**46.97**	**20614**	**90676**	**41.34**
Phanerochaete chrysosporium RP-78	**273507**	**Fungi**	**Basidiomycota**	**35.15**	**56.82**	**10048**	**48688**	**47.65**
Postia placenta MAD 698-R	**561896**	**Fungi**	**Basidiomycota**	**90.93**	**53.83**	**17173**	**99423**	**19.39**
Schizophyllum commune H4-8	**578458**	**Fungi**	**Basidiomycota**	**38.62**	**57.50**	**14652**		**58.31**
Serpula lacrymans lacrymans S7.9	**341189**	**Fungi**	**Basidiomycota**	**42.75**	**45.30**	**16257**		**46.04**
Pucciniomycotina								

Puccinia graminis f. sp. tritici CRL 75-36-700-3	418459	Fungi	Basidiomycota	88.72	43.35	20567	75271	25.54
Puccinia graminis tritici	413621	Fungi	Basidiomycota	88.64	43.35	15800	56898	21.08
Puccinia triticina 1-1 BBBD Race 1	630390	Fungi	Basidiomycota	162.95	46.34	11638	38114	8.60
Sporobolomyces roseus IAM 13481	365493	Fungi	Basidiomycota	21.17	53.75	5536	34375	44.30
Ustilaginomycotina								
Malassezia globosa CBS 7966	**425265**	**Fungi**	**Basidiomycota**	**8.96**	**52.06**	**4286**	**0**	**71.13**
Ustilago maydis 521	237631	Fungi	Basidiomycota	19.74	53.96	6689	4900	64.32
Ustilago maydis FB1	559306	Fungi	Basidiomycota	19.30	54.03	6950	3360	
Basal fungal lineages								
Blastocladiomycetes								
Allomyces macrogynus ATCC 38327	**578462**	**Fungi**	**Blastocladiomycota**	**57.06**	**61.59**	**18773**	**48417**	**47.48**
Chyridiomycetes								
Batrachochytrium dendrobatidis JAM81	684364	Fungi	Chytridiomycota	24.32	39.25	8732	28691	48.54
Batrachochytrium dendrobatidis JEL423	**403673**	**Fungi**	**Chytridiomycota**	**23.90**	**39.27**	**8818**	**29733**	**66.26**
Homoloaphlyctis polyrhiza JEL 142	166479	Fungi	Chytridiomycota	26.70			0	
Spizellomyces punctatus DAOM BR117	**645134**	**Fungi**	**Chytridiomycota**	**24.13**	**47.60**	**8952**	**47775**	**57.73**
Early diverging fungal lineages								
Mortierella verticillata NRRL 6337	1069443	Fungi	Mortierellomycotina	41.85	48.90	12457	35253	46.27
Phycomyces blakesleeanus NRRL 1555(-)	763407	Fungi	Mucoromycotina	55.85	35.72	14792	56260	31.32
Rhizopus oryzae RA 99-880	246409	Fungi	Mucoromycotina	46.09	35.60	17459	40515	38.98
Microsporidia								
Apansporoblastina								
Enterocytozoon bieneusi H348	**481877**	**Fungi**	**Microsporidia**	**3.86**	**33.70**	**3804**		**69.38**
Nosema ceranae BRL01	**578460**	**Fungi**	**Microsporidia**	**7.86**	**25.30**	**2124**		**25.79**
Encephalitozoon cuniculi GB-M1	**284813**	**Fungi**	**Microsporidiae**	**2.50**	**47.31**	**1996**	**0**	**86.39**
Encephalitozoon intestinalis ATCC 50506	876142	Fungi	Microsporidiae	2.01	41.53	1661	12	88.39
Nematocida parisi ERTm1	881290	Fungi	Microsporidiae	4.03	34.42	2668	9	72.16

Nematocida parisi ERTm2	944018	Fungi	Microsporidiae	4.65	38.34	2771	3	63.77
Nematocida parisi ERTm3	935791	Fungi	Microsporidiae	4.12	34.46	2726	8	69.22
Vavraia culicus	948595	Fungi	Microsporidiae	6.03	39.75	2780	4	46.90
Vittaforma corneae ATCC 50505	993615	Fungi	Microsporidiae	3.15	36.47	2248	5	66.99
Early diverging eukaryotes								
Monosiga brevicollis MX1	431895	Eukaryota	Choanoflagellida	41.63	54.89	9171	60511	39.73
Capsaspora Owczarzaki ATCC 30864	595528	Eukaryota	Ichthyosporea	27.97	53.82	8657	34000	58.71

Genomes shown in bold were used for genomic comparisons to C1A genome. All genomes (excluding the 2 early-diverging eukaryotic genomes *M. brevicollis*, and *C. owczarzaki*) were used for general genomic features comparisons shown in figures 1, and S2.

Table S6. Codon usage of strain C1A.

	T			**C**			**A**			**G**			
T	**Phe**	**222513**	**71%**	**Ser**	**123748**	**25%**	**Tyr**	**272297**	**81%**	**Cys**	**86836**	**78%**	**T**
	Phe	92195	29%	Ser	43322	9%	Tyr	65460	19%	Cys	24694	22%	**C**
	Leu	**325171**	**51%**	**Ser**	**158623**	**32%**	STOP	117583		STOP	63687		**A**
	Leu	89183	14%	Ser	18226	4%	STOP	37509		Trp	62565	100%	**G**
C	Leu	108453	17%	Pro	49391	26%	**His**	**102137**	**81%**	Arg	38442	17%	**T**
	Leu	25797	4%	Pro	9867	5%	His	23695	19%	Arg	3757	2%	**C**
	Leu	56194	9%	**Pro**	**122594**	**64%**	**Gln**	**188415**	**84%**	Arg	15915	7%	**A**
	Leu	30986	5%	Pro	10250	5%	Gln	36323	16%	Arg	5702	3%	**G**
A	**Ile**	**368709**	**52%**	**Thr**	**136253**	**43%**	**Asn**	**569337**	**86%**	**Ser**	**125877**	**25%**	**T**
	Ile	73551	10%	Thr	45815	15%	Asn	91768	14%	Ser	25415	5%	**C**
	Ile	**262623**	**37%**	**Thr**	**114251**	**36%**	**Lys**	**506477**	**77%**	**Arg**	**136988**	**60%**	**A**
	Met	176282	100%	Thr	18254	6%	Lys	155221	23%	Arg	26167	12%	**G**
G	**Val**	**143837**	**47%**	**Ala**	**99628**	**51%**	**Asp**	**282004**	**89%**	**Gly**	**114483**	**50%**	**T**
	Val	28636	9%	Ala	24755	13%	Asp	33617	11%	Gly	12215	5%	**C**
	Val	**96982**	**32%**	**Ala**	**62492**	**32%**	**Glu**	**374225**	**90%**	**Gly**	**89723**	**39%**	**A**
	Val	35875	12%	Ala	6799	4%	Glu	41111	10%	Gly	11008	5%	**G**

C1A codon usage pattern showed a bias towards AT-rich codons. Codon usage for the amino acids represented by only 2 codons (Phe, Tyr, His, Gln, Asn, Lys, Asp, Glu, and Cys) showed an apparent skew (70% or more) towards the AT-rich codon. Codon usage of Ile (represented by 3 codons) showed 90% usage skew to the 2 entirely AT-codons. Codon usage for amino acids represented by 4 codons (Val, Pro, Thr, Ala, and Gly) showed 80-90% usage of the 2 AT-rich codons (30-60% each). Arg and Leu (each represented by 6 codons) showed >50% usage skew to the codon with the highest AT content. And lastly, Ser (also represented by 6 codons) showed >80% usage to the 3 codons with the highest AT content.

Table S7. Transposable elements in the C1A genome.

TE class	Number of occurrences in C1A genome	Total length (bp)	Percentage genome coverage
Class I			
LTR			
Copia	2482	1,197,759	1.186
Gypsy	2752	1,533,722	1.519
Pao	49	5,972	0.006
Non-LTR			
LINEs			
L1	77	38,441	0.038
L2	76	28,505	0.028
CR1	3	492	0.0005
Rex	8	3,937	0.004
RTE	157	40,204	0.040
RTE-X	545	214,720	0.213
Class II			
hAT	9	1,926	0.002
MuDR	495	269,488	0.267
Polinton	11	1,552	0.002
Total	6664	3,336,718	3.307

Table S8. Simple sequence repeats (SSRs) of the C1A genome

Repeat Unit[a]		Genome	cDNA	Introns	Intergenic regions
Mononucleotide	Total number	170937	1018	14073	155210
	Total length	**3002239**	11228	273549	**2707450**
Dinucleotide	Total number	40102	81	6794	32857
	Total length	**678720**	1087	119492	**551668**
Trinucleotide	Total number	28346	2613	4023	21699
	Total length	**874957**	59802	150661	**666168**
Tetranucleotide	Total number	7202	0	982	5833
	Total length	**294312**	0	38817	**240399**
Pentanucleotide	Total number	1919	0	250	1565
	Total length	**75356**	0	10395	**60670**
Hexanucleotide	Total number	688	79	72	529
	Total length	**31411**	2960	3852	**24447**
Total number		249194	3791	26194	217693
Total length		**4956995**	75077	596766	4250802

[a]: Mononucleotide with at least 10 repeats; dinucleotides with at least 6 repeats; tri-, tetra-, penta- and hexanucleotides with at least 5 repeats were identified.

Table S9. Classification of C1A genes with no hits in Mycota (with bacterial or eukaryal orthologs), and with no hits in Dikarya (with basal fungal lineages orthologs)

	Proteins with no hits in Mycota		**Proteins with no hits in Dikarya**
Protein classification	**C1A proteins with bacerial orthologs**	**C1A proteins with eukaryal orthologs**	**C1A proteins with basal fungal orthologs**
Proteins with Panther class			
Isomerase	3	0	0
Ca-binding protein	0	4	1
Cell junction protein	0	0	1
Transmembrane receptor regulatory/adaptor protein	0	1	1
Defense/immunity protein	6	4	2
Cell adhesion molecule	2	0	3
Chaperone	2	1	4
Signaling molecule	0	4	4
Transfer/carrier protein	0	0	4
Viral protein	3	11	4
Structural protein	0	0	8
Oxidoreductase	4	9	9
Extracellular matrix protein	0	1	11
Ligase	4	3	12
Lyase	0	0	15
Phosphatase	22	9	20
Carbohydrate binding module	6	5	21
Protease	14	18	21
Membrane traffic protein	0	3	23
Transferase	34	9	28
Enzyme modulator	0	3	42
Receptor	0	26	44
Kinase	1	13	46
Transcription factor	0	5	48
Nucleic acid-binding protein	14	24	55
Cytoskeletal protein	0	8	82
Hydrolase	91	13	112
Transporter	63	9	244
Unclassified	125	145	341
Hypothetical proteins	236	250	313

Table S10. Axoneme-associated proteins of strain C1A

Axonemal protein	Subgroup/ Complex subunit	Gene ID/ IMG ID in Orpinomyces	Present in other basal fungi			Present in Dikarya Fungi
			Allomyces macrogynus ATCC 38327	Batrachochytrium dendrobatidis JEL423	Spizellomyces punctatus DAOM BR117	
A. Strcutural components						
1. Axonemal Dyneins	a. Heavy chain	Orpinomyces_1566	No	Yes	No	No
		Orpinomyces_16213	No	Yes	No	No
		Orpinomyces_16306	No	Yes	No	No
		Orpinomyces_16323	No	Yes	No	No
		Orpinomyces_16325	No	Yes	No	No
		Orpinomyces_16336	No	Yes	No	No
		Orpinomyces_16341	No	Yes	No	No
		Orpinomyces_16342	Yes	Yes	No	No
		Orpinomyces_1737	Yes	Yes	Yes	No
		Orpinomyces_3254	No	Yes	No	No
		Orpinomyces_4908	No	No	No	No
		Orpinomyces_6405	No	Yes	No	No
		Orpinomyces_8404	Yes	Yes	Yes	No
	b. intermediate chain	Orpinomyces_1377	Yes	Yes	Yes	No
	c. light chain	Orpinomyces_6968	Yes	Yes	Yes	No
2. Radial spokes		Orpinomyces_11986	Yes	Yes	Yes	No
		Orpinomyces_12260	Yes	Yes	Yes	No
		Orpinomyces_9559	Yes	Yes	Yes	No
3. Protein kinase A anchor protein		Orpinomyces_14681	No	No	No	No
4. Ribbon Rib43A		Orpinomyces_6911	Yes	Yes	Yes	No
5. Nexin-Dynein regulatory complex components		Orpinomyces_3201	No	Yes	Yes	No
		Orpinomyces_7282	Yes	Yes	Yes	No

6. Outer dynein arm docking complex component	protein 1	Orpinomyces_8906	Yes	Yes	Yes	No
7. Basal body assembly	a. B9 protein	Orpinomyces_9350	Yes	Yes	Yes	No
		Orpinomyces_8512	Yes	Yes	Yes	No
	b. Meckelin	Orpinomyces_8276	No	No	Yes	No
	c. FOP dimerization domain	Orpinomyces_12637	No	No	No	No
	d. Cep290	Orpinomyces_14841	Yes	Yes	Yes	No
B. IFT components						
1. Kinesin II	a. heterotrimeric	Orpinomyces_13919	Yes	Yes	Yes	No
		Orpinomyces_6530	Yes	Yes	Yes	No
	b. Kinesin-associated protein	Orpinomyces_2434	Yes	Yes	Yes	No
2. Cytoplasmic Dynein 2	a. heavy chain	Orpinomyces_3946	No	No	No	No
		Orpinomyces_16341	No	No	No	No
	b. intermediate chain	Orpinomyces_2387	No	Yes	Yes	No
	c. Light intermediate chain	Orpinomyces_904	Yes	Yes	Yes	No
	d. light chain	Orpinomyces_10954	Yes	Yes	Yes	No
3. IFT complex A	a. IFT144 (WDR19)	Orpinomyces_913	Yes	Yes	Yes	No
	b. IFT140	Orpinomyces_7315	Yes	Yes	Yes	No
	c. IFT139	Orpinomyces_2300	No	Yes	Yes	No
	d. IFT122 (WDR10)	Orpinomyces_9214	Yes	Yes	Yes	No
		Orpinomyces_15334	Yes	Yes	Yes	No

	e. IFT121 (WDR35)	2510866969	Yes	Yes	Yes	No
4. IFT complex B	a. OSM-1 (IFT172)	Orpinomyces_2822	No	Yes	Yes	No
	b. IFT88	Orpinomyces_5343	Yes	Yes	Yes	No
	c. IFT46	Orpinomyces_5553	Yes	Yes	Yes	No
	d. IFT57	Orpinomyces_1988	Yes	Yes	Yes	No
	e. IFT81	Orpinomyces_3636	Yes	Yes	Yes	No
	f. IFT80 (WDR56, CHE-2)	Orpinomyces_2046	Yes	Yes	Yes	No
	g. IFT74	Orpinomyces_4472	Yes	Yes	Yes	No
	h. IFT70	Orpinomyces_15501	Yes	Yes	Yes	No
	i. IFT54	Orpinomyces_12967	No	Yes	Yes	No
	j. IFT52	Orpinomyces_7475	Yes	Yes	Yes	No
	k. IFT27 (RABL4)	2510863227	Yes	Yes	Yes	Yes
	l. IFT25 (HSP-β11)	2510867340	No	Yes	Yes	No
	m. IFT22 (RABL5)	2510862755	Yes	Yes	Yes	Yes
	n. IFT20	2510856885	No	Yes	Yes	No
5. IFT complex A accessory	Tub family	Orpinomyces_15294	No	Yes	Yes	No
6. IFT complex B accessory	a. Clusterin-associated protein 1	Orpinomyces_10037	Yes	Yes	Yes	No
	b. DYF-13 (TTC26)	Orpinomyces_5282	Yes	Yes	Yes	No
		Orpinomyces_4873	Yes	Yes	Yes	No
7. Bardet-Biedl syndrome complex	a. BBS1	Orpinomyces_13448	No	No	No	No
	b. BBS2	Orpinomyces_12482	No	No	No	No
	c. BBS4	Orpinomyces_12384	No	No	No	No
	d. BBS5	2510866654	No	No	No	No
	e. BBS7	Orpinomyces_8448	No	No	No	No
	f. BBS8	Orpinomyces_2072	No	No	No	No
	g. BBS9	Orpinomyces_12104	No	No	No	No

Table S11. Protease inhibitors in the C1A genome

Gene ID	**Inhibitor family**	**Merops accession number**	**E-value**	**Annotation/ domains**	**Presence in other fungi**
Orpinomyces_14234	I04	MER018696	2.50E-39	serpin	*Piromyces* sp. E2
Orpinomyces_6138		MER018662	4.20E-38	IPR000215 Protease inhibitor I4, serpin IPR002883 Dockerin cellulose-binding IPR009034 Dockerin, cellulose docking	*Piromyces* sp. E2
Orpinomyces_2311		MER018662	6.00E-35	IPR000215 Protease inhibitor I4, serpin IPR002883 Dockerin cellulose-binding IPR009034 Dockerin, cellulose docking	*Piromyces* sp. E2
Orpinomyces_14684		MER019039	1.90E-12	serpin	*Piromyces* sp. E2
Orpinomyces_4452	I32	MER023234	6.60E-10	KOG1101 - Apoptosis inhibitor IAP1 and related BIR domain proteins	*Coccidioides immitis* RS, *Coccidioides posadasii* C735 delta SOWgp, *Coprinopsis cinerea* okayama7#130, *Fusarium oxysporum* lycopersici FGSC 4286, *Gibberella zeae* PH-1, *Nosema ceranae* BRL01, *Phanerochaete chrysosporium* RP-78, *Saccharomyces cerevisiae* AWRI1631, *Saccharomyces cerevisiae* YJM789, *Saitoella complicata* NRRL Y-17804, *Scheffersomyces stipitis* CBS 6054, *Spizellomyces punctatus* DAOM BR117, *Ustilago maydis* 521
2510865883	I42	MER090314	6.10E-11	Chagasin	*Allomyces macrogynus* ATCC 38327
2510871764		MER090314	2.60E-10	Chagasin	None
2510864005		MER047972	1.80E-09	Chagasin	None

Table S12. C1A proteins involved in hydrogenosomal protein import and their closest relatives in the nr database.

Gene ID/ IMG ID	**Protein product**	**Phylogenetic affiliation of Blastp first hit**	**Taxonomy of first hit**
2510863710	MIA complex componet; FAD dependent sulfhydryl oxidase Erv1	*Yarrowia lipolytica*	Fungi
2510870272	MIA complex componet; MIA40	*Yarrowia lipolytica*	Fungi
Orpinomyces_10445	Presequence Translocase-Associated Motor component Mge1	*Branchiostoma floridae*	Chordata
2510866045	Presequence Translocase-Associated Motor component mtHSP70	*Neocallimastix patriciarum*	Fungi
2510871141	Presequence Translocase-Associated Motor component mtHSP70	*Neocallimastix patriciarum*	Fungi
2510859873	Presequence Translocase-Associated Motor component PAM16	*Lachancea thermotolerans*	Fungi
2510856302	Presequence Translocase-Associated Motor component PAM18	*Drosophila erecta*	Arthropoda
2510870846	Presequence Translocase-Associated Motor component TIM44	*Schizophyllum commune* H4-8	Fungi
2510863027	SAM complex component; MDM10	*Batrachochytrium dendrobatidis* JAM81	Fungi
Orpinomyces_12490	SAM complex component; MDM12	*Aspergillus nidulans* FGSCA4	Fungi
Orpinomyces_12567	SAM complex component; Mim1	*Tuber melanosporum* Mel28	Fungi
Orpinomyces_9000	SAM complex component; Mim1	*Laccaria bicolor* S238N-H82	Fungi
Orpinomyces_593	SAM complex component; Sam37	*Acaryochloris marina* MBIC11017	Cyanobacteria
Orpinomyces_2572	SAM complex component; SAM50	*Tuber melanosporum* Mel28	Fungi
2510854209	small TIM; TIM10A	*Lachancea thermotolerans*	Fungi
2510854209	small TIM; TIM12	*Lachancea thermotolerans*	Fungi
Orpinomyces_7065	small TIM; TIM9A	*Laccaria bicolor* S238N-H82	Fungi
2510867063	TIM22 complex component; TIM18	*Tuber melanosporum* Mel28	Fungi
2510867256	TIM22 complex component; TIM18	*Tuber melanosporum* Mel28	Fungi
Orpinomyces_1577	TIM22 complex component; TIM22	*Tuber melanosporum* Mel28	Fungi
2510862636	TIM22 complex component; TIM54	*Trichophyton rubrum* CBS 118892	Fungi
2510859920	TIM23 complex component; TIM17	*Nematostella vectensis*	Cnidaria
2510874415	TIM23 complex component; TIM23	*Candida dubliniensis* CD36	Fungi
Orpinomyces_9320	TIM23 complex component; TIM50	*Coprinopsis cinerea* okayama7#130	Fungi
2510851173	TOM complex component; TOM20	*Coprinopsis cinerea* okayama7#130	Fungi
Orpinomyces_10011	TOM complex component; TOM20	*Coprinopsis cinerea* okayama7#130	Fungi
Orpinomyces_5414	TOM complex component; TOM22	*Ustilago maydis* 521	Fungi
Orpinomyces_14168	TOM complex component; TOM40	*Zea mays*	Plantae
2510863838	TOM complex component; TOM70	*Schizophyllum commune* H4-8	Fungi

Table S13. Hydrogenosomal proteins of strain C1A.

Gene ID/ IMG ID	**Product name**	**Pre-sequence (motif search)/ or Mitoprot cleavable sequence**	**Mitoprot prob**	**Comments**
Orpinomyces_2862	chaperonin GroEL	MLSAQRSL	0.996	Predicted using N-terminal sequence motifs then checked in mitoprot for probability of hydrogenosomal export. Pre sequences with W were removed. Motif sequence is shown.
Orpinomyces_10370	hypothetical protein	MSFFLFDDPIKESERSI	0.61	
2510855668	NADPH:quinone reductase and related Zn-dependent oxidoreductases	MLSSQTLIKNALARGI	0.996	
2510859021	Cytosine deaminase FCY1 and related enzymes	MLEITKESTQFMREA	0.744	
2510859416	hypothetical protein	MLFLLSATTRSI	0.73	
2510861873	Peptidase S26B, eukaryotic signal peptidase	MLLNSLFPGKTNREI	0.63	
Orpinomyces_5115	hypothetical protein	MLQKIIRNKLGLCLVALVLTPRSL	0.73	
Orpinomyces_8225	hypothetical protein	MSLFNKIDRAFG	0.66	
2510865002	hypothetical protein	MLYNRRFS	0.68	
2510865887	hypothetical protein	MLYKKNIRNL	0.595	
2510866001	Galactose binding lectin domain, D-galactoside/L-rhamnose binding SUEL lectin domain	MSLKNAMIITFNKIKRGA	0.97	
2510868169	ATPases involved in chromosome partitioning, nucleotide-binding.	MLRGF	0.898	
2510869114	Predicted ATPase, nucleotide-binding	MLRGF/ 0.95	0.82	
Orpinomyces_15932	SH3 domain.	MLRNSQNSTASSRSS	0.97	
Orpinomyces_16041	Ca2+-binding actin-bundling protein fimbrin/plastin (EF-Hand superfamily)	MLAQRRIFS	0.87	
2510872772	Guanine-nucleotide releasing factor	MLSRNI	0.68	

2510873335	Casein kinase (serine/threonine/tyrosine protein kinase)	MLRSI	0.7	
2510873379	Casein kinase (serine/threonine/tyrosine protein kinase)	MLRSI	0.7	
Orpinomyces_15557	dihydroxy-acid dehydratase, Dihydroxy-acid/6-phosphogluconate dehydratase, Aconitase/3-isopropylmalate dehydratase, swivel	MSLANIVAKRSL	0.94	
2510874047	acetolactate synthase, large subunit, biosynthetic type	MLSNTVKRSI	0.996	
2510874400	[FeFe] hydrogenase, group A	MLSSVLNKASANPKLTRSL	0.94	
2510870853	cysteine desulfurase	MQSLISNSIKSSIRLPTLSTVGKLGINQLKPTLFNSVRTFSRTSLTK	0.98	Enzymes of Fe-S cluster assembly. Probability predicted using mitoprot. Mitoprot predicted cleavable sequence is shown.
Orpinomyces_2766	iron-sulfur cluster assembly enzyme ISCU, mitochondrial-like	MNTLLNIAKPALVNSFKVQATRIIGTSTLNA	0.61	
2510861454	YAH1; mitochondrial matrix iron-sulfur protein	not predictable	0.99	
2510864168	frataxin, mitochondrial-like	not predictable	0.72	
2510856721	iron-sulfur cluster assembly protein IscA	not predictable	0.84	
Orpinomyces_10445	MGE1; mitochondrial grpE-type co-chaperone of the HSP70 system	not predictable	0.93	Chaperones and co-chaperones. Probability predicted using mitoprot. Mitoprot predicted cleavable sequence is shown.
Orpinomyces_9602	chaperonin GroEL, heat shock 60kDa protein 1	not predictable	0.78	
2510860701	chaperonin GroES, heat shock 10kDa protein 1	MFNSLVTRVNPIKMSLRF	0.74	
Orpinomyces_10834	mDj10; DnaJ homolog subfamily B member 12	MSSQITNSIPKVQAIVNASLINQAAHAVSPVSNRLNLHNSVKLNGKRNY	0.92	

Orpinomyces_1920	mitochondrial processing peptidase, beta	not predictable	0.87	Mitochondrial peptidases. Mitoprot predicted cleavable sequence is shown.
2510858001	mitochondrial intermediate peptidase	not predictable	0.69	
2510870042	Mitochondrial matrix Mmp37	not predictable	0.74	Mitochondrial matrix proteins. Probability predicted using mitoprot.
2510862602	succinate dehydrogenase, (ubiquinone) flavoprotein subunit	not predictable	0.83	Enzymes of pyruvate and energy metabolism. Probabilty predicted using mitoprot. Mitoprot predicted cleavable sequence is shown.
2510867063	succinate dehydrogenase (ubiquinone) membrane anchor subunit	MEIVNSNCGISNIRKIGSSILNS	0.67	
Orpinomyces_7860	NADH dehydrogenase (ubiquinone) flavoprotein 1	not predictable	0.96	
2510857800	malate dehydrogenase (oxaloacetate-decarboxylating)(NADP+)	MSTFNLIRNPLSNRG	0.692	
Orpinomyces_15581	malate dehydrogenase (oxaloacetate-decarboxylating)(NADP+)	MLAPIQTIARP	0.685	
2511052008	Pyruvate-formate lyase	not predictable	0.87	
Orpinomyces_7905	Acetyl-CoA hydrolase	MQALSNNLVKKTTFSGCQRGLASMSEAI	0.7	
2510848938	Succinyl-CoA synthetase, alpha subunit	MISNQVSSSAKVAASAVSKRLY	0.89	
2510871091	Succinyl-CoA synthetase, beta subunit	MLANVTRSTSKAAPALASIAQTAQKRFLSVHEYCS	0.94	
Orpinomyces_9924	Putative NADPH-quinone reductase (modulator of drug activity B)	not predictable	0.89	
2510867741	enoyl-[acyl carrier protein] reductase II	not predictable	0.87	
2510860831	Palmitoyl protein thioesterase	MKFNKFFIASFLIEASLAAKTLSKPKYRP	0.61	

Table S14. C1A proteins involved in Fe-S protein biogenesis.

Gene ID/ IMG ID	**Protein product**	**Comment**
Mitochondrial ISC machinery		
Orpinomyces_2048	Mrs3/4 (ferrous iron transporter)	Fe^{2+} import
2510870853	cysteine desulfurase NifS	Early ISC machinery. The genome lacks a homologue of Isd11. NifS is bacterial like.
Orpinomyces_2766	iron-sulfur cluster assembly enzyme ISCU, mitochondrial-like	
Orpinomyces_3968	YAH1; mitochondrial matrix iron-sulfur protein	
2510864168	frataxin, mitochondrial-like	
2510856721	iron-sulfur cluster assembly protein IscA	
2510856721	Fe-S cluster biosynthesis protein ISA1	
2510870984	Mitochondrial Fe-S cluster biosynthesis protein ISA2	
2510871141, 2510866045	Mitochondrial Hsp70, Ssq1	Late ISC macinery
2510855272	Fe-S protein assembly co-chaperone; Jac1	
Orpinomyces_10445	nucleotide exchange factor; Mge1	
2510859344	Monothiol gluatredoxin, Grx4	
Mitochondrial ISC export machinery		
Orpinomyces_11214, Orpinomyces_6161	Mitochondrial Fe/S cluster exporter Atm1, ABC superfamily	GSH aids in export of an unknown compound.
2510863710	FAD dependent sulfhydryl oxidase Erv1	
Cytosolic CIA machinery		
2510866123	P-loop containing nucleoside triphosphate hydrolases, Cfd1	Assembly and maturation of cytosolic Fe/S proteins.
2510867391	P-loop containing nucleoside triphosphate hydrolases, Nbp35	
Orpinomyces_13712	Nuclear architecture related protein, Nar-1	
Orpinomyces_10332	WD40 repeat protein Cia1	

Table S15. Transporters of strain C1A.

Classification	Superfamily/ Family	Number in transcriptome	Number in genome	Comments
1. ATP-dependent (primary active) transporters	The ATP-binding Cassette [ABC] Superfamily	211	103	Carbohydrate uptake , polar aa uptake, cellobiose and cellooligosaccharide uptake, Fe-chelate uptake, siderophore-Fe uptake, drug exporters, acetate exporter, multidrug resistance exporter (short chain FA phosphatidyl choline flippase), peroxisomal LCFA-CoA transporter, drug-conjugate transporter, heavy metal transporter, mitochondrial Fe-S cluster transporter.
	The H$^+$- or Na$^+$-translocating F-type, V-type and A-type ATPase [F-ATPase] Superfamily	10	4	H$^+$-translocating V-type ATPase, and vacuolar V-type ATPase
	The P-type ATPase [P-ATPase] Superfamily	70	7	Vacuolar Ca uptake, Golgi Ca, Mn efflux, ER Ca:H$^+$ antiporter, Ag ATPase, Cu ATPase, Golgi PS, PE translocase, PL uptake flippase, Inwardly directed PL and Lyso-PL flippase, Inwardly directed PC, PS, Lyso-PE flippase at the Plasma membrane involved in endocytosis, K$^+$ ATPase efflux, ER ATPase transporting PL
2. Ion channels/ pores	The Major Intrinsic Protein [MIP] Family	1	1	Water transporter positively regulated by PKA and PKC phosphorylation
	The Large Conductance Mechanosensitive Ion Channel [MscL] Family	2	1	Non-selective ion transport (slight bias for cations over anions)
	The Small Conductance Mechanosensitive Ion Channel [MscS] Family	2	0	Osmolyte efflux channel
	The Transient Receptor Potential Ca^{2+} Channel [TRP-CC] Family	100	85	Subfamilies C, V, A
	The Voltage-gated Ion Channel [VIC] Superfamily	34	15	Sodium, potassium, calcium, or non-selective
	The Mitochondrial and Plastid Porin [MPP] Family	2	1	VDAC large conductance chloride channel

	The CorA Metal Ion Transporter [MIT] Family	1	0	Al and Mn resistance
	The Ammonia Transporter Channel [Amt] Family	6	5	Ammonia uniporter both high and low affinity transporter
3. Secondary transporters	The Alanine or Glycine:Cation Symporter [AGCS] Family	1	1	Ala:Na symporter
	The Amino Acid-Polyamine-Organocation [APC] Family	2	2	Cystine transporter
	The Amino Acid/Auxin Permease [AAAP] Family	8	5	Symport with proton. Large neutral AA, glutamine, acidic AA.
	The Arsenite-Antimonite [ArsB] Efflux Family	1	1	Plasma membrane silicate efflux pump
	The Bile Acid:Na$^+$ Symporter [BASS] Family	1	0	Organic anion:Na symporter
	The Ca^{2+}:Cation Antiporter [CaCA] Family	4	1	Ca extrusion. Vacuolar Ca or Mn:H$^+$ antiporter, K$^+$-dependent Ca:Na antiporter.
	The Cation Diffusion Facilitator [CDF] Family	8	2	Heavy metal:proton or K$^+$ antiporter. Zn and Mn
	The Dicarboxylate/Amino Acid:Cation [Na$^+$ or H$^+$] Symporter [DAACS] Family	2	2	Asp:Na symporter, Ser/Thr:Na symporter
	The Divalent Anion:Na$^+$ Symporter [DASS] Family	5	0	Vacuolar PO$_4$ transporter, oxoglutarate:malate antiporter, tartarate:succinate antiporetr

	The Drug/Metabolite Transporter [DMT] Superfamily	23	7	Small multidrug resistance family (heterooligomeric drug efflux pump), triose phosphate transporter family (PEP), UDP-N-acetylglucosamine:UMP antiporter family (bifunctional Golgi nucleotide-sugar transporter), UDP-gal:UMP antiporter family (UDP-gal/UDP-glu:UDP antiporter in Golgi), CMP-sialate:CMP antiporter family (Golgi UDP-gal and UDP-NAcGal:UDP antiporter, Golgi UDP-NAcGlu transporter, ER/Golgi UDP-NAcGal/ UDP-NAcGlu transporter), GDP-mannose:GMP antiporter family (Golgi GDP-mannose transporter), GDP-fucose transporter family, Thiamine PP transporter family (DUF6 domain transporter)
	The Equilibrative Nucleoside Transporter [ENT] Family	1	1	Nucleoside transporters high and low affinity.
	The Formate-Nitrite Transporter [FNT] Family	1	1	Formate uptake permease
	The Glycoside-Pentoside-Hexuronide [GPH]:Cation Symporter Family	12	7	Symport with proton or sodium
	The Lysosomal Cystine Transporter [LCT] Family	1	0	Suppressor of the Lec15/Lec35 glycosylation mutation. Influence retention of lumenal ER proteins and glycosylation in Golgi.
	The Major Facilitator Superfamily [MFS]	84	52	Sugar:H^+ symporter, sugar uniporter, drug:H^+ antiporter, efflux pumps, polyol:H^+ symporter, nucleoside:H^+ symporter, amino acid uniporter, purine uniporter, oxalate:formate antiporter, acetyl-CoA:CoA antiporter.
	The Mitochondrial Carrier [MC] Family	26	7	ADP:ATP translocasr, hydrogenosomal ADP/ATP carrier, oxidocarboxylate carrier (C5-C7), Mitochondrial pH dependent Fe import, carnitine/acylcarnitine translocase, Ornithine/Arg carrier, mitochondrial folate transporter, dNDP transporter, peroxisomal ADP/AMP/ATP antiporter, citrate/oxoglutarate carrier, SAM importer.

	The Monovalent Cation:Proton Antiporter-1 [CPA1] Family	2	1	Na or K:H$^+$ antiporter
	The Monovalent Cation:Proton Antiporter-2 [CPA2] Family	2	0	K$^+$:H$^+$ antiporter
	The Multidrug/Oligosaccharidyl-lipid/Polysaccharide [MOP] Flippase Superfamily	13	9	Multiantimicrobial extrusion family (drug:Na antiporters), oligosaccharidyl-lipid flippase (Rft protein)
	The Neurotransmitter:Sodium Symporter [NSS] Family	5	3	Tryptophan:Na symporter
	The NhaC Na$^+$:H$^+$ Antiporter [NhaC] Family	4	2	Na$^+$:H$^+$ antiporter, Lysine transporter
	The Proton-dependent Oligopeptide Transporter [POT] Family	5	1	Di or tripeptide:H$^+$ symporter, cationic, neutral, and anionic peptide:H$^+$ symporter
	The Solute:Sodium Symporter [SSS] Family	1	1	Proline:Na symporter
	The Sulfate Permease [SulP] Family	6	1	Sulfate transporter, HCO$_3$:Cl exchange antiporter
	The Zinc [Zn^{2+}]-Iron [Fe^{2+}] Permease [ZIP] Family	11	9	Uptake dependent on proton motive force
	The Inorganic Phosphate Transporter [PiT] Family	0	1	Inorganic phosphate uptake coupled to Na symport
	The Chloride Carrier/Channel [ClC] Family	6	2	ClC voltage-gated Cl channel, intracellular endosomal Cl:H$^+$ antiporter
Total number		668	339	

Table S16. Sugar transporters of strain C1A

Family ID	**Gene ID /IMG ID**	**Sugar transported**	**Type of transport**	**Organismal type**	**BLAST result**	
					E-value	**First hit phylogeny**
I. Secondary transporters						
A. Major faciltator superfamily MFS (2.A.1)						
1. Subfamily (2.A.1.1) Sugar porters						
2.A.1.1.1	2511051723	Galactose, xylose	H^+ symport	Bacteria	2.00E-11	*Escherichia coli O6*
	2510859213				2.00E-19	*Bacillus subtilis*
2.A.1.1.2	Orpinomyces_2349	Arabinose, galactose, xylose	H^+ symport	Bacteria	7.00E-15	*Escherichia coli O157:H7*
2.A.1.1.3	2511056457	Xylose	H^+ symport	Bacteria	1.00E-122	*Escherichia coli O157:H7*
	2510870695				2.00E-120	*Escherichia coli O157:H7*
2.A.1.1.4	Orpinomyces_14384	Glucose	Uniport	Bacteria	5.00E-12	*Zymomonas mobilis*
2.A.1.1.5	Orpinomyces_11701	Hexose	Uniport	Yeast	7.00E-23	*Saccharomyces cerevisiae*
2.A.1.1.6	2511054810	Galactose, gucose, xylose	Uniport	Yeast	2.00E-08	*Saccharomyces cerevisiae*
2.A.1.1.19	Orpinomyces_14872	Low affinity glucose	Uniport	Yeast	2.00E-21	*Saccharomyces cerevisiae*
2.A.1.1.20	2511051114	Myoinositol	H^+ symport	Protozoa	4.00E-07	*Pyrenophora teres f. teres 0-1*
	Orpinomyces_15983				4.00E-07	*Pyrenophora teres f. teres 0-1*
2.A.1.1.21	2511051847	Hexose (glucose>fructose)	H^+ symport	Yeast	8.00E-18	*Saccharomyces cerevisiae*
2.A.1.1.22	2511057085	Hexose (fructose>glucose)	H^+ symport	Yeast	1.00E-06	*Schizosaccharomyces pombe*
2.A.1.1.23	2511051722	Gluconate	H^+ symport	Yeast	1.00E-05	*Schizosaccharomyces pombe*
	2511056106				9.00E-35	*Schizosaccharomyces pombe*
	2511059306				2.00E-12	*Arabidopsis thaliana*
	2510872284				2.00E-12	*Arabidopsis thaliana*
2.A.1.1.30	2511052453	Low affinity glucose	Uniport	Yeast	4.00E-52	*Saccharomyces cerevisiae*
	Orpinomyces_6091				2.00E-20	*Saccharomyces cerevisiae*
2.A.1.1.35	Orpinomyces_15569	Glucose, 2-deoxyglucose	Uniport	Bacteria	6.00E-07	*Arabidopsis thaliana*

2.A.1.1.39	2511052318	High affinity glucose	Uniport	Yeast	1.00E-18	*Candida albicans*
2.A.1.1.40	2511057329	Xylose	Uniport	Yeast	9.00E-16	*Saccharomyces cerevisiae*
	2510865509				2.00E-18	*Saccharomyces cerevisiae*
2.A.1.1.41	2511056052	D-xylose	H^+ symport	Bacteria	2.00E-13	*Arabidopsis thaliana*
	Orpinomyces_15569				2.00E-13	*Arabidopsis thaliana*
2.A.1.1.51	2511049304	Glucose, xylose	H^+ symport	Yeast	4.00E-41	*Kluyveromyces lactis*
	2511056275				7.00E-41	*Saccharomyces cerevisiae*
	2511060108				4.00E-47	*Saccharomyces cerevisiae*
	2510872025				8.00E-41	*Saccharomyces cerevisiae*
2.A.1.1.53	2510855196	Myoinositol	Uniport	Bacteria	4.00E-08	*Chlorella kessleri*
2.A.1.1.55	Orpinomyces_4098	L-arabinose, xylose, glucose	H^+ symport	Bacteria	3.00E-40	*Bacillus subtilis*
	2511056033				3.00E-35	*Bacillus subtilis*
	Orpinomyces_6091				1.00E-50	*Saccharomyces cerevisiae*
	2511059727				2.00E-16	*Bacillus subtilis*
	Orpinomyces_2876				3.00E-40	*Bacillus subtilis*
	Orpinomyces_8990				4.00E-31	*Arabidopsis thaliana*
	Orpinomyces_15353				3.00E-35	*Bacillus subtilis*
	Orpinomyces_15217				2.00E-33	*Bacillus subtilis*
	2510873278				2.00E-16	*Bacillus subtilis*
2.A.1.1.57	Orpinomyces_15235	High affinity glucose	H^+ symport	Fungi	7.00E-16	*Neurospora crassa*
2.A.1.1.58	2511048974	Low affinity glucose	H^+ symport	Fungi	2.00E-41	*Escherichia coli O157:H7*
	2511052287				8.00E-21	*Schizosaccharomyces pombe*
	2511052403				2.00E-43	*Kluyveromyces lactis*
	Orpinomyces_13994				5.00E-15	*Lactobacillus brevis*
	Orpinomyces_14872				2.00E-14	*Homo sapiens*
2.A.1.1.61	Orpinomyces_10621	High affinity monosaccharide (glucose, galactose, mannose, xylose)	H^+ symport	Plants	2.00E-24	*Arabidopsis thaliana*
2. Subfamily (2.A.1.4) Fucose:H+ symporter family (FHS)						

2.A.1.7.3	Orpinomyces_7618	Glucose, mannose, xylose	H^+ symport	Bacteria	2.00E-145	*Oribacterium sinus F0268*
	Orpinomyces_14909				2.00E-145	*Oribacterium sinus F0268*
B. Glycoside-Pentoside-Hexuronide:cation symporter family (GPH) (2.A.2)						
2.A.2.3.2	2511051033	Pentoside permease	Na^+ or H^+ symport	G^+ Bacteria	4.00E-11	*Bacillus subtilis*
	Orpinomyces_9871				2.00E-37	*Bacillus subtilis*
	2511054137				8.00E-14	*Bacillus subtilis*
	2511055215				7.00E-15	*Bacillus subtilis*
	2511062640				2.00E-06	*Bacillus subtilis*
	2511062641				3.00E-21	*Bacillus subtilis*
	2510863878				4.00E-24	*Bacillus subtilis*
	2510871901				4.00E-21	*Bacillus subtilis*
	2510872640				7.00E-15	*Bacillus subtilis*
	2510850345				3.00E-13	*Bacillus subtilis*
	2510863877				2.00E-06	*Bacillus subtilis*
	2510871900				2.00E-06	*Bacillus subtilis*
2.A.2.3.5	2510850800	Probable B-xyloside	Na^+ or H^+ symport	Bacteria	9.00E-09	*Bacillus subtilis*
2.A.2.6.1	Orpinomyces_10983	Maltose, sucrose	Na^+ or H^+ symport	Yeast	9.00E-55	*Schizosaccharomyces pombe*
	Orpinomyces_10983				1.00E-42	*Schizosaccharomyces pombe*
II. ATP-dependent primary active transporters						
A. The ATP-binding Cassette [ABC] Superfamily (3.A.1)						
1. The Carbohydrate Uptake Transporter-1 (CUT1) Family (3.A.1.1)						
3.A.1.1.17	Orpinomyces_1656	Trehalose, maltose, sucrose	ATP-dependent active transport	Proteobacteria	2.00E-36	*Ochrobactrum anthropi ATCC 49188*
	Orpinomyces_15808				5.00E-08	*Deinococcus radiodurans*
	2510874816				5.00E-08	*Deinococcus radiodurans*
	Orpinomyces_6050				3.00E-17	*Agrobacterium radiobacter K84*

	2510855885				4.00E-09	*Deinococcus radiodurans*
	Orpinomyces_5442				3.00E-10	*Chromohalobacter salexigens DSM 3043*
	Orpinomyces_15631				1.00E-10	*Brucella melitensis ATCC 23457*
	Orpinomyces_1570				2.00E-10	*Rhodospirillum centenum SW*
	2510870707				5.00E-07	*Deinococcus radiodurans*
3.A.1.1.19	2511054746	Palatinose (isomaltulose)	ATP-dependent active transport	Proteobacteria	2.00E-11	*Agrobacterium tumefaciens 5A*
	Orpinomyces_7769				5.00E-20	*Deinococcus radiodurans*
	2510867483				4.00E-33	*Deinococcus radiodurans*
	Orpinomyces_5584				6.00E-20	*Deinococcus radiodurans*
	2510850900				6.00E-12	*Deinococcus radiodurans*
	Orpinomyces_15365				4.00E-21	*Halanaerobium praevalens DSM 2228*
	2510856662				7.00E-09	*Mesorhizobium australicum WSM2073*
3.A.1.1.25	Orpinomyces_1401	Trehalose, maltose, sucrose, palatinose, low affinity glucose	ATP-dependent active transport	Bacteria	5.00E-13	*Desulfomonile tiedjei DSM 6799*
	Orpinomyces_1401				1.00E-12	*Desulfomonile tiedjei DSM 6799*
	Orpinomyces_5966				2.00E-13	*Deinococcus radiodurans*
2. The Peptide/Opine/Nickel Uptake Transporter (PepT) Family (3.A.1.5)						
3.A.1.5.9	2511050031	Cellobiose, cellooligosaccharide	ATP-dependent active transport	Archaea	1.00E-06	*Mus musculus*

Table S17. Proteases/Peptidases of strain C1A

Family/ subfamily	Comment	Presence in Fungi	Location
1. Aspartate protease			
A01A	Endopeptidases active at acidic pH	yes	The enzymes are mostly secreted from cells as inactive proenzymes that activate autocatalytically at acidic pH.
A11A	Endopeptidases processing retrotransposons-encoded polypeptides	yes	
A22A	Intramembrane endopeptidases component of gamma-secretase complex	Basal fungi	Intramembrane
A22B	Intramembrane endopeptidases component of gamma-secretase complex	Basal fungi	Intramembrane
2. Cysteine protease			
C01A	Endopeptidases of the lysosomal system	yes	Intracellular
C12	Ubiquitinyl hydrolases	yes	Intracellular
C13	Asparaginyl endopeptidases (lysosomal) and glycosylphosphatidylinositol:protein transamidase (ER)	yes	Intracellular
C14B	Caspases, cytosolic endopeptidases	yes	Intracellular
C15	Pyroglutamyl-peptidase, omega peptidases that release an N-terminal pyroglutamate residue	yes	Intracellular
C19	Ubiquitin-specific peptidase 14	yes	Intracellular
C26	Gamma-glutamyl hydrolase, lysosomal enzyme probably involved in the turnover of folyl poly-gamma-glutamates	yes	Intracellular
C44	Amidophosphoribosyltransferase self-processing precursor	yes	Intracellular
C48	SUMO (small ubiquitin-like modifier) deconjugating enzymes	yes	Intracellular
C50	Separase required for the separation of sister chromatids during mitosis	yes	Intracellular

C54	Endopeptidases with specificity for glycyl bonds. Activate and delipidate proteins important in a number of intracellular signalling pathways that involve transfer of proteins across membranes, including autophagy, intra-Golgi transport and receptor sorting	yes	Intracellular
C59	The only known peptidase activity is the autolytic cleavage of the penicillin V acylase precursor protein. The autolysis step produces a mature amidase with Cys1 as the nucleophile. All peptidase activity is lost on conversion to the mature peptidase. Exhibits great diversity in substrate specificity, to include enzymes such as choloyglycine hydrolases, acid ceramidases, isopenicillin N acyltransferases, and a subgroup of eukaryotic proteins with unclear function.	Only in *Gibberella zeae*	Possibly intracellular
C65	Isopeptidases that release ubiquitin from polyubiquitin	yes	Intracellular
C69	Dipeptidases and aminopeptidases. D-alanyl-D-alanine dipeptidase.	Only in *Gibberella zeae*	Possibly intracellular
C83	Exopeptidases. The synthesis of phytochelatins by the action of a gamma-glutamylcysteine dipeptidyl transpeptidase. Phytochelatins are the principle heavy-metal detoxifying compounds in the plant kingdom. They are linear polymers of the gamma-Glu-Cys portion of glutathione (GSH), with the general formula $(\text{Glu-Cys})_n$-Gly, where n= 2-11. The family also contains enzymes that hydrolyse the terminal Gly from glutathione (gamma-Glu-Cys-Gly) but lack the transferase activity necessary for the synthesis of phytochelatin. Glutathione is gamma-Glu-Cys-Gly, and its degradation by cleavage of the C-terminal Gly contrasts with the action of the gamma-glutamyltransfereses (e.g. T03.001) that remove the N-terminal residue.	*Puccinia graminis*, *Schizosaccharomyces japonicus*, *S. pombe*	Intracellular
C85	Deubiquitinylating peptidases	yes	Intracellular
C86	Deubiquitinylating peptidases	*Coprinopsis cinerea*, *Cryptococcus gattii*, *Filobasidiella neoformans*, *Laccaria bicolor*, *Malassezia*	Intracellular

		globosa, *Ustilago maydis*	
C88	Deubiquitinylating enzymes	yes	Intracellular
3. Metalloprotease			
M01	Aminopeptidases. The family also contains pyroglutamyl-peptidase II. Pyroglutamyl-peptidase II and aminopeptidase N are membrane-bound with N-terminal cytoplasmic and transmembrane domains.	yes	Intracellular. Some homologues have a signal peptide and transmembrane domain and so might be membrane bound.
M03A	Endopeptidase activity that is restricted to substrates of low molecule mass, and is therefore termed 'oligopeptidase'. One likely function of peptidases in family M3 is the intracellular degradation of oligopeptides. These could include cleaved signal peptides, and products of protein degradation.	yes	Intracellular
M13	Endopeptidase. Probably all peptidases in family M13 are restricted to acting on substrates of not more than about 40 residues. The enzymes act outside animal cells to degrade or convert polypeptide substrates.	yes	Extracellular
M14B	Metallocarboxypeptidases hydrolysing single, C-terminal amino acids from polypeptide chains.	yes	Intracellular
M16A	Zn metalloendopeptidase	yes	Intracellular
M16B	Mitochondrial processing peptidase beta-subunit	yes	Intracellular
M16C	Zn metalloendopeptidase	yes	Intracellular

M17	Leucyl aminopeptidase is cytosolic and involved in the breakdown of peptide products of intracellular proteinases.	*Batrachochytrium dendrobatidis*, *Coprinopsis cinerea*, *Laccaria bicolor* , *Malassezia globosa*, *Phanerochaete chrysosporium*, *Postia placenta*, *Puccinia graminis*, *Schizosaccharomyces japonicus*, *Schizosaccharomyces pombe*, *Serpula lacrymans*, *Sporisorium reilianum*, *Ustilago maydis*	Intracellular
M18	Aspartyl aminopeptidase is a cytosolic enzyme, and probably contributes to the catabolism of peptides including those produced by the proteasome.	yes	Intracellular
M20A	Glutamate carboxypeptidase. Exopeptidases. The peptidases hydrolyse the late products of protein degradation so as to complete the conversion of proteins to free amino acids.	yes	Intracellular
M20B	Peptidase T. exopeptidases. The peptidases hydrolyse the late products of protein degradation so as to complete the conversion of proteins to free amino acids.	yes	Intracellular
M20C	Xaa-His dipeptidase. Exopeptidases. The peptidases hydrolyse the late products of protein degradation so as to complete the conversion of proteins to free amino acids.	yes	Intracellular

M20D	Carboxypeptidase Ss1. Exopeptidases. The peptidases hydrolyse the late products of protein degradation so as to complete the conversion of proteins to free amino acids.	yes	Intracellular
M22	Endopeptidase that cleaves only proteins that are *O*-sialoglycosylated. Removal of the carbohydrate prevents cleavage	yes	Possibly intracellular
M23B	Endopeptidases that lyse bacterial cell wall peptidoglycans. The peptidases in the family cleave either the N-acylmuramoyl-Ala bond between the cell wall peptidoglycan and the cross-linking peptide or a bond within the cross-linking peptide.	yes	The peptidases are synthesized as precursors and are activated extracellularly
M24A	Methionyl aminopeptidase. Removal of the initiating N-terminal methionine from newly synthesized proteins acting co-translationally in association with the ribosomes.	yes	Intracellular
M24B	The X-Pro dipeptidase found in eukaryotes has a role in the cleavage of Xaa-Pro linkages found in dipeptides associated with collagen recycling.	yes	Intracellular
M28A	Aminopeptidase S. Glutaminyl-peptide cyclotransferase. Convert N-terminal glutamine to oxoproline. Catalyse the cyclization of free L-glutamine and N-terminal glutaminyl residues in proteins to pyroglutamate (5-oxoproline) and pyroglutamyl residues respectively.	yes	Possibly extracellular
M35	Fungal metalloendopeptidases	yes	Possibly extracellular
M36	Family M36 includes a number of endopeptidases from pathogenic fungi. These endopeptidases appear to be adapted for cleaving extracellular matrix proteins such as elastin and keratin	yes	Possibly extracellular
M41	ATP-dependent metalloendopeptidases. The physiological substrates of peptidases in family M41 are generally membrane proteins. The mitochondrial-AAA protease has a chaperone function important for the correct assembly of protein complexes in the mitochondrion, including elements of the respiratory chain and ATP-dependent enzymes. Both m-AAA and i-AAA endopeptidases are essential for the breakdown of uncomplexed components.	yes	Intracellular

M48A	Cleavage of prenylated proteins. CAAX prenyl protease 1. The peptide bond hydrolyzed can be designated -C-|-A-A-X in which C is an S-isoprenylated cysteine residue, A is usually aliphatic and X is the C-terminal residue of the substrate protein, and may be any of several amino acids. Subsequently, the S-isoprenylated cysteine residue that forms the new C-terminus is methyl-esterified and forms a hydrophobic membrane-anchor.	yes	Intracellular
M48B	May be involved in the degradation of abnormal proteins	yes	Intracellular
M50A	S2P peptidase	Basal fungi	Intramembrane
M67A	Isopeptidases that release ubiquitin from ubiquitinated proteins	yes	Intracellular
M67C	AMSH deubiquitinating peptidase. AMSH is associated with endosomes affecting the rate of turnover of receptor proteins which are ubiquitinated and sorted to the lysosome for degradation.	yes	Intracellular
4. Serine protease			
S01B	Glutamyl peptidase	yes	Intracellular
S08A	Serine endopeptidase subtilisin and its homologues. In bacteria, archaea and fungi, family members are probably involved in nutrition. Site-1 protease (S08.063) is a mammalian member of subfamily S8A which catalyzes the first step in the proteolytic activation of sterol regulatory element-binding protein.	yes	Most are secreted
S08B	In subfamily S8B, kexin (S08.070) processes the yeast alpha-mating factor and killer toxin precursors.	yes	Most are secreted
S09A	Prolyl oligopeptidase. Active mainly on oligopeptides. Important for the degradation of biologically active peptides	yes	Intracellular
S09B	Dipeptidyl-peptidase IV. Active mainly on oligopeptides. important for the degradation of biologically active peptides	yes	Intracellular
S09C	Acylaminoacyl-peptidase. Active mainly on oligopeptides. important for the degradation of biologically active peptides	yes	Intracellular
S09D	Glutamyl endopeptidase. Active mainly on oligopeptides. important for the degradation of biologically active peptides	yes	Intracellular

S10	Carboxypeptidases. peptidases are only active at acidic pH. The serine carboxypeptidases are synthesised as preproenzymes. Lysosomal carboxypeptidase A is essential for the correct assembly and function of the particle that contains beta-galactosidase and neuraminidase.	yes	Intracellular
S14	Endopeptidase Clp and its homologues. Proteasome components.	yes	Intracellular
S33	Prolyl aminopeptidase. Exopeptidases that act at the N-terminus of peptides. Prolyl aminopeptidase (S33.001) releases an N-terminal residue from a peptide, preferably (but not exclusively) a proline.	yes	Intracellular
S49	Signal peptide peptidase	Mainly bacterial. Found in fungi only in *Malassezia globosa, Ustilago maydis*. In eukaryotes found in plants, and protozoa.	In bacteria acts in the periplasm.
S54	Rhomboid intramembrane protease. Substrates include TGFalpha-like growth factor preprotein. Cleavage occurs in the Golgi, following translocation of the substrates from the endoplasmic reticulum membrane by Star, another transmembrane protein. The growth factors are then able to activate the epidermal growth factor receptor. In Saccharomyces cerevisiae the Pcp1 protein (S54.007) has a completely different function, and is a mitochondrial endopeptidase required for the activation of cytochrome c peroxidase and for the processing of the mitochondrial dynamin-like protein Mgm1.	yes	
S58	Aminopeptidase	yes	Intramembrane
S59	Autolytic endopeptidase. Autolytic processing to generate Nup98 and Nup96 is essential for correct targeting of these components to the nucleoplasmic side of the nuclear pore complex and for correct formation of the nuclear pore.	Yes	Intracellular

S66	Bacterial LD-carboxypeptidases. The physiological substrates are tetrapeptide peptidoglycan fragments that contain an L-configured residue (lysine or meso-diaminopimelic acid) to which is attached a C-terminal D-alanine residue. It is the bond to the C-terminal D-Ala residue that is hydrolysed.	Bacterial. Not in fungi but present in animals (*Acyrthosiphon pisum*, *Anopheles gambiae*), and plants	Intracellular
5. Threonine protease			
T01A	The component peptidases of the proteasome and related compound peptidases	Yes	Intracellular
T03	Gamma-glutamyltransferase 1. Peptidase family T3 contains self-processing proteins that express aminopeptidase as well as aminotransferase activities in their mature forms. Mammalian gamma-glutamyl transferase occurs as a heterodimeric enzyme that is highly glycosylated and processed in the endoplasmic reticulum and Golgi, and then becomes attached to the external surface of cell membranes. Its physiological function is the degradation of glutathione by cleavage of the gamma-glutamyl bond, either by hydrolysis or transpeptidation	Yes	Possibly extracellular
T05	Self-processing ornithine acetyltransferase precursor. Only the precursor of ornithine acetyltransferase has peptidase activity. The mature form of the enzyme released by the autoproteolytic cleavage of the precursor catalyses the reversible interchange of an acetyl group between N-acetylornithine and glutamate	Yes	Intracellular

Table S18. Carbohydrate-active enzymes (CAZymes) in the C1A genome

Family	**Gene ID**	**Associated domains**	**Degradation of Cell Wall Component**	**Predicted Fucntion**	**Top Hit phylogeny**	**Phylogeny color code***	**Top hit in Neocall**	
							Accession #	**Reference**
Glycoside hydrolases								
GH1	Orpinomyces_14808		Cellulose	β-glucosidase	Phytophthora infestans T30-4	🟩	AAP30745.1	(81)
GH1	Orpinomyces_16414		Cellulose	β-glucosidase	Picea sitchensis	🟨	AAP30745.1	(81)
GH1	Orpinomyces_16492		Cellulose	β-glucosidase	Camptotheca acuminate	🟨	AAD45834.1	(82)
GH1	Orpinomyces_16550		Cellulose	β-glucosidase	Magnaporthe oryzae 70-15	🟥	AAP30745.1	(81)
GH1	Orpinomyces_16576		Cellulose	β-glucosidase	Aspergillus clavatus NRRL 1	🟥	AAP30745.1	(81)
GH1	Orpinomyces_1791		Cellulose	β-glucosidase	Neosartorya fischeri NRRL 181	🟥	AAP30745.1	(81)
GH1	Orpinomyces_1810		Cellulose	β-glucosidase	Tribolium castaneum	🟫	CAC34952.1	(83)
GH1	Orpinomyces_3649		Cellulose	β-glucosidase	Camptotheca acuminate	🟨	AAD45834.1	(82)
GH1	Orpinomyces_8060		Cellulose	β-glucosidase	Neosartorya fischeri NRRL 181	🟥	AAD45834.1	(82)
GH2	Orpinomyces_14660		Hemicellulose	β-galactosidase	Ruminococcus flavefaciens FD-1	🟦		
GH2	Orpinomyces_14749		Hemicellulose	endo-1,4-β-mannosidase	Flammeovirga yaeyamensis	🟦		

GH2	Orpinomyces_4367		Hemicellulose	endo-1,4-β-mannosidase	Flammeovirga yaeyamensis	🟦		
GH2	Orpinomyces_9907		Hemicellulose	endo-1,4-β-mannosidase	Flammeovirga yaeyamensis	🟦		
GH3	Orpinomyces_11276		Cellulose	β-glucosidase	Roseburia hominis A2-183	🟦		
GH3	Orpinomyces_11433		Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AFI47447.1	(84)
GH3	Orpinomyces_11443		Cellulose	β-glucosidase	Haloplasma contractile SSD-17B	🟦		
GH3	Orpinomyces_1261		Cellulose	β-glucosidase	Roseburia intestinalis L1-82	🟦		
GH3	Orpinomyces_1293		Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AFI47447.1	(84)
GH3	Orpinomyces_14809		Cellulose	β-glucosidase	Rhizomucor miehei	🟥	AFI47447.1	(84)
GH3	Orpinomyces_15058		Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AFI47447.1	(84)
GH3	Orpinomyces_15232		Cellulose	β-glucosidase	Roseburia intestinalis L1-82	🟦		
GH3	Orpinomyces_15926		Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AFI47447.1	(84)
GH3	Orpinomyces_16425		Cellulose	β-glucosidase	Roseburia intestinalis L1-82	🟦		
GH3	Orpinomyces_16439		Cellulose	β-glucosidase	Roseburia intestinalis XB6B4	🟦		

GH3	Orpinomyces_16446		Cellulose	β-glucosidase	Roseburia intestinalis XB6B4	🟦		
GH3	Orpinomyces_16462		Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AFI47447.1	(84)
GH3	Orpinomyces_16466		Cellulose	β-glucosidase	Coprococcus comes ATCC 27758	🟦		
GH3	Orpinomyces_16467		Cellulose	β-glucosidase	Roseburia intestinalis L1-82	🟦		
GH3	Orpinomyces_16478		Cellulose	β-glucosidase	Roseburia intestinalis L1-82	🟦		
GH3	Orpinomyces_16497		Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AFI47447.1	(84)
GH3	Orpinomyces_16504		Cellulose	β-glucosidase	Roseburia intestinalis XB6B4	🟦		
GH3	Orpinomyces_16516		Cellulose	β-glucosidase	Roseburia intestinalis XB6B4	🟦		
GH3	Orpinomyces_16522		Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AFI47447.1	(84)
GH3	Orpinomyces_16534		Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AEX92706.1	(85)
GH3	Orpinomyces_16542		Cellulose	β-glucosidase	Glarea lozoyensis 74030	🟥	AFI47447.1	(84)
GH3	Orpinomyces_16568		Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AFI47447.1	(84)
GH3	Orpinomyces_16582		Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AEX92706.1	(85)

GH3	Orpinomyces_16587		Cellulose	β-glucosidase	Roseburia intestinalis L1-82	🟦		
GH3	Orpinomyces_16604		Cellulose	β-glucosidase	Roseburia intestinalis L1-82	🟦		
GH3	Orpinomyces_16608		Cellulose	β-glucosidase	Aspergillus nidulans FGSC A4	🟥	AEX92706.1	(85)
GH3	Orpinomyces_16619		Cellulose	β-glucosidase	Coprococcus comes ATCC 27758	🟦		
GH3	Orpinomyces_16631		Cellulose	β-glucosidase	Roseburia intestinalis XB6B4	🟦		
GH3	Orpinomyces_16736	Dock	Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AAO41704.1	(86)
GH3	Orpinomyces_2490	Dock	Cellulose	β-glucosidase	Phanerochaete chrysosporium	🟥	AEX92713.1	(85)
GH3	Orpinomyces_259		Cellulose	β-glucosidase	Roseburia intestinalis XB6B4	🟦		
GH3	Orpinomyces_3134		Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AEX92706.1	(85)
GH3	Orpinomyces_4249		Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AFI47447.1	(84)
GH3	Orpinomyces_4560		Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AAO41704.1	(86)
GH3	Orpinomyces_5676		Cellulose	β-glucosidase	Roseburia intestinalis L1-82	🟦		

GH3	Orpinomyces_6669		Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AEX92706.1	(85)
GH3	Orpinomyces_9603		Cellulose	β-glucosidase	Rhizopus oryzae RA 99-880	🟥	AEX92706.1	(85)
GH5	Orpinomyces_1071	Dock,	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Ktedonobacter racemifer DSM 44963	🟦		
GH5	Orpinomyces_11298	Dock, DUF3857	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Clostridiaceae bacterium AN-C16-KBRB	🟦	AAC06321.1	(87)
GH5	Orpinomyces_11551		Hemicellulose	endo-1,4-β-mannosidase	Bacillus clausii KSM-K16	🟦		
GH5	Orpinomyces_12843		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Clostridiaceae bacterium AN-C16-KBRB	🟦	AEO51792.1	cellulase [Orpinomyces sp. LT-3]
GH5	Orpinomyces_13832		Hemicellulose	endo-1,4-β-mannosidase	Thermotoga maritima MSB8	🟦	AAL01213.1	(88)
GH5	Orpinomyces_1444	Dock,	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Eubacterium siraeum V10Sc8a	🟦	Q12647.1	(89)
GH5	Orpinomyces_14471		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Acidothermus cellulolyticus 11B	🟦	AAL01213.1	(88)
GH5	Orpinomyces_1538		Hemicellulose	endo-1,4-β-mannosidase	Thermotoga neapolitana DSM 4359	🟦	AAL01213.1	(88)

GH5	Orpinomyces_16434		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Clostridiaceae bacterium AN-C16-KBRB	🟦	Q12647.1	(89)
GH5	Orpinomyces_16438		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Clostridiaceae bacterium AN-C16-KBRB	🟦	AEO51792.1	cellulase [Orpinomyces sp. LT-3]
GH5	Orpinomyces_16455		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Clostridiaceae bacterium AN-C16-KBRB	🟦	AEO51792.1	cellulase [Orpinomyces sp. LT-3]
GH5	Orpinomyces_16468		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Clostridiaceae bacterium AN-C16-KBRB	🟦	AAC05164.1	(90)
GH5	Orpinomyces_16486		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Clostridiaceae bacterium AN-C16-KBRB	🟦	Q12647.1	(89)
GH5	Orpinomyces_16514	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Ktedonobacter racemifer DSM 44963	🟦		
GH5	Orpinomyces_16616		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Clostridiaceae bacterium AN-C16-KBRB	🟦	AAC05164.1	(90)
GH5	Orpinomyces_16691	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Clostridium sp. BNL1100	🟦	AAC63094.1	(91)
GH5	Orpinomyces_16700	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Eubacterium siraeum DSM 15702	🟦		

GH5	Orpinomyces_16787	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Eubacterium siraeum DSM 15702	🟦		
GH5	Orpinomyces_204	Dock,	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Ruminococcus champanellensis 18P13	🟦	AAB69347.1	(90)
GH5	Orpinomyces_2345	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Paenibacillus curdlanolyticus YK9	🟦		
GH5	Orpinomyces_2510		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Clostridiaceae bacterium AN-C16-KBRB	🟦	Q12647.1	(89)
GH5	Orpinomyces_2543	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Eubacterium siraeum DSM 15702	🟦		
GH5	Orpinomyces_2605	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Clostridium sp. BNL1100	🟦	AAC63094.1	(91)
GH5	Orpinomyces_2784		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Clostridiaceae bacterium AN-C16-KBRB	🟦	AAC05164.1	(90)
GH5	Orpinomyces_3851		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Clostridiaceae bacterium AN-C16-KBRB	🟦	Q12647.1	(89)
GH5	Orpinomyces_3867	Polysaccharide deacetylase, AP_endonuc_2	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	bacterium Ellin514	🟦		

GH5	Orpinomyces_4162		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Eubacterium ventriosum ATCC 27560	🟦		
GH5	Orpinomyces_44		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Ruminococcus champanellensis 18P13	🟦	Q12647.1	(89)
GH5	Orpinomyces_5086		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Ruminococcus champanellensis 18P13	🟦	AAC05164.1	(90)
GH5	Orpinomyces_5811	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Ruminococcus champanellensis 18P13	🟦	AAC05164.1	(90)
GH5	Orpinomyces_5814		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Ruminococcus champanellensis 18P13	🟦	Q12647.1	(89)
GH5	Orpinomyces_6541	Dock, DUF3014	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Ralstonia syzygii R24	🟦		
GH5	Orpinomyces_7414	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Paenibacillus curdlanolyticus YK9	🟦		
GH5	Orpinomyces_8446	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Clostridiaceae bacterium AN-C16-KBRB	🟦	AAC06321.1	(87)
GH5	Orpinomyces_909	Dock,	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Ruminococcus champanellensis 18P13	🟦	AAD43818.1	(92)

GH5	Orpinomyces_9720	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Myceliophthora thermophila ATCC 42464	🟥	AAL01213.1	(88)
GH6	Orpinomyces_10295		Cellulose	exo-cellobiohydrolase (non-reducing end)	Neosartorya fischeri NRRL 181	🟥	AAF34679.1	(92)
GH6	Orpinomyces_10614		Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AAF34679.1	(92)
GH6	Orpinomyces_11639	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AEX92708.1	(85)
GH6	Orpinomyces_12853		Cellulose	exo-cellobiohydrolase (non-reducing end)	No hits found		AEX92716.1	(85)
GH6	Orpinomyces_1464	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Paenibacillus mucilaginosus K02	🟦	AEX92715.1	(85)
GH6	Orpinomyces_16419		Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AAL01211.1	(88)

GH6	Orpinomyces_16465	FRD2	Cellulose	exo-cellobiohydrolase (non-reducing end)	Neosartorya fischeri NRRL 181	🟥	AAF34679.1	(92)
GH6	Orpinomyces_16489		Cellulose	exo-cellobiohydrolase (non-reducing end)	Aspergillus clavatus NRRL 1	🟥	AAC09228.1	(93)
GH6	Orpinomyces_16505		Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AAF34679.1	(92)
GH6	Orpinomyces_16543	FRD2	Cellulose	exo-cellobiohydrolase (non-reducing end)	Neosartorya fischeri NRRL 181	🟥	AAF34679.1	(92)
GH6	Orpinomyces_16562		Cellulose	exo-cellobiohydrolase (non-reducing end)	Aspergillus niger	🟥	AAL01211.1	(88)
GH6	Orpinomyces_16626	ApoLp-III	Cellulose	exo-cellobiohydrolase (non-reducing end)	Neosartorya fischeri NRRL 181	🟥	AAO47726.1	Piromyces sp. KS11 cellulase
GH6	Orpinomyces_16627		Cellulose	exo-cellobiohydrolase (non-reducing end)	Phaeosphaeria nodorum SN15	🟥	AEX92708.1	(85)

GH6	Orpinomyces_16629		Cellulose	exo-cellobiohydrolase (non-reducing end)	Neosartorya fischeri NRRL 181	🟥	ABY52799.1	1,4-beta-D-glucan-cellobiohydrolase, partial [Piromyces rhizinflatus]
GH6	Orpinomyces_16639	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Paenibacillus mucilaginosus KNP414	🟦	AEX92708.1	(85)
GH6	Orpinomyces_16641	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AAL01211.1	(88)
GH6	Orpinomyces_16653	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Paenibacillus mucilaginosus KNP414	🟦	AEX92714.1	(85)
GH6	Orpinomyces_16656	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AEX92715.1	(85)
GH6	Orpinomyces_16776	Dock,	Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AEX92715.1	(85)

GH6	Orpinomyces_16924		Cellulose	exo-cellobiohydrolase (non-reducing end)	Neosartorya fischeri NRRL 181	🟥	AAR08200.1	CelA Neocallimastix patriciarum
GH6	Orpinomyces_2459		Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AAF34679.1	(92)
GH6	Orpinomyces_2901	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Pyrenophora tritici-repentis Pt-1C-BFP	🟥	AEX92715.1	(85)
GH6	Orpinomyces_3375		Cellulose	exo-cellobiohydrolase (non-reducing end)	No hits found		AEX92715.1	(85)
GH6	Orpinomyces_3449	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Xanthomonas oryzae pv. oryzae MAFF 311018	🟦	AEX92708.1	(85)
GH6	Orpinomyces_3530	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Xylella fastidiosa EB92.1	🟦	AEX92710.1	(85)
GH6	Orpinomyces_3583	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AEX92708.1	(85)

GH6	Orpinomyces_3813		Cellulose	exo-cellobiohydrolase (non-reducing end)	Lentinula edodes	🟥	AAR08200.1	CelA Neocallimastix patriciarum
GH6	Orpinomyces_3857	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AEX92708.1	(85)
GH6	Orpinomyces_4145	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AEX92708.1	(85)
GH6	Orpinomyces_4503		Cellulose	exo-cellobiohydrolase (non-reducing end)	Neosartorya fischeri NRRL 181	🟥	AAC09228.1	(93)
GH6	Orpinomyces_4642	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AEX92710.1	(85)
GH6	Orpinomyces_4748	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AEX92710.1	(85)
GH6	Orpinomyces_5216	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AEX92710.1	(85)

GH6	Orpinomyces_5544	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Aspergillus niger	🟥	AAM94167.1	(94)
GH6	Orpinomyces_6132		Cellulose	exo-cellobiohydrolase (non-reducing end)	Verticillium albo-atrum VaMs.102	🟥	AAC09228.1	(93)
GH6	Orpinomyces_6944	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AEX92708.1	(85)
GH6	Orpinomyces_7034	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Paenibacillus mucilaginosus K02	🟦	AEX92708.1	(85)
GH6	Orpinomyces_7615	Dock, Peptidase_M3	Cellulose	exo-cellobiohydrolase (non-reducing end)	Oreochromis niloticus	🟦	AEX92714.1	(85)
GH6	Orpinomyces_8084	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'	🟦	AEX92716.1	(85)
GH6	Orpinomyces_8266	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Paenibacillus mucilaginosus K02	🟦	AEX92710.1	(85)

GH6	Orpinomyces_8423	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Paenibacillus mucilaginosus K02	🟦	AEX92710.1	(85)
GH6	Orpinomyces_8802	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Verticillium dahliae VdLs.17	🟥	AAD51054.2	(92)
GH6	Orpinomyces_903	Dock	Cellulose	exo-cellobiohydrolase (non-reducing end)	Pyrenophora tritici-repentis Pt-1C-BFP	🟥	AEX92715.1	(85)
GH8	Orpinomyces_14432		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Hahella chejuensis KCTC 2396	🟦		
GH8	Orpinomyces_16605		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Butyrivibrio fibrisolvens 16/4	🟦		
GH9	Orpinomyces_16430		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Flavobacteriaceae bacterium HQM9	🟦	AAP30753.1	(81)
GH9	Orpinomyces_16648	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Kitasatospora setae KM-6054	🟦	AEX92719.1	(85)
GH9	Orpinomyces_16721	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Kitasatospora setae KM-6054	🟦	AAM81967.1	(95)

GH9	Orpinomyces_16732	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Kitasatospora setae KM-6054	🟦	AEX92719.1	(85)
GH9	Orpinomyces_16778	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Hahella chejuensis KCTC 2396	🟦	AEX92718.1	(85)
GH9	Orpinomyces_3361	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Kitasatospora setae KM-6054	🟦	AAM81966.1	(95)
GH9	Orpinomyces_3567	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Thermobifida fusca	🟦	AEX92718.1	(85)
GH9	Orpinomyces_5169	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Cellvibrio gilvu	🟦	AAM81966.1	(95)
GH9	Orpinomyces_5746		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Moorea producta 3L	🟦	AEX92719.1	(85)
GH9	Orpinomyces_6515		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	No hits found			
GH9	Orpinomyces_8987	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Kitasatospora setae KM-6054	🟦	AEX92719.1	(85)
GH9	Orpinomyces_9543	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Kitasatospora setae KM-6054	🟦	AAM81967.1	(95)

GH10	Orpinomyces_1160	Dock	Hemicellulose	endo-1,4-β-xylanase	Butyrivibrio proteoclasticus B316	🟦		
GH10	Orpinomyces_11830	Dock	Hemicellulose	endo-1,4-β-xylanase	Meiothermus ruber DSM 1279	🟦	AAB30669.1	(96)
GH10	Orpinomyces_13743		Hemicellulose	endo-1,4-β-xylanase	Meiothermus ruber DSM 1279	🟦	AAB30669.1	(96)
GH10	Orpinomyces_14446		Hemicellulose	endo-1,4-β-xylanase	Saccharomonospora paurometabolica YIM 90007	🟥	AAB30669.1	(96)
GH10	Orpinomyces_16413		Hemicellulose	endo-1,4-β-xylanase	Cellulomonas flavigena DSM 20109	🟦	AAB30669.1	(96)
GH10	Orpinomyces_16432	Dock	Hemicellulose	endo-1,4-β-xylanase	Streptomyces sviceus ATCC 29083	🟦	AAB30669.1	(96)
GH10	Orpinomyces_16474		Hemicellulose	endo-1,4-β-xylanase	Thermopolyspora flexuosa	🟦	AAB30669.1	(96)
GH10	Orpinomyces_16527		Hemicellulose	endo-1,4-β-xylanase	Saccharomonospora azurea NA-128	🟥	AAB30669.1	(96)
GH10	Orpinomyces_16669	Dock, Ricin_B_lectin	Hemicellulose	endo-1,4-β-xylanase	Streptomyces sviceus ATCC 29083	🟦	AAB30669.1	(96)
GH10	Orpinomyces_16687	Dock	Hemicellulose	endo-1,4-β-xylanase	Butyrivibrio proteoclasticus B316	🟦		

GH10	Orpinomyces_16690	Dock	Hemicellulose	endo-1,4-β-xylanase	Clostridium thermocellum DSM 2360	🟦	AAB30669.1	(96)
GH10	Orpinomyces_16698	Dock	Hemicellulose	endo-1,4-β-xylanase	Saccharomonospora paurometabolica YIM 90007	🟥	AAB30669.1	(96)
GH10	Orpinomyces_16766	Dock, Ricin_B_lectin	Hemicellulose	endo-1,4-β-xylanase	Eubacterium cellulosolvens 6	🟦	AEX92721.1	(85)
GH10	Orpinomyces_16857	Dock	Hemicellulose	endo-1,4-β-xylanase	Eubacterium cellulosolvens 6	🟦	AEX92721.1	(85)
GH10	Orpinomyces_16860	Dock	Hemicellulose	endo-1,4-β-xylanase	Butyrivibrio proteoclasticus B316	🟦		
GH10	Orpinomyces_16861	CBM1	Hemicellulose	endo-1,4-β-xylanase	Meiothermus ruber DSM 1279	🟦	AAB30669.1	(96)
GH10	Orpinomyces_16894	CBM1	Hemicellulose	endo-1,4-β-xylanase	No hits found			
GH10	Orpinomyces_16911	CBM1	Hemicellulose	endo-1,4-β-xylanase	Butyrivibrio proteoclasticus B316	🟦		
GH10	Orpinomyces_1983	Dock	Hemicellulose	endo-1,4-β-xylanase	Butyrivibrio proteoclasticus B316	🟦		
GH10	Orpinomyces_2071		Hemicellulose	endo-1,4-β-xylanase	Meiothermus ruber DSM 1279	🟦	AAB30669.1	(96)
GH10	Orpinomyces_5723	CBM1	Hemicellulose	endo-1,4-β-xylanase	Meiothermus ruber	🟦	AAB30669.1	(96)

GH10	Orpinomyces_5763	CBM1	Hemicellulose	endo-1,4-β-xylanase	Meiothermus ruber DSM 1279	🟦	AAB30669.1	(96)
GH10	Orpinomyces_5899	CBM1	Hemicellulose	endo-1,4-β-xylanase	Sreptomyces Lividans	🟦	AAB30669.1	(96)
GH10	Orpinomyces_6679	Dock	Hemicellulose	endo-1,4-β-xylanase	Butyrivibrio proteoclasticus B316	🟦		
GH10	Orpinomyces_6742	CBM1	Hemicellulose	endo-1,4-β-xylanase	No hits found		AAB30669.1	(96)
GH10	Orpinomyces_7494	CBM1	Hemicellulose	endo-1,4-β-xylanase	Verrucosispora maris AB-18-032	🟥	AAB30669.1	(96)
GH10	Orpinomyces_7686	CBM1	Hemicellulose	endo-1,4-β-xylanase	Butyrivibrio proteoclasticus B316	🟦		
GH10	Orpinomyces_808	Dock	Hemicellulose	endo-1,4-β-xylanase	Saccharomonospora paurometabolica YIM 90007	🟥	AAB30669.1	(96)
GH10	Orpinomyces_9123		Hemicellulose	endo-1,4-β-xylanase	Verrucosispora maris AB-18-032	🟥	AAB30669.1	(96)
GH10	Orpinomyces_938	CBM1	Hemicellulose	endo-1,4-β-xylanase	Actinomadura sp.	🟦	AAB30669.1	(96)
GH11	Orpinomyces_10724	Dock	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	ACL68347.1	(97)

GH11	Orpinomyces_11518		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)
GH11	Orpinomyces_1220	Dock	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	Q12667.1	(99)
GH11	Orpinomyces_13040		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	CAA57820.1	(100)
GH11	Orpinomyces_13274		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	CAA57820.1	(100)
GH11	Orpinomyces_1396	Dock	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	Q12667.1	(99)
GH11	Orpinomyces_15347		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAG18439.1	Piromyces communis xylanase
GH11	Orpinomyces_16423		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)

GH11	Orpinomyces_16454		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)
GH11	Orpinomyces_16458		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)
GH11	Orpinomyces_16510	Dock, Polysaccharide deacetylase, DUF915, Esterase_phd, Peptidase_S9	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	ACL68347.1	(97)
GH11	Orpinomyces_16518		Hemicellulose	endo-1,4-β-xylanase	No hits found		AAF14365.1	(98)
GH11	Orpinomyces_16520	Polysaccharide deacetylase, Herpes_BLLF1	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAG18439.1	Piromyces communis xylanase
GH11	Orpinomyces_16525		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)

GH11	Orpinomyces_16526		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)
GH11	Orpinomyces_16531		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAG18439.1	Piromyces communis xylanase
GH11	Orpinomyces_16563		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	CAA57820.1	(100)
GH11	Orpinomyces_16564		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)
GH11	Orpinomyces_16572		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAG18439.1	Piromyces communis xylanase
GH11	Orpinomyces_16574		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	Q12667.1	(99)
GH11	Orpinomyces_16625		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)

GH11	Orpinomyces_16651	Dock, Polysaccharide deacetylase	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	Q12667.1	(99)
GH11	Orpinomyces_16755	Dock	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	CAA57820.1	(100)
GH11	Orpinomyces_16793	Dock	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)
GH11	Orpinomyces_16866	CBM1	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)
GH11	Orpinomyces_16874	CBM1	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	Q12667.1	(99)
GH11	Orpinomyces_16909	CBM1	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)
GH11	Orpinomyces_16919	CBM1	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)

GH11	Orpinomyces_16926	CBM1	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)
GH11	Orpinomyces_16927	CBM1	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)
GH11	Orpinomyces_16928	CBM1	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)
GH11	Orpinomyces_16937	CBM1	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)
GH11	Orpinomyces_16938	CBM1	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	CAA57820.1	(100)
GH11	Orpinomyces_1946		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAG18439.1	Piromyces communis xylanase
GH11	Orpinomyces_2012		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)

GH11	Orpinomyces_2909		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAG18439.1	Piromyces communis xylanase
GH11	Orpinomyces_3040		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	Q12667.1	(99)
GH11	Orpinomyces_3043		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	CAA57820.1	(100)
GH11	Orpinomyces_4849		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	CAA57820.1	(100)
GH11	Orpinomyces_4868	CBM1	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)
GH11	Orpinomyces_5426		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	CAA57820.1	(100)
GH11	Orpinomyces_5512		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAG18439.1	Piromyces communis xylanase

GH11	Orpinomyces_5933		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	CAA57820.1	(100)
GH11	Orpinomyces_6955	CBM1	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	AAF14365.1	(98)
GH11	Orpinomyces_7134	Polysaccharide deacetylase	Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	CAA57820.1	(100)
GH11	Orpinomyces_7606		Hemicellulose	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85	🟦	CAA57820.1	(100)
GH13	Orpinomyces_11877		Starch	α-Amylase	Botryotinia fuckeliana	🟥		
GH13	Orpinomyces_16426		Starch	α-Amylase	Streptococcus gallolyticus subsp. gallolyticus ATCC 43143	🟦		
GH13	Orpinomyces_16441		Starch	α-Amylase	Paenibacillus sp. Aloe-11	🟦		
GH13	Orpinomyces_16495		Starch	α-Amylase	Grimontia hollisae CIP 101886	🟦		

GH13	Orpinomyces_16503		Starch	α-Amylase	Streptococcus dysgalactiae subsp. equisimilis SK1250	🟦		
GH13	Orpinomyces_16506	DUF3830	Starch	α-Amylase	Bacillus subtilis subsp. subtilis str. 168	🟦		
GH13	Orpinomyces_3185		Starch	α-Amylase	Bacillus sp. KR-8104	🟦		
GH13	Orpinomyces_3241		Starch	α-Amylase	Pyrenophora tritici-repentis Pt-1C-BFP	🟥		
GH13	Orpinomyces_3479		Starch	α-Amylase	Bacillus atrophaeus 1942	🟦		
GH13	Orpinomyces_5526	CBM25	Starch	α-Amylase	Streptococcus gallolyticus subsp. gallolyticus ATCC BAA-2069	🟦		
GH13	Orpinomyces_5986	CBM25	Starch	α-Amylase	Streptococcus gallolyticus subsp. gallolyticus TX20005	🟦		
GH13	Orpinomyces_675	CBM48	Starch	α-Amylase	Neurospora crassa OR74A	🟥		
GH13	Orpinomyces_7289		Starch	α-Amylase	Paenibacillus sp. Aloe-11	🟦		
GH16	Orpinomyces_14946		Hemicellulose	β-(1,3-1,4)-glucanase (Lichenase)	Maricaulis maris MCS10	🟦		

GH16	Orpinomyces_16436		Hemicellulose	β-(1,3-1,4)-glucanase (Lichenase)	Maricaulis maris MCS10	🟦		
GH16	Orpinomyces_16490		Hemicellulose	β-(1,3-1,4)-glucanase (Lichenase)	Maricaulis maris MCS10	🟦		
GH16	Orpinomyces_16614		Hemicellulose	β-(1,3-1,4)-glucanase (Lichenase)	Streptococcus equinus	🟦	AAQ09257.1	lichenase [Anaeromyces sp. W-98]
GH16	Orpinomyces_16615		Hemicellulose	β-(1,3-1,4)-glucanase (Lichenase)	Streptococcus equinus	🟦	AAQ09257.1	lichenase [Anaeromyces sp. W-98]
GH16	Orpinomyces_16635		Hemicellulose	β-(1,3-1,4)-glucanase (Lichenase)	Streptococcus equinus	🟦	AAQ09257.1	lichenase [Anaeromyces sp. W-98]
GH16	Orpinomyces_16636		Hemicellulose	β-(1,3-1,4)-glucanase (Lichenase)	Streptococcus equinus	🟦	AAQ09257.1	lichenase [Anaeromyces sp. W-98]
GH16	Orpinomyces_3308		Hemicellulose	β-(1,3-1,4)-glucanase (Lichenase)	Maricaulis maris MCS10	🟦		
GH16	Orpinomyces_5528		Hemicellulose	β-(1,3-1,4)-glucanase (Lichenase)	Chitinophaga pinensis DSM 2588	🟦		
GH18	Orpinomyces_10945	CBM18	Chitin	Chitinase	Paenibacillus elgii B69	🟦		
GH18	Orpinomyces_13782		Chitin	Chitinase	Bacillus sp. m3-13	🟦		

GH18	Orpinomyces_16580		Chitin	Chitinase	Pandora neoaphidis	🟥		
GH18	Orpinomyces_16609		Chitin	Chitinase	Batrachochytrium dendrobatidis JAM81	🟥		
GH18	Orpinomyces_16634		Chitin	Chitinase	Desmospora sp. 8437	🟦		
GH18	Orpinomyces_16683	Dock	Chitin	Chitinase	No hits found			
GH18	Orpinomyces_8445	CBM18	Chitin	Chitinase	Metarhizium acridum CQMa 102	🟦		
GH25	Orpinomyces_6304		Peptidoglycan	Lysozyme	Lachnospiraceae bacterium 3_1_57FAA_CT1	🟦		
GH26	Orpinomyces_16411		Hemicellulose	endo-1,4-β-mannosidase	Dictyoglomus thermophilum H-6-12	🟦		
GH26	Orpinomyces_16586		Hemicellulose	endo-1,4-β-mannosidase	Fibrobacter succinogenes subsp. succinogenes S85	🟦		
GH26	Orpinomyces_16611	DUF1208, Hydrolase_3, PepSY, Rho_GDI	Hemicellulose	endo-1,4-β-mannosidase	Neosartorya fischeri NRRL 181	🟥	AAC09228.1	(93)
GH26	Orpinomyces_16850	Dock,	Hemicellulose	endo-1,4-β-mannosidase	Acetivibrio cellulolyticus CD2	🟦	P55297.1	(99)

GH26	Orpinomyces_5113		Hemicellulose	endo-1,4-β-mannosidase	Dysgonomonas gadei ATCC BAA-286	🟦	P55296.1	(99)
GH31	Orpinomyces_10928		Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Trichoplax adhaerens	🟫		
GH31	Orpinomyces_1225	Trefoil	Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Amphimedon queenslandica	🟫		
GH31	Orpinomyces_13033		Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	No hits found			
GH31	Orpinomyces_15985	Trefoil	Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Mortierella alliacea	🟥		
GH31	Orpinomyces_16178	Trefoil	Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Amphimedon queenslandica	🟫		
GH31	Orpinomyces_16227	Trefoil	Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Amphimedon queenslandica	🟫		

GH31	Orpinomyces_16421	Trefoil	Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Amphimedon queenslandica	🟫		
GH31	Orpinomyces_16431	Trefoil	Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Podospora anserina S mat+	🟥		
GH31	Orpinomyces_16442		Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Mycosphaerella graminicola IPO323	🟥		
GH31	Orpinomyces_16498	Trefoil	Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Mortierella alliacea	🟥		
GH31	Orpinomyces_16517	Trefoil	Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Trichoplax adhaerens	🟫		
GH31	Orpinomyces_16528		Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Amphimedon queenslandica	🟫		
GH31	Orpinomyces_16537		Starch, (1→4)-linked α-D-glucose	α-glucosidase	Schizosaccharomyces pombe	🟥		

			oligosaccharides			🟥		
GH31	Orpinomyces_16541		Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Schizosaccharomyces pombe	🟥		
GH31	Orpinomyces_16548		Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Schizosaccharomyces pombe	🟥		
GH31	Orpinomyces_2368		Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Trichoplax adhaerens	🟫		
GH31	Orpinomyces_3125		Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Schizosaccharomyces pombe	🟥		
GH31	Orpinomyces_4334	Trefoil, Mucin	Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Amphimedon queenslandica	🟫		
GH31	Orpinomyces_8795	Trefoil	Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Trichoplax adhaerens	🟫		

GH31	Orpinomyces_9302	Trefoil	Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Trichoplax adhaerens	🟥		
GH31	Orpinomyces_9628		Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Trichoplax adhaerens	🟥		
GH31	Orpinomyces_9629	Trefoil	Starch, (1→4)-linked α-D-glucose oligosaccharides	α-glucosidase	Amphimedon queenslandica	🟥		
GH32	Orpinomyces_16603		Other (fructose-containing polysaccharides)	β-fructosidase	Butyrivibrio proteoclasticus B316	🟦		
GH32	Orpinomyces_1931	Cadherin-like	Other (fructose-containing polysaccharides)	Unknown	No hits found			
GH38	Orpinomyces_16429		Other	α-mannosidase	No hits found			
GH39	Orpinomyces_16452		Hemicellulose	β-xylosidase	Clostridium sp. DL-VIII	🟦		
GH39	Orpinomyces_16509		Hemicellulose	β-xylosidase	Clostridium sp. DL-VIII	🟦		

GH39	Orpinomyces_16769	Dock, Ricin-B-lectin	Hemicellulose	β-xylosidase	Clostridium sp. DL-VIII	🟦		
GH39	Orpinomyces_2939		Hemicellulose	β-xylosidase	Clostridium sp. DL-VIII	🟦		
GH43	Orpinomyces_1088	Dock, Ricin_B_lectin	Hemicellulose	alpha-N-arabinofuranosidase	Clostridium lentocellum DSM 5427	🟦	AEX92721.1	(85)
GH43	Orpinomyces_13498		Hemicellulose	β-xylosidase	Eubacterium cellulosolvens 6	🟦		
GH43	Orpinomyces_13906		Hemicellulose	β-xylosidase	Cellovibrio japonicus Ueda107	🟦		
GH43	Orpinomyces_14143		Hemicellulose	endo-1,4-β-xylanase	Cellulosilyticum ruminicola	🟦		
GH43	Orpinomyces_15144		Hemicellulose	β-xylosidase	Prevotella ruminicola 23	🟦		
GH43	Orpinomyces_16449		Hemicellulose	alpha-N-arabinofuranosidase	Clostridium lentocellum DSM 5427	🟦		
GH43	Orpinomyces_16453		Hemicellulose	endo-1,4-β-xylanase	Cellulosilyticum ruminicola	🟦		
GH43	Orpinomyces_16512	Ricin_B_lectin	Hemicellulose	Alpha-N-arabinofuranosidase	Ruminococcus flavefaciens FD-1	🟦		
GH43	Orpinomyces_16539		Hemicellulose	hypothetical protein	Spirochaeta thermophila DSM 6192	🟦		

GH43	Orpinomyces_16570		Hemicellulose	alpha-N-arabinofuranosidase	Clostridium lentocellum DSM 5427	🟦		
GH43	Orpinomyces_16579		Hemicellulose	β-xylosidase/ arabinofuranosidase	Butyrivibrio proteoclasticus B316	🟦		
GH43	Orpinomyces_16597		Hemicellulose	β-xylosidase/ arabinofuranosidase	Butyrivibrio proteoclasticus B316	🟦		
GH43	Orpinomyces_16617		Hemicellulose	β-xylosidase/ arabinofuranosidase	Butyrivibrio proteoclasticus B316	🟦		
GH43	Orpinomyces_16620		Hemicellulose	β-xylosidase/ arabinofuranosidase	Butyrivibrio proteoclasticus B316	🟦		
GH43	Orpinomyces_16799	Dock, Ricin_B_lectin	Hemicellulose	alpha-N-arabinofuranosidase	Clostridium lentocellum DSM 5427	🟦	AEX92721.1	(85)
GH43	Orpinomyces_16828	Dock, Ricin_B_lectin	Hemicellulose	alpha-N-arabinofuranosidase	Clostridium lentocellum DSM 5427	🟦	AEX92721.1	(85)
GH43	Orpinomyces_16879		Hemicellulose	alpha-N-arabinofuranosidase	Clostridium lentocellum DSM 5427	🟦	AEX92721.1	(85)
GH43	Orpinomyces_16895		Hemicellulose	No hits found	No hits found			
GH43	Orpinomyces_2136		Hemicellulose	alpha-N-arabinofuranosidase	Ruminococcus flavefaciens FD-1	🟦		

GH43	Orpinomyces_2841		Hemicellulose	alpha-N-arabinofuranosidase	Clostridium lentocellum DSM 5427	🟦	AEX92721.1	(85)
GH43	Orpinomyces_3146	Dock, Ricin_B_lectin/ Trypan_PARP	Hemicellulose	alpha-N-arabinofuranosidase	Clostridium lentocellum DSM 5427	🟦	AEX92721.1	(85)
GH43	Orpinomyces_3639		Hemicellulose	alpha-N-arabinofuranosidase	Clostridium lentocellum DSM 5427	🟦	AEX92721.1	(85)
GH43	Orpinomyces_4292		Hemicellulose	alpha-N-arabinofuranosidase	Clostridium lentocellum DSM 5427	🟦		
GH43	Orpinomyces_5980		Hemicellulose	putative lipoprotein	Ruminococcus albus 8	🟦		
GH43	Orpinomyces_6516		Hemicellulose	alpha-N-arabinofuranosidase	Ruminococcus flavefaciens FD-1	🟦		
GH43	Orpinomyces_843		Hemicellulose	β-xylosidase	Amycolatopsis mediterranei U32	🟦		
GH43	Orpinomyces_8634	Ricin_B_lectin	Hemicellulose	alpha-N-arabinofuranosidase	Clostridium lentocellum DSM 5427	🟦	AEX92721.1	(85)
GH45	Orpinomyces_16477	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Syncephalastrum racemosum	🟥	CAB92325.1	(101)
GH45	Orpinomyces_16488		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	No hits found			

GH45	Orpinomyces_16521		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae RA 99-880	🟥	CAB92325.1	(101)
GH45	Orpinomyces_16533	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae RA 99-880	🟥	CAB92325.1	(101)
GH45	Orpinomyces_16613		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	No hits found			
GH45	Orpinomyces_16658	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae RA 99-880	🟥	CAB92325.1	(101)
GH45	Orpinomyces_16718	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	No hits found		CAB92325.1	(101)
GH45	Orpinomyces_16773	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae RA 99-880	🟥	CAB92325.1	(101)
GH45	Orpinomyces_1749		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae	🟥	CAB92325.1	(101)
GH45	Orpinomyces_2252	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	No hits found		CAB92325.1	(101)
GH45	Orpinomyces_4393		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae	🟥		

GH45	Orpinomyces_524	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae RA 99-880	🟥	CAB92325.1	(101)
GH45	Orpinomyces_5778		Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae RA 99-880	🟥	CAB92325.1	(101)
GH45	Orpinomyces_6562	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae RA 99-880	🟥	CAB92325.1	(101)
GH45	Orpinomyces_7825	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	No hits found		CAB92325.1	(101)
GH45	Orpinomyces_7870	Dock	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae RA 99-880	🟥	CAB92325.1	(101)
GH47	Orpinomyces_16599		Other	alpha-mannosidase(ER)	Ixodes scapularis	🟫		
GH47	Orpinomyces_2189		Other	Unknown (Conserved hypothetical protein)	Schizophyllum commune H4-8	🟥		
GH48	Orpinomyces_10796	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	Streptomyces bingchenggensis BCW-1	🟦	AEX92722.1	(85)

GH48	Orpinomyces_1263		Cellulose	exo-cellobiohydrolase (reducing end)	Streptosporangium roseum DSM 43021	🟦	AEX92722.1	(85)
GH48	Orpinomyces_15099		Cellulose	exo-cellobiohydrolase (reducing end)	Streptosporangium roseum DSM 43021	🟦	AEX92722.1	(85)
GH48	Orpinomyces_16424		Cellulose	exo-cellobiohydrolase (reducing end)	Streptosporangium roseum DSM 43021	🟦	AEX92722.1	(85)
GH48	Orpinomyces_16646	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	Catenulispora acidiphila DSM 44928	🟦	AAN76735.1	(102)
GH48	Orpinomyces_16647	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	No hits found		AEX92722.1	(85)
GH48	Orpinomyces_16679	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	Streptomyces hygroscopicus subsp. jinggangensis 5008	🟦	AEX92722.1	(85)
GH48	Orpinomyces_16682	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	Streptomyces bingchenggensis BCW-1	🟦	AEX92722.1	(85)

GH48	Orpinomyces_16688	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	No hits found		AEX92722.1	(85)
GH48	Orpinomyces_16737	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	Streptomyces griseoaurantiacus M045	🟦	AAN76734.1	(102)
GH48	Orpinomyces_16738	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	Streptomyces griseoaurantiacus M045	🟦	AAN76734.1	(102)
GH48	Orpinomyces_16739	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	Streptomyces griseoaurantiacus M045	🟦	AEX92722.1	(85)
GH48	Orpinomyces_16756	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	Catenulispora acidiphila DSM 44928	🟦	AEX92722.1	(85)
GH48	Orpinomyces_16770	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	Catenulispora acidiphila DSM 44928	🟦	AAN76734.1	(102)
GH48	Orpinomyces_16779	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	Catenulispora acidiphila DSM 44928	🟦	AEX92722.1	(85)

GH48	Orpinomyces_16795	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	Streptomyces griseoaurantiacus M045	🟦	AEX92722.1	(85)
GH48	Orpinomyces_4269	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	Streptomyces griseoaurantiacus M045	🟦	AAN76734.1	(102)
GH48	Orpinomyces_7252	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	Streptosporangium roseum DSM 43021	🟦	AAN76735.1	(102)
GH48	Orpinomyces_7411	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	Ktedonobacter racemifer DSM 44963	🟦	AEX92722.1	(85)
GH48	Orpinomyces_8349	Dock	Cellulose	exo-cellobiohydrolase (reducing end)	Streptomyces avermitilis MA-4680	🟦	AEX92722.1	(85)
GH48	Orpinomyces_9842		Cellulose	exo-cellobiohydrolase (reducing end)	Streptomyces griseoaurantiacus M045	🟦	AEX92722.1	(85)
GH53	Orpinomyces_16670	Dock	Pectin	arabinogalactan endo-1,4-β-galactosidase	Ruminococcus albus 8	🟦		
GH53	Orpinomyces_36		Pectin	arabinogalactan endo-1,4-β-galactosidase	Candidatus Solibacter usitatus Ellin6076	🟦		

GH57	Orpinomyces_16437	Polysaccharide deacetylase, Chitin_bind_1 domain	Starch	α-glucosidase	Batrachochytrium dendrobatidis JAM81	🟥		
GH57	Orpinomyces_16484	Polysaccharide deacetylase, DUF963	Starch	α-glucosidase	Rhizopus oryzae RA 99-880	🟥		
GH57	Orpinomyces_3441	Polysaccharide deacetylase	Starch	α-glucosidase	Rhizopus oryzae RA 99-880	🟥		
GH67	Orpinomyces_16601		Hemicellulose	α-glucuronidase	Paludibacter propionicigenes WB4	🟦		
GH67	Orpinomyces_16632		Hemicellulose	α-glucuronidase	Paludibacter propionicigenes WB4	🟦		
GH88	Orpinomyces_16433		Pectin PL produced unsaturated oligosaccharides	unsaturated glucuronyl hydrolase (glycuronidase)	Ruminococcus albus 8	🟦		
Pectate Lyase								
PL1	Orpinomyces_16403	CBM1	Pectin	Pectate lyase	Streptomyces sp. SirexAA-E	🟦		
PL1	Orpinomyces_16404	CBM1	Pectin	Pectate lyase	Streptomyces sp. SirexAA-E	🟦		

PL1	Orpinomyces_16405	CBM1	Pectin	Pectate lyase	Neosartorya fischeri NRRL 181	🟥		
PL1	Orpinomyces_16408	CBM1	Pectin	Pectate lyase	Streptomyces sp. SirexAA-E	🟦		
PL1	Orpinomyces_16410	CBM1	Pectin	Pectate lyase	Streptomyces sp. SirexAA-E	🟦		
PL1	Orpinomyces_16400		Pectin	Pectate lyase	Streptomyces bingchenggensis BCW-1	🟦		
PL1	Orpinomyces_16402		Pectin	Pectate lyase	Streptomyces bingchenggensis BCW-1	🟦		
PL1	Orpinomyces_143		Pectin	Unknown	No hits found			
PL1	Orpinomyces_2377	CBM1	Pectin	Pectate lyase	Streptomyces sp. SirexAA-E	🟦		
PL1	Orpinomyces_10052	CBM1	Pectin	Pectate lyase	Streptomyces sp. SirexAA-E	🟦		
PL1	Orpinomyces_10317		Pectin	Pectate lyase	Streptomyces bingchenggensis BCW-1	🟦		
PL1	Orpinomyces_11041		Pectin	Pectate lyase	Streptomyces bingchenggensis BCW-1	🟦		
PL1	Orpinomyces_11295		Pectin	Pectate lyase	Streptomyces bingchenggensis BCW-1	🟦		
PL1	Orpinomyces_15112	Ricin-type beta-trefoil lectin domain	Pectin	Pectate lyase	Ruminococcus champanellensis 18P13	🟦		

PL1	Orpinomyces_16255	Ankyrin repeats (3 copies) /Ankyrin repeat	Pectin	Pectate lyase	No hits found			
PL1	Orpinomyces_14551	CBM1, LIM-domain binding protein	Pectin	Pectate lyase	Streptomyces sp. SirexAA-E	🟦		
PL1	Orpinomyces_14597	CBM1	Pectin	Pectate lyase	Streptomyces sp. SirexAA-E	🟦		
PL1	Orpinomyces_16407	CBM1	Pectin	Pectate lyase	Neosartorya fischeri NRRL 181	🟥		
PL1	Orpinomyces_16399		Pectin	Pectate lyase	Neosartorya fischeri NRRL 181	🟥		
PL1	Orpinomyces_10086	CBM1	Pectin	Pectate lyase	Aspergillus terreus NIH2624	🟥		
PL1	Orpinomyces_13937	CBM1	Pectin	Pectate lyase	Aspergillus terreus NIH2624	🟥		
PL3	Orpinomyces_1373		Pectin	Pectate lyase	Aspergillus nidulans	🟥		
PL3	Orpinomyces_8883		Pectin	Pectate lyase	Aspergillus nidulans FGSC A4			
PL3	Orpinomyces_12197		Pectin	pectate lyase	Glomerella graminicola M1.001	🟥		
Carboxylesterases								
CE1	Orpinomyces_14286	Abhydrolase_5, CBM_10, Esterase	N/A	feruloyl esterase	Cellulosilyticum ruminicola	🟦	ADZ47894.1	(103)

CE1	Orpinomyces_14389	Abhydrolase_5, CBM_10, Esterase	N/A	feruloyl esterase	Cellulosilyticum ruminicola	🟦	ADZ47894.1	(103)
CE1	Orpinomyces_14566	Abhydrolase_5, CBM_10, Esterase	N/A	feruloyl esterase	Cellulosilyticum ruminicola	🟦	ADZ47894.1	(103)
CE1	Orpinomyces_14680	Abhydrolase_5, CBM_10, Esterase	N/A	feruloyl esterase	Cellulosilyticum ruminicola	🟦	ADZ47894.1	(103)
CE1	Orpinomyces_7232	CBM_10, Esterase	N/A	feruloyl esterase	Cellulosilyticum ruminicola	🟦	ADZ47894.1	(103)
CE1	Orpinomyces_10331	Peptidase_S9	N/A	Lipase	Streptococcus gallolyticus subsp. gallolyticus TX20005	🟦		
CE1	Orpinomyces_11170	Abhydrolase_3	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_11371	Abhydrolase_3, Abhydrolase_5, AXE1, BAAT_C, DLH, DUF2811, Esterase_phd, Peptidase_S9	N/A	Lipase	Clostridium ramosum DSM 1402	🟦		
CE1	Orpinomyces_11412	Abhydrolase_3	N/A	Carboxyl esterase	Clostridium acetobutylicum ATCC 824	🟦		
CE1	Orpinomyces_11974	Peptidase_S9	N/A	Lipase	Streptococcus gallolyticus subsp. gallolyticus TX20005	🟦		

CE1	Orpinomyces_11995	Abhydrolase_3, Abhydrolase_5, DUF2305	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_11996	Abhydrolase_3, Abhydrolase_5, DUF2305	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_12320	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6, Chlorophyllase2, DUF676	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_1299	Abhydrolase_3, Abhydrolase_5, Peptidase_S9	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_13681	Abhydrolase_3, Abhydrolase_5	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_13737	Abhydrolase_3, Abhydrolase_5	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_13813	Abhydrolase_3, Abhydrolase_5	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_13818	Abhydrolase_3, Abhydrolase_5	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_13845	Abhydrolase_3, Abhydrolase_5	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_13848	Abhydrolase_3, Abhydrolase_5, Chlorophyllase	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		

CE1	Orpinomyces_13941	Abhydrolase_3	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_14107	Peptidase_S9	N/A	Lipase	Streptococcus gallolyticus subsp. gallolyticus ATCC BAA-2069	🟦		
CE1	Orpinomyces_15536	Abhydrolase_5, Peptidase_S9	N/A	Lipase	Clostridium ramosum DSM 1402	🟦		
CE1	Orpinomyces_16208	Abhydrolase_3, Abhydrolase_5, BAAT_C, DUF2811, Esterase_phd, Peptidase_S9	N/A	Lipase	Clostridium ramosum DSM 1402	🟦		
CE1	Orpinomyces_16350	Abhydrolase_3, Abhydrolase_5, Peptidase_S9	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16351	Abhydrolase_3, Abhydrolase_5, Peptidase_S9	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16352	Abhydrolase_3, Peptidase_S9	N/A	Carboxyl esterase	Clostridium acetobutylicum ATCC 824	🟦		
CE1	Orpinomyces_16354	Abhydrolase_3, Abhydrolase_5, Chlorophyllase2	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16355	Abhydrolase_3, Peptidase_S9	N/A	Carboxyl esterase	Clostridium acetobutylicum ATCC 824	🟦		

CE1	Orpinomyces_16356	AAA_10, ABC_membrane, ABC_tran, DUF258, FtsK_SpoIIIE, GSPII_E, Macoilin, Miro, SMC_N, TrwB_AAD_bind, Zeta_toxin	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16357	Abhydrolase_3, Abhydrolase_5	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16358	Abhydrolase_3, Peptidase_S9	N/A	Carboxyl esterase	Clostridium acetobutylicum ATCC 824	🟦		
CE1	Orpinomyces_16359	Abhydrolase_5, Esterase_phd, Peptidase_S9	N/A	Lipase	Streptococcus gallolyticus subsp. gallolyticus TX20005	🟦		
CE1	Orpinomyces_16360	Abhydrolase_3, Abhydrolase_5, DUF2305	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16361	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6, Lipase_3	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16363	Abhydrolase_3, Abhydrolase_5, Chlorophyllase, Peptidase_S9	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		

CE1	Orpinomyces_16364	Abhydrolase_3, Abhydrolase_5	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16365	Abhydrolase_3, Abhydrolase_5, Peptidase_S9	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16366	Abhydrolase_3, Abhydrolase_5	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16367	Abhydrolase_3	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16370	Abhydrolase_3, Abhydrolase_5, Peptidase_S9	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16372	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16375		N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16377	Abhydrolase_3, Abhydrolase_5, Chlorophyllase	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16379	Abhydrolase_3, Abhydrolase_5, Chlorophyllase2, DUF2305, Peptidase_S9	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16383	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6, Chlorophyllase2	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		

CE1	Orpinomyces_16384	Abhydrolase_3, Abhydrolase_5, Peptidase_S9, Abhydrolase_3, Abhydrolase_5, Chlorophyllase, Chlorophyllase2, Abhydrolase_3, Abhydrolase_5	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16385	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6, Chlorophyllase2, Lipase_3	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16386	Abhydrolase_3, Peptidase_S9	N/A	Carboxyl esterase	Clostridium acetobutylicum ATCC 824	🟦		
CE1	Orpinomyces_16390	Abhydrolase_3, Abhydrolase_5	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_16392	Abhydrolase_3, Abhydrolase_5	N/A	Conserved hypothetical protein	No hits found			
CE1	Orpinomyces_16395	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6, DLH	N/A	Conserved hypothetical protein	No hits found			
CE1	Orpinomyces_1977	Abhydrolase_3, Abhydrolase_5, Chlorophyllase2, LRR_1, LRR_4,	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		

		Peptidase_S9				🟥		
CE1	Orpinomyces_3901	Abhydrolase_3, Abhydrolase_5, DUF2305, Peptidase_S9	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		
CE1	Orpinomyces_4580	Abhydrolase_3, Abhydrolase_5, Esterase_phd, Peptidase_S9	N/A	Lipase	Mollicutes bacterium D7	🟦		
CE1	Orpinomyces_4687	Abhydrolase_5, Peptidase_S9	N/A	Lipase	Streptococcus gallolyticus subsp. gallolyticus TX20005	🟦		
CE1	Orpinomyces_4704	Abhydrolase_3, Abhydrolase_5, Chitin_bind_1, Esterase_phd, Peptidase_S9	N/A	Lipase	Clostridium ramosum DSM 1402	🟦		
CE1	Orpinomyces_8320	Abhydrolase_5, Esterase_phd, Peptidase_S9	N/A	Lipase	Streptococcus gallolyticus subsp. gallolyticus ATCC BAA-2069	🟦		
CE1	Orpinomyces_8977	Abhydrolase_5, Peptidase_S9	N/A	Lipase	Streptococcus gallolyticus subsp. gallolyticus TX20005	🟦		
CE1	Orpinomyces_9040	Abhydrolase_3, Abhydrolase_5, Peptidase_S9	N/A	Conserved hypothetical protein	Batrachochytrium dendrobatidis JAM81	🟥		

CE1	Orpinomyces_9108	Abhydrolase_3, Abhydrolase_5, Chitin_bind_1, Esterase_phd, Peptidase_S9, rRNA_methylase	N/A	Lipase	Clostridium ramosum DSM 1402	🟦		
CE10	Orpinomyces_13595	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6	N/A	Esterase	No hits found			
CE10	Orpinomyces_13726	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6, Chlorophyllase2, Lipase_3	N/A	Conserved hypothetical protein	Thalassiosira oceanica	🟩		
CE10	Orpinomyces_13752	Abhydrolase_5, Abhydrolase_6, Peptidase_S9	N/A	Conserved hypothetical protein	Physcomitrella patens subsp. Patens	🟨		
CE10	Orpinomyces_13911	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6	N/A	Lipase	No hits found			
CE10	Orpinomyces_14371	Abhydrolase_3, Chlorophyllase2	N/A	Carboxyl esterase	Gordonia terrae NBRC 100016	🟦		
CE10	Orpinomyces_16348	Abhydrolase_3, Abhydrolase_5	N/A	alpha, beta hydrolase domain-containing protein	Citromicrobium bathyomarinum JL354	🟦		
CE10	Orpinomyces_16353	Abhydrolase_3	N/A	ester hydrolase	uncultured prokaryote	🟦		
CE10	Orpinomyces_16373	Abhydrolase_3, Abhydrolase_5	N/A	conserved hypothetical protein	Perkinsus marinus ATCC 50983	🟩		

CE10	Orpinomyces_16374	Abhydrolase_3, Abhydrolase_5, Chlorophyllase2, Peptidase_S9	N/A	Conserved hypothetical protein	No hits found			
CE10	Orpinomyces_16378	Abhydrolase_3, Abhydrolase_5	N/A	alpha, beta hydrolase domain-containing protein	No hits found			
CE10	Orpinomyces_16380	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6, DUF915, Peptidase_S9	N/A	Conserved hypothetical protein	Physcomitrella patens subsp. Patens	🟨		
CE10	Orpinomyces_16388	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6	N/A	Lipase	No hits found			
CE10	Orpinomyces_16393	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6, Lipase_3	N/A	Lipase	No hits found			
CE10	Orpinomyces_16396	Abhydrolase_3, Abhydrolase_5, Chlorophyllase, Chlorophyllase2	N/A	unnamed protein product	Tetraodon nigroviridis	🟥		
CE10	Orpinomyces_2309	Abhydrolase_3, Abhydrolase_5, DUF2305	N/A	Conserved hypothetical protein	Dictyostelium fasciculatum	🟩		
CE10	Orpinomyces_4056	Abhydrolase_5, Abhydrolase_6, Hydrolase_4, Peptidase_S9	N/A	Conserved hypothetical protein	Physcomitrella patens subsp. Patens	🟨		

CE10	Orpinomyces_8981	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6, Chlorophyllase2, Lipase_3	N/A	Conserved hypothetical protein	Thalassiosira oceanica	🟩		
CE_UC	Orpinomyces_4874	Ank, Ank_2	N/A	Conserved hypothetical protein	Salpingoeca sp. ATCC 50818	🟩	ADZ47894.1	(103)
CE_UC	Orpinomyces_102		N/A	Unknown				
CE_UC	Orpinomyces_11068	Abhydrolase_5, Abhydrolase_6, Peptidase_S9	N/A	Conserved hypothetical protein	Rhizopus oryzae RA 99-880	🟥		
CE_UC	Orpinomyces_14170	Abhydrolase_3, Abhydrolase_5	N/A	Conserved hypothetical protein	Puccinia graminis f. sp. tritici CRL 75-36-700-3	🟥		
CE_UC	Orpinomyces_16349		N/A	Unknown				
CE_UC	Orpinomyces_16368		N/A	Conserved hypothetical protein	No hits found			
CE_UC	Orpinomyces_16369	Ank, Ank_2, SH3_1, SH3_2	N/A	Conserved hypothetical protein	Candidatus Amoebophilus asiaticus 5a2	🟦		
CE_UC	Orpinomyces_16381	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6, Chlorophyllase2, DUF2305, Abhydrolase_3, Abhydrolase_5	N/A	Conserved hypothetical protein	Phytophthora sojae	🟩		
CE_UC	Orpinomyces_16389	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6	N/A	Lipase	No hits found			

CE_UC	Orpinomyces_16391	Abhydrolase_3, Fibrinogen_BP, Peptidase_S9	N/A	Rab family GTPase	Entamoeba histolytica HM-1:IMSS	🟩		
CE_UC	Orpinomyces_16394	Abhydrolase_3, Abhydrolase_5, Abhydrolase_6, Chlorophyllase2, Peptidase_S9	N/A	esterase, lipase, thioesterase family protein	No hits found			
CE_UC	Orpinomyces_16397	Abhydrolase_5, DUF249, Peptidase_S9, Abhydrolase_3, Abhydrolase_5, Abhydrolase_6	N/A	Lipase	Streptococcus gallolyticus UCN34	🟦		
CE_UC	Orpinomyces_2306		N/A	mitochondrial endopeptidase	Serpula lacrymans var. lacrymans S7.9	🟥		
CE_UC	Orpinomyces_810	BP28CT, Esterase, HEAT, U3snoRNP10	N/A	Conserved hypothetical protein	Rhizopus oryzae RA 99-880	🟥		

Table S19. Comparison of CAZyme numbers in C1A and various fungal genomes

Phylum	**Species Name**	**GH**	**CE**	**PL**
Neocallimastigomycota	***Orpinomyces* sp. strain C1A**	**358**	**92**	**24**
Ascomycetes	**Eurotiomycetes**			
	Aspergillus nidulans	259	33	22
	Aspergillus niger	254	26	8
	Aspergillus Oryzae	306	30	23
	Penicillium chrysogenum	223	22	9
	Sordariomycetes			
	Magnaporthe grisea	240	47	5
	Myceliophthora thermophile	215	29	8
	Thielavia terrestris	224	27	4
	Podospora anserine	236	41	7
	Neurospora crassa	197	22	4
	Nectria haematococca	329	44	33
	Trichoderma reesei	200	22	5
	Saccharomycetes			
	Saccharomyces cerevisiae	57	3	5
	Schizosaccharomycetes			
	Schizosaccharomyces pombe	54	5	0
	Cryptococcus neoformans	83	7	3
Basidiomycetes	**Agaricomycetes**			

	Laccaria bicolor	163	17	7
	Postia placenta	144	10	6
	Ustilaginomycetes			
	Ustilago maydis	102	1	18
Mucoromycotina	*Mucor circinelloides*	97		
	Rhizopus orizae	116	41	6
Blastocladiomycota	*Allomyces macrogynus*	95	19	4
Chytridiomycota	*Batrachochytrium dendrobatidis* JAM81	36	4	1
	Spizellomyces punctatus	54	14	1
Oomycetes	*Phytophthora infestans*	278	27	47
	Phytophthora ramorum	257	22	49
	Phytophthora sojae	301	37	54

Table S20. Substrate degradtaion capabilities of strain C1A.

Substrate	**Activity**	**CAZY affiliation**	**No. of genes**	**Gene ID**
Cellulose	Endoglucanase	GH45	16	Orpinomyces_7825, Orpinomyces_16773, Orpinomyces_16658, Orpinomyces_524, Orpinomyces_2252, Orpinomyces_16477, Orpinomyces_16718, Orpinomyces_16533, Orpinomyces_6562, Orpinomyces_7870, Orpinomyces_16613, Orpinomyces_16488, Orpinomyces_1749, Orpinomyces_4393, Orpinomyces_16521, Orpinomyces_5778.
		GH5	34	Orpinomyces_8446, Orpinomyces_16691, Orpinomyces_2605, Orpinomyces_16787, Orpinomyces_16700, Orpinomyces_2543, Orpinomyces_16514, Orpinomyces_9720, Orpinomyces_2345, Orpinomyces_7414, Orpinomyces_5811, Orpinomyces_1444, Orpinomyces_1071, Orpinomyces_204, Orpinomyces_909, Orpinomyces_6541, Orpinomyces_11298, Orpinomyces_3867, Orpinomyces_14471, Orpinomyces_16434, Orpinomyces_16438, Orpinomyces_16455, Orpinomyces_16468, Orpinomyces_16486, Orpinomyces_16616, Orpinomyces_2510, Orpinomyces_2784, Orpinomyces_3851, Orpinomyces_12843, Orpinomyces_4162, Orpinomyces_16607, Orpinomyces_44, Orpinomyces_5086, Orpinomyces_5814
		GH8	2	Orpinomyces_16605, Orpinomyces_14432
		GH9	12	Orpinomyces_5169, Orpinomyces_16778, Orpinomyces_16732, Orpinomyces_16648, Orpinomyces_16721, Orpinomyces_3361, Orpinomyces_8987, Orpinomyces_9543, Orpinomyces_3567, Orpinomyces_16430, Orpinomyces_5746, Orpinomyces_6515

	Cellobiohydrolase	GH6	43	Orpinomyces_16626, Orpinomyces_3813, Orpinomyces_16924, Orpinomyces_10295, Orpinomyces_2459, Orpinomyces_6132, Orpinomyces_5544, Orpinomyces_1464, Orpinomyces_7034, Orpinomyces_8266, Orpinomyces_8423, Orpinomyces_16639, Orpinomyces_16653, Orpinomyces_903, Orpinomyces_2901, Orpinomyces_16641, Orpinomyces_16656, Orpinomyces_3583, Orpinomyces_3857, Orpinomyces_4145, Orpinomyces_4642, Orpinomyces_4748, Orpinomyces_5216, Orpinomyces_6944, Orpinomyces_8084, Orpinomyces_11639, Orpinomyces_8802, Orpinomyces_3449, Orpinomyces_3530, Orpinomyces_16776, Orpinomyces_7615, Orpinomyces_16465, Orpinomyces_16543, Orpinomyces_16489, Orpinomyces_16562, Orpinomyces_16629, Orpinomyces_4503, Orpinomyces_3375, Orpinomyces_12853, Orpinomyces_16627, Orpinomyces_16419, Orpinomyces_16505, Orpinomyces_10614
		GH48	21	Orpinomyces_16647, Orpinomyces_16688, Orpinomyces_16756, Orpinomyces_16770, Orpinomyces_16779, Orpinomyces_16646, Orpinomyces_7411, Orpinomyces_8349, Orpinomyces_16682, Orpinomyces_10796, Orpinomyces_16737, Orpinomyces_16738, Orpinomyces_16739, Orpinomyces_16795, Orpinomyces_4269, Orpinomyces_16679, Orpinomyces_7252, Orpinomyces_9842, Orpinomyces_16424, Orpinomyces_1263, Orpinomyces_15099
	Beta glucosidase	GH1	9	Orpinomyces_16576, Orpinomyces_16492, Orpinomyces_3649, Orpinomyces_16550, Orpinomyces_1791, Orpinomyces_8060, Orpinomyces_14808, Orpinomyces_16414, Orpinomyces_1810

		GH3	37	Orpinomyces_2490, Orpinomyces_16608, Orpinomyces_16466, Orpinomyces_16619, Orpinomyces_16542, Orpinomyces_11443, Orpinomyces_14809, Orpinomyces_16462, Orpinomyces_16497, Orpinomyces_16522, Orpinomyces_16534, Orpinomyces_16568, Orpinomyces_1293, Orpinomyces_4249, Orpinomyces_4560, Orpinomyces_6669, Orpinomyces_9603, Orpinomyces_11433, Orpinomyces_15926, Orpinomyces_16582, Orpinomyces_3134, Orpinomyces_15058, Orpinomyces_11276, Orpinomyces_16467, Orpinomyces_16478, Orpinomyces_16587, Orpinomyces_16604, Orpinomyces_1261, Orpinomyces_5676, Orpinomyces_15232, Orpinomyces_16425, Orpinomyces_16504, Orpinomyces_259, Orpinomyces_16439, Orpinomyces_16446, Orpinomyces_16516, Orpinomyces_16631
Hemicellulose				
Xylans (arabinoglucoronoxylan and arabinoxylan)	Endoxylanase	GH10	30	Orpinomyces_938, Orpinomyces_16894, Orpinomyces_16911, Orpinomyces_7686, Orpinomyces_16861, Orpinomyces_5763, Orpinomyces_5899, Orpinomyces_6742, Orpinomyces_5723, Orpinomyces_7494, Orpinomyces_16860, Orpinomyces_16687, Orpinomyces_1160, Orpinomyces_1983, Orpinomyces_6679, Orpinomyces_16690, Orpinomyces_16857, Orpinomyces_11830, Orpinomyces_16698, Orpinomyces_808, Orpinomyces_16432, Orpinomyces_16766, Orpinomyces_16669, Orpinomyces_16413, Orpinomyces_2071, Orpinomyces_13743, Orpinomyces_16527, Orpinomyces_14446, Orpinomyces_16474, Orpinomyces_9123

		GH11	46	Orpinomyces_16909, Orpinomyces_16919, Orpinomyces_16926, Orpinomyces_16927, Orpinomyces_16928, Orpinomyces_16937, Orpinomyces_16938, Orpinomyces_16866, Orpinomyces_16874, Orpinomyces_4868, Orpinomyces_6955, Orpinomyces_16755, Orpinomyces_16793, Orpinomyces_1220, Orpinomyces_1396, Orpinomyces_10724, Orpinomyces_16651, Orpinomyces_16510, Orpinomyces_7134, Orpinomyces_16520, Orpinomyces_16423, Orpinomyces_16454, Orpinomyces_16458, Orpinomyces_16525, Orpinomyces_16526, Orpinomyces_16531, Orpinomyces_16563, Orpinomyces_16564, Orpinomyces_16572, Orpinomyces_16574, Orpinomyces_16625, Orpinomyces_1946, Orpinomyces_2012, Orpinomyces_2909, Orpinomyces_3040, Orpinomyces_3043, Orpinomyces_4849, Orpinomyces_5426, Orpinomyces_5512, Orpinomyces_5933, Orpinomyces_7606, Orpinomyces_11518, Orpinomyces_13274, Orpinomyces_13040, Orpinomyces_15347, Orpinomyces_16518
	beta-xylosidase	GH43[a]	8	Orpinomyces_843, Orpinomyces_16579, Orpinomyces_16597, Orpinomyces_16617, Orpinomyces_16620, Orpinomyces_13906, Orpinomyces_13498, Orpinomyces_15144
		GH39	4	Orpinomyces_16769, Orpinomyces_16452, Orpinomyces_16509, Orpinomyces_2939
	Alpha-glucuronidase	GH67	2	Orpinomyces_16601, Orpinomyces_16632
	Alpha-N-arabinofuranosidase	GH43[a]	16	Orpinomyces_16799, Orpinomyces_16828, Orpinomyces_1088, Orpinomyces_3146, Orpinomyces_8634, Orpinomyces_16512, Orpinomyces_16579, Orpinomyces_16597, Orpinomyces_16617, Orpinomyces_16620, Orpinomyces_16449, Orpinomyces_16570, Orpinomyces_16879, Orpinomyces_2841, Orpinomyces_3639, Orpinomyces_4292

	Acetyl Xylan esterase	NA	26	Orpinomyces_2818, Orpinomyces_16853, Orpinomyces_8585, Orpinomyces_5186, Orpinomyces_973, Orpinomyces_4446, Orpinomyces_8334, Orpinomyces_9840, Orpinomyces_16910, Orpinomyces_2595, Orpinomyces_2, Orpinomyces_4, Orpinomyces_8, Orpinomyces_41, Orpinomyces_55, Orpinomyces_5186, Orpinomyces_8585, Orpinomyces_11850, Orpinomyces_9840, Orpinomyces_58, Orpinomyces_8334, Orpinomyces_14968, Orpinomyces_14969, Orpinomyces_8054, Orpinomyces_8172, Orpinomyces_11975, Orpinomyces_12793
	Polysaccharide deacetylase	GH57[b]	4	Orpinomyces_16501, Orpinomyces_3441, Orpinomyces_16437, Orpinomyces_16484
		GH5[b]	1	Orpinomyces_3867
		GH11[b]	4	Orpinomyces_16651, Orpinomyces_16510, Orpinomyces_7134, Orpinomyces_16520
		Others	12	Orpinomyces_14850, Orpinomyces_14663, Orpinomyces_15246, Orpinomyces_8088, Orpinomyces_10863, Orpinomyces_10749, Orpinomyces_13579, Orpinomyces_9527, Orpinomyces_13889, Orpinomyces_13095, Orpinomyces_13957, Orpinomyces_9796
	Feruloyl esterase/ cinnamoyl esterase	NA	29	Orpinomyces_7232, Orpinomyces_14286, Orpinomyces_14680, Orpinomyces_14566, Orpinomyces_14389, Orpinomyces_107, Orpinomyces_229, Orpinomyces_243, Orpinomyces_259, Orpinomyces_14286, Orpinomyces_14680, Orpinomyces_14566, Orpinomyces_14389, Orpinomyces_7232, Orpinomyces_11664, Orpinomyces_11461, Orpinomyces_476, Orpinomyces_529, Orpinomyces_569, Orpinomyces_14811, Orpinomyces_6926, Orpinomyces_7268, Orpinomyces_10141, Orpinomyces_13911, Orpinomyces_587, Orpinomyces_615, Orpinomyces_657, Orpinomyces_5469, Orpinomyces_6693
Mannans/ glucomannans	Mannanase, Mannosidase	GH5	3	Orpinomyces_11551, Orpinomyces_13832, Orpinomyces_1538

		GH26	5	Orpinomyces_16850, Orpinomyces_16611, Orpinomyces_16411, Orpinomyces_5113, Orpinomyces_16586
		GH2	3	Orpinomyces_4367, Orpinomyces_9907, Orpinomyces_14749
Mixed linkage beta glucan*	Lichenase	GH16	9	Orpinomyces_16490, Orpinomyces_3308, Orpinomyces_14946, Orpinomyces_5528, Orpinomyces_16436, Orpinomyces_16614, Orpinomyces_16615, Orpinomyces_16635, Orpinomyces_16636
Anionic polysaccharides and associated side chain				
Galactan and Arabinogalactan	arabinogalactan endo-1,4-beta-galactosidase	GH53	2	Orpinomyces_16670, Orpinomyces_36
	Beta galactosidase	GH2	1	Orpinomyces_14660
Pectin	unsaturated beta-glucuronyl hydrolase	GH88	1	Orpinomyces_16433
	Pectin/Pectate Lyase	PL1	21	Orpinomyces_16403, Orpinomyces_16404, Orpinomyces_16405, Orpinomyces_16408, Orpinomyces_16410, Orpinomyces_16400, Orpinomyces_16402, Orpinomyces_143, Orpinomyces_2377, Orpinomyces_10052, Orpinomyces_10317, Orpinomyces_11041, Orpinomyces_11295, Orpinomyces_15112, Orpinomyces_16255, Orpinomyces_14551, Orpinomyces_14597, Orpinomyces_16407, Orpinomyces_16399, Orpinomyces_10086, Orpinomyces_13937
		PL3	3	Orpinomyces_1373, Orpinomyces_8883, Orpinomyces_12197

Starch	Alpha amylase	GH13	13	Orpinomyces_675, Orpinomyces_5526, Orpinomyces_5986, Orpinomyces_16506, Orpinomyces_3479, Orpinomyces_3185, Orpinomyces_11877, Orpinomyces_16495, Orpinomyces_16441, Orpinomyces_7289, Orpinomyces_3241, Orpinomyces_16503, Orpinomyces_16426
	alpha-glucosidase	GH31	22	Orpinomyces_16421, Orpinomyces_1225, Orpinomyces_9629, Orpinomyces_16178, Orpinomyces_16227, Orpinomyces_16498, Orpinomyces_15985, Orpinomyces_16431, Orpinomyces_16517, Orpinomyces_8795, Orpinomyces_9302, Orpinomyces_4334, Orpinomyces_16528, Orpinomyces_16442, Orpinomyces_13033, Orpinomyces_16537, Orpinomyces_16541, Orpinomyces_16548, Orpinomyces_3125, Orpinomyces_2368, Orpinomyces_9628, Orpinomyces_10928,
		GH57	3	Orpinomyces_3441, Orpinomyces_16437, Orpinomyces_16484
Chitin	Chitinase	GH18	7	Orpinomyces_8445, Orpinomyces_10945, Orpinomyces_16683, Orpinomyces_13782, Orpinomyces_16609, Orpinomyces_16634, Orpinomyces_16580
Peptidoglycan	Lysozyme	GH25	1	Orpinomyces_6304
Other activities	Beta fructosidase	GH32	2	Orpinomyces_1931, Orpinomyces_16603
	Alpha Mannosidase	GH38	1	Orpinomyces_16429
		GH47	2	Orpinomyces_16599, Orpinomyces_2189
	Expansin module Proteins	NA	8	Orpinomyces_14816, Orpinomyces_7967, Orpinomyces_6036, Orpinomyces_12716, Orpinomyces_12826, Orpinomyces_9216, Orpinomyces_81, Orpinomyces_7911

a: Four annotated as having shared xylosidase, and alpha-N-arabinofuranosidase activities.

b: Genes with dual domains, a GH domain and a Polysaccharide deacetylase domain.

c: Although dedicated xyloglucanases were not identified, beta endogluconases could be utilized for xyloglucan degradation.

d: Although no clear alpha xylosidases were identified, many of the GH31 alpha 1,6 glucosidases could have this activity. Hard to differentiate based on sequence data.

Table S21. Genes with Dockerin domains in the C1A genome.

Model Name	**CAZY Domain**	**Other Non CAZY Domains**	**Predicted Function**	**Top hit phylogeny**
Orpinomyces_16683	GH18		Chitinase	Metarhizium acridum CQMa 102
Orpinomyces_16718	GH45		endo-1,4-β-D-glucanase (Cellulase)	Syncephalastrum racemosum
Orpinomyces_11830	GH10		endo-1,4-β-xylanase	Meiothermus ruber DSM 1279
Orpinomyces_16857	GH10		endo-1,4-β-xylanase	Eubacterium cellulosolvens 6
Orpinomyces_16687	GH10		endo-1,4-β-xylanase	Butyrivibrio proteoclasticus B316
Orpinomyces_16787	GH5		endo-1,4-β-D-glucanase (Cellulase)	Eubacterium siraeum DSM 15702
Orpinomyces_16766	GH10	Ricin-type beta-trefoil lectin domain	endo-1,4-β-xylanase	Eubacterium cellulosolvens 6
Orpinomyces_1160	GH10		endo-1,4-β-xylanase	Butyrivibrio proteoclasticus B316
Orpinomyces_1444	GH5		endo-1,4-β-D-glucanase (Cellulase)	Eubacterium siraeum V10Sc8a
Orpinomyces_6679	GH10		endo-1,4-β-xylanase	Butyrivibrio proteoclasticus B316
Orpinomyces_2543	GH5		endo-1,4-β-D-glucanase (Cellulase)	Eubacterium siraeum DSM 15702
Orpinomyces_16700	GH5		endo-1,4-β-D-glucanase (Cellulase)	Eubacterium siraeum DSM 15702
Orpinomyces_16860	GH10		endo-1,4-β-xylanase	Butyrivibrio proteoclasticus B316
Orpinomyces_6541	GH5	Protein of unknown function (DUF3014)	endo-1,4-β-D-glucanase (Cellulase)	Ralstonia syzygii R24
Orpinomyces_5544	GH6		exo-cellobiohydrolase (non-reducing end)	Aspergillus niger
Orpinomyces_16670	GH53		arabinogalactanendo-1,4-β-galactosidase	Ruminococcus albus 8

Orpinomyces_16669	GH10	Ricin-type beta-trefoil lectin domain	endo-1,4-β-xylanase	Streptomyces sviceus ATCC 29083
Orpinomyces_7414	GH5		endo-1,4-β-D-glucanase (Cellulase)	Paenibacillus curdlanolyticus YK9
Orpinomyces_524	GH45		endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae RA 99-880
Orpinomyces_7411	GH48		exo-cellobiohydrolase (reducing end)	Ktedonobacter racemifer DSM 44963
Orpinomyces_16647	GH48		exo-cellobiohydrolase (reducing end)	Caldicellulosiruptor sp. Tok7B.1
Orpinomyces_2345	GH5		endo-1,4-β-D-glucanase (Cellulase)	Paenibacillus curdlanolyticus YK9
Orpinomyces_8423	GH6		exo-cellobiohydrolase (non-reducing end)	Paenibacillus mucilaginosus K02
Orpinomyces_1071	GH5		endo-1,4-β-D-glucanase (Cellulase)	Ktedonobacter racemifer DSM 44963
Orpinomyces_16755	GH11		endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
Orpinomyces_16773	GH45		endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae RA 99-880
Orpinomyces_16793	GH11		endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
Orpinomyces_16779	GH48		exo-cellobiohydrolase (reducing end)	Catenulispora acidiphila DSM 44928
Orpinomyces_16691	GH5		endo-1,4-β-D-glucanase (Cellulase)	Clostridium sp. BNL1100
Orpinomyces_204	GH5		endo-1,4-β-D-glucanase (Cellulase)	Ruminococcus champanellensis 18P13
Orpinomyces_1464	GH6		exo-cellobiohydrolase (non-reducing end)	Paenibacillus mucilaginosus K02
Orpinomyces_2605	GH5		endo-1,4-β-D-glucanase (Cellulase)	Clostridium sp. BNL1100
Orpinomyces_8266	GH6		exo-cellobiohydrolase (non-reducing end)	Paenibacillus mucilaginosus K02

Orpinomyces_909	GH5		endo-1,4-β-D-glucanase (Cellulase)	Ruminococcus champanellensis 18P13
Orpinomyces_7034	GH6		exo-cellobiohydrolase (non-reducing end)	Paenibacillus mucilaginosus K02
Orpinomyces_3449	GH6		exo-cellobiohydrolase (non-reducing end)	Xanthomonas oryzae pv. oryzae MAFF 311018
Orpinomyces_10724	GH11		endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
Orpinomyces_16651	GH11	Polysaccharide deacetylase	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
Orpinomyces_16690	GH10		endo-1,4-β-xylanase	Clostridium thermocellum DSM 2360
Orpinomyces_16698	GH10		endo-1,4-β-xylanase	Saccharomonospora paurometabolica YIM 90007
Orpinomyces_7615	GH6	Peptidase family M3	exo-cellobiohydrolase (non-reducing end)	Oreochromis niloticus
Orpinomyces_16639	GH6		exo-cellobiohydrolase (non-reducing end)	Paenibacillus mucilaginosus KNP414
Orpinomyces_16756	GH48		exo-cellobiohydrolase (reducing end)	Catenulispora acidiphila DSM 44928
Orpinomyces_1396	GH11		endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
Orpinomyces_16799	GH43	Ricin-type beta-trefoil lectin domain	alpha-N-arabinofuranosidase	Eubacterium cellulosolvens 6
Orpinomyces_16828	GH43	Ricin-type beta-trefoil lectin domain	alpha-N-arabinofuranosidase	Streptomyces sviceus ATCC 29083
Orpinomyces_1088	GH43	Ricin-type beta-trefoil lectin domain	alpha-N-arabinofuranosidase	Clostridium lentocellum DSM 5427

Orpinomyces_3146	GH43	Ricin-type beta-trefoil lectin domain/Procyclic acidic repetitive protein (PARP)	alpha-N-arabinofuranosidase	Clostridium lentocellum DSM 5427
Orpinomyces_16769	GH39	Ricin-type beta-trefoil lectin domain	β-xylosidase	Clostridium sp. DL-VIII
Orpinomyces_16641	GH6		exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'
Orpinomyces_16646	GH48		exo-cellobiohydrolase (reducing end)	Catenulispora acidiphila DSM 44928
Orpinomyces_7252	GH48		exo-cellobiohydrolase (reducing end)	Streptosporangium roseum DSM 43021
Orpinomyces_16737	GH48		exo-cellobiohydrolase (reducing end)	Streptomyces griseoaurantiacus M045
Orpinomyces_16738	GH48		exo-cellobiohydrolase (reducing end)	Streptomyces griseoaurantiacus M045
Orpinomyces_16770	GH48		exo-cellobiohydrolase (reducing end)	Catenulispora acidiphila DSM 44928
Orpinomyces_4269	GH48		exo-cellobiohydrolase (reducing end)	Streptomyces griseoaurantiacus M045
Orpinomyces_16721	GH9		endo-1,4-β-D-glucanase (Cellulase)	Kitasatospora setae KM-6054
Orpinomyces_8446	GH5		endo-1,4-β-D-glucanase (Cellulase)	Clostridiaceae bacterium AN-C16-KBRB
Orpinomyces_11298	GH5	Domain of Unknown Function with PDB structure	endo-1,4-β-D-glucanase (Cellulase)	Clostridiaceae bacterium AN-C16-KBRB
Orpinomyces_16658	GH45		endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae RA 99-880
Orpinomyces_8802	GH6		exo-cellobiohydrolase (non-reducing end)	Verticillium dahliae VdLs.17
Orpinomyces_16732	GH9		endo-1,4-β-D-glucanase (Cellulase)	Kitasatospora setae KM-6054

Orpinomyces_16739	GH48		exo-cellobiohydrolase (reducing end)	Streptomyces griseoaurantiacus M045
Orpinomyces_16776	GH6		exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'
Orpinomyces_16778	GH9		endo-1,4-β-D-glucanase (Cellulase)	Hahella chejuensis KCTC 2396
Orpinomyces_16795	GH48		exo-cellobiohydrolase (reducing end)	Streptomyces griseoaurantiacus M045
Orpinomyces_16648	GH9		endo-1,4-β-D-glucanase (Cellulase)	Kitasatospora setae KM-6054
Orpinomyces_16653	GH6		exo-cellobiohydrolase (non-reducing end)	Paenibacillus mucilaginosus KNP414
Orpinomyces_16656	GH6		exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'
Orpinomyces_16679	GH48		exo-cellobiohydrolase (reducing end)	Streptomyces hygroscopicus subsp. jinggangensis 50
Orpinomyces_16682	GH48		exo-cellobiohydrolase (reducing end)	Streptomyces bingchenggensis BCW-1
Orpinomyces_2490	GH3		β-glucosidase	Phanerochaete chrysosporium
Orpinomyces_2901	GH6		exo-cellobiohydrolase (non-reducing end)	Pyrenophora tritici-repentis Pt-1C-BFP
Orpinomyces_3583	GH6		exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'
Orpinomyces_4145	GH6		exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'
Orpinomyces_4642	GH6		exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'
Orpinomyces_4748	GH6		exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'

Orpinomyces_5216	GH6		exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'
Orpinomyces_10796	GH48		exo-cellobiohydrolase (reducing end)	Streptomyces bingchenggensis BCW-1
Orpinomyces_11639	GH6		exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'
Orpinomyces_1220	GH11	Nuclear RNA-splicing-associated protein	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
Orpinomyces_16850	GH26		endo-1,4-β-mannosidase	Acetivibrio cellulolyticus CD2
Orpinomyces_7870	GH45		endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae RA 99-880
Orpinomyces_8084	GH6		exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'
Orpinomyces_8349	GH48		exo-cellobiohydrolase (reducing end)	Streptomyces avermitilis MA-4680
Orpinomyces_8987	GH9		endo-1,4-β-D-glucanase (Cellulase)	Kitasatospora setae KM-6054
Orpinomyces_9543	GH9		endo-1,4-β-D-glucanase (Cellulase)	Kitasatospora setae KM-6054
Orpinomyces_9720	GH5		endo-1,4-β-D-glucanase (Cellulase)	Myceliophthora thermophila ATCC 42464
Orpinomyces_16432	GH10		endo-1,4-β-xylanase	Streptomyces sviceus ATCC 29083
Orpinomyces_16477	GH45		endo-1,4-β-D-glucanase (Cellulase)	Syncephalastrum racemosum
Orpinomyces_16510	GH11	Polysaccharide deacetylase, DUF915, Esterase_phd, Peptidase_S9	endo-1,4-β-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
Orpinomyces_16514	GH5		endo-1,4-β-D-glucanase (Cellulase)	Ktedonobacter racemifer DSM 44963

Orpinomyces_16533	GH45		endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae RA 99-880
Orpinomyces_16688	GH48		exo-cellobiohydrolase (reducing end)	No hits found
Orpinomyces_16736	GH3		β-glucosidase	Rhizopus oryzae RA 99-880
Orpinomyces_1983	GH10		endo-1,4-β-xylanase	Butyrivibrio proteoclasticus B316
Orpinomyces_2252	GH45		endo-1,4-β-D-glucanase (Cellulase)	No hits found
Orpinomyces_3361	GH9		endo-1,4-β-D-glucanase (Cellulase)	Kitasatospora setae KM-6054
Orpinomyces_3530	GH6		exo-cellobiohydrolase (non-reducing end)	Xylella fastidiosa EB92.1
Orpinomyces_3567	GH9		endo-1,4-β-D-glucanase (Cellulase)	Thermobifida fusca
Orpinomyces_3857	GH6		exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'
Orpinomyces_5169	GH9		endo-1,4-β-D-glucanase (Cellulase)	Cellvibrio gilvu
Orpinomyces_5811	GH5		endo-1,4-β-D-glucanase (Cellulase)	Ruminococcus champanellensis 18P13
Orpinomyces_6562	GH45		endo-1,4-β-D-glucanase (Cellulase)	Rhizopus oryzae RA 99-880
Orpinomyces_6944	GH6		exo-cellobiohydrolase (non-reducing end)	Sorangium cellulosum 'So ce 56'

Orpinomyces_7825	GH45		endo-1,4-β-D-glucanase (Cellulase)	No hits found
Orpinomyces_808	GH10		endo-1,4-β-xylanase	Saccharomonospora paurometabolica YIM 90007
Orpinomyces_903	GH6		exo-cellobiohydrolase (non-reducing end)	Pyrenophora tritici-repentis Pt-1C-BFP
Orpinomyces_16803		Phosphatidylinositol-specific phospholipase C, X domain/Ricin-type beta-trefoil lectin domain	1-phosphatidylinositol phosphodiesterase	Bacillus thuringiensis serovar tochigiensis BGSC 4Y1
Orpinomyces_2818		GDSL-like Lipase/Acylhydrolase	acetylxylan esterase	Trametes versicolor
Orpinomyces_16853		Domain of unknown function (DUF303)	acetylxylan esterase	Cytophaga hutchinsonii ATCC 33406
Orpinomyces_8585		Domain of unknown function (DUF303)/Ricin-type beta-trefoil lectin domain	acetylxylan esterase	Algoriphagus sp. PR1
Orpinomyces_5186		Domain of unknown function (DUF303)	acetylxylan esterase	Algoriphagus sp. PR1
Orpinomyces_973		Domain of unknown function (DUF303)	acetylxylan esterase	Niastella koreensis GR20-10
Orpinomyces_4446		Domain of unknown function (DUF303)	acetylxylan esterase	Niastella koreensis GR20-10
Orpinomyces_8334		GDSL-like Lipase/Acylhydrolase	acetylxylan esterase	Ruminococcus flavefaciens
Orpinomyces_16782		Calcineurin-like phosphoesterase	conserved hypothetical protein with phosphoesterase and metallophosphatase domains	Lachnospiraceae bacterium 5 1 63FAA
Orpinomyces_16836		Calcineurin-like phosphoesterase	conserved hypothetical protein with phosphoesterase and metallophosphatase domains	Lachnospiraceae bacterium 5 1 63FAA
Orpinomyces_3805		Calcineurin-like phosphoesterase	conserved hypothetical protein with phosphoesterase and metallophosphatase domains	Lachnospiraceae bacterium 5 1 63FAA
Orpinomyces_9010			endo-1,4-beta- glucanase/ xyloglucanase	Thermobispora bispora DSM 43833

Orpinomyces_703		BNR/Asp-box repeat	cellulose-binding protein with ricin-type beta trefoil	Micromonospora sp. ATCC 39149
Orpinomyces_14684		Serpin (serine protease inhibitor)	cellulosomal serpin precursor	Ascaris suum
Orpinomyces_16822		Serpin (serine protease inhibitor)	cellulosomal serpin precursor	Ascaris suum
Orpinomyces_2311		Serpin (serine protease inhibitor)	cellulosomal serpin precursor	Branchiostoma lanceolatum
Orpinomyces_15403		Putative esterase/Ricin-type beta-trefoil lectin domain	Ferulic acid esterase	Clostridium thermocellum
Orpinomyces_16816			alpha-L-arabinofuranosidase B	Ruminococcus albus 7
Orpinomyces_13393			alpha-L-arabinofuranosidase B	Ruminococcus albus 7
Orpinomyces_16757			alpha-L-arabinofuranosidase B	Ruminococcus albus 7
Orpinomyces_15655			alpha-L-arabinofuranosidase B	Ruminococcus albus 7
Orpinomyces_16788			hypothetical protein with dockerin type I repeat	Clostridium cellulovorans 743B
Orpinomyces_16712			hypothetical protein with dockerin type I repeat	Clostridium cellulovorans 743B
Orpinomyces_16693		Rare lipoprotein A (RlpA)-like double-psi beta-barrel	expansin module family protein	Trichoderma atroviride IMI 206040
Orpinomyces_9216		Rare lipoprotein A (RlpA)-like double-psi beta-barrel/Pollen allergen	expansin module family protein	Trichoderma atroviride IMI 206040
Orpinomyces_7911		Rare lipoprotein A (RlpA)-like double-psi beta-barrel/Pollen allergen	expansin-like protein	Trichoderma virens Gv29-8
Orpinomyces_4725		Rare lipoprotein A (RlpA)-like double-psi beta-barrel	expansin-like protein	Trichoderma virens Gv29-8
Orpinomyces_14911			Catalytic NodB homology domain of the carbohydrate esterase 4 superfamily	Metarhizium acridum CQMa 102
Orpinomyces_8310			Catalytic NodB homology domain of the carbohydrate esterase 4 superfamily	Metarhizium anisopliae ARSEF 23
Orpinomyces_14389	CEX	Alpha/beta hydrolase family/Putative esterase	Ferulic acid esterase	Cellulosilyticum ruminicola
Orpinomyces_14566	CEX	Alpha/beta hydrolase family/Putative esterase	Ferulic acid esterase	Cellulosilyticum ruminicola

Orpinomyces_11460		Putative esterase	Ferulic acid esterase	Cellulosilyticum ruminicola
Orpinomyces_3084		Putative esterase	Ferulic acid esterase	Cellulosilyticum ruminicola
Orpinomyces_437		Alpha/beta hydrolase family/Putative esterase	Ferulic acid esterase	Cellulosilyticum ruminicola
Orpinomyces_14680		Alpha/beta hydrolase family/Putative esterase	Ferulic acid esterase	Cellulosilyticum ruminicola
Orpinomyces_11664		Alpha/beta hydrolase family/Putative esterase	Ferulic acid esterase	Cellulosilyticum ruminicola
Orpinomyces_14286	CEX	Alpha/beta hydrolase family/Putative esterase	Ferulic acid esterase	Cellulosilyticum ruminicola
Orpinomyces_7232	CEX	Putative esterase	Ferulic acid esterase	Cellulosilyticum ruminicola
Orpinomyces_16812			ABC-type oligopeptide/ dipeptide transporter, substrate-binding component	Desulfovibrio salexigens DSM 2638
ctg7180000056355.1		Putative esterase	feruloyl esterase A, partial	Butyrivibrio fibrisolvens 16/4
Orpinomyces_6926		Putative esterase	feruloyl esterase A, partial	Butyrivibrio fibrisolvens 16/4
Orpinomyces_10141		Putative esterase	feruloyl esterase A, partial	Butyrivibrio fibrisolvens 16/4
Orpinomyces_9349		CBM1, Mucin-like glycoprotein	hypothetical protein with GH61 and fungal CBM domains	Arthrobotrys oligospora ATCC 24927
Orpinomyces_16829		WD domain, G-beta repeat	similar to WD repeat domain 45	Batrachochytrium dendrobatidis JAM81
Orpinomyces_16835		Spc97 / Spc98 family	Spc97 / Spc98 family spindle pole body (SBP) components that form a complex with gamma-tubulin	Batrachochytrium dendrobatidis JAM81
Orpinomyces_16123			endo-1,4-beta-xylanase A	Clostridium leptum DSM 753

Orpinomyces_16704			Peptidase associated domain: C-terminal domain of M14 N/E carboxypeptidase, Por secretion system C-terminal sorting domain, and Poly(3-hydroxybutyrate) depolymerase domain	Fibrobacter succinogenes subsp. succinogenes S85
Orpinomyces_10902			Peptidase associated domain: C-terminal domain of M14 N/E carboxypeptidase, Por secretion system C-terminal sorting domain, and Poly(3-hydroxybutyrate) depolymerase domain	Fibrobacter succinogenes subsp. succinogenes S85
Orpinomyces_3766			Peptidase associated domain: C-terminal domain of M14 N/E carboxypeptidase, Por secretion system C-terminal sorting domain, and Poly(3-hydroxybutyrate) depolymerase domain	Fibrobacter succinogenes subsp. succinogenes S85
Orpinomyces_6346		Ricin-type beta-trefoil lectin domain	Peptidase associated domain: C-terminal domain of M14 N/E carboxypeptidase, Por secretion system C-terminal sorting domain, and Poly(3-hydroxybutyrate) depolymerase domain	Fibrobacter succinogenes subsp. succinogenes S85
Orpinomyces_11941		Ricin-type beta-trefoil lectin domain	Peptidase associated domain: C-terminal domain of M14 N/E carboxypeptidase, Por secretion system C-terminal sorting domain, and Poly(3-hydroxybutyrate) depolymerase domain	Fibrobacter succinogenes subsp. succinogenes S85
Orpinomyces_11115		Ricin-type beta-trefoil lectin domain	Peptidase associated domain: C-terminal domain of M14 N/E carboxypeptidase, Por secretion system C-terminal sorting domain, and Poly(3-hydroxybutyrate) depolymerase domain	Fibrobacter succinogenes subsp. succinogenes S85
Orpinomyces_16859			hypothetical protein	Fibrobacter succinogenes subsp. succinogenes S85

Orpinomyces_1054			hypothetical protein	Bacteroides eggerthii 1 2 48FAA
Orpinomyces_16637		Calcineurin-like phosphoesterase	conserved hypothetical protein with phosphoesterase and metallophosphatase domains	Megamonas funiformis YIT 11815
Orpinomyces_16794		Calcineurin-like phosphoesterase/Calcineurin-like phosphoesterase superfamily domain	conserved hypothetical protein with phosphoesterase and metallophosphatase domains	Megamonas funiformis YIT 11815
Orpinomyces_10090		Calcineurin-like phosphoesterase	conserved hypothetical protein with phosphoesterase and metallophosphatase domains	Megamonas funiformis YIT 11815
Orpinomyces_16837		Calcineurin-like phosphoesterase	conserved hypothetical protein with phosphoesterase and metallophosphatase domains	Megamonas funiformis YIT 11815
Orpinomyces_16800		Calcineurin-like phosphoesterase/Calcineurin-like phosphoesterase superfamily domain	conserved hypothetical protein with phosphoesterase and metallophosphatase domains	Megamonas funiformis YIT 11815
Orpinomyces_16824		Domain of unknown function (DUF4353)	unknown	Clostridium citroniae WAL-17108
Orpinomyces_16845		Domain of unknown function (DUF4353)	unknown	Clostridium citroniae WAL-17108
Orpinomyces_16815		Domain of unknown function (DUF4353)	unknown	Clostridium citroniae WAL-17108
Orpinomyces_16792		Domain of unknown function (DUF4353)	unknown	Clostridium citroniae WAL-17108
Orpinomyces_16811		Protein of unknown function (DUF3184)	unknown	Phytophthora sojae
Orpinomyces_8086			hypothetical protein with ankyrin repeats	Salpingoeca sp. ATCC 50818
Orpinomyces_16655		CotH protein	CotH protein	Ruminococcus flavefaciens FD-1
Orpinomyces_16677		CotH protein	CotH protein	Ruminococcus flavefaciens FD-1

Orpinomyces_8872			lamin-tail domain (LTD), Early set domain associated with the catalytic domain of sugar utilizing enzymes at either the N or C terminus, and cotH domain protein	Ruminococcus flavefaciens FD-1
Orpinomyces_14805			lamin-tail domain (LTD), Early set domain associated with the catalytic domain of sugar utilizing enzymes at either the N or C terminus, and cotH domain protein	Ruminococcus flavefaciens FD-1
Orpinomyces_16830		CotH protein	lamin-tail domain (LTD), Early set domain associated with the catalytic domain of sugar utilizing enzymes at either the N or C terminus, and cotH domain protein	Ruminococcus flavefaciens FD-1
Orpinomyces_16796		Polysaccharide deacetylase	Catalytic NodB homology domain of Colletotrichum lindemuthianum chitin deacetylase and similar proteins	Rhizopus oryzae RA 99-880
Orpinomyces_4953		Predicted AAA-ATPase	AAA-ATPase_like and Protein of unknown function (DUF1703) domain	Roseburia inulinivorans DSM 16841
Orpinomyces_16754			SGNH_hydrolase, or GDSL_hydrolase	Serpula lacrymans var. lacrymans S7.3
Orpinomyces_15761		Rhamnogalacturonate lyase family	Hypothetical protein with rhamnogalacturonan lyase domain	Sorghum bicolor
Orpinomyces_5376		Rare lipoprotein A (RlpA)-like double-psi beta-barrel	COG4305 Endoglucanase C-terminal domain/subunit and related proteins	Myxococcus xanthus DK 1622
Orpinomyces_15124		Leucine Rich Repeat/Leucine Rich repeats (2 copies)	Leucine-rich repeats (LRRs), ribonuclease inhibitor (RI)-like subfamily.	Ectocarpus siliculosus
Orpinomyces_16817			Unknown	No hits found
Orpinomyces_16640			Unknown	No hits found
Orpinomyces_1283		CBM_11	Unknown	No hits found
Orpinomyces_16760			Unknown	No hits found
Orpinomyces_16752			Unknown	No hits found

Orpinomyces_16848		Phosphatidylinositol-specific phospholipase C, X domain/Ricin-type beta-trefoil lectin domain	Phosphatidylinositol diacylglycerol-lyase	Yersinia frederiksenii ATCC 33641
Orpinomyces_12030		Cytolethal distending toxin A/C family/ PtdIns-specific phospholipase C, X domain	phosphatidylinositol-specific phospholipase C	Sphingobacterium spiritivorum ATCC 33861
Orpinomyces_14850		Polysaccharide deacetylase	Catalytic NodB homology domain of Colletotrichum lindemuthianum chitin deacetylase and similar proteins	Fibrobacter succinogenes subsp. succinogenes S85
Orpinomyces_13957		Polysaccharide deacetylase	Catalytic NodB homology domain of Colletotrichum lindemuthianum chitin deacetylase and similar proteins	Aspergillus oryzae RIB40
Orpinomyces_16846		Calcineurin-like phosphoesterase	Predicted phosphohydrolasesof the metallophosphatase superfamily	Megamonas hypermegale ART12/1
Orpinomyces_7271		Calcineurin-like phosphoesterase	Predicted phosphohydrolasesof the metallophosphatase superfamily	Megamonas hypermegale ART12/1
Orpinomyces_3499		Calcineurin-like phosphoesterase	Predicted phosphohydrolasesof the metallophosphatase superfamily	Megamonas hypermegale ART12/1
Orpinomyces_16751		Calcineurin-like phosphoesterase	Predicted phosphohydrolasesof the metallophosphatase superfamily	Megamonas hypermegale ART12/1
Orpinomyces_15540		Calcineurin-like phosphoesterase	Predicted phosphohydrolasesof the metallophosphatase superfamily	Megamonas hypermegale ART12/1
Orpinomyces_16834		Calcineurin-like phosphoesterase/Calcineurin-like phosphoesterase superfamily domain	Predicted phosphohydrolasesof the metallophosphatase superfamily	Megamonas hypermegale ART12/1
Orpinomyces_9959		Rare lipoprotein A (RlpA)-like double-psi beta-barrel	COG4305 Endoglucanase C-terminal domain/subunit and related proteins	Acidovorax radicis N35
Orpinomyces_16802		Ricin-type beta-trefoil lectin domain	Ferulic acid esterase	Schizophyllum commune H4-8

Orpinomyces_6693		Alpha/beta hydrolase family/Esterase PHB depolymerase	Ferulic acid esterase	Sorangium cellulosum 'So ce 56'
Orpinomyces_5469		Alpha/beta hydrolase family/Esterase PHB depolymerase	Poly(3-hydroxybutyrate) depolymerase	Sorangium cellulosum 'So ce 56'
Orpinomyces_16765		Alpha/beta hydrolase family/Esterase PHB depolymerase	Ferulic acid esterase	Sorangium cellulosum 'So ce 56'
Orpinomyces_16819		Calcineurin-like phosphoesterase superfamily domain	Ser/Thr protein phosphatase family protein	Mitsuokella multacida DSM 20544
Orpinomyces_2251		Calcineurin-like phosphoesterase	Ser/Thr protein phosphatase family protein	Mitsuokella multacida DSM 20544
Orpinomyces_4698		CotH protein	spore coat protein CotH	Ruminococcus albus 7
Orpinomyces_5354		CotH protein	spore coat protein CotH	Ruminococcus albus 7
Orpinomyces_15150		CotH protein	spore coat protein CotH	Ruminococcus albus 7
Orpinomyces_1846		CotH protein	spore coat protein CotH	Ruminococcus albus 7
Orpinomyces_16748		CotH protein	spore coat protein CotH	Ruminococcus albus 7
Orpinomyces_16722		CotH protein	spore coat protein CotH	Ruminococcus albus 7
Orpinomyces_7184		CotH protein	spore coat protein CotH	Ruminococcus albus 7
Orpinomyces_14816		Rare lipoprotein A (RlpA)-like double-psi beta-barrel	swollenin	Hypocrea jecorina
Orpinomyces_7967			swollenin	Hypocrea jecorina
Orpinomyces_16844		Rare lipoprotein A (RlpA)-like double-psi beta-barrel/Pollen allergen	swollenin like protein	Aspergillus fumigatus
Orpinomyces_16723			Extracellular tannase	Streptococcus gallolyticus subsp. gallolyticus ATCC 43143
Orpinomyces_7859			Unknown	No hits found
Orpinomyces_16841			ubiquitin carboxyl-terminal hydrolase 5	Coprinopsis cinerea okayama7#130
Orpinomyces_16643		Rare lipoprotein A (RlpA)-like double-psi beta-barrel	similar to extracellular endoglucanase precursor	Corallococcus coralloides DSM 2259

Table S22. Genes with CBM domains in the C1A genome.

Family	**Model name**	**Associated Domains**	**Predicted function**	**Top hit phylogeny**
Carb_bind	Orpinomyces_14946	GH16	β-(1,3-1,4)-glucanase (Lichenase)	Maricaulis maris MCS10
CBM_1	Orpinomyces_7740	Ankyrin repeat/Ankyrin repeats (3 copies)	ankyrin repeat protein	Trichomonas vaginalis G3
CBM_1	Orpinomyces_16903	Ankyrin repeat/Ankyrin repeats (3 copies) /Protein of unknown function (DUF2002)/ Protein of unknown function (DUF3344)/ Uncharacterised protein from bacillus cereus group	Comnserved hypothetical protein	Salpingoeca sp. ATCC 50818
CBM_1	Orpinomyces_2954	Calcineurin-like phosphoesterase superfamily domain	conserved hypothetical protein with phosphoesterase and metallophosphatase domains	Megamonas funiformis YIT 11815
CBM_1	Orpinomyces_6741	Dock, Ankyrin repeat/Ankyrin repeats (3 copies)/BRCA1 C Terminus (BRCT) domain/Domain of unknown function (DUF1771)/Uroporphyrinogen decarboxylase (URO-D)	Comnserved hypothetical protein	Salpingoeca sp. ATCC 50818
CBM_1	Orpinomyces_16893	Domain of unknown function (DUF946)	Unknown	No hits found
CBM_1	Orpinomyces_16091	GDSL-like Lipase/Acylhydrolase/Rhamnogalacturonate lyase family/Rhamnogalacturonase B, N-terminal	Hypothetical protein with rhamnogalacturonan lyase domain	Opitutaceae bacterium TAV5
CBM_1	Orpinomyces_16894	GH10	Spore coat assembly factor SafA	Bacillus thuringiensis serovar israelensis ATCC 35646
CBM_1	Orpinomyces_16911	GH10	endo-1,4-beta-xylanase	Butyrivibrio proteoclasticus B316
CBM_1	Orpinomyces_16861	GH10	endo-1,4-beta-xylanase	Meiothermus ruber DSM 1279
CBM_1	Orpinomyces_938	GH10	endo-1,4-beta-xylanase	Actinomadura sp.
CBM_1	Orpinomyces_5763	GH10	endo-1,4-beta-xylanase	Meiothermus ruber DSM 1279
CBM_1	Orpinomyces_5899	GH10	endo-1,4-beta-xylanase	Sreptomyces Lividans
CBM_1	Orpinomyces_7494	GH10	endo-1,4-beta-xylanase	Verrucosispora maris AB-18-032
CBM_1	Orpinomyces_5723	GH10	endo-1,4-β-xylanase	Meiothermus ruber

CBM_1	Orpinomyces_7686	GH10	endo-1,4-β-xylanase	Butyrivibrio proteoclasticus B316
CBM_1	Orpinomyces_6742	GH10	endo-1,4-β-xylanase	No hits found
CBM_1	Orpinomyces_16938	GH11	endo-1,4-beta-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
CBM_1	Orpinomyces_16909	GH11	endo-1,4-beta-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
CBM_1	Orpinomyces_16919	GH11	endo-1,4-beta-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
CBM_1	Orpinomyces_16926	GH11	endo-1,4-beta-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
CBM_1	Orpinomyces_16927	GH11	endo-1,4-beta-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
CBM_1	Orpinomyces_16928	GH11	endo-1,4-beta-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
CBM_1	Orpinomyces_16937	GH11	endo-1,4-beta-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
CBM_1	Orpinomyces_16866	GH11	endo-1,4-beta-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
CBM_1	Orpinomyces_4868	GH11	endo-1,4-beta-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
CBM_1	Orpinomyces_6955	GH11	endo-1,4-beta-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
CBM_1	Orpinomyces_16874	GH11	endo-1,4-beta-xylanase	Fibrobacter succinogenes subsp. succinogenes S85
CBM_1	Orpinomyces_9349	Mucin-like glycoprotein	hypothetical protein with GH61 and fungal CBM domains	Arthrobotrys oligospora ATCC 24927
CBM_1	Orpinomyces_9630	Pectinesterase	similar to pectinesterase family protein	Arabidopsis lyrata subsp. lyrata
CBM_1	Orpinomyces_7161	Pectinesterase	similar to pectinesterase family protein	Arthrobotrys oligospora ATCC 24927
CBM_1	Orpinomyces_14559	Pectinesterase	similar to pectinesterase family protein	Arthrobotrys oligospora ATCC 24927
CBM_1	Orpinomyces_6177	Pectinesterase	COG4677 Pectin methylesterase	Paenibacillus mucilaginosus K02

CBM_1	Orpinomyces_2350	Pectinesterase	Pectinesterase-2 precursor	Ricinus communis
CBM_1	Orpinomyces_14597	PL1	Pectate lyase	Streptomyces sp. SirexAA-E
CBM_1	Orpinomyces_16404	PL1	Pectate lyase/Amb allergen [Streptomyces sp. SirexAA-E]	Streptomyces sp. SirexAA-E
CBM_1	Orpinomyces_16407	PL1	Pectate lyase	Neosartorya fischeri NRRL 181
CBM_1	Orpinomyces_16410	PL1	Pectate lyase	Streptomyces sp. SirexAA-E
CBM_1	Orpinomyces_16405	PL1	Pectate lyase	Neosartorya fischeri NRRL 181
CBM_1	Orpinomyces_16403	PL1	Pectate lyase	Streptomyces bingchenggensis BCW-1
CBM_1	Orpinomyces_16408	PL1	Pectate lyase	Streptomyces sp SirexAA-E
CBM_1	Orpinomyces_11788	Polysaccharide deacetylase	carbohydrate deacetylase	Coprinopsis cinerea okayama7#130
CBM_1	Orpinomyces_13624	Polysaccharide deacetylase	carbohydrate deacetylase	Coprinopsis cinerea okayama7#130
CBM_1	Orpinomyces_13629	Polysaccharide deacetylase	carbohydrate deacetylase	Coprinopsis cinerea okayama7#130
CBM_1	Orpinomyces_16896	Polysaccharide deacetylase	hypothetical protein	Tuber melanosporum Mel28
CBM_1	Orpinomyces_16899	Rhamnogalacturonate lyase family domain	Unknown	No hits found
CBM_1	Orpinomyces_15651	Rhamnogalacturonate lyase family domain	Hypothetical protein with rhamnogalacturonan lyase domain	Opitutaceae bacterium TAV5
CBM_1	Orpinomyces_16883	Uncharacterised protein family (UPF0203)	Unknown	No hits found
CBM_1	Orpinomyces_16930		similar to pectinesterase family protein	Arthrobotrys oligospora ATCC 24927
CBM_1	Orpinomyces_15421		similar to pectinesterase family protein	Arthrobotrys oligospora ATCC 24927
CBM_1	Orpinomyces_2377		Pectate lyase	Aspergillus nidulans FGSC A4
CBM_1	Orpinomyces_10086		similar to pectate lyase C	Aspergillus terreus NIH2624
CBM_1	Orpinomyces_13692		similar to endoglucanase C	Aspergillus terreus NIH2624
CBM_1	Orpinomyces_13937		similar to pectate lyase C	Aspergillus terreus NIH2624
CBM_1	Orpinomyces_13664		similar to endoglucanase C	Aspergillus terreus NIH2624

CBM_1	Orpinomyces_3267		Hevein or type 1 chitin binding domain (ChtBD1), Catalytic NodB homology domain of Colletotrichum lindemuthianum chitin deacetylase and similar proteins	Batrachochytrium dendrobatidis JAM81
CBM_1	Orpinomyces_16944		hypothetical protein	butyrate-producing bacterium SS3/4
CBM_1	Orpinomyces_9071		Rhamnogalacturan_acetylesterase_like subgroup of SGNH-hydrolases	Clostridium cellulovorans 743B
CBM_1	Orpinomyces_2217		RNA-binding post-transcriptional regulator cip2; Regulates global gene expression after oxidative stress.	Coprinopsis cinerea okayama7#130
CBM_1	Orpinomyces_2595		acetylxylan esterase precursor	Dekkera bruxellensis AWRI1499
CBM_1	Orpinomyces_16933		Peptidase associated domain: C-terminal domain of M14 N/E carboxypeptidase, Por secretion system C-terminal sorting domain, and Poly(3-hydroxybutyrate) depolymerase domain	Fibrobacter succinogenes subsp. succinogenes S85
CBM_1	Orpinomyces_2129		N/A	N/A
CBM_1	Orpinomyces_5927		N/A	N/A
CBM_1	Orpinomyces_12188		N/A	N/A
CBM_1	Orpinomyces_15704		N/A	N/A
CBM_1	Orpinomyces_16910		acetylxylan esterase	Zobellia galactanivorans
CBM_1	Orpinomyces_9840		acetylxylan esterase	Bacteroides plebeius DSM 17135
CBM_1	Orpinomyces_16935		Unknown	No hits found
CBM_1	Orpinomyces_16869		Unknown	No hits found
CBM_1	Orpinomyces_16876		Unknown	No hits found
CBM_1	Orpinomyces_143		Unknown	No hits found
CBM_1	Orpinomyces_295		Unknown	No hits found
CBM_1	Orpinomyces_831		Unknown	No hits found

CBM_1	Orpinomyces_12180		Leucine-rich repeats (LRRs), ribonuclease inhibitor (RI)-like subfamily domain, and Poly(3-hydroxybutyrate) depolymerase domain	Ostreococcus tauri
CBM_1	Orpinomyces_16907		related to esterase D	Piriformospora indica DSM 11827
CBM_1	Orpinomyces_16922		hypothetical protein	Piriformospora indica DSM 11827
CBM_1	Orpinomyces_16934		hypothetical protein	Piriformospora indica DSM 11827
CBM_1	Orpinomyces_16875		hypothetical protein	Podospora anserina S mat+
CBM_1	Orpinomyces_16881		Proteophosphoglycan ppg4 includes the basic region and the leucine zipper region (bZIP) transcription factor domain	Rhodotorula glutinis ATCC 204091
CBM_1	Orpinomyces_16880		Hypothetical protein	Schizosaccharomyces pombe 972h-
CBM_11	Orpinomyces_16917	Minor capsid protein from bacteriophage/Phage T7 tail fibre protein	Unknown	No hits found
CBM_11	Orpinomyces_1283		Unknown	No hits found
CBM_18	Orpinomyces_8445	GH18	Chitinase, Metarhizium acridum CQMa 102	Paenibacillus elgii B69
CBM_18	Orpinomyces_10945	GH18	Chitinase	Grosmannia clavigera kw1407
CBM_21	Orpinomyces_16912		Protein phosphatase 2C	Apis mellifera
CBM_21	Orpinomyces_9610		Putative phosphatase regulatory subunit thought to be involved in the regulation of glycogen metabolism	Aspergillus terreus NIH2624
CBM_21	Orpinomyces_16932		protein phosphatase regulator	Coprinopsis cinerea okayama7#130
CBM_21	Orpinomyces_16885		Unknown	No hits found
CBM_21	Orpinomyces_16901		Unknown	No hits found
CBM_21	Orpinomyces_4347		putative phosphatase regulatory subunit	Paenibacillus curdlanolyticus YK9

CBM_21	Orpinomyces_16914		Hypothetical protein	Plasmodium berghei strain ANKA
CBM_21	Orpinomyces_16920		protein phosphatase regulatory subunit Gac1, Regulatory subunit for Glc7p type-1 protein phosphatase (PP1), tethers Glc7p to Gsy2p glycogen synthase	Talaromyces stipitatus ATCC 10500
CBM_25	Orpinomyces_10670	Ankyrin repeat/Ankyrin repeats (3 copies)	Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit B-like	Amphimedon queenslandica
CBN_25	Orpinomyces_5526		α-Amylase	Streptococcus gallolyticus subsp. gallolyticus ATCC BAA-2069
CBM_25	Orpinomyces_5986	GH13	α-Amylase	Streptococcus gallolyticus subsp. gallolyticus TX20005
CBM_25	Orpinomyces_16890		hypothetical protein with weak homolgy to amyE (alpha-amylase) of Bacillus subtilis 168	Bacillus subtilis subsp. natto BEST195
CBM_25	Orpinomyces_7742		Hypothetical protein with CBM25 domain	Halothermothrix orenii H 168
CBM_25	Orpinomyces_1900		α-Amylase-like protein	Streptosporangium roseum DSM 43021
CBM_48	Orpinomyces_675	GH13	1,4-alpha-glucan-branching enzyme	Neurospora tetrasperma FGSC 2508
CBM_48	Orpinomyces_9551		N/A	N/A
CBM_48	Orpinomyces_16904		pullulanase, type I	Paenibacillus sp. oral taxon 786 str. D14

Table S23. Absolute and relative transcripts per million (TPM) of the top highly transcribed glycoside hydrolase genes in cellobiose- and cellulose-grown cultures.

GH family	**Gene name**	**Substrate**	**Activity**	**Transcripts per million of cellobiose-grown cultures**		**Transcripts per million of cellulose-grown cultures**	
				Absolute	**Relative (%)**	**Absolute**	**Relative (%)**
GH45	Orpinomyces_7825	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	703.6	2.34	155	1.57
GH45	Orpinomyces_6562	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	89.2	0.30	62.9	0.64
GH45	Orpinomyces_524	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	87.7	0.29	58.1	0.59
GH45	Orpinomyces_9330	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	87.7	0.29	58.1	0.59
GH45	Orpinomyces_2252	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	80.2	0.27	26.3	0.27
GH45	Orpinomyces_7870	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	108.9	0.36	19.7	0.20
GH45	Orpinomyces_4559	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	62.3	0.21	0	0
GH5	Orpinomyces_6541	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	12	0.04	498.7	5.05
GH5	Orpinomyces_8446	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	35.9	0.12	434.7	4.40
GH5	Orpinomyces_2345	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	250.1	0.83	418.4	4.23
GH5	Orpinomyces_1071	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	18.3	0.06	260.8	2.64
GH5	Orpinomyces_11298	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	40	0.13	199.3	2.02

GH5	Orpinomyces_909	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	13.8	0.05	168.8	1.71
GH5	Orpinomyces_7414	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	99.6	0.33	124.8	1.26
GH5	Orpinomyces_1538	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	21.2	0.07	91.4	0.92
GH5	Orpinomyces_204	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	73	0.24	77.2	0.78
GH5	Orpinomyces_8093	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	73	0.24	77.2	0.78
GH9	Orpinomyces_8987	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	295.7	0.98	305	3.09
GH9	Orpinomyces_3567	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	212.9	0.71	267.8	2.71
GH9	Orpinomyces_3361	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	103.8	0.34	136.7	1.38
GH9	Orpinomyces_9543	Cellulose	endo-1,4-β-D-glucanase (Cellulase)	442.7	1.47	28.3	0.29
GH6	Orpinomyces_4145	Cellulose	exo-cellobiohydrolase (non-reducing end)	283.3	0.94	688.1	6.96
GH6	Orpinomyces_12853	Cellulose	exo-cellobiohydrolase (non-reducing end)	115.9	0.38	606.9	6.14
GH6	Orpinomyces_2459	Cellulose	exo-cellobiohydrolase (non-reducing end)	16.5	0.05	290.1	2.94
GH6	Orpinomyces_4748	Cellulose	exo-cellobiohydrolase (non-reducing end)	541.9	1.80	116.9	1.18
GH6	Orpinomyces_3449	Cellulose	exo-cellobiohydrolase (non-reducing end)	35	0.12	106.1	1.07
GH6	Orpinomyces_3530	Cellulose	exo-cellobiohydrolase (non-reducing end)	29.0	0.10	56	0.57
GH6	Orpinomyces_11639	Cellulose	exo-cellobiohydrolase (non-reducing end)	78.3	0.26	51.7	0.52
GH6	Orpinomyces_3375	Cellulose	exo-cellobiohydrolase (non-reducing end)	243.5	0.81	44.9	0.45
GH6	Orpinomyces_12896	Cellulose	exo-cellobiohydrolase (non-reducing end)	129.8	0.43	0	0

GH6	Orpinomyces_2592	Cellulose	exo-cellobiohydrolase (non-reducing end)	296.1	0.98	0	0
GH6	Orpinomyces_8802	Cellulose	exo-cellobiohydrolase (non-reducing end)	136.8	0.45	0	0
GH48	Orpinomyces_8349	Cellulose	exo-cellobiohydrolase (reducing end)	445.7	1.48	2426.4	24.56
GH48	Orpinomyces_9842	Cellulose	exo-cellobiohydrolase (reducing end)	504.3	1.67	1561.9	15.81
GH48	Orpinomyces_15099	Cellulose	exo-cellobiohydrolase (reducing end)	54.8	0.18	100.3	1.01
GH48	Orpinomyces_7252	Cellulose	exo-cellobiohydrolase (reducing end)	43.1	0.14	83.2	0.84
GH48	Orpinomyces_7411	Cellulose	exo-cellobiohydrolase (reducing end)	83.4	0.28	15.5	0.16
GH48	Orpinomyces_15550	Cellulose	exo-cellobiohydrolase (reducing end)	110.4	0.37	0	0
GH48	Orpinomyces_8396	Cellulose	exo-cellobiohydrolase (reducing end)	220.2	0.73	0	0
GH1	Orpinomyces_14808	Cellulose	β-glucosidase	1234.5	4.10	783.4	7.93
GH1	Orpinomyces_1810	Cellulose	β-glucosidase	870.2	2.89	402.1	4.07
GH1	Orpinomyces_1791	Cellulose	β-glucosidase	590.4	1.96	167.2	1.69
GH1	Orpinomyces_8060	Cellulose	β-glucosidase	438.9	1.46	120.2	1.22
GH1	Orpinomyces_14840	Cellulose	β-glucosidase	384.1	1.27	0	0
GH3	Orpinomyces_4560	Cellulose	β-glucosidase	329.4	1.09	298.2	3.02
GH3	Orpinomyces_2490	Cellulose	β-glucosidase	252	0.84	250.4	2.53
GH3	Orpinomyces_15232	Cellulose	β-glucosidase	28.9	0.10	147.3	1.49
GH3	Orpinomyces_15926	Cellulose	β-glucosidase	131.8	0.44	118.3	1.20

GH3	Orpinomyces_11433	Cellulose	β-glucosidase	109.5	0.36	8.6	0.09
GH3	Orpinomyces_6669	Cellulose	β-glucosidase	55.8	0.19	5	0.05
GH10	Orpinomyces_11830	Hemicellulose	endo-1,4-β-xylanase	77.5	0.26	268.3	2.72
GH10	Orpinomyces_6742	Hemicellulose	endo-1,4-β-xylanase	19.4	0.06	117.3	1.19
GH10	Orpinomyces_808	Hemicellulose	endo-1,4-β-xylanase	30.1	0.10	57.5	0.58
GH10	Orpinomyces_6679	Hemicellulose	endo-1,4-β-xylanase	6.6	0.02	53.5	0.54
GH13	Orpinomyces_675	Starch	α-Amylase	160.7	0.53	164.9	1.67
GH13	Orpinomyces_5986	Starch	α-Amylase	275.7	0.91	38.1	0.39
GH13	Orpinomyces_5526	Starch	α-Amylase	207.1	0.69	24.9	0.25
GH13	Orpinomyces_7289	Starch	α-Amylase	118.5	0.39	11.5	0.12
GH16	Orpinomyces_6782	Hemicellulose	β-(1,3-1,4)-glucanase (Lichenase)	637.8	2.12	0	0
GH18	Orpinomyces_10945	Chitin	Chitinase	283	0.94	2683.1	27.15
GH18	Orpinomyces_8445	Chitin	Chitinase	319.5	1.06	549.9	5.57
GH18	Orpinomyces_13782	Chitin	Chitinase	9	0.03	104.4	1.06
GH31	Orpinomyces_2368	Starch	α-glucosidase	1134.7	3.77	152.5	1.54
GH31	Orpinomyces_16178	Starch	α-glucosidase	542.7	1.80	71.5	0.72
GH31	Orpinomyces_9302	Starch	α-glucosidase	784.5	2.60	36	0.36
GH31	Orpinomyces_9629	Starch	α-glucosidase	124.5	0.41	30.6	0.31

GH31	Orpinomyces_9628	Starch	α-glucosidase	219.4	0.73	28.9	0.29
GH31	Orpinomyces_3125	Starch	α-glucosidase	232.1	0.77	7.7	0.08
GH31	Orpinomyces_8702	Starch	α-glucosidase	981	3.26	0	0
GH32	Orpinomyces_1931	Other	Unknown	5.7	0.02	54.8	0.55
GH43	Orpinomyces_13906	Hemicellulose	β-xylosidase	12	0.04	1346.3	13.62
GH43	Orpinomyces_2841	Hemicellulose	alpha-N-arabinofuranosidase	3.1	0.01	78.1	0.79
GH43	Orpinomyces_3639	Hemicellulose	alpha-N-arabinofuranosidase	11	0.04	73.1	0.74
GH57	Orpinomyces_3441	Starch	α-glucosidase	11	0.04	55.7	0.56

Table S24. Absolute and relative transcripts per million (TPM) of the top highly transcribed genes in cellobiose- and cellulose-grown cultures.

Gene name	**Function**	**Transcripts per million (TPM)**			
		Cellobiose-grown		**Cellulose-grown**	
		Absolute	**Relative**	**Absolute**	**Relative**
Orpinomyces_10123	Eukaryotic large ribosomal protein P2	3832	0.13	14603	1.48
Orpinomyces_10272	Ribosomal protein L15	15029	0.50	827	0.08
Orpinomyces_10344	Ribosomal protein L23	15088	0.50	36607	3.70
Orpinomyces_10352	Hypothetical protein	537	0.02	4593	0.46
Orpinomyces_10354	Ribosomal protein L31	10647	0.35	4357	0.44
Orpinomyces_10566	Ribosomal protein P2	12010	0.40	102	0.01
Orpinomyces_10952	Ribosomal protein L32	14262	0.47	901	0.09
Orpinomyces_11043	Hypothetical protein	11297	0.37	74	0.01
Orpinomyces_1124	Ribosomal_L16_L10e	12510	0.42	12999	1.32
Orpinomyces_11662	Glyceraldehyde-3-phosphate dehydrogenase	21275	0.71	7348	0.74
Orpinomyces_11762	Ribosomal protein S6e	29100	0.97	7958	0.81
Orpinomyces_1204	Sugar transporter	11759	0.39	786	0.08
Orpinomyces_12434	Ribosomal protein L1	12506	0.42	2255	0.23
Orpinomyces_1249	Nicotinic acetylcholine receptor, alpha2 subunit-like	5016	0.17	14	0.00
Orpinomyces_12508	Ribosomal protein S8	22505	0.75	6659	0.67
Orpinomyces_12818	Ribosomal protein L7	20044	0.67	2740	0.28
Orpinomyces_1327	Ribosomal protein L19e	15876	0.53	25178	2.55
Orpinomyces_13294	Pyruvate formate lyase	4488	0.15	2079	0.21
Orpinomyces_13622	Hypothetical protein	392	0.01	7683	0.78
Orpinomyces_14188	Fructose-bisphosphate aldolase, class II	9856	0.33	4542	0.46
Orpinomyces_14207	Reverse transcriptase	2325	0.08	4282	0.43
Orpinomyces_14466	Ribosomal protein L3	22573	0.75	9338	0.95
Orpinomyces_14786	Enolase	7530	0.25	1938	0.20
Orpinomyces_15581	Malic enzyme, hydrogenosomal	4148	0.14	2270	0.23
Orpinomyces_2376	Ribosomal protein L41	16630	0.55	43658	4.42

Orpinomyces_2836	Hypothetical protein	84	0.00	8730	0.88
Orpinomyces_2942	Hypothetical protein	2	0.00	9621	0.97
Orpinomyces_3183	Ribosomal protein L37ae	4205	0.14	427	0.04
Orpinomyces_3205	RNA-binding protein	1056	0.04	25696	2.60
Orpinomyces_3408	Ribosomal protein L22/L17e	21318	0.71	11599	1.17
Orpinomyces_3421	Actin, gamma 1 propeptide-like	2165	0.07	11524	1.17
Orpinomyces_3446	Ribosomal protein L18e/L15	17944	0.60	6714	0.68
Orpinomyces_3452	Ribosomal protein L11	17040	0.57	4807	0.49
Orpinomyces_3762	Vacuolar ATP synthase 16 kDa proteolipid subunit	4682	0.16	1495	0.15
Orpinomyces_3803	Actin	11265	0.37	10747	1.09
Orpinomyces_3898	G-protein complex beta subunit	15292	0.51	8285	0.84
Orpinomyces_3990	Ribosomal protein S16	5922	0.20	480	0.05
Orpinomyces_4054	Ribosomal protein L24e/L24	18076	0.60	14262	1.44
Orpinomyces_4413	Ribosomal protein rpl30	18248	0.61	4686	0.47
Orpinomyces_4514	Pyruvate formate lyase	829	0.03	12036	1.22
Orpinomyces_4779	Hypothetical protein	71	0.00	5227	0.53
Orpinomyces_4825	Ribosomal protein L18ae	23561	0.78	3728	0.38
Orpinomyces_4841	Telomere and ribosome associated protein Stm1	7759	0.26	14965	1.51
Orpinomyces_4949	Ribosomal protein L7Ae	21609	0.72	2978	0.30
Orpinomyces_5592	Ribosomal protein S24	6413	0.21	9327	0.94
Orpinomyces_5667	Ribosomal protein L2	27725	0.92	10499	1.06
Orpinomyces_6009	Ribosomal protein S2	27266	0.91	13454	1.36
Orpinomyces_6091	Sugar transporter of the major facilitator superfamily	21202	0.70	1782	0.18
Orpinomyces_6251	Ribosomal protein L34	19930	0.66	14998	1.52
Orpinomyces_6538	Ribosomal protein L37	16027	0.53	934	0.09
Orpinomyces_6893	Nascent polypeptide-associated complex alpha subunit-like	4136	0.14	2200	0.22
Orpinomyces_7291	Ribosomal protein L23a-like	19870	0.66	33949	3.44
Orpinomyces_7585	Ribosomal protein L4	13824	0.46	16158	1.64
Orpinomyces_7694	Pyruvate formate lyase	4644	0.15	996	0.10
Orpinomyces_8	Hypothetical protein	120	0.00	6566	0.66

Orpinomyces_8027	Ribosomal protein S27	6468	0.21	185	0.02
Orpinomyces_8358	Ribosomal protein S13	12546	0.42	6205	0.63
Orpinomyces_8829	Elongation Factor 1-alpha	26331	0.87	7368	0.75
Orpinomyces_8876	Succinyl-CoA synthetase alpha subunit	4631	0.15	1751	0.18
Orpinomyces_8897	Ribosomal_L16_L10e	12510	0.42	12999	1.32
Orpinomyces_8990	Similar to MFS sugar transporter	7230	0.24	389	0.04
Orpinomyces_9093	Translationally controlled tumor protein	9590	0.32	2179	0.22
Orpinomyces_919	Hypothetical protein	451	0.01	4350	0.44
Orpinomyces_9804	Ribosomal protein L8	18282	0.61	9680	0.98

Table S25. Summary of dry weight, cellulose, hemicellulose, xylan, lignin, and fermentable sugar losses in microcosms with different types of plant materials.

Substrate	Treat-ment	Dry weight		Cellulose		Hemicellulose		Xylan		Lignin		% Dry weight loss	% FS lost
		T_0	T_f	T_0	T_f	T_0	T_f	T_0	T_f	T_0	T_f		
Switch grass	None	0.40 ± 0.01	0.32 ± 0.07	0.16 ± 0	0.11 ± 0	0.15 ± 0	0.11 ± 0.01	0.12 ± 0	0.09 ± 0.03	0.09 ± 0	0.10 ± 0.01	18.6	28.71
	Acid	0.48 ± 0	0.39 ± 0.01	0.27 ± 0	0.19 ± 0.01	0.06 ± 0	0.05 ± 0	0.06 ± 0	0.04 ± 0	0.15 ± 0	0.16 ± 0	20.2	31.25
	NaOH	0.44 ± 0	0.27 ± 0	0.21 ± 0	0.11 ± 0	0.15 ± 0	0.07 ± 0	0.13 ± 0	0.06 ± 0	0.08 ± 0.01	0.09 ± 0	40.82	53.92
	Steam	0.50 ± 0	0.32 ± 0.01	0.29 ± 0	0.13 ± 0	0.02 ± 0	0.02 ± 0	0.02 ± 0	0.01 ± 0	0.16 ± 0	0.18 ± 0.01	32.19	45.66
Corn stover	None	0.44 ± 0.06	0.26 ± 0	0.20 ± 0	0.10 ± 0	0.16 ± 0	0.10 ± 0	0.13 ± 0.01	0.06 ± 0	0.08 ± 0	0.06 ± 0	39.32	47.45
	NaOH	0.47 ± 0.02	0.18 ± 0.04	0.26 ± 0.01	0.06 ± 0	0.16 ± 0	0.04 ± 0	0.13 ± 0	0.03 ± 0	0.04 ± 0	0.04 ± 0.01	60.94	75.19
Sorghum	None	0.41 ± 0.01	0.28 ± 0.01	0.18 ± 0	0.11 ± 0	0.13 ± 0	0.09 ± 0	0.11 ± 0	0.07 ± 0	0.1 ± 0	0.07 ± 0	32.11	39.26
	NaOH	0.48 ± 0.02	0.25 ± 0.03	0.26 ± 0	0.10 ± 0.01	0.15 ± 0	0.06 ± 0	0.12 ± 0	0.04 ± 0	0.07 ± 0	0.07 ± 0.01	46.85	59.92
Energy cane	None	0.32 ± 0.01	0.24 ± 0.01	0.13 ± 0	0.08 ± 0	0.11 ± 0	0.07 ± 0	0.08 ± 0	0.06 ± 0	0.08 ± 0	0.07 ± 0.01	34.05	28.85
	NaOH	0.49 ± 0.01	0.29 ± 0.01	0.16 ± 0	0.10 ± 0.01	0.16 ± 0	0.07 ± 0	0.12 ± 0	0.05 ± 0	0.10 ± 0	0.09 ± 0.01	41.14	33.35
Alfalfa	None	0.33 ± 0.02	0.24 ± 0.01	0.14 ± 0	0.08 ± 0	0.10 ± 0	0.07 ± 0	0.05 ± 0	0.05 ± 0	0.08 ± 0	0.07 ± 0	27.52	38.85
	NaOH	0.47 ± 0.02	0.29 ± 0.01	0.23 ± 0	0.10 ± 0	0.13 ± 0	0.06 ± 0	0.07 ± 0	0.03 ± 0	0.11 ± 0	0.12 ± 0	37.18	53.39
Bermuda	None	0.49 ± 0.001	0.45 ± 0.01	0.11 ± 0	0.09 ± 0	0.16 ± 0	0.15 ± 0	0.09 ± 0.	0.08 ± 0	0.2 ± 0	0.2 ± 0	7.96	11.1
	NaOH	0.49 ± 0.01	0.37 ± 0	0.15 ± 0	0.06 ± 0	0.12 ± 0.01	0.07 ± 0	0.08 ± 0	0.03 ± 0	0.23 ± 0	0.22 ± 0	25.51	50.33
Willow	None	0.46 ± 0.01	0.45 ± 0	0.21 ± 0	0.20 ± 0	0.11 ± 0.01	0.11 ± 0	0.09 ± 0	0.09 ± 0	0.14 ± 0	0.14 ± 0.01	2.39	5.66
	NaOH	0.50 ± 0.01	0.40 ± 0.01	0.21 ± 0	0.15 ± 0	0.12 ± 0.01	0.08 ± 0	0.09 ± 0	0.05 ± 0	0.18 ± 0	0.15 ± 0	19.56	26.28

T_0: Value at time zero.
T_f: Value at time final.
FS: fermentable sugar

Table S26. Ratios of end products produced (grams) per gram of plant biomass consumed by strain C1A.

Substrate	Lactate (g)/plant biomass consumed (g)	Formate (g) /plant biomass consumed (g)	Acetate (g) /plant biomass consumed (g)	Ethanol (g) /plant biomass consumed (g)
Unt. Switchgrass	0.187	0.491	0.506	0.053
Acid Switchgrass	0.1982	0.403	0.410	0.086
NaOH switchgrass	0.422	0.345	0.310	0.096
Hydrothermolysis switchgrass	0.344	0.349	0.293	0.087
Unt. Alfalfa	0.298	0.376	0.426	0.074
NaOH alfalfa	0.283	0.272	0.300	0.061
Unt. Bermuda	0.022	0.341	0.500	0.0454
NaOH Bermuda	0.185	0.377	0.400	0.015
Unt. Corn Stover	0.477	0.273	0.284	0.063
NaOH Corn Stover	0.384	0.204	0.176	0.062
Unt. Sorghum	0.404	0.316	0.353	0.066
NaOH Sorghum	0.458	0.233	0.185	0.048
Unt. Energy Cane	0.447	0.494	0.576	0.082
NaOH Energy Cane	0.556	0.295	0.290	0.063