



Published in final edited form as:

Mol Microbiol. 2013 May ; 88(3): 577–589. doi:10.1111/mmi.12210.

Functional requirements for bacteriophage growth: Gene essentiality and expression in Mycobacteriophage Giles

Rebekah M. Dedrick^a, Laura J. Marinelli^b, Gerald L. Newton^c, Kit Pogliano^c, Joseph Pogliano^c, and Graham F. Hatfull^a

Graham F. Hatfull: gfh@pitt.edu

^aDepartment of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, T:412 831 2734, F:412 624 4870

^cDivision of Biological Sciences University of California San Diego, La Jolla, CA

Summary

Bacteriophages represent a majority of all life forms, and the vast, dynamic population with early origins is reflected in their enormous genetic diversity. A large number of bacteriophage genomes have been sequenced. They are replete with novel genes without known relatives. We know little about their functions, which genes are required for lytic growth, and how they are expressed. Furthermore, the diversity is such that even genes with required functions – such as virion proteins and repressors – cannot always be recognized. Here we describe a functional genomic dissection of mycobacteriophage Giles, in which the virion proteins are identified, genes required for lytic growth are determined, the repressor is identified, and the transcription patterns determined. We find that although all of the predicted phage genes are expressed either in lysogeny or in lytic growth, 45% of the predicted genes are non-essential for lytic growth. We also describe genes required for DNA replication, show that recombination is required for lytic growth, and that Giles encodes a novel repressor. RNAseq analysis reveals abundant expression of a small non-coding RNA in a lysogen and in late lytic growth, although it is non-essential for lytic growth and does not alter lysogeny.

Keywords

Bacteriophage; Transcription; RNAseq

Introduction

The size, age, and dynamic nature of the bacteriophage population contribute to their vast genetic diversity (Hatfull & Hendrix, 2011). Not only do phages infecting hosts of different bacterial genera typically share little or no nucleotide sequence similarity, but phages infecting the same specific bacterial strain can also encompass large genetic diversity (Hatfull & Hendrix, 2011, Krupovic *et al.*, 2011). Moreover, it is common for large proportions of phage genes (>75%) to fail to have significant sequence similarity to genes outside of the close phage relatives, and given the massive size of the population, bacteriophages likely represent the largest reservoir of unexplored sequences in the biosphere (Mokili *et al.*, 2012). A central challenge in phage biology is thus to elucidate the functions of these unknown genes.

^bCurrent Address: Division of Dermatology, Department of Medicine and Department of Microbiology, Immunology, and Molecular Genetics, David Geffen School of Medicine University of California Los Angeles, Los Angeles, CA

The authors have no conflict of interest to declare.

Mycobacteriophages are viruses of mycobacterial hosts including *Mycobacterium tuberculosis* and *M. smegmatis*. Comparative analysis of over 220 completely sequenced genomes shows that they are mosaic, with DNA segments corresponding to single genes being pervasively exchanged among phages in the environment (Pedulla *et al.*, 2003, Hatfull, 2010). All of these genomes infect a single common host strain, *M. smegmatis* mc²155, and span considerable diversity and host range profiles (Jacobs-Sera *et al.*, 2012). To simplify the genomic analysis, closely related genomes (nucleotide sequence similarity spanning more than 50% of genome length) are grouped in clusters (Cluster A, B etc), with some clusters being further divided into subclusters reflecting genome nucleotide variation. Currently, the ~220 sequenced mycobacteriophage genomes deposited in GenBank are grouped into 15 clusters, and eight singletons, i.e. phages for which close relatives have yet to be identified (Hatfull, 2012a, Pope *et al.*, 2011a, Pope *et al.*, 2011b).

Giles is a singleton mycobacteriophage that is temperate in *M. smegmatis*, and contains a 53,746 bp genome with 14-bp 3' single-stranded extensions (Morris *et al.*, 2008); the DNA is presumably packaged by a *cos*-packaging mechanism. Twelve virion structure and assembly genes were identified but more than 50% of its predicted genes encode proteins with no close sequence similarity (>32.5% identity) to other mycobacteriophages and the functions of fewer than 30% of its genes can be predicted (Fig. 1) (Morris *et al.*, 2008). The Giles genome was established as a good substrate for development of the Bacteriophage Recombineering of Electroporated DNA (BRED) system that enables simple construction of phage mutants (Marinelli *et al.*, 2008, Marinelli *et al.*, 2012), and this was used to demonstrate roles of the LysA and LysB lysins (Marinelli *et al.*, 2008, Payne *et al.*, 2009). The full extent of the Giles host range is not known, but it forms plaques on other strains of *M. smegmatis* at a greatly reduced efficiency of plating (Jacobs-Sera *et al.*, 2012). It does not infect *M. tuberculosis*, but plaques can be recovered as infectious centers on lawns of *M. smegmatis* following introduction of Giles genomic DNA into *M. tuberculosis* by electroporation (Jacobs-Sera *et al.*, 2012).

Mycobacteriophages represent a rich resource of tools for mycobacterial genetics as well as novel strategies for rapid TB diagnosis and drug susceptibility testing (Jacobs *et al.*, 1993, Piuri *et al.*, 2009, Hatfull, 2012b). Mycobacterial-specific recombineering systems have been derived from mycobacteriophage Che9c (van Kessel & Hatfull, 2007, van Kessel & Hatfull, 2008), and a variety of integration-proficient vectors have been described, including those generated from phage Giles (Morris *et al.*, 2008, Pham *et al.*, 2007, Lee *et al.*, 1991, Pope *et al.*, 2011a). Further exploitation of mycobacteriophage genomes is limited by our poor understanding of the overall patterns of gene expression, regulation, gene function, and gene essentiality.

Using a combination of transcriptomic and functional genomic approaches we describe the transcription patterns of mycobacteriophage Giles and show that at least 35 of its 78 predicted genes are non-essential for lytic growth, including three virion-associated proteins. A small non-coding RNA (ncRNA) is expressed at high levels both in lysogeny and in late lytic growth, but is non-essential and has no known function.

Results

Giles structural proteins

Mycobacteriophage Giles has a siphoviral morphology and a genome architecture sharing features with the large group of siphoviral phages including phage λ (Morris *et al.*, 2008); the *attP* attachment site is located near the center of the genome and defines the left and right arms (genes 1–28 and 29–78 respectively). The left arm encodes the rightwards-transcribed virion structure and assembly functions, interrupted by three leftwards-transcribed genes

between the terminase small and large subunit genes (Fig. 1). Of the 11 virion proteins identified previously (Morris *et al.*, 2008) one (gp36) is unusually encoded within the genome right arm (Fig. 1). Mass spectrometry of whole virion particles identified a total of 20 proteins (Table 1), 18 of which are encoded in the left arm, (Fig. 1) as well as a second right-arm encoded protein, gp37 (Table 1, Fig. 1). Although only one peptide of gp37 was identified, the predicted protein is small (50 aa) and generates only two possible peptides of significant complexity (>7 amino acids) by trypsin digestion (DVTNSQWTAHTQQMNR and LLEAEG LQQTGK). The latter peptide was identified with a 99% confidence from its mass spectrum and represents 24% of the protein sequence. The peptide is complex and the fragmentation pattern of b and y ions closely matches that expected from the sequence. The presence of both protease (gp7) and scaffold (gp8) suggests that these are incompletely removed from proheads following DNA packaging, although we cannot rule out contamination of the phage preparation with incompletely assembled particles.

Most of the products encoded in the left arm (1–28) are virion associated, although seven are not, including the putative small and large terminase subunits (gp1 and gp4 respectively, which are required for packaging), and gp17, gp18 and gp19, which may function in tail assembly (Morris *et al.*, 2008). However, the organization of the region between the Giles major tail subunit (15) and tape measure (20) genes is a departure from that of other phages where two ORFs [presumed to be Giles 16 and 17 (Morris *et al.*, 2008)] coding for tail assembly chaperones are expressed by a programmed translational frameshift (Hatfull & Sarkis, 1993, Xu *et al.*, 2004). We also did not find gp23 or gp26, which are encoded among the other tail genes (Fig. 1).

Giles genes required for lytic growth

We determined which Giles genes are required for lytic growth using the BRED strategy described previously (Marinelli *et al.*, 2008). In this method, phage genomic DNA is co-electroporated into a recombinering strain of *M. smegmatis* with a DNA substrate (typically about 200 bp long) that contains the mutant allele – either a specific gene deletion or a point mutation – and plaques recovered on *M. smegmatis* plating cells after a short recovery period. Each plaque is thus derived from a single cell that has taken up phage DNA, and at least 10% of these typically contain a mixture of the wild-type and mutant alleles, from which a homogenous mutant of a non-essential gene (i.e. a gene that is not required to form a visible plaque) can be recovered after further purification. Because the overall process is efficient, mutants can be identified by physical characterization (PCR) without the need for selection. If a mutation is deleterious to lytic growth, then a mixed primary plaque can usually be recovered – because of complementation by wild-type particles in the same plaque – but cannot be recovered after subsequent purification.

Of the 78 predicted Giles genes, we selected 54 for deletion avoiding most of the virion structure and assembly genes, which are expected to be essential for lytic growth (i.e. is required to form a visible plaque) (Hendrix *et al.*, 1983). Of the 54 genes tested, 35 (65%) were determined to be non-essential for lytic growth (and can thus be isolated as a homogenous mutant population; Table 2, 50 Figs. 1 and 2A–C). Most of these form normal sized plaques containing similar numbers of particles as wild-type Giles (Table 2; Fig. S2); several mutants show mild losses in fecundity but only Δ shows a large reduction (Table 2) suggesting it plays an important role, even though it is non-essential for lytic growth.

Nineteen ORFs are predicted to be essential for lytic growth (and cannot be recovered as a homogenous mutant population), and we attempted to recover each of the mutants using plasmid-mediated complementation. For two of these (genes 31 and 64) the complementation was successful [the gene 31 deletion was reported in a previous study (Payne *et al.*, 2009)], and the mutant was isolated and purified in the complementing strain

(Fig. 2D–G). The $\Delta 64$ mutant was shown to only form plaques when plated on the complementation strain, but not on a wild-type strain. Giles gene *64* is thus essential for lytic growth (Fig. 2G). The remaining 17 mutants identified in initial BRED platings could not be further propagated even when plasmid-encoded genes were provided (Table 2), and for some we constructed plasmids with pairs of complementing genes (such as *62* and *63*) but still failed to recover the mutant. Complementation fails presumably due to poor expression of the complementing gene or expression at inappropriate stoichiometry, loss of an essential *cis*-acting element, or because of genetic polarity. Nonetheless, the simple conclusion from the BRED approach is that each of these genes is required for lytic growth.

In one case, we were able to isolate mixed primary plaques for a gene *67* deletion but a pure mutant was difficult to isolate. As this gene was likely to be required for lytic growth, a complementation plasmid was constructed and a pure mutant was isolated on the complementing strain. We found that the $\Delta 67$ mutant phage does form plaques on *M. smegmatis* mc²155 in the absence of complementation, but the plaques are extremely small and barely visible (Fig. S1). The gene is therefore designated as being non-essential for lytic growth, although it is clearly important. Two of the deletion mutants ($\Delta 32$ and $\Delta 29$) required removal of ~1.2 kbp DNA representing a reduction of genome size of about 2%. Because these were successfully generated, this reduction has no significant impact on DNA packaging. We note that we were able to successfully generate a double mutant that removes both gene *73* and *74*; other double mutants were not attempted.

Finally, some genes might appear to be essential in this assay as a consequence of misannotation of reading frames that are adjacent to essential components, including *cis*-acting elements. One example of this emerged through the failure to construct a deletion mutant of gene *75*. However, there is a short non-coding gap between genes *74* and *75*, which reduces confidence in the choice of the translation initiation codon in the current annotation (Fig. 1), and RNAseq data suggests that this is incorrect (see below). We therefore used an alternative substrate to remove the 3'-half of the originally annotated gene *75*, assuming use of an alternative initiation codon at coordinate 52,251. This mutant was created successfully, and we conclude that the revised *75* is non-essential. A second example is a gene originally annotated as *48*, which we were unable to delete. However, functional characterization of the flanking genes *47* and *49* coupled with transcriptomic analysis (see below) showed that the *48* open reading frame lies in an intergenic regulatory region, accounting for the inability to remove it. We have thus removed gene *48* from the genome annotation. These corrections are included in an updated Genbank file (accession number EU203571.3).

Roles of Giles genes 50, 64 and 67 in DNA replication

Although few genes in the Giles right arm have known functions, it is likely that at least some are involved in phage DNA replication. Because the $\Delta 64$ phage could be constructed and propagated on a complementing strain, we tested whether it is defective in DNA replication in a non-complementing host. Using qPCR, we observed no replication of $\Delta 64$ phage DNA following infection of a wild-type host, and the defect was largely restored in the complementing strain (Fig. 3A). This is consistent with gp64 having weak but significant similarity (Probability=95.18, E-value=0.025) to the G39P helicase loader of *Bacillus* phage SPP1 shown by HHPred (Soding *et al.*, 2005). The higher level of Giles DNA replication in strain mc²155pRMD1 compared to mc²155 could be a consequence of a higher level of gp64 expression. This gene is not being controlled in its native state (in the phage) and its dysregulation might contribute to elevated DNA replication. In contrast, DNA replication of the $\Delta 67$ mutant was observed in a wild-type strain although it is reduced from the parent phage and is only modestly enhanced by complementation (Fig. 3B). This is consistent with the predicted role of gp67 as a RuvC-like protein involved in Holliday Junction (HJ)

resolution. Giles $\Delta 50$ is viable on a non-complementing strain but shows a marked defect in fecundity and produces small plaques (Table 2) and like $\Delta 64$, it too shows a strong defect in DNA replication (Fig. 3B). There are no bioinformatic clues as to its specific function.

Identification of the Repressor

Although Giles is a temperate phage, and its integration system has been characterized (Morris *et al.*, 2008), its repressor has not been identified. Because wild-type Giles plaques are only lightly turbid, making it hard to distinguish clear plaque mutant phenotypes, we screened each of the deletion mutants for the ability to form lysogens on phage-seeded plates (Table 2). We observed significant defects in lysogeny in only two mutants, $\Delta 47$ and $\Delta 49$ (Fig. 4A and Table 2), and in both cases, lysogeny was reduced to below 0.01%. To determine which of these encodes the repressor, *47* and *49* were cloned, expressed and tested for the ability to confer immunity to Giles superinfection (Fig. 4B). Gene *47* confers strong immunity and we conclude that gp47 is the phage repressor. We note that gp47 has no close homologues and does not contain any readily identifiable DNA binding motifs. Gene *49* does not confer immunity and its role is unclear, although it could regulate repressor expression, similar to λ cII; gp49 has no close homologues but HHPred predicts a small zinc finger domain and it is likely a DNA binding protein.

Transcription of the Giles genome

We used RNAseq to determine transcription profiles in early and late Giles infections, 30 minutes and 2.5 hours after adsorption, respectively, according to the infection patterns described previously (Payne *et al.*, 2009) as well as in a Giles lysogen, and compared these to uninfected *M. smegmatis* (Figs. 1, 6, 7). Several datasets were generated to optimize the number of non-rRNA sequence reads, and all are in general agreement but differ significantly in overall quality (see Materials and Methods). We also used qRT-PCR to amplify each of the gene junctions in lysogeny and lytic growth (Fig. 5).

The transcription profile of a Giles lysogen is relatively simple. The strongest signal spans gene *47* with transcription initiating in the *47–49* intergenic region (Fig. 5, 6, 7G), consistent with gp47 being the phage repressor. The transcription start site was determined by 5' RACE (Fig. 4C), showing that initiation occurs at coordinate 36,674, in strong agreement with the RNAseq data. The promoter does not obviously correspond to σ -70 like promoters described in other mycobacteriophages and presumably a different sigma factor is used. The transcriptional level of the repressor is unusually high, with expression equivalent to the top 0.5th percentile of host genes, similar to ribosomal protein L2 and RpoC, in notable contrast to the relatively low but autoregulated levels of cI transcription in a λ lysogen (Ptashne *et al.*, 1980). Fusion of this promoter to an mCherry reporter gene shows strong expression in lysogens but little or no expression in wild-type cells, indicating that it strongly requires activation (data not shown). There is some transcription of the genes downstream of *47*, and qRT-PCR indicates genes 44–46 are also expressed in the lysogen (Fig. 5, 6, 7G). Strong transcription also is observed in a short (~100 bp) non-coding segment between genes *74* and *75*, and is discussed in further detail below (Fig. 7H). A modest level of genes *2–4* transcription was also observed (Figs. 5, 6, 7A).

Only modest levels of Giles transcription were observed during early lytic growth. The strongest transcription initiates between coordinates 36,975 and 36,990 in front of the rightwards-transcribed gene *49*, and gradually diminishes across the downstream operon up to gene *59*, at which point transcription is barely detectable. Both the repressor (*47*) and integrase (*29*) are expressed at low or barely detectable levels, although several other leftwards-transcribed genes are expressed, including genes *43* and *44*, *39–41*, and *2–3* (Figs. 5, 6, 7F). Transcription initiates within the putative RDF gene, suggesting an alternative start

site, (*30*; coordinate ~26,649; Fig. 7D, S3) but diminishes within lysin A (*31*). There are barely detectable levels of expression of any of the virion structure and assembly genes at this time. qRT-PCR analyses (Fig. 5) are in general agreement with the RNAseq data but good signals are observed all the way through to the end of the right end of the genome, suggesting that RNAseq data may somewhat under represent regions at the 3' end of long operons.

Late in infection the transcription pattern is markedly different. Interestingly, much of the early transcription pattern is retained, and the early and late levels from gene *49* to the right end of the genome are almost superimposable (Fig. 6). The notable exception is the very high level of transcription of the *74–75* intergenic region, a similar segment to that transcribed from the prophage (see below). The virion structure and assembly genes are expressed at high levels although with some apparent variation across the left arm (Fig. 6). Very high levels of expression begin with the portal gene (*6*) but diminish at the tape measure gene (*20*), return to high levels at gene *26* (Fig. 6, 7B–D), and stop near the factor-independent terminator following gene *28*. The right-arm genes *30–37* – including the two virion gene *36* and *37* and the lysis cassette – are also expressed late in lytic growth, but transcription rises sharply between genes *30* and *31* (Fig. 6, 7E–F).

These transcriptomic data suggest there are at least three early lytic promoters, upstream of genes *4*, *30*, and *49*, (Figs 1, 6). Transcription initiates upstream of gene *49* at around 36,973, although there are no obvious σ -70 like sequences and presumably another sigma factor is used (Fig. 4C); presumably this promoter is directly regulated by the gp47 repressor as it is not active in a lysogen. These same observations apply to a probable promoter upstream of gene *30*. There is a predicted σ -70 leftwards promoter upstream of gene *4* with a start site at coordinate 1,828, which is the first base of the ATG start codon, and other examples of leaderless mRNAs have been reported (Broussard *et al.*, 2012). There are several plausible promoters active in late lytic growth located upstream of genes *78*, *6*, *26*, *31*, and transcribed rightwards, and one for expression of *37* and/or *38*, although it is unclear which strand is being transcribed. None of these contain an obvious σ -70 like promoter and we assume that transcription from these requires an as yet unidentified Giles-encoded transcriptional activator.

Expression and role of a small non-coding RNA

High levels of a small (~100 nucleotide) transcript are observed from the *74–75* intergenic region in both late lytic and lysogenic growth (Fig. 6, 7H). This was confirmed by qRT-PCR which also demonstrated that it is expressed in the rightwards direction (Fig. 8A). To determine if it is required for lysogeny or lytic growth, we using BRED mutagenesis to delete coordinates 51,631 – 51,728 such that the flanking genes were not interrupted. The mutant was readily constructed, and we have been unable to identify any defect in either lytic growth or lysogeny. Fusion of the upstream region (coordinates 51,565 – 51,620) to a mCherry reporter gene showed an active promoter, although it is about 10-fold more active in a Giles lysogen than a non-lysogen (Fig. 8B). Metabolomic comparisons of lysogens carrying the deletion mutation and wild-type Giles revealed no significant differences in growth. We also found no evidence of the RNA being incorporated into Giles particles.

Discussion

The massive increase in bacteriophage genomics over the past ten years has revealed their enormous genetic diversity and a rich abundance of novel gene sequences. While a great deal is known about the detailed biology of a small number of phages such as λ , T4 and T7, the question arises as to what extent this knowledge applies to the broader population of phages, especially those that infect hosts other than *Escherichia coli*. Moreover, methods for

determining phage gene functions and transcriptomic profiles are not well established. Giles presents an excellent model system for functional genomic analysis as most Giles genes have no close relatives and few functions can be predicted. BRED mutagenesis provides a simple method for determining gene essentiality for a large proportion of its genome, but is likely to be effective not only for other mycobacteriophages, but also for phages of any other host in which recombineering is available.

The proportion of Giles genes required for lytic growth was not predictable bioinformatically. In phage λ at least 18 genes are non-essential – including nine in the b2 region (Hoess & Landy, 1978, Hendrix *et al.*, 1983) – and it is surprising that as many as 35 Giles genes – 45% of its genome – are dispensable for lytic growth. Other than the three virion-associated proteins and lysin B (Payne *et al.*, 2009) it is unclear what roles these play, and bioinformatic analyses provide little insight. The one exception is gp67, a putative RuvC-like HJ resolvase, but loss of 67 leads to barely viable phage suggesting that unresolved HJs interfere with DNA packaging. This indicates that recombination is active in Giles replication, and the essentiality of the RecE/T-like recombination proteins (gp52 and gp53) suggests that recombination is required, as in T4 (Kreuzer & Brister, 2010). Nonetheless, if extrapolated to the larger collection of 220 sequenced mycobacteriophages, these data suggest there are over 11,300 non-essential genes in almost 1,500 unique sequence families (Hatfull, 2012a).

All of the non-essential genes are expressed either in lytic or lysogenic growth. The most dramatic perhaps is the ncRNA between 74 and 75, expressed at high levels both in a lysogen and in late lytic growth. This ncRNA is not required for either lytic growth or lysogeny and its role remains unclear. However, this situation is not common, as we have also identified many protein-coding genes that are also not required for either lytic or lysogenic growth and for which a function has not yet been assigned. One attractive explanation is that at least some of these RNA and protein functions could confer protection from competing phage infections, as seen with restriction-modification and rogue immunity acquisition (Pope *et al.*, 2011b). However, this is difficult to examine directly unless an excluded phage can be identified, and because mycobacteriophage diversity is extremely high, such phage(s) may not as yet have been isolated. Alternatively, these genes could mediate changes to host gene expression, although there are only subtle differences between the lysogenic and non-lysogenic transcriptomes. We note that genes 2–4, which are expressed in both lytic and lysogenic growth, share a similar location to gene 2 of *Streptomyces* phage ϕ C31, between the small and large terminase subunits and oriented in the opposite direction to the structural genes (Smith *et al.*, 1999), so perhaps these have similar functions.

Nineteen ORFs are essential and the requirement for WhiB is surprising as the WhiB-like protein of phage TM4 is non-essential (Rybniker *et al.*, 2010). However, many mycobacteriophages encode WhiB-like proteins – some with several copies – and they are a highly diverse perhaps providing a variety of particular functions. Because the DNA methylase is essential, it is plausible that it provides a modification component of a restriction system, although the best candidate for a restriction endonuclease (based on HHPred analysis, which also shows similarity to an HNH nuclease and an HJ resolvase) is gp76, which is, however, also essential. HHPred also suggests that the essential gp42 is implicated in replication initiation, and that gp40 (with a C-terminal DNA binding motif) may be a regulator.

The identification of gp47 as the phage repressor – it is well-expressed in a lysogen and is both required for lysogeny and for superinfection immunity – is surprising given its complete lack of known DNA binding motifs. This illustrates the amazing genetic diversity

of the phage population in that even this thoroughly well-studied class of proteins cannot be readily predicted bioinformatically in phage genomes. Presumably gp47 regulates the rightwards early lytic promoter upstream of 49, and identifying what class of promoter is used and how it is regulated will be of interest.

Experimental Procedures

Bacterial strains and media

M. smegmatis mc²155, lysogens, and recombinants were grown and recombineering cells prepared as described previously (Morris *et al.*, 2008, Marinelli *et al.*, 2008). Tween was omitted and 1mM CaCl₂ was included in all media for phage infections. Plasmids used are listed in Table S1 and oligonucleotides in Table S2. A revised mycobacteriophage Giles GenBank file has been submitted to NCBI under accession number EU203571.3. See Supporting Information for further details regarding annotation revisions.

Construction of gene deletions

Giles deletions were constructed using BRED as described previously (Marinelli *et al.*, 2008). Briefly, PCR was used to produce 200bp substrates containing 100bp of homology to the upstream and downstream regions of the gene to be deleted. A 9-base unique bar coding 'tag' was inserted in place of the deleted coding region to facilitate mutant identification. Giles DNA and the target substrate were electroporated into *M. smegmatis* recombineering cells (van Kessel & Hatfull, 2007) and plated in an infectious center assay. In general, flanking primer (FP) PCR worked well in identifying mixed plaques and pure mutants, but tag-specific PCR was important to identify hard-to-isolate mutants in which FP PCR generated weak signals. In general this strategy was effective, although there was one example (gene 77) for which we were not able to recover a mixed primary plaque unless a substrate was used that lacked the 9-base tag.

For each mutant construction, we typically screened 16–20 primary plaques and mixed plaques were identified from 3% to 94%; there was no obvious correlation between the number of mixed plaques isolated and the essentiality of the gene. Secondary plaques were screened similarly, picking 16–20 for PCR validation. When a pure mutant could not be recovered from the secondary plating, several pools of 5–10 plaques were screened, and if this still did not verify that a mutant was present, several mixed plaque lysate dilutions were screened. If a mutant band was present in all lysate dilution plates, suggesting that the mutant was present and nonessential, then further single plaque PCR screenings were performed. For essential genes, mutant plaques were only seen at the lowest dilution of the mixed plaque lysate and not the higher dilutions. This suggested that the mutant could not survive without the presence of wild-type phage particles acting as helpers. Primers are listed in Table S2.

For complementation, mc²155 containing plasmids were grown to OD₆₀₀=0.4, e-caprolactam (Sigma) added at varying concentrations (0.2–1%), and cultures grown for 3 hrs at 37°C. Following infection with mixed primary plaques, plaques were recovered and screened by PCR.

Lysogeny Assays

Lysogeny frequencies were measured by plating dilutions of *M. smegmatis* (10⁴–10⁷ cfu) on agar plates seeded with 10⁹ pfu of each phage, and determining the number of colonies relative to non-seeded plates.

RNA analyses

Total RNA was isolated from a log-phase culture of an $OD_{600} = 1.0$. For every 500ul of culture, 1 ml of RNeasy Protect reagent (Qiagen) was added. Samples were pelleted and the RNeasy Mini Kit (Qiagen) was completed. Cells were broken in Matrix B (MP Biomedicals) using a Beadbeater twice for 45 seconds on max speed with 1min incubation on ice in between. After DNase I treatment (Invitrogen) the samples regularly contained 1 ug/ul of RNA. RiboZero (Epicentre) rRNA removal kit was used according to the manufacturers instructions; mRNA was then retreated with DNaseI. Removal of rRNA and concentrations of mRNA were confirmed using Agilent 2100 Bioanalyzer. Purified samples were stored at -80°C . Samples were prepared for RNA-seq using the TruSeq RNA Sample Preparation kit (Illumina #15026495) according to the manufacturers instructions. Samples were sent to Tufts Genomics Core Facility (Tufts University, MA, USA) where single-end libraries were subjected to 50-bp read Hi-Seq Illumina sequencing. Data was analyzed using Galaxy (Penn State University). First, Bowtie was used to map the RNAseq reads to the reference genome. This file was then filtered of unmapped reads (Filter SAM), and converted into a BAM file format (SAM-to-BAM). The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar *et al.*, 2002) accessible through GEO Series accession number GSE43434.

Primers for qRT-PCR were generated using PrimerExpress software (Applied Biosystems). cDNA was generated using random hexamers and Maxima reverse transcriptase (Fermentas), and qRT-PCR performed using the Maxima SYBR Green qPCR Master Mix (Fermentas). The Applied Biosystems 7300 Real-Time PCR System was the instrument used with cycling conditions of: 50°C for 2', 95°C for 10', and 40 cycles of 95°C for 15'', 60°C for 1'. Dissociation curves were produced to verify amplification. Raw data was analyzed using ABI 7300 software and other calculations were performed with Microsoft Excel. RT-PCR primers were designed manually and standard 50ul PCR reactions were used with 10ng of RT product as template. Thermocycler conditions were: 95°C for 5', 20 cycles of 95°C for 30'', 61.7°C for 30'', and 72°C for 1', followed by 72°C for 7' and a final hold at 4°C .

5' RACE

mRNA from uninfected, early and late infected, and lysogen samples were treated with Tobacco Acid Pyrophosphatase for 30min at 37°C . The samples were extracted with phenol/chloroform and precipitated with EtOH and 0.3M sodium acetate. A 5' RNA adaptor (5'-AUAUGCGCGAAUCCUGUAGAACGAACACUAGAAGAAA-3') was ligated with T4 RNA ligase (Fermentas) at 17°C overnight. Again, samples were extracted with phenol/chloroform and precipitated with EtOH and 0.3M sodium acetate. The samples were reverse transcribed using a gene 49 specific oligo (5'-CAGGTACTTATCGCGGTG-3') and Maxima RT (Fermentas). Samples were treated with RNase H for 20min at 37°C . PCR amplification was completed using an adaptor specific oligo (5'-GCGCAATTCCTGTAGA-3') and a gene 49 specific oligo (5'-CAGGTACTTATCGCGGTG-3') at the following conditions: 95°C for 10', 35 cycles of 95°C for 40'', 58°C for 40'', and 72°C for 40'', 72°C for 7' and 4°C hold. PCR products were gel purified and subjected to TOPO-TA cloning (Invitrogen). *E. coli* transformation candidates were verified to contain an insert by PCR and sequenced (Genewiz) using T7 and T3 primers.

DNA replication analyses

M. smegmatis infected with phages at a multiplicity of infection of 1 were incubated at room temperature for 15min; tween-80 added to a final concentration of 0.1%, and cells were shaken for 1min. Infected cells were centrifuged and the supernatant containing unadsorbed

phage was removed. The pellet was washed once with 7H9/ADC/tween 0.1% and once with 7H9/ADC/tween 0.05%. Cells were resuspended in 7H9/ADC, incubated at 37°C, and sample taken every hour for qPCR analysis.

Sample trypsinization and LC-MS/MS analysis

Sample preparation was performed as described in Guttman *et al.* (Guttman *et al.*, 2009). Mycobacteriophage Giles CsCl₂ preparations were concentrated to $>1 \times 10^{11}$ phage/ml using a Speed Vac for in solution protein digestion. RapiGest SF reagent (Waters Corp.) was added to the 0.1 ml phage sample to a final concentration of 0.1% and samples were boiled for 5 min. TCEP (Tris (2-carboxyethyl) phosphine) was added to 1 mM (final concentration) and the samples were incubated at 37°C for 30 min. Subsequently, the samples were carboxymethylated with 0.5 mg/ml of iodoacetamide for 30 min at 37°C followed by neutralization with 2 mM TCEP (final concentration). Proteins samples prepared as above were digested with Promega sequencing grade modified trypsin (trypsin:protein ratio - 1:50) overnight at 37°C. RapiGest was degraded and removed by treating the samples with 250 mM HCl at 37 °C for 1 h followed by centrifugation at $15,800 \times g$ for 30 min at 4°C. The soluble fraction was then added to a new tube and the peptides were extracted and desalted using a 1 ml SepPak C18 solid phase extraction columns (Waters).

Trypsin-digested peptides were analyzed by high pressure liquid chromatography (HPLC) coupled with tandem mass spectroscopy (LC-MS/MS) using nano-spray ionization as described by McCormack *et al.* (McCormack *et al.*, 1997) with these changes. The nanospray ionization experiments were performed using a QSTAR-Elite hybrid mass spectrometer (ABSCIEX) interfaced with nano-scale reversed-phase HPLC (Tempo) using a 10 cm-100 micron ID glass capillary packed with 5 μm C18 Zorbax™ beads (Agilent Technologies, Santa Clara, CA). Peptides were eluted from the C18 column into the mass spectrometer using a linear gradient (5–60%) of ACN (Acetonitrile) at a flow rate of 400 μl/min for 1 h. The buffers used to create the ACN gradient were: Buffer A (98% H₂O, 2% ACN, 0.2% formic acid, and 0.005% TFA) and Buffer B (100% ACN, 0.2% formic acid, and 0.005% TFA). MS/MS data were acquired in a data-dependent manner in which the MS1 data was acquired at *m/z* of 400 to 1800 Da and the MS/MS data was acquired from *m/z* of 50 to 2,000 Da. Finally, the collected data were analyzed using MASCOT® (Matrix Sciences) and Protein Pilot 4.0 (ABSCIEX) for peptide identifications. The LC MS/MS analysis was performed in the UCSD Biomolecular and proteomics Mass Spectrometry Facility by Majid Ghassemian.

mCherry fusions

A promoterless mCherry vector was digested with *NotI* and *KpnI*. Primers were designed to amplify the region of interest from the Giles genome and contained *NotI* and *KpnI* restriction sites. After amplification of the target region, the reaction was cleaned (Qiagen), digested and cloned upstream of the reporter gene mCherry. Clones were verified by sequencing and transformed into wild-type mc²155 and the Giles lysogen. Liquid cultures were analyzed using an Image Reader FLA5000.

Metabolic analysis

Wild-type mc²155 and the Giles lysogen were sent to Biolog, Inc (Hayward, CA) for phenotypic microarray services. A small change in menadione resistance was noted in mc²155(Giles), but was not reproducible upon further testing. Nonetheless, the small ncRNA deletion lysogen was tested for resistance to menadione, but none was observed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Christina Ferreira and Carlos Guerrero for excellent technical assistance and Daniel Russell for assistance with RNA-seq data processing. We also thank Amrita Balachandran and the 2008 Gene Team (University of Pittsburgh) for isolating mixed plaques of mutants $\Delta 29$, $\Delta 51$ and $\Delta 61$, Anna Mansueto for isolating a mixed plaque of mutant $\Delta 68$ and constructing its complementation plasmid, and Chiara Ricci-tam for protein sample preparation. Greg Broussard provided comments on the manuscript. This work was supported by a National Institutes of Health training grant fellowship 5T32AI049820 to RMD, and Grant GM093901 to GFH.

References

- Broussard GW, Oldfield LM, Villanueva VM, Lunt BL, Shine EE, Hatfull GF. Integration-Dependent Bacteriophage Immunity Provides Insights into the Evolution of Genetic Switches. *Mol Cell*. 2012
- Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics*. 2011; 12:395. [PubMed: 21991981]
- Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002; 30:207–210. [PubMed: 11752295]
- Guttman M, Betts GN, Barnes H, Ghasseman M, van der Geer P, Komives EA. Interactions of the NPXY microdomains of the low density lipoprotein receptor-related protein 1. *Proteomics*. 2009; 9:5016–5028. [PubMed: 19771558]
- Hatfull GF. Mycobacteriophages: genes and genomes. *Annu Rev Microbiol*. 2010; 64:331–356. [PubMed: 20528690]
- Hatfull GF. Complete Genome Sequences of 138 Mycobacteriophages. *J Virol*. 2012a; 86:2382–2384. [PubMed: 22282335]
- Hatfull GF. The secret lives of mycobacteriophages. *Adv Virus Res*. 2012b; 82:179–288. [PubMed: 22420855]
- Hatfull GF, Hendrix RW. Bacteriophages and their Genomes. *Current Opinions in Virology*. 2011; 1:298–303.
- Hatfull GF, Sarkis GJ. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. *Mol Microbiol*. 1993; 7:395–405. [PubMed: 8459766]
- Hendrix, RW.; Roberts, JW.; Stahl, FW.; Weisberg, RA. *Lambda II*. Cold Spring Harbor Press; Cold Spring Harbor, NY: 1983.
- Hoess RH, Landy A. Structure of the lambda att sites generated by int-dependent deletions. *Proc Natl Acad Sci U S A*. 1978; 75:5437–5441. [PubMed: 364480]
- Jacobs WR Jr, Barletta RG, Udani R, Chan J, Kalkut G, Sosne G, Kieser T, Sarkis GJ, Hatfull GF, Bloom BR. Rapid assessment of drug susceptibilities of *Mycobacterium tuberculosis* by means of luciferase reporter phages. *Science*. 1993; 260:819–822. [PubMed: 8484123]
- Jacobs-Sera D, Marinelli LJ, Bowman C, Broussard GW, Guerrero Bustamante C, Boyle MM, Petrova ZO, Dedrick RM, Pope WH, Modlin RL, Hendrix RW, Hatfull GFG. Science Education Alliance Phage Hunters Advancing, P. Evolutionary Science Sea-Phages. On the nature of mycobacteriophage diversity and host preference. *Virology*. 2012
- Kreuzer KN, Brister JR. Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the *Virology Journal* series on bacteriophage T4 and its relatives. *Virol J*. 2010; 7:358. [PubMed: 21129203]
- Krupovic M, Prangishvili D, Hendrix RW, Bamford DH. Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiol Mol Biol Rev*. 2011; 75:610–635. [PubMed: 22126996]
- Lee MH, Pascopella L, Jacobs WR Jr, Hatfull GF. Site-specific integration of mycobacteriophage L5: integration-proficient vectors for *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, and bacille Calmette-Guerin. *Proc Natl Acad Sci U S A*. 1991; 88:3111–3115. [PubMed: 1901654]

- Marinelli LJ, Hatfull GF, Piuri M. Recombineering: A powerful tool for modification of bacteriophage genomes. *Bacteriophage*. 2012; 2:5–14. [PubMed: 22666652]
- Marinelli LJ, Piuri M, Swigonova Z, Balachandran A, Oldfield LM, van Kessel JC, Hatfull GF. BRED: a simple and powerful tool for constructing mutant and recombinant bacteriophage genomes. *PLoS ONE*. 2008; 3:e3957. [PubMed: 19088849]
- McCormack AL, Schieltz DM, Goode B, Yang S, Barnes G, Drubin D, Yates JR 3rd. Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal Chem*. 1997; 69:767–776. [PubMed: 9043199]
- Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. *Current opinion in virology*. 2012; 2:63–77. [PubMed: 22440968]
- Morris P, Marinelli LJ, Jacobs-Sera D, Hendrix RW, Hatfull GF. Genomic characterization of mycobacteriophage Giles: evidence for phage acquisition of host DNA by illegitimate recombination. *J Bacteriol*. 2008; 190:2172–2182. [PubMed: 18178732]
- Payne K, Sun Q, Sacchettini J, Hatfull GF. Mycobacteriophage Lysin B is a novel mycolylarabinogalactan esterase. *Mol Microbiol*. 2009; 73:367–381. [PubMed: 19555454]
- Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, Jacobs-Sera D, Falbo J, Gross J, Pannunzio NR, Brucker W, Kumar V, Kandasamy J, Keenan L, Bardarov S, Kriakov J, Lawrence JG, Jacobs WR, Hendrix RW, Hatfull GF. Origins of highly mosaic mycobacteriophage genomes. *Cell*. 2003; 113:171–182. [PubMed: 12705866]
- Pham TT, Jacobs-Sera D, Pedulla ML, Hendrix RW, Hatfull GF. Comparative genomic analysis of mycobacteriophage Tweety: evolutionary insights and construction of compatible site-specific integration vectors for mycobacteria. *Microbiology*. 2007; 153:2711–2723. [PubMed: 17660435]
- Piuri M, Jacobs WR Jr, Hatfull GF. Fluoromycobacteriophages for rapid, specific, and sensitive antibiotic susceptibility testing of *Mycobacterium tuberculosis*. *PLoS ONE*. 2009; 4:e4870. [PubMed: 19300517]
- Pope WH, Ferreira CM, Jacobs-Sera D, Benjamin RC, Davis AJ, DeJong RJ, Elgin SCR, Guilfoile FR, Forsyth MH, Harris AD, Harvey SE, Hughes LE, Hynes PM, Jackson AS, Jalal MD, MacMurray EA, Manley CM, McDonough MJ, Mosier JL, Osterbann LJ, Rabinowitz HS, Rhyan CN, Russell DA, Saha MS, Shaffer CD, Simon SE, Sims EF, Tovar IG, Weisser EG, Wertz JT, Weston-Hafer KA, Williamson KE, Zhang B, Cresawn SG, Jain P, Piuri M, Jacobs WR Jr, Hendrix RW, Hatfull GF. Cluster K Mycobacteriophages: Insights into the Evolutionary Origins of Mycobacteriophage TM4. *PLoS ONE*. 2011a; 6:e26750. [PubMed: 22053209]
- Pope WH, Jacobs-Sera D, Russell DA, Peebles CL, Al-Atrache Z, Alcoser TA, Alexander LM, Alfano MB, Alford ST, Amy NE, Anderson MD, Anderson AG, Ang AAS, Ares M Jr, Barber AJ, Barker LP, Barrett JM, Barshop WD, Bauerle CM, Bayles IM, Belfield KL, Best AA, Borjon A Jr, Bowman CA, Boyer CA, Bradley KW, Bradley VA, Broadway LN, Budwal K, Busby KN, Campbell IW, Campbell AM, Carey A, Caruso SM, Chew RD, Cockburn CL, Cohen LB, Corajod JM, Cresawn SG, Davis KR, Deng L, Denver DR, Dixon BR, Ekram S, Elgin SCR, Engelsen AE, English BEV, Erb ML, Estrada C, Filliger LZ, Findley AM, Forbes L, Forsyth MH, Fox TM, Fritz MJ, Garcia R, George ZD, Georges AE, Gissendanner CR, Goff S, Goldstein R, Gordon KC, Green RD, Guerra SL, Guiney-Olsen KR, Guiza BG, Haghghat L, Hagopian GV, Harmon CJ, Harmson JS, Hartzog GA, Harvey SE, He S, He KJ, Healy KE, Higinbotham ER, Hildebrandt EN, Ho JH, Hogan GM, Hohenstein VG, Holz NA, Huang VJ, Hufford EL, Hynes PM, Jackson AS, Jansen EC, Jarvik J, Jasinto PG, Jordan TC, Kasza T, Katelyn MA, Kelsey JS, Kerrigan LA, Khaw D, Kim J, Knutter JZ, Ko CC, Larkin GV, Laroche JR, Latif A, et al. Expanding the Diversity of Mycobacteriophages: Insights into Genome Architecture and Evolution. *PLoS ONE*. 2011b; 6:e16329. [PubMed: 21298013]
- Ptashne M, Jeffrey A, Johnson AD, Maurer R, Meyer BJ, Pabo CO, Roberts TM, Sauer RT. How the lambda repressor and cro work. *Cell*. 1980; 19:1–11. [PubMed: 6444544]
- Rybniiker J, Nowag A, van Gumpel E, Nissen N, Robinson N, Plum G, Hartmann P. Insights into the function of the WhiB-like protein of mycobacteriophage TM4--a transcriptional inhibitor of WhiB2. *Mol Microbiol*. 2010; 77:642–657. [PubMed: 20545868]
- Smith MC, Burns RN, Wilson SE, Gregory MA. The complete genome sequence of the *Streptomyces* temperate phage phiC31: evolutionary relationships to other viruses. *Nucleic Acids Res*. 1999; 27:2145–2155. [PubMed: 10219087]

- Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 2005; 33:W244–248. [PubMed: 15980461]
- van Kessel JC, Hatfull GF. Recombineering in *Mycobacterium tuberculosis*. *Nature Methods.* 2007; 4:147–152. [PubMed: 17179933]
- van Kessel JC, Hatfull GF. Efficient point mutagenesis in mycobacteria using single-stranded DNA recombineering: characterization of antimycobacterial drug targets. *Mol Microbiol.* 2008; 67:1094–1107. [PubMed: 18221264]
- Xu J, Hendrix RW, Duda RL. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Mol Cell.* 2004; 16:11–21. [PubMed: 15469818]

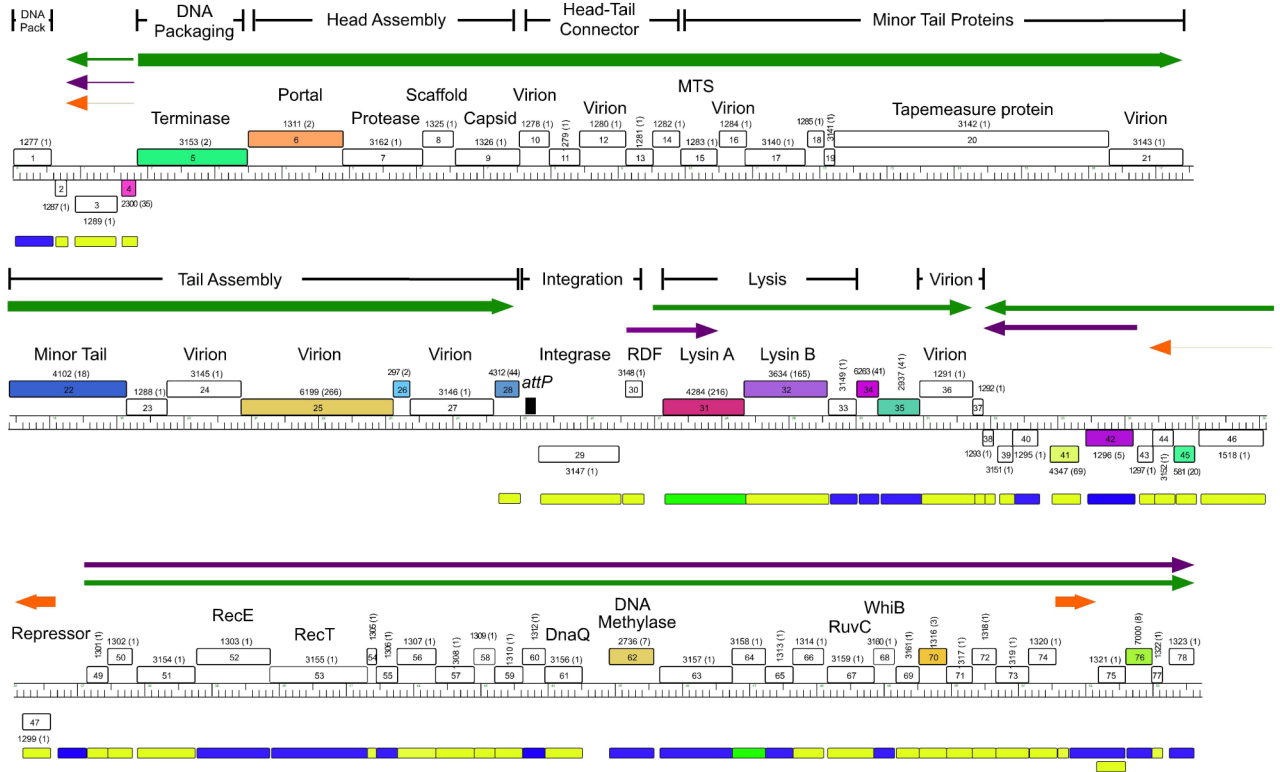


Fig. 1. Mycobacteriophage Giles gene essentiality

A map of the mycobacteriophage Giles shows genes as boxes above or below the genome ruler to indicate those that are transcribed in the rightwards and leftwards direction, respectively. The pham family designations are shown above each gene with the number of pham family members in parentheses; genes are colored according to pham designation with orphans (those with no close mycobacteriophage relatives) shown in white. Boxes below the genome indicate whether the gene is non-essential for lytic growth (yellow), likely essential (blue), or essential (green). Arrows indicate genes expressed in lysogeny (red), early (green) or late (purple) lytic growth, with line thickness reflecting transcription strength, summarized from data shown in Figure 6. The map was generated using Phamerator (Cresawn *et al.*, 2011) with the database Mycobacteriophage_220.

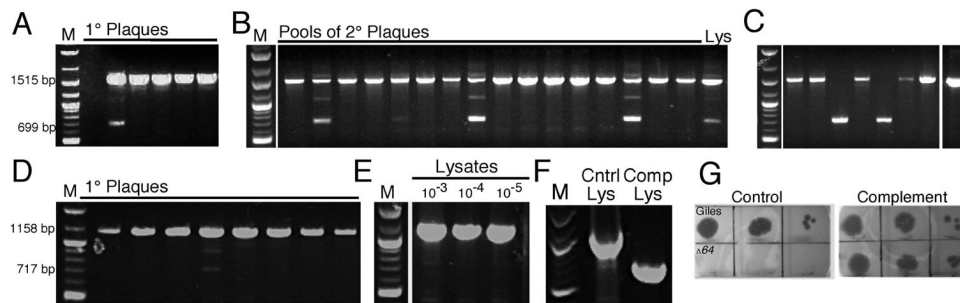


Fig. 2. Examples of Non-essential and Essential Gene Deletions

A. A 200 bp dsDNA substrate introducing a 816 bp deletion in Giles *51* was co-electroporated with Giles DNA into recombineering cells, and primary plaques were tested by flanking primer PCR. The wild-type band is 1515 bp and the mutant band is 699 bp.

B. The primary mixed plaque was plated, but the initial screening of secondary plaques did not reveal any mutants; however, the mutant band was visible from a plate lysate (data not shown). Therefore, secondary plaques were pooled (in groups of 5–10 plaques) and examined by flanking primer PCR, revealing four that contained the mutant allele.

C. Pools containing the deletion mutant were re-plated, and individual plaques were examined by flanking primer PCR. From one pool, two of eight plaques screened were mutant, indicating that gp*51* is not required for plaque formation. However, the low frequency at which the mutant was recovered may suggest that the mutant particles are somewhat impaired relative to wild-type Giles.

D. A 200 bp dsDNA substrate introducing a 450 bp deletion in Giles *64* was co-electroporated with Giles DNA into recombineering cells, and primary plaques were tested using flanking-primer PCR. The wild-type band is 1158 bp and the mutant band is 717 bp.

E. A mixed plaque was re-plated and flanking primer PCR did not reveal any mutants in the secondary plaques (data not shown) and therefore plate lysates were analyzed. The mutant product was not detected, suggesting that *64* is essential.

F. The mixed primary plaque was re-plated on either *M. smegmatis* (Control) or a recombinant strain expressing Giles gp*64* (Complement), and lysates were harvested from plates containing ~2000 plaques. Screening by flanking PCR shows that the mutant is only propagated in the complementing strain.

G. The *64* mutant was isolated from the complementing strain, and serial dilutions of mutant and wild-type phage were spotted onto lawns seeded with either control cells or the complementation strain. The mutant is only able to form plaques on the complementing strain, suggesting the gene is essential.

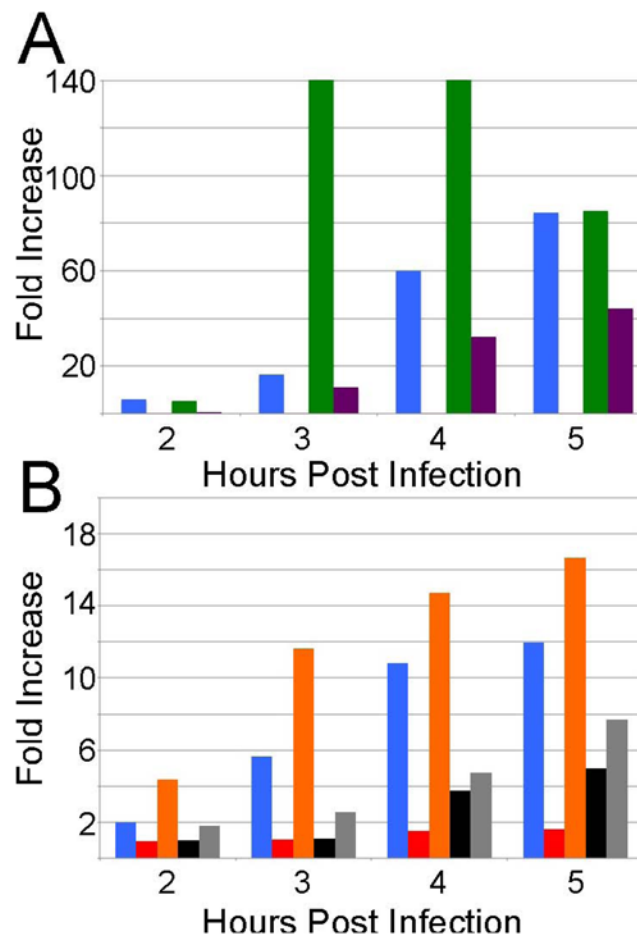


Fig. 3. Giles gp64 and gp50 are required for DNA replication

Following infection, the amount of phage DNA present at different timepoints after infection was determined by qPCR.

A. Infection of mc^2155 with wild-type Giles (blue) or $\Delta 64$ (red), and $mc^2155pRMD1$ infected with wild-type Giles (green), or $\Delta 64$ (purple).

B. Infection with wild type Giles (blue), $\Delta 50$ (red), $\Delta 58$ (orange), $\Delta 67$ (black), and $mc^2155pLAM83$ (expressing gp67) infected with $\Delta 67$ (grey).

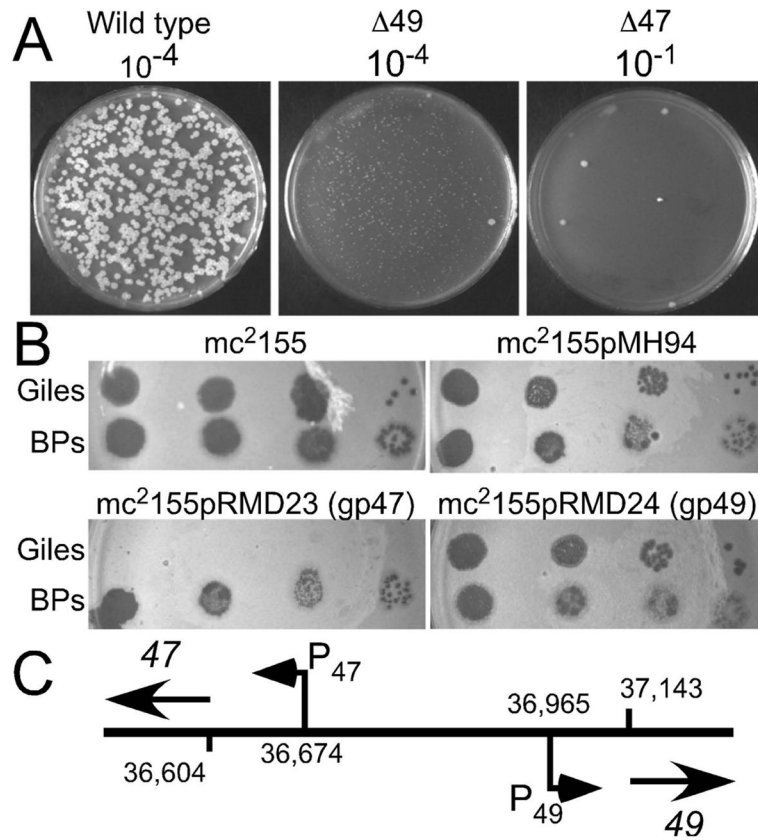


Fig. 4. Giles gp47 is the Repressor

A. Frequencies of lysogeny were determined by plating 100 μ l of dilutions of *M. smegmatis* (as indicated) on agar plates seeded with wild-type Giles, $\Delta 47$ or $\Delta 49$ phages. Colonies represent lysogens which are formed with wild-type Giles at a frequency of $\sim 7\%$, and with both $\Delta 49$ and $\Delta 47$ at $<0.01\%$.

B. Infection of *M. smegmatis* lawns with phage Giles. A recombinant strain expressing Giles gp47 ($mc^2155pRMD23$) confers immunity to superinfection by phage Giles, but strains expressing Giles gp49 ($mc^2155pRMD24$), or containing the pMH94 vector, do not. The unrelated control phage, BPs, plates at equivalent efficiencies on all strains.

C. Schematic representation of the Giles 47–49 intergenic region and the location of transcription start sites mapped by RNAseq and 5'–RACE. The Giles genome coordinates for the predicted start codons of genes 47, 49, and the transcription start sites for the leftwards (P₄₇) and rightwards (P₄₉) promoters are shown.

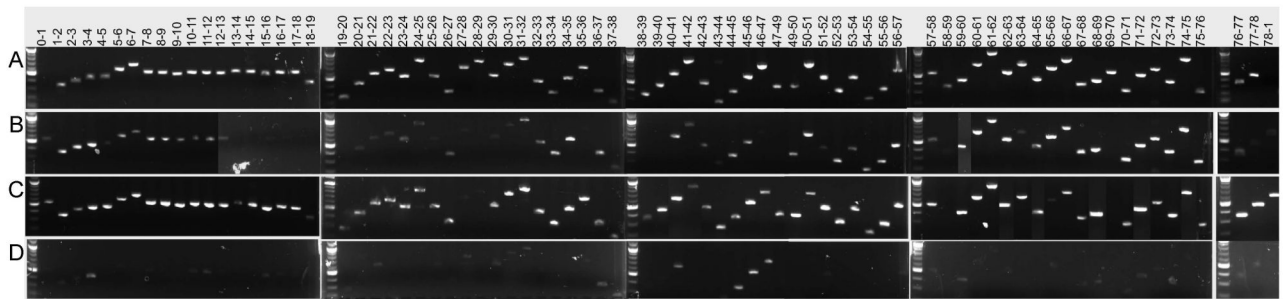


Fig. 5. Expression of gene boundary regions in Giles lytic growth and lysogeny

RT-PCR reactions were performed using primer pairs that traverse Giles gene boundaries (as indicated), and control Giles DNA (A), RNA isolated from an early Giles infection (B), RNA isolated from a late Giles infection (C), and RNA isolated from a Giles lysogen (D). DNA markers are indicated (M).

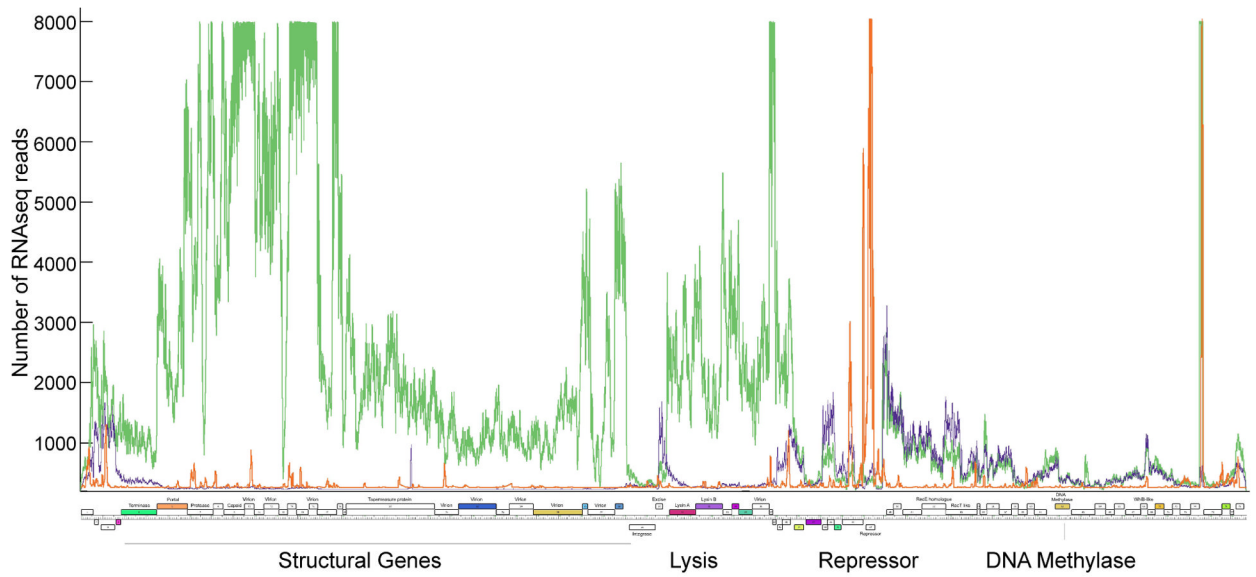


Fig. 6. Transcription of the Giles genome determined by RNAseq

RNA was isolated during early Giles infection (purple), late Giles infection (green), and from a Giles lysogen (orange), and sequenced using Illumina high throughput sequencing. The number of times each nucleotide was sequenced is plotted for each Giles genome coordinate. A map of the Giles genome with several functions noted is shown below.

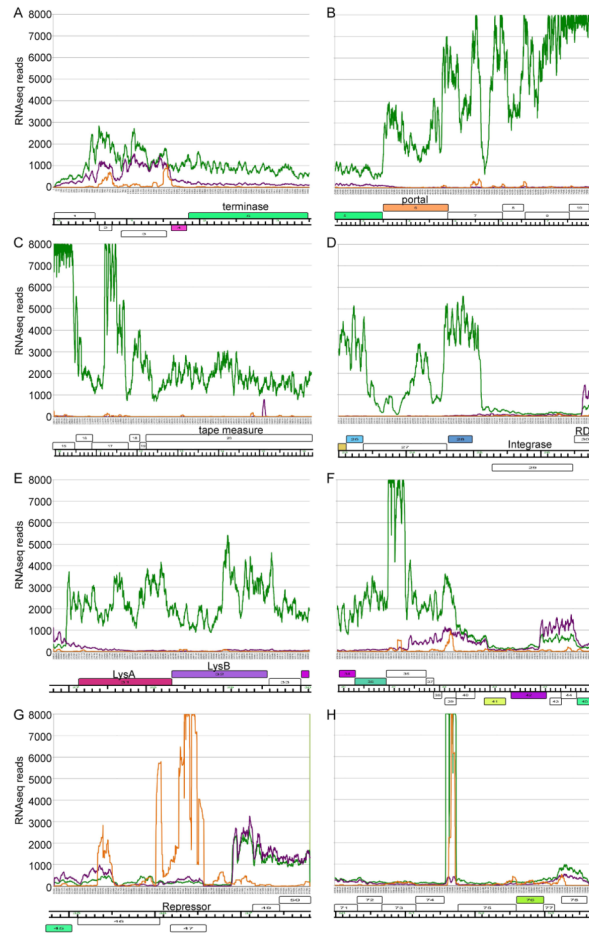


Fig. 7. Transcription patterns at specific loci in the Giles genome determined by RNAseq
The data are the same as though shown in Fig. 6 but with greater resolution at each of eight regions within the Giles genomes: the extreme left end (A), the capsid and capsid assembly genes (B), tail genes (C), integration region (D), lysis cassette (E), convergence of the leftwards and rightwards transcription units at genes 37-38 (F), the repressor (G), and the small non-coding RNA between genes 74 and 75 (H). RNAseq hits per nucleotide generated with RNA from early lytic transcripts (purple), late lytic transcripts (green), and a lysogen (orange) are shown.

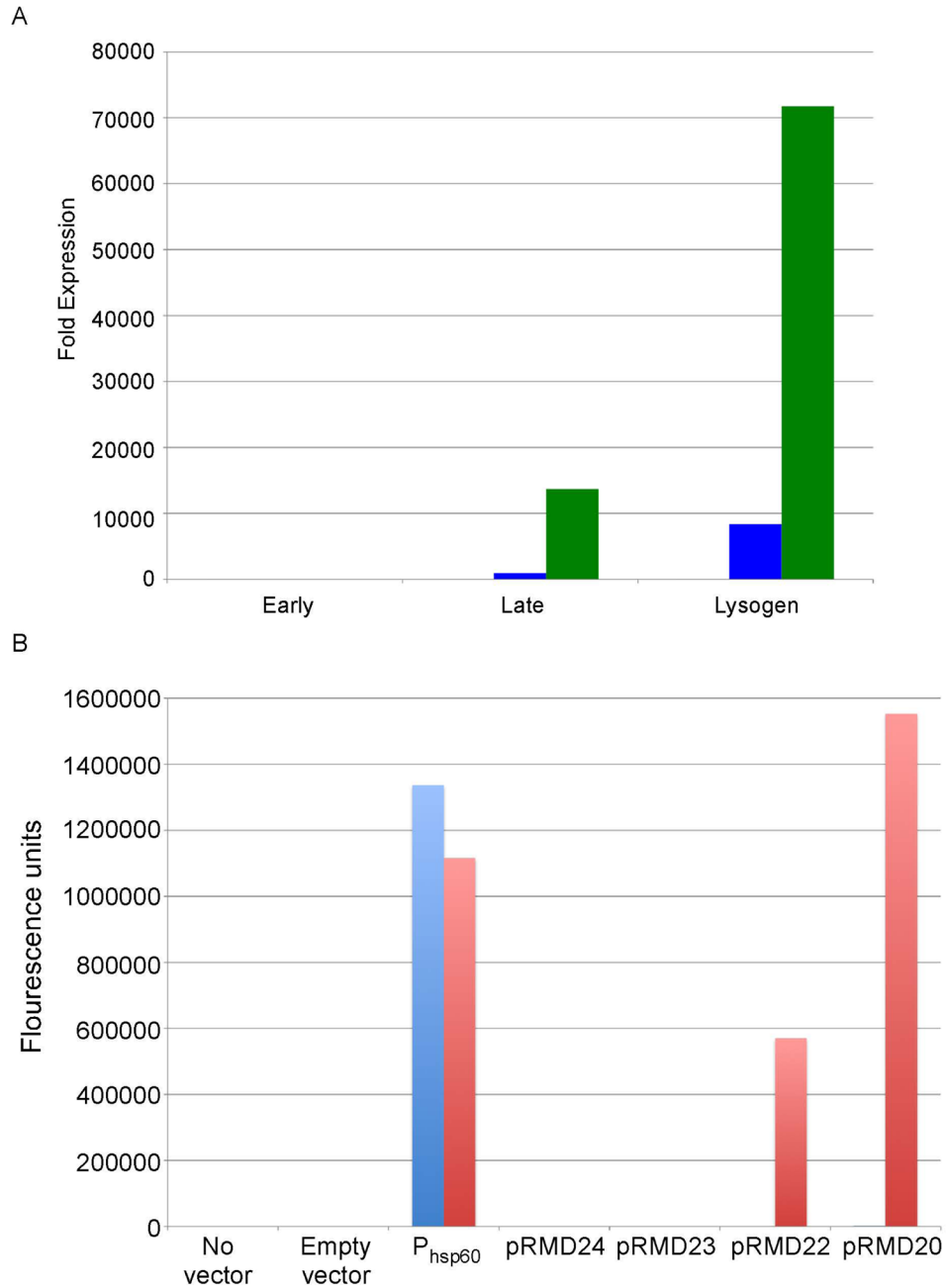


Fig. 8. Expression of a small non-coding RNA encoded between genes 74 and 75

A. The strand coding for the RNA signal between genes 74 and 75 shown in Figure 6H was determined by strand-specific qRT-PCR, using primers reporting rightwards (blue) or leftwards (red) transcription, or random hexamers (green). The y-axis (fold expression) indicates the fold change in mRNA expression compared to wildtype.

B. Fusion of the DNA segment upstream and adjacent to the 74–75 intergenic region encoding the ncRNA to an mCherry reporter gene indicates promoter activity in a lysogen. The vector containing mCherry controlled by the P_{hsp60} promoter (pLMO87) was used as a positive control for both the wild-type mc^2155 and the Giles lysogen. The vector containing mCherry controlled by the P_{rep} promoter (pRMD20) was used as a positive control for the

lysogen. Plasmid pRMD22 contains 56bp (51,565-51,620) of the intergenic region cloned upstream of mCherry, pRMD23 contains 131bp (51,565-51,695) of the intergenic region cloned upstream of mCherry and pRMD24 contains 200bp (51,565-51,746) of the intergenic region cloned upstream of mCherry. Only pRMD22 showed promoter activity, and only in the Giles lysogen. Wild-type (blue), lysogen (red).

Table 1

Giles virion proteins determined by mass spectrometry

Gene	Function	% Coverage	# Peptide Hits (>95%conf)
6	Portal	58.9%	29
7	Clp-protease	27.5%	4
8	scaffold	94.0%	67
9	capsid	96.2%	186
10	H-T Connector	88.4%	41
11	H-T Connector	76.2%	16
12	H-T Connector	75.9%	23
13	H-T Connector	61.8%	5
14	H-T Connector	35.8%	4
15	Major Tail Subunit	80.0%	15
16	Tail	95.6%	37
20	TMP	55.7%	45
21	Minor Tail	60.6%	21
22	Minor Tail	50.1%	23
24	Minor Tail	85.9%	37
25	Gly-rich	31.8%	13
27	Gly-rich	58.1%	12
28	Minor Tail	78.2%	13
36	Virion protein	74.5%	22
37	Virion protein	23.5%	1 ^{<i>l</i>}

^{*l*}This is a longer peptide with a unique sequence that is well represented by y and b ions in the MS/MS spectra. It is identified with a >99% confidence level

Table 2

Giles genes essential for lytic and lysogenic growth

Gene	Coordinates ¹	Ess ²	Prim ³	Sec ⁴	Fec (%) ⁵	Lyso (%) ⁶
1	14-571	Yes	1/14	0/16 ⁹	-	-
2	630-800	No	1/16	1/16	68	86
3	930-1553	No	1/16	4/12	220	45
4	1617-1826	No	1/16	16/16	179	73
28	24633-24989	No	12/16	3/16	241	6
29	25285-26478	No	3/59	3/20	56	27
30	26505-26822	No	16/34	1/45	243	39
31	27131-28339	Yes	9/103	N/A	-	-
32	28339-29574	No	-	-	-	-
33	29596-30006	Yes	3/24 ⁷	N/A	-	-
34	30016-30336	Yes	10/25 ⁷	N/A	-	-
35	30329-30949	Yes	17/18 ⁷	N/A ⁹	-	-
36	30952-31740	No	5/16	3/16	76	47
37	31742-31894	No	1/16	4/16	106	133
38	31891-32046	No	2/17	1/16	104	82
39	32112-32336	No	1/4	4/10	161	46
40	32333-32707	Yes	3/28 ⁷	0/18 ⁹	-	-
41	32891-33316	No	2/5	1/7 ⁸	13	31
42	33425-34126	Yes	1/17	0/135	-	-
43	34195-34422	No	15/16	5/16	177	76
44	34419-34724	No	1/18	6/12	177	66
45	34736-35041	No	3/18	6/16	110	121
46	35107-36063	No	2/18	2/10	271	41
47	36188-36604	No	4/16	1/16	165	0
49	37143-37457	No	1/11	1/15	17	0.0002
50	37454-37819	No	3/36	1/17	0.04	21
51	37891-38754	No	1/20	2/8	20	100

Gene	Coordinates ^I	Ess ²	Prim ³	Sec ⁴	Fec (%) ⁵	Lys ⁶ (%)
52	38780–39871	Yes	1/16	N/A ⁹	-	-
53	39868–41316	Yes	3/14	0/34 ⁹	-	-
54	41313–41453	No	4/32	0/39 ⁹	91	100
55	41450–41767	Yes	12/16	0/31 ⁹	-	-
56	41760–42338	No	1/2	2/16	100	63
57	42335–42907	No	5/17	1/6	131	69
58	42904–43215	No	2/18	1/5 ⁸	7	100
59	43212–43631	No	5/18	1/30	68	128
60	43624–43962	Yes	7/13	0/34 ⁹	-	-
61	43959–44513	Yes	2/17	2/16	-	-
62	44916–45578	Yes	1/39	N/A ⁹	-	-
63	45667–46746	Yes	2/18	0/34 ⁹	-	-
64	46743–47240	Yes	2/32	0/115	-	-
65	47237–47650	Yes	2/18	0/120	-	-
66	47647–48102	No	7/16	3/11	204	62
67	48164–48856	No	2/17	0/15 ⁹	-	68
68	48849–49157	Yes	1/18	1/28 ⁹	-	-
69	49185–49526	No	4/18	1/16	6	74
70	49523–49936	No	4/20	1/55	96	45
71	49933–50319	No	6/24	3/32 ⁸	54	96
72	50316–50672	No	5/18	6/14	17	93
73–74	50669–51564	No	2/6	1/11	41	333
75	51765–52598	Yes	4/45	N/A ⁹	-	-
75	52251–52598	No	2/16	3/14	-	-
76	52603–52992	Yes	6/34 ⁷	0/18 ⁹	-	-
77	52989–53147	No	4/16	4/16	274	46
78	53248–53613	Yes	1/16	0/32	-	-

^I Coordinates of deleted segment

Dedrick et al.

Page 26

- ² Essential gene (yes/no)
- ³ Number of positive primary plaques/total number of plaques
- ⁴ Number of positive secondary plaques/total number of plaques
- ⁵ % fecundity relative to wild-type Giles (100%)
- ⁶ % lysogeny relative to wild-type Giles (100%)
- ⁷ Isolated using tag-specific PCR (DADA-PCR)
- ⁸ Mutant found in pool of plaques (>50 plaques checked)
- ⁹ Lysate dilutions were also checked for the mutant
N/A; not available