Extremophiles (2018) 22:203–209
https://doi.org/10.1007/s00792-017-0988-1

ORIGINAL PAPER

Description of a cryptic thermophilic (pro)phage, CBP1 from *Caldibacillus debilis* strain GB1

Scott Wushke[1] · Zimo Jin[1] · Victor Spicer[2] · Xiang Li Zhang[3] · Brian Fristensky[3] · Oleg V. Krokhin[4] · David B. Levin[5] · Richard Sparling[1]

Received: 10 October 2017 / Accepted: 8 December 2017 / Published online: 30 January 2018

Abstract

This study characterizes a cryptic (pro)phage-related sequence within the *Caldibacillus debilis* GB1 genome, designated CBP1.CBP1 is a Siphoviridae-like genome highly related to GBVS1 from *Geobacillus* sp. 6k51. The CBP1genome is a 37,315 bp region containing 69 putative ORFs with a GC content of 42% flanked on both sides by host DNA integrated into the main bacterial chromosome (contig 16). Bioinformatic analyses identified cassettes of genes within the CBP1 genome that were similar in function, yet distinct in sequence, from genes previously identified in GBVS1. All of CBP1 genes had less than 60% amino acid sequence identity with GBVS1by tBLASTx, with the exception of the TMP repeat gene. CBP1 possessed all the necessary genes to undergo a temperate/lytic phage life cycle, including excision, replication, structural genes, DNA packaging, and cell lyses. Proteomic analysis of CBP1 revealed the expression of 5 proteins. One of the expressed proteins was a transcriptional regulator protein homologous to the bacteriophage λ repressor protein (*cI*) expressed in high amounts from the CBP1 region, consistent with a lysogenic phage in a repressed state. The CBP1 protein expression profile during host growth provides unique insight into thermophilic *Siphoviridae*-like phages in the repressed state within their host cells.

Keywords Phage · Proteome · Genome analysis

Introduction

Caldibacillus debilis strain GB1 was isolated from a marsh contaminated by seasonal manure run-off from a cattle barn (Wushke et al. 2013). Physiological comparison of the type strain *C. debilis* DSM 16016 against *C. debilis* GB1 (Wushke et al. 2015) revealed several distinct phenotypic differences, including differences in yields of end-product amounts, differing amounts of cell lysis in stationary, and the ability to grow anaerobically. Genomic and proteomic characterization of *C. debilis* GB1 was conducted previously with a focus on core metabolism (Wushke et al. 2017).

Phage – host interactions are important in affecting microbial physiology, biogeochemical and ecological processes, biofilm formation in the environmental communities, and horizontal gene transfer (Bohannan and Lenski 2000; Ochman et al. 2000; Sutherland et al. 2004; Weitz and Wilhelm 2012). Within the domain Bacteria, thermophilic phages represent a novel area of study, with only a handful isolated from thermophilic Firmicutes, such as bacteriophages GBVS1, D6E, GVE1, GVE2, with only GVE2 being lytic (Doi et al. 2013; Wang and Zhang 2010; Liu et al.

Communicated by H. Atomi.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00792-017-0988-1) contains supplementary material, which is available to authorized users.

✉ Richard Sparling
richard.sparling@ad.umanitoba.ca

[1] Department of Microbiology, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

[2] Department Physics and Astronomy, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

[3] Department Plant Science, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

[4] Department of Internal Medicine and Manitoba Centre for Proteomics and Systems Biology, University of Manitoba, Winnipeg, MB R3E 3P4, Canada

[5] Department of Biosystems Engineering, University of Manitoba, Winnipeg, MB R3T 3V6, Canada

2006). There have been few thermophilic tailed *Siphoviridae* bacteriophages characterized within the literature (Liu et al. 2009; Liu and Zhang 2008; Doi et al. 2013). Thermophilic Firmicutes with thermophilic (pro)phage genomes are rarely reported, described, and characterized in the literature (Sidhu 2000).

Caldibacillus debilis may have uses in biofuel applications (Wushke et al. 2015). Part of evaluating an organism for industrial usefulness is determining its genomic stability. In general, genomic stability is negatively affected by mobile genomic elements like integrative phages and transposons (Ochman et al. 2000). Finding, describing, and understanding potential genome destabilizing elements are a key to evaluate an organism for industrial usefulness.

Bioinformatic analyses identified putative genes encoded by the *Siphoviridae*-like bacteriophage CBP1 within the *C. debilis* genome. We have characterized gene cassette of CBP1, acryptic (pro)phage detected within the *C. debilis* GB1 genome, and demonstrated that some gene products are actively expressed during early exponential phase of cell growth.

Methods

Cell growth

For all experiments, *C. debilis* strain GB1, DSM 29516 (Wushke et al. 2013) was grown on cellobiose in a modified 1191 medium (Islam et al. 2006), with a lower concentration of yeast extract (0.76 g/L) and with the initial pH adjusted to 7.2 and at a temperature of 60 °C. Aerobic environments in Balch tubes were prepared as previously described by Wushke et al. (2015, 2017). Plating was conducted using cellobiose as a substrate on modified 1191 media as previously described by Wushke et al. (2015, 2017). *C. debilis* strain Tf, DSM16016, was grown using the same media and conditions as *C. debilis* GB1 and used to create a lawn for plaque assays.

Plaque assay

Caldibacillus debilis GB1 was grown to stationary phase(24 h post inoculum). All plating was done in triplicate. *C. debilis* GB1 supernatant (0.1 mL) was filtered – sterilized (0.22 µm filter) and mixed with liquid 1% agar m-1191 inoculated with 1% *C. debilis* Tf DSM 16061 and incubated for 10 min at 60 °C to allow phage absorption. Liquid 1% agar m-1191 was then poured into plates and allowed to solidify. These plates were then incubated for 12, 24 and 48 h. In an attempt to induce the lytic cycle, half of the plates were exposed to UV light for 30 s after 12 h of incubation using the UV light in the biosafety cabinet (1300 Series A2 from Thermo Scientific). The UV-induced plates were put back in the incubator and checked every 3 h until 48 h for the presence of plaque formation.

Genomic analysis

The *C. debilis* GB1 genome description was published previously (accession number: AZRV00000000; Wushke et al. 2017). The genome sequence of wild-type *C. debilis* (strain Tf) is available at the NCBI database (accession number: ARVR00000000). PHAST allowed identification of putative integration sites as well as phage-associated open reading frames (ORFs), which typical bacterial annotation pipelines may not interpret correctly (Zhou et al. 2011). Complete cryptic (pro)phage genomes in *C. debilis* GB1 and *C. debilis* Tf were identified using the Phage Search Tool (PHAST) (Zhou et al. 2011).

Proteomic analysis

The proteome of *C. debilis* GB1 under aerobic and anaerobic growth conditions was previously described by Wushke et al. (2017). The methods used to extract, purify, and analyze the proteome are also described by Wushke et al. (2017). The total ion current (TIC) for each protein under aerobic and anaerobic conditions were pooled and the value expressed as Log_2(TIC) as shown in Table 1. Data generated by that study were further analyzed with a focus on bacteriophage-associated genes. These methods and analysis were applied to the CBP1 (pro)phage genome (accession MF595878).

Results

Identification of phage genomes in *C. debilis* GB1

Analysis of the *C. debilis* GB1 genome (Wushke et al. 2017) revealed several distinct regions where exogenous DNA was integrated into the main bacterial chromosome. Due to the significant differences in physiology between *C. debilis* GB1 and *C. debilis* Tf, and the presence of significant amounts of exogenous DNA in the *C. debilis* GB1 genome, an attempt was made to identify and characterize any (pro)phage(s) using the 'omics information previously generated for *C. debilis* GB1.

One potential whole phage genome appeared to be integrated in the *C. debilis* GB1 genome (Table 1). This cryptic prophage was designated CBP1. The chromosomal region containing the cryptic prophage was putatively identified within Contig 16 (bp position 43,270–80,585). Analysis of Contig 16 with the Phage Search Tool (PHAST) identified a cryptic prophage region of ~ 37,315 bp with a GC content of 42% (Fig. 1), distinct from *C. debilis* GB1 which has a

Table 1 Annotation of the CBP1 genome

ORF in CBP1[a]	Locus tag in *C. debilis* GB1 DSM 29516[b]	Top PHAST hit[c,d]	*e* value against top PHAST hit	Functional annotation as given by PHAST	Proteome Log_2 TIC
1	Cdeb_02783	PHAGE_Thermu_OH2_NC_021784: SMF family protein[c]	3e−07	Helix turn helix?	–
2	Not annotated	Hypothetical	0.0	Unknown	–
3	Not annotated	Hypothetical	0.0	Unknown	–
4	Not annotated	Hypothetical	0.0	Unknown	19.42
5	Cdeb_02784	LuxR family transcriptional regulator (*Bacillus infantis* NRRL B-14911)	7e−19	Transcriptional regulator	–
6	Cdeb_02785	Hypothetical	0.0	Unknown	–
7	Cdeb_02786	PHAGE_Bacill_SPBc2_NC_001884: hypothetical protein SPBc2p066	2e−19	Unknown	11.38
8	Not annotated	Hypothetical	0.0	Unknown	–
9	Cdeb_02787	PHAGE_Bacill_IEBH_NC_011167: helix-turn-helix domain protein	2e−15	Transcriptional regulator	10.36
10	Cdeb_02788	PHAGE_Lister_B054_NC_009813: gp42	2e−06	Transcriptional regulator	–
11	Cdeb_02789	PHAGE_Bacill_PM1_NC_020883: hypothetical protein	4e−70	Prophage antirepressor	–
12	Not annotated	Hypothetical	0.0	Unknown	–
13	Not annotated	Hypothetical	0.0	Unknown	–
14	Cdeb_02790	PHAGE_Lister_A118_NC_003216: gp43	7e−11	Unknown	–
15	Cdeb_02791	Hypothetical	0.0	Unknown	–
16	Cdeb_02792	Hypothetical	0.0	Unknown	–
17	Cdeb_02793	Hypothetical	0.0	Unknown	–
18	Cdeb_02794	Hypothetical protein LMOSLCC7179_2546 (*Listeria monocytogenes* SLCC7179)	2e−83	Unknown	–
19	Cdeb_02795	PHAGE_Lactob_LF1_NC_019486: hypothetical protein	1e−41	Unknown	–
20	Cdeb_02796	PHAGE_Bacill_IEBH_NC_011167: hypothetical protein IEBH_gp06	2e−06	Unknown	–
21	Cdeb_02797	PHAGE_Cronob_phiES15_ NC_018454: putative replication protein O	8e−14	Phage replication protein O	–
22	Cdeb_02798	hypothetical	0.0	Unknown	–
23	Cdeb_02799	PHAGE_Brocho_NF5_NC_015252: gp45	4e−40	Single-strand binding protein	–
24	Cdeb_02800	Hypothetical	0.0	Unknown	–
25	Cdeb_02801	PHAGE_Paenib_PG1_NC_021558: dUTPase	2e−36	dUTPase	–
26	Cdeb_02802	PHAGE_Pseudo_LUZ24_ NC_010325: hypothetical protein	1e−22	HNH endonuclease	–
27	Cdeb_02803	PHAGE_Geobac_virus_E2_ NC_009552: hypothetical protein GBVE2_gp053	1e−45	Unknown	–
28	Cdeb_02804	Hypothetical protein ACP_0371 [*Acidobacterium capsulatum* ATCC 51196]	9e−22	Unknown	–
29	Cdeb_02805	Hypothetical	0.0	Unknown	–
30	Cdeb_02806	Hypothetical	0.0	Unknown	–
31	Not annotated	Hypothetical	0.0	Unknown	20.85
32	Cdeb_02807	Hypothetical	0.0	Unknown	–

Table 1 (continued)

ORF in CBP1[a]	Locus tag in *C. debilis* GB1 DSM 29516[b]	Top PHAST hit[c,d]	*e* value against top PHAST hit	Functional annotation as given by PHAST	Proteome Log_2 TIC
33	Cdeb_02808	PHAGE_Staphy_JS01_NC_021773: hypothetical protein	5e−12	Unknown	–
34	Cdeb_02809	PHAGE_Thermu_P74_26_NC_009804: metal-dependent hydrolase	1e−17	Unknown	–
35	Not annotated	Hypothetical	0.0	Unknown	–
36	Cdeb_02810	Hypothetical		Unknown	–
37	Cdeb_02811	Hypothetical protein gbs1222 (*Streptococcus agalactiae* NEM316) gi|25011271|ref|NP_735666.1|	3e−06	Unknown	–
38	Not annotated	hypothetical	0.0	Unknown	–
39	Cdeb_02812	PHAGE_Paenib_PG1_NC_021558: hypothetical protein	2e−20	Unknown	–
40	Cdeb_02813	PHAGE_Clostr_phiMMP04_NC_019422: terminase	4e−144	Phage terminase, protein, large subunit	–
41	Cdeb_02814	PHAGE_Clostr_phiMMP04_NC_019422: phage portal protein	1e−144	Phage portal protein	–
42	Cdeb_02815	PHAGE_Clostr_phiSM101_NC_008265: putative endopeptidase	1e−63	Clp protease	–
43	Cdeb_02816	PHAGE_Geobac_GBSV1_NC_008376: phage major capsid protein	7e−11	Phage major capsid protein	–
44	Not annotated	PHAGE_Staphy_52A_NC_007062: ORF045	6e−06	Unknown	–
45	Cdeb_02817	PHAGE_Paenib_PG1_NC_021558: DNA packaging protein	3e−18	DNA packaging protein	–
46	**Cdeb_02818**	**PHAGE_Geobac_GBSV1_NC_008376: hypothetical protein GPGV1_gp23**	**4e−12**	**Unknown**	**–**
47	**Cdeb_02819**	**PHAGE_Geobac_GBSV1_NC_008376: hypothetical protein GPGV1_gp24**	**9e−23**	**Tail-component**	**–**
48	**Cdeb_02820**	**PHAGE_Geobac_GBSV1_NC_008376: aminopeptidase**	**1e−21**	**Aminopeptidase**	**–**
49	**Cdeb_02821**	**PHAGE_Geobac_GBSV1_NC_008376: major tail protein**	**7e−41**	**Phage major tail protein**	**–**
50	**Cdeb_02822**	**PHAGE_Geobac_GBSV1_NC_008376: hypothetical protein GPGV1_gp27**	**4e−36**	**Unknown**	**–**
51	**Not annotated**	**PHAGE_Geobac_GBSV1_NC_008376: hypothetical protein GPGV1_gp28; PP_00112; phage(**	**2e−09**	**Unknown**	**–**
52	**Cdeb_02823**	**PHAGE_Geobac_GBSV1_NC_008376: TMP repeat protein**	**1e−115**	**Unknown**	**–**
53	**Cdeb_02824**	**PHAGE_Geobac_GBSV1_NC_008376: TMP repeat protein**	**3e−37**	**TMP repeat**	**–**
54	**Cdeb_02825**	**PHAGE_Geobac_GBSV1_NC_008376: TMP repeat protein**	**2e−20**	**TMP repeat**	**–**
55	**Cdeb_02826**	**PHAGE_Geobac_GBSV1_NC_008376: hypothetical protein GPGV1_gp30**	**3e−07**	**Phage tail component**	**–**
56	Cdeb_02827	PHAGE_Lactoc_KSY1_NC_009817: gp061	2e−13	Endopolygalacturonase	–

Table 1 (continued)

ORF in CBP1[a]	Locus tag in *C. debilis* GB1 DSM 29516[b]	Top PHAST hit[c,d]	*e* value against top PHAST hit	Functional annotation as given by PHAST	Proteome Log_2 TIC
57	**Cdeb_02828**	**PHAGE_Geobac_GBSV1_NC_008376: hypothetical protein GPGV1_gp32**	**2e−08**	**Phage minor structural protein**	**–**
58	Cdeb_02829	Hypothetical protein BAS0458 [*Bacillus anthracis* str. Sterne]	3e−09	Unknown	–
59	Cdeb_02830	Hypothetical	0.0	Unknown	–
60	Cdeb_02831	PHAGE_Bacill_SPBc2_NC_001884: possibly involved in bacteriocin production or immunity	3e−17	Unknown	–
61	Cdeb_02832	PHAGE_Bacill_phBC6A51_NC_004820: holin	9e−12	Phage holin	–
62	**Cdeb_02833**	**PHAGE_Geobac_GBSV1_NC_008376: *N*-acetylmuramoyl-L-alanine amidase**	**2e−72**	***N*-acetylmuramoyl-L-alanine amidase**	**–**
63	Cdeb_02834	Hypothetical	0.0	PIN domain?	14.4
64	Cdeb_02835	Hypothetical	0.0	Unknown	–
65	Cdeb_02836	Hypothetical	0.0	Unknown	–
66	Cdeb_02837	PHAGE_Bacill_SPBc2_NC_001884: IMPB/MUCB/SAMB family protein	2e−86	Nucleotidyltransferase/DNA polymerase involved in DNA repair	–
67	Cdeb_02838	PHAGE_Bacill_SPBc2_NC_001884: IMPB/MUCB/SAMB family protein	2e−31	Nucleotidyltransferase/DNA polymerase involved in DNA repair	–
68	Not annotated	Hypothetical	0.0	Unknown	–
69	Cdeb_02839	PHAGE_Bacill_phBC6A51_NC_004820: site-specific recombinase	6e−125	Site-specific recombinase, DNA invertase Pin-like	–

[a]CBP1 ORFs can be found in the prophage genome (accession number: MF595878)

[b]Locus tags for *C. debilis* GB1 accession AZRV00000000, not annotated corresponds to genes that were not called using the NCBI bacterial annotation pipeline but were called by PHAST analysis

[c]Rows in gray match the viral and prophage database at the protein level, rows with a white background matched the bacterial or genebank database at the protein level

[d]Bolded rows most closely match GBVS1

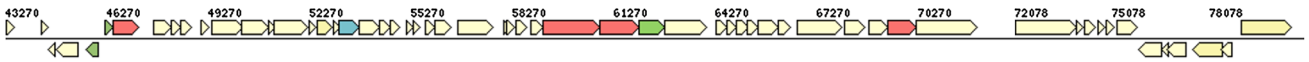

Fig. 1 A schematic representation of the cryptic prophage CBP1 genome on Contig 16 (host DNA omitted) produced by IMG. The directions of the arrows indicate the putative direction of transcription (white: unknown function, red: COG X, blue: COG F, light tan: COG L, green: COG O)

GC content ~ of 51%. The CBP1 genome was submitted to NCBI under accession number MF595878.

Plaques

Caldibacillus debilis DSM 16016 was used as a lawn in an attempt to isolate plaques as genome analysis revealed no phage/prophage genomes were present. No zones of clearing were observed when plaque assays were done using GB1 supernatant from late stationary. Plates appeared to grow as robustly as when no GB1 supernatant was added. When GB1 was used to create a lawn, UV was used in an attempt to induce the phage lytic cycle; treatment with UV did not result in plaques forming or lyses of the lawn.

Analysis of the CBP1 genome

Many of the putative ORFs encoded by the CBP1 genome had the highest nucleotide sequence identity to another thermophilic *Siphoviridae* genome, GBVS1, found in

Geobacillus sp. 6k51 (Liu et al. 2009). A comparison of the GBSV1 and CBP1 genomes revealed regions with greater than 30% nucleotide average sequence identity per 100 bp (Figure S1). A total of 69 putative ORFs (Table 1) were identified in the CBP1 genome using PHAST, 62 of which were transcribed in the same direction, and 8 in the opposite direction. Twenty-seven (27) of the ORFs most closely matched the bacterial or genebank database and 42 ORFs most closely matched the viral and prophage database at the protein level. Twenty-three (23) of the ORF's protein sequences showed extreme variability compared to the databases with *e* values of 0. Forty (40) of the putative ORFs top tblastn hits were to other phages with Gram positive bacteria (Table 1), with many matching ORFs in GBVS1. Moreover, ORFs 45–55 were functionally syntenic between CBP1 and GBVS1 upon inspection of PHAST results.

Thermophilic *Siphoviridae* phages are presumed to survive and replicate via both lytic and temperate cycles (Liu and Zhang 2008). CBP1 appears to have all the necessary functions encoded for a lytic/temperate life cycle including gene homologues for DNA replication (ORF 21, 23), DNA recombinase/invertase (ORF 66, 67, 69), lytic genes (ORF 41, 61,62), phage capsid structural genes (ORF 40, 43, 47, 49, 55, 57), and packaging (ORF 45). The general functions as annotated by PHAST are shown in Figure S2. Several of the PHAST hits in CBP1 were similar to not only other Bacilli-related phages, but also to *Clostridium*-related phages. Three *Clostridium*-related phage genes were identified in the genome of phage CBP1: a terminase, phage portal protein, and clp protease (Table 1; ORFS 40, 41, 42). Several of the closest related phage proteins were found to be from mesophilic hosts per PHAST analysis in Table 1.

Proteomic expression of phage genes

Of the 69 putative ORFs in CBP1, 5 proteins corresponding to these ORFs were detected within the proteome (under both anaerobic and aerobic growth conditions). The genes were not equally expressed and displayed different Log_2TIC scores, shown in Table 1. Notably, ORF 4 and 31 had much higher Log_2TIC scores (19.42 and 20.85, respectively) than ORFs 7, 9, and 63 (11.38, 10.36, 14.4, respectively). The high Log_2TIC scores of ORFs 4 and 31 were similar in expression value (Log_2TIC) to those proteins expressed in central metabolism of *C. debilis* GB1 during growth (Wushke et al. 2017). Three of the expressed proteins (ORFs 4, 31, 63) had e-values of 0, suggesting these proteins are unique and in areas of possible hypervariability of the CBP1 genome. Expression of a transcriptional regulator (CBP1 ORF 9), a protein homologous to the bacteriophage λ repressor protein (cI), was expressed during cell growth. This is consistent with a prophage that would be repressed at the time of sampling (Ackers et al. 1982). The other 4 expressed proteins were hypothetical proteins with undetermined function as characterized by PHAST. To further characterize detected phage-associated proteins InterProScan analysis was used, shown in Table S1 (Zdobnov and Apweiler 2001). InterProScan analysis did not provide significant further insight compared to PHAST analysis. Detection of five phage-associated proteins shows that the phage genes are active in some capacity. Proteomic analyses of purified phages from *Geobacillus* typically only identify the phage structural proteins (Liu and Zhang 2008; Liu et al. 2009), whereas we have observed protein expression from the prophage during host growth (Wushke et al. 2017).

Discussion

The genus *Caldibacillus* (formerly within the genus *Geobacillus*) is a single species sister genus to *Geobacillus* (Coorevits et al. 2012). *Caldibacillus* and *Geobacillus* are both thermophilic Bacilli (Banat et al. 2004; Coorevits et al. 2012). *Geobacillus* have been looked at explicitly for industrial uses. Thus, identifying and characterizing potential genome destabilizing elements in a bacterial strain are important. Bacteriophages in the Family Siphoviridae are known to infect a broad range of Firmicutes including *Clostridium* hosts (Horgan et al. 2010), and several genes encoded by the bacteriophage CBP1 and GBSV1 genome showed their highest sequence identity to genes encoded by phages isolated from Clostridia (Hargreaves et al. 2013; Yoon and Hyo 2011). This fits with the fact that *C. debilis* and *C. thermocellum* have been found within the same environment (Wushke et al. 2013). From the PHAST analysis, three *Clostridium*-related phage genes were identified in the CBP1 genome, matching phages phiMMP04 and phiSM101 (Hargreaves et al. 2013; Nariya et al. 2011). These genes could assist in the life cycle of thermophilic *Siphoviridae* in a *Clostridium* host. This would support the notion that thermophilic Firmicutes act as a natural reservoir of thermophilic *Siphoviridae* and that they have multiple host targets (Lucchini et al. 1998; Brüssow et al. 2001). The thermophilic *Geobacillus* and *Caldibacillus* have been isolated from mesophilic environments (Banat et al. 2004; Wushke et al. 2013); it is possible this phage could be transferred between mesophilic hosts as well. Bacteriophages of this type may represent a vector for genetic transfer between these organisms. Nothing was found (or found to be omitted) at the genomic level, by our analyses, that would preclude the ability of CBP1 to form viable phage particles under the right conditions. The expression of 5 proteins in the CBP1 genome, with one expressed (ORF 9) protein appearing similar to the λ repressor protein (cI), suggests that CBP1 could be a phage in a repressed state. The function of the other 4 proteins observed is unknown, but their expression

may also be associated specifically with the repressed state of the CBP1 phage.

Acknowledgements This work was funded by Genome Canada, through the Applied Genomics Research in Bioproducts or Crops (ABC) program for the grant titled, "Microbial Genomics for Biofuels and Co-Products from Biorefining Processes" as well as a Natural Science and Engineering Research Council (NSERC) Discovery grant to RS. SW was also supported through a University of Manitoba Graduate Fellowship (UMGF) award.

References

Ackers GK, Johnson AD, Shea MA (1982) Quantitative model for gene regulation by lambda phage repressor. Proc Nat Acad Sci 79:1129–1133

Banat I, Marchant R, Rahman T (2004) *Geobacillus debilis* sp. nov., a novel obligately thermophilic bacterium isolated from a cool soil environment, and reassignment of *Bacillus pallidus* to *Geobacillus pallidus* comb. nov. Int J Syst Evol Microbiol 54:2197–2201. https://doi.org/10.1099/ijs.0.63231-0

Bohannan BJ, Lenski RE (2000) Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. Ecol Lett 3:362–377

Brüssow H, Desiere F (2001) Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages. Mol Microbiol 39:213–223. https://doi.org/10.1046/j.1365-2958.2001.02228.x

Coorevits A, Dinsdale AE, Halket G, Lebbe L, De Vos P, Van Landschoot A, Logan NA (2012) Taxonomic revision of the genus Geobacillus: emendation of Geobacillus, *G. stearothermophilus*, *G. jurassicus*, *G. toebii*, *G. thermodenitrificans* and *G. thermoglucosidans* (nom. corrig., formerly 'thermoglucosidasius'); transfer of *Bacillus thermantarcticus* to the genus as *G. thermantarcticus* comb. nov.; proposal of *Caldibacillus debilis* gen. nov., comb. nov.; transfer of *G. tepidamans* to Anoxybacillus as *A. tepidamans* comb. nov.; and proposal of *Anoxybacillus caldiproteolyticus* sp. nov. Int J Syst Evol Microbiol 62:1470–1485. https://doi.org/10.1099/ijs.0.030346-0

Doi K, Mori K, Martono H, Nagayoshi Y, Fujino Y, Tashiro K, Ohshima T (2013) Draft genome sequence of *Geobacillus kaustophilus* GBlys, a lysogenic strain with bacteriophage ϕOH2. Gen Ann 1:e00634-13. https://doi.org/10.1128/genomea.00634-13

Hargreaves KR, Colvin HV, Patel KV, Clokie JJP, Clokie MR (2013) Genetically diverse *Clostridium difficile* strains harboring abundant prophages in an estuarine environment. Appl Environ Microbiol 79:6236–6243

Horgan M, O'Sullivan O, Coffey A, Fitzgerald GF, van Sinderen D, McAuliffe O, Ross RP (2010) Genome analysis of the *Clostridium difficile* phage ΦCD6356, a temperate phage of the Siphoviridae family. Gene 462:34–43. https://doi.org/10.1016/j.gene.2010.04.010

Islam R, Cicek N, Sparling R, Levin D (2006) Effect of substrate loading on hydrogen production during anaerobic fermentation by clostridium thermocellum 27405. Appl Microbiol Biotechnol 72:576–583. https://doi.org/10.1007/s00253-006-0316-7

Liu B, Zhang X (2008) Deep-sea thermophilic *Geobacillus* bacteriophage GVE2 transcriptional profile and proteomic characterization of virions. Appl Microbiol Biotechnol 80:697–707. https://doi.org/10.1007/s00253-008-1575-2

Liu B, Suijie W, Qing S, Xiaobo Z, Lianhui X (2006) Two novel bacteriophages of thermophilic bacteria isolated from deep-sea hydrothermal fields. Curr Microbiol 53:163–166. https://doi.org/10.1007/s00284-005-0509-9

Liu B, Zhou F, Wu S, Xu Y, Zhang X (2009) Genomic and proteomic characterization of a thermophilic *Geobacillus* bacteriophage GBSV1. Res Microbiol 160:166–171. https://doi.org/10.1016/j.resmic.2008.12.005

Lucchini S, Desiere F, Brüssow H (1998) The structural gene module in *Streptococcus thermophilus* bacteriophage φSfi11 shows a hierarchy of relatedness to Siphoviridae from a wide range of bacterial hosts. Virology 246:63–73. https://doi.org/10.1006/viro.1998.9190

Nariya H, Miyata S, Tamai E, Sekiya H, Maki J, Okabe A (2011) Identification and characterization of a putative endolysin encoded by episomal phage phiSM101 of *Clostridium perfringens*. Appl Microbiol Biotechnol 90:1973–1979

Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304

Sidhu SS (2000) Phage display in pharmaceutical biotechnology. Curr Opin Biotechnol 11:610–616. https://doi.org/10.1016/S0958-1669(00)00152-X

Sutherland IW, Hughes KA, Skillman LC, Tait K (2004) The interaction of phage and biofilms. FEMS Microbiol Lett 232:1–6. https://doi.org/10.1016/S0378-1097(04)00041-2

Wang Y, Zhang X (2010) Genome analysis of deep-sea thermophilic phage D6E. Appl Environ Microbiol 76:7861–7866

Weitz JS, Wilhelm SW (2012) Ocean viruses and their effects on microbial communities and biogeochemical cycles. Biology Rep 4:17

Wushke S, Levin DB, Cicek N, Sparling R (2013) Characterization of enriched aerotolerant cellulose-degrading communities for biofuels production using differing selection pressures and inoculum sources. Can J Microbiol 59:679–683. https://doi.org/10.1139/cjm-2013-0430

Wushke S, Levin DB, Cicek N, Sparling R (2015) Characterization of the facultative anaerobe *Caldibacillus debilis* GB1 and its use in a designed aerotolerant, cellulose degrading, co-culture with *Clostridium thermocellum*. Appl Environ Microbiol. https://doi.org/10.1128/AEM.00735-15

Wushke S, Spicer V, Zhang X, Fristensky B, Krokhin OV, Levin DB, Cicek N, Sparling R (2017) Understanding aerobic/anaerobic metabolism in *Caldibacillus debilis* through a comparison with model organisms. Syst Appl Microbiol 40(5):245–253. https://doi.org/10.1016/j.syapm.2017.03.004

Yoon BH, Hyo IC (2011) Complete genomic sequence of the Lactobacillus temperate phage LF1. Arch Virol 156(10):1909

Zdobnov EM, Apweiler R (2001) InterProScan—an integration platform for the signature recognition methods in InterPro. Bioinformatics 17:847–848

Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS (2011) PHAST: a fast phage search tool. Nucleic Acids Res. https://doi.org/10.1093/nar/gkr485