Functional Diversity of Carbohydrate-Active Enzymes Enabling a Bacterium to Ferment Plant Biomass
Magali Boutard, Tristan Cerisy, Pierre-Yves Nogue, Adriana Alberti, Jean Weissenbach, Marcel Salanoubat, Andrew C. Tolonen

To cite this version:

HAL Id: cea-01138771
https://hal-cea.archives-ouvertes.fr/cea-01138771
Submitted on 2 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Functional Diversity of Carbohydrate-Active Enzymes Enabling a Bacterium to Ferment Plant Biomass

Magali Boutard1,2,3, Tristan Cerisy1,2,3, Pierre-Yves Nogue1,2,3, Adriana Alberti1, Jean Weissenbach1,2, Marcel Salanoubat1,2,3, Andrew C. Tolonen1,2,3*

1 Genoscope, CEA, DSV, IG, Évry, France, 2 CNRS-UMR8030, Évry, France, 3 Department of Biology, Université d’Évry Val d’Essonne, Évry, France

Abstract

Microbial metabolism of plant polysaccharides is an important part of environmental carbon cycling, human nutrition, and industrial processes based on cellulosic bioconversion. Here we demonstrate a broadly applicable method to analyze how microbes catabolize plant polysaccharides that integrates carbohydrate-active enzyme (CAZyme) assays, RNA sequencing (RNA-seq), and anaerobic growth screening. We apply this method to study how the bacterium Clostridium phytofermentans ferments plant biomass components including glucans, mannans, xylans, galactans, pectins, and arabinoxylans. These polysaccharides are fermented with variable efficiencies, and diauxies prioritize metabolism of preferred substrates. Strand-specific RNA-seq reveals how this bacterium responds to polysaccharides by up-regulating specific groups of CAZymes, transporters, and enzymes to metabolize the constituent sugars. Fifty-six up-regulated CAZymes were purified, and their activities show most polysaccharides are degraded by multiple enzymes, often from the same family, but with divergent rates, specificities, and cellular localizations. CAZymes were then tested in combination to identify synergies between enzymes acting on the same substrate with different catalytic mechanisms. We discuss how these results advance our understanding of how microbes degrade and metabolize plant biomass.

Introduction

Plants annually produce 200 billion tons of lignocellulosic biomass [1], which is metabolized by specialized microbes in diverse environments. For instance, recycling of plant biomass by soil [2] and marine [3] microbes is a key part of the global carbon cycle and intestinal bacteria ferment indigestible plant fiber to short chain fatty acids that constitute 60–85% of calories in ruminants and 5–10% in humans [4]. Further, as only 2% of cellulose is currently used by humans [5], it is a vast potential feedstock that industrial microbes could convert into energy and commodities. Elucidating how microbes depolymerize and metabolize plant biomass is thus important to understand carbon flow in the environment, to promote healthy human nutrition and prevent disease [6], and to develop industrial processes based on cellulosic bioconversion.

Most of plant biomass is in the cell wall, a macromolecular network of phenolic lignin and three types of polysaccharides (cellulose, hemicelluloses, and pectins) whose relative abundances vary widely among species and tissues (Table S1). The load bearing structure of the cell wall consists of cellulose fibrils tethered by various types of hemicellulose. Hemicellulose is enriched in xylan [7] and xyloglucan [8] in dicots, arabinoxylan in monocots [9], and galacto- and glucomannans in gymnosperms [10]. Outside the cell wall, mannans also act as storage polysaccharides in seeds [11], similar to starch. Pectins are cross-linked galacturonic acid-based polysaccharides that act in cellular adhesion and primary wall extension. More than 60% of pectin is often homogalacturan (HG) [12], which is esterified with methanol to various degrees. Rhamnogalacturan I (RGI) [13], the second most abundant pectin, can have galactan and arabinan side chains on the rhamnose residues [14]. Because plant tissues are composed of such heterogeneous polysaccharides, plant-degrading microbes express a myriad of carbohydrate-active enzymes (CAZymes) [15], each of which modifies or cleaves a specific type of sugar linkage.

Here we demonstrate a strategy for systematic analysis of the enzymatic machinery used by microbes to degrade and metabolize plant polysaccharides. Among these microbes, the plant-fermenting clostridia are of particular interest for being a dominant group in the human gut microbiome [6] and top candidates to transform cellulose biomass into fuels and commodities [16], [17]. We studied Clostridium phytofermentans [18], a soil bacterium with 171 CAZyme-encoding genes (Table S2) including 116 glycoside hydrolases in 44 different families. We first quantified growth on comprehensive panel of plant polysaccharides and sugars (Table S3). Strand-specific RNA sequencing revealed all genes whose expression changed on the various substrates. In particular, we focused on up-regulated CAZyme genes and determined how they are organized into regulons that respond to specific
Author Summary

Plant-fermenting bacteria are important for the global carbon cycle, human nutrition, and industrial production of renewable fuels and commodities from cellulosic biomass. Plants are primarily composed of heterogeneous polysaccharides, requiring plant-degrading microbes to encode many carbohydrate-active enzymes (CAZymes) to cleave different sugar linkages. Here we develop a broadly applicable method to study how microbes catalyze plant biomass by determining the combination of CAZymes that depolymerize each polysaccharide into sugars, how the cell alters global mRNA expression, and the efficiency with which each polysaccharide is metabolized. We apply this method to investigate how Clostridium phytofermentans, a bacterium encoding 171 CAZymes, ferments polysaccharides. We assimilate our results into a genetic model of how this bacterium metabolizes plant biomass and discuss how these results further our understanding of microbial plant fermentation.

Results/Discussion

Growth on polysaccharides and sugars

We developed a high resolution, microtiter anaerobic growth assay that shows C. phytofermentans ferments diverse plant polysaccharides (Fig. 1) and their constituent monosaccharides (Fig. S1), but with widely varying cell yields and growth rates (Table S4). It also forms colonies on solid medium containing each polysaccharide except arabinoxylan II (AGII) (Fig. S2). Growth was fastest on HG (Fig. 1A, generation time 0.70h), similar to rumen microbes that digest pectin more rapidly than cellulose and hemicellulose [19]. Although C. phytofermentans ferments both galacturonic acid (Fig. S1F) and rhamnose (Fig. S1H), cell yield was low on RGI (Fig. 1B). C. phytofermentans grows well on galactan (Fig. 1C), xylans (Fig. 1F–G), mannans (Fig. 1H–I), xyloglucan (Fig. 1J), and starch (Fig. 1I). Limited growth on AGII (Fig. 1E) relative to galactan supports that C. phytofermentans cleaves β-1,4 galactan, but not the β-1,3 and β-1,6-galactose bonds in AGII. Poor growth on arabinan (Fig. 1D) is similar to arabinose (Fig. S1G), suggesting this sugar is transported or metabolized inefficiently. C. phytofermentans grows well on cellulose plates (Fig. S2) and solubilizes cellulose substrates such as filter paper and raw corn stover (Fig. S3), but weak growth on carboxymethylcellulose (CMC) might result from either lack of a suitable endoglucanase or carboxymethyl side groups inhibiting its metabolism.

C. phytofermentans shows diauxic growth on the mixed sugar polysaccharides galactomannan (Fig. II) and xyloglucan (Fig. IJ). For each of these substrates, one of the component sugars (galactose or glucose) supports faster growth than the other (mannose or xylose) (Fig. S1, Table S4). Growth on various mixtures of galactose/mannose (Fig. S4) and of glucose/xylose (Fig. S5) shows rapid metabolism of the preferred sugar followed by slower growth on the other one. However, in both cases when the favored sugar reached 75% of the total, the other sugar does not appear to be metabolized.

Similar to some ruminal [19] and human gut microbes [20], C. phytofermentans often grows faster on polysaccharides than the constituent sugars (Table S4). When presented with mixtures of xylan and xylose, this bacterium shows diauxic growth with preferential metabolism of xylan (Fig. S6), which is surprising because xylan must be cleaved to xylose to be metabolized. Growth on polysaccharides could be energetically favorable if significant ATP is saved by simultaneous transport of multiple sugar units in a single oligosaccharide [21] or by intracellular phosphorolysis of oligosaccharides [22]. C. phytofermentans encodes at least a dozen phosphorylases [23] [24], which cleave oligosaccharides without using ATP. Although the mechanisms regulating sugar metabolism in C. phytofermentans are unknown, diauxic growth supports carbon catabolite repression prioritizes growth on preferred sugars and polysaccharides.

Gene expression

We quantified mRNA expression by strand-specific RNA sequencing during log-phase growth on 8 polysaccharides, 3 monosaccharides, and raw corn stover as a complex biomass substrate. An average of 17.3 million mRNA reads were mapped per sample (Table S5), yielding expression (RPKM) values (Table S6) that were highly correlated (r² = 0.96–0.99) between duplicate cultures for all conditions (Fig. S7). The reads were also highly strand-specific (Fig. S8), which will facilitate their future use for de novo transcriptome assembly, gene annotation and detection of antisense transcription. The fraction of reads mapping to CAZymes during growth on glucose was 2.0%, but this increased greatly on polysaccharides, especially cellulose (11.9%) and stover (31.0%). We assessed which genes were significantly differentially expressed on each polysaccharide relative to glucose using DESeq [25] (Table S7). Expression of CAZyme genes on polysaccharides relative to glucose (Fig. 2) shows that between 15 (cellobiose) and 40 (stover) CAZymes were significantly up-regulated per treatment (Table S8) with a total of 92 CAZymes up-regulated on at least one polysaccharide.

The differentially expressed CAZymes are putatively classified by the CAZy database as 67 glycoside hydrolases, 6 carbohydrate esterases, 4 polysaccharide lyases, 14 carbohydrate esterases, and 2 CBM proteins. We analyzed the specificity of the CAZyme transcripational response by K-means clustering the expression profiles of these genes (Fig. 3, Table S9). Cluster A consists six genes that were highly up-regulated on multiple substrates: the GH26 cphy1071, the GH11 cphy2105, two GH18 chitinases cphy1799 and cphy1800 [26], the GH9 cellulase cphy3367 [27] [28] and the GH48 cellulase cphy3368 [29]. Clusters B–F respond to specific polysaccharides such as homogalacturonan (clusters B,C), starch (cluster D), xylan (cluster E), and cellulose/arabinan (cluster F). C. phytofermentans thus perceives signals from individual polysaccharides and responds by up-regulating specific transcriptional regulons that enable it to tailor its complement of CAZymes to the polysaccharide substrate.

CAZyme activities

A set of 56 CAZymes up-regulated on polysaccharides were His-tagged, overexpressed, purified, and their abilities to bind and cleave polysaccharides were quantified. The CAZy database classifies these enzymes putatively as 47 glycoside hydrolases, 4 polysaccharide lyases, and 4 carbohydrate esterases (Table S2); putative glycosyltransferases were not examined as they are not involved in polysaccharide catabolism [6]. Thirty-two enzymes have significant cleavage or binding activities (Fig. 4, Table S1).
Some substrates such as β-1,4-galactan appear to be cut by a single, highly active enzyme, while multiple CAZymes from the same family degrade other substrates such as xylan (GH10), mannan (GH26), starch (GH13), and HG (PL9). CAZymes from multiple families together depolymerize substrates such as xyloglucan (GH2,5,12,31), glucomannan (GH5,GH9,GH26) and galactomannan (GH5,GH26).

We found 15 CAZymes that bind insoluble polysaccharides, most commonly cellulose and mannan (Fig. 4B). Unexpectedly, no CAZyme bound corn stover, suggesting that partial digestion of raw biomass is needed to facilitate enzyme binding. Nine enzymes that bound substrates have carbohydrate binding modules (CBM), but some enzymes such as the cellulase Cphy1163 can bind their substrate without one. While CBM are known to discriminate between polysaccharides such as cellulose and mannan [30], we observed overlap with cellulase CBMs binding mannan and vice versa. Further, CBM from xylanases can bind cellulose and mannan, but with lower affinity, showing that CBM often bind a range of polysaccharides. Consistent with their cleavage activities, GH13 were the only enzymes to bind starch. Enzymes with CBM usually also have catalytic modules, but Cphy1713, a CAZyme with a CBM32 and no catalytic module, binds galactomannan. CBM32 are known to bind galactose and this protein may function similar to one in Yersinia that is proposed to bind oligosaccharides to prevent them from leaking out of the cell [31]. Thirty-two CAZy families have multiple members, which often have divergent cleavage activities and cellular localizations. Cphy1510 has the highest activity among the four GH10 active on xylan (Fig. 5A). Cphy3010, the GH10 with lowest activity, is the only one lacking a secretion signal, supporting it acts intracellularly on xylo-oligosaccharides while the other GH10 are extracellular. Members of the GH5 family act on a wide range of polysaccharides [32]. C. phytofermentans encodes 3 GH5 enzymes, among which one is active on galactomannan and two on xyloglucan (Fig. 5B). The GH5 Cphy1163 has no activity on either of these substrates, but is the most active on cellulose and glucomannan. The 3 GH26 also vary in substrate specificities (Fig. 5C); all the GH26 are similarly active on β-mannan, but only Cphy1071 has cellulase activity and it has lower activity on gluco- and galactomannan. Sequenced-based families are thus useful to make general substrate predictions for CAZymes, but experiments are needed to determine substrate range and catalytic efficiency.

CAZymes mixtures can degrade polysaccharides more efficiently than individual enzymes. We assessed pairwise interactions between each CAZyme and a second enzyme on cellulose (Cphy3367), xylan (Cphy2105), glucomannan (Cphy1071), and

![Figure 1. C. phytofermentans growth on pectic A-E, hemicellulosic F-J, and glucan K-L. Polysaccharides: homogalacturonan A, rhamnogalacturonan I B, galactan C, arabinan D, arabinogalactan II E, xylan F, arabinoxylan G, glucomannan H, galactomannan I, xyloglucan J, carboxymethylcellulose K, starch L. Growth was measured as OD600 every 15 minutes. Each point is the mean of six cultures; red lines show one standard deviation. doi:10.1371/journal.pgen.1004773.g001](image)
homogalacturonan (Cphy1687) (Fig. 5D–G). Similar to results showing synergy between the GH9 Cphy3367 and a B. subtilis GH5 [33], we found that a mix of Cphy3367 and the GH5 Cphy1163 has higher activity on cellulose than either enzyme alone (Fig. 5D), supporting they have complementary roles in cellulolysis. CAZymes can also potentiate other enzymes that have no activity by themselves. For example, the xylanase Cphy2105 activates the putative xylosidases Cphy3009 and Cphy3207 on xylan (Fig. 5E). Similarly, Cphy1071 activates the putative mannosidase Cphy1719 on glucomannan (Fig. 5F). Activities of the GH28 Cphy2567 and Cphy3310 are enhanced by the carbohydrate esterase Cphy1687 (Fig. 5G), supporting this enzyme demethylesterifies homogalacturonan to facilitate its degradation. This carbohydrate esterase did not, however, increase cleavage by the PL9 enzymes that were the most active on homogalacturonan.

Global correlations between CAZyme mRNA expression and cleavage activities were weak for all polysaccharides (Fig. S12), mostly because many CAZyme genes are up-regulated on substrates upon which they have no activity. CAZymes up-regulated on multiple substrates (Fig. 3, cluster A) may act as ‘carbon scouts’ [34] that degrade complex substrates into inducing...
molecules used to fine-tune the expression of hydrolytic enzymes. As described above, some CAZymes such as xylosidases (Fig. 5D) are inactive on intact xylan, but are potentiated by other xylanases. The GH18 Cphy1799 and Cphy1800 are the most highly-upregulated CAZymes on cellulose (Fig. S12), but are chitinases with no activity on cellulose or other plant substrates.
As such, the set of up-regulated CAZymes is useful to identify active enzymes, but strong up-regulation does not necessarily indicate activity on a given substrate.

Conclusions

We assimilated our results into a model of *C. phytofermentans* polysaccharide catabolism that shows degradation by active CAZymes and uses mRNA expression profiles to predict how these substrates are transported and metabolized (Fig. 6). Unlike other clostridia that transport sugars with numerous phospho-transferase systems (PTS) [35] [36], *C. phytofermentans* encodes a single, lowly expressed PTS and also lacks the symporters to transport xylose and arabinose [37]. Instead, *C. phytofermentans* responds to carbon sources by up-regulating between two (galacturonic acid) and twenty-two (arabinan) ABC transporters (Fig. 6). Expression changes support that oligosaccharides and monosaccharides are uptaken by distinct transporters. For example, different ABC transporters are up-regulated on xylose and xylan. Similarly, different transporters respond to glucose, cellobiose, and cellulose. Intracellular cellodextrins are cleaved by at least one cellodextrin phosphorylase (GH94); hexoses are phosphorylated, likely by a ROK hexokinase (Cphy0329) and a putative galactokinase (Cphy2237), and fed into glycolysis. While hexokinases may have wide substrate activity [38], poor growth on...
mannose could be due to inefficient mannose phosphorylation. The pentoses xylose and arabinose are isomerized and metabolized by the pentose phosphate pathway (PPP). Weak growth on arabinose could be due to inefficient transport or the lack of the phosphoketolase in the PPP enabling rapid L-arabinose metabolism by C. acetobutylicum [39].

Plant degrading microbes differ widely in their abilities to depolymerize and metabolize polysaccharides, likely reflecting niche differentiation to alleviate resource competition. Among soil clostridia, C. thermocellum ferments cellulose, but not xylan [40]. C. cellulolyticum grows faster on xylose than xylan and faster on cellobiose than glucose [41], both of which differ from C. acetobutylicum.

Figure 5. Members of the same CAZy family vary in polysaccharide cleavage activities and CAZymes can be potentiated by other enzymes. A Variation in cleavage activities of GH10 enzymes on xylan. B GH5 and C GH26 family members differ in their activities and substrate specificities on amorphous cellulose (red), glucomannan (green), xyloglucan (violet), galactomannan (yellow), mannan (gray). Enzyme activities in A–C are nmol reducing sugar released per milligram enzyme per minute. D–G CAZyme mixtures have higher activities than the individual enzymes. D Cphy1163 and Cphy3367 alone and together on amorphous cellulose. E Cphy2105, Cphy3009, and Cphy3207 alone and the latter two enzymes plus Cphy2105 on xylan. F Cphy1719 and Cphy1071 alone and together on glucomannan. G Cphy1687, Cphy2567, and Cphy3310 alone and the latter two enzymes plus Cphy1687 on homogalacturonan. In D–G, enzyme activities are shown as reducing sugar (nmol) produced by individual and combined enzymes. The fraction of the reducing sugar produced by the mixed enzymes that exceeds the sum of the individual enzymes is shown in green.

doi:10.1371/journal.pgen.1004773.g005
phytofermentans. Similar specialization exists in the human gut microbiome where microbes catabolize different glycans in dietary fiber [20]. The strategy presented here of high-resolution anaerobic growth measurements, RNA sequencing, and CAZyme assays complements other methods such as proteomics [42] and metagenomics. Elucidating how microbes metabolize polysaccharides is key to understanding the function of plant-degrading microbial communities and to develop improved enzyme mixtures and recombinant microbes for industrial processing of plant biomass.

Materials and Methods

Growth measurements

*C. phytofermentans* was cultured anaerobically at 30°C in GS2 medium [43]. Growth was quantified in medium containing 3 g l⁻¹ mono-, di-, or polysaccharides (Table S3 for product numbers) in 100-well microtiter plates (Bioscreen 9502550) that were sealed by press-fitting adhesive sheets (Qiagen 1018104) under the lids. Growth was measured every 15 minutes as OD₆₀₀ at 30°C using a Thermo Scientific Bioscreen C with 30 seconds shaking before each reading. Growth was not measured for β-mannan and amorphous cellulose cultures because these low solubility substrates occluded optical density measurements. Growth on insoluble substrates (15 g l⁻¹ of 0.5 x 5 cm strips of filter paper cellulose or raw corn stover) was measured in 10 ml cultures as substrate consumption by collection on 11 mm filters by vacuum filtration, drying overnight at 65°C, and weighing. Growth on solid GS2 medium with 3 g l⁻¹ polysaccharide and 15 g l⁻¹ agar was tested by incubating plates anaerobically for 10 days at 30°C (Fig. S2). Regenerated amorphous cellulose (RAC) for GS2 agar plates and enzyme assays was prepared from Avicel PH-101 by phosphoric acid treatment [44].

RNA-seq

Cultures containing soluble substrates were sampled for RNA-seq in mid-log phase. Filter paper cellulose cultures were sampled under the same conditions as in [45]; corn stover cultures were sampled after 3 days. Cells were arrested with RNA stabilization buffer, collected by centrifugation (5 kg, 5 min, 4°C), and RNA was isolated with an Ambion Ribopure Bacteria Kit (AM1925). Twenty µg total RNA was treated with Turbo DNase (Invitrogen AM1907), phenol-chloroform extracted, ethanol precipitated, and resuspended in 15 µl 10 mM Tris-Cl, pH 8.5. Ribosomal RNA was depleted from 10 µg RNA using a MICROBExpress kit (Invitrogen AM1905), giving a typical yield of 1–2 µg RNA. 500 ng of rRNA-depleted RNA was fragmented with magnesium buffer (NEB E6101) for 2.5 minutes at 94°C, ethanol precipitated, and resuspended in 10 µl water. RNA was resolved on an Agilent Bioanalyzer 2100 to confirm it was 200–800 bp.

Single stranded cDNA was made from 500 ng fragmented RNA by Superscript III first strand synthesis (Invitrogen 18080-051) including 200 ng/µl actinomycin D (Invitrogen 11805017) and 120 ng/µl random hexamers (Invitrogen 48190-011). Single stranded DNA was twice phenol-chloroform extracted, ethanol precipitated, and resuspended in 52 µl water. The second cDNA strand was synthesized by the Invitrogen double stranded cDNA

Figure 6. Model of polysaccharide degradation and metabolism by *C. phytofermentans*. CAZymes (shown as the number of enzymes in CAZY families) are based on purified activities and are intra- or extracellular based on putative secretion signals. Metabolic enzymes are shown as NCBI numbers and are proposed based on mRNA expression. Rhamnose transport and assimilation is based on pathway from [55]. Abbreviations are D-galacturonic acid (GA), L-rhamnose (R), D-mannose (M), D-glucose (Gc), D-galactose (G), D-xylose (X), L-arabinose (A), fructose (F), phosphate (P), pentose phosphate pathway (PPP), dihydroxyacetone-phosphate (DHAP), glyceraldehyde-3-phosphate (G3P). For each substrate, the number of significantly up-regulated extracellular solute binding proteins (ESB) and ABC transporters (ABC) are shown. Shaded regions show metabolism of glucose (green), mannose (blue), xylose and arabinose (yellow), rhamnose (orange), and galacturonic acid (red). doi:10.1371/journal.pgen.1004773.g006
synthesis kit (Invitrogen 11917-010) using 250 μM dNTP (dATP, dCTP, dGTP, dUTP). Double stranded cDNA was purified using 1.8 volumes of Solid Phase Reversible Immobilisation (SPRI) beads [46] prepared from carboxylate-modified microparticles [47] (ThermoScientific 6513-2103-050250) and resuspended in 50 μl 10 mM Tris-Cl, pH 8.5. DNA was resolved on an Agilent Bioanalyzer to confirm cDNA was 100–300 bp.

Sequencing libraries were prepared from ~250 ng cDNA using the Illumina Truseq DNA Kit (Illumina EC-121-2001). DNA was purified with 1.8 volumes of SPRI beads after end repair. Following adapter ligation, 200–500 bp DNA was selected by incubating with 1 unit USER enzyme (NEB NEB M550S) at 37°C for 15 minutes, followed by 5 minutes at 95°C. Single stranded DNA was PCR amplified and size selected with 1.8 volumes of SPRI beads. Multiplexed libraries were normalized to 2 ng μl⁻¹, pooled, and resolved on an Agilent Bioanalyzer to confirm the peak was at 200–500 bp. Libraries were sequenced on a Illumina HiSeq2000 sequencer. RNA-seq reads were mapped to the C. phytofermentans genome [NCBI NC_010001.1] using Bowtie [48] to report the single, best alignment (see Table S5 for parameters). Gene expression was measured as RPMK (sum of nucleotides in gene per million mapped nucleotides per gene kb) using RSEQtools [49] (Table S6), excluding rRNA reads from RPMK calculations. Differentially expressed genes (Tables S7, S8) were identified by DESeq [25] using the statistical thresholds defined in [50] of >4-fold expression and a p-value <10⁻⁵, which corresponds to a p-value <0.05 adjusted for multiple testing of the 3902 genes in C. phytofermentans genome by Bonferroni correction.

Enzyme purification and activities

A set of 56 CAZymes (Table S2) that were up-regulated on plant polysaccharides were His-tagged, overexpressed, and purified. Primers (Table S10) were designed to clone CAZyme genes by Ligation-Independent Cloning [51] into pET-22B(+). Genes were cloned with C-terminal His-tags except for cphy2105, which were cloned with N-terminal tags to improve yields. The 26 enzymes predicted to be secreted by SignalP [52] were cloned as their mature forms, omitting the N-terminal secretion signals. Forward and reverse sequences of the 56 genes cloned in pET-22B(+) were confirmed by sequencing (Fig. S9). Plasmids were transformed into E. coli BL21(DE3) (Novagen 70235), grown in 50 ml TB medium to OD₆₀₀ = 1.0, and induced with 500 μM IPTG and incubating overnight at 20°C. Cells were pelleted, resuspended in lysis buffer: 50 mM phosphate buffer pH = 8, 0.5M NaCl, 10 mM imidazole, 15% glycerol, 1 mM pefabloc (Sigma 76307). Cells were lysed by sonication (Cole-Parmer Vibracell CV33) in the presence of lysozyme (Novagen 71230). His-tagged proteins were purified from 50 ml culture on Ni-NTA spin columns (Qiagen 31014) and quantified by Bradford assay, giving an average yield of 3 mg protein. Purified proteins were visualized on 12% SDS-PAGE gels (Novex). His-tagged proteins were purified from 50 ml culture on Ni-NTA spin columns (Qiagen 31014) and quantified by Bradford assay, giving an average yield of 3 mg protein.

Supporting Information

Figure S1 C. phytofermentans growth curves on 3 g l⁻¹ di- and monosaccharides: A D-cellobiose, B D-glucose, C D-galactose, D D-xyllose, E D-mannose, F D-galacturonic acid, G D-arabinose, H L-rhamnose. Blue curve is mean density (OD₆₀₀) of 6 cultures; red curves show one standard deviation.

Figure S2 C. phytofermentans growth on solid GS2 medium containing 3 g l⁻¹ polysaccharides. Plates were incubated anaerobically at 30°C for 10 days. Colonies were observed on all substrates except arabinogalactan II.

Figure S3 C. phytofermentans growth on 15 g l⁻¹ A raw corn stover and B filter paper cellulose as a sole carbon source in GS2 medium. Growth was measured as consumption of insoluble substrate. Data are means of triplicate cultures; error bars are one standard deviation. Gray bars show when samples were taken for RNA-seq. The cellulose sample for RNA-seq was taken under the same conditions as those for proteomics in [45].

Figure S4 C. phytofermentans growth on mixtures of galactose and mannose (3 g l⁻¹ total for all treatments): A Galactomannan (80% mannose, 20% galactose), B D-galactose, C D-mannose, D 30% D-mannose and 20% D-galactose, E 50% D-mannose and 50% D-galactose, F 20% D-mannose and 80% D-galactose. Blue line shows mean OD₆₀₀ of 6 cultures, red lines show range of one standard deviation. Growth on individual sugars shows that C. phytofermentans grows faster and to higher density on D-galactose than D-mannose. Growth is diauxic on galactomannan and sugar mixtures A, D, E, and F supporting that galactose is metabolized preferentially before mannose.

Figure S5 C. phytofermentans growth on mixtures of glucose and xylose (3 g l⁻¹ total for all treatments): A xyloglucan (45% glucose, 35% xylose), B D-glucose, C D-xylose, D 75% D-glucose and 25% D-xylose, E 50% D-glucose and 50% D-xylose, F 25% D-glucose and 75% D-xylose. Blue line shows mean OD₆₀₀ of 6 cultures, red lines show range of one standard deviation. Growth on individual sugars shows that C. phytofermentans grows faster and to higher density on D-glucose than on D-xylose. Growth is diauxic on xyloglucan and mixtures of glucose and xylose A, D, E, F supporting that D-glucose is metabolized preferentially before D-xylose.

Figure S6 C. phytofermentans growth on mixtures of xylan and xylose (3 g l⁻¹ total): A xylan, B D-xylose, C 25% D-xylose and...
75% xylan, D 50% D-xylose and 50% xylan, E 75% D-xylose and 25% xylan. Black curve is the mean OD_{600} of 6 cultures; the red curves show the range of 1 standard deviation. Growth curves show that xylan is metabolized before its constituent monomer, D-xylose.

(PNG)

Figure S7  Expression values (log_{2}(RPKM)) for RNA sequencing of replicate cultures of all carbon source treatments are highly correlated. A–H are single-end reads and I–M are paired-end reads. Spearman correlation coefficients are shown on each panel. (PNG)

Figure S8  Reads from dUTP sequencing are highly strand-specific with an average of 33,715 times more reads mapping to the expected strand for each of the 8 rRNA operons (16S-5S-23S) in the glucose samples. Operons 1–3 are transcribed in the positive direction; Operons 4–8 are in the reverse direction. Reads mapping to the expected strand are in magenta and the opposing strand are green. Note, dUTP sequencing reads map to the opposite strand. Thus, genes transcribed in the positive orientation are sequenced with reads on the reverse strand. (PNG)

Figure S9  Forward and reverse sequence alignments of 56 CAZyme genes cloned into pET-22B(+) confirm that genes had the correct sequences. (PDF)

Figure S10  Purified CAZyme visualized on 12% SDS-PAGE gels (Nupage bis-Tris novex gel IM-8042). Masses (kDa) of full proteins are shown for each CAZyme; masses with asterisks are secreted proteins for which the N-terminal secretion signal was not cloned, resulting in an expected size slightly smaller than the full protein mass shown. Mass ladders with blue borders are from images that were cropped by omitting intermediate lanes. (PNG)

Figure S11  Sugar standard curves from DNS assays used to convert polysaccharide DNS readings to sugar concentrations. (PNG)

Figure S12  Comparison of mRNA expression versus enzyme activity for purified CAZymes on A homogalacturonan, B galactan, C arabian, D xylan, E galactomannan, F cellulose, G starch. mRNA expression is expressed as log_{2}(RPKM) on the polysaccharide relative to glucose. Enzyme activity is log_{2}(umol sugar per mg enzyme per minute). (PNG)

Table S1  Plant polysaccharides used in this study: natural abundances, plant source of tested compound, chemical structure, and sugar composition. References: [36] [57] [58] [59] [60] [61] [62] [63]. (XLS)

Table S2  *C. phytofermentans* CAZymes in the CAZy database [15] including gene name, NCBI accession, gene (bp) and protein (amino acid) length, predicted protein molecular weight (kDa), N-terminal secretion signal cleavage position (SignalP [52] Y-max position), CAZy classification, and annotation. The 56 CAZymes purified in this study are shown in blue. (XLS)

Table S3  Vendor product information for mono-, di-, and polysaccharides used in this study. (XLS)

Table S4  *C. phytofermentans* growth rates (generation time in hours) and cell yields (maximum OD_{600}) growing on polysaccharides and di-/monosaccharides. Growth rate on each substrate was calculated by log-transforming the mean growth curve (Fig. 1S1) and calculating the slope of a linear regression fit to the portion of the curve describing exponential growth. Regressions were fit to each of the two growth phases on galactomannan and xyloglucan. (XLS)

Table S5  Mapping of RNA-seq reads to the *C. phytofermentans* genome, excluding reads mapping to rRNA operons. Reads were mapped using Bowtie [48] to report the single, best alignment between RNA-seq reads in.fastq files and the *C. phytofermentans* NCBI genome file (NC_010001.gbk). Reads were mapped using the following command: bowtie –best -k 1 [bowtie database of NC_010001.fna] –un unmappedReads.txt -q [fastq file] mappedReads.txt. (XLS)

Table S6  mRNA expression (RPKM) of all *C. phytofermentans* genes (page 1) and of only CAZymes (page 2) during steady state growth on sugars and polysaccharides. RPKM is defined as the sum of nucleotides from the reads that overlap with a given annotation entry normalized per million mapped nucleotides and the length of the annotation item. RPKM were calculated using RSEQtools [49] from mapped reads using MRFrameant files. rRNA reads were removed from mrq files before running mrQuant so that expression levels are normalized to mRNA reads. Samples are labeled whether they were sequenced using single-end (S) or paired-end (P) reads. The 56 CAZymes purified in this study are shown in blue on page 2. (XLS)

Table S7  Differential expression of all *C. phytofermentans* genes on nine polysaccharides relative to glucose. Differentially expressed genes (DEseq p-value<10^{-5} and a >4-fold differential expression) are shown in blue. Sheets show differential expression relative to glucose of the following carbon sources: homogalacturonan, galactan, arabian, xylan, galactomannan, cellulose, cellobiose, starch, stover. Paired-end sequenced polysaccharide samples are compared to the paired-end glucose sample; Single-end sequenced polysaccharide samples are compared to the single-end glucose sample (see Table S6). (XLS)

Table S8  Differentially expressed CAZyme genes on nine polysaccharides relative to glucose. Page 1 shows all 92 differentially expressed CAZymes and the polysaccharides upon which they were up-regulated. Subsequent pages show the expression values of CAZymes up-regulated on specific polysaccharides. Purified CAZymes are shown in blue. Proportion of CAZyme genes with significant mRNA expression changes that were purified: homogalacturonan (25/27), galactan (8/13), arabian (37/54), xylan (19/19), galactomannan (28/45), cellobiose (15/15), cellulose (26/27), starch (14/32), stover (33/40). (XLS)

Table S9  Membership of CAZyme genes in K-means clusters. The 92 differentially-expressed CAZymes were clustered based on their expression (log_{2} expression ratios relative to glucose) as shown in Fig. 3. (XLS)

Table S10  Primers for LIC cloning of 56 CAZymes. LIC sequences that overlap with the pET-22B(+) cloning plasmid are shown in red. His-tag sequences are in green. Genes of putatively secreted proteins were cloned without their predicted N-terminal secretion signals (Table S1). Genes were cloned with C-terminal...
His-tags, except for Cphy1687 and Cphy2105, were cloned with N-terminal His-tags to improve yield.

**Table S11** Polysaccharide cleavage activities (nmol reducing sugar per mg enzyme per minute) of 56 purified CAZymes on 14 polysaccharides. Enzyme (50 μg) was incubated with 0.25% polysaccharide in 25 mM potassium phosphate buffer pH = 7 for 30 minutes at 37°C. Activities: >160 red, >80 orange, >40 yellow, >20 green.

**Acknowledgments**

We are grateful to the following people for their technical assistance: S Leschine, J Blanchard and T Warnick for microbiology, J Levin and K Labadie for RNA sequencing, and S Gas and CEH Tolonen for informatics. We thank the Biopolymers Facility at Harvard Medical School and the Genoscope-CEA for RNA sequencing.

**Author Contributions**

Conceived and designed the experiments: ACT. JW MS. Performed the experiments: MB TC PYN AA. Analyzed the data: MB TC PYN AA. Wrote the paper: ACT.

**References**