HHS Public Access

Author manuscript

Microbiol Spectr. Author manuscript; available in PMC 2018 December 06.

Published in final edited form as:

Microbiol Spectr. 2018 October ; 6(5): . doi:10.1128/microbiolspec.GPP3-0026-2018.

Mycobacteriophages

Graham F. Hatfull

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, gfh@pitt.edu

Abstract

Mycobacteriophages are viruses that infect mycobacterial hosts. A large number of mycobacteriophages have been isolated and genomically characterized, providing insights into viral diversity and evolution, as well as fueling development of tools for mycobacterial genetics. Mycobacteriophages have intimate relationships with their hosts, and provide insights into the genetics and physiology of the mycobacteria, and tools for potential clinical applications such as drug development, diagnosis, vaccines, and potentially therapy.

INTRODUCTION

Because of the importance of mycobacterial diseases such as tuberculosis (TB) and leprosy, mycobacteriophages have been long-studied, with the first ones being isolated in the 1940's (1–3). At that time it was recognized that bacteriophages display particular profiles of specificity for their bacterial hosts, nearly always distinguishing between bacteria in different genera, and sometimes distinguishing between strains of the same bacterial species (4–7). Mycobacteriophages thus presented a plausible means of identifying (i.e. typing) clinical isolates by scoring for phage sensitivity profiles, with the prospects of obtaining data considerably faster than standard protocols that required extensive growth of the host. *Mycobacterium tuberculosis* – the causative agent of human TB – has a remarkably slow growth rate (24-hour doubling time), so phage typing can speed up diagnosis by several weeks. This general concept of exploiting the relative rapid propagation of mycobacteriophages has been a common theme in their subsequent development (8–10).

Although the important role of *M. tuberculosis* in human health was known from the early studies by Koch, our knowledge of the organism was limited for many years because of the inability to genetically transform it with the uptake of DNA (11). An important breakthrough in the late 1980's occurred when it was demonstrated that phage DNA could be taken up by mycobacterial spheroplasts, followed by conversion into infectious units and plaque formation (12). The approach illustrated three additional aspects of working with the mycobacteria and their phages. First, Jacobs and colleagues used *Mycobacterium smegmatis* as a surrogate host, because of its much faster growth rate (~3 hour doubling time), coupled with mycobacteriophages which had been shown to infect both *M. smegmatis* and *M. tuberculosis*. Secondly, the construction of shuttle phasmids – that replicate as a large plasmid in *Escherichia coli* and as a phage in mycobacteria – facilitated the introduction of foreign DNA, as the shuttle phasmids could be manipulated readily in *E. coli*, and then used as vectors in *M. smegmatis* and *M. tuberculosis*. Shuttle phasmids have found extensive uses for introduction of reporter genes, transposons, and allelic exchange substrates into a variety

of mycobacterial hosts (11). Lastly, shuttle phasmids built with temperate phages enable stable introduction of gene cassettes that can be used to evaluate expression and activities of foreign genes such as those for antibiotic resistance (13).

Mycobacteriophages entered the genomics era in the early 1990's (14), opening up a variety of new tools that can be applied to studying the mycobacteria (15). As sequenced mycobacteriophage genomes emerged over the next twenty-five years, the extent of their very considerable genetic diversity became apparent, culminating in a current collection of over 9,600 individual phages, of which 1,500 have been completely sequenced (16). Because all of these (with the exception of DS6A) infect a single bacterial strain (*M. smegmatis* mc^2155), they represent a unique perspective on viral diversity and bacteriophage evolution. These also encompass more than 150,000 genes, and together provide a tremendous resource for advancing our understanding and control of their mycobacterial hosts.

In this chapter I will review our current understanding of mycobacteriophages, their roles in mycobacterial genetics, and the insights gained from their diversity. A number of other reviews on mycobacteriophages are available (8, 9, 15, 17–26) that provide a variety of perspectives including a historical view of this field. Here, I will focus primarily on recent developments that have not been extensively covered in these prior reviews, the most recent of which was four years ago (27).

GENERAL MYCOBACTERIOPHAGE FEATURES

Current genomic scope

As noted above, the total number of mycobacteriophages isolated is nearing 10,000, and over 1,500 have been completely sequenced. Further details about how these were isolated and characterized are provided below, but for the most part we will constrain the discussion here to the subset of genomes that are available in the public databases, totaling 899 genomes. As should be apparent from the discrepancies in these numbers, at the time of writing there are almost 700 genomes that have been sequenced but are not yet fully annotated and submitted to GenBank. The complete and up-to-date list of sequenced phages and their current status is available at the PhagesDB (http://phagesdb.org) and Phamerator (http://phamerator.org) databases (16). It should be noted that the annotation, quality control review, revision, and GenBank submission represents a current bottleneck in the pipeline, and is relatively time consuming compared to sequence determination. Nonetheless, the dataset of 899 genomes offers a representative view of the overall diversity, which does not change substantially if the larger dataset is considered.

It is also worth noting that there is a growing collection of genomically sequenced phages that infect other bacterial hosts within the phylum Actinobacteria. This includes substantial numbers of phages of *Arthrobacter*, *Corynebacteria*, *Gordonia*, *Microbacterium*, *Propionibacterium*, *Rhodoccocus*, and *Streptomyces* (16). These will not be discussed in detail here but are especially relevant in considering the processes by which bacteriophages evolve.

Phage Discovery and Genomics as a Platform for Advancing Science Education

Isolation of bacteriophages using a simple plaque assay has been performed successfully for the past 100 years (28). However, genomic analyses of these phages has only been available since about the early 1990's (14), and DNA sequencing technology has advanced rapidly since then. The coupling of phage isolation and genomic analysis facilitates defining viral diversity and provides insights into understanding their evolution (29). The finding that there are so many different phage genomes underpins the use of phage isolation and genomics as a powerful platform for integrating our missions in science education with those in phage exploration (30).

The Phage Hunters Integrating Research and Education (PHIRE) program was developed at the University of Pittsburgh in 2002 with the intent of using phage discovery and genomics to introduce high school and undergraduate students to authentic scientific research (31). The PHIRE program provides strong evidence not only that students can readily successfully engage in research, but that it is inclusive and does not require either prerequisite understanding of bacteriophages, or technical skills in microbiology (30, 32). Since its inception, students in the PHIRE program have isolated over 300 individual mycobacteriophages, about one-half of which have been sequenced. However, PHIRE is essentially a local program mostly confined to the Pittsburgh region where our research laboratory is located.

The demonstration that this discovery-rich platform is well-suited to scientific novices provided the rationale for broader dissemination and expansion to a nationwide program organized as a course-based research experience for first-year undergraduate students. In 2008, in collaboration with the Science Education Alliance (SEA) program at the Howard Hughes Medical Institute (HHMI), we developed the Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program (33). SEA-PHAGES is an example of a general model for transforming science education – an Inclusive Research-Education Community (iREC) – in which a program infrastructure supports inclusion of different types of institutions and students from all types of backgrounds (Fig. 1). In its first year (2008) there were 12 participating institutions, but additional schools have joined SEA-PHAGES each year, and currently there are over 120 member schools, with more than 4,500 student researchers each year (33, 34). The SEA-PHAGES experience strongly promotes student engagement in studying science, with project ownership and networking being prominent contributing elements (34). It has also had a major impact on our understanding of bacteriophages, and SEA-PHAGES students and faculty have isolated a total of 9,600 actinobacteriophages of which 1,880 have been sequenced. Of these, 8,800 are mycobacteriophages, of which over 1,200 have been sequenced (16).

In general, we have adopted a non-systematic scheme for naming these newly isolated phages. The reasons are three-fold. First, the genomes are clearly constructed as architectural mosaics, with individual genomes composed of specific assemblages of individual genes (35) (see below). It is thus unclear how to use a systematic naming system that reflects the genomic constitution of the phages. Second, naming a newly discovered phage is interesting and stimulating for PHIRE and SEA-PHAGES students, and provides an

opportunity for them to define their discoveries and their contributions, expanding the sense of project ownership that is important to the impact of these programs (34). Thirdly, from a purely practical perspective, it is much easier to remember names such as Corndog, Rosebush, Patience, Daenerys, RedRock Phayonce, or Snazzy, than some tedious series of meaningless letters and numbers. Science doesn't have to be dull and boring all the time!

Databases

The large sizes of both the SEA-PHAGES program and the growing phage collections require data management that is accomplished with three inter-connected databases with front-end web interfaces. One is available at http://seaphages.org and contains information about the SEA-PHAGES program including institutional participation, contact information, protocols, forums, news events etc. The second is at http://phagesdb.org (16) and contains all of the information associated with individual phages, including genome sequences, GPS location coordinates, Blast functions, and many other resources. It is a place where information on newly isolated phages is deposited, and where data on all of the phages is available.

The third is the Phamerator database at http://phamerator.org. The database was designed as a means of organizing mycobacteriophage-encoded genes into 'phamilies' or sequence-related groups, together with the nucleotide sequence data (36). This information can then be used for comparative genomics, such as displaying sets of genome maps in which the genes are colored according to their phamily membership, and pairwise nucleotide sequence similarity can be displayed. The Phamerator database includes some genome-associated information that is not present in the other databases, and a web browser interface provides facile database usage. It liaises with both phagesdb.org and GenBank to help maintain data consistency, and is also used by other programs such as Starterator and PECAAN. As such, Phamerator plays a central role in the annotation pipeline, from genome sequence to GenBank submission (Fig. 1). Phamerator databases containing different numbers of genomes can be downloaded and 'frozen' such that they can be archived and used for publication preparation.

Mycobacteriophage morphologies

All of the mycobacteriophages isolated to date are tailed phages with dsDNA genomes; no ssDNA or RNA phages have been found. Three families of such *Caudovirales* have been described for other bacterial hosts – *Siphoviridae*, *Myoviridae*, and *Podoviridae* – but only siphoviral and myoviral morphotypes are present among the mycobacteriophages (Fig. 2). Because such a large number have been morphologically examined, it seems unlikely that mycobacteriophages with podoviral morphologies exist. Interestingly, phages with podoviral morphologies have been described for other actinobacterial hosts, including *Arthrobacter* and *Streptomyces* (37, 38), suggesting that the absence of podoviral mycobacteriophages is due to a physical blockage presented by mycobacterial cell walls, rather than being evolutionarily inaccessible.

The mycobacteriophage myoviruses – having contractile tails – are all confined to a single genomic type, with phages grouped within Cluster C (Fig. 2; see below). All of the other

mycobacteriophages are siphoviruses – with long flexible non-contractile tails – and these span considerable genetic diversity (Fig. 2). Most mycobacteriophages have isometric capsids (i.e. like a soccer ball) with sizes varying from a diameter of about 40 nm to about 80 nm, correlating with their genome sizes. However, some have prolate heads with the length:width ratios varying from about 2.5:1 to about 4:1 (Fig. 2). Tail length also varies considerably, ranging from about 135 nm to more than 350 nm.

MYCOBACTERIOPHAGE GENOMIC DIVERSITY

Clusters, Subclusters, and Singletons

When the first few mycobacteriophage genome sequences were determined in the 1990's, it became apparent that they do not simply represent variants of a common sequence type (14, 35, 39–41). As the number of sequenced genomes grew, it became clear that these could be grouped according to their overall nucleotide sequence similarity. These groups are referred to as 'clusters', and in general phages within a cluster share evident DNA sequence similarity to each other that spans more than 50% of their genome lengths; phages in different clusters share little if any DNA sequence similarity (42). When 30 genome sequences were available (42), these could be assembled into six clusters (Clusters A – F); nine phages were not closely related to the others and are referred to as singletons. Within four years, the number of genomes had doubled, and these 60 phages were grouped into nine clusters (Cluster A – I), and five singletons (43). It also became clear that within some clusters there are boundaries justifying a separation into subclusters, and average nucleotide identity (ANI) values reflect these divisions. Thus, for the nine clusters formed with 60 sequenced mycobacteriophages, five could be divided into subclusters. For example, Cluster A contained a total of 13 phages at that time, eight of which are in Subcluster A1, and five are in Subcluster A2 (43). As noted at the time (43), the nature of the relationships within clusters is not uniform, such that the apparent divisions revolve around different ANI values for different clusters. A similar pattern has been reported for phages of the *Enterobacteriaceae* (44).

The 899 mycobacteriophages considered here assemble into a total of 28 clusters together with five singletons (Table 1). These clusters are designated Clusters A – Z and Clusters AB and AC (note that Cluster AA is constituted with genomes that are yet to be submitted to GenBank and thus excluded from the current list). Twelve of these clusters (A, B, C, D, F, G, H, I, K, L, M, P) are currently divided into subclusters (Table 1). The largest of the clusters is Cluster A with 322 phages, and this is split into 18 subclusters (A1 – A14, A16 – A19; note that A15 is made up solely of *Gordonia* phages). The largest of the clusters that is not subdivided is Cluster E, with 52 members (Table 1). These all share substantial nucleotide sequence similarity but are far from identical, and differ with numerous small insertions/deletions/substitutions relative to each other, often constrained to a single gene or a small group of genes. Although it typically requires two or more members to constitute a cluster, three of the clusters in the dataset discussed here (X, Y, AC) contain only a single member, and the other members of these are sequenced but are not yet publically available in GenBank (Table 1).

Cluster assignments simplify the analyses and descriptions of phage genomes, but clusters represent groupings of convenience rather than hard biological boundaries, consistent with them having mosaic architectures (45). As more phage genomes are sequenced, the boundaries between different clusters and subclusters have become more obscure. Thus these divisions likely reflect unequal sampling of phages from an underlying continual spectrum of diversity, of which not all parts are equally prevalent in the biosphere (45). Thus, while the vast majority of mycobacteriophages within a cluster do indeed share nucleotide sequence similarity spanning greater than 50% of the genome lengths, some phages have genomes that fall only just below that, complicating their cluster assignment. This has become especially apparent in the comparative genomic analyses of *Gordonia* phages (46, 47). As a consequence, for those genomes that have partially shared sequences where nucleotide-based assignments are difficult, it is recommended that the encoded genes be first grouped into 'phamilies' using Phamerator, and the proportions of shared genes calculated (http://phagesdb.org). Phages sharing more than 35% of their genes should be grouped into the same cluster, and those with fewer than 35% should be in different clusters (46).

Allocation of cluster designations

When only a relatively modest number of mycobacteriophages were sequenced and the first few clusters were designated, their naming was simple, using ordered alphabetical names (Cluster A, Cluster B, etc). Three events prompted further consideration of this naming system. First, as the numbers of clusters grew, the possibility arose that new clusters would be designated even after Cluster Z was named. How should these be designated? Secondly, as our attention shifted to actinobacterial hosts in other genera, how should these be termed? Thirdly, phages isolated on different host genera might group into the same cluster either frequently, or very rarely. How should this be accommodated?

We decided to implement a scheme in which blocks of letters are set aside for clusters of phages isolated on particular hosts (Table 2). In brief, the alphabetical series is extended beyond Clusters A – Z into two-letter (e.g. AA, AB, AC, etc), and then three-letter names (AAA, AAB, AAC, etc) as needed, and a block assigned to each host. For example, phages isolated on mycobacteria have a reserved block of Clusters A – Z and AA – AJ. Currently only clusters up to AC are actually used, so there are seven unoccupied clusters. The size of the initial block (40 for mycobacteriophages; Table 2) is mostly a guess based on an estimate of how many phages may be isolated for that host, and their predicted diversity. For example, a relatively small block is set aside for *Cutibacterium* (including the formerly named *Propionibacterium acnes*) phages, as the phages of *C. acnes* are very similar to each other and are all in a single cluster (48). If a block of cluster allocations becomes fully used, then additional blocks can be added, even though they are likely to be non-sequential. For example, phages of *Arthrobacter* were initially assigned Clusters AK – AZ, but these were quickly used, and an additional block of FA – FZ was added (Table 2). Similarly, *Gordonia* phages were found to be quite diverse, and the initial allocation of Clusters CQ – CZ was quickly populated. However, this could be easily extended to Clusters DA – DZ, as these had not been previously assigned (Table 2), preserving a series of consecutive letters.

This scheme works in large part because most of the phages isolated on a host of a different genus form new clusters from those previously described. However, we do not expect a strict correlation between genus of the host and phage cluster designation. An excellent example is provided by Cluster A. Cluster A is highly populated (Table 1) with individual phages, by far the majority of which were isolated on *M. smegmatis*. However, some Cluster A phages were isolated on *Gordonia* strains, all of which constitute Subcluster A15; it is noteworthy, however, that in spite of the genomic relationships, Subcluster A15 phages do not infect *M. smegmatis*, and other Cluster A phages that have been tested do not infect *Gordonia terrae*. Likewise, we note that several *Gordonia* phages described previously also infect *Nocardia* strains (49). In general, we predict that the more closely related the bacterial hosts are phylogenetically, the more likely it is that their phages may be grouped into the same clusters.

We have constrained these cluster allocations to phages of hosts in the phylum Actinobacteria. However, this could plausibly be extended to phages of hosts in other phyla using a similar scheme and designations, using a prefix to denote the phylum in circumstances where confusion might arise.

Mycobacteriophage gene phamilies

The central questions in deciphering these mycobacteriophage genomes are how are they architecturally structured and how have they evolved? To simplify the comparative genomic analyses, we can take the total gene set, compare each gene to all other genes pairwise at the amino acid sequence level, and place similar genes into groups that we call ‘phamilies’ (43), often abbreviated to ‘phams’. This is a different strategy to using a Hidden Markov Model, in part in an attempt to accommodate the complexity arising from the substantial number of modular genes. When the dataset was small, phamilies were constructed using BlastP and ClustalW, using cutoff values of 10^{-50} (E-value) and 32.5% (aa identity), respectively (36). These values were set such as to capture distantly-related homologues into the same phamily, but to minimize false-positive inclusion of non-homologues in phamilies (36). As the dataset expanded, the use of BlastP and ClustalW became computationally prohibitive, and were substituted with a kmer-based method for phamily assembly (45). Different phams are identified by numbers (Pham 1, Pham 2, Pham 3, etc), and we attempt to maintain pham designations as additional phages are added to the databases. The program Phamerator uses k-clust to sort genes into phamilies, and displays genome maps that include pham designations, and other genome and gene information (36). The web-based version of Phamerator is available at http://phamerator.org, which uses a default database (Actino_Draft) that contains information on all sequenced phages of Actinobacterial hosts. The Actino_Draft database is constantly updated as new phages are isolated and their genomes sequenced, and as such contains genomes of several different status classes including ‘Draft’ (autoannotated) annotations, ‘Final’ annotations that have been manually reviewed and revised, and ‘Gbk’ who annotation history is unknown.

A Phamerator database (Mycobacteriophage_899) was constructed for the 899 phages discussed here, which contains only those genomes that are available in GenBank. This database contains 96,979 genes, which are assorted into 6,046 phamilies. The average

number of genes per phamily is 16.03, and approximately 25% (1,506) contain only a single gene and are designated as ‘orphams’. The largest phamily has 337 gene members.

MYCOBACTERIOPHAGE GENOME ORGANIZATION AND EXPRESSION

Genome annotation and gene naming

When phage genomes are sequenced within the PHIRE or SEA-PHAGES programs they can be autoannotated simply using Glimmer (50) and GeneMark (51) using the program DNA Master (http://cobamide2.bio.pitt.edu/). Although these programs are quite good at identifying coding regions and predicting translational start sites, they do this incorrectly for about 10% of the genes. To obtain as well-supported an annotation as possible, each of the predicted genes is manually inspected and putative functional assignments are made. The genomes are also inspected for missing genes and for tRNA genes and other features. All annotated genomes are reviewed through a quality control process prior to submission to GenBank (Fig. 1).

To simplify the gene naming process, each open reading frame or RNA gene is assigned a gene name as a number, sequentially starting with ‘1’ at the left end of the genome and proceeding rightwards. It is important to note that gene names (e.g. 1, 2, 3, etc) and their gene products (e.g. gp1, gp2, gp3, etc) have no functional meaning. Thus, for example, both phage Corndog have Rosebush have a gene *20*, each coding for a protein, gp20, but Corndog gp20 and Rosebush gp20 are unrelated at the sequence level, and are not functionally related.

Genome Organization

Although there is great sequence diversity among the mycobacteriophages, there are several common themes arising in their overall organization and genomic architecture. An example is shown in Figure 3. First, there is a common organization of virion structural genes in all of the phages with siphoviral morphologies, which includes all of the clusters and singletons with the exception of Cluster C. The genomes are typically presented such that these structural genes are at the left end of the genome and transcribed rightwards (Fig. 3). The gene order is typically: terminase (small and large subunit genes), portal, capsid maturation protease, scaffold, capsid, tail capping and head-tail connectors, major tail subunit, tail assembly chaperones (typically expressed *via* a programmed translational frameshift), tape measure protein, and minor tail genes (of which there may be 4–10) (Fig. 3).

In some genomes, usually those that are relatively short (40–45 kbp), the 20–25 closely packed virion structure and assembly genes span about 22–23 kbp, occupying ~50% of the whole genome (Fig. 3). In such phages, the terminase small subunit gene is usually very close to the physical left end of the genome (in phage Taheera, the translation start site of the small terminase subunit genes is predicted to be at coordinate 47). In other genomes, the gene order may be conserved, but the virion structure and assembly genes are interspersed with non-structural genes. In phage Omega and its relatives (52), the virion structure and assembly genes span over 35kbp, and the inserted genes include a glycosyl transferase, an O-methyl transferase, and an IS110-like transposon. In phage Marvin and its relatives two of

the minor tail protein genes are atypically located near the right end of the genome, separated by about 20 kbp from the other structural genes (53).

All of the phages code for a lysis system that typically includes three components, a holin, an endolysin (Lysin A) and a serine-esterase, referred to as Lysin B or Lysterase (see Fig. 3). The lysis cassette is located in one of two genomic positions. In most phages it is immediately downstream (to the right) of the virion structural genes (Fig. 3). However, in the Cluster A phages it is positioned to the left of the structural genes, between the leftmost *cos* end, and the terminase genes (14, 54). Within the lysis cassettes, the holin gene is sometimes difficult to assign, because of there being several genes with predicted transmembrane domains near the Lysin A gene; it is plausible that several of them play some role in the regulation of lysis. The endolysin genes are constructed modularly, with individual modules shared between otherwise distantly related phages (55). The Lysin B genes code for esterases that cleave the mycolic acids from the cell wall to promote efficient lysis, and at least in phage Giles the Lysin B is not essential for plaque formation, although it forms smaller plaques with fewer particles (56). Some of the mycobacteriophages appear to lack a Lysin B gene altogether (e.g. Rosebush), and it is unclear if other genes substitute for its function (56).

Over one-half of the clusters of mycobacteriophages appear to be temperate, and contain an integration cassette (e.g. Clusters A, E, F, G, I, J, K, L, M, N, O, P, Q, T, X, Y, Z; Singletons DS6A, Dori, Kumao, MooMoo, and Sparky). Typically, the integration cassette is located near the center of the genome and contains an integrase gene and an *attP* site (see Fig. 3). Both major families of integrases – tyrosine integrases and serine-integrases – are represented, although the serine-integrases are typically found in Clusters A and K genomes. Over a dozen different *attB* sites are predicted to be used for prophage integration (22). Interestingly, some phages – predominantly members of Cluster A – do not have an integration cassette and instead code for a *parABS* partitioning system (57, 58). In lysogeny, these phages replicate extrachromosomally at low copy number, although the replication functions have yet to be determined (58).

The mycobacteriophages have a variable number of genes that are not involved in virion structure and assembly, lysis, or integration, depending on the overall genome size. Thus, the smallest of the mycobacteriophages (e.g. Clusters G, N, T) may have only 25–30 additional genes (Fig. 3), whereas larger genomes (e.g. Cluster J) may have over 150 such genes. A few of these have predicted functions in DNA metabolism including replication and recombination, but the vast majority are relatively small and are of unknown function. In phage Giles, it has been shown that about half of the non-structural genes are dispensable for lytic growth (59). Little is known about the essentiality of any of the predicted DNA replication functions including the Pol I type genes and Pol III alpha genes present in some of the phages. Many mycobacteriophages don’t encode any DNA polymerase, so those that do may need them to synthesize genomes with altered chemistry, or because the host polymerases are degraded or inactivated. However, this has yet to be investigated.

Mycobacteriophage gene expression

A general picture of mycobacteriophage gene expression is emerging, although only a rather small subset of phages has been examined. Lytic gene expression has been described for several Cluster A phages including L5, D29, StarStuff, Kampy, and SWU1 (54, 60, 61), for Giles (Cluster Q) (59), and Fruitloop (Cluster F) (62), and several Cluster N phages (63). In general, there appears to be two major types of transcription that occur, in early and late periods. Early gene expression occurs in the first 30 minutes, with late transcription initiating about 30 minutes after infection and continuing until lysis, typically about 180 mins after infection. These general profiles are in agreement with patterns of phage protein synthesis detected by radiolabeling (14, 40).

Early gene transcription usually encompasses non-structural genes in the right arm, and the virion structure and assembly genes and the lysis cassette are expressed late. In phages such as Fruitloop (Fig. 4), late transcription starts ~5 kbp from the right genome end (in the 113 bp intergenic region between genes *91* and *92*) and proceeds rightwards through *cos* into the structural genes (62) (Fig. 4). Of the 11 right arm genes (*92-102*) at the beginning of the late operon, four of the gene products have predicted functions. One (gp92) is a predicted transcriptional regulator, two (gp99 and gp102) are putative glycosyltransferases, and gp100 is a putative kinase. It is plausible that several of these may be involved in post-replication DNA modification as protection against host defenses. Currently very little is understood about the factors needed to activate late transcription in any of the mycobacteriophages. Presumably they are encoded as early genes but they have not been identified.

Curiously, in the Cluster A phages that have been transcriptionally characterized (L5, D29, StarStuff, RedRock, Kampe, SWU1), a non-coding region near the right genome end is transcribed extremely high very early in lytic growth (54, 58, 60, 61). There may be several small transcripts made, but their functions are not known. However, they appear to be quite toxic to *M. smegmatis*, and even *E. coli*, somewhat complicating their study (54). It was previously shown that host gene expression is shut down during L5 early lytic growth (14), and a role for these early RNA's in host shutdown would be consistent with their severe toxicity.

Lysogenic gene expression has been explored in a few mycobacteriophages, primarily those in Clusters A, F, Q, and N (58, 59, 62, 63). Not unexpectedly, the repressor is typically expressed, and the lytic genes are switched off. However, it is not uncommon for additional genes – typically located near the repressor – to also be expressed (see below). The Cluster A phages are unusual in that they contain a large number of putative repressor binding sites (54, 64–66), spanning the entire genome. These sites are 13–14 bp long, asymmetric, and predominantly present in short intergenic regions. Furthermore, the sites are mostly oriented in one direction relative to the direction of transcription (64). It is proposed that at least one of these sites acts as a true operator site for regulation of the early lytic promoter P_{left} (67), whereas the others – termed 'stoperators' – act to promote transcription termination, primarily to prevent expression of toxic genes during lysogeny (64).

Unusual life cycle regulation

Some of the mycobacteriophages employ an interesting and non-canonical system for the establishment of lysogeny that differs from the well-studied paradigms such as in coliphage lambda. Examples of integration-dependent immunity are found in mycobacteriophages in Clusters G, N, and P, and the best studied is that of phage BPs (Cluster G) (68). However, these systems have specific bioinformatic features that are readily recognizable in a variety of other temperate phages (46).

Perhaps the most striking feature of the integration-dependent immunity systems is that the phage attachment site (*attP*) is located within the coding region of the immunity repressor (Fig. 5). At first, this is seemingly a bizarre organization, because integrase-mediated site-specific recombination between the phage *attP* and host *attB* sites will result in disruption of the repressor gene, even though repressor synthesis is absolutely required for maintenance of lysogeny (68). A rationalization of this emerges from a second common feature, a C-terminal tag in the virally-encoded repressor gene product that promotes degradation, typically *via ssrA*-like proteolysis. Because of this *ssrA*-like tag, the virally-encoded form of the repressor is rapidly turned over, and insufficient repressor is made to maintain lysogeny or confer superinfection immunity (68). Integration results in removal of the C-terminal 30–35 amino acids including the *ssrA*-like tag, such that the prophage-encoded form of the repressor is active for maintaining lysogeny and immunity to superinfection (Fig. 5). These phages thus require integration as a prerequisite for lysogeny, and the efficiency of lysogeny is determined by the level of integration, which in turn is modulated by a tag for proteolysis at the C-terminus of integrase (68).

These features impose constraints on other aspects of immunity regulation and are accompanied by some notable consequences. First, we should note that in all of the systems studied to-date, the integrases are of the tyrosine-recombinase family, and use an *attB* site that overlaps a host tRNA gene (Fig. 5). The common core – a region of sequence identity between *attP* and *attB* is typically 25–45 bp, and contains not only the region at which strand-exchange occurs, but also the sequences corresponding to the 3’ end of the tRNA gene; tRNA gene integrity is thus maintained at the attachment junction (*attL* or *attR*) following integration (Fig. 5). In phage BPs, the common core is 35 bp long and positioned 33 codons from the 3’ end of the viral repressor gene. Following integration, the very first base of the bacterial sequence joined at *attL* changes a cysteine codon (5’-TGT) to a translation terminal codon (5’-TGA), so the prophage-encoded form of the repressor lacks the C-terminal 33 amino acid residues of the virally-encoded form (68).

Canonical tyrosine integration systems such as in phage lambda and mycobacteriophage L5 (69, 70) use complex *attP* sites containing secondary integrase-binding sites flanking the common core (i.e. arm-type Int-binding sites), and thus span up to 250 bp. Not only is there no evidence of similar arm-type binding sites in BPs or other integration-dependent systems, but the integrase proteins lack an N-terminal arm-type binding domain (68). As such, these are simple site-specific recombination systems reminiscent of FLP or Cre. This in turn raises the question as to how directionality (i.e. integration *vs* excision) is regulated, because without the ability to use arm-type binding to form higher-order protein-DNA architectures, there is no obvious role for a recombination directionality factor (RDF) such as Xis (71). It

is likely that directionality is regulated at least in part by the availability of active integrase, because the integrases in these systems also contain an *ssrA*-like C-terminal proteolysis tag. Removing the tag leads to higher rates of excisive recombination even in the absence of additional phage-encoded proteins (68).

The relationship between the *attP* site and the repressor gene also results in an unusual expression pattern (Fig. 6). This is illustrated by RNAseq analysis of an *M. smegmatis* lysogen carrying a prophage the Cluster N phage Panchino, which integrates at the *attB*-4 site overlapping a tRNA-lys gene (22). The truncated prophage form of the repressor (gene *33*) is expressed as expected, but the host tRNA-lys gene is also expressed, generating RNA on both strands within the repressor gene (Fig. 6). These rightwards transcripts terminate before they proceed through to gene *34*, which is a candidate for a *cro* gene, whose expression could result in interference with lysogeny and induction of lytic growth. It is somewhat surprising that the transcripts originating from the tRNA do not act as anti-sense inhibitors of repressor expression. This unusual expression pattern is observed in all the integration-dependent systems that have been examined by RNAseq (63).

Finally, these immunity systems can be used to construct integration vectors for inserting genes into mycobacterial chromosomes at *attB* loci, as has been described for other integration vectors (72–74). However, they transform at very low frequency because of the *ssrA*-like tags on the integrase. Removal of these tags gives higher transformation frequencies, but this may be accompanied by instability of the integrated vector if the integrase gene is still present in the recombinant strains (68).

INSIGHTS INTO PHAGE EVOLUTION

Mosaicism: The hallmark of phage genomes

Comparative genomic analysis of phages clearly shows them to have modular structures, with segments assembled to form distinct genomes (35, 75). This is evident at the nucleotide sequence level when comparing phages grouped within particular clusters or subclusters, where otherwise closely related segments are interspersed with regions that are unrelated or more distantly related. An example illustrated by some genomes grouped in Subcluster A2 is shown in Figure 7. It is strikingly common for the boundaries between shared and non-shared sequences to coincide with gene boundaries. A vast number of discontinuities are identifiable, and rarely are there signs of sequence homology that could have promoted recombination through a homology-dependent mechanism, although such events have been described (54). This has led to the suggestion that the recombination events are likely to predominantly occur without extensive sequence dependence, but with selection for function (76). It is plausible that very small regions of homology (3–5 bp) are favored, and these could be mediated by phage-encoded recombinases, which often have relaxed sequence requirements (77–79).

Genomic mosaicism is also evident when comparing genes at the amino acid sequence level, where distant homologues can be identified (Fig. 8). There are numerous examples in which gene members of a phamily are present in two otherwise unrelated genomes, such that the flanking genes to its left and its right are quite different (Fig. 8). Although sometimes two or

more genes may be co-evolving and are shared between genomes, it is very common to observe single genes as units within the mosaic architecture. This mosaicism is particularly pervasive among the large number of small non-structural genes, most of which are of unknown function. However, it is also evident in some of the virion structure and assembly genes, especially those located between the capsid and major tail subunit genes (Fig. 8).

Mobile elements

Although it is anticipated that mobile genetic elements contribute to mosaicism, transposons, introns and inteins are relatively rare in mycobacteriophage genomes. There is an IS*110*-like element in phage Omega and some other Cluster J phages (52), IS*605*-like elements in phage Llij and several other Subcluster F1 phages – although inserted at several distinct locations – an IS*1096*-like element in the Cluster G phage LouisV14, and an IS*1380*-like element in Guillsminger (Subcluster K5). There are instances of introns (e.g. in Omega and other Cluster J phages)(52), and inteins, especially in terminase large subunit genes (65). However, there are a greater number of genes appearing to code for stand-alone homing endonucleases or HNH proteins (65). These are expected to be mobile, and it is not unusual to find HNH genes sharing high levels of amino acid sequence and sometimes nucleotide sequence similarity located in different genomic contexts, indicative of recent mobility.

The most common mobile elements are small mycobacteriophage mobile elements (MPME) and form two main classes, MPME1 and MPME2, which share about 85% nucleotide identity with each other (80). These were first identified by comparison of genomes in Cluster G, where some pairs of genomes are very closely-related and differ primarily by insertion of the MPME element (80). These comparisons reveal the sequence of the pre-integration site, and thus the consequences of insertion. The MPME elements are small (~440 bp) and have short (11 bp) imperfectly conserved inverted repeats (IR) at each end (IR right, IR-L, and IR left, IR-L), and IR-R is joined to the target DNA. However, the junction at IR-L contains six additional base pairs that are not evidently part of the transposon and are not derived from the target sequence. Neither the origin of these nucleotides nor the mechanism giving rise to their inclusion is known. The MPME elements contain a small opening reading frame coding for a putative 125-residue protein, although it has no motifs commonly found in other transposases or recombinases (80). Many instances of these MPME elements have been identified (over 120 MPME1 and over 35 MPME2) in a variety of genomes (including those in Clusters F, G, I, O, T, and W); within a group they are either completely identical or differ by one nucleotide difference. Thus, even though there has been no experimental demonstration of mobility, there is strong evidence that they are indeed mobile, and their small size and unusual features suggest they warrant further investigation.

Multiple modes of evolution

The pervasive mosaicism of the genomes likely arises because of a relative high level of horizontal genetic exchange (HGE) relative to the mutational clock. However, the rates have not been simple to quantify, in part because of the challenges in determining the number of HGE events and estimating mutation rates that likely differ among different phages, especially those that may deploy their own DNA replication machinery. Pairwise comparisons of NCBI phage reference sequences (whose annotations should be reasonably

reliable), determining both nucleotide distances and shared gene content, sheds some light on these parameters. Phages of all hosts were included (not just mycobacteriophages) (81). Traditional comparison methods are too computationally intensive to measure nucleotide distances, so a kmer-based Mash method was used (82). Although there is not universal agreement that Mash distances reflect nucleotide divergence, a variety of validation strategies supported its application in this analysis (81). Because phage genome diversity is high, most genomes are not related and pairwise comparisons have Mash scores above 0.5 (Fig. 9) and are insufficiently similar to be informative (81). The second parameter is an estimate of gene content similarity (Fig. 9), which was determined by grouping the genes into phams, and then determining what proportion of phams that are shared pairwise between genome pairs (81).

This analysis provides an interesting insight, in that there appears to be two major modes of phage genome evolution, one in which gene content varies broadly proportion to nucleotide distances (Low Gene Content Flux, LGCF) and a second in which gene content appears to vary substantially more over shorter nucleotide distance changes (High Gene Content Flux, HGCF) (81) (Fig. 9A). In general, lytic phages follow the LGCF evolutionary mode, whereas temperate phages distribute between the LGCF and HGCF modes. Moreover, phage genomes fall into one of the two modes based on their genomic type, such that clusters of phages tend to follow one or the other, but not both modes (81). This is well illustrated by three clusters of mycobacteriophages (Fig. 9B, C, D). Cluster B phages are lytic, do not form stable lysogens, and do not contain integrase or repressor genes. These generally distribute throughout the LGCF profile (Fig. 9B). In contrast, Clusters F and K are both temperate (Fig. 9C, D), stable lysogens can be recovered, and repressors and integrases have been characterized (73, 83). Cluster F genomes clearly distribute in the HGCF mode (Fig. 9C), whereas Cluster K phages lie in the LGCF mode (Fig. 9D). The drivers of these evolutionary modes are unclear, but it is plausible that the HGCF temperate phages strongly participate in the acquisition of genes that benefit their host when expressed in lysogeny, and may contribute directly to pathogenesis (81).

Evolution and host range

The great genetic diversity of the mycobacteriophages with many types sharing little or no DNA sequence information is unexpected, given that they all infect the same host bacterial strain and therefore should be in genetic communication with each other. A simple model to explain this is that phage host preferences are quite mutable, and there is a low barrier to switching from one host to another (84). Moreover, there must be strong pressure to do so, as phages must mutate to survive viral attack, and phages must co-evolve to access hosts for replication. We thus envisage pathways by which phages migrate across bacterial host landscapes, such that phages pursuing different paths have different access to the large common gene pool (Fig. 10).

Several lines of evidence support this model. First, there are several examples demonstrating that mycobacteriophages can expand their host ranges at reasonably high frequencies (e.g. 10^{-5}), including acquiring the ability to infect *M. tuberculosis*, or a different strain of *M. smegmatis* (84). In each case, this requires just a single base pair change resulting in single

amino acid substitutions in phage tail proteins. Second, characterization of Cluster U phage Patience identified features consistent with the hypothesis that it acquired the ability to infect mycobacteriophage hosts relatively recently (85). Patience has an unusually low G+C content (50.3%), substantially different than most other mycobacteriophages (Table 1) or its *M. smegmatis* host (67.3%). Perhaps not surprisingly, it also has a radically different codon usage profile than *M. smegmatis*, including preferred use of some codons that are only rarely used in the host. A seemingly obvious route to adapting to this situation would be for Patience to acquire tRNA genes that would recognize the rare codons, but this has not happened, and Patience has just a single tRNA gene (tRNA-gln). Instead, the genome appears to be in the process of adapting to its new host, such that the most highly expressed genes are more quickly adapting to match the coding profiles of host genes (85).

DRIVERS OF MYCOBACTERIAL AND MYCOBACTERIOPHAGE DYNAMICS

Prophage-mediated defense against viral attack

Because bacteria are under constant attack from viral infection there is a strong selection for survival and acquisition of systems for viral defense. These include the well-studied restriction-modification and CRISPR-Cas systems, but there are numerous examples of defense by abortive infection (abi) and a plethora of new defense systems have been identified (86, 87). Some mycobacterial strains – including *M. tuberculosis* – have CRISPR-Cas-like arrays, although their functionality in defense and spacer acquisition is unclear (88); mycobacterial restriction-modifications have also been described (89–92). However, *M. smegmatis* mc^2155 which has been used extensively for phage isolation appears to be free of both restriction and CRISPR-Cas systems (93).

Mycobacteriophages have been useful for demonstrating that phage defense systems can also be expressed from the prophages of temperate phages (63). This phenomenon has also been described for *Pseudomonas* phages and is likely to be widespread (94). Although it has been known for some time that phages and plasmids can carrying restriction and abi systems (95, 96), the number and variety of prophage-encoded systems is quite impressive. These have best been characterized in the mycobacteriophages within Cluster N, but are predicted to be present in many other temperate phages and other phages of actinobacterial hosts (63). It is simple to envisage how such defense systems would play key roles in microbial dynamics, as lysogeny is common, and prophage acquisition and recruitment of critical defense systems can occur at relatively high frequency in bacterial populations (97).

The presence of prophage-mediated defense systems is suggested from the expression patterns of lysogenic strains, as indicated by RNAseq analysis. The expression pattern of part of the Panchino genome is shown in Figure 6, which clearly illustrates that the repressor is not the only phage gene being expressed during lysogeny. Up to four additional genes are expressed (28 – 31), and the expression of gene *28* is several times greater than the repressor (Fig. 6). Bioinformatic analysis strongly predicts the gp28 is a multi-subunit restriction system, containing hsdR, hsdM, and hsdS-like domains (63). Genes *30* and *31* are expressed at more modest levels and are of unknown function, although gp31 is a predicted membrane protein. Although prophage-expressed genes could mediate a multitude of physiological

responses in the general category of lysogenic conversion, the restriction system is strongly implicated in viral defense.

The availability of the large collection of diverse phages that infect *M. smegmatis* mc^{2}155 facilitates the characterization of such defense systems. For example, determination of the plating efficiencies of about 100 phages shows that about one-third of them have a reduction of 10^{-5}-fold or greater plating efficiency on the Panchino lysogen relative to the parent strain. When Panchino *28* alone is expressed from a plasmid, the plating efficiencies of the Panchino lysogen are largely reproduced (63). The Cluster N phages are homoimmune and show strong superinfection immunity to each other (Fig. 11A), and we note that Panchino gp28 does not alter infection of Panchino itself or other Cluster N phages; thus this defense system is heterotypic with regard to its specificity for the attacking phages. Presumably, genes *29*, *30*, and *31* could be involved in viral defense too, but if a phage(s) targeted by these is not present in the set that was tested, it could be readily overlooked.

The Panchino *28 – 31* genes are centrally located in the genome between the immunity cassette and the lysis genes (63). Even though the genomes in Cluster N are closely-related to each other, this central region is highly variable, and many of these genes are expressed from the prophages (63). Approximately 3–8 genes are expressed from each of 11 Cluster N genomes, corresponding to over 25 different phamilies, and the determination of plating efficiencies shows that all of the 11 prophages defend against at least one other phage (63). In the case of phage Charlie, the specificity is quite impressive, and the Charlie prophage defends against only one of the tested phages, Che9c. A Charlie derivative in which gene *32* has been deleted using BRED engineering (98) loses its defense against Che9c, and a recombinant plasmid expressing gp32 recapitulates the defense (63). Charlie gp32 is a small (121-residue) protein that is predicted to be membrane localized, and could plausibly act by exclusion, preventing Che9c DNA injection. Consistent with this is the observation that Charlie gp32 not only inhibits Che9c plaque formation, but also interferes with Che9c lysogenization. It is plausible that Charlie gp32 would defend against other phages that use similar host proteins for DNA entry, but that these are rare in our collection.

Several of the Cluster N prophages defend against phage Tweety (Cluster F), including Phrann, and three closely related phages Xerxes, Carcharodon, and MichelleMyBell (Fig. 11) (63). For this second group, a gene responsible was suggested from comparative genomic analysis, as Pipsqueaks is a closely related phage, but has a small deletion in the region corresponding to MichelleMyBell gene *29* (and its relatives), and Pipsqueaks does not defend against Tweety. Deletion of genes *29 and 30* from MichelleMyBell abrogates Tweety defense, and a plasmid carrying genes *29* and *30* reproduces the defense. When using extrachromosomal vectors, it was not possible to clone MichelleMyBell gene *29* without *30*, suggesting that high levels of gp29 are toxic, and that gp30 alleviates the toxicity, akin to the behaviors of toxin-antitoxin (TA) systems. MichelleMyBell gp30 is predicted to be membrane localized, but the biochemical function of gp30 is unclear.

Genetic analysis of phage Phrann showed that Phrann *29* and *30* are required for defense against Tweety (note that the similarity of gene names is coincidental and does not indicate homology) (Fig. 11). These share some of the behaviors of the MichelleMyBell defense

genes in that Phrann gp29 is toxic in the absence of Phrann gp30, but they are not related at the sequence level. Phrann gp29 is strongly predicted to be a (p)ppGpp synthetase, with similarity to RelA family proteins (63). This poses a conundrum in that (p)ppGpp is used by bacteria as an alarmone to promote arrest of cell growth, and yet it is expressed well in a Phrann lysogen (63). However, the lysogen grows normally, and it is probable that gp30 interacts with gp29 and maintains it in an inactive state, consistent with its anti-toxic behavior. This model is attractive as the Phrann gp29/30 proteins would then be poised to be activated upon infection by an attacking phage (Fig. 11). Because the defense is specific to only a subset of the phages tested (including TM4 and Gaia, in addition to Tweety), these phages presumably code for a function that dissociates Phrann gp29 and gp30, resulting in activation of gp29, synthesis of (p)ppGpp, growth arrest, and inhibition of phage replication (63). Interestingly, Phrann gp29/30 defends against Tweety lytic growth but does not interfere with Tweety lysogenization, consistent with a post-DNA injection sensing mechanism. It is unclear what genes or sequences identify Tweety, TM4, and Gaia for targeting by this defense system, but early lytic genes are implicated (see below) (63).

MichelleMyBell gp29/30 likely acts similarly to Phrann, in spite of the lack of sequence similarity between the products (Fig. 11B). In an unusual set of relationships, the C-terminal half of MichelleMyBell gp29 is closely related (92% amino acid identity) to gp29 encoded by phage Squirty (Cluster F). However, the N-terminal half of Squirty gp29 is 100% identical to Phrann gp29 (Fig. 11B). Thus, although Phrann gp29 and MichelleMyBell gp29 are not related to each other, it is plausible that MichelleMyBell is also involved in alarmone synthesis (63).

Mycobacteriophage counter-defense

Phages often carry counter defense systems such as anti-restriction and anti-CRISPR genes (99, 100), so it should not be surprising that mycobacteriophages have systems to counter prophage-mediated defenses. A particularly intriguing system is encoded by Tweety, which was revealed by isolating defense escape mutants (DEMs) that overcome either Phrann-mediated or MichelleMyBell-mediated defense (63). Many of these DEMs have genomic changes in gene *54*, a peculiar gene coding for a protein with multiple copies of a tetrapeptide repeat (Fig. 12). The DEMs typically have changes in the numbers of repeats, usually fewer, but sometimes more (63). DEMs isolated for their escape of Phrann defense remain subject to MichelleMyBell defense, and vice versa (Christian Gauthier and GFH, unpublished observations), suggesting that gp54 is being ‘tuned’ to specifically counter the different defenses. The hypothesis that these mutants are gain-of-function derivatives rather than loss-of-activation mutants is illustrated by the observation that gene *54* can be deleted and Tweety is still subject to both Phrann and MichelleMyBell defenses (63).

Tweety gp54 plays no known role other than in its counter-defense capacity, but homologues are present in about 40 other Cluster F phages, all with different numbers of tetrapeptide repeats. The repeats are composed of the sequence Alanine-Alanine-XX, where XX is either WS, GY, QS, GS or WY (Fig. 12). The three-dimensional structure of this type of repeat is not known, but because the DEMs have variations in the numbers of repeats, it is plausible that it forms an extended structure, positioning the unique N- and C-terminal domains in

configurations that can bind to gp29 and gp30 and prevent their dissociation. The mechanism by which expansion and contraction of the repeat number happens is also not known, but it is noteworthy that in addition to changes in repeat number, many of the mutants also have changes in the sequence of the repeat itself (Fig. 12). However, the changes in the repeat are always in the XX motif and always substituted by one of the tetrapeptide repeats found elsewhere. For example, DEM201, a mutant that overcomes Phrann defense has lost 11 copies of the repeat, but six of the individual repeats are altered (Fig. 12). Not only are the six repeats not contiguous, but only one pair of the substituted sequences is present elsewhere, suggesting that five distinct mutational events created that derivative, in addition to the deletion of 11 copies. Because the repeated nature is also reflected as a 12 bp nucleotide sequence repeat, these variations could arise from replication errors and slippage, although it would appear to occur at a relatively high frequency. Whether the repeat substitutions play any functional role in counter-defense is unclear.

Finally, we note that although Tweety *54* homologues are predominantly found in Cluster F phages, there is a related gene in the Subcluster A1 phage, Nerujay (Nerujay gene *74*) (Fig. 13). The unique N- and C-terminal regions are closely related to Tweety gp54 (90% and 80% aa identity, respectively) but Nerujay contains only 10 copies of the repeat and only two tetrapeptide sequences are present, AAWS and AAGS. Nerujay is the only Cluster A phage to carry such a gene, and it is notably absent from phage DD5, which is closely related in this region at the DNA sequence level (Fig. 13). A simple evolutionary explanation is that Nerujay acquired gene *74* relatively recently as insertion of a 796 bp fragment, perhaps derived from a Cluster F phage similar to Ibhubesi, whose homologue (54) is identical in length to Nerujay *74* and shares 90% nucleotide sequence identity. Nerujay gp74 may also be active in counter-defense, but identifying which defense system it counters represents a substantial challenge.

Exclusion during lytic growth

Many viruses have been reported to exclude superinfection of other viruses during either lytic or lysogenic growth. Some well-studied examples include the T4 *imm* gene and P22 *sieA*, both of which are small membrane proteins that confer homotypic exclusion (101, 102). Coliphage HK97 also codes for a small membrane protein that is prophage expressed and prevents HK97 superinfection (103). It is not unusual for some of the small genes of unknown function in mycobacteriophages to be predicted membrane proteins that could perform such functions (e.g. there are eight predicted membrane proteins in the right arm of phage Nerujay), but a specific role for these in exclusion has not been shown. However, a WhiB-like regulator encoded in phage TM4 is implicated in homotypic exclusion, plausibly through regulation of host genes (104). In a recent example, superinfection exclusion was shown to be mediated by interference of a host protein used for efficient infection by some other phages (Fig. 14) (62).

This superinfection exclusion system was discovered through use of a screen for mycobacteriophage genes that are toxic to *M. smegmatis* when overexpressed (62). Approximately 5–10% of genes are expected to have this toxic property (105–108), and we identified gene *52* of phage Fruitloop (Cluster F) as being highly toxic when overexpressed.

Fruitloop gp52 interacts directly with the essential host protein Wag31 (DivIVA) and inactivates it, resulting in cell death (Fig. 14). Fruitloop *52* is expressed early in lytic growth and is not essential for plaque formation, and its toxicity is unlikely to be relevant for phage growth. Rather, it was reasoned that some mycobacteriophages depend on Wag31 for efficient infection, as it is polar localized and present both cytoplasmically and extracytoplasmically (109, 110). A search for phages inhibited by Fruitloop gp52 expression identified two Subcluster B2 phages, Hedgerow and Rosebush that are affected, and presumably use Wag31 for efficient infection (62). Hedgerow and Rosebush are not related to Fruitloop at the nucleotide sequence level, and it is not evident that they have followed similar evolutionary paths to Fruitloop. It is plausible that mycobacteriophages code for many similar systems, targeting different aspects of mycobacteria that other phages exploit to gain entry.

Compatibility and Incompatibility

Most of the temperate mycobacteriophages can be grouped into homoimmune classes, and have closely-related repressors. The large majority of these carry integration systems, targeting a variety of *attB* sites. In general, if two phages are heteroimmune and use different *attB* sites for integration, they are fully compatible and there is no obvious impediment to the formation of double lysogens. This is not unexpected, and the sequences of large numbers of bacterial genomes shows that they commonly carry two or more prophages. If phages are heteroimmune but use the same *attB* site, then they may compete with other for stable lysogeny.

A subset of temperate mycobacteriophages do not have integration systems, but instead encode *parABS* systems similar to those used by plasmids to promote segregation into daughter cells at division (57, 58). These phages – all of which group in Cluster A but span 10 different subclusters – replicate extrachromosomally at low copy number, and as such are subject to incompatibility mediated potentially by both prophage replication system, and the partitioning systems themselves. None of the replication systems have yet to be described, but there is good evidence of incompatibility of among phages with similar *parABS* systems (58). For example, a double lysogen of phages RedRock and Alma – both of which have *parABS* systems – can be constructed, but not stably maintained without loss of one of the prophages (58). In contrast, a double lysogen of RedRock and Bxb1 – which has a canonical integration system – stably maintains both prophages.

Repressor theft

The immunity repressors have been well studied in the Cluster A phages (107) although there is substantial sequence variation reflecting a variety of heteroimmune groups. Unexpectedly, a variety of other phages in other clusters carry closely related copies of some of these repressors (65). These related gene copies are found in lytic phages that do not form lysogens, including some Cluster C phages, and also in temperate phages (e.g. Cluster F and Cluster J), where they are present in addition to their canonical repressors. For example, Fruitloop (Cluster F1) has a canonical repressor (gene *44*) which confers immunity to itself and some other Cluster F1 phages, and the stolen repressor (37) which shares 97% amino acid identity with the Bxb1 repressor. Fruitloop gene *37* is expressed in a lysogen, although

at a somewhat lower level than the canonical repressor (62), but confers immunity to at least some Cluster A1 phages such as SkiPole and Pinto (unpublished observations). Because the Cluster C phages do not form lysogens, the repressor may block superinfection of Cluster A1-like phages during lytic growth.

EXPLOITATION OF MYCOBACTERIOPHAGES

Using recombinant phages for efficient delivery

Bacteriophages have the inherent ability to efficiently inject their DNA into host bacterial cells. Mycobacteria do not undergo natural transformation at high frequency, and electroporation is effective but only a subset of cells take up DNA. In contrast, infection with mycobacteriophages at a multiplicity of infection greater than three generally results in DNA injection into every cell in the culture. This property was exploited early on in the development of mycobacterial genetic systems (12), and has been further developed for the delivery of transposons, reporter genes, and allelic exchange substrates (111–115). Phages are useful tools for the delivery of reporter genes such as luciferase or florescent proteins in rapid diagnostic assays for tuberculosis drug susceptibility in the clinic (116–119), but are also useful for understanding mycobacterial physiology. For example, dual reporter phages have been exploited to identify and analyze persister cells in *M. tuberculosis* (120).

Mycobacteriophage tools for mycobacterial genetics

Mycobacteriophages also offer a rich reservoir of genes and cassettes for use in mycobacterial genetics. One application that has found widespread use is the development of integration-proficient vectors that transform mycobacteria by stable chromosomal integration using phage integrases. Integration vectors have been constructed from a variety of different phages, including Adephagia, Brujita, L5, Bxb1, Tweety, Giles, Omega, and Ms6 (52, 68, 72–74, 83, 121, 122). These typically transform efficiently and the transformants are quite stable, except if newly introduced genes are deleterious to growth and the integrase gene is retained in the recombinants. Most of these vectors work in both fast- and slow-growing strains, as the *attB* sites – especially those used by the tyrosine integrases – tend to be well conserved.

A second tool with broad applicability is a mycobacterial-specific recombineering system utilizing mycobacteriophage-encoded recombination systems (123). The most widely used system is that derived from mycobacteriophage Che9c, which has been used for both dsDNA and ssDNA recombineering (78, 79, 124). For dsDNA recombineering both the exonuclease and DNA pairing genes (Che9c *60* and *61*, respectively) are required, but for ssDNA recombineering only *61* is needed. The recombineering systems work in both *M. smegmatis* and *M. tuberculosis*, and recombination is sufficiently efficient that it can be used to introduce point mutations into the genome of *M. tuberculosis* (125). Many newly identified recombination systems are present in the multitude of mycobacteriophages sequenced since the Che9c system was developed, and some of these may be worth investigating for improved ease of use or efficiency.

The recombineering system has been especially useful for engineering the phages themselves (98, 123, 126). When phage DNA and a mutagenic DNA substrate are co-electroporated into recombineering-proficient cells, recombination is efficient, and following recovery, individual plaques can be screened by PCR for the presence of the mutant allele. Typically ~10% of progeny are mixed and contain both wild-type and mutant alleles, and a homogenous mutant derivative can be identified by PCR of plaques after re-plating. Point mutations can be introduced relatively easily, provided a PCR scheme is used that can distinguish between the mutant and wild-type alleles (68). The main limitation of the Bacteriophage Recombineering of Electroporated DNA (BRED) approach is whether plaques can be recovered following transfection in *M. smegmatis* (98). Many phages do this quite efficiently, and sufficient plaques can be readily recovered for screening. However, some phages transfect at very low efficiency or not at all, and these tend to be phages with larger genomes. Whether genome size *per se* is the principle factor, or whether some of these phages use encapsidated phage proteins for infection is unclear.

Mycobacteriophages have also provided non-antibiotic based selectable markers for mycobacteria, taking advantage of immunity repressor genes that confer immunity to superinfection, and lytic derivatives of the temperate phages that efficiently kill mycobacteria. The most well developed system is based on the Cluster K phage Adephagia, which infects both *M. smegmatis* and *M. tuberculosis* (127). Plasmids carrying the Adephagia repressor (gene *43*) can be introduced by electroporation and selected in the presence of an Adephagia derivative from which gene *43* has been deleted. The transformants will be carrying an integrated copy of the Adephagia genome, with the additional advantage that the plasmid cannot be lost without lytic induction and death, thus ensuring plasmid stability. If the presence of the prophage is undesirable, a second phage derivative can be used in which both the repressor (43) and the integrase gene (41) have been deleted (127).

Mycobacterial plasmids based of the *oriM* origin of replication derived from plasmid pAL5000 (128–131) are somewhat unstable in the absence of selection (72), especially if recombinant genes they are carrying are not well-tolerated. Addition of the phage-encoded *parABS* systems to these plasmids adds considerable stability to their maintenance (58).

Therapeutic potential

The possibility of using mycobacteriophages for the treatment of mycobacterial infections has been postulated for some time, although little progress has been made (10). Although tuberculosis remains a major public health concern in many parts of the world, effective treatment could be achieved with current antibiotics, a robust health care infrastructure, and monitoring. However, the emergence of multiply and extensively drug resistance tuberculosis infections demands new approaches, and phage applications may warrant attention. It would seem unlikely that mycobacteriophages could achieve sterilization from an *M. tuberculosis* infection and a long-term cure. But some gains might be achieved from combinations with antibiotics that could extend their effectiveness, lower required dosages, reduce the prevalence of resistance, provide relief in late stages of infection, or help prevent transmission. An alternative approach would be to use phages prophylactically, to protect

co-workers or family members from infection when in close contact with an infected patient (10).

Therapeutic utility is likely to require a cocktail of phages so as to minimize the incidence of phage resistance, just as antibiotics must be used in combination for TB treatment. However, the key property of phages in such a cocktail is that resistance to one phage does not confer resistance to the other phages. Unfortunately, very little is known about either the specific receptors used for mycobacteriophage infection, or the mechanisms of resistance (132), and this is an area that needs further investigation. Furthermore, relatively few of the phages isolated on *M. smegmatis* infect *M. tuberculosis*, and they generally group in a small number of cluster or subclusters (84). As such, they may use a rather limited set of receptors, and cross-resistance may limit the numbers that could be combined together into a suitable cocktail. Moreover, early phage typing studies suggested differences in phage sensitivities among clinical isolates of *M. tuberculosis*, and this needs to be revisited now that we have a deeper understanding of the genetic diversity of *M. tuberculosis* strains (133).

Infections of non-tuberculosis mycobacteria (NTM) may also represent possible targets for phage therapy. Attractive targets are *Mycobacterium abscessus* infections associated with cystic fibrosis (CF), which are often highly refractory to antibiotic therapies. Relatively little is known about the phage susceptibility profiles of *M. abscessus* clinical isolates, but a substantial proportion of sequenced *M. abscessus* strains carry one or more prophage (134, 135), which could code for viral defense systems. Suitable mycobacteriophage cocktails could conceivably be used to reduce the bacterial load in CF patients, or to prevent infection following lung transplantation.

PERSPECTIVES

Mycobacteriophages have provided a wealth of insights into viral diversity and evolution, and played indispensable roles in the development of mycobacterial genetics. We anticipate that the advances in mycobacteriophage genomics will slow somewhat, in large part because we are nearing saturation in the isolation of phages that infect *M. smegmatis*. However, as the emphasis shifts to phages that infect other mycobacterial strains and other Actinobacteria, new and different types of genomes are likely to be discovered. We anticipate that the coming years will see some progress towards defining the functions of more mycobacteriophage genes. The number of uncharacterized genes is vast, but the methods are available for their genetic, biochemical, and structural characterization, and the limiting factor may be the development of suitable assays. Surrogate approaches such as screening for toxicity (62, 106) and finding interacting partners (136, 137) should be useful strategies. Because many of the phage genes may play roles in microbial dynamics, the availability of such a large collection offers a very substantial resource.

Finally, the mycobacteriophages occupy a unique niche in how they have been isolated and characterized, mostly by early-career undergraduate students and faculty at institutions spread across the United States and the world (34). These students have not only made fantastic contributions to this field, but the fun, curious, and sometimes entertaining names of the phages they discovered will continue to amuse. The number of phages that infect

Actinobacterial hosts likely far exceeds the numbers of students who will isolate and characterize them over the coming years, and I can't wait to see the novelties, surprises, and potential utilities that emerge.

ACKNOWLEDGEMENTS

I would like to thank Bekah Dedrick, Krista Freeman, Becky Garlena, Christian Gauthier, Carlos Guerrero, Debbie Jacobs-Sera, Ching-Chung Ko, Travis Mavrich, Welkin Pope, Dan Russell, and Katie Wetzel for comments on the manuscript and for other help and insights. The Mycobacteriophage_899 database is available at http://phamerator.webfactional.com/databases_Hatfull. I would like to thank the National Institutes of Health and the Howard Hughes Medical Institute for support, and to all the student phage hunters and SEA-PHAGES faculty and staff for their amazing contributions to mycobacteriophageology.

REFERENCES

1. Gardner GM, Weiser RS. 1947 A Bacteriophage for *Mycobacterium smegmatis*. Proc Soc Exp Biol Med 66:205–206. [PubMed: 20270730]
2. Whittaker E 1950 Two bacteriophages for Mycobacterium smegmatis. Can J Public Health 41:431–436. [PubMed: 14778057]
3. Froman S, Will DW, Bogen E. 1954 Bacteriophage active against Mycobacterium tuberculosis I. Isolation and activity. Am J Pub Health 44:1326–1333.
4. Grange JM. 1975 Proceedings: Bacteriophage typing of strains of Mycobacterium tuberculosis isolated in south-east england. J Med Microbiol 8:ix (2).
5. Jones WD, Jr. 1975 Phage typing report of 125 strains of "Mycobacterium tuberculosis". Ann Sclavo 17:599–604. [PubMed: 820285]
6. Snider DE, Jr., Jones WD, Good RC. 1984 The usefulness of phage typing Mycobacterium tuberculosis isolates. Am Rev Respir Dis 130:1095–1099. [PubMed: 6439088]
7. Goode D 1983 Bacteriophage typing of strains of Mycobacterium tuberculosis from Nepal. Tubercle 64:15–21. [PubMed: 6405518]
8. Hatfull GF. 2006 Mycobacteriophages, p 602–620. *In* Calendar R (ed), The Bacteriophages. Oxford University Press, New York, NY.
9. Hatfull GF. 2010 Mycobacteriophages: genes and genomes. Annu Rev Microbiol 64:331–356. [PubMed: 20528690]
10. Hatfull GF. 2014 Mycobacteriophages: windows into tuberculosis. PLoS Pathog 10:e1003953. [PubMed: 24651299]
11. Jacobs WR, Jr. 2014 Gene Transfer in Mycobacterium tuberculosis: Shuttle Phasmids to Enlightenment. Microbiol Spectr 2.
12. Jacobs WR, Jr., Tuckman M, Bloom BR. 1987 Introduction of foreign DNA into mycobacteria using a shuttle phasmid. Nature 327:532–535. [PubMed: 3473289]
13. Snapper SB, Lugosi L, Jekkel A, Melton RE, Kieser T, Bloom BR, Jacobs WR, Jr. 1988 Lysogeny and transformation in mycobacteria: stable expression of foreign genes. Proc Natl Acad Sci U S A 85:6987–6991. [PubMed: 2842799]
14. Hatfull GF, Sarkis GJ. 1993 DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. Mol Microbiol 7:395–405. [PubMed: 8459766]
15. Hatfull GF. 1994 Mycobacteriophage L5: A toolbox for tuberculosis. ASM News 60:255–260.
16. Russell DA, Hatfull GF. 2017 PhagesDB: the actinobacteriophage database. Bioinformatics 33:784–786. [PubMed: 28365761]
17. Mizuguchi Y 1984 Mycobacteriophages, p 641–662. *In* Kubica GP, Wayne LG (ed), The Mycobacteria: A Sourcebook, vol Part A. Marcel Dekker, Inc., New York.
18. Hatfull GF. 2008 Bacteriophage genomics. Curr Opin Microbiol 11:447–453. [PubMed: 18824125]
19. Hatfull GF. 1999 Mycobacteriophages, p 38–58. *In* Ratledge C, Dale J (ed), Mycobacteria: Molecular Biology and Virulence. Chapman and Hall, London.

20. Hatfull GF. 2000 Molecular Genetics of Mycobacteriophages, p 37–54. *In* Hatfull GF, Jacobs WR, Jr. (ed), Molecular Genetics of the Mycobacteria. ASM Press, Washington, DC.
21. Hatfull GF. 2004 Mycobacteriophages and tuberculosis, p 203–218. *In* Eisenach K, Cole ST, Jacobs WR, Jr., McMurray D (ed), Tuberculosis. ASM Press, Washington, DC.
22. Hatfull GF. 2012 The secret lives of mycobacteriophages. Adv Virus Res 82:179–288. [PubMed: 22420855]
23. Hatfull GF, Jacobs WR, Jr. 1994 Mycobacteriophages: Cornerstones of mycobacterial research, p 165–183. *In* Bloom BR (ed), Tuberculosis: Pathogenesis, Protection and Control. ASM, Washington.
24. McNerney R 1999 TB: the return of the phage. A review of fifty years of mycobacteriophage research. Int J Tuberc Lung Dis 3:179–184. [PubMed: 10094316]
25. Jacobs WR, Jr., Snapper SB, Tuckman M, Bloom BR. 1989 Mycobacteriophage vector systems. Rev Infect Dis 11 Suppl 2:S404–410. [PubMed: 2652256]
26. Sarkis GJ, Hatfull GF. 1998 Mycobacteriophages. Methods Mol Biol 101:145–173. [PubMed: 9921476]
27. Hatfull GF. 2014 Molecular Genetics of Mycobacteriophages. Microbiology Spectrum 2:1–36.
28. Rohwer F, Segall AM. 2015 In retrospect: A century of phage lessons. Nature 528:46–48. [PubMed: 26632584]
29. Hatfull GF. 2015 Innovations in Undergraduate Science Education: Going Viral. J Virol 89:8111–8113. [PubMed: 26018168]
30. Hanauer DI, Jacobs-Sera D, Pedulla ML, Cresawn SG, Hendrix RW, Hatfull GF. 2006 Inquiry learning. Teaching scientific inquiry. Science 314:1880–1881. [PubMed: 17185586]
31. Hatfull GF, Racaniello V. 2014 PHIRE and TWiV: Experiences in Bringing Virology to New Audiences. Annual Review of Virology 1:37–53.
32. Hanauer D, Hatfull GF, Jacobs-Sera D. 2009 Active Assessment: Assessing Scientific Inquiry. Springer, New York, NY.
33. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, Dennehy JJ, Denver DR, Dunbar D, Elgin SC, Findley AM, Gissendanner CR, Golebiewska UP, Guild N, Hartzog GA, Grillo WH, Hollowell GP, Hughes LE, Johnson A, King RA, Lewis LO, Li W, Rosenzweig F, Rubin MR, Saha MS, Sandoz J, Shaffer CD, Taylor B, Temple L, Vazquez E, Ware VC, Barker LP, Bradley KW, Jacobs-Sera D, Pope WH, Russell DA, Cresawn SG, Lopatto D, Bailey CP, Hatfull GF. 2014 A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. MBio 5:e01051–01013. [PubMed: 24496795]
34. Hanauer DI, Graham MJ, Sea P, Betancur L, Bobrownicki A, Cresawn SG, Garlena RA, Jacobs-Sera D, Kaufmann N, Pope WH, Russell DA, Jacobs WR, Jr., Sivanathan V, Asai DJ, Hatfull GF. 2017 An inclusive Research Education Community (iREC): Impact of the SEA-PHAGES program on research outcomes and student learning. Proc Natl Acad Sci U S A 114:13531–13536. [PubMed: 29208718]
35. 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. 2003 Origins of highly mosaic mycobacteriophage genomes. Cell 113:171–182. [PubMed: 12705866]
36. Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2011 Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics 12:395. [PubMed: 21991981]
37. Klyczek KK, Bonilla JA, Jacobs-Sera D, Adair TL, Afram P, Allen KG, Archambault ML, Aziz RM, Bagnasco FG, Ball SL, Barrett NA, Benjamin RC, Blasi CJ, Borst K, Braun MA, Broomell H, Brown CB, Brynell ZS, Bue AB, Burke SO, Casazza W, Cautela JA, Chen K, Chimalakonda NS, Chudoff D, Connor JA, Cross TS, Curtis KN, Dahlke JA, Deaton BM, Degroote SJ, DeNigris DM, DeRuff KC, Dolan M, Dunbar D, Egan MS, Evans DR, Fahnestock AK, Farooq A, Finn G, Fratus CR, Gaffney BL, Garlena RA, Garrigan KE, Gibbon BC, Goedde MA, Guerrero Bustamante CA, Harrison M, Hartwell MC, Heckman EL, et al. 2017 Tales of diversity: Genomic and morphological characteristics of forty-six Arthrobacter phages. PLoS One 12:e0180517. [PubMed: 28715480]

38. Klyczek KK, Jacobs-Sera D, Adair TL, Adams SD, Ball SL, Benjamin RC, Bonilla JA, Breitenberger CA, Daniels CJ, Gaffney BL, Harrison M, Hughes LE, King RA, Krukonis GP, Lopez AJ, Monsen-Collar K, Pizzorno MC, Rinehart CA, Program S-P, Program P, Staples AK, Stowe EL, Garlena RA, Russell DA, Cresawn SG, Pope WH, Hatfull GF. 2018 Complete Genome Sequences of 44 Arthrobacter Phages. Genome Announc 6.
39. Ford ME, Sarkis GJ, Belanger AE, Hendrix RW, Hatfull GF. 1998 Genome structure of mycobacteriophage D29: implications for phage evolution. J Mol Biol 279:143–164. [PubMed: 9636706]
40. Ford ME, Stenstrom C, Hendrix RW, Hatfull GF. 1998 Mycobacteriophage TM4: genome structure and gene expression. Tuber Lung Dis 79:63–73. [PubMed: 10645443]
41. Mediavilla J, Jain S, Kriakov J, Ford ME, Duda RL, Jacobs WR, Jr., Hendrix RW, Hatfull GF. 2000 Genome organization and characterization of mycobacteriophage Bxb1. Mol Microbiol 38:955–970. [PubMed: 11123671]
42. Hatfull GF, Pedulla ML, Jacobs-Sera D, Cichon PM, Foley A, Ford ME, Gonda RM, Houtz JM, Hryckowian AJ, Kelchner VA, Namburi S, Pajcini KV, Popovich MG, Schleicher DT, Simanek BZ, Smith AL, Zdanowicz GM, Kumar V, Peebles CL, Jacobs WR, Jr., Lawrence JG, Hendrix RW. 2006 Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. PLoS Genet 2:e92. [PubMed: 16789831]
43. Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, Weber RJ, Patel MC, Germane KL, Edgar RH, Hoyte NN, Bowman CA, Tantoco AT, Paladin EC, Myers MS, Smith AL, Grace MS, Pham TT, O'Brien MB, Vogelsberger AM, Hryckowian AJ, Wynalek JL, Donis-Keller H, Bogel MW, Peebles CL, Cresawn SG, Hendrix RW. 2010 Comparative Genomic Analysis of 60 Mycobacteriophage Genomes: Genome Clustering, Gene Acquisition, and Gene Size. J Mol Biol 397:119–143. [PubMed: 20064525]
44. Grose JH, Casjens SR. 2014 Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. Virology 468–470:421–443.
45. Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG, Jacobs WR, Hendrix RW, Lawrence JG, Hatfull GF, Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science, Phage Hunters Integrating Research and Education, Mycobacterial Genetics Course. 2015 Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. Elife 4:e06416. [PubMed: 25919952]
46. Pope WH, Mavrich TN, Garlena RA, Guerrero-Bustamante CA, Jacobs-Sera D, Montgomery MT, Russell DA, Warner MH, Science Education Alliance-Phage Hunters Advancing G, Evolutionary S, Hatfull GF. 2017 Bacteriophages of Gordonia spp. Display a Spectrum of Diversity and Genetic Relationships. MBio 8.
47. Pope WH, Montgomery MT, Bonilla JA, Dejong R, Garlena RA, Guerrero Bustamante C, Klyczek KK, Russell DA, Wertz JT, Jacobs-Sera D, Hatfull GF. 2017 Complete Genome Sequences of 38 Gordonia sp. Bacteriophages. Genome Announc 5.
48. Marinelli LJ, Fitz-Gibbon S, Hayes C, Bowman C, Inkeles M, Loncaric A, Russell DA, Jacobs-Sera D, Cokus S, Pellegrini M, Kim J, Miller JF, Hatfull GF, Modlin RL. 2012 Propionibacterium acnes Bacteriophages Display Limited Genetic Diversity and Broad Killing Activity against Bacterial Skin Isolates. MBio 3.
49. Dyson ZA, Tucci J, Seviour RJ, Petrovski S. 2015 Lysis to Kill: Evaluation of the Lytic Abilities, and Genomics of Nine Bacteriophages Infective for Gordonia spp. and Their Potential Use in Activated Sludge Foam Biocontrol. PLoS One 10:e0134512. [PubMed: 26241321]
50. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007 Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673–679. [PubMed: 17237039]
51. Borodovsky M, Lomsadze A. 2011 Gene identification in prokaryotic genomes, phages, metagenomes, and EST sequences with GeneMarkS suite. Curr Protoc Bioinformatics Chapter 4:Unit 4.5 1–17.
52. Pope WH, Jacobs-Sera D, Best AA, Broussard GW, Connerly PL, Dedrick RM, Kremer TA, Offner S, Ogiefo AH, Pizzorno MC, Rockenbach K, Russell DA, Stowe EL, Stukey J, Thibault SA, Conway JF, Hendrix RW, Hatfull GF. 2013 Cluster J mycobacteriophages: intron splicing in capsid and tail genes. PLoS One 8:e69273. [PubMed: 23874930]

53. Mageeney C, Pope WH, Harrison M, Moran D, Cross T, Jacobs-Sera D, Hendrix RW, Dunbar D, Hatfull GF. 2012 Mycobacteriophage Marvin: a new singleton phage with an unusual genome organization. J Virol 86:4762–4775. [PubMed: 22357284]
54. Dedrick RM, Mavrich TN, Ng WL, Hatfull GF. 2017 Expression and evolutionary patterns of mycobacteriophage D29 and its temperate close relatives. BMC Microbiol 17:225. [PubMed: 29197343]
55. Payne KM, Hatfull GF. 2012 Mycobacteriophage endolysins: diverse and modular enzymes with multiple catalytic activities. PLoS One 7:e34052. [PubMed: 22470512]
56. Payne K, Sun Q, Sacchettini J, Hatfull GF. 2009 Mycobacteriophage Lysin B is a novel mycolylarabinogalactan esterase. Mol Microbiol 73:367–381. [PubMed: 19555454]
57. Stella EJ, de la Iglesia AI, Morbidoni HR. 2009 Mycobacteriophages as versatile tools for genetic manipulation of mycobacteria and development of simple methods for diagnosis of mycobacterial diseases. Rev Argent Microbiol 41:45–55. [PubMed: 19391526]
58. Dedrick RM, Mavrich TN, Ng WL, Cervantes Reyes JC, Olm MR, Rush RE, Jacobs-Sera D, Russell DA, Hatfull GF. 2016 Function, expression, specificity, diversity and incompatibility of actinobacteriophage parABS systems. Mol Microbiol 101:625–644. [PubMed: 27146086]
59. Dedrick RM, Marinelli LJ, Newton GL, Pogliano K, Pogliano J, Hatfull GF. 2013 Functional requirements for bacteriophage growth: gene essentiality and expression in mycobacteriophage Giles. Mol Microbiol 88:577–589. [PubMed: 23560716]
60. Halleran A, Clamons S, Saha M. 2015 Transcriptomic Characterization of an Infection of Mycobacterium smegmatis by the Cluster A4 Mycobacteriophage Kampy. PLoS One 10:e0141100. [PubMed: 26513661]
61. Fan X, Duan X, Tong Y, Huang Q, Zhou M, Wang H, Zeng L, Young RF, 3rd, Xie J 2016 The Global Reciprocal Reprogramming between Mycobacteriophage SWU1 and Mycobacterium Reveals the Molecular Strategy of Subversion and Promotion of Phage Infection. Front Microbiol 7:41. [PubMed: 26858712]
62. Ko CC, Hatfull GF. 2018 Mycobacteriophage Fruitloop gp52 inactivates Wag31 (DivIVA) to prevent heterotypic superinfection. Mol Microbiol doi:10.1111/mmi.13946.
63. Dedrick RM, Jacobs-Sera D, Bustamante CA, Garlena RA, Mavrich TN, Pope WH, Reyes JC, Russell DA, Adair T, Alvey R, Bonilla JA, Bricker JS, Brown BR, Byrnes D, Cresawn SG, Davis WB, Dickson LA, Edgington NP, Findley AM, Golebiewska U, Grose JH, Hayes CF, Hughes LE, Hutchison KW, Isern S, Johnson AA, Kenna MA, Klyczek KK, Mageeney CM, Michael SF, Molloy SD, Montgomery MT, Neitzel J, Page ST, Pizzorno MC, Poxleitner MK, Rinehart CA, Robinson CJ, Rubin MR, Teyim JN, Vazquez E, Ware VC, Washington J, Hatfull GF. 2017 Prophage-mediated defence against viral attack and viral counter-defence. Nat Microbiol 2:16251. [PubMed: 28067906]
64. Brown KL, Sarkis GJ, Wadsworth C, Hatfull GF. 1997 Transcriptional silencing by the mycobacteriophage L5 repressor. Embo J 16:5914–5921. [PubMed: 9312049]
65. 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, et al. 2011 Expanding the Diversity of Mycobacteriophages: Insights into Genome Architecture and Evolution. PLoS ONE 6:e16329. [PubMed: 21298013]
66. Bandhu A, Ganguly T, Jana B, Mondal R, Sau S. 2010 Regions and residues of an asymmetric operator DNA interacting with the monomeric repressor of temperate mycobacteriophage L1. Biochemistry 49:4235–4243. [PubMed: 20377203]
67. Nesbit CE, Levin ME, Donnelly-Wu MK, Hatfull GF. 1995 Transcriptional regulation of repressor synthesis in mycobacteriophage L5. Mol Microbiol 17:1045–1056. [PubMed: 8594325]
68. Broussard GW, Oldfield LM, Villanueva VM, Lunt BL, Shine EE, Hatfull GF. 2013 Integration-dependent bacteriophage immunity provides insights into the evolution of genetic switches. Mol Cell 49:237–248. [PubMed: 23246436]

69. Landy A 2015 The lambda Integrase Site-specific Recombination Pathway. Microbiol Spectr 3:MDNA3-0051-2014.
70. Peña CE, Lee MH, Pedulla ML, Hatfull GF. 1997 Characterization of the mycobacteriophage L5 attachment site, attP. J Mol Biol 266:76–92. [PubMed: 9054972]
71. Lewis JA, Hatfull GF. 2001 Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. Nucleic Acids Res 29:2205–2216. [PubMed: 11376138]
72. Lee MH, Pascopella L, Jacobs WR, Jr., Hatfull GF. 1991 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 88:3111–3115. [PubMed: 1901654]
73. Pham TT, Jacobs-Sera D, Pedulla ML, Hendrix RW, Hatfull GF. 2007 Comparative genomic analysis of mycobacteriophage Tweety: evolutionary insights and construction of compatible site-specific integration vectors for mycobacteria. Microbiology 153:2711–2723. [PubMed: 17660435]
74. Morris P, Marinelli LJ, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2008 Genomic characterization of mycobacteriophage Giles: evidence for phage acquisition of host DNA by illegitimate recombination. J Bacteriol 190:2172–2182. [PubMed: 18178732]
75. Hatfull GF. 2015 Dark Matter of the Biosphere: the Amazing World of Bacteriophage Diversity. J Virol 89:8107–8110. [PubMed: 26018169]
76. Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF. 1999 Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc Natl Acad Sci U S A 96:2192–2197. [PubMed: 10051617]
77. Martinsohn JT, Radman M, Petit MA. 2008 The lambda red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism. PLoS Genet 4:e1000065. [PubMed: 18451987]
78. van Kessel JC, Hatfull GF. 2007 Recombineering in *Mycobacterium tuberculosis*. Nature Methods 4:147–152. [PubMed: 17179933]
79. van Kessel JC, Hatfull GF. 2008 Efficient point mutagenesis in mycobacteria using single-stranded DNA recombineering: characterization of antimycobacterial drug targets. Mol Microbiol 67:1094–1107. [PubMed: 18221264]
80. Sampson T, Broussard GW, Marinelli LJ, Jacobs-Sera D, Ray M, Ko CC, Russell D, Hendrix RW, Hatfull GF. 2009 Mycobacteriophages BPs, Angel and Halo: comparative genomics reveals a novel class of ultra-small mobile genetic elements. Microbiology 155:2962–2977. [PubMed: 19556295]
81. Mavrich TN, Hatfull GF. 2017 Bacteriophage evolution differs by host, lifestyle and genome. Nat Microbiol 2:17112. [PubMed: 28692019]
82. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016 Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 17:132. [PubMed: 27323842]
83. 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. 2011 Cluster K Mycobacteriophages: Insights into the Evolutionary Origins of Mycobacteriophage TM4. PLoS ONE 6:e26750. [PubMed: 22053209]
84. Jacobs-Sera D, Marinelli LJ, Bowman C, Broussard GW, Guerrero Bustamante C, Boyle MM, Petrova ZO, Dedrick RM, Pope WH, Science Education Alliance Phage Hunters Advancing G, Evolutionary Science Sea-Phages P, Modlin RL, Hendrix RW, Hatfull GF. 2012 On the nature of mycobacteriophage diversity and host preference. Virology 434:187–201. [PubMed: 23084079]
85. Pope WH, Jacobs-Sera D, Russell DA, Rubin DH, Kajee A, Msibi ZN, Larsen MH, Jacobs WR, Jr., Lawrence JG, Hendrix RW, Hatfull GF. 2014 Genomics and proteomics of mycobacteriophage patience, an accidental tourist in the Mycobacterium neighborhood. MBio 5:e02145. [PubMed: 25467442]

86. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R. 2018 Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359.

87. Ofir G, Melamed S, Sberro H, Mukamel Z, Silverman S, Yaakov G, Doron S, Sorek R. 2018 DISARM is a widespread bacterial defence system with broad anti-phage activities. Nat Microbiol 3:90–98. [PubMed: 29085076]

88. He L, Fan X, Xie J. 2012 Comparative genomic structures of Mycobacterium CRISPR-Cas. J Cell Biochem 113:2464–2473. [PubMed: 22396173]

89. Jones WD, Jr., Greenberg J 1977 Host modification and restriction with a mycobacteriophage isolated from a pseudolysogenic Mycobacterium chelonei. J Gen Microbiol 99:389–395. [PubMed: 874454]

90. Shankar S, Tyagi AK. 1992 MhaAI, a novel isoschizomer of PstI from Mycobacterium habana recognizing 5'-CTGCA/G-3'. Nucleic Acids Res 20:2891. [PubMed: 1614878]

91. Shankar S, Tyagi AK. 1992 MfoAI, a novel isoschizomer of HaeIII from Mycobacterium fortuitum recognizing 5'-GG/CC-3'. Nucleic Acids Res 20:2890. [PubMed: 1614877]

92. Shankar S, Tyagi AK. 1993 MchAI and MchAII, two class-II restriction endonucleases from Mycobacterium chelonei. Gene 132:119–123. [PubMed: 8406034]

93. Snapper SB, Melton RE, Mustafa S, Kieser T, Jacobs WR, Jr. 1990 Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. Mol Microbiol 4:1911–1919. [PubMed: 2082148]

94. Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, Maxwell KL. 2016 Prophages mediate defense against phage infection through diverse mechanisms. ISME J 10:2854–2866. [PubMed: 27258950]

95. Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GP. 2009 The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proc Natl Acad Sci U S A 106:894–899. [PubMed: 19124776]

96. Chopin MC, Chopin A, Bidnenko E. 2005 Phage abortive infection in lactococci: variations on a theme. Curr Opin Microbiol 8:473–479. [PubMed: 15979388]

97. Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobian-Guemes AG, Coutinho FH, Dinsdale EA, Felts B, Furby KA, George EE, Green KT, Gregoracci GB, Haas AF, Haggerty JM, Hester ER, Hisakawa N, Kelly LW, Lim YW, Little M, Luque A, McDole-Somera T, McNair K, de Oliveira LS, Quistad SD, Robinett NL, Sala E, Salamon P, Sanchez SE, Sandin S, Silva GG, Smith J, Sullivan C, Thompson C, Vermeij MJ, Youle M, Young C, Zgliczynski B, Brainard R, Edwards RA, Nulton J, Thompson F, Rohwer F. 2016 Lytic to temperate switching of viral communities. Nature 531:466–470. [PubMed: 26982729]

98. Marinelli LJ, Piuri M, Swigonova Z, Balachandran A, Oldfield LM, van Kessel JC, Hatfull GF. 2008 BRED: a simple and powerful tool for constructing mutant and recombinant bacteriophage genomes. PLoS ONE 3:e3957. [PubMed: 19088849]

99. Tock MR, Dryden DT. 2005 The biology of restriction and anti-restriction. Curr Opin Microbiol 8:466–472. [PubMed: 15979932]

100. Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR. 2013 Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature 493:429–432. [PubMed: 23242138]

101. Lu MJ, Henning U. 1989 The immunity (imm) gene of Escherichia coli bacteriophage T4. J Virol 63:3472–3478. [PubMed: 2746737]

102. Susskind MM, Botstein D, Wright A. 1974 Superinfection exclusion by P22 prophage in lysogens of Salmonella typhimurium. III. Failure of superinfecting phage DNA to enter sieA+ lysogens. Virology 62:350–366. [PubMed: 4610992]

103. Cumby N, Edwards AM, Davidson AR, Maxwell KL. 2012 The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. J Bacteriol 194:5012–5019. [PubMed: 22797755]

104. Rybniker J, Nowag A, van Gumpel E, Nissen N, Robinson N, Plum G, Hartmann P. 2010 Insights into the function of the WhiB-like protein of mycobacteriophage TM4--a transcriptional inhibitor of WhiB2. Mol Microbiol 77:642–657. [PubMed: 20545868]

105. Rybniker J, Plum G, Robinson N, Small PL, Hartmann P. 2008 Identification of three cytotoxic early proteins of mycobacteriophage L5 leading to growth inhibition in Mycobacterium smegmatis. Microbiology 154:2304–2314. [PubMed: 18667563]

106. Rybniker J, Krumbach K, van Gumpel E, Plum G, Eggeling L, Hartmann P. 2011 The cytotoxic early protein 77 of mycobacteriophage L5 interacts with MSMEG_3532, an L-serine dehydratase of Mycobacterium smegmatis. J Basic Microbiol 51:515–522. [PubMed: 21656815]

107. Donnelly-Wu MK, Jacobs WR, Jr., Hatfull GF. 1993 Superinfection immunity of mycobacteriophage L5: applications for genetic transformation of mycobacteria. Mol Microbiol 7:407–417. [PubMed: 8459767]

108. Liu J, Dehbi M, Moeck G, Arhin F, Bauda P, Bergeron D, Callejo M, Ferretti V, Ha N, Kwan T, McCarty J, Srikumar R, Williams D, Wu JJ, Gros P, Pelletier J, DuBow M. 2004 Antimicrobial drug discovery through bacteriophage genomics. Nat Biotechnol 22:185–191. [PubMed: 14716317]

109. Meniche X, Otten R, Siegrist MS, Baer CE, Murphy KC, Bertozzi CR, Sassetti CM. 2014 Subpolar addition of new cell wall is directed by DivIVA in mycobacteria. Proc Natl Acad Sci U S A 111:E3243–3251. [PubMed: 25049412]

110. Mukherjee P, Sureka K, Datta P, Hossain T, Barik S, Das KP, Kundu M, Basu J. 2009 Novel role of Wag31 in protection of mycobacteria under oxidative stress. Mol Microbiol 73:103–119. [PubMed: 19496931]

111. Bardarov S, Bardarov S, Jr., Pavelka MS, Jr., Sambandamurthy V, Larsen M, Tufariello J, Chan J, Hatfull G, Jacobs WR, Jr. 2002 Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in Mycobacterium tuberculosis, M. bovis BCG and M. smegmatis. Microbiology 148:3007–3017. [PubMed: 12368434]

112. Bardarov S, Jr., Dou H, Eisenach K, Banaiee N, Ya S, Chan J, Jacobs WR, Jr., Riska PF. 2003 Detection and drug-susceptibility testing of M. tuberculosis from sputum samples using luciferase reporter phage: comparison with the Mycobacteria Growth Indicator Tube (MGIT) system. Diagn Microbiol Infect Dis 45:53–61. [PubMed: 12573551]

113. Bardarov S, Kriakov J, Carriere C, Yu S, Vaamonde C, McAdam RA, Bloom BR, Hatfull GF, Jacobs WR, Jr. 1997 Conditionally replicating mycobacteriophages: a system for transposon delivery to Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 94:10961–10966. [PubMed: 9380742]

114. Jacobs WR, Jr., Barletta RG, Udani R, Chan J, Kalkut G, Sosne G, Kieser T, Sarkis GJ, Hatfull GF, Bloom BR. 1993 Rapid assessment of drug susceptibilities of Mycobacterium tuberculosis by means of luciferase reporter phages. Science 260:819–822. [PubMed: 8484123]

115. Piuri M, Jacobs WR, Jr., Hatfull GF. 2009 Fluoromycobacteriophages for rapid, specific, and sensitive antibiotic susceptibility testing of Mycobacterium tuberculosis. PLoS ONE 4:e4870. [PubMed: 19300517]

116. O'Donnell MR, Pym A, Jain P, Munsamy V, Wolf A, Karim F, Jacobs WR, Jr., Larsen MH. 2015 A Novel Reporter Phage To Detect Tuberculosis and Rifampin Resistance in a High-HIV-Burden Population. J Clin Microbiol 53:2188–2194. [PubMed: 25926493]

117. Jain P, Hartman TE, Eisenberg N, O'Donnell MR, Kriakov J, Govender K, Makume M, Thaler DS, Hatfull GF, Sturm AW, Larsen MH, Moodley P, Jacobs WR, Jr. 2012 phi(2)GFP10, a high-intensity fluorophage, enables detection and rapid drug susceptibility testing of Mycobacterium tuberculosis directly from sputum samples. J Clin Microbiol 50:1362–1369. [PubMed: 22278833]

118. Yu X, Gu Y, Jiang G, Ma Y, Zhao L, Sun Z, Jain P, O'Donnell M, Larsen M, Jacobs WR, Jr., Huang H. 2016 Evaluation of a High-Intensity Green Fluorescent Protein Fluorophage Method for Drug- Resistance Diagnosis in Tuberculosis for Isoniazid, Rifampin, and Streptomycin. Front Microbiol 7:922. [PubMed: 27379052]

119. Rondon L, Piuri M, Jacobs WR, Jr., de Waard J, Hatfull GF, Takiff HE. 2011 Evaluation of fluoromycobacteriophages for detecting drug resistance in Mycobacterium tuberculosis. J Clin Microbiol 49:1838–1842. [PubMed: 21346042]

120. Jain P, Weinrick BC, Kalivoda EJ, Yang H, Munsamy V, Vilcheze C, Weisbrod TR, Larsen MH, O'Donnell MR, Pym A, Jacobs WR, Jr. 2016 Dual-Reporter Mycobacteriophages (Phi2DRMs) Reveal Preexisting Mycobacterium tuberculosis Persistent Cells in Human Sputum. MBio 7.

121. Kim AI, Ghosh P, Aaron MA, Bibb LA, Jain S, Hatfull GF. 2003 Mycobacteriophage Bxb1 integrates into the Mycobacterium smegmatis groEL1 gene. Mol Microbiol 50:463–473. [PubMed: 14617171]
122. Freitas-Vieira A, Anes E, Moniz-Pereira J. 1998 The site-specific recombination locus of mycobacteriophage Ms6 determines DNA integration at the tRNA(Ala) gene of Mycobacterium spp. Microbiology 144:3397–3406. [PubMed: 9884232]
123. van Kessel JC, Marinelli LJ, Hatfull GF. 2008 Recombineering mycobacteria and their phages. Nat Rev Microbiol 6:851–857. [PubMed: 18923412]
124. van Kessel JC, Hatfull GF. 2008 Mycobacterial recombineering. Methods Mol Biol 435:203–215. [PubMed: 18370078]
125. Singh V, Dhar N, Pato J, Kolly GS, Kordulakova J, Forbak M, Evans JC, Szekely R, Rybniker J, Palcekova Z, Zemanova J, Santi I, Signorino-Gelo F, Rodrigues L, Vocat A, Covarrubias AS, Rengifo MG, Johnsson K, Mowbray S, Buechler J, Delorme V, Brodin P, Knott GW, Ainsa JA, Warner DF, Keri G, Mikusova K, McKinney JD, Cole ST, Mizrahi V, Hartkoorn RC. 2017 Identification of aminopyrimidine-sulfonamides as potent modulators of Wag31-mediated cell elongation in mycobacteria. Mol Microbiol 103:13–25. [PubMed: 27677649]
126. Marinelli LJ, Hatfull GF, Piuri M. 2012 Recombineering: A powerful tool for modification of bacteriophage genomes. Bacteriophage 2:5–14. [PubMed: 22666652]
127. Petrova ZO, Broussard GW, Hatfull GF. 2015 Mycobacteriophage-repressor-mediated immunity as a selectable genetic marker: Adephagia and BPs repressor selection. Microbiology 161:1539–1551. [PubMed: 26066798]
128. Gicquel-Sanzey B, Moniz-Pereira J, Gheorghiu M, Rauzier J. 1989 Structure of pAL5000, a plasmid from M. fortuitum and its utilization in transformation of mycobacteria. Acta Leprol 7:208–211. [PubMed: 2503992]
129. Ranes MG, Rauzier J, Lagranderie M, Gheorghiu M, Gicquel B. 1990 Functional analysis of pAL5000, a plasmid from Mycobacterium fortuitum: construction of a “mini” mycobacterium-Escherichia coli shuttle vector. J Bacteriol 172:2793–2797. [PubMed: 2158981]
130. Rauzier J, Moniz-Pereira J, Gicquel-Sanzey B. 1988 Complete nucleotide sequence of pAL5000, a plasmid from Mycobacterium fortuitum. Gene 71:315–321. [PubMed: 3224826]
131. Villar CA, Benitez J. 1992 Functional analysis of pAL5000 plasmid in Mycobacterium fortuitum. Plasmid 28:166–169. [PubMed: 1409973]
132. Chen J, Kriakov J, Singh A, Jacobs WR, Jr., Besra GS, Bhatt A. 2009 Defects in glycopeptidolipid biosynthesis confer phage I3 resistance in Mycobacterium smegmatis. Microbiology 155:4050–4057. [PubMed: 19744987]
133. Boritsch EC, Brosch R. 2016 Evolution of Mycobacterium tuberculosis: New Insights into Pathogenicity and Drug Resistance. Microbiol Spectr 4.
134. Sassi M, Bebeacua C, Drancourt M, Cambillau C. 2013 The First Structure of a Mycobacteriophage, the Mycobacterium abscessus subsp. bolletii Phage Araucaria. J Virol 87:8099–8109. [PubMed: 23678183]
135. Sassi M, Gouret P, Chabrol O, Pontarotti P, Drancourt M. 2014 Mycobacteriophage-drived diversification of Mycobacterium abscessus. Biol Direct 9:19. [PubMed: 25224692]
136. Mehla J, Dedrick RM, Caufield JH, Siefring R, Mair M, Johnson A, Hatfull GF, Uetz P. 2015 The Protein Interactome of Mycobacteriophage Giles Predicts Functions for Unknown Proteins. J Bacteriol 197:2508–2516. [PubMed: 25986902]
137. Mehla J, Dedrick RM, Caufield JH, Wagemans J, Sakhawalkar N, Johnson A, Hatfull GF, Uetz P. 2017 Virus-host protein-protein interactions of mycobacteriophage Giles. Sci Rep 7:16514. [PubMed: 29184079]

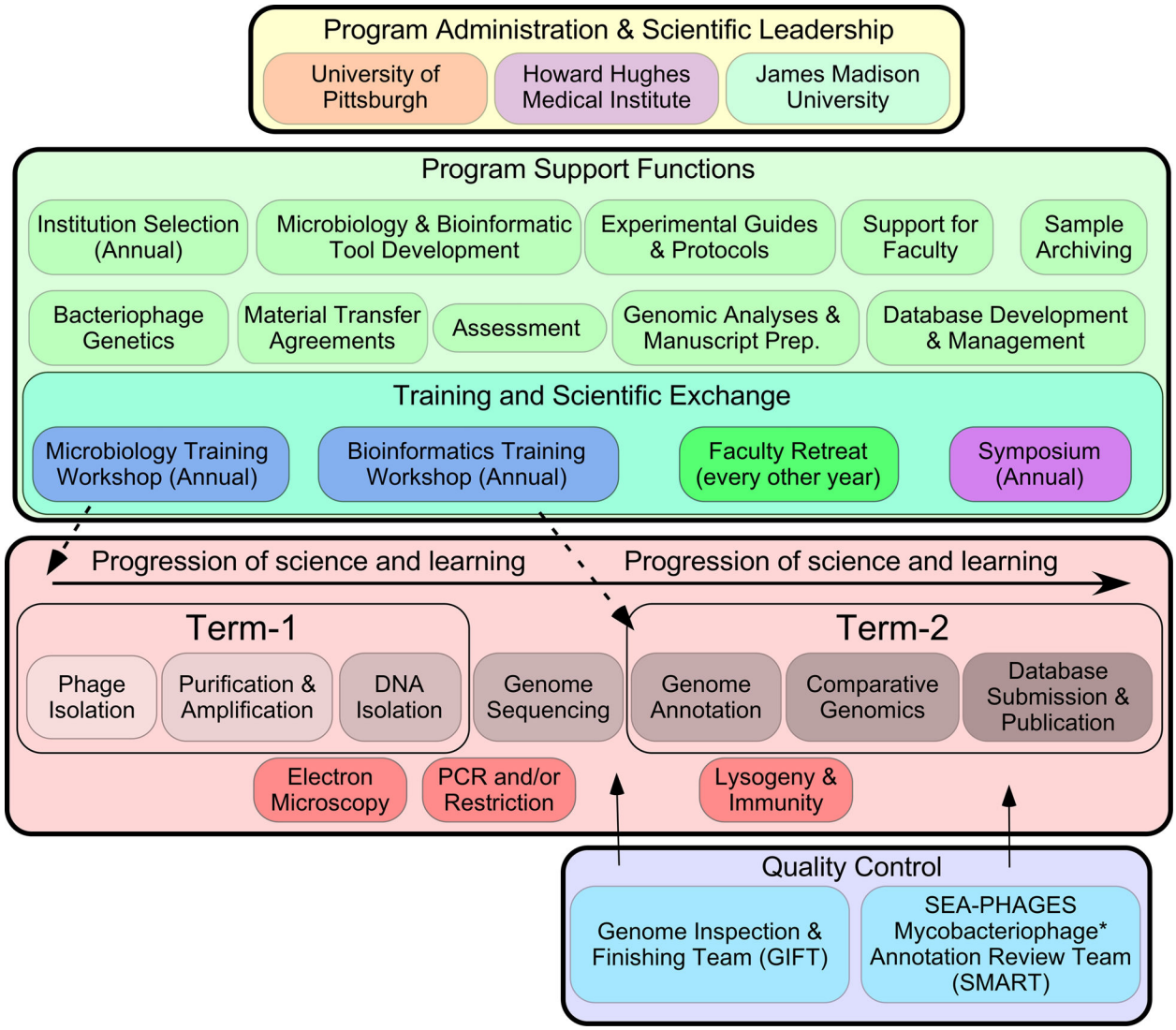

Figure 1. Organization and structure of the SEA-PHAGES program.

SEA-PHAGES program administrators (yellow box, top) oversee support components critical to program implementation (green box, upper middle). Typical two-term course structure (red box, lower middle) includes phage isolation through comparative genomics; additional characterization includes electron microscopy, PCR/restriction analysis, and lysogeny assays (red ovals). Sequence and annotation quality control is shared with SEA-PHAGES faculty teams (purple box, bottom right). Reproduced with permission from reference 34.

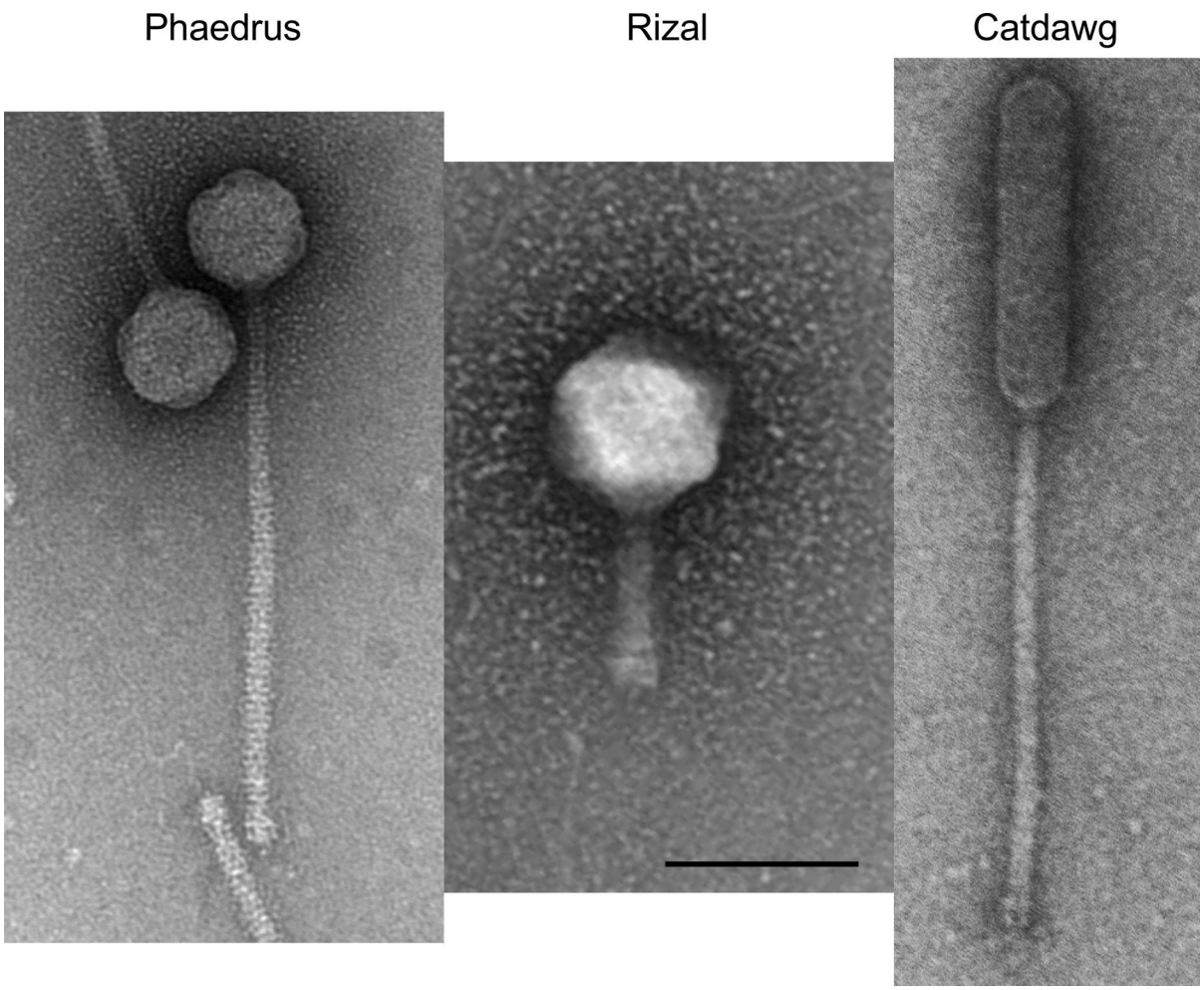

Figure 2. Mycobacteriophage viral morphotypes.

Mycobacteriophages generally fall into two morphological families, the *Siphoviridae* with long flexible tails (e.g. Phaedrus and Catdawg), and the *Myoviridae* with contractile tails (e.g. Rizal). Most of the siphoviral phages have isometric heads (e.g. Phaedrus), but some (e.g. Catdawg) have prolate heads. Scale marker is 100 nm.

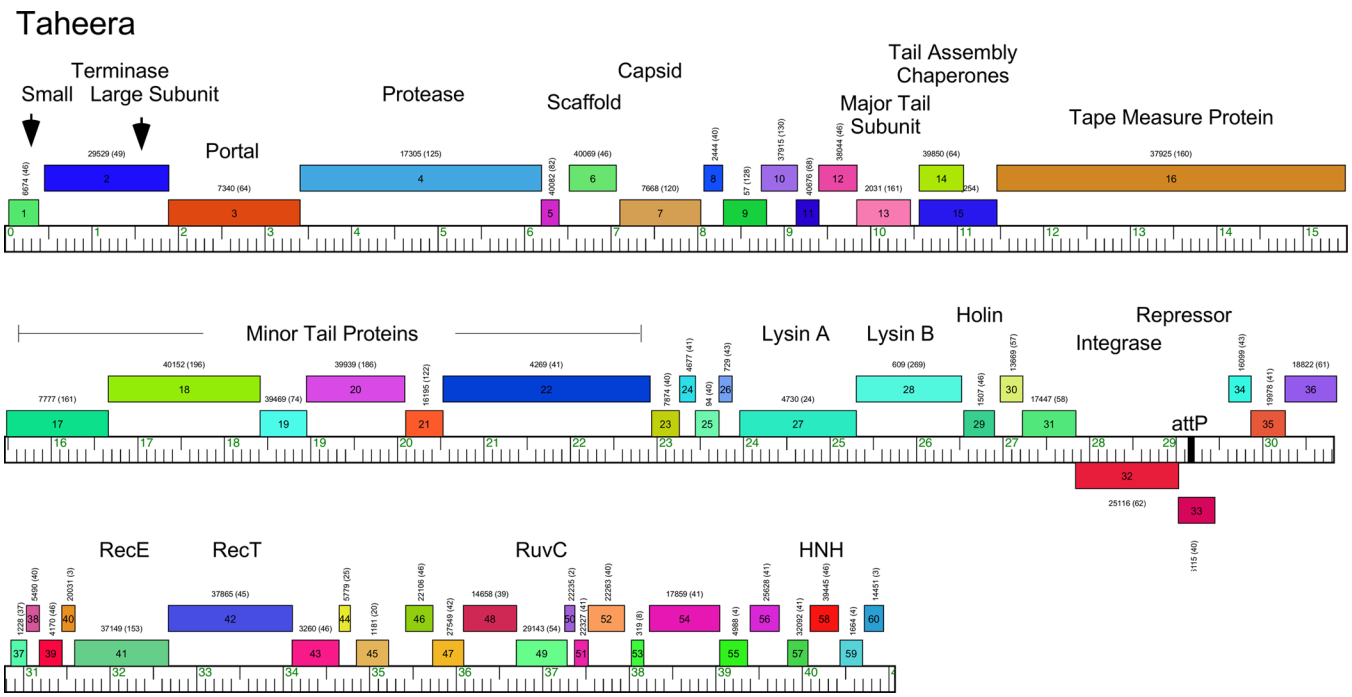

Figure 3. Genome organization of mycobacteriophage Taheera.

The genome of phage Taheera (Cluster G) is represented as a scale bar (major intervals: 1 kbp) with predicted genes shown as boxes either above (rightwards transcribed) or below (leftwards transcribed). Gene numbers are shown within each box and the phamily designation is shown either above or below with the number of phamily members shown in parentheses. Putative gene functions are indicated.

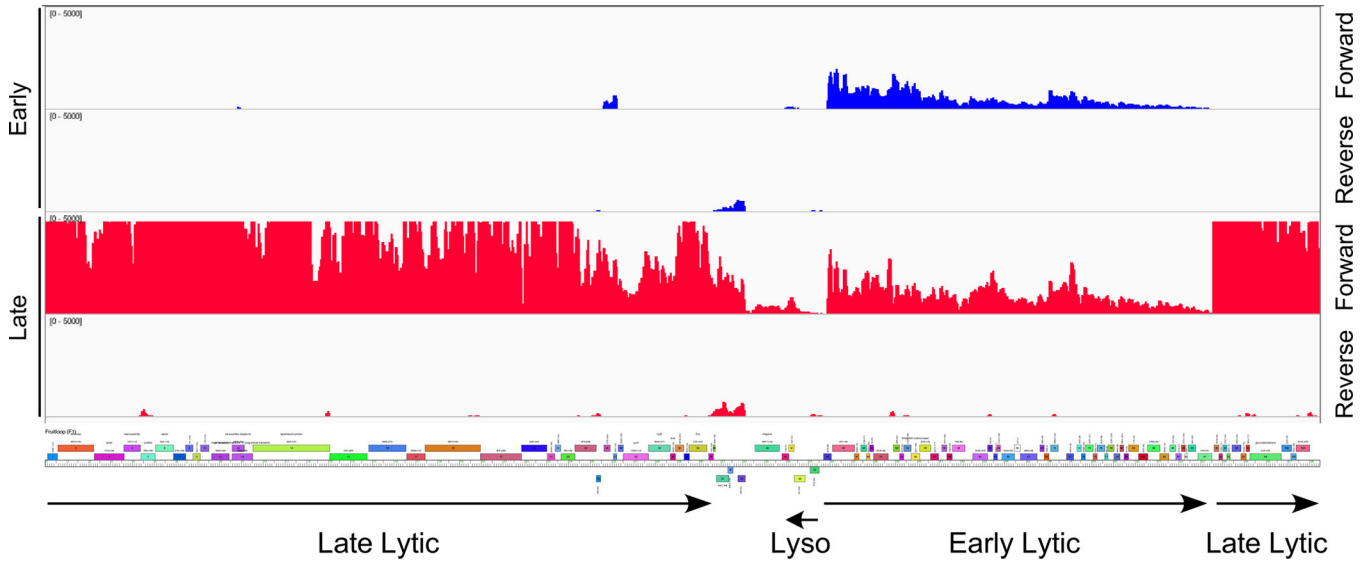

Figure 4. Gene expression in mycobacteriophage Fruitloop.

RNA was isolated from *M. smegmatis* infected with phage Fruitloop at either early (30mins) or late (150 mins) time points after infection. Following strand-specific RNAseq analysis, sequence reads were mapped to the Fruitloop genome (shown below). Reads mapping to the forward and reverse strands are indicated on the right. Scales for the numbers of sequence reads are from 0–5000. Reproduced with permission from reference 62.

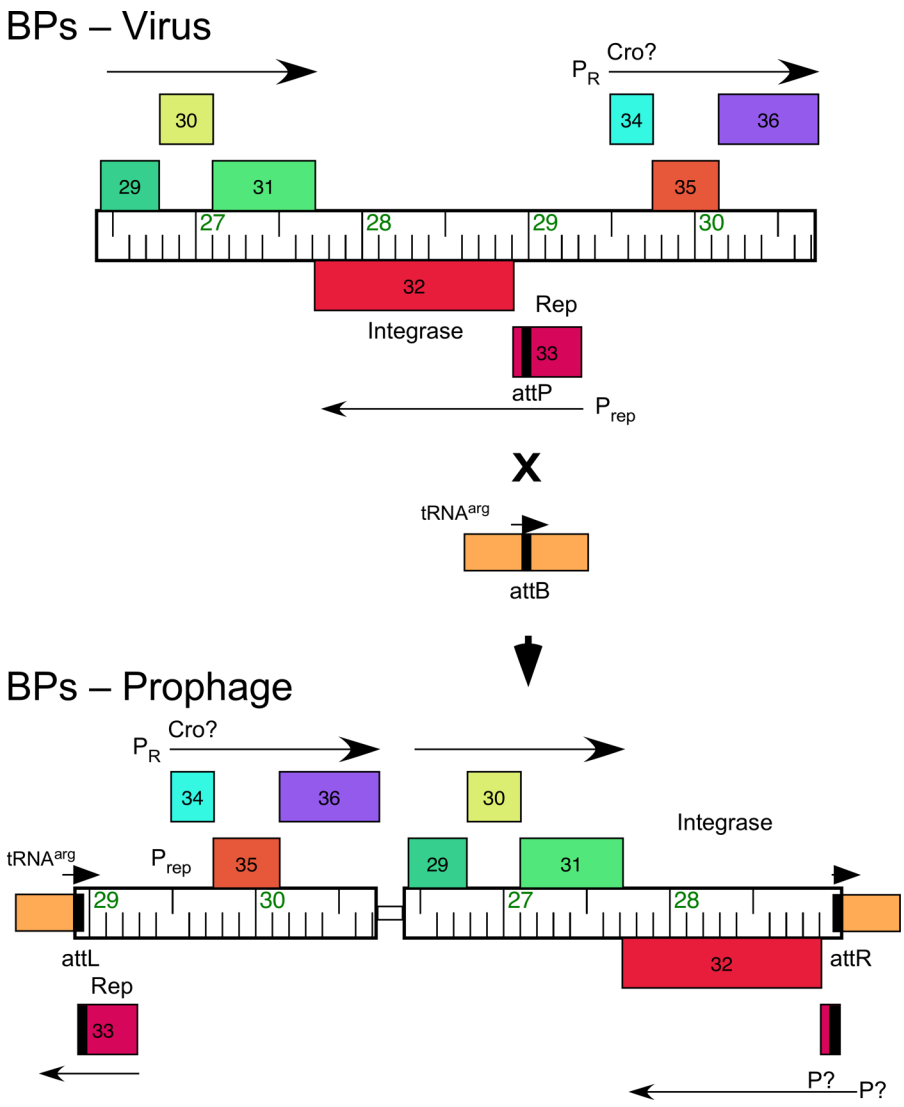

Figure 5. Integration-dependent immunity regulation.

BPs is a Cluster G phage, and a central ~4 kbp segment of the genome is shown at the top. Genes are shown as colored boxes, transcribed rightwards or leftwards as depicted above or below the genome respectively. The repressor and integrase genes (*33* and *32*, respectively) are transcribed leftwards from the P_{rep} promoter. The *attP* site (black bar) is located within the repressor open reading frame, and can recombine with an *attB* site overlapping an *M. smegmatis* $tRNA^{arg}$ gene. Establishment of lysogeny involves formation of an integrated prophage in which an active form of the repressor is expressed from near *attL*, and the 3' end of the gene is near *attR*. The virally-encoded form of the repressor contains a C-terminal *ssrA*-like tag that targets it for degradation and it fails to confer immunity. Integration results in removal of the C-terminal tag such that the prophage-expressed form of the repressor is stable and confers immunity.

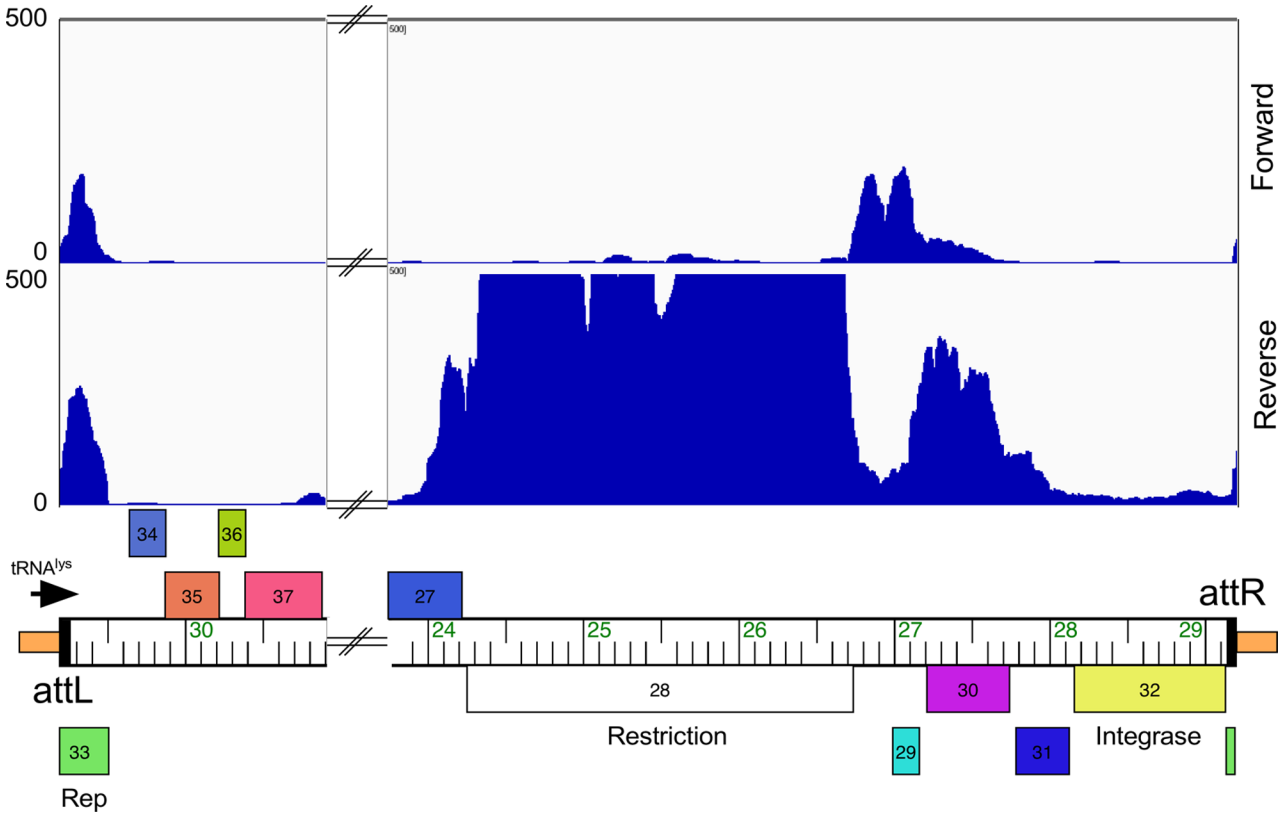

Figure 6. Gene expression in a Panchino lysogen of *M. smegmatis*.

RNA was isolated from a lysogen of mycobacteriophage Panchino and strand-specific RNAseq reads were mapped to the Panchino prophage. Only the genome segments at the ends of the prophage near *attL* and *attR* are shown. Sequence reads for Forward and Reverse DNA direction are shown, and the scales are 0–500 reads. Genes below the genome are transcribed leftwards, and those above are transcribed rightwards. Panchino uses an integration-dependent immunity system similar to that in phage BPs (see Fig. 5). Genes *31-28* are also expressed; the lytic genes are not expressed (not shown). Reproduced with permission from reference 63.

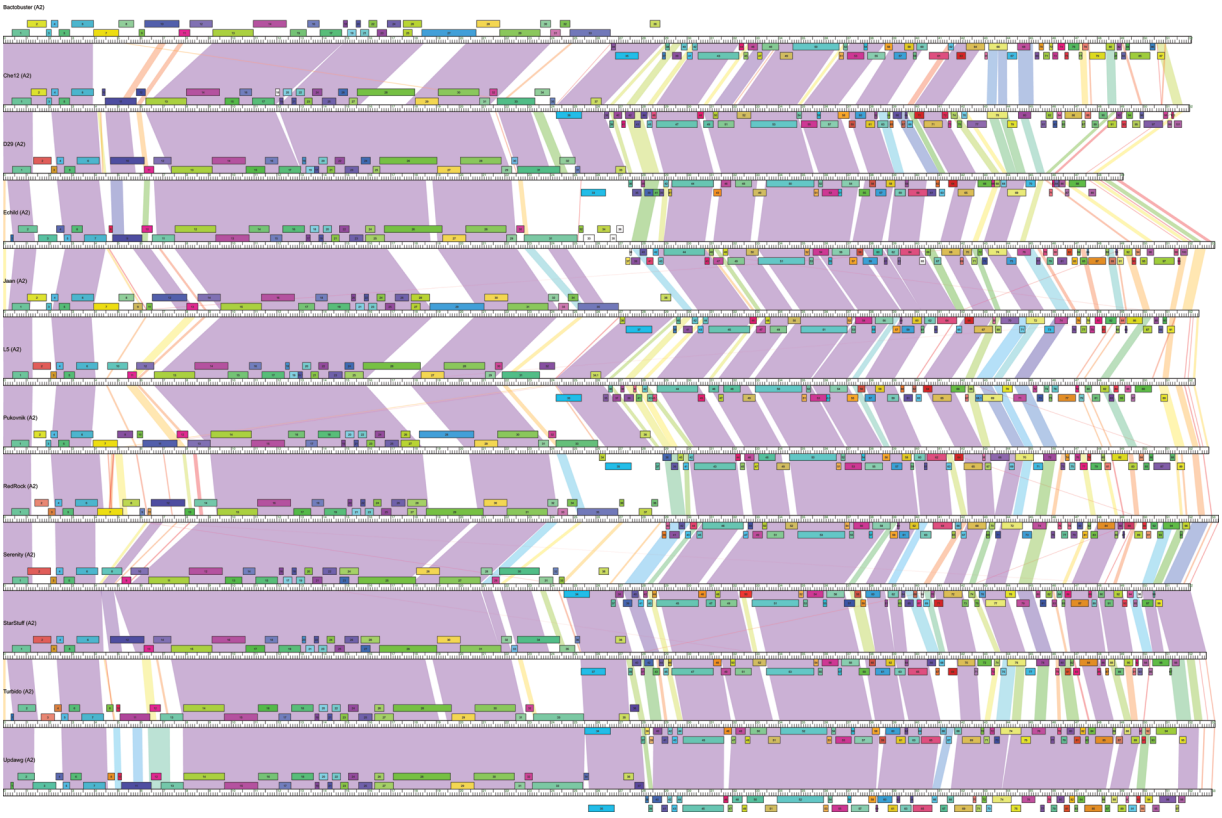

Figure 7. Genomic mosaicism in Subcluster A2 phages.

Mycobacteriophages are characteristically mosaic, with shared genome segments interspersed with non-homologous regions. The genomes of 12 Subcluster A2 phages are shown (Bactobuster, Che12, D29, Echild, Jaan, L5, Pukovnik, RedRock, Serenity, StarStuff, Turbido, and Updawg), with pairwise nucleotide sequence similarities shown as shadings between the genomes; shading is spectrum colored with violet being the most closely related, and red being just above the threshold E-value of 10^{-5}. Genes are shown as colored boxes, and genes within the same phamily are similarly colored. Note the segments with close sequence similarity (violet-shaded regions) interspersed with dissimilar regions (white-shaded segments).

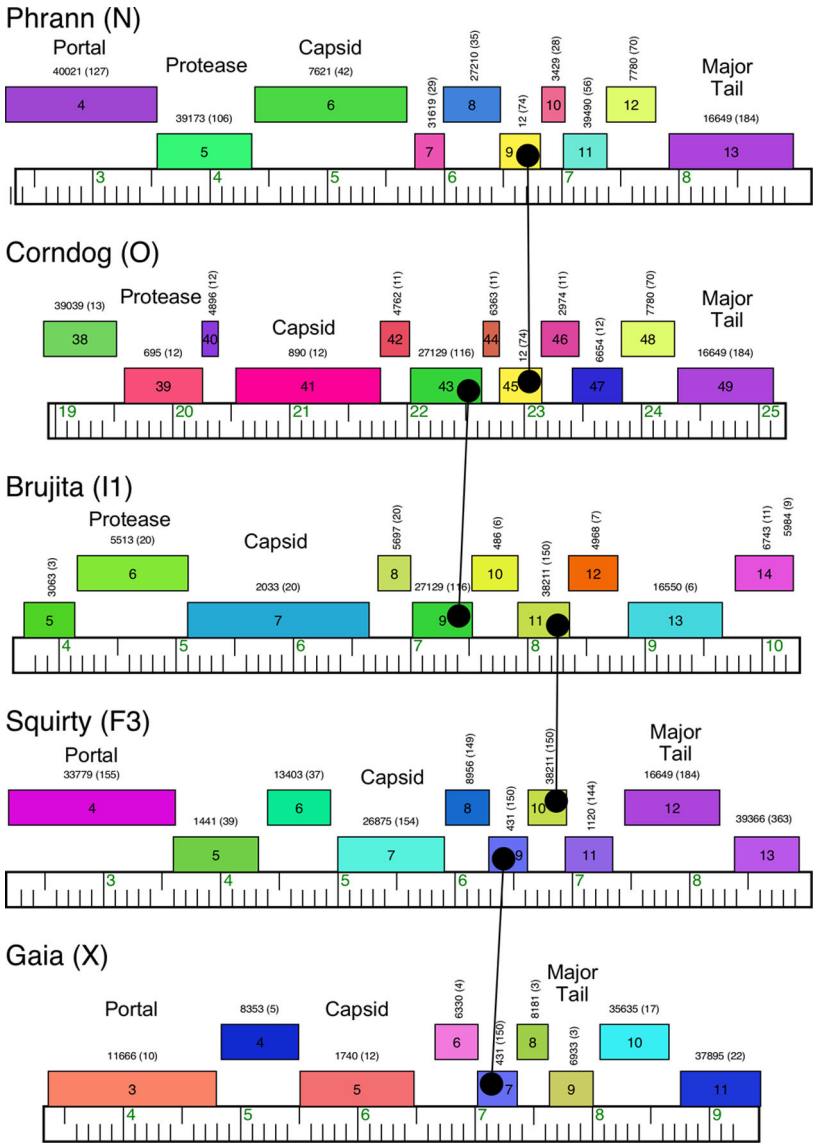

Figure 8. Single-gene mosaics in mycobacteriophage genomes.

Segments of five phage genomes (Phrann, Corndog, Brujita, Squirty, and Gaia) are shown, which are unrelated at the nucleotide sequence level and are grouped in different clusters, as shown (in Clusters, N, O, I1, F3, and X, respectively). Genes are shown as colored boxes, with genes in the same phamilies colored similarly, and pham numbers shown above with the number of phamily members in parentheses. Pairwise-shared homologues are indicated with dumbbells. Note that the homologues are situated in different genomic contexts and the flanking genes are not homologous.

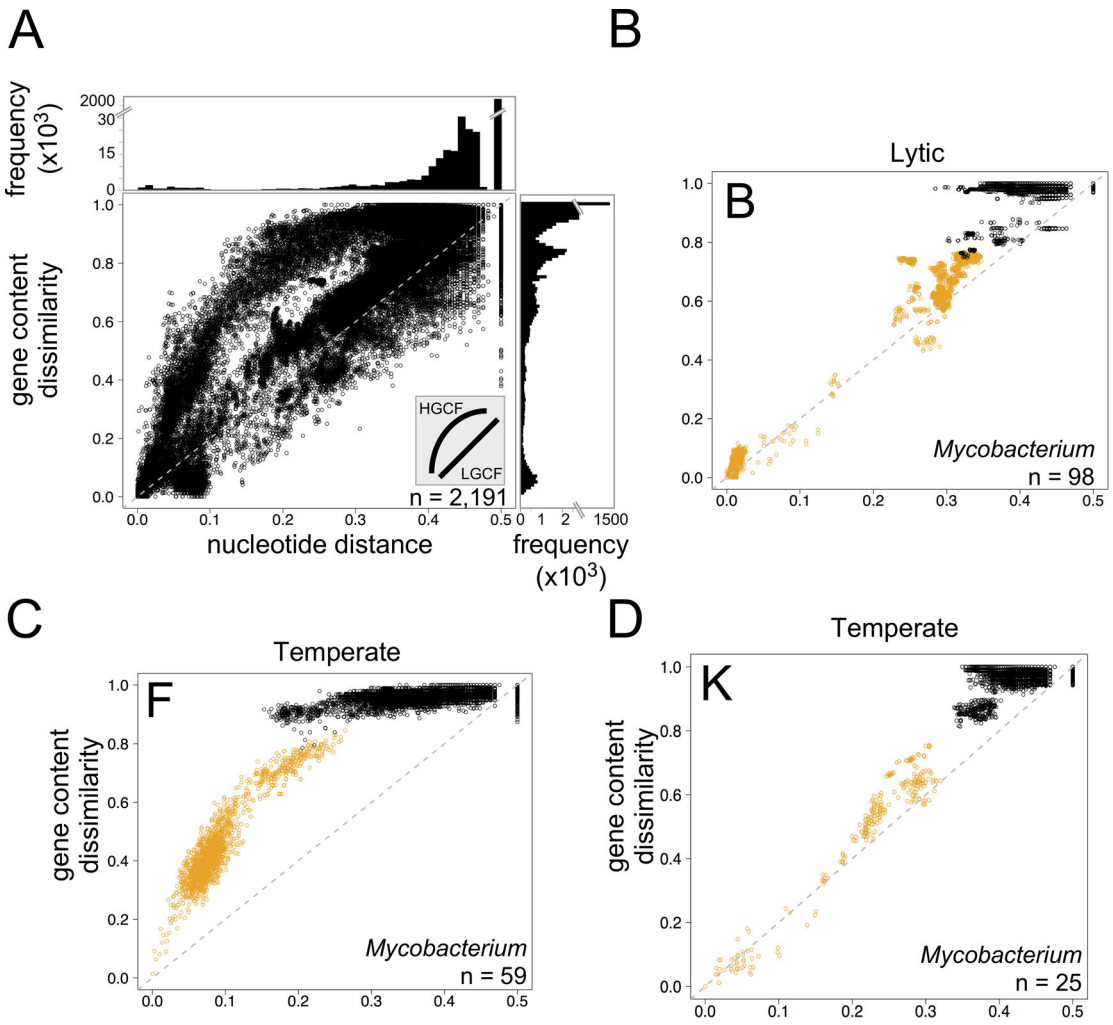

Figure 9. Modes of phage evolution.

A. Nucleotide distance (using Mash) and gene content dissimilarity (using phams from Phamerator) are plotted for ~2.4 × 10^6 dsDNA phage comparisons to reveal two evolutionary modes, HGCF and LGCF (inset). The line at y=2× is plotted for reference. Marginal frequency histograms emphasize densely plotted regions, with truncated y axes for viewability. **B-D**. Cluster-specific intra-cluster (orange) and inter-cluster (black) comparisons are plotted as in Fig. 1a for actinobacteriophage Clusters B (panel B), F (panel C), and K (panel D). Phages in Clusters F and K are temperate, and phages in Cluster B are lytic, as indicated. n, number of phages present in the specific cluster. Reproduced with permission from reference 81.

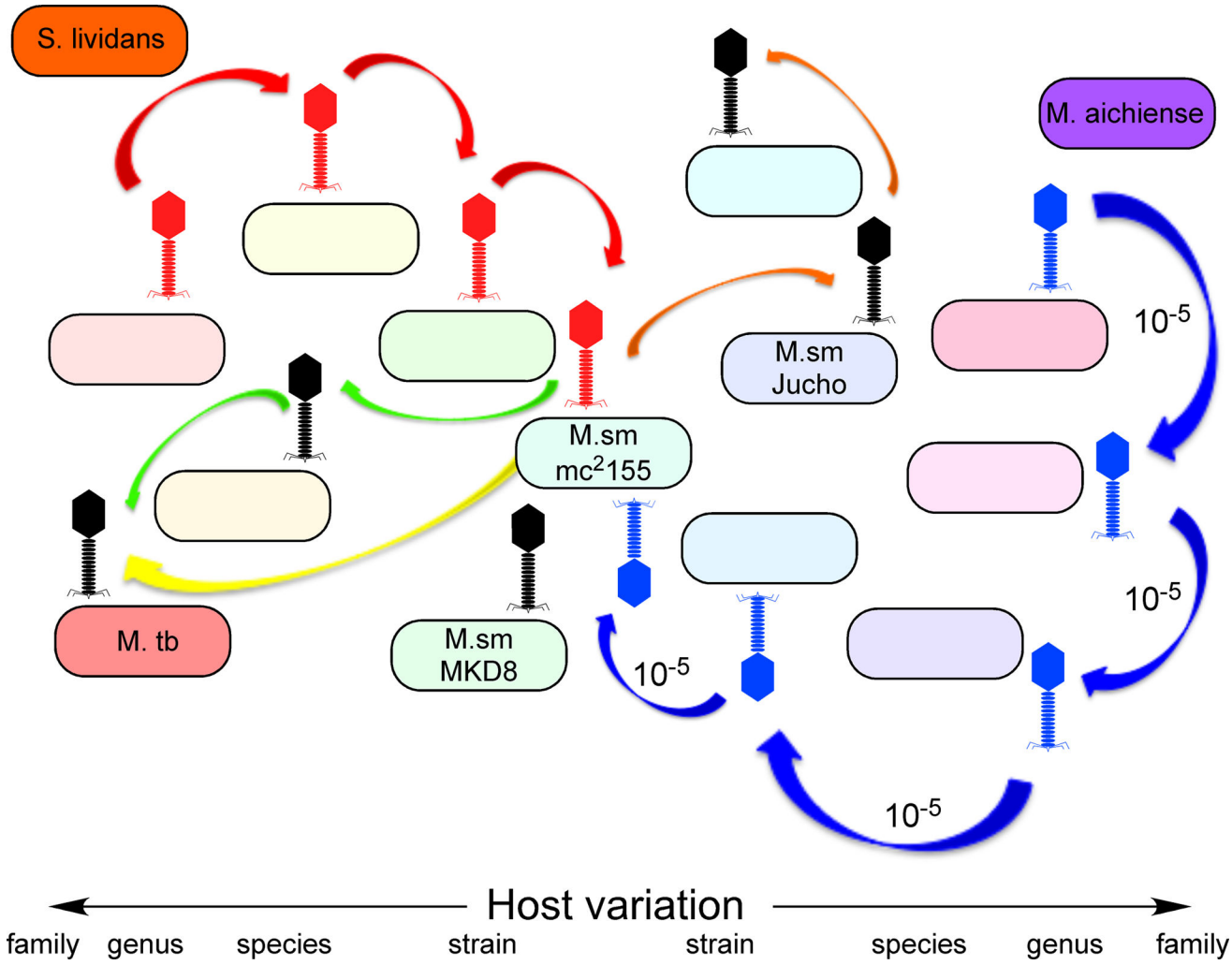

Figure 10. A model for mycobacteriophage diversity.

The diversity of mycobacteriophages isolated on *M. smegmatis* mc^{2}155 can be explained by a model in which a) phages can readily infect a new bacterial host – either by a switch or an expansion of host range – and b) a highly diverse bacterial population, including many closely-related strains, in the environments from which the phages are isolated. As such, phages with distinctly different genome sequences and GC% contents infecting distantly related bacterial hosts – such as those to the left (red) or right (blue) extremes of a spectrum of hosts – can migrate across a microbial landscape through multiple steps. Each host switch occurs at a relatively high frequency (~1 in 10^5 particles, or an average of about one every 10^3 bursts of lytic growth), and much faster than either amelioration of phage GC% to its new host, or genetic recombination. Two phages (such as those shown in red and blue) can thus 'arrive' at a common host (*M. smegmatis* mc^{2}155) but be of distinctly different types (clusters, subclusters, and singletons). The variety of hosts is shown two dimensionally for simplicity, and the actual relationships among bacteria in environments such as soil and compost is likely to be considerably more complicated. Because host range variation is a common feature of bacteriophages, the model predicts that a high degree of phage diversity will be seen for any particular host if the microbial population from which the phages are

isolated from is also highly diverse and rich in closely-related strains. Reproduced with permission from reference 84.

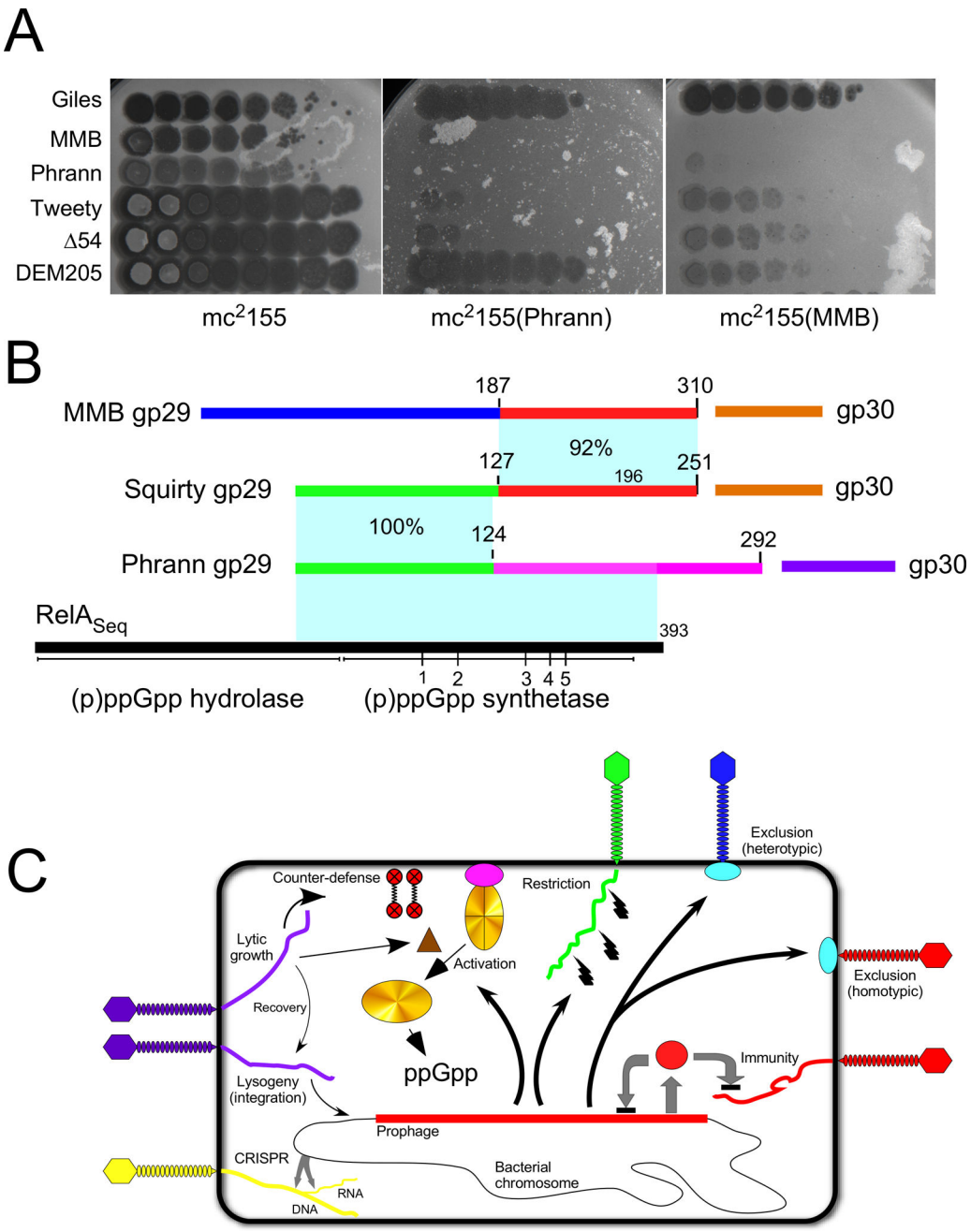

Figure 11. Prophage-mediated defense against phage attack.

A. Phage defense plaque assays. Ten-fold serial dilutions of phages Giles (control), MichelleMyBell (MMB), Phrann, and Tweety were plated onto lawns of a non-lysogen (mc^2155) and *M. smegmatis* lysogens of Phrann [mc^2155(Phrann)] and MichelleMyBell [mc^2155(MMB)]. The Phrann and MMB lysogens confer superinfection immunity to Phrann and MMB, as these are part of Cluster N, and all of the Cluster N phages are homoimmune, coding for closely-related repressors. Tweety (Cluster F) is unrelated to Phrann and MMB, and the Phrann and MMB lysogens defend against Tweety and the efficiency of plating is greatly reduced relative to the wild type strain. A defense escape

mutant (DEM205) overcomes the Phrann defense and plates efficiently on the Phrann lysogen but remains subject to MMB-mediated defense. DEM205 contains a mutational change in gene *54*, but it is a gain of function mutant, as a Tweety mutant in which gene 54 has been deleted (Δ54) is still targeted by the Phrann defense. **B**. Relationships among phage-encoded (p)ppGpp synthetase-like proteins. Phrann gp29 is a homologue of RelA/SpoT proteins with similarity to the (p)ppGpp synthetase domain of *Streptococcus equimilis* RelA including 5 conserved motifs (1–5). The Cluster F phage Squirty encodes a related protein (gp29) sharing the N-terminal 124 residues with Phrann gp29, but with divergent C-termini. MMB gp29 is not predicted to be related to Rel_{Seq}, but shares its C-terminus with Squirty gp29. MMB gp30 and Squirty gp30 are closely-related and are both predicted to be membrane localized. **C**. Models for prophage-mediated viral defense. An integrated prophage (red line) confers defense against viral attack through numerous mechanisms, either homotypically (i.e. against the same or closely-related viruses) or heterotypically (against unrelated phages). Homotypic defense includes repressor-mediated immunity (repressor, red circle) and superinfection exclusion (blue circle) against itself (red phage). Heterotypic defense includes an exclusion-like system illustrated by Charlie gp32 defense against Che9c (blue phage), and restriction against many viruses (illustrated by the green phage) by Panchino gp28. Defense is also mediated by a predicted (p)ppGpp synthetase (e.g. Phrann gp29; gold circle), which we propose is kept in an inactive form (gold circle with crossed lines) by an inhibitor (purple circle), which for Squirty gp30 is membrane localized. Lytic growth by specific phages activates the defense through early lytic protein, which is proposed to dissociate the (p)ppGpp synthetase from its inhibitor, enabling rapid accumulation of (p)ppGpp and growth arrest. Tweety encodes a counter-defense system (gp54) that may prevent activation of (p)ppGpp synthesis. Reproduced with permission from reference 63.

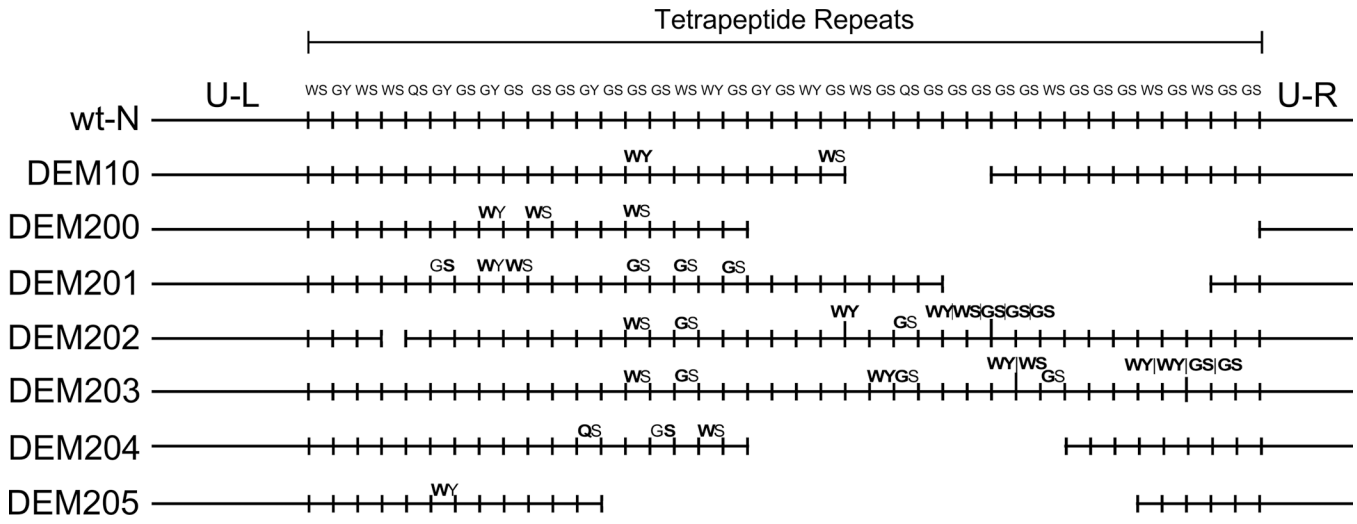

Figure 12. Variations in tetrapeptide repeats in Tweety gp54.

Gene *54* of phage Tweety codes for a 374-residue protein in which unique N-terminal (U-L) and C-terminal (U-R) flank a long series of tetrapeptide repeats. The number of repeats in the original sequenced genome is 48, although the derivative used for defense studies (wt-N) shown here only has 40. Each of the repeats has a sequence AAXX, where XX is either WS, GY, QS, GS or WY, and the sequence at each repeat is shown. The organization of nine defense escape mutants (DEM) are shown below, each of which efficiently overcomes defense mediated by the Phrann prophage. They all remain subject to defense mediated by the MichelleMyBell prophage. Six of the DEM mutants (DEM10, DEM200, DEM201 DEM202, DEM204, DEM205) have lost some of the repeats, but also have altered repeat sequences, as indicated. DEM202 also has an insertion of five additional repeats copies. DEM203 has two insertions, one of two repeat units, and the other with four repeat units.

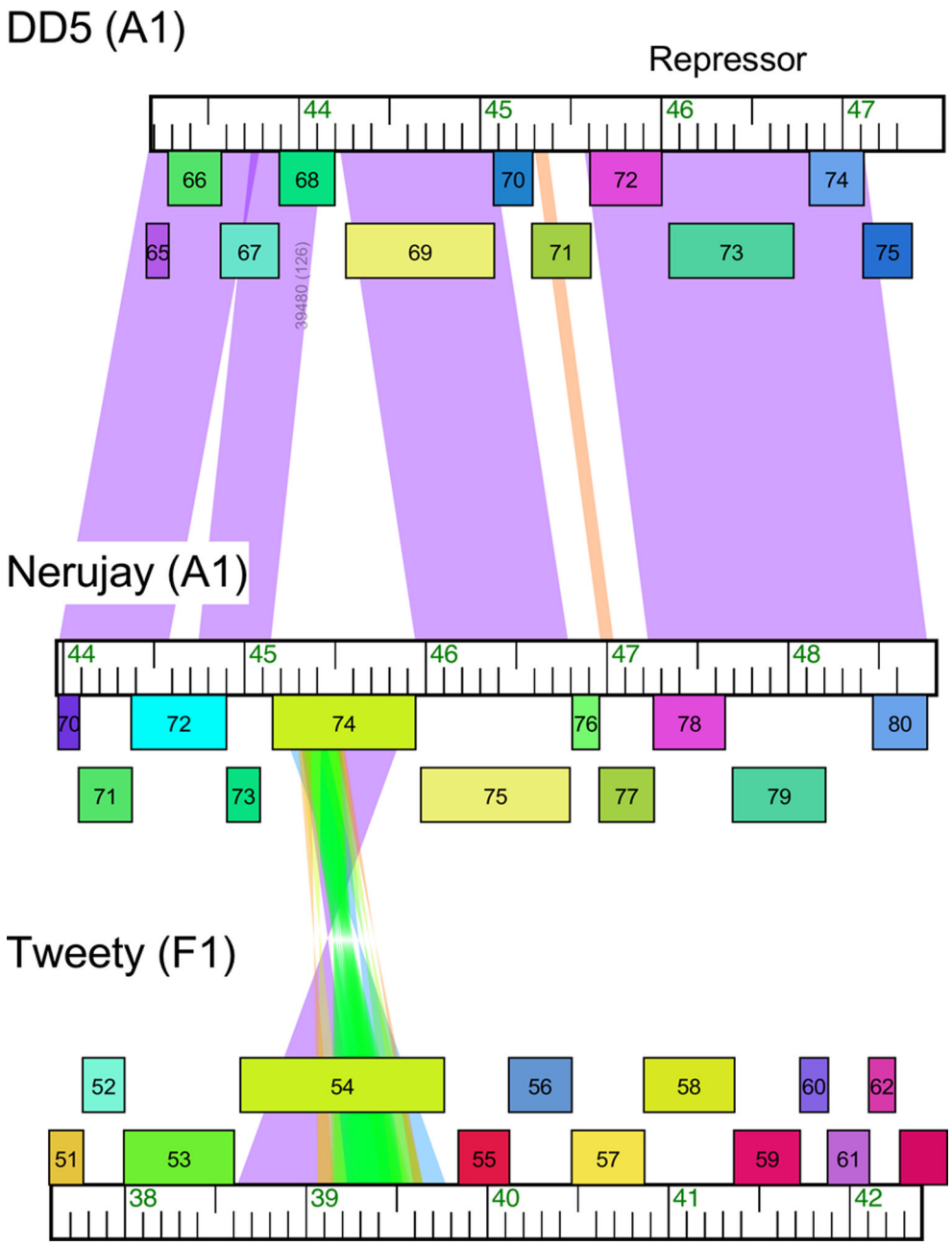

Figure 13. Genomes with potential counter-defense systems.

Tweety gene *54* is active in counter defense against Phrann (see Figs. 11, 12), and there are homologues in many other Cluster F phages. The only phage outside of Cluster F with a closely-related homologue is the Subcluster A1 phage, Nerujay. All other Cluster A phages, including DD5, lack a Tweety *54* homologue. Segments of the genome maps are aligned with pairwise nucleotide sequence similarity shown as colored shading, as described for Figure 7. In this region, all of the DD5 and Nerujay genes are transcribed leftwards, and the Tweety genes are transcribed rightwards.

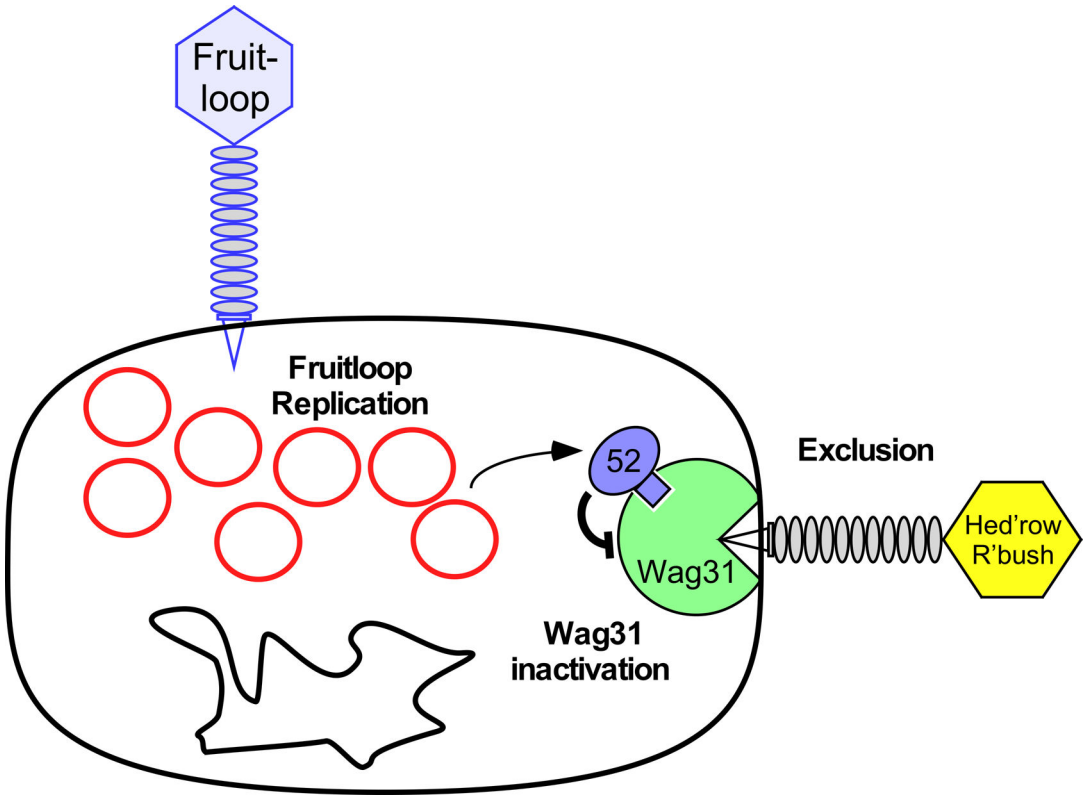

Figure 14. A model for Fruitloop gp52-mediated superinfection exclusion of Wag31-dependent phages.

This model proposes a role for gp52 expressed by lytically growing phage Fruitloop in binding to and inactivating the host Wag31 (DivIVA) protein, which is essential and localizes to the growing pole of the cell. Some phages, such as Subcluster B2 Hedgerow and Rosebush depend on Wag31 for efficient infection, such that gp52-mediated inactivation of Wag31 excludes them from superinfecting. Replicating Fruitloop genomes are shown as red circles alongside the bacterial chromosome (black circle), and Fruitloop gp52 is shown in blue, binding to Wag31 shown in green. Not drawn to scale. Reproduced with permission from reference 62.

Table 1.

Mycobacteriophage clusters and subclusters

Cluster	# phages	# Subclusters[1]	Avg genome length	Avg G+C%
A	318	18	51656	63.3
B	147	6	68705	67.2
C	49	2	155858	64.7
D	10	2	64965	59.4
E	52	-	75446	63.0
F	84	3	57427	61.5
G	28	4	42429	66.9
H	6	2	69292	57.3
I	5	2	51129	66.3
J	17	-	109962	61.0
K	71	6	60000	66.8
L	27	3	74978	59.0
M	7	2	81539	61.2
N	14	-	43200	66.2
O	9	-	70805	65.4
P	17	2	47901	67.0
Q	6	-	53753	67.5
R	5	-	71424	56.0
S	3	-	64989	63.4
T	5	-	42847	66.2
U	2	-	69942	50.4
V	2	-	78262	56.8
W	3	-	60820	67.5
X	1	-	90460	56.8
Y	1	-	77832	67.3
Z	2	-	50806	66.0
AB	2	-	47976	58.7
AC	1	-	68869	49.1
Dori[2]	-	N/A	64613	66.0
Kumao[2]	-	N/A	70373	62.1
MooMoo[2]	-	N/A	55178	62.0
DS6A[2]	-	N/A	60588	68.4
Sparky[2]	-	N/A	63334	65.2

[1]The numbers of subclusters are shown but are not applicable (N/A) for singleton phages.

[2]Singleton phages with no close relatives. Phage names are provided.

Table 2.

Allocation of cluster designations

Host Phylum	Host Genus	Designations	Number
Actinobacteria	Mycobacterium	A – Z, AA – AJ	40
	Arthrobacter	AK – AZ, FA – FZ	42
	Streptomyces	BA – BT	20
	Propionibacterium	BU – BZ	6
	Rhodococcus	CA – CP	14
	Gordonia	CQ – DZ	36
	Microbacterium	EA – EM	13
	Corynebacterium	EN – EZ	13