19 Apr 19 sbs

Caldicellulosiruptor bescii regulates pilus expression in response to the polysaccharide, xylan

Asma M.A.M. Khan, Valerie J. Hauk, Thomas Raffel, and Sara E. Blumer-Schuette*
Department of Biological Sciences
Oakland University, Rochester, MI 48309

Submitted to: *Appl Environ Microbiol (Apr 2019)*

Running Title: *Caldicellulosiruptor bescii pilin expression*

Keywords: *Caldicellulosiruptor*, extreme thermophile, type IV pilus, xylan

* Address Correspondence to: Sara E. Blumer-Schuette
Dept. of Biological Sciences
Oakland University
118 Library Drive
374 Dodge Hall
Phone: (248) 370-3168
Fax: (248) 370-4225
Email: blumerschuette@oakland.edu

ABSTRACT

Biological hydrolysis of cellulose at high temperatures relies on microorganisms that either secrete free enzymes or assemble cellulosomes. While these enzymatic systems appear to be opposites of one another, they may share an underlying mechanism of attachment. Extreme thermophile *Caldicellulosiruptor bescii* is highly cellulolytic, due in part to its freely secreted modular, multi-functional carbohydrate acting enzymes. Additionally, *C. bescii* also employs non-catalytic carbohydrate binding proteins, which likely evolved as a mechanism to compete against other heterotrophs in the carbon limited biotopes that these bacteria inhabit. Prior analysis of the *Caldicellulosiruptor* pangenome identified that a type IV pilus (T4P) locus is conserved among all *Caldicellulosiruptor* species. Interestingly, T4P loci are evolutionarily divergent between the highly and weakly cellulolytic members of the genus *Caldicellulosiruptor*. In this study, we sought to determine if *C. bescii* T4P plays a role in attachment to plant polysaccharides. Based on pilin-like protein domains, transcriptomics and protein expression data, we identified the major pilin (*pilA*) encoded for by the *C. bescii* genome. Using immunodot blots, we identified xylan as the main inducer of PilA production, in comparison to other representative plant polysaccharides. The extracellular location of PilA was further confirmed by immunofluorescence microscopy. Furthermore, recombinant PilA specifically disrupted *C. bescii* cell adhesion to xylan and crystalline cellulose in competitive cell binding assays. Based on these observations, we propose that PilA, the major *C. bescii* pilin, and by extension the T4P, plays a role in *Caldicellulosiruptor* cell attachment to plant biomass.

IMPORTANCE

Most microorganisms are capable of attaching to surfaces either to persist or take advantage of an energy source. Here, we describe that the thermophilic, plant degrading bacterium, *Caldicellulosiruptor bescii*, uses type IV pili to attach to polysaccharides found in plant biomass. This ability is likely key to survival in environments where carbon sources are limiting, allowing *C. bescii* to compete against other plant degrading microorganisms. Interestingly, the polysaccharide that induced the highest expression of pilin protein was xylan, a hemicellulose that is not the majority polysaccharide in plant biomass. We also demonstrate that attachment to polysaccharides can be disrupted by the addition of recombinant pilin. This mechanism mirrors those recently described in pathogenic Gram-positive bacteria, and may indicate the ancient origins of type IV pilus systems.

INTRODUCTION

35 Thermophilic microorganisms capable of hydrolyzing all, or part of lignocellulosic plant biomass have been under considerable interest for biotechnological applications of their enzymes. Of note are cellulolytic microorganisms which produce the enzymes necessary to hydrolyze recalcitrant plant biomass. The Gram-positive, anaerobic, extremely thermophilic genus *Caldicellulosiruptor* employ an array of mechanisms for deconstruction of plant biomass 40 (reviewed in 1). One hallmark of this genus is modular, multifunctional enzymes comprised of both catalytic and binding domains. *Caldicellulosiruptor bescii* is a highly cellulolytic member of the genus (2, 3) capable of attaching to and degrading plant biomass at temperatures as high as 90°C (4). Notably, *C. bescii* is able to degrade insoluble cellulose along with various other plant polysaccharides like xylan (5) and pectin (6), and can grow efficiently on untreated plant biomass 45 with high lignin content (7-9). Modular enzymes with multiple catalytic domains, found primarily in the glucan degradation locus (GDL) (10) diversifies the substrates that these enzymes can hydrolyze (11-15).

In multiple studies, *Caldicellulosiruptor* cells have been observed adhering to plant biomass during growth (4, 9, 16-19), presumably as an adaptation to efficiently degrade 50 lignocellulosic biomass. Given that the genus *Caldicellulosiruptor* does not produce a cellulosome, other proteins have been implicated in mediating this attachment. All members of genus *Caldicellulosiruptor* produce one or more S-layer bound proteins and enzymes (9, 10). Two S-layer located proteins from *Caldicellulosiruptor saccharolyticus* were demonstrated to adhere to cellulose (17). Additionally, S-layer associated enzymes from *Caldicellulosiruptor* 55 *kronotskyensis* facilitated attachment to xylan when heterologously expressed in *C. bescii* (9). Aside from S-layer located proteins, other potential plant polysaccharide-interacting proteins are also produced by the genus *Caldicellulosiruptor*, including substrate-binding proteins, flagella, a type IV pilus (T4P) and uncharacterized hypothetical proteins which were enriched in a cellulose-bound fraction through a proteomics screen (10). Additional *C. bescii* substrate binding proteins 60 implicated in attachment to plant biomass have also been identified through extracellular protein identification (20), including an expanded proteomics dataset comparing extracellular proteins produced during growth on plant biomass versus crystalline cellulose (21). Both these proteomics studies corroborate the significance of non-catalytic proteins in the process of lignocellulosic plant biomass deconstruction by the genus *Caldicellulosiruptor*. Structurally unique proteins called 65 tāpirins are another mechanism by which strongly to weakly cellulolytic *Caldicellulosiruptor* species attach to cellulose from plant biomass (19, 22). Genes encoding for tāpirins are located

directly downstream of the T4P locus in all cellulolytic *Caldicellulosiruptor* (10, 19, 22), however it remains to be determined if they interact with the T4P.

Protein expression studies using ruminal cellulolytic bacteria identified pilins as cellulose-binding proteins through comparison of binding-deficient mutants versus wild type for *Fibrobacter succinogenes* (23) and *Ruminococcus flavefaciens* (24). Celluose affinity pull-down assays using extracellular proteins from *R. albus* 8 (25) and extracellular proteome analysis of non-cellulosomal binding deficient *R. albus* 20 mutant also identified pilin-like proteins, further implicating pili in cellular attachment to cellulose (26). Transcriptomic analysis, however, determined that pilin genes encoded by *R. albus* 7 were not differentially expressed on cellulose in comparison to cellobiose, and may indicate that other polysaccharides act as the inducer for pilin expression (27). Taken together, these studies indicate that T4 pilins from other Gram-positive, cellulolytic bacteria, like the genus *Caldicellulosiruptor*, may facilitate the attachment of cells to cellulose.

Genes required for assembly of a Gram-negative like T4P are fairly widespread throughout the genus *Clostridium* (28, 29), including the pathogens *Clostridium perfringens* (28), and *C. difficile* (30). Major pilins from both *C. perfringens* (31) and *C. difficile* (32) have been demonstrated to play a role in adhesion, and heterologous expression of the *C. perfringens* major pilin gene, (*pilA2*), in T4P-deficient *Neisseria gonorrhoea* mutants resulted in attachment to muscle cells (31). By gaining a novel cell-adherence specificity in the complemented *N. gonorrhea* strain, but not restoration of motility, illustrates a limitation to the conserved structure and function of pilins. Recently, the ATPases required for twitching motility in *C. perfringens* were found to be upregulated in response to colonization and necrosis of murine muscle tissue (33) supporting the role of *C. perfringens* T4P adherence to muscle cells. Furthermore, *C. difficile* Δ*pilA1* mutants, lacking T4P, were significantly reduced in their ability to attach to epithelial cells (32).

Gram-negative like T4P genetic loci are also present in the genomes of plant biomass degrading genus, *Caldicellulosiruptor* (10). Among the strongly cellulolytic *Caldicellulosiruptor*, this locus is located upstream of the tāpirins and modular, multi-functional cellulases (10). Available transcriptomics (4, 8, 34) and proteomics data (4, 10, 21, 35) indicate that this T4P locus is strongly upregulated, and that pilin peptides are also present during growth on plant biomass or plant-derived polysaccharides. Considering this data, along with compelling evidence from ruminal and pathogenic Firmicutes that indicate the involvement of T4P in adherence (23-25, 31, 32) we propose that the *C. bescii* T4P plays a role in attachment to plant polysaccharides during plant biomass deconstruction. Through analysis of publicly available transcriptomics and proteomics data, we have identified the major pilin in *C. bescii.* Our findings also reveal that xylan is the key inducer of the *C. bescii* major pilin, and the pilin protein is located to the extracellular

surface of *C. bescii* during growth on complex polysaccharides. Recombinant *C. bescii* pilin also specifically inhibited adherence of *C. bescii* cells to plant polysaccharides, supporting the role of *C. bescii* T4P in attachment to plant biomass.

RESULTS

105 ***A type IV pilus is encoded for by the genome from Caldicellulosiruptor bescii***

Based on genome data available for *C. bescii* (4) we confirmed that all essential genes required for assembly of a Gram negative-like type IV pilus (T4P) are present (**Fig. 1a**). Essential genes for T4P assembly in a Gram-positive bacterium include a pilin subunit, a pre-pilin peptidase, an assembly ATPase, and a membrane protein that recruits the ATPase (29).
110 Moreover, like other Gram-positive bacteria (36) these genes appear to be arranged in an operon, as there are no gaps larger than 71bp between the T4P genes, and only a single hairpin sequence is predicted within the pre-pilin peptidase coding sequence in the T4P locus (**Table S1**).

When comparing the organization of the T4P locus of *C. bescii* to 12 other sequenced *Caldicellulosiruptor* species, we observed that their T4P locus organization was highly conserved.
115 Despite this, orthologous pilins shared between 41.8 to 100% percent amino acid sequence identity (**Table S2**), indicating that there may be some evolutionary adaptations among pilins between the strongly cellulolytic versus weakly cellulolytic species. A phylogenetic tree built from alignment of concatenated T4P loci from 12 sequenced *Caldicellulosiruptor* species was constructed to assess if there were correlations between cellulolytic ability and the T4P (**Fig. 1b)**.
120 As expected, genes from the T4P operons clustered by cellulolytic ability of the corresponding *Caldicellulosiruptor* species. This lends credence to our primary hypothesis that the T4P, on account of its proximity to modular, multi-domain glycoside hydrolases, plays a role in plant biomass attachment preceding enzymatic deconstruction.

125 ***Five putative pilins are encoded in the C. bescii genome***

In order to identify putative pilins in the T4P locus, we screened genes for an *N*-terminal pre-pilin cleavage site that is post translationally modified by the prepilin peptidase (PilD) prior to pilus assembly (37). We also screened for typical Gram-positive sortase dependent amino acid motifs (LPxTG, 38) however none were identified in the T4P locus. Based on the presence of a
130 pre-pilin cleavage site (*N*-terminal amino acid motif: GFxxxE), we identified five genes (Athe_1872, Athe_1876, Athe_1877, Athe_1880 and Athe_1881) as encoding for putative pilins (**Table 1**). Predicted protein lengths for these range in size from 130aa (Athe_1880) to 277aa (Athe_1872), furthermore, when we analyzed the leader peptide for all five predicted pilins, they all were of variable length, ranging from 5 up to 21 amino acids long (**Table 1, Fig. S1**). Based on
135 the total predicted amino acid length of Athe_1880 and Athe_1881, these proteins are typical of T4a pilins (39), however, Athe_1880 has a leader peptide 15 amino acids in length which is not typical for T4a pilins.

Predicted secondary structures of these putative pilins indicate that these pilins share some common regions with the Gram negative T4 pilins as shown in **(Fig. 2)** (39, 40). These common regions include an N-terminal helix: α1 and a globular C-terminal domain. The N-terminal half of α1 is referred to as α1-N (peach) and the C terminal half as α1-C (cyan) regions, the αβ loop (red) separating the α helix from the antiparallel β sheets (yellow) of the C terminus (**Fig. 2**). Gram negative T4 pilins typically have cysteine residues that define and stabilize the D region. Such a D region cannot be defined for *C. bescii* putative pilins as they lack the cysteine residues. A predicted transmembrane domain (TMD) is located within the α1-N terminal region of each putative pilin (**Fig. 2**).

Predicted secondary structure and the amino acid length of mature Athe_1880 suggest that it resembles a T4a pilin, however the major pilin from *C. difficile* was originally predicted to be a T4a pilin on the basis of structural prediction (29), but its solved structure (PDB accessions, 4TSM, 4OGM, and 4PE2) is more closely aligned with T4b pilins (41). Given the prior inaccuracy of predicted protein models for Gram positive pilins, we first used *C. bescii* putative pilins as queries to search for amino acid similarity from proteins with crystal structures available on the Protein Data Bank website. Athe_1880 (37%) and Athe_1881 (59%) shared limited homology with poor E-values in the α1-N region with representative T4a pilin, *Dichelobacter nodosus* FimA (PDB accession 3SOK), so we did not attempt to use these structures to predict the tertiary structure of any *C. bescii* pilins. Surprisingly, none of the putative pilins shared significant homology with *Clostridioides difficile* (formerly '*Clostridium difficile'*) pilin PilA1 (PDB accessions, 4TSM, 4OGM, and 4PE2) (41) or *Streptococcus sanguinis* PilE1 (PDB accession 6I20) (42). Firmicutes that encode for Gram negative-like T4P may use different pilins as their major pilin, therefore, we widened our BLASTp analysis to determine if any *C. bescii* putative pilins matched other predicted pilin amino acid sequences from *C. difficile*, *C. perfringens* or *S. sanguinis*. Of the five putative pilins, only Athe_1880 and Athe_1881 shared low homology (28–29% identity over >70% query coverage) with *C. difficile* pilins, none of which were homologs of the major *C. difficile* pilin, PilA1 (30). While we could not find acceptable templates to build a predicted structure of *C. bescii* pilins with, these findings do suggest that in cases were a genome encodes for multiple pilins, orthologous pilins may not necessarily serve as the major pilin subunit across functionally diverse species.

Athe_1880 is the major pilin (PilA) based on transcriptomics and proteomics evidence

We expect that the major *C. bescii* pilin would be highly expressed and would constitute the majority of the T4P structure, therefore we examined publicly available transcriptomics and

proteomics data available for *C. bescii*. The major *C. bescii* pilin should either be upregulated as determined by transcriptomics data, or enriched as peptide fragments in proteomics data. Datasets from three independent comparative transcriptomic studies of *C. bescii* grown on
175 switchgrass versus glucose (8), filter paper versus glucose (4) and microcrystalline cellulose versus switchgrass (34) respectively indicated that Athe_1880 and Athe_1881 are the most highly upregulated genes among the candidate pilins (**Table 2**). Proteomics data for protein abundance on microcrystalline cellulose confirmed that Athe_1880 is the most abundant of all of the candidate pilins across three independent sets of proteomics data (4, 10, 35). These data were
180 further corroborated by a recent study on the extracellular proteome of *C. bescii* where Athe_1880 was found to have a fold change greater than 2x on complex substrates like xylan, switchgrass and Avicel compared to simple substrates like xylose, glucose and cellobiose (21). Given both the transcriptomic and proteomic evidence, we confirm that Athe_1880 is the major pilin represented in the T4P of *C. bescii*, that we will refer to as PilA. We then sought to produce
185 soluble, recombinant PilA using rational design, informed by predicted secondary protein structures (see **Fig. 2**). Soluble, recombinant PilA protein was produced by truncating Athe_1880 to remove the α-1N region of the alpha-helix (**Fig. 3**).

Xylan is the key inducer of PilA in C. bescii

190 Previously published transcriptomics and proteomics data confirmed that *pilA* gene expression and protein production were upregulated when *C. bescii* was grown on polysaccharides or plant biomass versus mono- or disaccharides (4, 8, 10, 21, 34, 35). Based on this evidence it is natural to assume that cellulose would be the main polysaccharide regulating the T4P operon. Since the other representative plant polysaccharides had yet to be tested, we
195 sought to examine if hemicellulose polysaccharides played a role in the regulation of the T4P by monitoring T4P pilus biogenesis through immuno-dot blots. *C. bescii* cultures were cultured on 5 different sugars: cellulose, pectin, glucomannan, xylan, and xylose and sampled during early, mid-, and late exponential phase. After normalization to cell density, xylan induced a 10-fold higher amount of pilin production (7.7 pg ml^{-1} $cell^{-1}$) compared to glucomannan (0.75 pg ml^{-1} $cell^{-1}$) or xylose (0.51 pg ml^{-1} $cell^{-1}$) at late exponential growth (**Fig. 4**). Furthermore, the amount of pilin protein quantified also increased in a growth phase dependent manner for all sugars tested, with no detectable PilA present during early exponential phase for growth on glucomannan and xylose. The most dramatic increase was noted during growth on xylan, with over a 3.5-fold increase in PilA from early to late exponential growth (**Fig. 4**). Interestingly, while cellulose is the
205 major component of plant biomass, it was the least effective polysaccharide for inducing T4P

production, and extracellular PilA below measurable limits, as was also the case for pectin (data not shown). Given that xylan induced the highest levels of PilA protein, we concluded that it is in fact, xylan, rather than cellulose that is the main inducer of PilA. This is not completely unsurprising, considering that after cellulose, xylan is the second most common polysaccharide present in secondary plant cell walls (43).

PilA is present on the extracellular surface of C. bescii cells

Although immunoblots confirmed the presence of PilA protein, we wanted to confirm that this protein is located to the external surface of the cell, and not secreted into the media. To visualize PilA, we labelled *C. bescii* cells grown on xylan with anti-PilA, using a fluorescently-tagged secondary antibody for visualization. Immunofluorescence microscopy images confirm the extracellular localization of PilA, as a green signal is present around DAPI-stained cells (**Fig. 5B**). Furthermore, the same samples processed with only the secondary antibody lack any fluorescent signal, establishing that non-specific binding was not the cause (**Fig. 5A, C**). Furthermore, we also attempted to detect PilA on *C. bescii* cells grown on xylose which is a monomer of xylan (**Fig. 5D**) to determine if PilA production would be induced by the monosaccharide component of xylan. Based on our immunofluorescence micrographs, PilA was not observed when *C. bescii* was grown on xylose. As a control, abiotically treated xylan was processed similar to biotic cultures, and the lack of fluorescence confirms that xylan was not interacting with either of the antibodies in a non-specific manner (**Fig. 5E**). However, when abiotically treated xylan was incubated with recombinant PilA, we observed a weak signal on xylan (**Fig. 5F**). This was surprising, as we could not detect any interaction of recombinant PilA with xylan or crystalline cellulose using pull down assays. Incubating rPilA with xylan, microcrystalline cellulose and filter paper (cellulose fiber) did not detect interactions as observed as eluted protein in an SDS-PAGE gel (**Fig. 6**). The overall amount of rPilA bound to xylan as observed in **Fig. 5F** is likely to be very low, as it is below the level of detection for stained SDS-PAGE gels (**Fig. 6**).

PilA inhibits binding of C. bescii cells to xylan and cellulose

Given the proximity of the T4P locus to major cellulases used by *C. bescii* (10), we originally hypothesized that the major pilin was functioning as an adhesin, by binding to plant polysaccharides. Based on the apparent weak affinity of PilA for xylan (**Fig. 5F**), we examined its putative role in adherence by assessing if it would interfere with *C. bescii* attachment to insoluble polysaccharides. We measured the ability of *C. bescii* planktonic cells to attach to insoluble polysaccharides using 2x2 factorial design (**Fig. S2**). Factorial design allowed us to test whether

240 the presence of PilA influenced the attachment of planktonic cells to insoluble substrates, by testing for a statistical interaction between the "Protein" (PilA presence) and "Substrate" (xylan or cellulose presence) treatment variables. Planktonic cell densities (PCDs) were measured after incubation with an insoluble substrate and/ or recombinant PilA, with a reduction in PCD after treatment is indicative of cell attachment. Based on the increased production of PilA protein in
245 response to xylan, we tested *C. bescii* cells grown on xylan for their ability to attach to insoluble xylan or cellulose (**Fig. 7 A, C**). As a comparison, we also tested if cells grown on cellulose behaved similarly (**Fig. 7B**).

In all cases, the PCD after exposure to insoluble substrate (**Fig. 7**, dotted lines) were lower than the PCD of cells in buffer alone (**Fig. 7**, solid line), indicating that *C. bescii* cells were
250 attaching to xylan and cellulose, as expected. Interestingly, while the attachment of *C. bescii* cells to xylan after growth on xylan (26% attachment, **Fig. 7A**) is expected, a higher proportion of cells were attaching to cellulose after growth on xylan (69% attachment, **Fig. 7C**). Moreover, after addition of rPilA (**Fig. 7 A-C**, PilA), PCDs increase indicating that the presence of rPilA during the attachment process is inhibiting the ability of cells to interact with the substrate. We observed
255 complete inhibition of cell attachment to xylan (**Fig. 7A**) compared to only partial inhibition of cell attachment to cellulose (**Fig. 7 B, C**) with rPilA. Two-way ANOVA was then used to test for interaction effects between our independent variables, substrate and protein (**Table S3**). The interaction between treatments (substrate and PilA) were statistically significant with p-values below 0.05 (**Table S3**). This interaction is also specific, as a control protein, BSA, did not interfere
260 with attachment based on a lack of statistical interaction in this experiment (**Fig. 7D, Table S3**). By extension, our cell binding data support that *C. bescii* T4P play an integral role in cellular attachment to xylan, but additional mechanisms are used for cellular attachment to crystalline cellulose.

265 DISCUSSION

In this study, we present an initial functional characterization of the major pilin from a Gram-negative T4P locus encoded for by *C. bescii*. Similar to other members of the class Clostridia (24, 29), the *C. bescii* genome encodes for the genes required to produce a Gram negative-like T4P in a single locus (**Fig. 1A**), that is likely arranged as an operon (**Table S1**).
270 Additionally, based on nucleotide sequence homology, the concatenated *C. bescii* T4P locus clusters with other T4P loci from strongly cellulolytic *Caldicellulosiruptor* species (**Fig. 1B**). Initial bioinformatics analyses identified five putative pilin genes located in the *C. bescii* T4P locus (**Table 1**). Pathogenic Firmicutes, such as *Streptococcus sanguinis* (44, 45) and *C. perfringens*

(28), also encode for multiple pilins in their respective T4P loci, and a hypervirulent *C. difficile* strain encodes for as many as nine pilins distributed across five genomic loci (30). Unexpectedly, none of the *C. bescii* putative pilins share similarity with the known crystal structures of major pilins from T4P producing Gram-positive bacteria, *C. difficile* and *S. sanguinis.* However, Athe_1880 and Athe_1881 align with other annotated pilin proteins from *C. difficile.* These analyses indicate that regardless of amino acid sequence homology between Firmicute pilins, the major pilin appears to be unique to each species.

Based on our phylogenetic analysis, combined with past proteomics data (10), we hypothesized that the *C. bescii* T4P is likely used to facilitate cell attachment to polysaccharides found in plant biomass. To test this hypothesis, we first used transcriptomics and proteomics data (**Table 2**) to identify the *C. bescii* major pilin, (PilA) for molecular cloning and analysis. Direct pilin-mediated attachment to cellulose was previously observed in cellulosomal and non-cellulosomal members of the genus *Ruminococcus* (26, 46), and we sought to determine if members of the noncellulosomal genus *Caldicellulosiruptor* used T4P in a similar manner to facilitate attachment to plant polysaccharides. Pilins from *C. bescii* are unlikely to mediate direct interaction with plant biomass, we did not detect any adsorption of rPilA to crystalline cellulose or xylan, in qualitative binding assays (**Fig. 6**) in contrast to the tāpirins, which bind with moderate affinity to crystalline cellulose (19).

Regardless, previous proteomics studies have identified PilA peptides produced by *C. bescii* when grown on cellulose (4, 10, 21, 35) implying that the T4P overall play a potential role in attachment to plant polysaccharides. We used rPilA to generate polyclonal antibodies against PilA to monitor the presence of *C. bescii* pilins during growth on various plant polysaccharides. In this study, we cultured *C. bescii* on a chemically defined medium ensuring that the carbohydrate provided is the only available carbon source. Overall, the highest PilA production during growth was observed on xylan in comparison to glucomannan, xylose (**Fig. 4**), cellulose, or pectin (data not shown), implying that xylan is the main inducer of T4P locus expression in *C. bescii*. Immunofluorescence microscopy further confirmed that PilA is extracellularly localized during growth on xylan (**Fig. 5**), similar to observations on the subcellular localization of *C. perfringens* T4P (28). It was surprising that we could not detect any extracellular PilA from *C. bescii* cells grown on cellulose using immunoblots, and this discrepancy with prior data may be explained by their use of complex media in previous proteomics studies (4, 10, 35).

Response of a T4P locus to xylan has yet to be described, as *R. flavefaciens* and *F. succinogenes* pilins were identified in cultures grown on cellulose (23, 24) and/or glucose (23). However, response to soluble xylooligosaccharides in carbon-limited biotopes would be a likely

adaptation to sense plant biomass in terrestrial hot springs. For example, *Ruminiclostridium thermocellum* (formerly '*Clostridium thermocellum'*) uses extracytoplasmic factor anti-sigma factors to upregulate genes encoding for cellulosomes-associated enzymes in response to soluble xylooligosaccharides in its environment (47, 48).

While the major *C. bescii* pilin does not appreciably adhere to xylan, planktonic *C. bescii* cells cultured on xylan attached to a greater extent to cellulose than to xylan. This adherence was, in part, mediated by T4P in a specific manner (**Fig. 7**), as supplemented rPilA interfered with the attachment of planktonic *C. bescii* cells to crystalline cellulose. Since rPilA did not adhere to xylan or cellulose in polysaccharide pulldown assays (**Fig. 6**), it is not likely that this observation is result of rPilA blocking binding sites for *C. bescii* cells. One explanation is that PilA may undergo a conformational change after pilus assembly, exposing a binding domain, similar to epitopes exposed by *Neisseria* T4P subjected to force (49, 50). However, this explanation does not account for the role of rPilA, as we do not expect N-terminus truncations of PilA to self assemble in a manner similar to the native T4P. Another possible explanation is that the *C. bescii* T4 pili are not directly interacting with insoluble substrates, but are attaching to cellulose through association with other proteins, such as the tāpirins, or to xylan using an as-of-yet unidentified xylan adhesin. This explanation is most plausible as it would explain the apparent lack of PilA affinity for polysaccharides, but the significant role of rPilA in interfering with *C. bescii* cell attachment. A similar mechanism has been recently described for a T4b pilus from enterotoxicogenic *Escherichia coli* (ETEC*)*, where a secreted adhesin (CofJ) protein functions as a type of "scout" that would first associate with receptors on the host cell, and then interact with minor pilins present at the tip of the T4b pilus (51). Based on previously published data (19) and our findings, we propose a model (**Fig. 8**) for the putative role of the T4P in attachment to cellulose or xylan. Overall, our data supports the notion that non-cellulosomal species, like those from the genus *Caldicellulosiruptor*, have evolved alternate mechanisms to adhere to insoluble substrates.

MATERIALS AND METHODS

Identification of Pilin Genes from *C. bescii* and bioinformatics analysis

Putative pilin genes in *C. bescii* were identified using the Joint Genome Institute Integrated Microbial Genomes (IMG) database (52). Functional annotation of the predicted amino acid sequences for each pilin used the Pfam database (53) available within IMG. Signal peptides were predicted by the SignalP 4.1 Server (54) and transmembrane domains were predicted using TMHMM, version 2.0 (55). Hairpin sequences that may serve as transcriptional terminators were

identified using TransTermHP (56). BLASTp, hosted within the IMG database, was used to identify *Caldicellulosiruptor* orthologs to all genes in the *C. bescii* T4P operon (Athe_1872 to Athe_1886) (52, 57). Nucleotide sequences of orthologs were aligned using the MUSCLE program with partial gaps at 95%, and a phylogenetic tree was built using the Maximum Likelihood method and Hasegawa-Kishino-Yano model, and the results were confirmed using bootstrapping values of 500 replicates using MEGA version 6 (58). Secondary structures of putative *C. bescii* pilins were predicted using PSIPRED program, version 3.3 (59).

Media, growth conditions and estimation of cell densities

Minimal, low osmolality defined (LOD) medium (60) was used for culturing *C. bescii* DSM 6725 strains on various carbon sources. *C. bescii* DSM 6725 was provided by Robert M. Kelly (North Carolina State University, Raleigh, NC). *C. bescii* cultures were grown at 75°C under anaerobic conditions with one of the following substrates, all at 1 g/l: xylose, Sigmacell (20 µm, Sigma), xylan (beechwood, Sigma), glucomannan (konjac root, NOW Foods). Epifluorescence microscopy (Nikon Eclipse E400) was used to enumerate glutaraldehyde fixed cells stained with acridine orange as previously described (61).

Cloning, production and purification of recombinant PilA

Two truncated versions of PilA (GenBank accession number WP_015908263.1) were cloned into the pET-28b protein expression vector (Novagen). Construct Athe_1880C was truncated after glycine similar to the way mature *C. bescii* pilins are processed in the cells, whereas PilA was truncated after the hydrophobic transmembrane domain as predicted by TMHMM server described above. Oligonucleotide primer sequences involved in the cloning process are listed in **Table 3**. Both the constructs were cloned with an *N*-terminal histidine tag for immobilized nickel affinity purification using 1ml HisTrap columns (GE Healthcare) per the manufacturer's protocol. Both recombinant proteins were produced using auto-induction medium (62) using 50 µg/ml kanamycin and 34 µg/ml chloramphenicol for selection. PilA was eluted using a linear gradient, and eluted fractions were confirmed for purity by SDS-PAGE. PilA was then dialyzed against 50mM sodium phosphate (pH 7.2) using SnakeSkin dialysis tubing (Thermo-Scientific) per the manufacturer's protocol. Final concentrations of PilA was determined using the bicinchoninic acid assay (BCA, Thermo Scientific) per the manufacturer's protocol.

375 **Immunoblots**

Actively growing cultures of *C. bescii* on plant polysaccharides xylan, glucomannan, cellulose and the sugar xylose were sampled throughout growth until cultures reached stationary phase. Growth rates for *C. bescii* on the substrates mentioned above were calculated from this data. Cultures were sampled at early, mid and late exponential growth and diluted 1:1 with sterile
380 glycerol prior to storage at -20° C until further use. Purified recombinant PilA was applied on each membrane as a standard curve for quantitative immunoblot analysis. PVDF membranes used for protein immobilization (Amersham Hybond 0.2 PVDF, GE) were pre-wet with 100% methanol per the manufacturer's instructions and then equilibrated in 1X PBS buffer. Samples were applied to the pre-wetted membrane blocked in PBS-T blocking buffer (PBS with 0.1% Tween-20 (v/v), 5%
385 milk (w/v)) for one hour on an orbital shaker at room temperature. Afterwards, the membrane was rinsed with PBS-T washing buffer. Primary antibody (chicken anti-PilA, GeneTel) was diluted 1:1,000 in PBST and incubated at room temperature. Afterwards, the membrane was washed in PBS-T. Secondary antibody (rabbit anti-chicken HPR conjugate, Immunoreagents) was diluted 1:1000 PBS-T and again incubated at room temperature. The membrane was then washed in
390 PBS-T as previously described. Chemiluminescent detection of secondary antibody used ECL western blotting detection reagents (GE), per the manufacturer's protocol. Membranes were imaged (ChemiDoc Touch Imaging System, Bio-Rad) using standard chemiluminescent settings and analyzed using Image Lab 5.2.1 (Bio-Rad). Three independent replicates were blotted for each time point. Cell densities determined while plotting growth curves were used to calculate the
395 number of cells applied to the membrane at each time point. Protein concentrations were determined using a standard curve (R-squared = 0.98). Protein concentrations were then normalized to the number of cells and compared using a two-sample t-test in R studio statistical software (63) (v.3.3.3).

400 **Immunofluorescence microscopy**

Immunolabelling of *C. bescii* cells followed methods from Conway et al. (9) with modifications. *C. bescii* cells were grown on 50 mL LOD medium in 125 ml serum bottles with either xylan or xylose as the carbon source and harvested at late exponential phase. Cells were pelleted using centrifugation 5000 x g for 10 minutes at room temperature (same conditions for
405 all of the centrifugation steps). Anti-PilA polyclonal antibodies were raised in chicken (chicken anti-PilA, GeneTel), and used as the primary antibody. The secondary antibody is a goat, anti-chicken IgY polyclonal antibody conjugated to DyLight 488 (Novus Biologicals). After the final

washing step, cells were stained with 1µg/ ml DAPI in 1X PBS for 5 minutes at 4°C. Cells were then vacuum-filtered onto 0.2µm polycarbonate track etched membrane filters (GVS Life Sciences) and mounted in Vectashield mounting medium for imaging (Vector laboratories). Epifluorescence imaging used a Nikon eclipse E400 microscope and images were captured using an Infinity 3 Lumenera camera. Control images were captured from secondary antibody labeled samples which were not incubated with the primary antibody. Biomass controls (uninoculated xylan in LOD) were processed similar to samples. Recombinant PilA labeled, xylan was incubated with 30 µM PilA at 300 rpm for an hour at room temperature. PilA was discarded and xylan was processed in the same manner described above. All assays had biological triplicates.

Qualitative and quantitative protein binding assays

For PilA binding assays, the substrates (cellulose or xylan) used were washed with distilled water five times followed by 16h air drying and then at 70°C for two hours. For pilin pull down assays, 10 mg of the substrate was incubated with 100 µl of PilA (30 µM) in the binding buffer, 50 mM sodium phosphate, pH 7.2 (same binding buffer used for all other assays) in a thermomixer (Eppendorf) at 700 rpm for four hours at room temperature. Insoluble substrates were pelleted by centrifugation for one minute (15000 x g). The supernatant (~ 70µl) represented unbound protein. Pelleted substrate was washed with 1 ml of the binding buffer five times. After the final wash, the pellet was resuspended in 70µl of binding buffer representing the bound protein. Both the bound and unbound fractions were analyzed using SDS-PAGE.

Cell binding assays

C. bescii cells cultures were grown to early stationary phase on either xylan or cellulose (1 g l^{-1}) and harvested at 5000 x g for 10 minutes at room temperature. Cells were resuspended in 50 mM sodium phosphate, pH 7.2 to a density of 10^9 cells ml^{-1}. Each experimental condition consisted of a total volume of 1.2 ml comprised of 1 ml *C. bescii* cells, and 0.2 ml of either 30 µM rPilA, BSA or buffer. Washed xylan or cellulose (10 mg) was added to measure the amount of cells bound, and no binding substrate were added to the negative controls. All binding assays were incubated at room temperature with gentle rotary shaking at 100 rpm for one hour. After incubation, planktonic cells were enumerated using epifluorescence microscopy as described above. Each of the four treatments in all cell binding assays had at least three biological replicates. Results from each experiment were analyzed with a two-way ANOVA, using the functions “lm” and “Anova” in Program R (64). Each model tested for main and interactive effects of binding substrate (presence/absence of xylan or cellulose) and protein (presence/absence of rPilA or

BSA) on the planktonic cell density. In this modeling framework, the significance of the interaction term (Substrate×Protein) indicates whether the presence of the protein (rPilA or BSA) influenced the binding affinity of cells to the substrate (xylan or cellulose).

445

ACKNOWLEDGEMENTS

This study was supported by start-up funds to S.E.B.-S. (Oakland University). A.K. was supported in part through funds from the Center for Biomedical Research (OU).

450 **REFERENCES**

1. **Blumer-Schuette SE, Brown SD, Sander KB, Bayer EA, Kataeva I, Zurawski JV, Conway JM, Adams MW, Kelly RM.** 2014. Thermophilic lignocellulose deconstruction. FEMS Microbiol Rev **38:**393-448.
2. **Svetlichnyi VA, Svetlichnaya TP, Chernykh NA, Zavarzin GA.** 1990. *Anaerocellum thermophilum* gen. nov. sp. nov. an extremely thermophilic cellulolytic eubacterium isolated from hot springs in the Valley of Geysers. Microbiology **59:**598-604.
3. **Conway JM, McKinley BS, Seals NL, Hernandez D, Khatibi PA, Poudel S, Giannone RJ, Hettich RL, Williams-Rhaesa AM, Lipscomb GL, Adams MWW, Kelly RM.** 2017. Functional analysis of the glucan degradation locus in *Caldicellulosiruptor bescii* reveals essential roles of component glycoside hydrolases in plant biomass deconstruction. Appl Environ Microbiol **83**.
4. **Dam P, Kataeva I, Yang SJ, Zhou F, Yin Y, Chou W, Poole FL, 2nd, Westpheling J, Hettich R, Giannone R, Lewis DL, Kelly R, Gilbert HJ, Henrissat B, Xu Y, Adams MW.** 2011. Insights into plant biomass conversion from the genome of the anaerobic thermophilic bacterium *Caldicellulosiruptor bescii* DSM 6725. Nucleic Acids Res **39:**3240-3254.
5. **Su XY, Han YJ, Dodd D, Moon YH, Yoshida S, Mackie RI, Cann IKO.** 2013. Reconstitution of a thermostable xylan-degrading enzyme mixture from the bacterium *Caldicellulosiruptor bescii*. Applied and Environmental Microbiology **79:**1481-1490.
6. **Chung D, Pattathil S, Biswal AK, Hahn MG, Mohnen D, Westpheling J.** 2014. Deletion of a gene cluster encoding pectin degrading enzymes in *Caldicellulosiruptor bescii* reveals an important role for pectin in plant biomass recalcitrance. Biotechnol Biofuels **7**.
7. **Yang SJ, Kataeva I, Hamilton-Brehm SD, Engle NL, Tschaplinski TJ, Doeppke C, Davis M, Westpheling J, Adams MWW.** 2009. Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe "*Anaerocellum thermophilum*" DSM 6725. Appl Environ Microbiol **75:**4762-4769.
8. **Kataeva I, Foston MB, Yang S-J, Pattathil S, Biswal AK, Poole FL, Basen M, Rhaesa AM, Thomas TP, Azadi P, Olman V, Safford TD, Mohler KE, Lewis DL, Doeppke C, Zeng Y, Tschaplinski TJ, York W, Davis M, Mohnen D, Xu Y, Ragauskas AJ, Ding S-Y, Kelly RM, Hahn MG, Adams MWW.** 2013. Carbohydrate and lignin are simultaneously solubilized from unpretreated switchgrass by microbial action at high temperature. Energy Environl Sci **6:**2186-2195.
9. **Conway JM, Pierce WS, Le JH, Harper GW, Wright JH, Tucker AL, Zurawski JV, Lee LL, Blumer-Schuette SE, Kelly RM.** 2016. Multidomain, surface layer-associated glycoside hydrolases contribute to plant polysaccharide degradation by *Caldicellulosiruptor* species. J Biol Chem **291:**6732-6747.
10. **Blumer-Schuette SE, Giannone RJ, Zurawski JV, Ozdemir I, Ma Q, Yin Y, Xu Y, Kataeva I, Poole FL, 2nd, Adams MW, Hamilton-Brehm SD, Elkins JG, Larimer FW, Land ML, Hauser LJ, Cottingham RW, Hettich RL, Kelly RM.** 2012. *Caldicellulosiruptor* core and pangenomes reveal determinants for noncellulosomal thermophilic deconstruction of plant biomass. J Bacteriol **194:**4015-4028.
11. **Su X, Mackie RI, Cann IKO.** 2012. Biochemical and mutational analyses of a multi-domain cellulase/mannanase from *Caldicellulosiruptor bescii*. Appl Environ Microbiol doi:10.1128/AEM.06814-11.
12. **Ye L, Su X, Schmitz GE, Moon YH, Zhang J, Mackie RI, Cann IKO.** 2012. Molecular and biochemical analyses of the GH44 Module of CbMan5B/Cel44A, a bifunctional enzyme from the hyperthermophilic bacterium *Caldicellulosiruptor bescii*. Appl Environ Microbiol **78:**7048-7059.

13. **Brunecky R, Alahuhta M, Xu Q, Donohoe BS, Crowley MF, Kataeva IA, Yang S-J, Resch MG, Adams MWW, Lunin VV, Himmel ME, Bomble YJ.** 2013. Revealing nature's cellulase diversity: The digestion mechanism of *Caldicellulosiruptor bescii* CelA. Science **342:**1513-1516.
14. **Yi Z, Su X, Revindran V, Mackie RI, Cann I.** 2013. Molecular and biochemical analyses of CbCel9A/Cel48A, a highly secreted multi-modular cellulase by *Caldicellulosiruptor bescii* during growth on crystalline cellulose. PLoS ONE **8:**e84172.
15. **Xue X, Wang R, Tu T, Shi P, Ma R, Luo H, Yao B, Su X.** 2015. The N-Terminal GH10 domain of a multimodular protein from *Caldicellulosiruptor bescii* is a versatile xylanase/beta-glucanase that can degrade crystalline cellulose. Appl Environ Microbiol **81:**3823-3833.
16. **VanFossen AL, Ozdemir I, Zelin SL, Kelly RM.** 2011. Glycoside hydrolase inventory drives plant polysaccharide deconstruction by the extremely thermophilic bacterium Caldicellulosiruptor saccharolyticus. Biotechnol Bioeng **108:**1559-1569.
17. **Ozdemir I, Blumer-Schuette SE, Kelly RM.** 2012. S-layer homology domain proteins Csac_0678 and Csac_2722 are implicated in plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. Appl Environ Microbiol **78:**768-777.
18. **Wang Z-W, Hamilton-Brehm SD, Lochner A, Elkins JG, Morrell-Falvey JL.** 2011. Mathematical modeling of hydrolysate diffusion and utilization in cellulolytic biofilms of the extreme thermophile *Caldicellulosiruptor obsidiansis*. Bioresour Technol**102:**3155-3162.
19. **Blumer-Schuette SE, Alahuhta M, Conway JM, Lee LL, Zurawski JV, Giannone RJ, Hettich RL, Lunin VV, Himmel ME, Kelly RM.** 2015. Discrete and structurally unique proteins (tapirins) mediate attachment of extremely thermophilic *Caldicellulosiruptor* species to cellulose. J Biol Chem **290:**10645-10656.
20. **Yokoyama H, Yamashita T, Morioka R, Ohmori H.** 2014. Extracellular secretion of noncatalytic plant cell wall-binding proteins by the cellulolytic thermophile Caldicellulosiruptor bescii. J Bacteriol **196:**3784-3792.
21. **Poudel S, Giannone RJ, Basen M, Nookaew I, Poole FL, 2nd, Kelly RM, Adams MWW, Hettich RL.** 2018. The diversity and specificity of the extracellular proteome in the cellulolytic bacterium *Caldicellulosiruptor bescii* is driven by the nature of the cellulosic growth substrate. Biotechnol Biofuels **11:**80.
22. **Lee LL, Hart WS, Lunin VV, Alahuhta M, Bomble YJ, Himmel ME, Blumer-Schuette SE, Adams MWW, Kelly RM.** 2019. Comparative biochemical and structural analysis of novel cellulose binding proteins (tapirins) from extremely thermophilic *Caldicellulosiruptor* species. Appl Environ Microbiol **85**.
23. **Jun HS, Qi M, Gong J, Egbosimba EE, Forsberg CW.** 2007. Outer membrane proteins of *Fibrobacter succinogenes* with potential roles in adhesion to cellulose and in cellulose digestion. J Bacteriol **189:**6806-6815.
24. **Vodovnik M, Duncan SH, Reid MD, Cantlay L, Turner K, Parkhill J, Lamed R, Yeoman CJ, Miller ME, White BA, Bayer EA, Marinsek-Logar R, Flint HJ.** 2013. Expression of cellulosome components and type IV pili within the extracellular proteome of *Ruminococcus flavefaciens* 007. PLoS One **8:**e65333.
25. **Pegden RS, Larson MA, Grant RJ, Morrison M.** 1998. Adherence of the gram-positive bacterium *Ruminococcus albus* to cellulose and identification of a novel form of cellulose-binding protein which belongs to the Pil family of proteins. J Bacteriol **180:**5921-5927.
26. **Rakotoarivonina H, Jubelin G, Hebraud M, Gaillard-Martinie B, Forano E, Mosoni P.** 2002. Adhesion to cellulose of the Gram-positive bacterium *Ruminococcus albus* involves type IV pili. Microbiology **148:**1871-1880.

27. **Christopherson MR, Dawson JA, Stevenson DM, Cunningham AC, Bramhacharya S, Weimer PJ, Kendziorski C, Suen G.** 2014. Unique aspects of fiber degradation by the ruminal ethanologen *Ruminococcus albus* 7 revealed by physiological and transcriptomic analysis. BMC Genomics **15:**1066.
28. **Varga JJ, Nguyen V, O'Brien DK, Rodgers K, Walker RA, Melville SB.** 2006. Type IV pili-dependent gliding motility in the Gram-positive pathogen *Clostridium perfringens* and other Clostridia. Mol Microbiol **62:**680-694.
29. **Melville S, Craig L.** 2013. Type IV pili in Gram-positive bacteria. Microbiol Mol Biol Rev **77:**323-341.
30. **Piepenbrink KH, Maldarelli GA, de la Pena CF, Mulvey GL, Snyder GA, De Masi L, von Rosenvinge EC, Gunther S, Armstrong GD, Donnenberg MS, Sundberg EJ.** 2014. Structure of *Clostridium difficile* PilJ exhibits unprecedented divergence from known type IV pilins. J Biol Chem **289:**4334-4345.
31. **Rodgers K, Arvidson CG, Melville S.** 2011. Expression of a *Clostridium perfringens* type IV pilin by Neisseria gonorrhoeae mediates adherence to muscle cells. Infect Immun **79:**3096-3105.
32. **McKee RW, Aleksanyan N, Garrett EM, Tamayo R.** 2018. Type IV Pili promote *Clostridium difficile* adherence and persistence in a mouse model of infection. Infect Immun **86:**e00943-00917.
33. **Low LY, Harrison PF, Gould J, Powell DR, Choo JM, Forster SC, Chapman R, Gearing LJ, Cheung JK, Hertzog P, Rood JI.** 2018. Concurrent host-pathogen transcriptional responses in a *Clostridium perfringen*s murine myonecrosis infection. MBio **9**.
34. **Zurawski JV, Conway JM, Lee LL, Simpson HJ, Izquierdo JA, Blumer-Schuette S, Nookaew I, Adams MW, Kelly RM.** 2015. Comparative analysis of extremely thermophilic *Caldicellulosiruptor* species reveals common and unique cellular strategies for plant biomass utilization. Appl Environ Microbiol **81:**7159-7170.
35. **Lochner A, Giannone RJ, Rodriguez M, Shah MB, Mielenz JR, Keller M, Antranikian G, Graham DE, Hettich RL.** 2011. Use of label-free quantitative proteomics to distinguish the secreted cellulolytic systems of *Caldicellulosiruptor bescii* and *Caldicellulosiruptor obsidiansis*. Appl Environ Microbiol **77:**4042 -4054.
36. **Pelicic V.** 2008. Type IV pili: e pluribus unum? Mol Microbiol **68:**827-837.
37. **Friedrich C, Bulyha I, Sogaard-Andersen L.** 2014. Outside-in assembly pathway of the type IV pilus system in *Myxococcus xanthus*. J Bacteriol **196:**378-390.
38. **Khare B, S VLN.** 2017. Pilus biogenesis of Gram-positive bacteria: Roles of sortases and implications for assembly. Protein Sci **26:**1458-1473.
39. **Giltner CL, Nguyen Y, Burrows LL.** 2012. Type IV pilin proteins: versatile molecular modules. Microbiol Mol Biol Rev **76:**740-772.
40. **Craig L, Li J.** 2008. Type IV pili: paradoxes in form and function. Curr Opin Struct Biol **18:**267-277.
41. **Piepenbrink KH, Maldarelli GA, Martinez de la Pena CF, Dingle TC, Mulvey GL, Lee A, von Rosenvinge E, Armstrong GD, Donnenberg MS, Sundberg EJ.** 2015. Structural and evolutionary analyses show unique stabilization strategies in the type IV pili of *Clostridium difficile*. Structure **23:**385-396.
42. **Berry JL, Gurung I, Anonsen JH, Spielman I, Harper E, Hall AMJ, Goosens VJ, Raynaud C, Koomey M, Biais N, Matthews S, Pelicic V.** 2019. Global biochemical and structural analysis of the type IV pilus from the Gram-positive bacterium *Streptococcus sanguinis*. J Biol Chem doi:10.1074/jbc.RA118.006917.
43. **Timell TE.** 1967. Recent progress in the chemistry of wood hemicelluloses. Wood Sci Technol **1:**45-70.

44. **Gurung I, Spielman I, Davies MR, Lala R, Gaustad P, Biais N, Pelicic V.** 2016. Functional analysis of an unusual type IV pilus in the Gram-positive *Streptococcus sanguinis*. Mol Microbiol **99:**380-392.
45. **Chen Y-YM, Chiang Y-C, Tseng T-Y, Wu H-Y, Chen Y-Y, Wu C-H, Chiu C-H.** 2019. Molecular and functional analysis of the type IV pilus gene cluster in *Streptococcus sanguinis* SK36. Appl Environ Microbiol **85:**e02788-02718.
46. **Rakotoarivonina H, Larson MA, Morrison M, Girardeau JP, Gaillard-Martinie B, Forano E, Mosoni P.** 2005. The *Ruminococcus albus pilA1-pilA2* locus: expression and putative role of two adjacent pil genes in pilus formation and bacterial adhesion to cellulose. Microbiology-Sgm **151:**1291-1299.
47. **Nataf Y, Bahari L, Kahel-Raifer H, Borovok I, Lamed R, Bayer EA, Sonenshein AL, Shoham Y.** 2010. *Clostridium thermocellum* cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. Proc Natl Acad Sci U S A **107:**18646-18651.
48. **Bahari L, Gilad Y, Borovok I, Kahel-Raifer H, Dassa B, Nataf Y, Shoham Y, Lamed R, Bayer EA.** 2011. Glycoside hydrolases as components of putative carbohydrate biosensor proteins in *Clostridium thermocellum*. J Ind Microbiol Biotechnol **38:**825-832.
49. **Biais N, Higashi DL, Brujic J, So M, Sheetz MP.** 2010. Force-dependent polymorphism in type IV pili reveals hidden epitopes. Proc Natl Acad Sci U S A **107:**11358-11363.
50. **Brissac T, Mikaty G, Dumenil G, Coureuil M, Nassif X.** 2012. The meningococcal minor pilin PilX is responsible for type IV pilus conformational changes associated with signaling to endothelial cells. Infect Immun **80:**3297-3306.
51. **Oki H, Kawahara K, Maruno T, Imai T, Muroga Y, Fukakusa S, Iwashita T, Kobayashi Y, Matsuda S, Kodama T, Iida T, Yoshida T, Ohkubo T, Nakamura S.** 2018. Interplay of a secreted protein with type IVb pilus for efficient enterotoxigenic *Escherichia coli* colonization. Proc Natl Acad Sci U S A **115:**7422-7427.
52. **Chen IA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, Huntemann M, Varghese N, White JR, Seshadri R, Smirnova T, Kirton E, Jungbluth SP, Woyke T, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC.** 2019. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. Nucleic Acids Res **47:**D666-D677.
53. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M.** 2014. Pfam: the protein families database. Nucleic Acids Res **42:**D222-230.
54. **Petersen TN, Brunak S, von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods **8:**785-786.
55. **Krogh A, Larsson B, von Heijne G, Sonnhammer ELL.** 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol **305:**567-580.
56. **Kingsford CL, Ayanbule K, Salzberg SL.** 2007. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. Genome Biol **8:**R22.
57. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.** 2009. BLAST+: architecture and applications. BMC Bioinformatics **10:**421.
58. **Kumar S, Stecher G, Tamura K.** 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol **33:**1870-1874.
59. **Jones DT.** 1999. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol **292:**195-202.

60. **Farkas J, Chung D, Cha M, Copeland J, Grayeski P, Westpheling J.** 2013. Improved growth media and culture techniques for genetic analysis and assessment of biomass utilization by *Caldicellulosiruptor bescii*. J Ind Microbiol Biotechnol **40:**1-9.
61. **Hobbie JE, Daley RJ, Jasper S.** 1977. Use of nuclepore filters for counting bacteria by fluorescence microscopy. Appl Environ Microbiol **33:**1225-1228.
62. **Studier FW.** 2005. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif **41:**207-234.
63. **R Core Team.** 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria., Vienna, Austria. https://www.R-project.org/.
64. **Fox J, Weisberg S.** 2011. An R Companion to Applied Regression, 2nd ed. Sage Publications, Thousand Oaks, CA.

FIGURE LEGENDS

Figure 1. *Caldicellulosiruptor* Type IV pilus genomic context and evolutionary divergence. (A) *C. bescii* T4P locus organization includes the following genes annotated as (from left to right): H.P., hypothetical protein (Athe_1886); response regulator (Athe_1885); PilB, ATPase (Athe_1884), PilT, ATPase (Athe_1883); PilC, secretion system protein (Athe_1882); PilD, pre pilin peptidase (Athe_1879); ComFB, late competence development protein (Athe_1878); PilM, pilus assembly protein (Athe_1875); PilN, pilus assembly protein (Athe_1874); PilO, pilus assembly protein (Athe_1873). Hypothetical proteins with predicted prepilin-type N-terminal cleavage domains (green arrows): Athe_1881, Athe_1880, Athe_1877, Athe_1876, and Athe_1872 and tāpirins (purple arrows) Athe_1871, Athe_1870. Modified from (10). (B) Evolutionary analysis of concatenated type IV pilus for the genus *Caldicellulosiruptor*. A concatenated phylogenetic tree was compiled using the MEGA7 software package. Nucleotide sequences homologous to all genes in the *C. bescii* T4P operon (Athe_1886 to Athe_1872) with > 31% identity, based on BLASTn were selected. Nucleotide sequences were then aligned using MUSCLE with partial gaps at 95%, and a phylogenetic tree was created using the Maximum Likelihood method. Branches are the consensus of 500 bootstraps. Dark red indicates strongly cellulolytic, and blue indicates weakly cellulolytic *Caldicellulosiruptor* species.

Figure 2. Schematic representation of predicted secondary and tertiary structures for putative *C. bescii* pilins. α1-N helix: peach, α1-C helix: aqua, αβ loop: red, β strands: yellow. Extremely hydrophobic regions are represented by magenta lines above the protein backbone. Transmembrane domains in the α1N region are indicated by olive green lines.

Figure 3. Conserved domains present in the predicted major *C. bescii* pilin gene. Athe_1880: complete gene along with sequence encoding for the pre pilin-type N-terminal cleavage/methylation domain- GFxxxE (cyan hexagon), as well as the highly hydrophobic regions (magenta) and the predicted hydrophobic transmembrane domain (green). Athe_1880C: gene was truncated after glycine in the pre-pilin-type N-terminal cleavage/methylation domain so as to resemble the mature pilin. PilA: truncated after the predicted hydrophobic transmembrane domain.

Figure 4. Detection of *C. bescii* PilA protein levels during growth on representative plant polysaccharides. Normalized PilA concentrations are plotted against the y-axis during early, mid and late exponential phases of growth. Each bar is the average from three independent replicates

(± standard error), normalized to the number of cells blotted for each time point. * p-value < 0.05 compared to glucomannan group, # p-value < 0.05 compared to xylose group. Representative plant polysaccharides include xylan (green), glucomannan (purple) and xylose (orange). No PilA
700 was detected during growth on glucomannan and xylose during early exponential phase.

Figure 5: Immunofluorescence detection of extracellular PilA from *C. bescii* cultures grown on xylan. (A) Immunofluorescence microscopy of *C. bescii* cells grown on xylan and labelled with secondary antibody only. (B) Immunofluorescence microscopy of *C. bescii* cells grown on xylan
705 and labelled with primary antibody followed by secondary antibody. (C) Immunofluorescence microscopy of *C. bescii* cells grown on xylose and labelled with secondary antibody. (D) Immunofluorescence microscopy of *C. bescii* cells grown on xylose and labelled with primary antibody followed by secondary antibody. (E) Immunofluorescence microscopy of abiotically-treated xylan, labelled with primary antibody followed by secondary antibody. (F)
710 Immunofluorescence microscopy of abiotically-treated xylan incubated with 30 μM rPilA, labelled with primary antibody followed by secondary antibody. All samples were counterstained with DAPI (blue). Primary antibody: chicken anti-PilA; secondary antibody: DyLight488 conjugated goat anti-chicken (green).

715 **Figure 6. Polysaccharide pulldown assay using *C. bescii* rPilA.** B, bound fraction of PilA , U: unbound fraction of PilA. Polysaccharides tested were xylan, microcrystalline cellulose (Sigmacell) and filter paper. SDS-PAGE images are representatives of the results from assays run in triplicate.

720 **Figure 7. Cell binding assays showing inhibition effects of recombinant PilA (or BSA) on binding of *C. bescii* cells to a xylan or cellulose substrate, in four separate experiments**. Each panel displays planktonic cell densities as a function of two predictor variables: binding-substrate presence/absence and protein presence/absence (e.g., “No PilA” versus “PilA”). (A) *C. bescii* cells from cultures grown on xylan and incubated with xylan as a binding substrate. (B) *C.*
725 *bescii* cells from cultures grown on cellulose and incubated with cellulose as a binding substrate. (C, D) *C. bescii* cells from cultures grown on xylan and incubated with cellulose as a binding substrate. Solid lines indicate planktonic cell densities (PCDs) without substrate whereas dotted lines indicate PCDs with substrate. No PilA/BSA, PCD values when no recombinant PilA or BSA was added; PilA/BSA, either recombinant PilA or BSA was added to the binding assay. Higher
730 PCD values correlate with fewer cells attaching to substrate versus lower values of PCDs which

correlate with more cells attaching to substrate. Two-way ANOVA revealed significant main and interactive effects of binding substrate and PilA on cell densities (all p-values < 0.02; Table S3), indicating that PilA significantly altered the binding-substrate effect (i.e., binding affinity) in all three experiments. In contrast, there was no significant main or interactive effect of BSA on cell density.

Figure 8. Model of the potential role(s) of type IV pili during deconstruction of plant biomass by *C. bescii*. Imported xylooligosaccharides induce expression of the type IV pilus operon (brown arrow, PilA; purple arrows, tāpirins). 1, Tāpirins interact directly with cellulose. 2, Type IV pili interact indirectly with cellulose via tāpirins. 3, Type IV pili interact indirectly with xylan via an as-of-yet unidentified adhesin. Cream colored hexagons, cellooligosaccharides; pink pentagons, xylooligosaccharides**.**

SUPPLEMENTARY DATA

Table S1. Hairpin sequences present in the type IV pilus locus

Table S2. Percent identity (amino acid) of each PilA homolog across 13 sequenced *Caldicellulosiruptor* genomes.

Table S3. Cell attachment analysis of variance results

Figure S1. Alignment of *Caldicellulosiruptor bescii* hypothetical proteins with putative pilin leader sequences. The conserved glycine (G, purple) and glutamic acid (E, green) from the pre-pilin N-terminus cleavage domain (GFxxxE) are highlighted.

Figure S2. Experimental design for the *C. bescii* cell binding assay. Blue rod: *C. bescii* cell, green circle: substrate. Four different treatment groups were set up to test the effect of two independent variables- substrate and protein. For the substrate variable the two treatments levels were: cells with binding buffer (50 mM sodium phosphate pH= 7.2) alone and cells with binding buffer and substrate. For the protein variable the two treatment levels were, cells with PilA alone and cells with PilA and substrate.

Table 1. Gene annotations for candidate pilins in *C. bescii*

Locus tag (Athe_)	IMG[a]	Pfam classification	Pfam ID	Length (aa)[c]	MW (kDa)	LP (aa)[d]
1872	Hypothetical protein	Prokaryotic N-terminal methylation motif	07963	277	31.49	5
1876	Hypothetical protein	NA[b]		175	19.91	15
1877	Hypothetical protein	Prokaryotic N-terminal methylation motif	07963	179	19.87	5
1880	T4P assembly protein PilA	Prokaryotic N-terminal methylation motif	07963	130	14.2	15
1881	T4P assembly protein PilA	Prokaryotic N-terminal methylation motif	07963	133	14.99	21

[a] IMG, Integrated Microbial Genomes database
[b] NA, not assigned
[c] Protein length (aa, amino acids) and MW, molecular weight are for the predicted mature pilin
[d] LP, leader peptide sequence length in apoprotein (see **Fig. S1**)

Table 2. Predicted *C. bescii* pilin gene and protein expression

Locus tag (Athe_)	Gene Expression (LSMeans)[a]				Protein Abundance (Rank)[b]			
	MCC[c]	SWG[c]	FP[d]	Glu[d]	MCC[e]	MCC[f]	Cel[d]	Xy[d]
1872	1.3	1.1	-0.9	0.7	185	124	163	n.d.[g]
1876	0.4	1.1	-2.5	-3.6	552	n.d.	n.d.	n.d.
1877	1.6	2.2	-0.1	0.0	251	250	n.d.	n.d.
1880	4.2	4.9	1.8	2.8	4	5	44	24
1881	5.7	6.0	3.4	6.2	99	n.d.	n.d.	n.d.

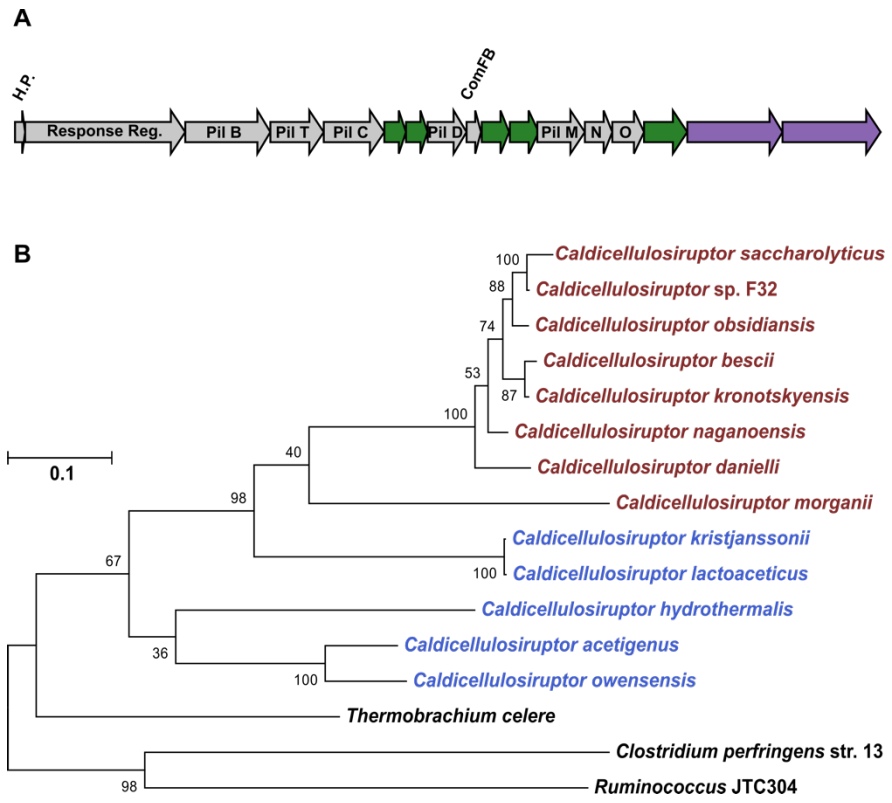
[a] MCC, microcrystalline cellulose; SWG, switchgrass; FP, filter paper; Glu, glucose
[b] MCC, microcrystalline cellulose; Cel, cellulose; Xy, xylan
[c] Gene expression data from Zurawski et al., (34)
[d] Gene expression and proteomics data from Dam et al., (4)
[e] Proteomics data from Blumer-Schuette et al., (10)
[f] Proteomics data from Lochner et al., (35)
[g] n.d., not detected

765

Table 3. Primers designed to amplify and clone PilA

Primer[a]	Primer Sequence[b]	Description	Plasmid
pVJB2-F	CGCCATATGTTTACATTAATTGAGATGGTTATAG	PC truncation	pVJB2
pVJB3-F	CGCCATATGCAAGTATTAAAACAGATAAAC	TMD truncation	pVJB3
pVJB2/3-R	GCTACTCGAGTTACTTGTAGTTTGGATCTGG	Reverse	pVJB2/3

[a] Abbreviations used: F, forward; R, reverse; TMD, transmembrane domain; PC, predicted pilin cleavage site
[b] Restriction enzyme recognition sites for NdeI (CATATG) and XhoI(CTCGAG) are underlined.



770 **Figure 1.** *Caldicellulosiruptor* Type IV pilus genomic context and evolutionary divergence. (A) *C. bescii* T4P locus organization includes the following genes annotated as (from left to right): H.P., hypothetical protein (Athe_1886); response regulator (Athe_1885); PilB, ATPase (Athe_1884), PilT, ATPase (Athe_1883); PilC, secretion system protein (Athe_1882); PilD, pre pilin peptidase (Athe_1879); ComFB, late competence development protein (Athe_1878); PilM, pilus assembly
775 protein (Athe_1875); PilN, pilus assembly protein (Athe_1874); PilO, pilus assembly protein (Athe_1873). Hypothetical proteins with predicted prepilin-type N-terminal cleavage domains (green arrows): Athe_1881, Athe_1880, Athe_1877, Athe_1876, and Athe_1872 and tāpirins (purple arrows) Athe_1871, Athe_1870. Modified from (10). (B) Evolutionary analysis of concatenated type IV pilus for the genus *Caldicellulosiruptor*. A concatenated phylogenetic tree
780 was compiled using the MEGA7 software package. Nucleotide sequences homologous to all genes in the *C. bescii* T4P operon (Athe_1886 to Athe_1872) with > 31% identity, based on BLASTn were selected. Nucleotide sequences were then aligned using MUSCLE with partial gaps at 95%, and a phylogenetic tree was created using the Maximum Likelihood method. Branches are the consensus of 500 bootstraps. Dark red indicates strongly cellulolytic, and blue
785 indicates weakly cellulolytic *Caldicellulosiruptor* species.

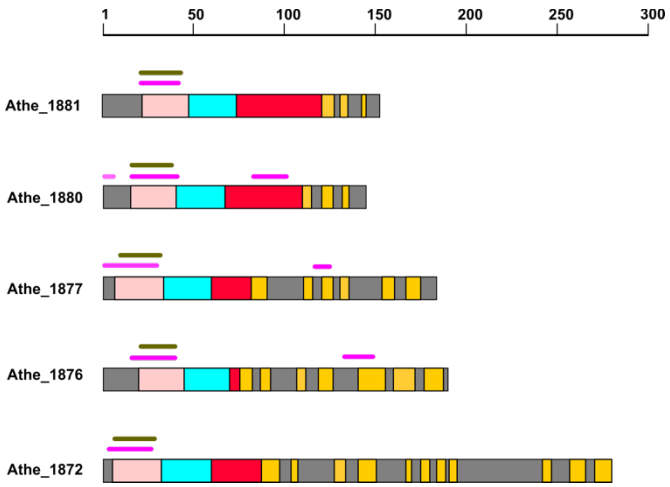


Figure 2. Schematic representation of predicted secondary and tertiary structures for
790 **putative *C. bescii* pilins.** α1-N helix: peach, α1-C helix: aqua, αβ loop: red, β strands: yellow. Extremely hydrophobic regions are represented by magenta lines above the protein backbone. Transmembrane domains in the α1N region are indicated by olive green lines.

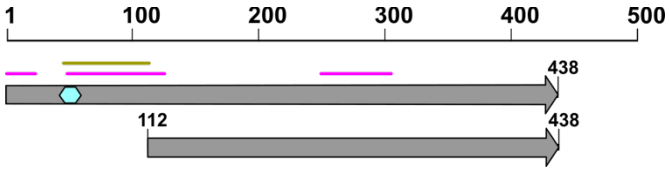
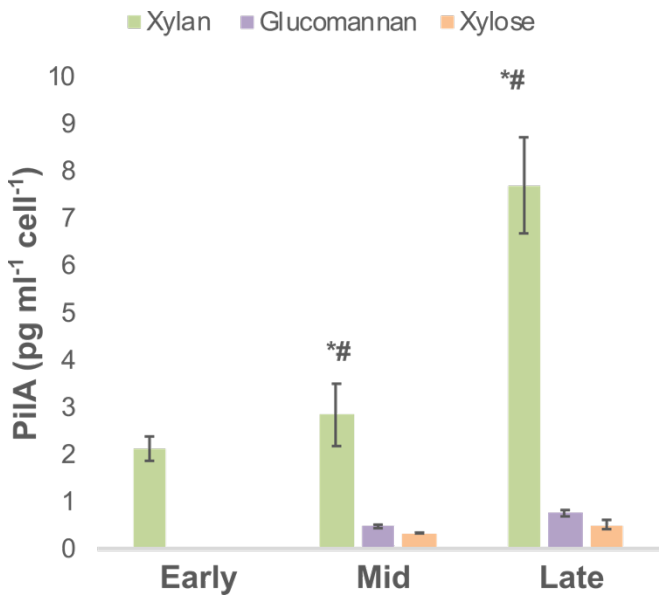
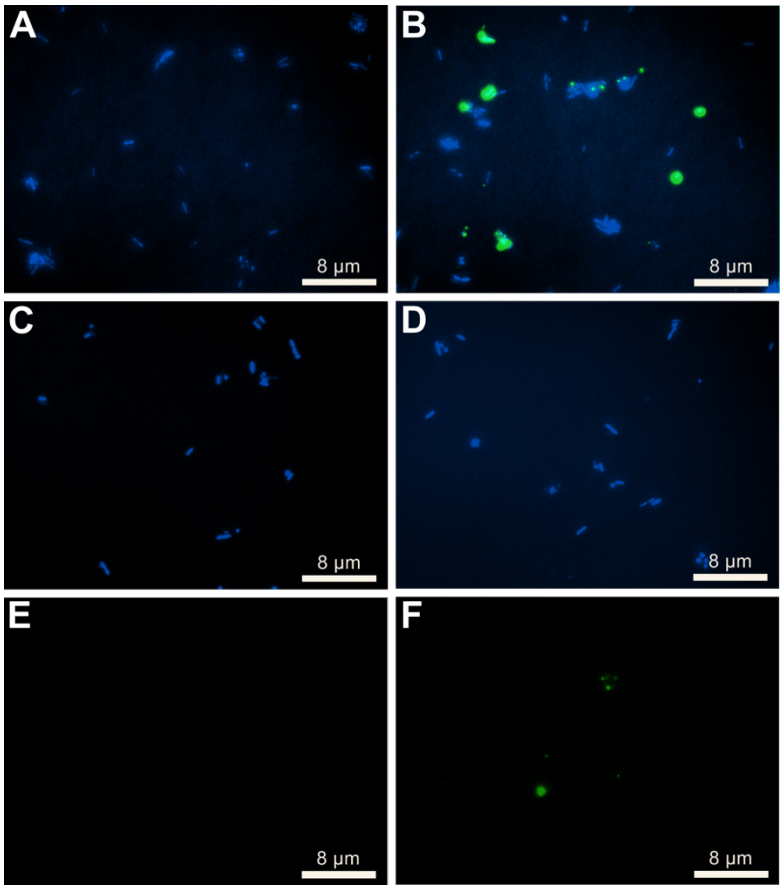


Figure 3. Conserved domains present in the predicted major *C. bescii* pilin gene. Athe_1880: complete gene along with sequence encoding for the pre pilin-type N-terminal cleavage/methylation domain- GFxxxE (cyan hexagon), as well as the highly hydrophobic regions (magenta) and the predicted hydrophobic transmembrane domain (green). Athe_1880C: gene was truncated after glycine in the pre-pilin-type N-terminal cleavage/methylation domain so as to resemble the mature pilin. PilA: truncated after the predicted hydrophobic transmembrane domain.



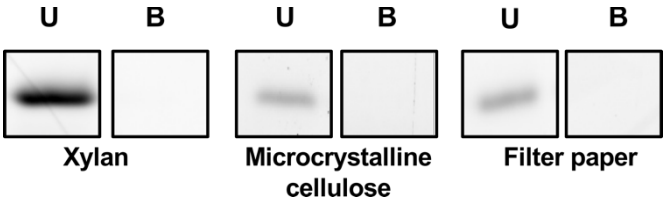
805

Figure 4. Detection of *C. bescii* PilA protein levels during growth on representative plant polysaccharides. Normalized PilA concentrations are plotted against the y-axis during early, mid and late exponential phases of growth. Each bar is the average from three independent replicates (± standard error), normalized to the number of cells blotted for each time point. * p-value < 0.05
810 compared to glucomannan group, # p-value < 0.05 compared to xylose group. Representative plant polysaccharides include xylan (green), glucomannan (purple) and xylose (orange). No PilA was detected during growth on glucomannan and xylose during early exponential phase.



815

Figure 5. Immunofluorescence detection of extracellular PilA from *C. bescii* cultures grown on xylan. (A) Immunofluorescence microscopy of *C. bescii* cells grown on xylan and labelled with secondary antibody only. (B) Immunofluorescence microscopy of *C. bescii* cells grown on xylan and labelled with primary antibody followed by secondary antibody. (C)
820 Immunofluorescence microscopy of *C. bescii* cells grown on xylose and labelled with secondary antibody. (D) Immunofluorescence microscopy of *C. bescii* cells grown on xylose and labelled with primary antibody followed by secondary antibody. (E) Immunofluorescence microscopy of abiotically-treated xylan, labelled with primary antibody followed by secondary antibody. (F) Immunofluorescence microscopy of abiotically-treated xylan incubated with 30 µM rPilA,
825 labelled with primary antibody followed by secondary antibody. All samples were counterstained with DAPI (blue). Primary antibody: chicken anti-PilA; secondary antibody: DyLight488 conjugated goat anti-chicken (green).



830

Figure 6. Polysaccharide pulldown assay using *C. bescii* rPilA. B, bound fraction of PilA , U: unbound fraction of PilA. Polysaccharides tested were xylan, microcrystalline cellulose (Sigmacell) and filter paper. SDS-PAGE images are representatives of the results from assays run in triplicate.

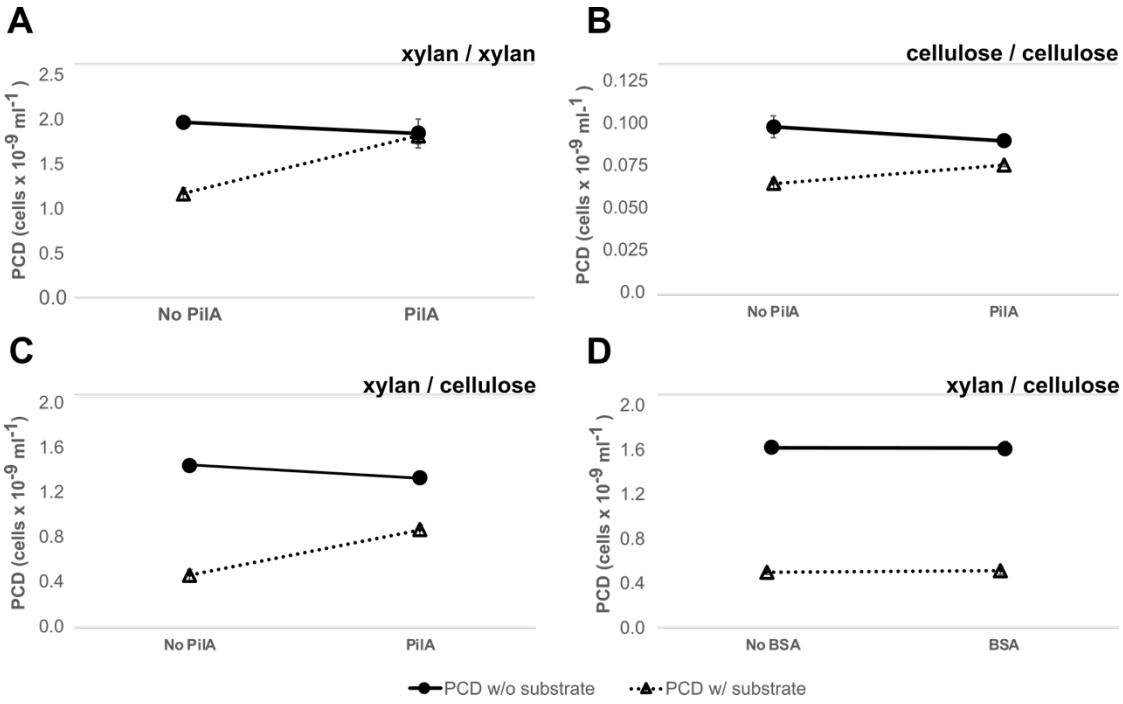
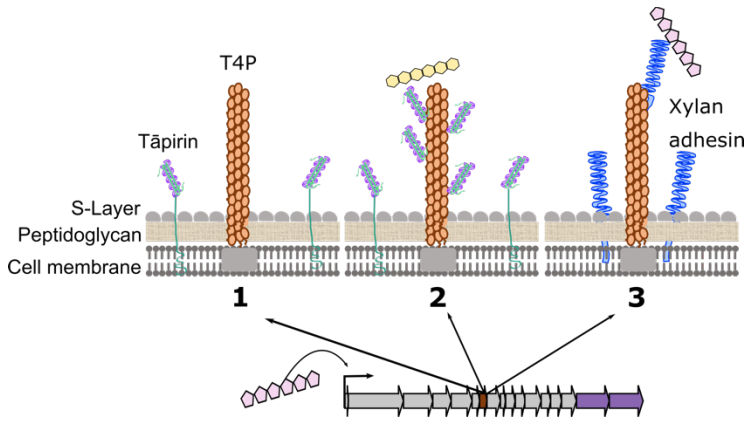


Figure 7. Cell binding assays showing inhibition effects of recombinant PilA (or BSA) on binding of *C. bescii* cells to a xylan or cellulose substrate, in four separate experiments. Each panel displays planktonic cell densities as a function of two predictor variables: binding-substrate presence/absence and protein presence/absence (e.g., “No PilA” versus “PilA”). (A) *C. bescii* cells from cultures grown on xylan and incubated with xylan as a binding substrate. (B) *C. bescii* cells from cultures grown on cellulose and incubated with cellulose as a binding substrate. (C, D) *C. bescii* cells from cultures grown on xylan and incubated with cellulose as a binding substrate. Solid lines indicate planktonic cell densities (PCDs) without substrate whereas dotted lines indicate PCDs with substrate. No PilA/BSA, PCD values when no recombinant PilA or BSA was added; PilA/BSA, either recombinant PilA or BSA was added to the binding assay. Higher PCD values correlate with fewer cells attaching to substrate versus lower values of PCDs which correlate with more cells attaching to substrate. Two-way ANOVA revealed significant main and interactive effects of binding substrate and PilA on cell densities (all p-values < 0.02; Table S3), indicating that PilA significantly altered the binding-substrate effect (i.e., binding affinity) in all three experiments. In contrast, there was no significant main or interactive effect of BSA on cell density.



855

Figure 8. Model of the potential role(s) of type IV pili during deconstruction of plant biomass by *C. bescii*. Imported xylooligosaccharides induce expression of the type IV pilus operon (brown arrow, PilA; purple arrows, tāpirins). 1, Tāpirins interact directly with cellulose. 2,
860 Type IV pili interact indirectly with cellulose via tāpirins. 3, Type IV pili interact indirectly with xylan via an as-of-yet unidentified adhesin. Cream colored hexagons, cellooligosaccharides; pink pentagons, xylooligosaccharides.