# Genomic Characterization of Mycobacteriophage Giles: Evidence for Phage Acquisition of Host DNA by Illegitimate Recombination<sup>∀</sup>†

Peter Morris,<sup>‡</sup> Laura J. Marinelli, Deborah Jacobs-Sera, Roger W. Hendrix, and Graham F. Hatfull<sup>\*</sup> Department of Biological Sciences and Pittsburgh Bacteriophage Institute, University of Pittsburgh, Pittsburgh, Pennsylvania 15241

Received 12 October 2007/Accepted 22 December 2007

A characteristic feature of bacteriophage genomes is that they are architecturally mosaic, with each individual genome representing a unique assemblage of individual exchangeable modules. Plausible mechanisms for generating mosaicism include homologous recombination at shared boundary sequences of module junctions, illegitimate recombination in a non-sequence-directed process, and site-specific recombination. Analysis of the novel mycobacteriophage Giles genome not only extends our current perspective on bacteriophage genetic diversity, with more than 60% of the genes unrelated to other mycobacteriophages, but offers novel insights into how mosaic genomes are created. In one example, the integration/excision cassette is atypically situated within the structural gene operon and could have moved there either by illegitimate recombination or more plausibly via integrase-mediated site-specific recombination. In a second example, a DNA segment has been recently acquired from the host bacterial chromosome by illegitimate recombination, providing further evidence that phage genomic mosaicism is generated by nontargeted recombination processes.

The total estimated number of 10<sup>31</sup> bacteriophage particles in the biosphere suggests that phages represent a majority of biological entities (14). The complete genome sequences of approximately 400 double-stranded DNA tailed phages have been determined, which show, together with viral metagenomic studies (1, 5, 8), that the phage population has very high genetic diversity and is replete with genetic information that either is very distant from other known sequences or has a distinct evolutionary origin. Furthermore, the completely sequenced phage genomes have characteristic mosaic architectures, so that modules, composed of either single genes or several genes, are shared between individual genomes (6, 15, 18, 28); each phage genome can thus be considered a unique assemblage of individual modules. While the number of different modules remains ill-defined, the combinatorial possibilities of these modules far exceeds the estimated total number of phage particles in the biosphere (12).

The evolution of mosaic genomes requires the creation of novel junctions at module boundaries. These module junctions often correspond closely with gene boundaries, and two distinct models have been proposed to explain this mosaicism. The first was proposed by Susskind and Botstein (34) and supposes that there are short conserved sequences at gene boundaries that act as recombination targets. Such boundary sequences have been described in some *Escherichia coli* phages (6), but these appear to be exceptional rather than typical examples. The second model proposes that module junctions are a result of illegitimate recombination events that occur without DNA sequence identity (beyond a few nucleotides in length), accompanied by selection for gene function and genome length of a packageable size (15). While such events are likely to be rare, the large population size, a long evolutionary history, and active phage growth (estimated at a global rate of  $10^{24}$  infections/second) (16) provide ample opportunity for such low-frequency events to play key evolutionary roles. Moreover, this is a creative process generating novel genetic junctions that are expected to remain stable within the population for long periods (i.e., as molecular fossils) but that can be further moved around by homologous recombination between shared gene sequences.

The absence of shared sequences at most presumed module junctions suggests that the boundary sequence model is unlikely to be generally applicable for creating new module junctions. However, there is also only rather poor direct evidence in support of the illegitimate-recombination model, since there are few examples of recombination events that have occurred sufficiently recently in evolutionary time for us to be able to interpret where the recombination events took place. One plausible example of such an event has been described among mycobacteriophage genomes (28), where a short DNA sequence of near identity is shared by two otherwise dissimilar genomes and the recombinant junctions can be inferred. However, the illegitimate-recombination model also suggests that genetic exchange would occur between the host and phage genomes (since extensive homology is not required) and that the comparatively large size of bacterial chromosomes would make these relatively frequent events. While "bacterial" genes are indeed commonly found in phage genomes, there are few examples that we are aware of where the mechanism of acquisition has been described, apart from the classic incorporation of gal or bio genes in the generation of lambda specialized transducing phages (4).

In this paper, we describe a novel mycobacteriophage (Giles), a distant relative of other phages that infect the same *Mycobacterium smegmatis* host, and its comparative genome

<sup>\*</sup> Corresponding author. Mailing address: Department of Biological Sciences and Pittsburgh Bacteriophage Institute, University of Pittsburgh, Pittsburgh, PA 15260. Phone: (412) 624 4350. Fax: (412) 624-4870. E-mail: gfh@pitt.edu.

<sup>‡</sup> Present address: UG2, School of Biological Sciences, University of East Anglia, Norwich, United Kingdom NR4 7TJ.

<sup>†</sup> Supplemental material for this article may be found at http://jb.asm.org/.

<sup>&</sup>lt;sup>v</sup> Published ahead of print on 4 January 2008.

analysis. While more than 30 mycobacteriophage genomes have been sequenced (12, 31), making them, together with groups of phages that infect dairy bacteria (3), staphylococcus (21), and pseudomonas (22), one of the largest groups of bacteriophages that infect a common host, mycobacteriophage Giles has a unique genomic architecture and is highly mosaic, and more than 50% of its putative genes are novel. Among the many interesting features of the Giles genome, two aspects are particularly illuminating in the process of phage evolution. First, Giles carries an active integration cassette that is located in an atypical position within the structural gene operon, suggesting that site-specific recombination events can contribute to noncanonical genome structures. Secondly, Giles carries a short DNA segment at the right end of the genome that is 100% identical to the host metE gene, providing direct evidence for the role of illegitimate recombination between phage and host genomes in generating mosaic genomic architectures.

#### MATERIALS AND METHODS

**Phage isolation and purification.** Mycobacteriophage Giles was identified by direct plating on lawns of *M. smegmatis* using an extract of soil and debris recovered from under a log pile near Pittsburgh, PA, with phage buffer (10 mM Tris-HCI [pH 7.5], 10 mM MgSO<sub>4</sub>, 1 mM CaCl<sub>2</sub>, 68.5 mM NaCl). The extract was filtered through a 0.22- $\mu$ m filter, and 50  $\mu$ l of this sample was plated with 1 ml of late-log-phase *M. smegmatis* mc<sup>2</sup>155 and 4.5 ml of 7H9 agar (Middlebrook 7H9 broth base; Difco Laboratories, Detroit, MI) supplemented with 1 mM CaCl<sub>2</sub>. Following several rounds of plaque purification, a high-titer stock was prepared and used for subsequent studies.

Sequence determination. To prepare a library of Giles, genomic DNA was sheared using HydroShear (Gene Machines, Inc.) and repaired, and 1- to 3-kbp DNA fragments were purified by gel extraction. The DNA fragments were cloned into the EcoRV site in the pBluescript vector and recovered in *Escherichia coli* XL1 Blue cells. DNAs from approximately 480 individual clones were prepared and sequenced from both ends of each insert using forward and reverse primers. The sequences of these clones were assembled into a single contig, and ambiguous regions were resolved by sequencing directly from Giles DNA with oligonucleotide primers.

**Construction of Giles integration-proficient vectors.** Giles integration-proficient plasmid vectors, pGH1000A and pGH1000B, were constructed by PCR amplification of a 1.7-kbp fragment of the Giles genome using primers 5'-TGA CGATCAACTCCGCGGGGCCGGGGCCA and 5'-GGAATGATGATGATCGCCG CGGTGACACAATCGGCG and cloning this fragment into the DraI site of plasmid pMosBlueHyg, a derivative of pMosBlue in which a hygromycin resistance cassette had been inserted. Electroporations of *M. smegmatis* mc<sup>2</sup>155 with pGH1000A and pGH1000B were performed with approximately 50 ng DNA.

Identification of Giles virion proteins. CsCl-banded Giles virions were dialyzed into phage buffer and subjected to three rounds of freezing on dry ice and thawing on wet ice. This sample was mixed with sodium dodecyl sulfate (SDS) sample buffer and heated to 95°C for 3 minutes. Proteins were separated by SDS-polyacrylamide gel electrophoresis (PAGE) on a 10% gel and visualized by staining them with colloidal Coomasie (Invitrogen). Bands were excised and sequenced using NanoLC-MS/MS peptide-sequencing technology (ProTech, Inc.). Briefly, protein bands were in-gel digested with sequencing grade modified trypsin (Promega). The peptide mixtures were then analyzed by a liquid chromatography-tandem mass spectrometry system (Thermo), consisting of highpressure liquid chromatography with a 75-µm-inner-diameter reverse-phase C18 column on-line coupled to an ion trap mass spectrometer. The spectrometric data were searched against a database consisting of all predicted Giles open reading frames, and the output was manually verified. When not otherwise indicated, all other chemicals used in proteolytic digestion and high-pressure liquid chromatography were obtained from Sigma. N-terminal sequencing confirmed the identities of the phage-encoded proteins gp10 and gp15.

Nucleotide sequence accession number. The GenBank accession number of the Giles genome is EU203571.



FIG. 1. Morphology of mycobacteriophage Giles virions. Shown is an electron micrograph of mycobacteriophage Giles particles negatively stained with uranyl acetate. Scale bar = 100 nm.

### RESULTS

**Isolation and genomic sequencing of mycobacteriophage Giles.** Mycobacteriophage Giles was isolated from a location near Pittsburgh, PA, by direct plating of a soil extract onto a lawn of *M. smegmatis*. Following plaque purification, a hightiter stock was prepared, and Giles virions were examined by electron microscopy. Giles particles have an isometric head approximately 55 nm in diameter and a long flexible tail approximately 205 nm long (Fig. 1). This siphoviridal morphology is the most common among the characterized mycobacteriophages.

DNA was extracted from Giles particles and examined by restriction digestion and agarose gel electrophoresis; the patterns obtained were clearly distinct from all of the previously sequenced mycobacteriophage genomes. The Giles genome was sequenced using a shotgun approach as described previously (28). The final complete Giles genome is 54,512 bp in length and has a G+C content of 67.3%. The ends of the virion DNA have 14-bp 3' single-stranded extensions (3'-ATTCGG CGCGCCAT), which are longer than of any of the 31 previously sequenced mycobacteriophage genomes. Both BlastN and Dotter analysis showed that Giles has no extensive nucleotide sequence similarity to other characterized phages (data not shown).

Genome organization of mycobacteriophage Giles. Annotation of the Giles genome revealed 79 open reading frames and no tRNA or other small RNA genes (Fig. 2 and Table 1). Remarkably, only 31 of these putative open reading frames (39%) have identifiable sequence similarity to other mycobacteriophage genes at the amino acid sequence level (Table 1), although at least two that do not match other mycobacteriophage genes have other database matches. There is an easily

# Giles



FIG. 2. Organization of the mycobacteriophage Giles genome. The Giles genome is represented by horizontal lines, with putative genes shown as boxes above (transcribed rightward) or below (transcribed leftward); the number of each gene is shown within its box. All genes have been sorted into phamilies (Phams) of related sequences using the computer program "Phamerator" (S. Cresawn, R. W. Hendrix, and G. F. Hatfull, unpublished data); the phamily number is displayed above each gene, and the boxes are color coordinated accordingly. Note that the Pham numbers differ from those described previously (12). Putative gene functions are noted. The positions of putative transcription promoters as identified by DNA Master (http://cobamide2.bio.pitt.edu) are shown as arrows.

identifiable gene encoding a tyrosine integrase (gp29) located near the center of the genome and transcribed leftward, and the adjacent gene on its right, which is transcribed rightward, encodes a putative excise protein. As with other integrasecontaining phage genomes, we will refer to the segment to the left of the integration cassette as the "left arm" and the genes to the right as the "right arm" (Fig. 2). Giles is unusual among the mycobacteriophages in that virion structural genes are present in both the left and right arms (see below).

The majority of Giles genes are transcribed in the rightward direction, and only 16 are transcribed leftward (Fig. 2). These include three genes at the extreme left end of the left arm (2 to 4), the integrase gene itself (29), the rightmost gene (79), and a cluster of genes (38 to 48) in the middle of the right arm. Overall, there are few noncoding spaces, with the notable exceptions of  $\sim$ 300 bp to the left of the integrase gene (which, as described below, contains the *attP* site),  $\sim$ 300 bp to the right of the excise gene (30),  $\sim$ 400 bp between genes 61 and 62, and 520 bp between the last gene (79) and the right end of the genome (Fig. 2). While accurate promoter prediction remains challenging (11), a promoter subset with sequence similarity to the  $\sigma^{70}$  promoters of *E. coli* are recognizable and have been shown to be important in other mycobacteriophages (27). We have identified four such putative promoters in the Giles genome, two that are linked to leftward-transcribed genes at the left end of the genome and two at the extreme right end, one facing leftward and the other rightward (Fig. 2). All three of the leftward-facing promoters are of interest in that they suggest that they, along with their accompanying open reading

frames, are parts of morons (18) and have been acquired relatively recently.

Giles virion structure and assembly genes. One of the most striking features of the Giles left arm is that of the 28 putative genes, only 12 are clearly related to those of other mycobacteriophages (Fig. 2). These include genes encoding a large terminase subunit (gp5), a prohead protease (gp7), the major capsid subunit (gp9), the tail tape measure protein (gp20), and five minor tail proteins (gp21, gp22, gp24, gp25, and gp27). From their genomic positions, we also predict that genes 4, 6, and 8 encode a putative small terminase subunit, a portal protein, and a capsid assembly protein, respectively. Curiously, no sequence relationship between the putative Giles capsid subunit (gp9) and any other mycobacteriophage capsid proteins could be detected, although after several rounds of PSI-BLAST analysis, matches to capsid proteins of several streptococcal and staphylococcal phages were detected. This assignment was confirmed by analysis of Giles virion proteins (Fig. 3) showing that gp9 is one of the most abundant structural components; we note, however, that the gp9 subunits are not covalently cross-linked, as is common among other mycobacteriophages (9, 10, 13, 25). Among the 31 previously sequenced mycobacteriophage genomes, putative capsid subunits can be identified in 24, and they fall into four distinct sequence families (12). The Giles capsid gene thus further expands the diversity of this group of proteins.

The putative tape measure protein of Giles (gp20) is encoded by the longest open reading frame in the genome and contains 1,360 residues. Since the size of the tape measure

| Gene     | F/R <sup>a</sup> | Start  | Stop  | MW           | Comments                      |
|----------|------------------|--|-------|--------------|-------------------------------|
| 1        | F                | 14   | 571   | 19.9         |                               |
| 2        | R                | 630  | 800   | 6.4          |                               |
| 3        | R                | 930  | 1553  | 22.3         |                               |
| 4        | R                | 1617   | 1826  | 7.8          |                               |
| 5        | F                | 1855   | 3489  | 60.1<br>51.5 | Terminase<br>Detential portal |
| 0        | Г<br>Г           | 3300   | 4915  | 51.5<br>41.0 | Potential portal              |
| 8        | г<br>F           | 4908<br>6100   | 6552  | 41.9         | Potential virion protein      |
| 9        | F                | 6586   | 7542  | 34.5         | Capsid subunit                |
| 10       | F                | 7539   | 7979  | 15.3         | Virion protein                |
| 11       | F                | 7983   | 8435  | 16.2         | · F                           |
| 12       | F                | 8435   | 9118  | 24.3         | Virion protein                |
| 13       | F                | 9121   | 9528  | 14.6         | *                             |
| 14       | F                | 9521   | 9922  | 14.5         |                               |
| 15       | F                | 9938   | 10477 | 20.0         | Major tail subunit            |
| 16       | F                | 10511  | 10915 | 13.9         | Potential virion protein      |
| 17       | F                | 10893  | 11789 | 33.3         |                               |
| 18       | F                | 11827  | 12069 | 9.5          |                               |
| 20       | Г<br>Г           | 12071  | 12229 | 141.2        | Tanamaasura protain           |
| 20       | F                | 16315  | 17409 | 41.0         | Minor tail protein            |
| 22       | F                | 17406  | 19148 | 64.8         | Minor tail protein            |
| 23       | F                | 19148  | 19753 | 22.3         | white the protein             |
| 24       | F                | 19750  | 20853 | 39.3         | Minor tail protein            |
| 25       | F                | 20853  | 23117 | 76.4         | Minor tail protein            |
| 26       | F                | 23120  | 23365 | 8.9          | *                             |
| 27       | F                | 23371  | 24609 | 39.4         | Minor tail protein            |
| 28       | F                | 24633  | 24989 | 11.9         | -                             |
| 29       | R                | 25285  | 26478 | 43.7         | Integrase                     |
| 30       | F                | 26505  | 26822 | 11.3         | Excise                        |
| 31       | F                | 27131  | 28339 | 43.4         | Lysin A                       |
| 32       | F                | 28339  | 29574 | 43.9         | Lysin B                       |
| 22       | F                | 29590  | 30000 | 14.4         |                               |
| 34       | г<br>F           | 30320  | 30330 | 23.1         |                               |
| 36       | F                | 30952  | 31740 | 27.8         | Virion protein                |
| 37       | F                | 31742  | 31894 | 5.9          | villon protein                |
| 38       | R                | 31891  | 32046 | 6.0          |                               |
| 39       | R                | 32112  | 32336 | 8.4          |                               |
| 40       | R                | 32333  | 32707 | 13.1         |                               |
| 41       | R                | 32891  | 33316 | 15.8         |                               |
| 42       | R                | 33425  | 34126 | 25.4         | Similar to Mmcs_5465          |
| 43       | R                | 34195  | 34422 | 7.8          |                               |
| 44       | R                | 34419  | 34724 | 10.7         |                               |
| 45       | R                | 34/30  | 35041 | 11.0         |                               |
| 40       | D                | 26188  | 36604 | 54.9<br>15.2 |                               |
| 47<br>48 | R                | 36712  | 37137 | 15.5         |                               |
| 49       | F                | 37143  | 37457 | 11.3         |                               |
| 50       | F                | 37454  | 37819 | 13.2         |                               |
| 51       | F                | 37891  | 38754 | 31.9         |                               |
| 52       | F                | 38780  | 39871 | 40.2         | RecE homologue                |
| 53       | F                | 39868  | 41316 | 52.1         | RecT-like protein             |
| 54       | F                | 41313  | 41453 | 4.8          |                               |
| 55       | F                | 41450  | 41767 | 10.9         |                               |
| 56       | F                | 41760  | 42338 | 20.1         |                               |
| 58       | F<br>E           | 42333  | 42907 | 21.0<br>11.0 |                               |
| 50<br>50 | r<br>F           | 4∠904<br><u></u> <u></u> | 43631 | 11.0         |                               |
| 60       | F                | 43624  | 43962 | 17.4         |                               |
| 61       | F                | 43959  | 44513 | 20.6         |                               |
| 62       | F                | 44916  | 45578 | 23.7         | Putative DNA methylase        |
| 63       | F                | 45667  | 46746 | 39.4         |                               |
| 64       | F                | 46743  | 47240 | 18.2         |                               |
| 65       | F                | 47237  | 47650 | 15.5         |                               |
| 66       | F                | 47647  | 48102 | 16.6         |                               |
| 67       | F                | 48164  | 48856 | 25.6         |                               |
| 68       | F                | 48849  | 49157 | 11.7         | WhiB-like protein             |
| 69<br>70 | F                | 49185  | 49526 | 12.2         |                               |
| 70       | F                | 49523  | 49936 | 15.4         |                               |
| /1 72    | F<br>E           | 49933  | 50672 | 14.0         |                               |
| 12<br>73 | r<br>E           | 50660  | 51157 | 12./         |                               |
| 73       | г<br>F           | 51154  | 51567 | 16.0         |                               |
| 75       | F                | 51765  | 52598 | 30.1         |                               |
| 76       | F                | 52603  | 52992 | 14.8         |                               |
| 77       | F                | 52989  | 53147 | 5.9          |                               |
| 78       | F                | 53248  | 53613 | 13.5         |                               |
| 79       | R                | 53686  | 53991 | 11.6         | MetE segment                  |
|          |                  |  |       |              |                               |

<sup>a</sup> F, forward; R, reverse.



FIG. 3. Virion proteins of mycobacteriophage Giles. Purified Giles particles were denatured, and the proteins were separated by SDS-PAGE. Markers (M) are designated by their masses in kDa. Putative assignments of bands to Giles gene products were determined by mass spectrometry and N-terminal sequence analysis of individual bands.

protein typically correlates closely with the length of the phage tail, with each amino acid corresponding to 1.5 Å of tail length (19, 28), we would predict a tail of 204 nm, in excellent correlation with the observed morphology (Fig. 1). No protein of the predicted mass of gp20 (141 kDa) is seen among Giles virion proteins (Fig. 3), although a band approximately 70 kDa in mass contains peptides spanning a substantial part of the protein (Fig. 3), and the simplest explanation is that gp20 is proteolytically processed near its center. However, we cannot rule out the potential for more complex internal processing, such as that observed with the Listeria monocytogenes phage PSA (38), or nonspecific protein degradation. The Giles tape measure protein is also of interest because it contains two of the three short sequence motifs (motifs 1 and 3 at positions 980 to 1082 and 852 to 968, respectively) that have been identified in other mycobacteriophages and that are related to families of small bacterial proteins (28). One of these two motifs (motif 1) is related to the Rpf family proteins that have been shown to function as peptidoglycan hydrolases and in growth resuscitation of dormant bacteria (26, 35) and is also present in the tape measure protein of mycobacteriophage Barnyard. The second motif (motif 3) is also related to peptidoglycan hydrolases and is present in several other mycobacteriophages, including TM4, in which it has been demonstrated to be required for efficient infection of stationary-phase mycobacterial cells (32). Giles is the first mycobacteriophage that contains both motifs 1 and 3.

Twelve Giles virion proteins (gp6, gp9, gp10, gp12, gp15, gp20, gp21, gp22, gp24, gp25, gp27, and gp36) were identified by mass spectrometry following SDS-PAGE separation (Fig. 3); some peptides from both gp8 and gp16 were identified in



FIG. 4. Predicted pseudoknot at the end of the gene 16 mRNA. Bioinformatic analysis suggests that there is a programmed +1 translational frameshift near the end of gene 16 that would shift ribosomes translating gene 16 into the gene 17 reading frame (see the text). We found the illustrated potential pseudoknot and Shine-Dalgarno-like ("S-D") sequence in the immediate vicinity of the end of gene 16.

the gp15 and gp10 samples, suggesting that they are also virion proteins. The second most abundant protein after the putative capsid subunit gp9 appears to be gp15, which therefore is a good candidate for the major tail subunit, and PSI-BLAST analysis revealed that it is a distant relative of the putative major tail subunit of mycobacteriophage Che9c gp14. In all other genomes of noncontractile tailed mycobacteriophages, the genes between the major tail subunit and the tape measure protein include a pair of overlapping open reading frames that are expressed via a programmed translational frameshift (12); this is one of the most highly conserved features of doublestranded DNA tailed phages (37), and it results in the production of assembly chaperones for the tail. In Giles, there are four genes (16, 17, 18, and 19) between the major tail subunit and the tape measure protein gene rather than the more usual two genes, making it unclear a priori which two genes might be participating in a frameshift. However, gp17 shows weak but significant sequence similarity to genes occupying the second position of the frameshifting pair in other mycobacteriophages (Bxb1 gene 21, Bxz2 gene 25, and Llij gene 13), so we tentatively identified Giles genes 16 and 17 as the gene pair that participates in programmed translational frameshifting in the phage. The two open reading frames overlap and are related in such a way that a +1 frameshift would be required for the ribosome to shift between them. We did not find any clear candidates for a "slippery sequence" in the overlap region, though we note that such sequences are less apparent than the signals for the more commonly encountered -1 frameshifts. We did find a "Shine-Dalgarno-like" sequence and strong potential for an RNA pseudoknot, both in the immediate vicinity of the end of gene 16 (Fig. 4). Such sequence features have been implicated in potentiating frameshifting in other systems (2), including two cases of +1 frameshifting at the ends of structural genes in the Listeria phage PSA that are associated

with pseudoknots of somewhat different topology from what we see here (38).

Giles nonstructural genes. Giles genes 38 to 79 are organized into three apparent operons, containing genes 38 to 48 (leftward transcribed), genes 49 to 78 (rightward transcribed), and gene 79 (leftward transcribed) (Fig. 2). Only 12 of these 42 genes are recognizable homologues of other mycobacteriophage genes, and putative functions can be assigned to only 5. Of particular note are genes 52 and 53, which appear to encode homologues of mycobacteriophage recombination proteins that are similar to the E. coli Rac prophage proteins RecE and RecT. These are quite rare among mycobacteriophage genomes, and only Che9c and Halo encode similar proteins. Moreover, the degree of sequence similarity is low, and Giles gp52 (RecE) is too distant from Che9c gp60 and Halo gp42 to be identified using a BlastP cutoff score of 0.001. Giles gp53 (RecT) shares 32% amino acid sequence identity with Halo gp43, but neither is sufficiently similar to Che9c gp61 to be recognized by this method. We note that the Che9c RecE and RecT homologues have proven to be useful tools for recombineering of mycobacteria (36), and Giles 52 and 53 may be similarly useful.

Among the other genes in the 38-to-79 segment of the Giles genome, the functions of only two additional genes can be deduced. Giles gene 62 encodes a putative DNA methylase that is likely part of a restriction-modification system, and while a putative restriction gene is not obvious, gene 61 is a possible candidate. Giles gene 68 encodes a WhiB-like regulator protein, and although the specific function of gp68 is not clear, we note that *whiB*-like genes are quite prevalent among the mycobacteriophages, and four of the nine *whiB*-containing genomes contain two copies. Giles gene 79 encodes a protein with similarity to MetE and is discussed in further detail below. Few of the genes in the 38-to-79 segment are related to other mycobacteriophage genes, but those that are identifiable are related to different genomes, consistent with a mosaic genome architecture (Fig. 2).

Is Giles a temperate phage? Giles forms plaques on lawns of *M. smegmatis* that are neither absolutely clear (for example, D29 [9]) nor as turbid as those of the obvious temperate phages, such as L5 and Bxb1 (Fig. 5A) (13, 17). Many mycobacteriophages are in this category and apparently do not form lysogens of *M. smegmatis* at high frequency (12). However, many of these, like Giles, have an integration cassette, suggesting that they either are competent to form lysogens (albeit at a reduced frequency in *M. smegmatis*) or possibly are recent derivatives of temperate parents that have lost their lysogenic functions during the isolation procedures; it is also possible that they efficiently lysogenize bacterial hosts other than *M. smegmatis*.

To determine whether lysogens could be recovered from a Giles infection, cells from a spot where Giles particles had infected a lawn of *M. smegmatis* were recovered and grown on solid media. Two independent colonies (i.e., potential lysogens) were restreaked on solid media two more times to remove contaminating viral particles from the initial infection, and at each stage, isolated colonies were patched onto *M. smegmatis* lawns to test for phage release; all of the colonies recovered from both putative lysogens showed phage release (data not shown). Both of the putative lysogens were shown to



FIG. 5. Analysis of Giles lysogens of *M. smegmatis*. (A) Top agar lawns were prepared with either a putative Giles lysogen or nonlysogenic mc<sup>2</sup>155 (as indicated), and 5  $\mu$ l of serial dilutions of phage L5 or Giles was spotted onto the lawn. Giles lysogens are immune to superinfection by Giles, and killing is only observed at the highest phage dose (corresponding to 10<sup>10</sup> PFU). (B) A Giles lysogen was similarly tested for immunity to a collection of other mycobacteriophages, all of which infect the lysogen and nonlysogen similarly. A key to the phages tested is shown below.

be immune to superinfection by Giles (Fig. 5A), although one showed a large number of individual plaques throughout the lawn, suggesting that the prophage was not stably maintained (data not shown), and this strain was not examined further. We conclude that Giles is competent to form lysogens in *M. smegmatis* but does not share immunity with any other phage that we have tested (Fig. 5B).

To determine the frequency of lysogenization under a set of standard conditions, we determined the proportion of bacterial colonies recovered following the plating of dilutions of an *M. smegmatis* culture on solid media seeded with  $10^9$  Giles particles. Using dilutions that yielded approximately 2,000, 200, and 20 colonies on control medium, colonies were recovered at a frequency of 2%, reflecting the frequency of lysogenization. It has been shown previously using similar assays that L5 lysogenizes at a frequency of about 22% but that a mutant with a clear plaque morphology lysogenizes at only 7% (7, 33). Thus, even this L5 mutant lysogenized at a higher frequency than

Giles. The molecular basis for the low frequency of Giles lysogenization is not clear.

While phage integration functions are relatively easy to identify in most phages that encode them, the immunity functions are diverse. For example, while the repressors of L5 and Bxb1 have been identified experimentally (7, 17), and related copies are present in mycobacteriophages Bxz2, Che12, U2, and Bethlehem (12), there are no identifiable homologues in any other phage or bacterial genome sequences. Examination of the Giles genome revealed no homologues of the L5/Bxb1 repressor group or of any other phage repressors (Fig. 2). An intriguing candidate, and the only gene related to other transcriptional regulators, is gene 68, which encodes a WhiB family protein; moreover, this gene is located in a region (i.e., 10 to 20% of the genome length to the left of the right end) where phage repressors are commonly positioned. However, repressor genes are typically associated with closely linked upstream regulatory sequences, and the nearest likely location is more than 3.5 kbp away, in the gene 61-to-62 noncoding region (Fig. 2). Interestingly, there are WhiB family members in a number of other mycobacteriophage genomes (e.g., Tweety, Llij, Che9d, PMC, and Che8), many of which also have integration cassettes.

The unusual location and organization of the Giles integration cassette. Sixteen of the previously characterized mycobacteriophages have an integrase gene that is typically located close to the center of the genome and separates the structural genes to the left (i.e., the left arm) and the nonstructural genes to the right (i.e., the right arm) (12); in all of these genomes, the lysis functions are encoded to the left of the integrase gene. Giles represents a significant departure from this pattern, since the group of seven rightward-transcribed genes to the right of the excise gene (30) encode the lysis functions (genes 31 and 32) (Fig. 2), and genes 34 and 35 are homologues of genes found in the structural operons of mycobacteriophages Halo, PG1, Cooper, and Orion (12); in Halo, genes 30 and 31 (homologues of Giles 34 and 35, respectively) are situated immediately to the left of the integrase gene. Furthermore, the product of Giles gene 36 (gp36) is likely a structural protein and is present in Giles virions (Fig. 3).

A simple interpretation of the Giles organization is that the integrase/excise cassette has relocated to a position within the "left arm" of the genome so that it now sits amid the structural genes (Fig. 6). This organization is additionally puzzling, since a putative rightward transcription terminator is located between the *attP* core and the integrase gene (see below), so that the 30-to-37 group of genes must be expressed either from an additional promoter or via an antitermination mechanism. It is unclear what mechanism gave rise to the repositioning of the integration cassette, although it is plausible that it occurred by an integrase-mediated event utilizing a secondary attachment site within the left arm rather than by an illegitimate-recombination process that could also give rise to other phage rearrangements (15, 28). This raises the questions of whether the integration cassette is functional and whether the lysogens described above contain an integrated prophage.

The Giles integrase (gp29) is of interest because it is not closely related to any of the other mycobacteriophage integrases and is more closely related to integrases encoded by a variety of bacterial prophages (see Fig. S1 in the supplemental



FIG. 6. Positioning of integration functions in the Giles structural operon. The Giles integration functions are atypically positioned among structural protein genes and the lysis genes. Shared Giles and Halo genes are joined by red shading, and known Giles virion proteins are labeled. While Halo and Giles are not globally related, the alignment illustrates the typical location of the integration cassette at the end of the structural gene operon in Halo, with the lysis and structural genes all to the left of *int*. In Giles, the integration cassette has moved into an atypical position within the structural operon, with the lysis genes to its right. Pham designations are shown above the genes as described in the legend to Fig. 2.

material). The closest relative is an integrase located in the genome of Mycobacterium sp. strain MCS, a strain whose genome is quite similar to that of *M. smegmatis*  $mc^{2}155$  (see Fig. S1 in the supplemental material). We propose that the Giles attachment site (attP) is located in the intergenic space to the left of the integrase gene (Fig. 2), and this region contains a 46-bp region (coordinates 25134 to 25179) that is nearly identical to a segment of the M. smegmatis genome overlapping the 3' end of the tRNA<sup>Pro</sup> gene (Msmeg 3734) (Fig. 7A). This locus is thus a strong candidate for the chromosomal *attB* site for Giles integration, with the common core at coordinates 3800548 to 3800593 in the M. smegmatis genome; we note that although there is a 1-base difference between the putative attP and attB core sequences, it lies within the variable loop of tRNA<sup>Pro</sup> (Fig. 7B). Interestingly, this putative attB site is occupied by prophage-like elements in *Mycobacterium* sp. strains MCS and KMS, and although the nearby int gene (i.e., Mmcs 2923) is the closest homologue of Giles gp29, it shares only 51% amino acid sequence identity. To confirm that the putative attB site is utilized for Giles integration during prophage formation, we characterized the lysogenic strain described above. In this strain, the predicted *attL* fragment, but not *attB*, could be amplified by PCR, confirming the use of an attB site at Msmeg 3734 (Fig. 7D).

To further analyze the putative Giles *attP* site, we searched for short 10- to 12-bp repeated sequences located near the 46-bp common core; in the well-characterized phage *attP* sites, such as of phage lambda (23) and mycobacteriophage L5 (30), these correspond to arm-type integrase binding sites and often are present as adjacent pairs of sites (30). In the Giles genome, we identified two pairs of repeats of the 12-bp sequence 5'-T TTGACCCCAAT, which we have designated P1 to P4 (Fig. 7A), on each site of the common core that likely correspond to arm-type binding sites for the Giles integrase.

Several mycobacteriophage genomes (such as L5) (13) contain transcriptional-terminator-like structures near the integration apparatus. Giles also contains a stem-loop terminator-like structure in this region, but it is atypically located within the putative *attP* site, positioned between the 46-bp common core and the P3/P4 pair of arm-type sites (Fig. 7A). While this is an unusual organization, there may be a strong selection for a terminator immediately to the right of the *attL* junction sequence in an integrated Giles prophage to prevent expression originating from the chromosomal tRNA gene from extending through to the *xis* gene (30), as well as the lysis functions. Errant expression of *xis* in a Giles prophage would likely lead to instability of lysogens, while premature expression of the lysis genes could likely be toxic to a Giles lysogen.

Construction of integration-proficient vectors. To determine whether the Giles integration cassette is functional, we constructed integration-proficient vectors containing the Giles attP site and integrase gene. Plasmids pGH1000A and pGH1000B were constructed that contained a 1.7-kbp attP-int cassette (coordinates 24922 to 26617) inserted in opposite orientations in a hygromycin-resistant nonmycobacterial replicating plasmid vector; all of *attP*, including the putative arm-type sites, was included, as well as the entire integrase gene and the 28-to-29 intergenic region (where *int* expression signals may reside), but not all of gene 30, which encodes the putative excise function. When electroporated into M. smegmatis, these plasmids generated transformants with an efficiency of  $3 \times 10^5$ to  $6 \times 10^5$  transformants/µg DNA, and analysis of the transformants by PCR showed that all contained the plasmid integrated at the predicted attB site (Fig. 7D); while the transformation frequencies are similar to those with extrachromosomally replicating plasmids, they are modestly lower (approximately 5- to 10-fold) than those with L5 and Bxb1 integration vectors. These vectors thus add to the repertoire of mycobacterial integration-proficient vectors that are compatible and stable and that generate single-copy recombinants (20, 21, 31).

Acquisition of host DNA by illegitimate recombination. Giles gene 79 is located at the right end of the genome but is transcribed leftward, in contrast to the genes to its immediate left. Database searches revealed that the protein product is closely related to parts of bacterial MetE proteins, although the related parts are confined to a small segment in the central





FIG. 7. Integration functions of mycobacteriophage Giles. (A) Nucleotide sequence of the intergenic region between Giles genes 28 and 29 (coordinates 24990 to 25285) that includes the *attP* site (Fig. 2). The 46-bp common core shared with the *M. smegmatis* genome is shown as a thick horizontal line with the single base difference boxed; the coordinates of the common core in Giles are 25134 to 25179; putative arm-type integrase binding sites (P1 to P4) are also shown. Located between the common core and the right arm-type binding sites is a putative stem-loop terminator for rightward transcription. (B) Structure of *M. smegmatis* tRNA<sup>Pro</sup> (Msmeg\_3734), which includes the Giles *attB* site with an arrow indicating the position corresponding to the 5' position of the common core. The tRNA position that changes following Giles integration as a consequence of the base difference is circled. (C) Organization of integration-proficient plasmid pGH1000A and its integration into the *M. smegmatis* chromosome. (D) PCR amplification of attachment junctions of Giles lysogen and integrated plasmids. Isolated colonies from transformations with Giles 100 µl sterile distilled H<sub>2</sub>O and boiled at 95°C for 6 min, and 1 µl was used as a template in PCRs with primer pairs that amplify either a 770-bp *attL* fragment or an 842-bp *attB* fragment. The *attL* product is amplified from the Giles lysogen and from all Giles transformants; the *attB* product is obtained with control colonies and wild-type mc<sup>2</sup>155.

part of MetE (Fig. 8A). Interestingly, gene 79 contains a 203-bp segment that is 100% identical to the *M. smegmatis metE* gene (Msmeg\_6638), although the alignment can be modestly extended at the right end by deletion of a single base

in the Giles 79 sequence (Fig. 8B). Because the shared sequence has not yet diverged, the acquisition of it by Giles must have occurred quite recently, and thus the junctions between conserved and flanking sequences likely correspond closely to



FIG. 8. Acquisition of host DNA by illegitimate recombination. (A) Giles gene 79 encodes a putative 103-residue protein with sequence similarity to *M. smegmatis* MetE; residues 20 to 86 of Giles gp79 are identical to residues 493 to 559 of MetE. (B) Nucleotide sequence comparison showed that the central part of Giles gene 79 contains a 203-bp sequence block that is 100% identical to a central part of the *M. smegmatis metE* gene (Msmeg\_6638). It is not known if Giles gp79 is functional, but the common DNA sequence has apparently been acquired from a mycobacterial chromosome (possibly *M. smegmatis*) by illegitimate recombination. (C) A transition in GC% at the right end of Giles gene 79 likely reflects an illegitimate-recombination event at the right end of the shared nucleotide block. The GC%s of both the *M. smegmatis* and Giles genomes are 63 to 65%, but the right end of the Giles genome has a substantially lower GC%. The panel corresponds to approximately 1.5 kbp at the extreme right end of the Giles genome. The reduction in GC% occurs at coordinate ~54000, and the GC% of the segment between there and the end of the genome is about 50%.

the sites of recombination. Since it is not obvious that this sequence could have been acquired by any targeted process, we conclude that it occurred through illegitimate recombination. Furthermore, we note that there is a segment of low-GC% DNA at the end of the Giles genome, and the transition between high and low GC% appears to correspond closely to the right junction of the conserved sequences (Fig. 8C), consistent with an illegitimate-recombination event at this site. While the acquisition of bacterial genomic sequences has been proposed to occur by such a process (15, 28), this is-to our knowledgethe first instance where an early stage in the process has been captured. In considering which specific host genome gave rise to the Giles sequence, it is likely that it came directly from either *M. smegmatis* or a very closely related strain. We note, for example, that the next closest bacterial *metE* sequences are in Mycobacterium avium and Saccharopolyspora erythraea, which have only 85% nucleotide identity to Giles gene 79. It seems unlikely that gp79 has any functional features of MetE, since MetE is a large protein (771 residues in *M. smegmatis*), and only 66 of the residues are present in gp79. While this segment includes several residues involved in the binding of folate (29), it is not obvious from the three-dimensional structure of MetE that a complete folate binding domain could be formed (29). While gp79 may indeed be nonfunctional, its acquisition from the bacterial chromosome could have occurred to satisfy a required packaging size of the genome. It seems reasonable to assume that this sequence could serve as a substrate for subsequent homologous-recombination events with the chromosome that would lead to creation of a functional gene.

## DISCUSSION

We have described here the isolation and genomic characterization of the novel mycobacteriophage Giles. While more than 30 mycobacteriophage genomes have been completely sequenced, their genomic diversity appears to be sufficiently high that the characterization of the new Giles sequence sheds considerable new light on phage diversity and evolution. Furthermore, with more than 50% of the putative Giles open reading frames encoding proteins with no apparent sequence similarity to other mycobacteriophage products, we are clearly far from a description of the complete repertoire of mycobacteriophage genes.

The pervasive mosaicism among bacteriophage genomes can generally be accounted for by assuming not only that illegitimate recombination has occurred widely throughout viral evolution, but that it represents an especially creative process. A more interesting question that then arises concerns who actually participates in this orgy of illicit genetic exchange. Analysis of the Giles genome suggests that it involves a broad variety of phage genomes, as illustrated by the large number of individual genes that are not related to other known phages, as well as the host chromosome. Such recombination events between phages and their bacterial hosts have been postulated previously, but Giles represents one of only very few examples where the event giving rise to this acquisition can be inferred. Thus, a picture of microbial evolution emerges that is messy, in which genetic variation is generated by illegitimate recombination, homologous recombination, and site-specific recombination and persists within the population either due to direct selection or as the nonadaptive consequence of other pressures, such as viral genome size selection.

A puzzling aspect of the phages of *M. smegmatis* is the predominance of those that form plaques that are neither obviously turbid nor clear but rather have plaque morphologies that are somewhere in between. The genomes of many of these phages show the presence of integration cassettes, suggesting that they either can form lysogens or are derivatives of temperate parents (12). While Giles has a similar plaque morphology, in this case, it is clear that the phage is competent to form lysogens carrying an integrated prophage, and the low frequency of lysogenic establishment accounts for the barely turbid plaque type. The location of the phage immunity functions remains unclear from bioinformatic analyses and will need to be determined empirically, as was done for mycobacteriophage L5 (7).

A notable feature of mycobacteriophage genomes is that while many contain integration cassettes, the predicted sites for chromosomal integration vary considerably (31). As expected, phages that share considerable similarity typically have similar integration systems, although at least 10 different integration sites have been postulated (31), with Giles choosing yet another site within a tRNA<sup>Pro</sup> gene. Interestingly, this site appears to be used by a yet-uncharacterized prophage that is present in the sequenced mycobacterial genomes MCS and KMS. We also note that the Giles *attB* site is the closest of any of these to the predicted terminus of DNA replication, which is estimated to be located at approximately 3.41 Mbp in the M. smegmatis mc<sup>2</sup>155 genome (H. Hendrickson, J. van Kessel, and J. Lawrence, unpublished data). Giles-derived integration-proficient vectors may thus be useful for inserting genes into the terminus-proximal region and using them to study expression and recombination in this part of the genome. It seems likely that many more integration sites will be identified for other mycobacteriophages, and it seems reasonable that all or most of the 47 M. smegmatis tRNA genes may be utilized. This makes the development of integration-proficient vector systems especially attractive, since multiple integration events can be used to construct complex genetic variants (31).

Finally, while mycobacteriophages represent one of the beststudied groups of phages that infect a common bacterial host, the apparently high degree of genetic diversity suggests that the genomic characterization of mycobacteriophages will continue to reveal interesting surprises and insights into phage diversity and evolution. Questions regarding the origins of genomic mosaicism, the number of exchangeable modules, the identities of recombining participants, and the frequency of genetic exchanges remain largely unanswered. However, characterization of the Giles genome suggests that continued phage genomic characterization will throw considerable light upon these questions.

### ACKNOWLEDGMENTS

We are grateful to Amy Vogelsberger and Molly Scanlon for excellent technical support and to Julia van Kessel for comments on the manuscript. Amy Vogelsberger also assisted with isolation of Giles lysogens, and Jennifer Houtz and Alexis Smith assisted with DNA sequencing and analysis. We also thank Tom Harper for help with electron microscopy, John Hempel for N-terminal sequencing, and Steven Cresawn for help with map construction and sequence analyses.

This work was supported by a grant to the University of Pittsburgh from the Howard Hughes Medical Institute (HHMI) in support of Graham Hatfull under HHMI's Professors program and by a grant from NIH to R.W.H. (GM51975).

### REFERENCES

- Angly, F. E., B. Felts, M. Breitbart, P. Salamon, R. A. Edwards, C. Carlson, A. M. Chan, M. Haynes, S. Kelley, H. Liu, J. M. Mahaffy, J. E. Mueller, J. Nulton, R. Olson, R. Parsons, S. Rayhawk, C. A. Suttle, and F. Rohwer. 2006. The marine viromes of four oceanic regions. PLoS Biol. 4:e368.
- Baranov, P. V., R. F. Gesteland, and J. F. Atkins. 2002. Recoding: translational bifurcations in gene expression. Gene 286:187–201.
- Brussow, H. 2001. Phages of dairy bacteria. Annu. Rev. Microbiol. 55:283– 303.
- Campbell, A. 1958. The different kinds of transducing particles in the lambda-gal system. Cold Spring Harbor Symp. Quant. Biol. 23:83–84.
- Casas, V., and F. Rohwer. 2007. Phage metagenomics. Methods Enzymol. 421:259–268.
- Clark, A. J., W. Inwood, T. Cloutier, and T. S. Dhillon. 2001. Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. J. Mol. Biol. 311:657–679.
- Donnelly-Wu, M. K., W. R. Jacobs, Jr., and G. F. Hatfull. 1993. Superinfection immunity of mycobacteriophage L5: applications for genetic transformation of mycobacteria. Mol. Microbiol. 7:407–417.
- Edwards, R. A., and F. Rohwer. 2005. Viral metagenomics. Nat. Rev. Microbiol. 3:504–510.
- Ford, M. E., G. J. Sarkis, A. E. Belanger, R. W. Hendrix, and G. F. Hatfull. 1998. Genome structure of mycobacteriophage D29: implications for phage evolution. J. Mol. Biol. 279:143–164.
- Ford, M. E., C. Stenstrom, R. W. Hendrix, and G. F. Hatfull. 1998. Mycobacteriophage TM4: genome structure and gene expression. Tuber. Lung Dis. 79:63–73.
- Gomez, M., and I. Smith. 2000. Determinants of mycobacterial gene expression, p. 111–147. *In* G. F. Hatfull and W. R. Jacobs, Jr. (ed.), Molecular genetics of the mycobacteria. ASM Press, Washington, DC.
- 12. Hatfull, G. F., M. L. Pedulla, D. Jacobs-Sera, P. M. Cichon, A. Foley, M. E. Ford, R. M. Gonda, J. M. Houtz, A. J. Hryckowian, V. A. Kelchner, S. Namburi, K. V. Pajcini, M. G. Popovich, D. T. Schleicher, B. Z. Simanek, A. L. Smith, G. M. Zdanowicz, V. Kumar, C. L. Peebles, W. R. Jacobs, Jr., J. G. Lawrence, and R. W. Hendrix. 2006. Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. PLoS Genet. 2:e92.
- Hatfull, G. F., and G. J. Sarkis. 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. Mol. Microbiol. 7:395–405.
- Hendrix, R. W. 2002. Bacteriophages: evolution of the majority. Theor. Popul. Biol. 61:471–480.
- Hendrix, R. W., M. C. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA 96:2192– 2197.
- Hendrix, R. W. 2003. Bacteriophage genomics. Curr. Opin. Microbiol. 6:506– 511.
- Jain, S., and G. F. Hatfull. 2000. Transcriptional regulation and immunity in mycobacteriophage Bxb1. Mol. Microbiol. 38:971–985.
- Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. 299:27–51.

- Katsura, I., and R. W. Hendrix. 1984. Length determination in bacteriophage lambda tails. Cell 39:691–698.
- Kim, A., P. Ghosh, M. A. Aaron, L. A. Bibb, S. Jain, and G. F. Hatfull. 2003. Mycobacteriophage Bxb1 integrates into the *Mycobacterium smegmatis* groEL1 gene. Mol. Microbiol. 50:463–473.
- Kwan, T., J. Liu, M. Dubow, P. Gros, and J. Pelletier. 2006. Comparative genomic analysis of 18 *Pseudomonas aeruginosa* bacteriophages. J. Bacteriol. 188:1184–1187.
- Kwan, T., J. Liu, M. DuBow, P. Gros, and J. Pelletier. 2005. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. Proc. Natl. Acad. Sci. USA 102:5174–5179.
- Landy, A. 1989. Dynamic, structural, and regulatory aspects of lambda sitespecific recombination. Annu. Rev. Biochem. 58:913–949.
- 24. Lee, M. H., L. Pascopella, W. R. Jacobs, Jr., and G. F. Hatfull. 1991. Site-specific integration of mycobacteriophage L5: integration-proficient vectors for *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, and bacille Calmette-Guerin. Proc. Natl. Acad. Sci. USA 88:3111–3115.
- Mediavilla, J., S. Jain, J. Kriakov, M. E. Ford, R. L. Duda, W. R. Jacobs, Jr., R. W. Hendrix, and G. F. Hatfull. 2000. Genome organization and characterization of mycobacteriophage Bxb1. Mol. Microbiol. 38:955–970.
- Mukamolova, G. V., A. S. Kaprelyants, D. I. Young, M. Young, and D. B. Kell. 1998. A bacterial cytokine. Proc. Natl. Acad. Sci. USA 95:8916– 8921.
- Nesbit, C. E., M. E. Levin, M. K. Donnelly-Wu, and G. F. Hatfull. 1995. Transcriptional regulation of repressor synthesis in mycobacteriophage L5. Mol. Microbiol. 17:1045–1056.
- 28. Pedulla, M. L., M. E. Ford, J. M. Houtz, T. Karthikeyan, C. Wadsworth, J. A. Lewis, D. Jacobs-Sera, J. Falbo, J. Gross, N. R. Pannunzio, W. Brucker, V. Kumar, J. Kandasamy, L. Keenan, S. Bardarov, J. Kriakov, J. G. Lawrence, W. R. Jacobs, R. W. Hendrix, and G. F. Hatfull. 2003. Origins of highly mosaic mycobacteriophage genomes. Cell 113:171–182.
- Pejchal, R., and M. L. Ludwig. 2005. Cobalamin-independent methionine synthase (MetE): a face-to-face double barrel that evolved by gene duplication. PLoS Biol. 3:e31.
- Peña, C. E., M. H. Lee, M. L. Pedulla, and G. F. Hatfull. 1997. Characterization of the mycobacteriophage L5 attachment site, *attP*. J. Mol. Biol. 266:76–92.
- 31. Pham, T. T., D. Jacobs-Sera, M. L. Pedulla, R. W. Hendrix, and G. F. Hatfull. 2007. Comparative genomic analysis of mycobacteriophage Tweety: evolutionary insights and construction of compatible site-specific integration vectors for mycobacteria. Microbiology 153:2711–2723.
- Piuri, M., and G. F. Hatfull. 2006. A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells. Mol. Microbiol. 62:1569–1585.
- Sarkis, G. J., W. R. Jacobs, Jr., and G. F. Hatfull. 1995. L5 luciferase reporter mycobacteriophages: a sensitive tool for the detection and assay of live mycobacteria. Mol. Microbiol. 15:1055–1067.
- Susskind, M. M., and D. Botstein. 1978. Molecular genetics of bacteriophage P22. Microbiol. Rev. 42:385–413.
- 35. Telkov, M. V., G. R. Demina, S. A. Voloshin, E. G. Salina, T. V. Dudik, T. N. Stekhanova, G. V. Mukamolova, K. A. Kazaryan, A. V. Goncharenko, M. Young, and A. S. Kaprelyants. 2006. Proteins of the Rpf (resuscitation promoting factor) family are peptidoglycan hydrolases. Biochemistry 71:414–422.
- van Kessel, J. C., and G. F. Hatfull. 2007. Recombineering in Mycobacterium tuberculosis. Nat. Methods 4:147–152.
- Xu, J., R. W. Hendrix, and R. L. Duda. 2004. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. Mol. Cell 16:11–21.
- 38. Zimmer, M., E. Sattelberger, R. B. Inman, R. Calendar, and M. J. Loessner. 2003. Genome and proteome of *Listeria monocytogenes* phage PSA: an unusual case for programmed + 1 translational frameshifting in structural protein synthesis. Mol. Microbiol. 50:303–317.