Positive Selection Determination and Twitching Motility Efficacy on Various Carbohydrates by *Caldicellulosirputor bescii*

Submitted by

Hwaylee Downey

Biology

To

The Honors College

Oakland University

In partial fulfillment of the

requirement to graduate from

The Honors College

Mentor: Dr. Sara Blumer-Schuette

Department of Biological Sciences and Bioengineering

Oakland University

3 December 2021

TABLE OF CONTENTS

ABSTRACT

Caldicellulosiruptor bescii is a species of hyperthermophilic, gram-positive bacteria most notable for its ability to hydrolyze plant biomass. It is hypothesised that protein glycosylation plays a role in the ability of *C. bescii* to attach to plant polysaccharides and form cell-to-cell connections. Recent studies demonstrated the genetic conservation of type IV pili (T4P) and plant biomass deconstruction-related proteins. This project aims to further confirm the conservation by identifying positively selected *C. bescii* proteins involved in attachment and if they are glycosylated. It will also identify the functional significance of T4P in twitching motility using well-established polysaccharide binding and plating assays. The major pilin PilA was not observed to be glycosylated. The protein sequence analysis suggests an insignificant amount of positive selection of carbohydrate deconstruction-related proteins in the *Caldicellulosiruptor* genus overall. I also observed a significant increase in colony formation by twitching motility on xylan compared to other carbohydrate treatments. Overall, this knowledge may contribute to our understanding of twitching motility as a whole.

Keywords: Glycosylation, anaerobe, type IV pili, twitching motility

INTRODUCTION

Caldicellulosiruptor bescii is a species of gram-positive bacteria that was isolated from terrestrial hot springs, and as a result grows at high temperatures anaerobically (Yang et al., 2009). *C. bescii* is most notable for its ability to hydrolyse plant biomass and can serve as a model extreme thermophile that produces a type IV pilus. *C. bescii* hydrolyzes polysaccharides, primarily cellulose and xylan that are abundantly found in plant biomass, for energy (Dam et al., 2011). It utilizes type IV pili (T4P) (Khan et al., 2020) and tāpirins (Blumer-Schuette et al., 2015) to adhere to such polysaccharides, and cellulase enzymes to deconstruct cellulose when bound (Brunecky et al., 2013, Conway et al., 2018). The species also relies on many extracellular proteins for binding and breakdown of many complex polysaccharides (Rodoinov et al., 2021). *C. bescii* has fewer genes encoding glycoside hydrolases used by other species in the *Caldicellulosiruptor* genus to catabolise plant biomass (Khan et al., 2019). Therefore the evolution and mechanisms of T4P and glycoside hydrolases are subjects of interest when analyzing *C. bescii*'s capacity to adhere to and hydrolyse plant biomass.

Although it has been demonstrated that *C. bescii* cellulases largely contribute to plant biomass hydrolysis, there is uncertainty as to whether glycosylation of these enzymes impacts their effectiveness. Glycosylation has been observed to protect proteins subject to proteolytic degradation (Chung et al., 2019). And in other bacteria, pilins and flagellins are glycosylated, lending the possibility that *C. bescii* may also have glycosylated T4P.

Previous studies suggest that T4P are an essential component in *C. bescii*'s ability to adhere to various carbohydrates (Blumer-Schuette et al., 2012, Khan et al., 2020). Other gram-positive bacteria, primarily in the order *Clostridiales*, rely on T4P for motility and

adhesion, and components of the pilus are seemingly conserved (Piepenbrink et al., 2016). The type IV pilus gene locus was indeed observed to be highly upregulated and conserved in highly cellulolytic species of the genus *Caldicellulosiruptor* (*C. bescii*, *C. kronotskyensis*, *C. obsidiansis*, and *C. saccharolyticus*), signifying its importance in their physiology (Khan et al., 2020; Blumer-Schuette et al., 2012). Performing a positive selection analysis using PosiGene would further confirm whether the genes are conserved or evolving.

It has been experimentally determined that the main type IV pilin interacts with polysaccharides at high, physiological temperatures (~75˚C) rather than the standard state temperature of 25°C (Khan et al., 2020). In such high temperature conditions, T4P interacts directly with xylan, but appears to work synchronously with tāparins to adhere to cellulose (Khan et al., 2020). Another bacterial species, *Acinetobacter baumannii* produces a T4P whose main pilin, PilA, was found responsible for twitching motility, biofilm formation, and contributed to host cell adherence (Ronish et al., 2018). It is possible that pili in *C. bescii* may also rely on T4P for twitching motility and possibly surface adherence. Plating *C. bescii* on solid media will determine whether the species is capable of twitching motility. *C. bescii* have stringent growth requirements on solid media and therefore the plates will be optimised based on their need for high temperatures, anaerobic conditions, and ability to form 3-D colonies.

This study will evaluate the presence of positively selected protein-coding genes among 14 species of *Caldicellulosiruptor* to determine the extent of pilin and sugar adherence gene conservation. It is predicted that genes responsible for motility or sugar adherence would be highly conserved within the *Caldicellulosiruptor* genus. The anaerobic plating optimization for twitching motility on various carbohydrates, cellobiose, cellulose, corn core xylan, xylose, will help identify to which sugars *C. bescii* has the greatest affinity. I hypothesize that the T4P major

pilin in *C. bescii* is glycosylated and that twitching motility will be more effective on complex carbohydrates like xylan and cellulose as the species has developed specialized mechanisms (T4P and tāpirins) for deconstructing plant biomass.

METHODS AND MATERIALS

Genetic Analysis

Phylogenetic Tree Assembly. Protein coding sequences for all 14 *Caldicellulosiruptor* species were obtained from the NCBI Assemblies GenBank database (Appendix 1). The protein FASTA (.faa) sequences were aligned and analysed using PosiGene v0.1. (Sahm et al., 2017) with default settings.

Positive Selection Test. *Caldicellulosiruptor* FASTA headers were shortened to print only the locus tag using awk and sed text editors (Appendix 1). A slurm script (Appendix 2) to run PosiGene was submitted to Oakland University's high performance computing cluster (HPCC), Matilda, following a program tutorial as described by Sahm et al. (2017). The command included file locations of coding sequences (CDS) for -reference-species (-rs) and -non-homologene species-by-reference (-nhsbr). File names were called for -target-species (-ts) and -anchor-species (-as). *C. bescii* was used as the reference and anchor species in all runs. The first run called for *C. bescii* as the target species only. The second run included both *C. bescii* and *C. owensensis* as target species, which tested their most recent common ancestor (CBO LCA). The third run included *C. bescii* and *C. changbaiensis* as target species to provide their last common ancestor (CBC LCA). The output files included locus tags, FDR value, P-value, the number of species represented in the alignment, and the number of sites under positive selection.

Positive Selection Analysis. All outputs from PosiGene were filtered by positive selection with a confidence of $p < 0.05$. The remaining genes from the three PosiGene runs were

added to JVenn (Bardou et al., 2014) to identify shared positively selected genes. Locus tags were compared to protein annotations provided in GenBank Accession GCA_000022325.1.

Twitching Motility Analysis

Bacterial Media and Culturing. *C. bescii* DSMZ 6725 stored in a -80°C freezer was inoculated into low osmolarity complex growth medium (LOC) with cellobiose and incubated at 70°C to be revived. 0.5mL of the inoculant was subcultured into low osmolarity defined growth medium (LOD), as defined by Farkas et al. (2013) with cellobiose (CB) and incubated at 75°C. Revived *C. bescii* was passaged from cellobiose to corn core xylan (XyCC) via 2 0.5mL subcultures into LOD 1:1 cellobiose:XyCC before subculturing on LOD XyCC. Cultures were incubated for 15-24 hours. All media went through a round of vacuum filtration prior to 3 rounds of nitrogen and vacuum cycling to make them anaerobic. The inclusion of resazurin indicated anaerobic conditions via color. Revived *C. bescii* cultures were subcultured into other carbohydrates based on polysaccharide: XyCC subcultured into xylose media; CB subcultured into cellulose.

Cell Counting. For microscopy, *C. bescii* cells were fixed using 2.5% glutaraldehyde and stained with 1g/L acridine orange (Hobbie et al., 1977) . After filtering the solution, cells were counted using a Nikon C-SHG1 phase contrast microscope (Appendix 3). Ten counts were obtained per sample for 9 successive subcultures after 8-15 hours of incubation. and averaged to calculate cell density as described by the formula (Average cell count * 2.1 x 10^7/Volume of cells added (μL)). Average maintenance cell density was calculated using subcultures #6-9.

Glycostaining. T4P glycoprotein staining was done using an enhanced Coomassie-stained SDS-PAGE Gel following the G-Biosciences protocol (2012).

Swimming Motility Plating. Plating for swimming motility typically uses ~0.3% agar (Ramos-Aires et al., 2004). I used a 0.34% agar concentration with LOD to observe swimming motility. Each plate contained 25mL of agar solution that rested for a minimum of 30 minutes before adding *C. bescii* suspension. The suspensions underwent 2 rounds of serial dilution by adding 100µL of culture into 900µL of 0.9% NaCl blanks and tested all 3 concentrations. 100µL of the suspensions were added to the set plates and spread using 5 glass beads. A top layer of LOD + 0.34% agar (5mL) was added to each plate. To maintain anaerobic conditions, Mitsubishi Anaeropack • CO_2 was added to a rubber-sealed glass container with petroleum applied to the rims. They were incubated anaerobically at 60°C for 7 days. The agar consistently lost its form during incubation and the swimming motility plates were thus not viable.

Twitching Motility Plating. My original twitching motility plating methods involved culturing *C. bescii* on solid media in accordance with Farkas et al. (2013). I tested agar concentrations of 1.0%, 1.3%, and 1.5%. Solidified plates containing 25mL of bottom agar were half inoculated using a micropipette with 1µL of serial diluted (tested original concentration, dilution #1, dilution #2, and dilution #3) *C. bescii* liquid suspension while the other plates were inoculated using a toothpick. A 5mL layer of top agar at the same concentration was added. Once solidified, the plates were inverted and placed into a rubber-sealed glass container with a Mitsubishi Anaeropack • CO_2 packet and incubated at 60°C for 5 days. Visible colonies were transferred to new plates containing 25mL of solidified 1.0% agar using a toothpick. The toothpicks were pushed down to the bottom of the plate. Plates were placed in an anaerobic rubber-sealed glass container and incubated at 60°C for 7 days. Resulting growth was limited and inconsistent.

Optimized Twitching Motility Plating. Each anaerobic plate included 25mL of LOD plus 1.0% agar solution (Ramos-Aires et al., 2004). The plates rested for 30 minutes until the agar solidified. Four quadrants were drawn on each plate, each containing a single well as described by Biswas (2019). Wells were created by inserting a 1mL micropipette tip through the agar to the bottom of the plate and gently swirling the tip. The removal of the pipette tip resulted in an empty well for the cell suspension. Cell suspension (5μL) was added to the bottom of each well. All plates were placed into petroleum-covered, rubber-sealed glass containers containing a Mitsubishi Anaeropack • CO_2 and incubated at 60°C for 7 days.

Twitching Motility Staining. After incubation**,** the plates were allowed to cool to room temperature. The agar was removed from the plates by dragging a 1mL pipette tip around the edge of the dishes. A 0.1% crystal violet stain was created by dissolving 0.1g of crystal violet in 20mL of ethanol and adding 80mL DI water (Biswas et al., 2019). 5mL of crystal violet stain were added to the empty petri dishes and incubated at room temperature for 1 hour. The plates were repeatedly rinsed with 10mL Tris-buffered saline until the stain only remained on the culture. The plates were left to air dry for 1 day. There were 25-33 replicates of each treatment; replicates with no growth were not considered in the data analysis.

ImageJ Area Calculations. Images of each dried twitching plate were taken with a ruler within the frame and uploaded to ImageJ (Appendix 4a). The ruler was used to set the scale for each photo by using the *Straight Line* tool and drawing a line between 0-1cm. *Known Distance* (1) and *Unit of Length* (cm) were adjusted within *Set Scale*. The image *Colour Threshold* was adjusted to highlight growth areas (Appendix 4b). Any gaps within the bacterial growth area were filled in using the *Paintbrush*. Selections from the *Wand Tracing* tool were measured via *Measure*.

Statistical Calculations. Twitching surface area values were analysed using one-way ANOVA with post-hoc LSD tests. Standard errors were calculated to determine uncertainty.

RESULTS

Genetic Analysis

Phylogenetic Tree. The computed unrooted phylogenetic tree contained 3 groups of sister species most related to one another: *C. owensensis* and *C. obsidiansis*, *C. bescii* and *C. kronotskyensis*, *C.* sp. F32 and *C. saccharolyticus* (Fig 1). There were 12 nodes representing common ancestors of the identified species, similar to other genomic studies. CBO and CBC LCA both include weakly (*C. owensensis* and *C. hydrothermalis*) and strongly cellulolytic species (*C. obsidiansis, C. bescii, C kronotskyensis* and *C. obsidiansis, C. bescii, C kronotskyensis, C. danielii, C. morganii, C. naganoensis, C. changbaiensis, C.* spp. F32*, C. saccharolyticus*, respectively). CBC LCA encompasses North American, Russian, Japanese, Chinese, and New Zealand descendant strains while CBO LCA only involves North American and Russian strains (Blumer-Schuette et al., 2012; Khan et al., 2019). CBC LCA thus has a higher sequence difference compared to the more narrowed CBO LCA. Comparing these two ancestors to *C. bescii* would help pinpoint where the species' highly cellulolytic lifestyle evolved if it was positively selected. At a protein sequence level, the differences in highly vs weakly cellulolytic *Caldicellulosiruptor* species are significant (Khan et al., 2019), and therefore I'd expect to observe positive selection of pilin-related proteins.

Positive Selection. PosiGene analysis would identify whether pilin evolution was a result of positive selection. If T4P-related proteins were not subject to positive selection, it suggests that the highly cellulolytic characteristics of *C. bescii* may otherwise be a result of genetic drift, or that I did not compare the correct ancestor that *C. bescii* evolved from. The PosiGene analysis results identified 134 total positively selected amino acid sequences: 47 proteins in *C. bescii*, 37

proteins in CBO LCA, and 50 in CBC LCA. Two main protein groups were identified from and isolated: motility and metabolism ($p < 0.05$, Table 1). A glycoside hydrolase family 2 protein (CEP58_RS02380) appeared to be positively selected along both *C. bescii* (3 sites; FDR: 1.90×10^{-2}, $p = 1.99 \times 10^{-3}$) and CBC LCA (17 sites; FDR: 1.90×10^{-4}, $p = 144 \times 10^{-4}$) nodes. An endo-1, 4-beta-xylanase in the CBO LCA run appeared to have 8 sites under positive selection (FDR: 1.53×10^{-2}, $p = 2.75 \times 10^{-3}$). A PilM pilus assembly protein was positively selected along the *C. bescii* node(1 site; FDR: 1.95×10^{-2}, $p = 2.75 \times 10^{-3}$). T4P twitching motility protein PilT was not found to be positively selected in any of the 3 runs.

Motility Analysis

Growth on Different Sources of Xylan. *C. bescii* was subcultured from beechwood xylan (XyBW) to XyCC liquid cultures. XyBW is insoluble and therefore would potentially hinder plating, whereas XyCC is soluble and can be easily incorporated into plates. The two types of xylan have structural differences since XyCC originates from a monocot and XyBW is extracted from a hardwood tree. Previous experiments observed global transcriptome differences when *C. saccharolyticus* was grown on oat spelt xylan vs. birchwood xylan (Sara Blumer-Schuette, personal communication, December 1, 2021), suggesting XyCC and XyBW would result in similar differences. The chemical differences in XyCC and XyBW were great enough that it took 6 passages before *C bescii* grew at cell densities we would expect on XyBW (Figure 2).

Glycostain. To test whether or not the *C. bescii* native T4 pilin was glycosylated, we used a glycan staining procedure. Native *C. bescii* pilin preparations were kindly provided by Asma Khan. Native PilA was not stained by the glycostain, and as expected the recombinant form produced in *E. coli* was also negative for glycans (Figure 3). Large proteins, likely

modular, multi-functional enzymes, including cellulases, were glycosylated as expected (Conway et al., 2018). Since the null hypothesis was supported, I instead focused on whether the overexpression of the T4 pilins in response to xylan would result in differences in twitching motility.

Twitching Surface Area. Because *C. bescii* T4P expression varies based on the type of carbohydrate present, specifically xylan, I expected there to be greater indication of twitching motility on xylan, and the polysaccharide cellulose due it also being a highly crystalline carbohydrate. The results did demonstrate significant levels of twitching motility when *C. bescii* was grown on xylan, but not on cellulose or any other sugar treatment. Surface area on XyCC treatments (0.59 ± 0.04; range: 0.84) was significantly greater compared to xylose (0.45 ± 0.03; range: 0.66), CB; (0.47 ± 0.05; range: 0.78), and cellulose (0.42 ± 0.04, range: 0.95), which did not differ ($F_{(3,112)} = 4.18$, $p = 0.007$; LSD test, $p = 0.05$).

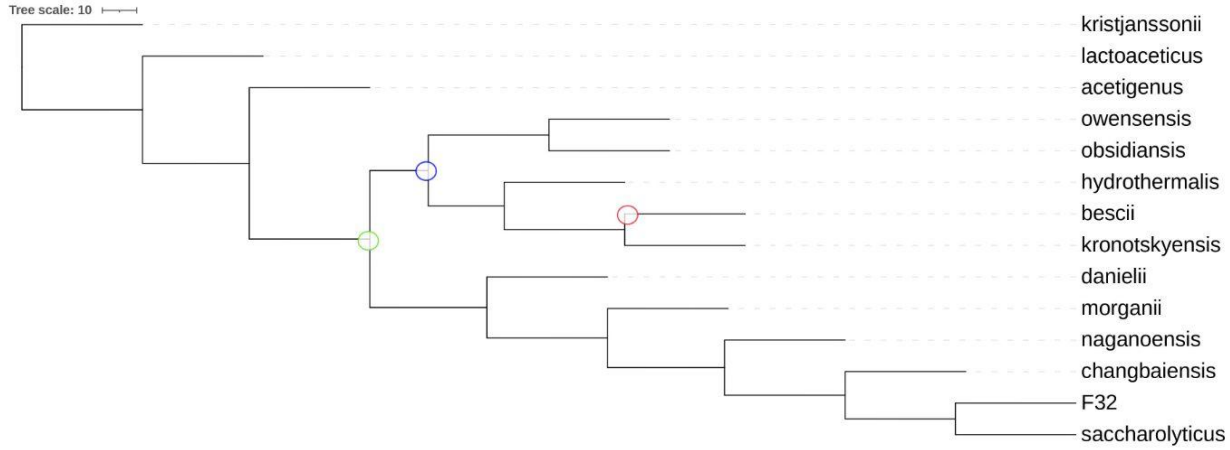


Figure 1 Unrooted phylogenetic tree using protein coding sequences of the genus *Caldicellulosiruptor* as predicted by PosiGene. *C. bescii* is located at the red circle. *C. bescii*/*owensensis* (CBO) LCA is located at the blue circle. *C. bescii*/*changbaiensis* (CBC) LCA is found at the green circle.

Table 1 Positively selected genes relevant to motility and metabolism out of a total of 134 positively selected proteins determined by PosiGene ($P < 0.05$). The presence of more than one row with the same gene indicates it was found to be positively selected along 2 nodes.

Species	**Locus Tag**	**# Sites Under Positive Selection**	**Protein Name**
C. bescii	CEP58_RS02380	3	glycoside hydrolase family 2 protein
	CEP58_RS10315	4	glycosyltransferase
	CEP58_RS10380	1	glycosyltransferase family 2 protein
	CEP58_RS10715	1	pilus assembly protein PilM
	CEP58_RS02115	2	extracellular solute-binding protein
CBO LCA	CEP58_RS02155	3	extracellular solute binding protein
	CEP58_RS02160	8	endo-1,4-beta-xylanase
	CEP58_RS12155	10	flagellar type III secretion system protein FliR
CBC LCA	CEP58_RS02380	17	glycoside hydrolase family 2 protein
	CEP58_RS03995	4	glycogen/starch/alpha-glucan phosphorylase
	CEP58_RS09570	3	flagellar protein FlaG
	CEP58_RS05545	6	alpha-glucuronidase

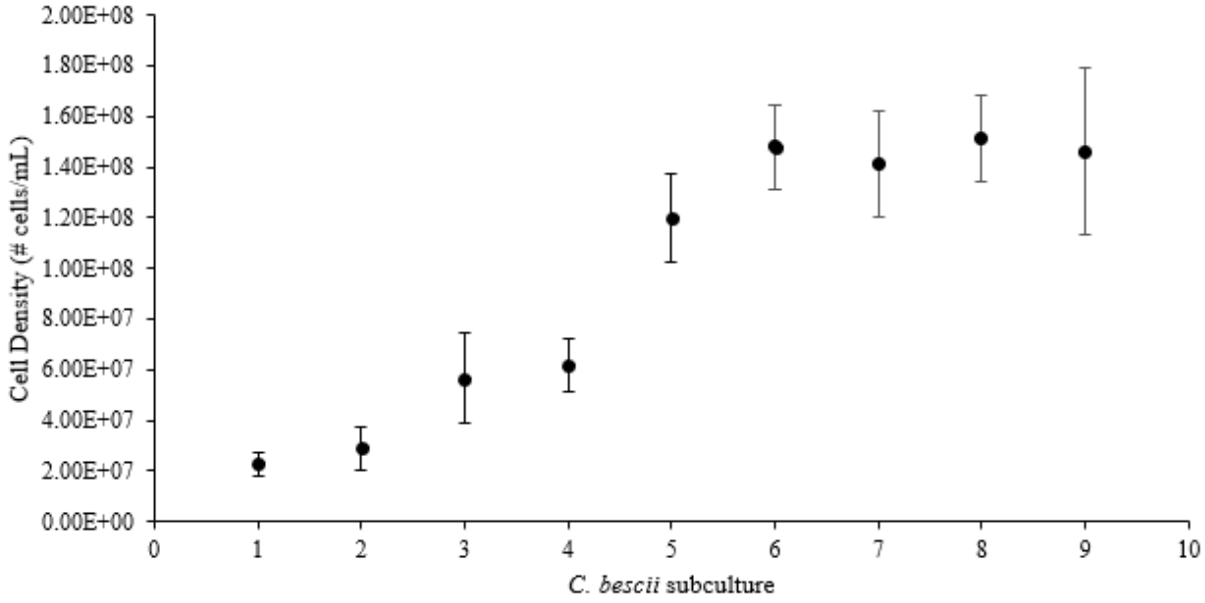


Figure 2 Calculated cell density of different subcultures of *C. bescii* in liquid 2X LOD XyCC media. Subculture 1 was inoculated with *C. bescii* grown in XyBW liquid media.

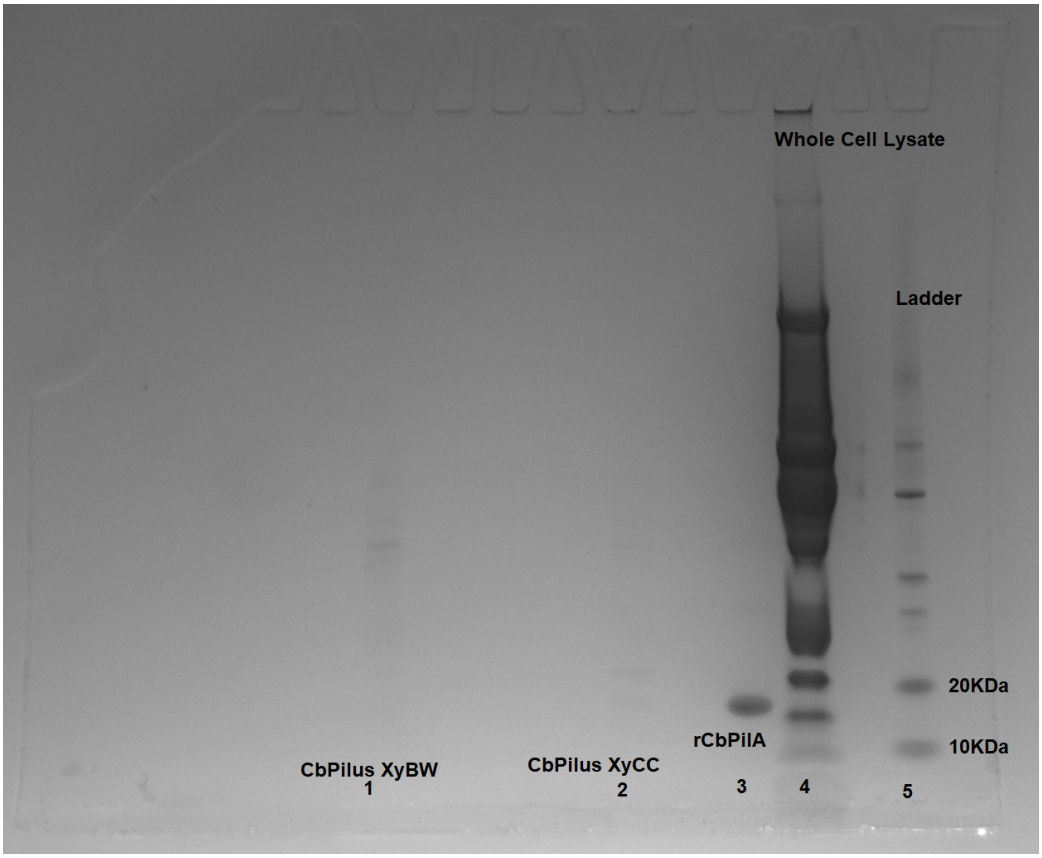


Figure 3 Coomassie and glycoprotein stained SDS-PAGE gel. Lanes: 1) *C. bescii* native pilin fraction from cultures grown on birchwood xylan (XyBW). 2) *C. bescii* native pilin fraction from cultures grown on corn core xylan (XyCC). 3) Purified, recombinant *C. bescii* major pilin. 4) Whole cell lysate from *C. bescii* grown on xylan. 5) Molecular mass marker.

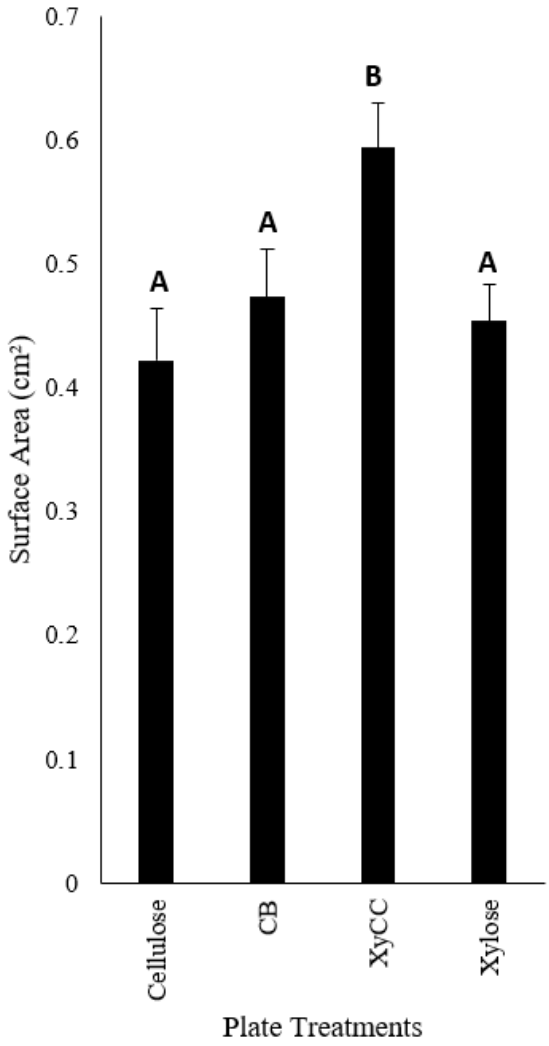


Figure 4 Mean surface area of *C. bescii* colonies on 4 carbohydrates. XyCC treatment demonstrated significantly higher colonization compared to all other treatments. Different letters indicate significant differences determined by one-way ANOVA with post-hoc LSD tests.

DISCUSSION

The phylogenetic tree constructed by PosiGene using protein coding sequences was used as the basis of all positive selection analyses. The tree was compared to the phylogenomic tree constructed using PhyloPhlAn in the Khan et al. (2019) comparative genomics study. Both trees agreed that *C. bescii* was most closely related to *C. kronotskyensis*, which makes sense since they were isolated from Kamchatka, Russia. Both phylogenetic trees agree on the placement of the CBC and CBO LCA. The CBC LCA includes 11 species and CBO LCA includes 5 species. Out of the 134 positively selected protein sequences, there were only 2 indications of protein types that were explicitly relevant to general carbohydrate hydrolysis and motility: a glycoside hydrolase (CEP58_RS02380) and pilus assembly protein (CEP58_RS10715). CEP58_RS02380 appeared to be positively selected in 2 runs and CEP58_RS10715 in one, with a combined frequency of 3 out of 142 total positively selected genes among all 3 PosiGene runs suggesting that carbohydrate interaction proteins are relatively more conserved than others. The glycoside hydrolase protein was identified in *C. bescii* and CBC LCA runs, but the latter had significantly more sites under positive selection (3 vs. 17), which would be expected since the CBC LCA is older. The high degree of positive selection in that protein may have played a role in speciation of CBC LCA to the 11 sequenced *Caldicellulosiruptor* species isolated globally. Though the intermediate species CBO LCA did not demonstrate a statistically significant probability for positive selection, the recurrence may still signify that the gene is evolving continuously. Because *Caldicellulosiruptor* species have multiple mechanisms for plant biomass deconstruction, it is possible that the particular glycoside hydrolase is relatively not as essential.

The T4P operon locus tag range includes locus tags CEP58_RS10690 - 10765 . PosiGene determined that only one related gene, pilus assembly protein PilM, in *C. bescii*

(CEP58_RS10715) was subject to positive selection. PilM has been described as an essential part of the T4P assembly sub-complex characteristic of many gram-positive bacteria that has a conserved functional domain (Pelicic et al., 2019). Since my analysis identified only 1 site under positive selection in that gene, it suggests that the gene is overall relatively conserved. Further genetic analysis would have to be done to determine if the site was in a functional domain. Other relevant T4P genes like the retraction ATPase PilT and major pilin PilA were not found to be positively selected in any runs. The lack of positive selection in all but one T4P-related protein suggests there is little variation among and within species and is essential. This supports previous claims that the T4P locus is highly conserved (Khan et al., 2020).

C. bescii demonstrated twitching motility significantly more on xylan than any other tested carbohydrate. Since *C. bescii* produces more PilA when grown on xylan (Khan et al., 2020), the observed significant preference for xylan is supported. However, when grown on cellulose, the observed surface area was not as large. Since *Caldicellulosiruptor* species have a wide range of extracellular cellulose-binding proteins (Rodionov et al., 2021), I expected there to be greater twitching motility in response to cellulose. The difficulties of working with an insoluble, crystalline sugar like cellulose may also have contributed to the insignificant results. However, *C. bescii* grown on cellobiose, a fully soluble disaccharide, also didn't demonstrate a significant twitching response. The xylose treatment had similar results, suggesting that the sugars present must be in long chains, on top of being soluble. *C. bescii* receptors that initiate T4P expression appear to be xylan-specific. As a result, *C. bescii* may instead rely on a different motility system, flagella, on cellulose-rich surfaces. Blumer-Schuette et al. (2012) observed flagellar peptides adhered to spent cellulose, indicating the use of swimming motility. Swimming motility plates were difficult to optimize because *C. bescii* can only form colonies

between surfaces, while the swimming motility assay requires the microbe to swim in the thin film of water present at the surface of an agar plate. Furthermore, incubating the low-viscosity plates at the high temperatures needed for the species to form colonies resulted in the agar melting, and losing its form.

The differences in *C. bescii* adhesion to and movement on different carbohydrates characteristic of plant biomass demonstrate preferences for specific complex carbohydrates. The degradation of such complex carbohydrates, like insoluble fibers, is relevant to the study of microbial symbiosis with the human digestive system. T4P evolution and twitching motility efficacy by *C. bescii* may additionally contribute to our understanding of biofilm formation by other bacterial species that rely on twitching motility.

APPENDIX

Appendix 1 Species and their respective GenBank accession number used for PosiGene analysis

***Caldicellulosiruptor* species**	**GenBank Accession Number**
C. acetigenus	GCA_000421725.1
C. bescii	GCA_000022325.1
C. changbaiensis	GCA_003999255.1
C. danielii	GCA_000955725.1
C. spp F32	GCA_000404025.1
C. hydrothermalis	GCA_000166355.1
C. kristjanssonii	GCA_000166695.1
C. kronotskyensis	GCA_000166775.1
C. lactoaceticus	GCA_000193435.3
C. morganii	GCA_000955745.1
C. naganoensis	GCA_000955735.1
C. obsidiansis	GCA_000145215.1
C. owensensis	GCA_000166335.1
C. saccharolyticus	GCA_000016545.1

Appendix 2 Example of the use of unix programs, awk and sed to manipulate FASTA files to only print the locus tag in the FASTA header

```
cat GCF_000016545.1_ASM1654v1_cds_from_genomic.fna|awk '/^>/ {if
($2 ~ /locus_tag/) print $2; if ($3 ~ /locus_tag/) print
$3;next}1'|sed 's/[^=>]*=//'|sed '/^>/s/.\{1\}$//'|sed 's/.$//'
> awksaccharolyticus.fa
```

Appendix 3 Slurm script of PosiGene run with *C. bescii* as target species

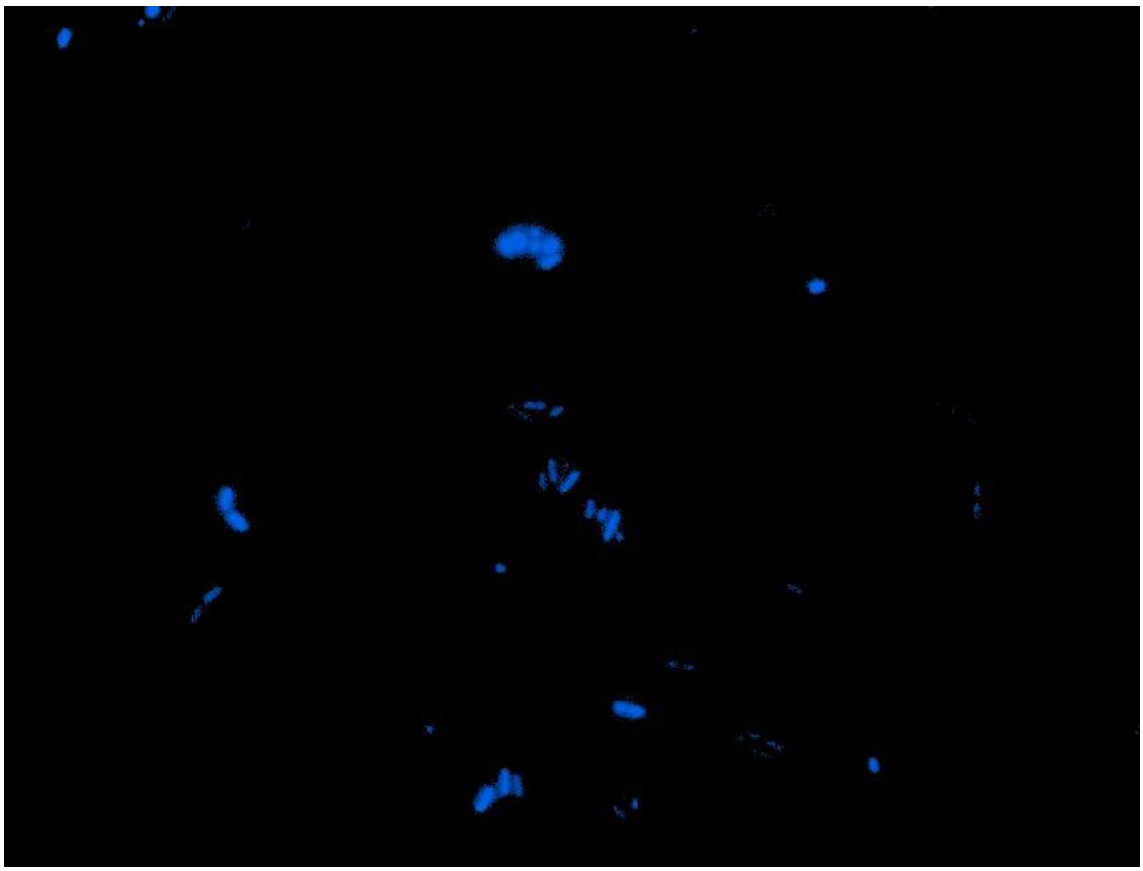
```
#!/bin/bash --login
########## SBATCH Lines for Resource Request ##########

#SBATCH --time=72:00:00            # limit of wall clock time -
how long the job will run (same as -t)
#SBATCH --nodes=1                  # number of different nodes -
could be an exact number or a range of nodes (same as -N)
#SBATCH --ntasks=1                 # number of tasks - how many
tasks (nodes) that you require (same as -n)
#SBATCH --cpus-per-task=16         # number of CPUs (or cores)
per task (same as -c)
#SBATCH --mem-per-cpu=8G           # memory required per
allocated CPU (or core) - amount of memory (in bytes)
#SBATCH --job-name PosiGene1       # you can give your job a
name for easier identification (same as -J) NO SPACES

########## Command Lines to Run ##########
module load Perl/5.32.0
module load PosiGene/0.1
cd /projects/sbs-lab/Caldi_GB
perl /cm/shared/apps/PosiGene/0.1/PosiGene.pl -o=Caldi_Final
-genetic_code=11 -as=bescii -tn=16
-rs=bescii:/projects/sbs-lab/Caldi_GB/_bescii.fa -ts=bescii
-nhsbr=saccharolyticus:/projects/sbs-lab/Caldi_GB/_saccharolytic
us.fa,bescii:/projects/sbs-lab/Caldi_GB/_bescii.fa,changbaiensis
:/projects/sbs-lab/Caldi_GB/_changbaiensis.fa,kronotskyensis:/pr
ojects/sbs-lab/Caldi_GB/_kronotskyensis.fa,kristjanssonii:/proje
cts/sbs-lab/Caldi_GB/_kristjanssonii.fa,hydrothermalis:/projects
/sbs-lab/Caldi_GB/_hydrothermalis.fa,lactoaceticus:/projects/sbs
-lab/Caldi_GB/_lactoaceticus.fa,obsidiansis:/projects/sbs-lab/Ca
ldi_GB/_obsidiansis.fa,owensensis:/projects/sbs-lab/Caldi_GB/_ow
ensensis.fa,naganoensis:/projects/sbs-lab/Caldi_GB/_naganoensis.
fa,acetigenus:/projects/sbs-lab/Caldi_GB/_acetigenus.fa,danielii
:/projects/sbs-lab/Caldi_GB/_danielii.fa,morganii:/projects/sbs-
lab/Caldi_GB/_morganii.fa,F32:/projects/sbs-lab/Caldi_GB/_F32.fa
```

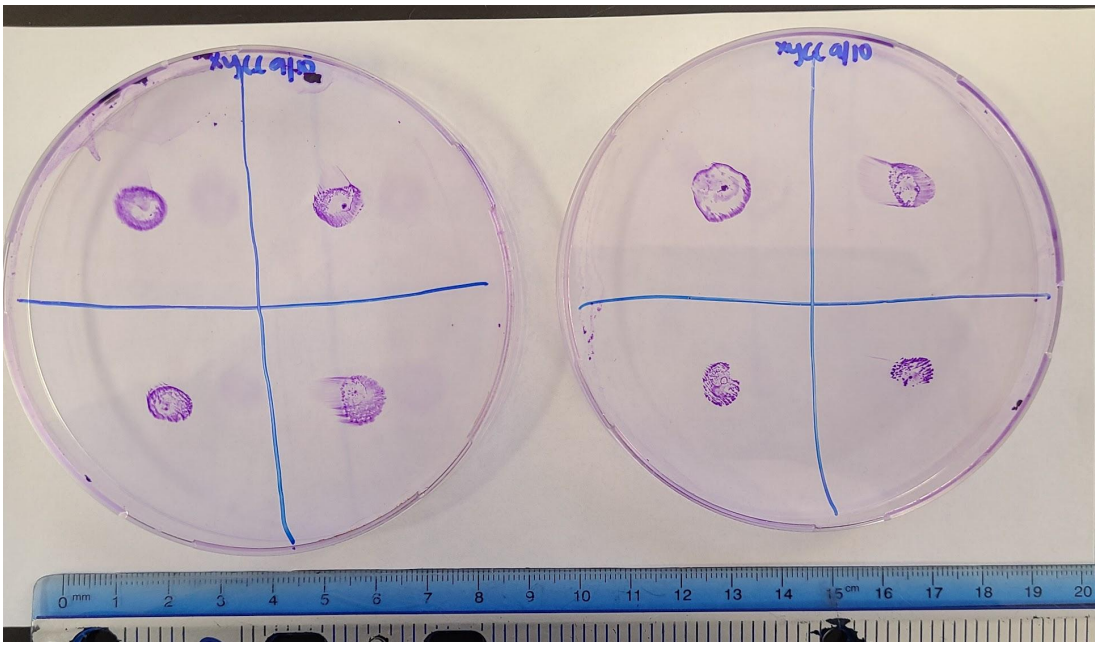
Appendix 5 Image of vacuum manifold used to make anaerobic media

Appendix 4 Example of a fluorescent micrograph of *C. bescii* stained with DAPIused for cell density determination

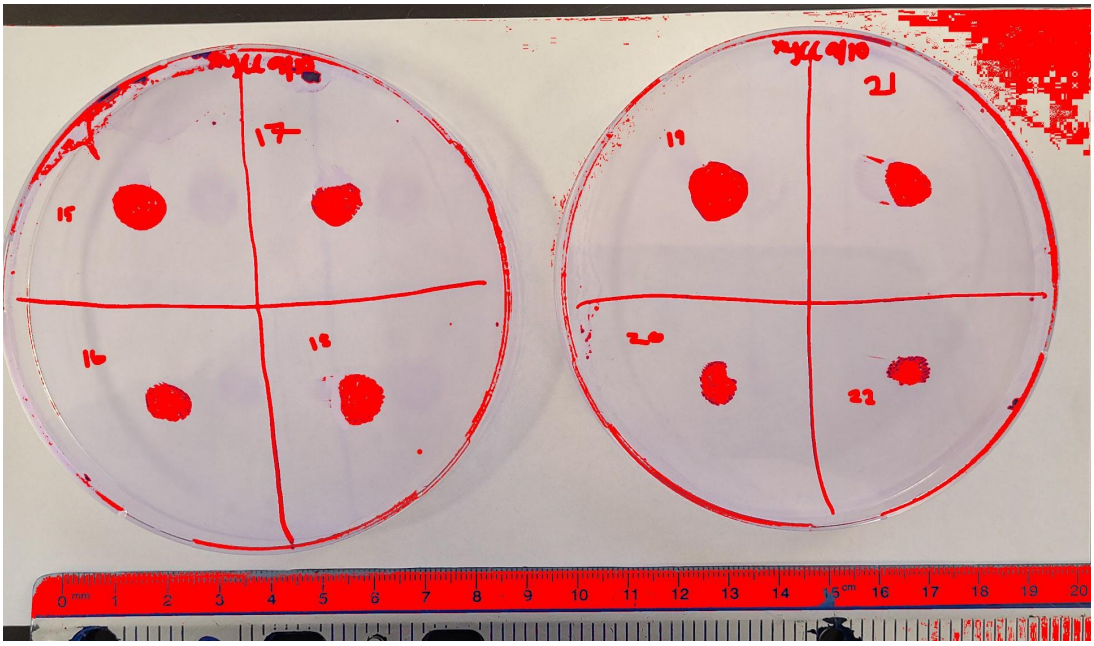


Appendix 5 (a) Image of plates used to set scale measurements. (b) Image adjusted for color threshold used to determine surface area measurements.

(a)



(b)



REFERENCES (APA)

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., & Klopp, C. (2014). Jvenn: An interactive venn diagram viewer. *BMC Bioinformatics*, *15*(1), 293. https://doi.org/10.1186/1471-2105-15-293

Biswas I., Machen A., Mettlach J. (2019) In Vitro Motility Assays for Acinetobacter Species. In: Biswas I., Rather P. (eds) *Acinetobacter baumannii. Methods in Molecular Biology*, vol 1946. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-9118-1_17

Blumer-Schuette, S. E., Alahuhta, M., Conway, J. M., Lee, L. L., Zurawski, J. V., Giannone, R. J., Hettich, R. L., Lunin, V. V., Himmel, M. E., & Kelly, R. M. (2015). Discrete and structurally unique proteins (Tāpirins) mediate attachment of extremely thermophilic caldicellulosiruptor species to cellulose. *Journal of Biological Chemistry*, *290*(17), 10645–10656. https://doi.org/10.1074/jbc.m115.641480

Blumer-Schuette, S. E., Giannone, R. J., Zurawski, J. V., Ozdemir, I., Ma, Q., Yin, Y., Xu, Y., Kataeva, I., Poole, F. L., 2nd, Adams, M. W., Hamilton-Brehm, S. D., Elkins, J. G., Larimer, F. W., Land, M. L., Hauser, L. J., Cottingham, R. W., Hettich, R. L., & Kelly, R. M. (2012). *Caldicellulosiruptor* core and pangenomes reveal determinants for noncellulosomal thermophilic deconstruction of plant biomass. *Journal of Bacteriology*, 194(15), 4015–4028. https://doi.org/10.1128/JB.00266-12

Conway, Jonathan M., et al. (2018)Parsing in Vivo and in Vitro Contributions to Microcrystalline Cellulose Hydrolysis by Multidomain Glycoside Hydrolases in the *Caldicellulosiruptor bescii* Secretome. *Biotechnology and Bioengineering*, 115(10), 2426–2440., doi:10.1002/bit.26773.

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*. *Journal of Industrial Microbiology and Biotechnology*, *40*(1), 41–49. https://doi.org/10.1007/s10295-012-1202-1

G-Biosciences. (2012). *Glycoprotein Staining Kit With Rapidstain™ for Enhanced Glycoprotein & Non Glycoproteins Staining (Cat. # 786-254): Protocol.* USA: Author.

Hobbie, J. E., Daley, R. J., & Jasper, S. (1977). Use of nuclepore filters for counting bacteria by fluorescence microscopy. *Applied and Environmental Microbiology*, *33*(5), 1225–1228. https://doi.org/10.1128/aem.33.5.1225-1228.1977

Khan, A. M., Hauk, V. J., Ibrahim, M., Raffel, T. R., & Blumer-Schuette, S. E. (2020). Caldicellulosiruptor bescii adheres to polysaccharides via a type IV pilin-dependent mechanism. *Applied and Environmental Microbiology*, *86*(9), e00200-20. https://doi.org/10.1128/aem.00200-20

Khan, A. M., Mendoza, C., Hauk, V. J., & Blumer-Schuette, S. E. (2019). Genomic and physiological analyses reveal that extremely thermophilic caldicellulosiruptor changbaiensis deploys uncommon cellulose attachment mechanisms. *Journal of Industrial Microbiology and Biotechnology*, *46*(9-10), 1251–1263. https://doi.org/10.1007/s10295-019-02222-1

Pelicic, V. (2019). Monoderm bacteria: The new frontier for type IV pilus biology. *Molecular Microbiology*, *112*(6), 1674–1683. https://doi.org/10.1111/mmi.14397

Piepenbrink, K. H., & Sundberg, E. J. (2016). Motility and adhesion through type IV pili in gram-positive bacteria. *Biochemical Society Transactions*, *44*(6), 1659–1666. https://doi.org/10.1042/bst20160221

Rodionov, D. A., Rodionova, I. A., Rodionov, V. A., Arzamasov, A. A., Zhang, K., Rubinstein, G. M., Tanwee, T. N., Bing, R. G., Crosby, J. R., Nookaew, I., Basen, M., Brown, S. D., Wilson, C. M., Klingeman, D. M., Poole, F. L., Zhang, Y., Kelly, R. M., & Adams, M. W. (2021). Transcriptional regulation of plant biomass degradation and carbohydrate utilization genes in the extreme thermophile *Caldicellulosiruptor bescii*. *MSystems*, *6*(3), e01345-20. https://doi.org/10.1128/msystems.01345-20

Ronish, L. A., Lillehoj, E., Fields, J. K., Sundberg, E. J., & Piepenbrink, K. H. (2019). The structure of Pila from *Acinetobacter baumannii* AB5075 suggests a mechanism for functional specialization in *Acinetobacter* type IV pili. *Journal of Biological Chemistry*, *294*(1), 218–230. https://doi.org/10.1074/jbc.ra118.005814

Sahm, A., Bens, M., Platzer, M., & Szafranski, K. (2017). Posigene: Automated and easy-to-use pipeline for genome-wide detection of positively selected genes. *Nucleic Acids Research*, *45*(11), e100. https://doi.org/10.1093/nar/gkx179

Yang, S.-J., Kataeva, I., Hamilton-Brehm, S. D., Engle, N. L., Tschaplinski, T. J., Doeppke, C., Davis, M., Westpheling, J., & Adams, M. W. (2009). Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe "*Anaerocellum thermophilum*" DSM 6725. *Applied and Environmental Microbiology*, *75*(14), 4762–4769. https://doi.org/10.1128/aem.00236-09