Short report

Genome sequence of the *Bacteroides fragilis* phage ATCC 51477-B1

Shawn A Hawkins*†1,3, Alice C Layton†2,3, Steven Ripp2,3, Dan Williams3 and Gary S Sayler2,3

Address: 1Department of Biosystems Engineering and Soil Science, The University of Tennessee, Knoxville, TN, USA, 2Department of Microbiology, The University of Tennessee, Knoxville, TN, USA and 3The Center for Environmental Biotechnology, The University of Tennessee, Knoxville, TN, USA

Email: Shawn A Hawkins* - shawkins@utk.edu; Alice C Layton - alayton@utk.edu; Steven Ripp - saripp@utk.edu; Dan Williams - dwilli34@utk.edu; Gary S Sayler - sayler@utk.edu

* Corresponding author   †Equal contributors

Abstract

The genome of a fecal pollution indicator phage, *Bacteroides fragilis* ATCC 51477-B1, was sequenced and consisted of 44,929 bases with a G+C content of 38.7%. Forty-six putative open reading frames were identified and genes were organized into functional clusters for host specificity, lysis, replication and regulation, and packaging and structural proteins.

Findings

Bacteriophages infecting *Escherichia coli* and *Bacteroides fragilis* serve as fecal pollution indicators (1). *Bacteroides* phages are more attractive fecal indicators than *E. coli* because *Bacteroides* are more abundant in fecal matter, provide higher host specificity, and are anaerobic and thus less likely to reproduce in aquatic environments than *E. coli* [1-4]. Phage ATCC 700786-B1, infecting *B. fragilis* RYC2056, serves as an ISO standard reference phage, but the host is also susceptible to phages in other animal feces [5,6]. Phage ATCC 51477-B1 infects the host *B. fragilis* HSP40, which is reported to be only susceptible to phages in human feces and surface water polluted with municipal/septic wastewater [7].

A drawback to using *Bacteroides* phages as water quality indicators results from the requirement to cultivate them on an anaerobic host [6]. This can be circumvented with direct phage detection via PCR [8], but assay design is difficult because only one gene for one *Bacteroides* bacteriophage is publically available. The lack of ability to detect *Bacteroides* bacterial phage is startlingly inadequate because the host genera is dominant in the human gut and contains important antibiotic resistant pathogens [3,9]. In this study, the genome of phage ATCC 51477-B1 was sequenced to promote fecal source tracking assay development and aid construction of a *B. fragilis* phage bioreporter [10].

Phage ATCC 51477-B1 was propagated on *B. fragilis* HSP40 (ATCC 51477) at 37°C in anaerobic *Bacteroides* phage recovery medium (BPRM) amended with kanamycin, nalidixic acid, bile salts, and Oxyrase (Oxyrase; Mansfield, OH) [11,12]. Phage were purified and concentrated from lysate using polyethylene glycol [13] and displayed a non-elongated icosohedral head and non-contractile tail consistent with the family Siphoviridae and other *B. fragilis* phages isolated from municipal wastewater (Figure 1) [14]. Structural dimensions were similar to *B. fragilis* phage B40-8 with respect to head diameter (60 ± 4.0 μm) and tail length (162 ± 17.0 μm), but the tail diameter was greater (13.4 ± 1.1 μm versus 9.3 ± 0.4 μm) [13].
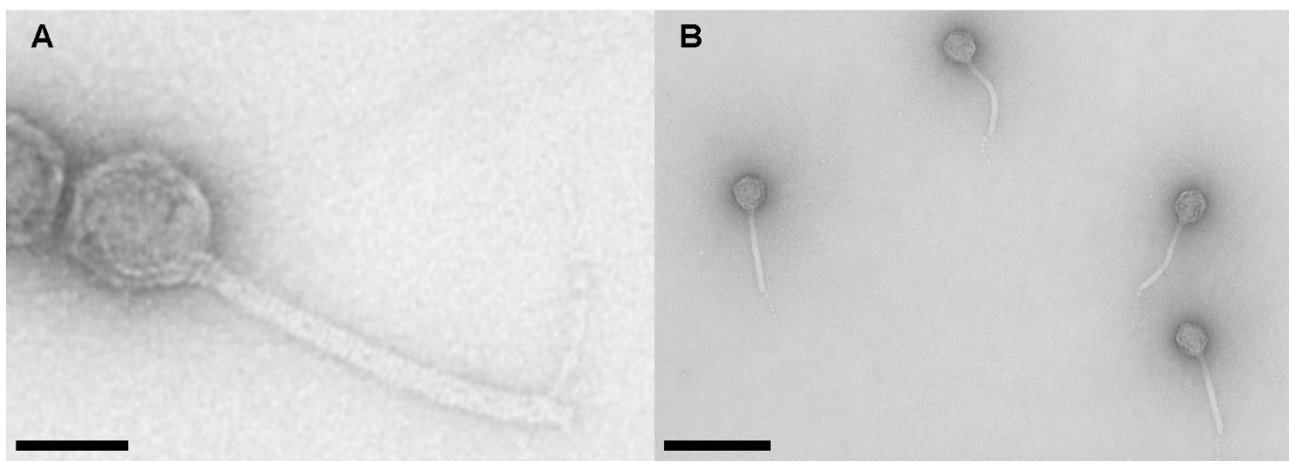
Figure 1
Transmission electron micrographs of phage ATCC 51477. (A) Magnified view of a single phage (bar = 50 nm). (B) Four additional phage all displaying tail fibers (bar = 200 nm).

Phage DNA was extracted with Lambda minipreps (Promega, Madison, WI) and digested with *Hin*dIII. Several *Hin*dIII restriction fragments were cloned into pUC19 [15] and sequenced using primer walking. PCR reactions bridging the cloned *Hin*dIII restriction fragments were cloned and sequenced to confirm fragment order. Non-redundant areas of the genome that were difficult to clone were directly sequenced. In total, dideoxynucleotide sequencing provided 2× coverage except at the 5' and 3' ends. Confirmation of sequence data was sought by GS FLX pyrosequencing (MWG Biotech, Inc., High Point, NC) which produced 23,263 reads, 58,730 base calls, and assembly of 62 contigs using the GS *De Novo* Assembler. Thirty eight of the contigs with more than 3 reads were co-assembled with the dideoxynucleotide sequences using DNASTAR Lasergene 7.1 software suite (DNASTAR, Inc., Madison, WI). A total of 21,935 pyrosequencing reads in 27 contigs were maintained in the final assembly which aligned 91% similarity for 100× coverage of the phage genome, including the 5' and 3' ends. The final assembled genome was 44,929 bp with a 38.7% G+C content (Figure 2).

Open reading frames (ORFs) were identified using Gene-Mark [16,17] and the NCBI ORF Finder and examined for known protein functions, structures, and motifs using a conserved domain database [18] (Additional file 1, Figure 2). Ten of the 46 putative ORFs contained amino acid sequences with predictable functions or motifs (Additional file 1). The 5' ends of three ORFs (ORF39, ORF40, and ORF43) displayed translated similarity to previously published N-terminal amino acid sequences for *B. fragilis* phage B40-8 head (MP1 and MP3) and tail (MP2) proteins at 100, 80, and 90% similarity, respectively [13]. The

B40-8 MP2 gene (the only *B. fragilis* phage gene present in GenBank), was present in ATCC 51477-B1 but contained a 119 bp insertion and was split into three misarranged sections, one of which was inverted (Figure 2). This suggested the putative ATCC 51477-B1 tail protein was chimeric with respect to B40-8.

The ATCC 51477-B1 genome contained gene functional clusters for host specificity (tail fiber), lysis, replication and regulation, and packaging and structural proteins (Figure 2) [19]. The host specificity region contained ORF7 with translated similarity to the tail fiber of *Enterococcus* phage phiEF24C (Additional file 1) and a large size (1,897 amino acids) suggesting it may be a tape measure gene [20]. A putative M-15 type peptidase (ORF10) was the only gene clearly linked with the lysis region. The replication and regulation region included phage anti-repressors (ORF22 and ORF26), DNA modifying enzymes (ORF15 and ORF16), and replication proteins (ORF22, ORF31, and ORF32). Within the packaging and structural cluster, an ORF with similarity to the TerL protein was identified (ORF 38), as well as ORFs with N-terminal sequences similar to the three phage B40-8 structural proteins previously mentioned (ORF39, ORF40, and ORF43) [13].

Four potential promoters were identified on the positive strand of the genome based on similarity to promoters found in <u>*B. fragilis*</u> [21]. Two were in the regulation and replication region, 5' of ORFs 17 and 22. The other two potential promoters were near the beginning of the genome, 5' of ORF2, and in the lysis region, 5' of ORF12. A tandem repeats finder [22] revealed a 23 bp repeat within an 83 bp segment having two perfect and two
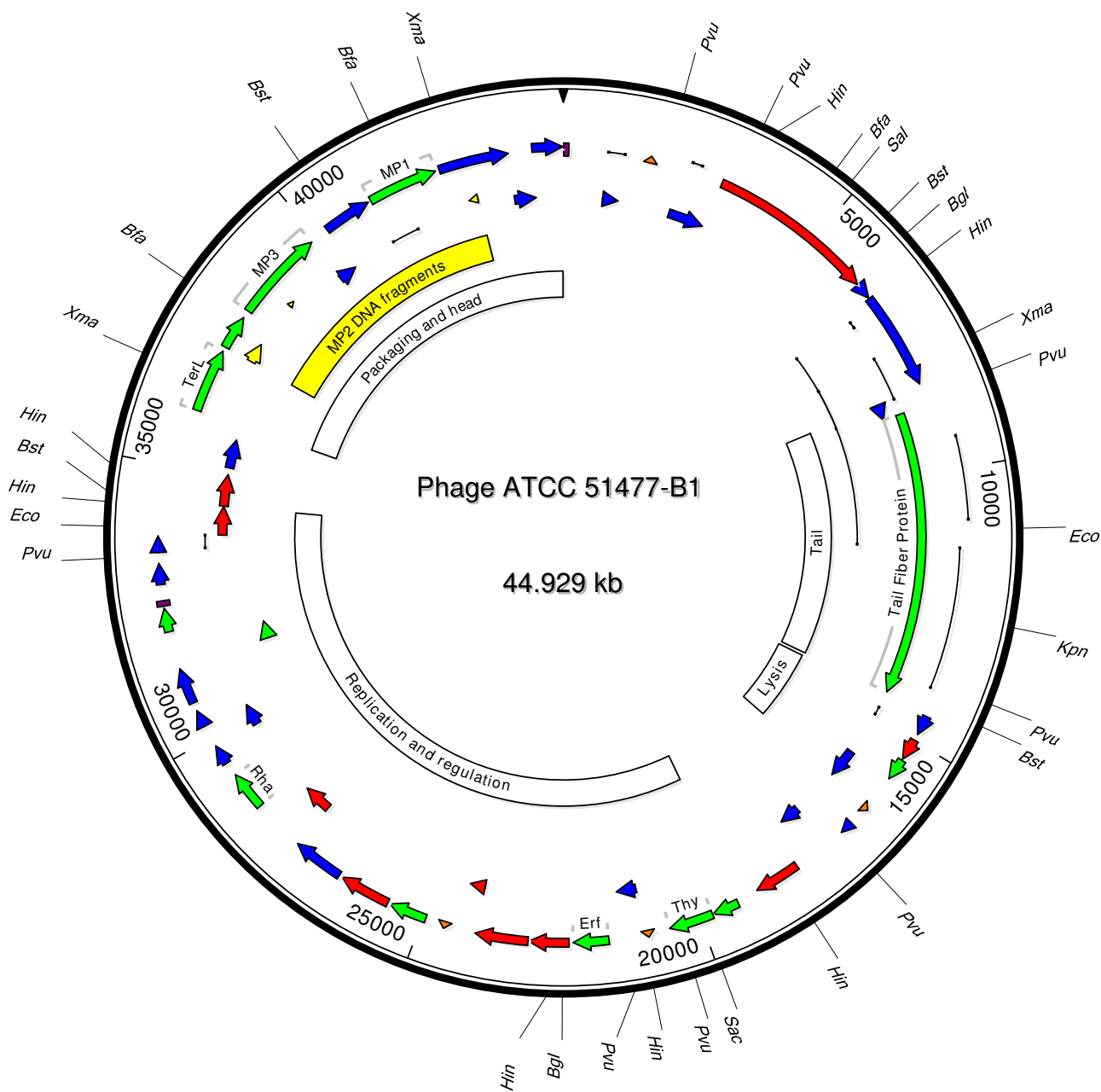
Figure 2
Phage ATCC 51477-B1 genome map. The direction and size of 46 putative ORFs are illustrated with arrows. ORFs labeled in green display high similarity to proteins with known function. ORFs labeled in red display high similarity to other putative open reading frames without assigned functions. ORFs labeled in blue have low similarity (< 0.1) to other putative open reading frames. Sequences with high DNA similarity to the phage B40-8 structural proteins are shown in yellow. Putative *B. fragilis* promoters are shown as inverted orange triangles, the location of repeats are shown as purple boxes, and polymorphisms are indicated with black lines.

degenerate repeats in the replication and regulation region. Another 96 bp perfect repeat sequence was identified at the beginning of the phage genome and again at the end of the replication and regulation region (Figure 2).

Eight polymorphic regions occurring within ORFs and displaying sequence variability from 4% to 13% were identified with the extensive pyrosequencing data. These regions ranged in size from 200 bp to greater than 1,000 bp. ORF7, the putative tail fiber protein, displayed the most polymorphisms, including a pyrosequencing contig with a deletion relative to the final assembly. Tail fiber variability modifies the phage host range [23] and may, for this phage, reflect the antigenic variability of *B. fragilis* surface components [21,24].

The sequence for *B. fragilis* phage ATCC 51477-B1 was deposited in GenBank with accession number FJ008913.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
SAH cloned and sequenced the genome by primer walking, assisted in the pyrosequence data analysis and annotation, and drafted the manuscript. ACL analyzed pyrosequencing data, annotated the genome, and assisted in drafting the manuscript. SR organized the study and provided final editing for the manuscript. DW provided genome cloning. GSS participated in the study design and coordination. All authors read and approved the final manuscript.

Additional material

Additional file 1
Phage ATCC 51477-B1 genome organization.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1743-422X-5-97-S1.pdf]

References
1. Bernhard AE, Field KG: **A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA.** *Applied and Environmental Microbiology* 2000, **66:**4571-4574.
2. Dick LK, Bernhard AE, Brodeur TJ, Domingo JWS, Simpson JM, Walters SP, Field KG: **Host distributions of uncultivated fecal *Bacteroidales* bacteria reveal genetic markers for fecal source identification.** *Applied and Environmental Microbiology* 2005, **71:**3184-3191.
3. Matsuki T, Watanabe K, Fujimoto J, Takada T, Tanaka R: **Use of 16S rRNA gene-targeted group-specific primers for real-time PCR analysis of predominant bacteria in human feces.** *Applied and Environmental Microbiology* 2004, **70:**7220-7228.
4. Tartera C, Jofre J: **Bacteriophages active against *Bacteroides fragilis* in sewage-polluted waters.** *Appl Environ Microbiol* 1987, **53(7):**1632-1637.
5. Puig A, Jofre J, Araujo R: **Bacteriophages infecting various *Bacteroides fragilis* strains differ in their capacity to distinguish human from animal faecal pollution.** *Water Science and Technology* 1997, **35:**359-362.
6. ISO: **International Standard 10705-4:2001(E). Water quality – Detection and enumeration of bacteriophages – Part 4: Enumeration of bacteriophages infecting *Bacteroides fragilis*.** International Organization for Standardization; 2001:12-15.
7. Tartera C, Jofre J, Lucena F: **Relationship between numbers of enteroviruses and bacteriophages infecting *Bacteroides fragilis* in different environmental samples.** *Environmental Technology Letters* 1988, **9:**407-410.
8. Puig M, Pina S, Lucena F, Jofre J, Girones R: **Description of a DNA amplification procedure for the detection of bacteriophages of *Bacteroides fragilis* HSP40 in environmental samples.** *Journal of Virological Methods* 2000, **89:**159-166.
9. Champoux JJ, Drew WL, Neidhardt FC, Plorde JJ, (Eds): **Sherris Medical Microbiology.** New York: The McGraw-Hill Companies, Inc; 2004.
10. Ripp S, Jegier P, Birmele M, Johnson CM, Daumer KA, Garland JL, Sayler GS: **Linking bacteriophage infection to quorum sensing signalling and bioluminescent bioreporter monitoring for direct detection of bacterial agents.** *Journal of Applied Microbiology* 2006, **100:**488-499.
11. Tartera C, Araujo R, Michel T, Jofre J: **Culture and decontamination methods affecting enumeration of phages infecting *Bacteroides fragilis* in sewage.** *Appl Environ Microbiol* 1992, **58(8):**2670-2673.
12. Araujo R, Muniesa M, Mendez J, Puig A, Queralt N, Lucena F, Jofre J: **Optimisation and standardisation of a method for detecting and enumerating bacteriophages infecting *Bacteroides fragilis*.** *Journal of Virological Methods* 2001, **93:**127-136.
13. Puig M, Girones R: **Genomic structure of phage B40-8 of *Bacteroides fragilis*.** *Microbiology* 1999, **145:**1661-1670.
14. Queralt N, Jofre J, Araujo R, Muniesa M: **Homogeneity of the morphological groups of bacteriophages infecting *Bacteroides fragilis* strain HSP40 and strain RYC2056.** *Current Microbiology* 2003, **46:**163-168.
15. Sambrook J, Russell DW: *Molecular Cloning: A Laboratory Manual* Cold Spring Harbor, NY: Cold Spring Harbor Press; 2001.
16. Besemer J, Borodovsky M: **GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses.** *Nucleic Acids Research* 2005, **33:**W451-W454.
17. Besemer J, Borodovsky M: **Heuristic approach to deriving models for gene finding.** *Nucleic Acids Research* 1999, **27:**3911-3920.
18. Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWweese-Scott C, Geer LY, Gwadz M, He SQ, Hurwitz DI, Jackson JD, Ke ZX, *et al.*: **CDD: a conserved domain database for protein classification.** *Nucleic Acids Research* 2005, **33:**D192-D196.
19. Ploeg JR van der: **Genome sequence of *Streptococcus mutans* bacteriophage M102.** *FEMS Microbiology Letters* 2007, **275:**130-138.
20. Piuri M, Hatfull GF: **A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells.** *Molecular Microbiology* 2006, **62:**1569-1585.
21. Weinacht KG, Roche H, Krinos CM, Coyne MJ, Parkhill J, Comstock LE: **Tyrosine site-specific recombinases mediate DNA inversions affecting the expression of outer surface proteins of *Bacteroides fragilis*.** *Molecular Microbiology* 2004, **53:**1319-1330.
22. Benson G: **Tandem repeats finder: a program to analyze DNA sequences.** *Nucleic Acids Research* 1999, **27:**573-580.
23. Inoue Y, Matsuura T, Ohara T, Azegami K: **Bacteriophage OP$_1$, lytic for *Xanthomonas oryzae* pv. *oryzae*, changes its host range by duplication and deletion of the small domain in the deduced tail fiber gene.** *Journal of General Plant Pathology* 2006, **72:**111-118.

24. Cerdeno-Tarraga AM, Patrick S, Crossman LC, Blakely G, Abratt V, Lennard N, Poxton I, Duerden B, Harris B, Quail MA, *et al.*: **Extensive DNA inversions in the *B. fragilis* genome control variable gene expression.** *Science* 2005, **307:**1463-1465.