Cell, Vol. 113, 171–182, April 18, 2003, Copyright ©2003 by Cell Press

Origins of Highly Mosaic Mycobacteriophage Genomes

Marisa L. Pedulla,[1] Michael E. Ford,[1] Jennifer M. Houtz,[1] Tharun Karthikeyan,[1] Curtis Wadsworth,[1] John A. Lewis,[1] Debbie Jacobs-Sera,[2] Jacob Falbo,[2] Joseph Gross,[2] Nicholas R. Pannunzio,[1] William Brucker,[1] Vanaja Kumar,[3] Jayasankar Kandasamy,[3] Lauren Keenan,[4] Svetsoslav Bardarov,[5] Jordan Kriakov,[5] Jeffrey G. Lawrence,[1] William R. Jacobs, Jr.,[5] Roger W. Hendrix,[1] and Graham F. Hatfull[1,*]

[1]Department of Biological Sciences and
Pittsburgh Bacteriophage Institute
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

[2]Greater Latrobe High School
131 Arnold Palmer Drive
Latrobe, Pennsylvania 15650

[3]Tuberculosis Research Center
ICMR
Chetput
Chennai, India 600031

[4]Pelham Memorial High School
Colonial and Corlies Avenues
Pelham, New York 10803

[5]Howard Hughes Medical Institute
Department of Microbiology and Immunology
Albert Einstein College of Medicine
1300 Morris Park Avenue
Bronx, New York 10461

Summary

Bacteriophages are the most abundant organisms in the biosphere and play major roles in the ecological balance of microbial life. The genomic sequences of ten newly isolated mycobacteriophages suggest that the bacteriophage population as a whole is amazingly diverse and may represent the largest unexplored reservoir of sequence information in the biosphere. Genomic comparison of these mycobacteriophages contributes to our understanding of the mechanisms of viral evolution and provides compelling evidence for the role of illegitimate recombination in horizontal genetic exchange. The promiscuity of these recombination events results in the inclusion of many unexpected genes including those implicated in mycobacterial latency, the cellular and immune responses to mycobacterial infections, and autoimmune diseases such as human lupus. While the role of phages as vehicles of toxin genes is well established, these observations suggest a much broader involvement of phages in bacterial virulence and the host response to bacterial infections.

Introduction

Bacteriophages may be the dark matter of the biological universe. The vast number of tailed phage particles on Earth—estimated at 10^{31} (Wommack and Colwell, 2000)—has been appreciated only recently, but it is increasingly clear that these phages exert enormous influence over the microbial world (Brussow and Hendrix, 2002; Wilhelm and Suttle, 1999). The overall diversity of this population appears to be great: no genomically defined phage has been isolated more than once, and the relatively few sequenced phage genomes are highly varied. This is perhaps not unexpected since the phage population is wonderfully dynamic, turning over rapidly through constant attrition and subsequent amplification in permissive hosts (Garza and Suttle, 1998; Short and Suttle, 1999). On a global scale, we estimate that $\sim 10^{25}$ phages initiate an infection every second, and in each of those infections the phage encounters DNA—of bacterial or prophage origin—with which it can potentially recombine to generate new genomic arrangements. It is likely that this process has been underway for over 3 billion years (Hendrix et al., 1999). This is combinatorial chemistry on a grand scale.

Mycobacteriophages—phages of the mycobacteria—have proven useful for diagnosis of mycobacterial infections such as tuberculosis, and in the development of tools for mycobacterial genetics (Eltringham et al., 1999; Hatfull, 2000; Jacobs et al., 1993; Jones, 1990). While many mycobacteriophages have been isolated (Hatfull and Jacobs, 1994), only four have been characterized genomically: L5, D29, Bxb1, and TM4 (Ford et al., 1998a, 1998b; Hatfull, 2000; Hatfull and Sarkis, 1993; Mediavilla et al., 2000). Although these were isolated at different times and in different locations, they have many features in common. For example, they are morphologically similar (see Figure 1), have similarly sized genomes (49.1–52.8 kbp), and have similar arrangements of structure and assembly genes, many of which encode related gene products. The two most closely related genomes are those of L5 and D29, which share over 75% of their genes (as determined by amino acid sequence similarity) and have extensive similarity at the DNA sequence level (Ford et al., 1998a). Bxb1 shares only little DNA sequence similarity with these although more than 40% of the predicted protein products are related to those of L5 and D29 (Mediavilla et al., 2000). All three also have a common genomic architecture with genes transcribed from the cohesive termini toward an attachment site (*attP*) in the center of the genome. TM4 represents a departure from these in both genome organization and sequence. All of the genes are transcribed in the same direction and relatively few genes ($\sim$10%) are shared with L5, D29, or Bxb1; nevertheless, these four phages appear to share a common block of genes involved in virion structure and assembly arranged in a similar order (Ford et al., 1998b).

Questions about how phages generate genomic diversity were first addressed (albeit at low resolution) 35 years ago with the "lambdoid" phages of enteric hosts (Simon et al., 1971; Westmoreland et al., 1969). These experiments showed for the first time that lambdoid phage genomes are extensively mosaic with respect to each other. It was clear, furthermore, that the mosaic

*Correspondence: gfh@pitt.edu

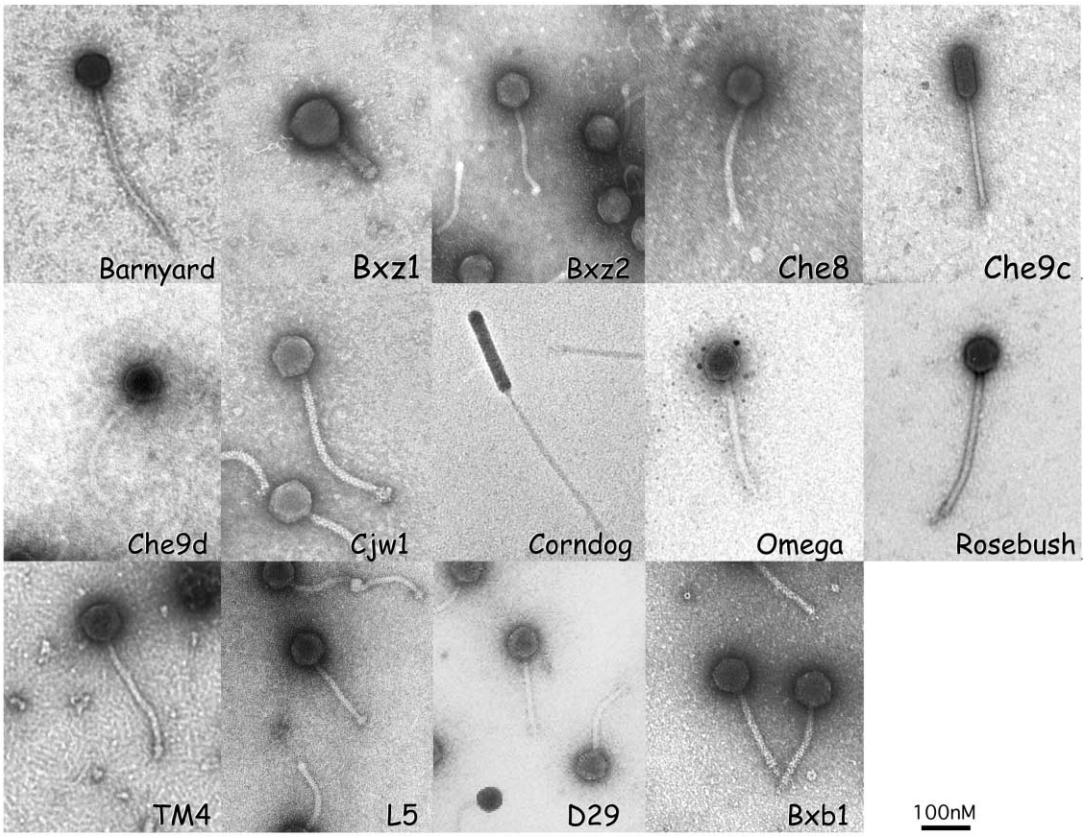


Figure 1. Mycobacteriophage Virion Morphologies

Electron micrographs of representative particles of fourteen mycobacteriophages illustrate a variety of virion morphologies.

joints—presumably sites of ancestral recombination—were not positioned randomly across the genome but rather recurred at certain locations, possibly between genes or conserved clusters of genes. These results led to the "modular evolution" model (Susskind and Botstein, 1978), in which it was postulated that phages evolve by genetic exchange at special intergenic sites, either through homologous recombination or by a site-specific mechanism. The recent availability of complete genome sequences for several lambdoid phages has allowed reexamination of these questions at higher resolution, and this has led to a different picture of mosaic formation (Hendrix, 2002; Juhala et al., 2000). In this view, illegitimate recombination takes place quasi-randomly along the recombining genomes, generating an unholy mélange of recombinant types, almost all of which will be defective for phage growth as a consequence of their misplaced recombination. Natural selection eliminates all but the tiny minority of recombinants in which biological function is intact—for the most part, phages with mosaic joints that lie between genes—thus giving rise to an observable population in which the sites of recombination are anything but random.

It has not been clear to what degree this picture of rampant horizontal exchange extends to other phage groups. For the Dairy phages (Brussow, 2001), although there is clear evidence of horizontal exchange, the many sequenced genomes of these phages are remarkable for their similarity of size, organization, and sequence (Brussow and Desiere, 2001; Brussow and Hendrix, 2002). This may be a consequence of the narrow range of habitats from which most of these phages were derived (commercial dairy fermentors), or alternatively, it may indicate a fundamental difference in their evolutionary mechanisms. In the case of the four previously sequenced mycobacteriophages, described above, there is also clearly substantial horizontal exchange, but the overall diversity is again somewhat limited. Regardless of the resolution of such questions about the nature and magnitude of genetic exchange within groups of closely related phages, it can be shown that exchange also happens, albeit at a much lower rate, across the entire range of characterized tailed phages (Hendrix et al., 1999), implying that all tailed phages are partaking of a common gene pool.

Phages acquire genes from, and contribute genes to, not only other phage genomes but also bacterial genomes, and they are thus powerful forces in the evolution, physiology, and pathogenicity of their hosts (Boyd et al., 2001; Dobrindt and Reidl, 2000; Wagner and Waldor, 2002). The role of phages in the virulence of mycobacterial pathogens is unclear and the genomes of *M. tuberculosis* and *Mycobacterium leprae* contain no full-length prophages. However, both of the two sequenced genomes of *M. tuberculosis* contain two small prophage-like elements, ϕRv1 and ϕRv2, at least one

of which (ϕRv1) has an active integration system (Bibb and Hatfull, 2002); both ϕRv1 and ϕRv2 contain capsid genes and could in principle form virus-like particles. However, while all clinical isolates of *M. tuberculosis* appear to contain at least one of these elements, it is unclear whether they play any role in the physiology of the host. Furthermore, the broader question as to what role mycobacteriophages might play in the evolution and physiology of their numerous mycobacterial hosts remains unresolved and largely unaddressed.

We describe here the isolation of ten additional mycobacteriophages and their morphological and genomic characterization. Surprisingly, the diversity of the mycobacteriophage population appears to be substantially greater than what could have been inferred from the four previously sequenced mycobacteriophages. Many of these phages share very few genes with other mycobacteriophages and a very high proportion of genes do not match existing database entries; of those that do match, many are unexpected, not having been identified previously within phage genomes and having no obvious role in viral growth. A comparative analysis of all 14 mycobacteriophages reveals their pervasively mosaic nature and illuminates the underlying evolutionary mechanisms that generate new viruses. Moreover, the presence of genes involved in host responses to bacterial infections as well as autoimmune diseases such as lupus suggests a more central role of bacteriophages in human diseases than previously recognized.

Results and Discussion

Phage Isolation

Ten bacteriophages were isolated from a variety of sources and geographical locations, using *M. smegmatis* as a host, without amplification. Phages were identified as individual plaques on lawns of *M. smegmatis*, purified, amplified, and characterized further. The plaque morphologies of these phages vary considerably. Bxz2 forms turbid plaques—similar to those of L5 and Bxb1—from which stable lysogens can be recovered, but most of the other phages form plaques with a hazy appearance, not obviously either clear (like D29) or turbid. Stable lysogens could not be recovered from these hazy plaques. One of the phages (Cjw1) forms hazy plaques at 37°C but turbid plaques at 42°C from which cold-sensitive lysogens can be propagated. Although *M. smegmatis* is saprophytic, it cannot be assumed that it—rather than one or more other bacterial species—were recent hosts for these phages in their natural environment. Thus, these plaque morphologies are more likely to reflect how these phages interact with *M. smegmatis* rather than revealing fundamental aspects of their viral life cycles.

Particles of these phages were characterized by electron microscopy and viral morphologies compared with those of L5, D29, Bxb1, and TM4 (Figure 1). Seven of the new phages (Bxz2, Che8, Che9c, Che9d, Barnyard, Rosebush, and Cjw1) have similar morphologies to the comparison group, with isometric, icosahedral heads approximately 60 nM in diameter and long flexible tails (a common phage morphology; Ackermann, 2001). Two phages (Che9c and Corndog) have flexible tails but have more unusual prolate heads, and one (Bxz1) has a larger head and a contractile tail. These viral morphologies alone suggest that the mycobacteriophage diversity may be greater than previously indicated, although recent genomic analyses suggest that virion morphology may not be as reliable an indicator of phylogenetic relationships as was once thought (Ackermann, 1987; Lawrence et al., 2002).

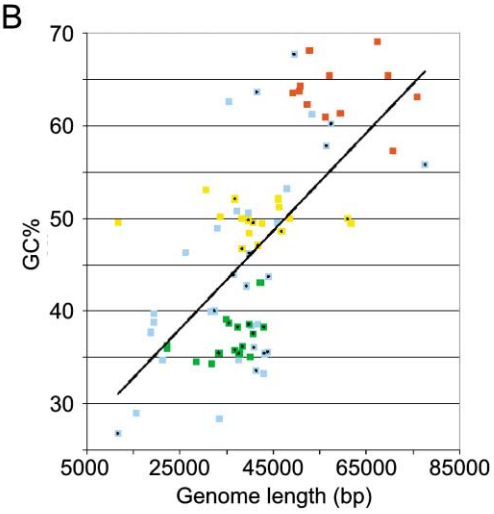
Mycobacteriophage Genometrics

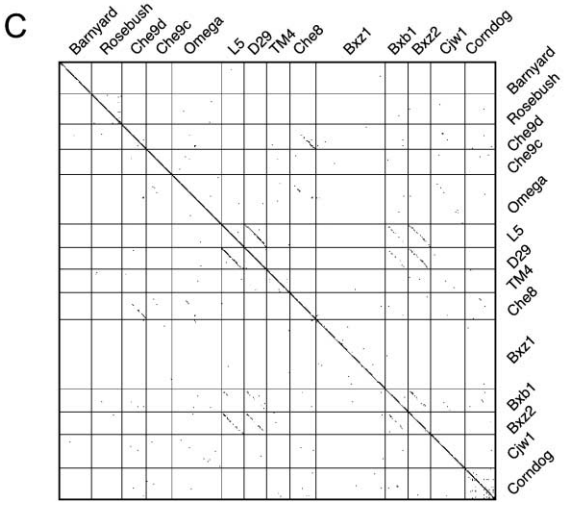
The complete genome sequences were determined for the ten mycobacteriophages and putative genes identified by a variety of bioinformatic analyses. The lengths of these genomes vary considerably, but with the exception of Bxz2 (50.9 kbp), all are larger than the previously sequenced mycobacteriophages (Figure 2A). Moreover, the genome lengths do not fall into discrete size classes, but rather form a continuum between the smallest (49.1 kbp) and the largest (156 kbp). Six of the newly sequenced genomes (Bxz1, Omega, Cjw1, Barnyard, Corndog, and Rosebush) rank among the ten largest completely sequenced phage genomes (the other four are phiKZ [280 Kbp], T4 [169 Kbp], SPBc2 [134 Kbp], and HF2 [78 Kbp]). The reason why mycobacteriophage genomes are among the largest is not clear, although it seems unlikely that growth in a mycobacterial host imposes greater genetic demands than in other bacteria, particularly since genome size (and the total number of genes) varies considerably within the mycobacteriophage group (Figure 2A). Thus, the determinants of phage genome length remain obscure. We note, however, that while mycobacteriophage %GC (which varies from 57.3% to 69%, similarly to their mycobacterial hosts; 57% and 65.6% GC in *M. leprae* and *M. tuberculosis* respectively; Cole et al., 1998, 2001; Figure 2A) does not correlate with their genome lengths, there is a surprising statistically significant positive correlation between %GC content and genome size when all sequenced tailed phage genomes are examined (excluding the 5 phages with genomes larger than 100 Kbp; Figure 2B). Interestingly, the *E. coli*, Dairy, and mycobacteriophages represent discrete groups with nonoverlapping %GC contents and distinct average genome sizes. Whether this relationship among bacteriophages arises from DNA packaging constraints, the effect of nucleotide composition on genome stability or some other parameter remains to be elucidated. Nevertheless, it provides a context for trying to understand why these mycobacteriophage genomes are large relative to other dsDNA phages.

The extent of nucleotide sequence similarity shared among all fourteen mycobacteriophages varies considerably (Figure 2C). Only one of the newly characterized phages (Bxz2) has detectable sequence similarity with L5, D29, and Bxb1. Phages Che9d, Che8, and Omega appear to share limited segments of nucleotide sequence similarity, but the other phages appear unrelated to any others at this level. There is no obvious relationship between geographical origin and sequence similarity, consistent with the generally rapid turnover rates of the phage population as a whole (Garza and Suttle, 1998; Short and Suttle, 1999). We note that the average open reading frame (ORF) length is small (200 codons) relative

A

Phage	Genome Size (bp)	G+C%	tRNAs (#)	Orfs (#)	Av. orf size (bp)
L5	52297	62.3	3	87	601
D29	49136	63.5	5	77	638
TM4	52797	68.1	0	92	574
Bxb1	50550	63.7	0	86	588
Bxz1	156102	64.8	26	225	694
Che8	59471	61.3	0	112	531
Bxz2	50913	64.2	3	86	599
Cjw1	75931	63.1	1	141	546
Corndog	69777	65.4	0	122	572
Che9c	57050	65.4	0	84	671
Omega	110857	61.4	2	237	466
Che9d	56276	60.9	0	111	507
Barnyard	70797	57.3	0	109	650
Rosebush	67480	69.0	0	90	750
Total	979434		40	1659	
Average	69960	63.6	2.9	118.5	599





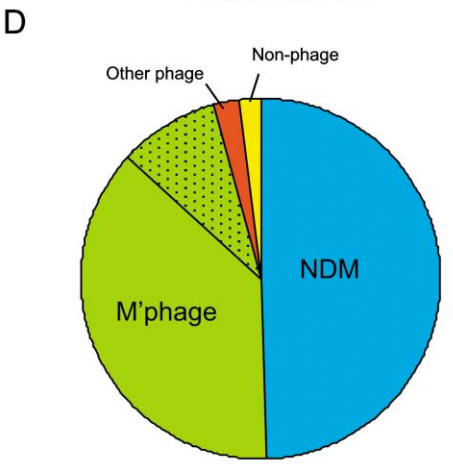


Figure 2. Mycobacteriophages Genometrics

(A) Genome length, %GC, and gene content is shown for 14 mycobacteriophages.

(B) Genome length of dsDNA tailed phages and their %GC content share a positive correlation with an R^2 value of 0.4336. The slope inferred by linear regression does not change significantly if the Dairy phage, mycobacteriophage, or coliphage group is eliminated from the analysis. Symbols are as follows: (red), mycobacteriophages; (yellow) *E. coli* and *Salmonella* phages; (green) Dairy phages; (pale blue) other dsDNA phages. Dots in symbols indicate phages sharing a gene with a mycobacteriophage that is not present in other sequenced mycobacteriophage genomes. The 82 data points represent all completely sequenced dsDNA tailed phages with the exception of 5 with genomes larger than 100 Kbp.

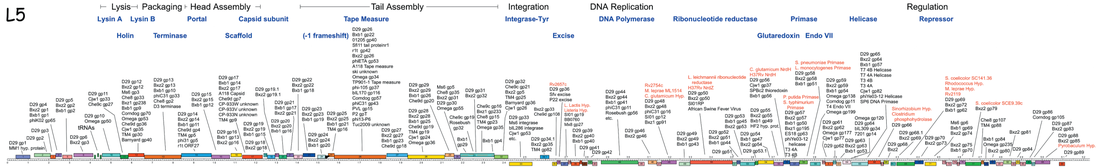
(C) The extent of DNA sequence similarity among the mycobacteriophages is illustrated in a Dotter plot using a sliding window of 25 bp (Sonnhammer and Durbin, 1995).

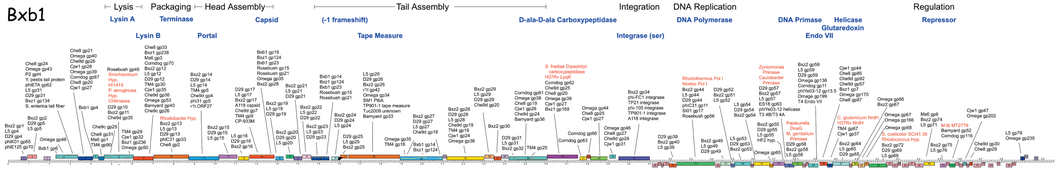
(D) The distribution of all 1659 mycobacteriophage genes according to their database matches. Approximately half (blue) have no database match (NDM), and most of the remainder (green) matches other mycobacteriophage genes; a subset (stippled) of these also has non-mycobacteriophage homologs. Mycobacteriophage genes matching either only other phage genes or only host genes (i.e., not mycobacteriophages) are shown in red and yellow, respectively.

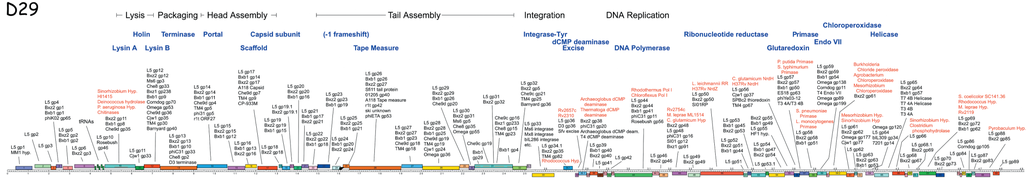
to bacterial genomes (339 codons in *M. tuberculosis*; Cole et al., 1998); this does not appear to arise from either pseudogenes or errant annotation since homologs of many small genes are present elsewhere within other mycobacteriophage genomes (see below). Six of the phages carry tRNA genes and the 26 in Bxz1 (with anticodons corresponding to 15 amino acids and a putative suppressor) are the most identified in any virus (Figure 2A). A more detailed picture of the relationships among these genomes is revealed by comparison of the predicted gene products, and this is described below.

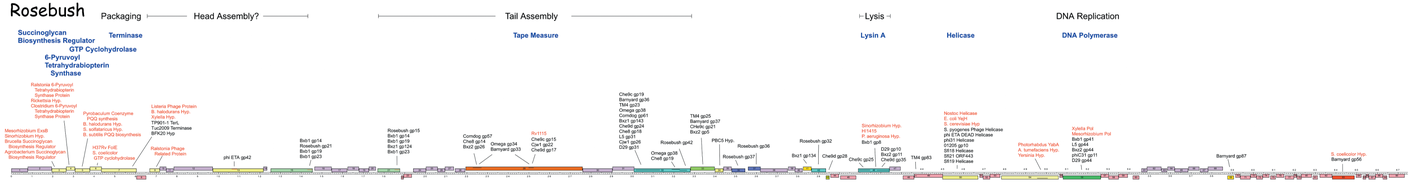
Mycobacteriophage Genome Organization

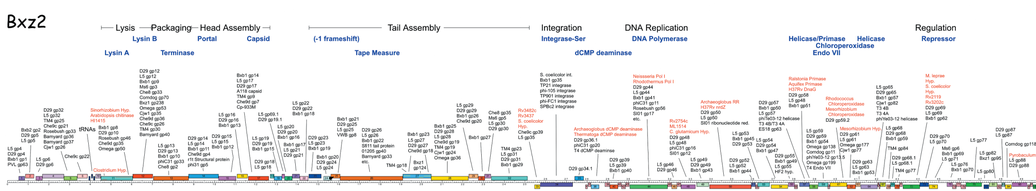
Genome maps of the ten mycobacteriophages along with the four described previously are shown in Figure 3 (see foldout). All fourteen genomes contain high numbers of closely packed protein-coding genes with few intergenic spaces. Generally, these genes are organized into long cotranscribed operons that are presumably expressed from a small number of promoters (Figure 3) as shown previously in L5 (Hatfull and Sarkis, 1993; Nesbit et al., 1995). However, the transcriptional organizations—as assessed by gene order and direction—are

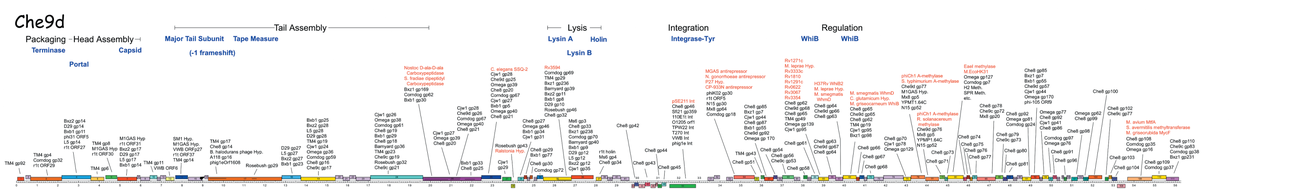

L5
Lysis
Packaging
Head Assembly
Tail Assembly
Integration
DNA Replication
Regulation
Lysin A
Lysin B
Holin
Terminase
Portal
Capsid subunit
Scaffold
(-1 frameshift)
Tape Measure
Integrase-Tyr
DNA Polymerase
Excise
Ribonucleotide reductase
Glutaredoxin
Primase
Endo VII
Helicase
Repressor

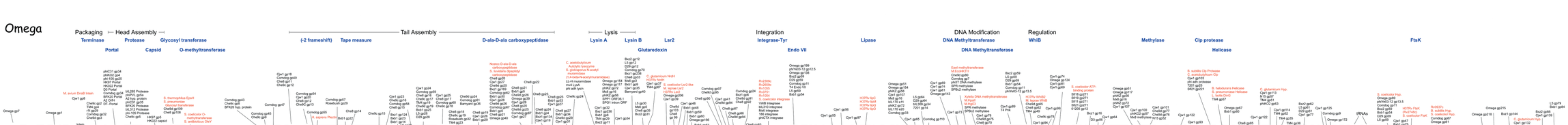

Bxb1
Lysis
Packaging
Head Assembly
Tail Assembly
Integration
DNA Replication
Regulation
Lysin A
Terminase
Capsid
(-1 frameshift)
Lysin B
Portal
Tape Measure
D-ala-D-ala Carboxypeptidase
Integrase (ser)
DNA Polymerase
DNA Primase
Helicase
Glutaredoxin
Endo VII
Repressor


D29
Lysis
Packaging
Head Assembly
Tail Assembly
Integration
DNA Replication
Holin
Terminase
Portal
Capsid subunit
(-1 frameshift)
Lysin A
Lysin B
Scaffold
Tape Measure
Integrase-Tyr
dCMP deaminase
Excise
DNA Polymerase
Ribonucleotide reductase
Glutaredoxin
Primase
Endo VII
Chloroperoxidase
Helicase

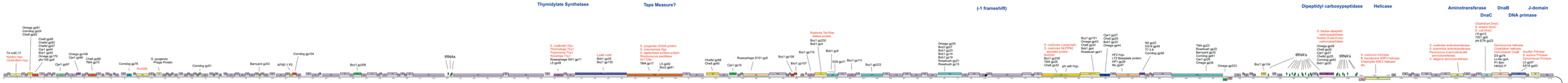

Rosebush
Packaging
Head Assembly?
Tail Assembly
Lysis
DNA Replication
Succinoglycan Biosynthesis Regulator
Terminase
GTP Cyclohydrolase
6-Pyruvoyl Tetrahydrobiopterin Synthase
Tape Measure
Lysin A
Helicase
DNA Polymerase


Bxz2
Lysis
Packaging
Head Assembly
Tail Assembly
Integration
DNA Replication
Regulation
Lysin A
Lysin B
Terminase
Portal
Capsid
(-1 frameshift)
Tape Measure
Integrase-Ser
dCMP deaminase
DNA Polymerase
Helicase/Primase
Helicase
Chloroperoxidase
Endo VII
Repressor

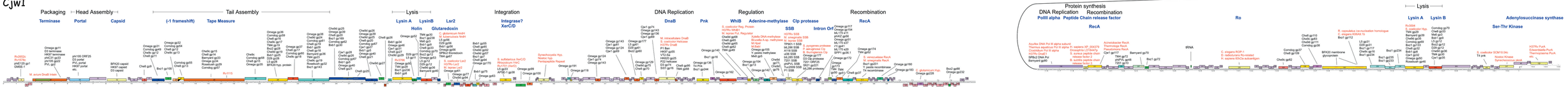

Che9d
Packaging
Head Assembly
Tail Assembly
Lysis
Integration
Regulation
Terminase
Capsid
Major Tail Subunit
Tape Measure
(-1 frameshift)
Portal
Lysin A
Holin
Lysin B
Integrase-Tyr
WhiB
WhiB


Omega
Packaging
Head Assembly
Tail Assembly
Integration
DNA Modification
Regulation
Terminase
Protease
Glycosyl transferase
Portal
Capsid
O-methyltransferase
(-2 frameshift)
Tape measure
D-ala-D-ala carboxypeptidase
Lysin A
Lysin B
Lsr2
Glutaredoxin
Integrase-Tyr
Endo VII
Lipase
DNA Methyltransferase
WhiB
Methylase
Clp protease
Helicase
FtsK

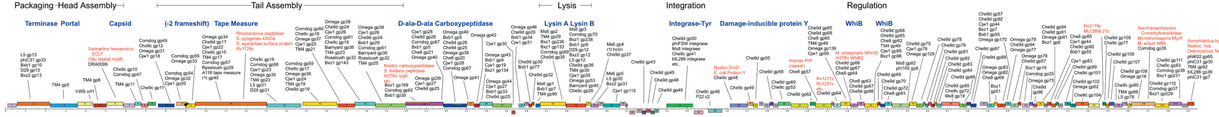
Bxz1



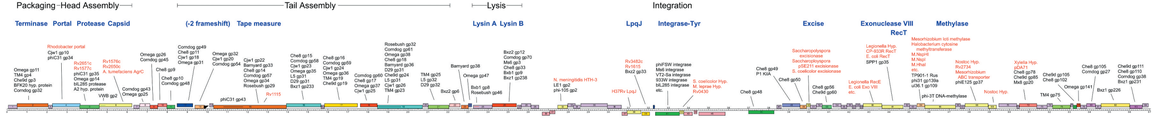
Cjw1



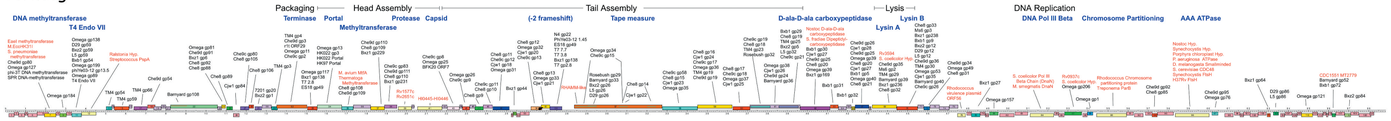
Che8



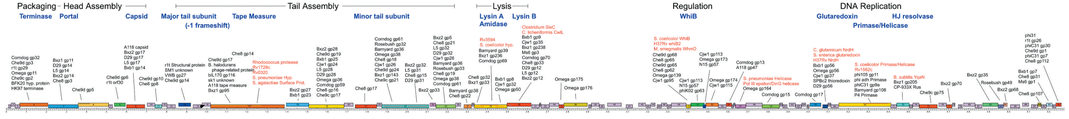
Che9c



Corndog



TM4



Barnyard

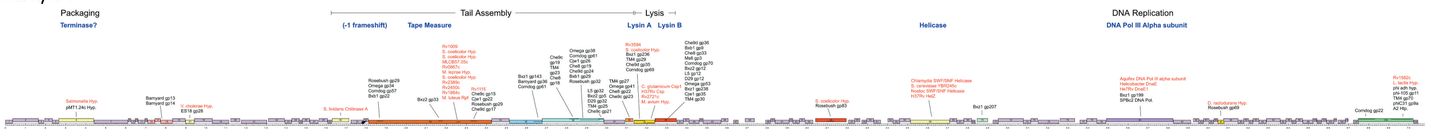


Figure 3. Genome Maps of Mycobacteriophages

Maps of L5, D29, Bxb1, Bxz2, TM4, Che8, Che9c, Che9d, Cjw1, Corndog, Barnyard, Rosebush, Omega, and Bxz1 are shown with numbered markers at 1 kb intervals. Open reading frames are shown as colored boxes above or below each genome, corresponding to rightward and leftward transcription, respectively. Color-coding of gene boxes reflects relationships to other genes: mycobacteriophage homologs are similarly colored; no database match, pale purple and pale pink (transcribed rightward and leftward, respectively); genes with non-mycobacteriophage homologs, pale yellow. Homologs of phage genes and non-phage genes are written in black and red text, respectively, listed in each set according to the degree of similarity with the highest at the top. Probable gene functions are shown in blue text. The maps of L5, D29, TM4, and Bxb1 are re-annotated from those described previously (Ford et al., 1998a, 1998b; Hatfull and Sarkis, 1993; Mediavilla et al., 2000).

highly varied, and apart from the common clustering of structural genes there is no single global architectural feature that characterizes this group of genomes. In many of the newly sequenced genomes (with Bxz1 being the notable exception), a structural gene cluster can be identified which occupies 20–25 Kbp of the genome, similar to that described previously in L5, D29, Bxb1, TM4, and many other dsDNA tailed phages. However, some of these genomes (e.g., Barnyard and Rosebush) share so few genes with the other mycobacteriophages that the identity and organization of the structural genes remains unclear. Nevertheless, a notable consequence of the relatively large sizes of these newly sequenced genomes is that they contain rather large expanses of genes that are likely involved in processes other than assembly and structure of their virions. The functions of these genes thus becomes of primary interest.

Comparative Genomic Analysis

The nucleotide sequences of the ten mycobacteriophage genomes clearly suggest that this group of viruses is substantially more diverse than suggested from the smaller sample sequenced previously. However, the manner in which all fourteen phages are related to each other can be further explored through a comparative analysis of the putative gene products. Using BLAST and PsiBLAST to compare all 1659 ORF's with databases of currently available sequences and a separate database of bacteriophage sequences, the following conclusions emerged. First, there is a high proportion of unique genes ($\sim$50%) that are unrelated to those of other mycobacteriophages or any other previously sequenced organism (Figure 2D); these $>$800 genes outnumber all of the genes without database matches identified in the *M. tuberculosis* genome (Cole et al., 1998). Furthermore, three-quarters of the remaining 50% of genes match only other mycobacteriophage genes such that 87% of all mycobacteriophage genes are unrelated to the gene pool outside of the mycobacteriophages. Clearly, not only must the diversity of the mycobacteriophage population as a whole be substantially greater than represented by these 14 examples, but if this accurately reflects the broader phage population, bacteriophages perhaps represent the biggest unexplored reservoir of sequence information in the biosphere.

Secondly, a high proportion ($>$90%) of the genes with identifiable homologs are shared by one or more mycobacteriophages (Figure 2D), which argues that genetic exchange occurs more frequently within the mycobacteriophage population than between mycobacteriophages and either other bacteriophage or host genomes. This is supported by the observation that a high proportion of genes that do match either host or other phage genes are also found elsewhere in the mycobacteriophage group (Figure 2D), suggesting that acquisition of genes by the mycobacteriophage pool is relatively rare compared to the rate at which they disseminate within the population once acquired. If it is assumed that genes with homologs outside the mycobacteriophages, but which are not present elsewhere within the mycobacteriophage group, are those that have been recently acquired, then it appears as though genes are picked up at similar rates from the host chromosome as they are from other phages (Figure 2D). We recognize, however, that the number of genes examined is still small and a considerably larger group of phage genomes must be examined to verify the relative rates of gene exchange between and within groups of phages.

Thirdly, these phages cannot be phylogenetically ordered into any single hierarchical relationship. For example, when the numbers of genes shared between pairs of genomes are examined (Table 1), it is clear that these phages are related in a reticulate rather than hierarchical manner (Lawrence et al., 2002; Rohwer and Edwards, 2002). Using moderately stringent criteria for similarity, the pattern of pairwise shared genes reveals that this group is not homogenously diverse and that two phage groups emerge, a closely knit one containing L5, D29, Bxz2, and Bxb1 and a more diverse group containing Corndog, Cjw1, Che8, Che9c, Che9d, and Omega; 12 pairs of phages have no closely related genes at all (Table 1). Using more relaxed parameters, it is evident that all phages share at least some genes and some (e.g., Bxb1, Bxz1, and Omega) share a significant number of genes with both groups.

Mechanisms of Phage Evolution and the Origins of Genomic Mosaicism

Previous comparison of the four sequenced mycobacteriophages showed that the genomes share a mosaic relationship (Ford et al., 1998a, 1998b; Hatfull and Sarkis, 1993; Mediavilla et al., 2000). For example, the cognate copies of genes *36* and *38* in L5 and D29 are clearly related (90% and 38% amino acid sequence identity, respectively) but the intervening genes (*37* in L5 and *36.1* in D29) are apparently unrelated. D29 *36.1* encodes a dCTP deaminase that shares 52% amino acid identity with gp20 of *Streptomyces* phage ϕC31 (Smith et al., 1999); L5 gene *37* is of unknown function but its product shares 55% identity with gp19 of Roseophage SI01 (Rohwer et al., 2000). As expected, the ten newly sequenced mycobacteriophage genomes also participate in a mosaic relationship, and the extent of this mosaicism is quite striking. Frequently, each module in the mosaic is a single gene, related at the amino acid level of its product to genes in one or more other phages (Figure 4); module and gene boundaries correspond closely as reported for lambdoid and dairy phages (Brussow, 2001; Juhala et al., 2000). Each of the fourteen mycobacteriophage genomes can thus be thought of as a unique assemblage of individual mosaic modules.

The availability of fourteen highly mosaic mycobacteriophage genomes provides an opportunity for elucidating the evolutionary mechanisms that generate this characteristic mosaicism. There are two alternative models to consider. One explanation is that genetic modules are reassorted by homologous recombination at short conserved boundary (or linker) sequences, as suggested initially by Susskind and Botstein (1978) and proposed to play a role in the evolution of coliphages HK620 and P27 (Clark et al., 2001; Recktenwald and Schmidt, 2002). An alternative model is that illegitimate exchange plays the major role, recombining viral and nonviral DNA molecules in a sequence-independent

Table 1. Mycobacteriophage Shared Genes

	L5	D29	Bxz2	Bxb1	Che9c	Che9d	Che8	Omega	Cjw1	Corndog	TM4	Rosebush	Barnyard	Bxz1
L5	**87**	67	53	36	4	8	5	9	8	5	7	2	2	4
D29	*58*	**77**	51	33	4	7	4	8	5	3	7	4	2	2
Bxz2	*32*	*37*	**86**	37	4	7	3	10	5	5	9	3	4	3
Bxb1	*24*	*26*	*27*	**86**	2	15	12	22	11	7	8	6	3	10
Che9c	*0*	*0*	*1*	*1*	**84**	9	19	15	9	14	5	3	3	4
Che9d	*1*	*3*	*3*	*6*	*2*	**113**	66	1	17	17	12	4	3	8
Che8	*1*	*1*	*1*	*5*	*10*	*44*	**114**	28	22	21	10	3	4	11
Omega	*4*	*4*	*4*	*6*	*6*	*8*	*17*	**239**	49	25	13	2	4	8
Cjw1	*1*	*1*	*1*	*5*	*3*	*6*	*9*	*22*	**143**	12	9	2	3	7
Corndog	*3*	*3*	*2*	*5*	*8*	*11*	*13*	*13*	*4*	**122**	9	3	6	10
TM4	*1*	*1*	*3*	*2*	*0*	*4*	*3*	*3*	*2*	*5*	**92**	3	5	6
Rosebush	*1*	*2*	*2*	*2*	*2*	*1*	*0*	*1*	*0*	*0*	*0*	**90**	5	3
Barnyard	*0*	*0*	*0*	*0*	*1*	*3*	*1*	*1*	*2*	*3*	*2*	*2*	**109**	5
Bxz1	*1*	*1*	*1*	*2*	*2*	*4*	*3*	*2*	*3*	*5*	*3*	*0*	*2*	**229**

Values represent the numbers of genes shared by pairs of mycobacteriophages. The numbers in the top part (in plain type) and bottom parts (in italics) are derived using BLAST cutoff E values of 10^{-4} and 10^{-20}, respectively. Where numbers for reciprocal searches are different the higher number is shown. Self-matches are shown in bold.

manner that generates mostly genomic trash that is either incorrectly sized for packaging into capsids or lacks required genes (Hendrix, 2002). The viable genomes that pass this filter for function and size will retain recombinant boundaries that have had minimal impact on gene function, occurring either at or close to gene boundaries. While both models can account for the generation of mosaic junctions, it is important to note that such junctions will be subsequently reassorted by homologous recombination between flanking sequences.

The four previously sequenced mycobacteriophages provide little evidence to distinguish between these possibilities, since while there is no evidence of linker sequences, it is also impossible to determine where possible illegitimate recombination events could have occurred, obscured by the passage of evolutionary time. The more

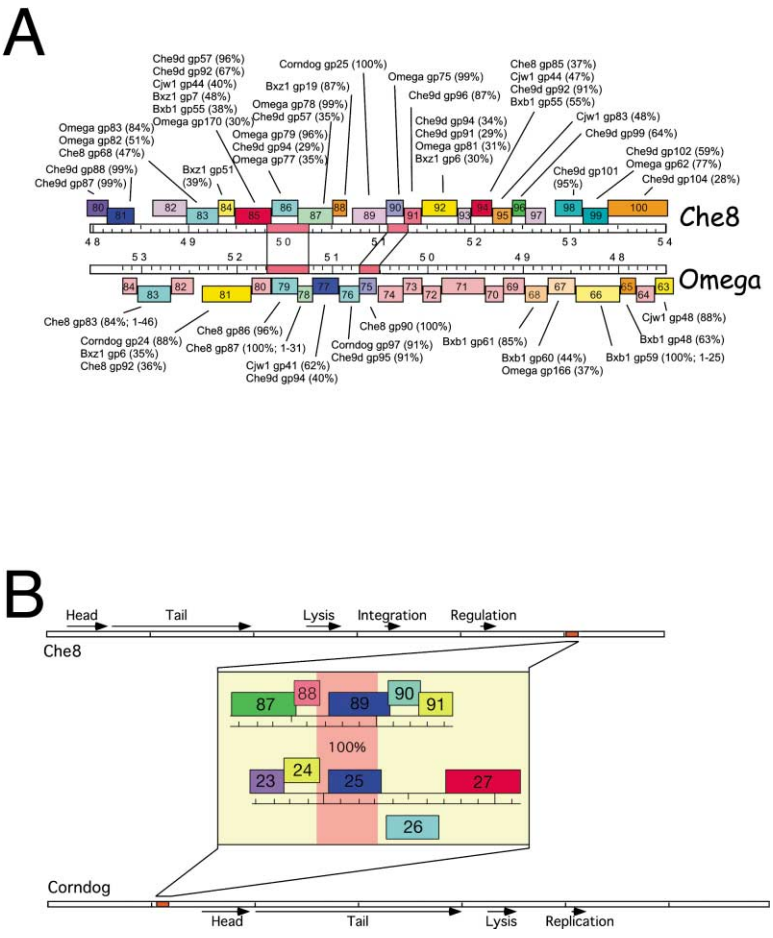


Figure 4. Genomic Mosaicism and Its Origins

(A) Highly mosaic segments of Che8 and Omega genomes showing homologous genes, levels of amino acid identity, and (where appropriate) matching regions. Red genomic segments are shared at >98% nucleotide identity. The leftmost of the four junctions is close to the start codons of genes Che8 *86* and Omega *79*, although weaker sequence similarity extends about 30 bp into the previous genes suggesting that this is the site of recombination; the second junction is 234 bp and 20 bp in from the termination codons of Che8 *87* and Omega *78*. The third junction coincides precisely with the start codons of Che8 90 and Omega 75 and the fourth is located 40 bp in from the start codons of Che8 *91* and Omega *74*.

(B) Part of the Che8 genome shown in (A) is also shared by phage Corndog. The 378 bp shared regions are 100% identical and their joint appearance in these phages must be derived from a recent evolutionary event.

substantial body of genomic evidence provided by the additional mycobacteriophage genomes provides new insights. First, even though a large number of module boundaries can be identified (>500) there is no evidence of linker sequences and this model can thus be excluded as playing a major role in mycobacteriophage evolution. Second, we can identify at least three recombination events that appear to have arisen by illegitimate recombination, having occurred sufficiently recently that the recombinant boundaries are unlikely to have been obscured by further evolutionary events (Figure 4A). Two of these are in a highly mosaic segment of Che8 and contain nucleotide segments shared by phage Omega with greater than 98% nucleotide identity. Three of the four junctions are close to the ends of genes, but do not correspond precisely. The fourth is within Che8 and Omega genes *87* and *78* respectively and results in gene products in which only the N-terminal segment is shared (Figure 4A). The third event—involving phages Che8 and Corndog—is yet more compelling in that the exchange involves quite different parts of the genomes but the common segments are 100% identical (Figure 4B). The recombination event is unlikely to have occurred more than a few thousand years ago (see Experimental Procedures) and the junctions have probably remained unaltered; they are located 28 bp in from the 3′ ends of genes *88* and *24* of Che8 and Corndog respectively, and within the adjacent genes *89* and *25*. The observation that all three of these recent exchange events apparently result from illegitimate recombination, give rise to new module junctions and generate coding sequences, indicates that this is a very creative process and—given the profuse mosaicism—suggests it is a dominant force in genome evolution.

Phage Genes Involved in Bacterial Infections and Human Diseases

Illegitimate recombination can readily account for how phages acquire genes that are not usually involved in viral propagation, but which may be retained if they confer some selective advantage upon their bacterial host. For example, there are many examples of toxin genes present in phage and prophage genomes (Wagner and Waldor, 2002). However, such toxin genes have not been implicated in mycobacterial virulence and none have been identified in the previously sequenced mycobacterial or mycobacteriophage genomes. We have also been unable to find toxin genes in any of the newly sequenced mycobacteriophages. These genomes do, however, contain many unexpected genes that were not previously thought to be phage-encoded; some of which have connections in other contexts to human disease. If these genes move into bacterial genomes with the help of the mycobacteriophages, as appears to be the case for toxin genes in the case of other phages, then bacteriophages may play a larger role in human diseases than previously recognized.

Table 2 shows a list of over fifty genes present in the ten mycobacteriophages sequenced here that were not previously phage-associated, several of which are present in more than one of the mycobacteriophages. In light of the discussion above, it should be noted that since bacteriophages of one host do not evolve independently of those of other hosts, all of these genes also have the potential to reside in phage genomes of other hosts. We also note that the presence of these genes alone does not confirm that they confer a selective advantage; their presence could be transitory and could have arisen from alternative selective forces (such as for an appropriate genome size). Nevertheless, they are intriguing and we will comment further on three that implicate bacteriophages in novel aspects of bacterial pathogenesis and human disease.

First, Cjw1 and Omega encode close homologs (genes *39* and *61*, respectively) of the leprosy and tuberculosis immunodominant antigen Lsr2 that is a potent stimulator of both cellular and humoral immune responses (Laal et al., 1991; Oftung et al., 2000), suggesting a possible role for phages in mycobacterial virulence. The function of the Lsr2 protein is unknown but it seems likely that these phages could influence immune responses of their hosts through the introduction of this gene. Secondly, phage Rosebush contains two genes (*4* and *6*) encoding homologs of enzymes involved in biosynthesis of tetrahydrobiopterin, a cofactor for a key enzyme in the host defense against mycobacterial infections, nitric oxide synthase (Roman et al., 2002; Scanga et al., 2001). Whether these genes promote the synthesis of the cofactor itself or an inactive form of the cofactor is not clear, but these functions have the potential to influence a key player in the host response to mycobacterial infections. Lastly, Bxz1 gp220 encodes a homolog ~35% identical to the human Ro protein, a major target of the autoimmune response in Lupus and Sjogren's diseases (Harley et al., 1992; McCauliffe et al., 1989). The function of the Ro ribonucleoprotein is not known, but it is implicated in 5S RNA processing in *Xenopus*, dauer formation in *Caenorhabditis elegans*, and resistance to UV light in *Deinococcus radiodurans* (Chen et al., 2000; Labbe et al., 2000; Shi et al., 1996). Although bacterial infections have been considered as playing a role in the onset of autoimmune infections, there is a rather poor correlation between these events (Fessler, 2002); the presence of a Ro homolog in Bxz1 raises the possibility that bacteriophages could act in concert with their hosts to stimulate autoimmunity.

Phage Tape Measures as Signaling Molecules

A particularly intriguing finding that emerges from the analysis of these ten mycobacteriophages is that specific phage tail proteins may act as signaling molecules to awaken dormant bacterial hosts. The tail proteins involved are the tape-measure proteins (TMP) found in virtually all phages with flexible noncontractile tails; in coliphage λ, the length of the virion tail shaft is determined by the size of the TMP which spans and measures the length of the tail as an α helix (Katsura, 1987; Katsura and Hendrix, 1984). While TMPs are typically diverse in their sequences, the genes can usually be identified due to their large size (>2000 bp) and their location immediately downstream from a pair of genes expressed via a translational frameshift (Levin et al., 1993; Xu, 2000). The TMP gene homologs for the 13 mycobacteriophages with simple (noncontractile) tails could thus readily be identified (see Figure 3, foldout) and as expected a close correspondence between TMP gene

Table 2. Unexpected Genes in Mycobacteriophage Genomes

Gene Product	Homologue (E value)	Function	Mycobacteriophage Homologues
Barnyard gp30	*S. lividans* Chitinase A (E-6)	Chitinase	
Barnyard gp33	*M. tb* Rv1009 (E-22)	Bacterial cytokine	
Barnyard gp39	*C. glutamicum* Esterase (E-17)	Esterase	
Barnyard gp65	*C. trachomatis* Hel. (E-27)	SWF/SNF Helicase	
Barnyard gp87	*D. radiodurans* Hyp. (0.003)	Unknown	Rosebush gp69
Bxb1 gp30	*S. fradiae* Carboxypeptidase (E-12)	D-ala-D-ala Carboxypeptidase	Corndog gp62, Che9d gp25, Che8 gp20, Omega gp39, Cjw1 gp27, Bxz1 gp169
Bxb1 gp62	*S. coelicolor* Hyp. (E-7)	Unknown	Bxz2 gp72, D29 gp69, L5 gp69
Bxb1 gp72	*M. tb* MT2779 (0.014)	Unknown	Corndog gp116
Bxz1 gp179	*S. coelicolor* Helicase (E-7)	Helicase	
Bxz1 gp188	*S. coelicolor* Transferase (E-4)	Aminotransferase	
Bxz1 gp200	Human Hyp (E-10)	Release Factor	
Bxz1 gp220	*C. elegans* ROP-1 (E-81)	Ro autoantigen	
Bxz1 gp247	Nostoc sp. Kinase (E-6))	Ser-Thr Kinase	
Bxz1 gp250	*C. perfringens* PurA (E-25)	Adenylosuccinase synthase	
Bxz1 gp29	*M. tb* Rv3090 (E-61)	Unknown	
Bxz2 gp33	*M. tb* Rv3482c (E-6)	Unknown	Che9c gp39, L5 gp35
Bxz2 gp85	*P. aerophilum* Hyp. (2.E-6)	Unknown	L5 gp88, D29 gp88
Che8 gp101	*M. tb* Rv2179c (E-24)	Unknown	
Che8 gp49	Nostoc sp. DinD (E-18)	DinD	Che9c gp48
Che8 gp59	Human Prp (E-8)	Prion	Che8 gp57
Che8 gp61	*M. tb* Rv1271c (E-6)	Unknown	Che9d gp64
Che9c gp38	*M. tb* LpqJ (E-17)	LpqJ	
Che9c gp42	*S. coelicolor* Hyp. (E-10)	Unknown	
Che9c gp60	*L. pneomiphila* OrfB (E-13)	ExoVIII	
Che9c gp69	Nostoc sp. Hyp. (E-7)	Unknown	
Che9d gp28	*R. solanacearum* Hyp. (0.005)	Unknown	
Cjw1 gp39	*S. coelicolor* Lsr2 (E-13)	Antigen	Omega gp61
Cjw1 gp59	Synechocystis sp. Hyp. (E-15)	Unknown	
Cjw1 gp95	*S. coelicolor* WhiB (E-12)	Regulator	Che9d gp68, Che8 gp65, Che9d gp65, Che8 gp62, TM4 gp49, Omega gp139
Corndog gp116	*M. tb* MT2779 (E-5)	Unknown	Bamyard gp52, Bxb1 gp72
Corndog gp12	*R. solanacearum* Hyp. (0.002)	Unknown	
Corndog gp35	*M. avium* MtfA (E-100)	O-Methyltransferase	Che8 gp108, Omega gp16, Che9d gp109
Corndog gp39	*M. tb* Rv1577c (0.7)	Unknown	
Corndog gp76	*R. equi* Hyp. (E-7)	Unknown	
Corndog gp84	*S. coelicolor* DnaN (E-8)	DNA Replication	
Corndog gp90	*R. equi* Par protein (E-62)	Partitioning	
Corndog gp96	Nostoc sp. Hyp. (E-56)	AAA ATPase	
D29 gp59.2	*B. capacia* Peroxidase (7.E-13)	Chloride Peroxidase	Bxz2 gp61
D29 gp61	*B. japonicum* Hyp. (E-6)	Unknown	L5 gp61, Bxz2 gp62, Omega gp177, Cjw1 gp77
D29 gp66	*C. glutamicum* Hyp. (E-27)	Unknown	L5 gp66
Omega gp 228	*C. glutamicum* Hyp. (E-9)		Cjw1 gp132
Omega gp105	*M. tb* LipC (E-49)	Lipase	
Omega gp16	*S. thermophilus* EpsH (E-6)	Glycosyl transferase	Che9d gp109, Che8 gp108
Omega gp164	*B. halodurans* Helicase (E-8)	Helicase	TM4 gp57
Omega gp17	*S. olivaceus* Meth. (E-17)	O-methyltransferase	
Omega gp203	*M. tb* FtsK (E-108)	Recombination	
Omega gp206	*M. tb* Rv0937c (E-35)	Unknown	Corndog gp87, Omega gp61
Rosebush gp29	*M. tb* Rv1115 (0.003)	Unknown	Che9c gp15, Cjw1 gp22, Che9d gp17, Barnyard gp33
Rosebush gp3	*M. loti* ExsB (E-45)	Regulation	
Rosebush gp4	*R. solanacearum* Synthase (E-5)	6-Pyruvoyl Tetrahydrobiopterin Synthase	
Rosebush gp5	*P. aerophilum* Hyp. (E-17)	PQQ Biosynthesis	
Rosebush gp54	Photorhabdus YabA (E-24)	phosphopantetheinyl transferase	
Rosebush gp6	*M. tb* FolE (E-27)	GTP Cyclohydrolase	
Rosebush gp83	*S. coelicolor* Hyp. (E-28)	Unknown	Bamyard gp56
TM4 gp17	*R. equi* Peptidase (E-29)	Peptidase	Bxz1 gp95

Genes shown are those not previously identified in phage genomes. This list is not comprehensive and does not include genes that do not have close non-phage homologues but which encode functions that are typically found in phage genomes.

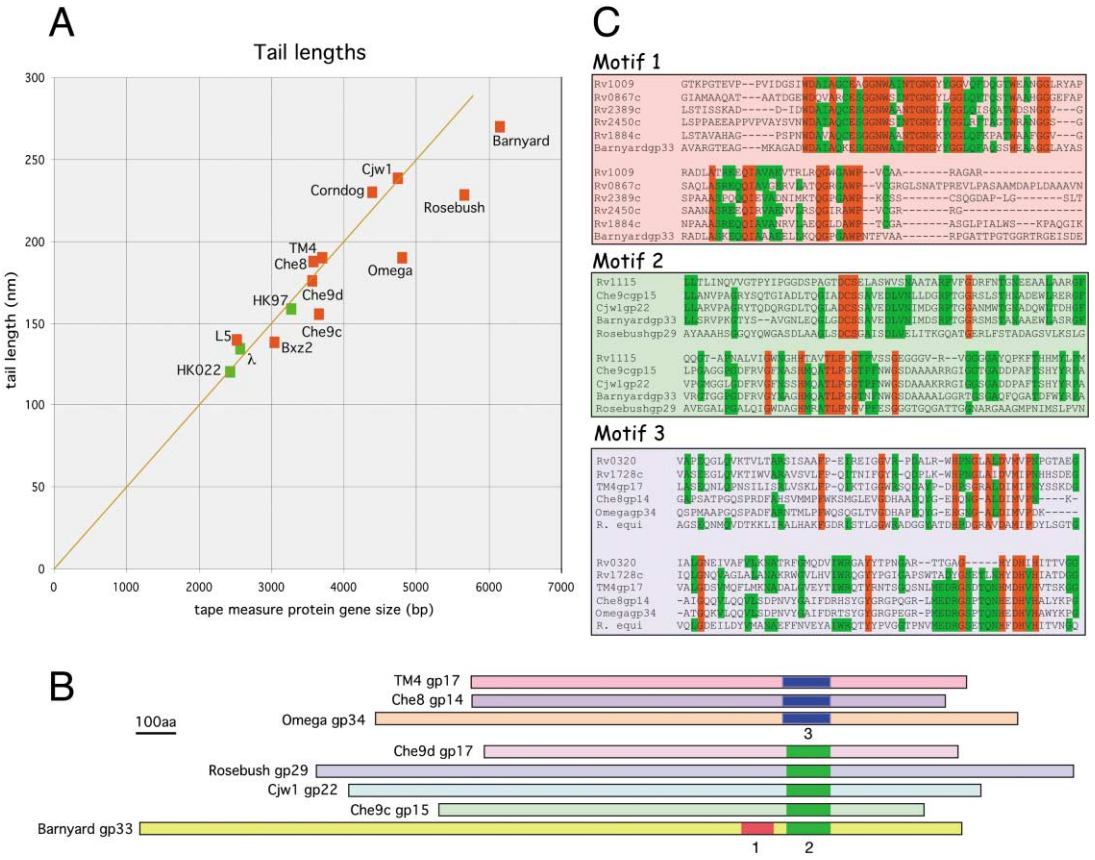


Figure 5. Phage Tape Measure Proteins as Signaling Molecules

(A) Relationship between tail length and tape measure size. Red symbols represent the mycobacteriophages; green symbols represent three *E. coli* phages, including λ. The line shows the dimensions of an α helix (1.5 Å length per amino acid). Error in tail length determination from all sources is estimated to be ±10%.

(B) The tape measure proteins of mycobacteriophages are shown with numbered boxes representing sequence motifs 1, 2, and 3.

(C) Sequence alignments of three motifs in phage tape-measure proteins. Motifs 2 and 3 are less well conserved but do not reflect simply an overall relatedness between the gene products since motif 2 is the only segment of similarity between Rosebush gp29 and Cjw1 gp22.

length and phage tail length is observed (Figure 5A). The relationship is consistent with length determination by an extended α-helical protein, although for some of the phages (Omega, Rosebush, and Barnyard), the TMP is somewhat longer than needed to span the tail as an α helix, suggesting that these proteins may have segments that do not participate in the length determination.

The tape measure protein of phage Barnyard (gp33) contains a 70 residue segment with strong similarity to a family of proteins that includes the resuscitation promoting factor (Rpf) of *Micrococcus luteus*, a secreted protein that promotes regrowth of dormant cells (Figures 5B and 5C) (Kell and Young, 2000). There are multiple copies of related genes in *M. tuberculosis*, *M. leprae*, *Streptomyces coelicolor*, and *Corynebacterium glutamicum* (five, three, three, and four genes, respectively). While these are encoded as small primary polypeptides (typically smaller than 200 aa; Rv0867c is the largest at 407 aa) that are proteolytically processed to their mature forms (Mukamolova et al., 1998), the homologous segment in Barnyard gp33 is embedded within a primary gene product that is over 2000 residues long (Figure 5B). One may infer that this motif signals metabolic activation of dormant mycobacterial hosts during phage infection, facilitating a productive viral infection. During infection, the tape measure protein must be ejected from the tail, presumably entering the bacterial cytoplasm, and is thus ideally suited for this function (Roessner and Ihler, 1984).

Since the vast majority of bacterial cells are in a dormant or nongrowing state in their natural environment (Kell and Young, 2000), the use of such a portable alarm clock presents a clever viral strategy and might be expected to be a more widespread phenomenon. Barnyard is the only sequenced mycobacteriophage to encode this particular motif (motif 1), but other mycobacteriophage tape measure proteins contain other motifs in similar parts of their tape measures (Figure 5). These motifs fall into two groups, one (motif 2) related to *M. tuberculosis* protein Rv1115, and another (motif 3) related to *M. tuberculosis* Rv0320, Rv1728c, and a putative protease of *Rhodococcus equi* (Figure 5C). The functions of these proteins are unknown, but we predict that they may also act in cellular signaling. Since *M. tuberculosis* latency is of clinical importance it is of par-

ticular interest whether these proteins play roles in regulating the growth patterns of this pathogen (Flynn and Chan, 2001; Shleeva et al., 2002).

Concluding Remarks

The isolation and genomic characterization of additional mycobacteriophages reveals that these phages are substantially more varied and diverse than was apparent from the four previously sequenced mycobacteriophage genomes. These mycobacteriophage genomes not only reveal the unrepresentative nature of L5, D29, Bxb1, and TM4, but also show that the overall mycobacteriophage diversity is greater than that represented by this group of fourteen sequenced genomes. Thus, the extent of mycobacteriophage diversity remains unclear and many more genomes must be sequenced before this question can be fully addressed. Furthermore, this situation extends to the phage population as a whole; bacteriophage genomics is clearly in its infancy.

The abundance of phages attests to their evolutionary success, and their enormous diversity results from an environment in which random recombination events generate novel genomes. While horizontal exchange is now recognized to play a central role in the evolution of bacterial genomes (Ochman et al., 2000), the rampant mosaicism of phage genomes illustrates the powerful creativity of this process. Moreover, the variety of genes that participate in this process is large and includes many genes without database matches (and whose functions are unknown) as well as many "bacterial" genes. Some of these genes suggest that bacteriophages may play previously unrecognized roles in bacterial infections and human diseases, influencing cellular and immune responses in addition to conferring virulence determinants. Given the size and diversity of the population, bacteriophages represent a rich and largely uncharted territory for genomic exploration.

Experimental Procedures

Phage Isolation

Phages were isolated from the following sources: Barnyard and Rosebush, Latrobe, PA; Bxz1 and Bxz2, the grounds near the Zebra House and Monkey House pits respectively at the Bronx Zoo, Bronx, NY; Omega and Corndog, Pittsburgh, PA; Cjw1, Edinburg, OH; Che8 was isolated from soil near a cow shed on the banks of Cuum river; and Che9c and Che9d from the soil near the TB clinic of the Tuberculosis Research Centre in Chennai, India. L5 was isolated in Japan in the 1960's (Doke, 1960); D29 was isolated in California in 1954 (Froman et al., 1954); TM4 was isolated in Colorado in 1984 (Timme and Brennan, 1984); and Bxb1 was isolated in Bronx, NY in the 1990's (Mediavilla et al., 2000). Samples of soil, compost, roadside dirt, animal waste, and other sources were extracted with phage buffer and plated directly on solid overlays containing 0.35% agar and *Mycobacterium smegmatis* mc^2155 and incubated at 37°C for 24 hr. Individual plaques were picked and purified.

Genome Sequencing and Analysis

Approximately 10 μg of purified phage DNA was sheared hydrodynamically and repaired, and 1–3 Kbp fragments were inserted into plasmid pBluescript. Individual clones were sequenced using an ABI377 or ABI3100 instrument and assembled (Gordon et al., 1998). At approximately 8-fold redundancy oligonucleotide primers were used with genomic template to generate a single contig and resolve sequence ambiguities. Genome termini were often identifiable as an overabundance of clone ends and a comparison of the sequences generated using primers annealed to ligated and unligated phage DNA provided an unambiguous determination of the ends; no termini could be identified for Barnyard or Rosebush and we assume these have circularly permuted genomes. The Bxz1 genome assembled into a single linear contig with each end having a string of G residues that cause premature termination of sequencing reactions. PCR amplification indicated that there are approximately 150 bp between the unique regions that we assume are all G:C base pairs.

Rates of Divergence

The 378 bp region that is 100% identical in Che8 and Omega contains 87 synonymous sites. Given that synonymous sites in bacterial chromosomal genes with little or no codon usage bias diverge by 3% per million years, a maximal likely separation time would be ~300,000 years. However, the 46 intergenic bases experience far less selection, likely only at the Shine Dalgarno sequences of Corndog *25* and Che8 *89*; since these sites evolve more quickly, both by point mutation and by insertion/deletion, this time of divergence is an overestimate by 10–100-fold, yielding a maximal time of divergence no more than 3,000–30,000 years. A caveat to these estimates is the unknown time spent by bacteriophage particles without replication, making comparisons with bacterial rates of divergence tenuous at best.

Acknowledgments

This work was supported in part by a grant to the University of Pittsburgh by the Howard Hughes Medical Institute (HHMI) in support of Graham Hatfull under HHMI's Professors program. Support was also provided by grants from NIH to R.W.H. (GM5197508), G.F.H. (AI28927), and W.R.J. (AI26170), from the David and Lucille Packard Foundation (J.G.L.), and the Ellison Medical Foundation (W.R.J.). We thank David Gordon and Phil Green for the Phred, Phrap, and Consed assembly programs and Jeffrey Brodsky and Rick Relyea for helpful comments on the manuscript. We also thank Tom Harper and Aleisha Dobbins for help with the illustrations.

Received: November 18, 2002
Revised: February 7, 2003
Accepted: February 20, 2003
Published: April 17, 2003

References

Ackermann, H.W. (1987). Bacteriophage taxonomy in 1987. Microbiol. Sci. *4*, 214–218.

Ackermann, H.W. (2001). Frequency of morphological phage descriptions in the year 2000. Brief review. Arch. Virol. *146*, 843–857.

Bibb, L.A., and Hatfull, G.F. (2002). Integration and excision of the *Mycobacterium tuberculosis* prophage-like element, phiRv1. Mol. Microbiol. *45*, 1515–1526.

Boyd, E.F., Davis, B.M., and Hochhut, B. (2001). Bacteriophage-bacteriophage interactions in the evolution of pathogenic bacteria. Trends Microbiol. *9*, 137–144.

Brussow, H. (2001). Phages of dairy bacteria. Annu. Rev. Microbiol. *55*, 283–303.

Brussow, H., and Desiere, F. (2001). Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol. Microbiol. *39*, 213–222.

Brussow, H., and Hendrix, R.W. (2002). Phage genomics: small is beautiful. Cell *108*, 13–16.

Chen, X., Quinn, A.M., and Wolin, S.L. (2000). Ro ribonucleoproteins contribute to the resistance of *Deinococcus radiodurans* to ultraviolet irradiation. Genes Dev. *14*, 777–782.

Clark, A.J., Inwood, W., Cloutier, T., and Dhillon, T.S. (2001). Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. J. Mol. Biol. *311*, 657–679.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd, et al. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature *393*, 537–544.

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Garnier, T., Churcher, C., Harris, D., et al. (2001). Massive gene decay in the leprosy bacillus. Nature *409*, 1007–1011.

Dobrindt, U., and Reidl, J. (2000). Pathogenicity islands and phage conversion: evolutionary aspects of bacterial pathogenesis. Int. J. Med. Microbiol. *290*, 519–527.

Doke, S. (1960). Studies on mycobacteriophages and lysogenic mycobacteria. J Kumamoto Med Soc *34*, 1360–1373.

Eltringham, I.J., Wilson, S.M., and Drobniewski, F.A. (1999). Evaluation of a bacteriophage-based assay (phage amplified biologically assay) as a rapid screen for resistance to isoniazid, ethambutol, streptomycin, pyrazinamide, and ciprofloxacin among clinical isolates of *Mycobacterium tuberculosis*. J. Clin. Microbiol. *37*, 3528–3532.

Fessler, B.J. (2002). Infectious diseases in systemic lupus erythematosus: risk factors, management and prophylaxis. Best Pract Res Clin Rheumatol. *16*, 281–291.

Flynn, J.L., and Chan, J. (2001). Tuberculosis: latency and reactivation. Infect. Immun. *69*, 4195–4201.

Ford, M.E., Sarkis, G.J., Belanger, A.E., Hendrix, R.W., and Hatfull, G.F. (1998a). Genome structure of mycobacteriophage D29: implications for phage evolution. J. Mol. Biol. *279*, 143–164.

Ford, M.E., Stenstrom, C., Hendrix, R.W., and Hatfull, G.F. (1998b). Mycobacteriophage TM4: genome structure and gene expression. Tuber. Lung Dis. *79*, 63–73.

Froman, S., Will, D.W., and Bogen, E. (1954). Bacteriophage active against *Mycobacterium tuberculosis* I. Isolation and activity. Am J Public Health *44*, 1326–1333.

Garza, D.R., and Suttle, C.A. (1998). The effect of cyanophages on the mortality of synechococcus spp., and selection for UV resistant viral communities. Microb. Ecol. *36*, 281–292.

Gordon, D., Abajian, C., and Green, P. (1998). Consed: a graphical tool for sequence finishing. Genome Res. *8*, 195–202.

Harley, J.B., Scofield, R.H., and Reichlin, M. (1992). Anti-Ro in Sjogren's syndrome and systemic lupus erythematosus. Rheum. Dis. Clin. North Am. *18*, 337–358.

Hatfull, G.F. (2000). Molecular Genetics of Mycobacteriophages. In Molecular Genetics of the Mycobacteria, G.F. Hatfull, and W. R. Jacobs Jr., eds. (Washington D.C.: ASM Press), pp. 37–54.

Hatfull, G.F., and Jacobs, W.R., Jr. (1994). Mycobacteriophages: cornerstones of mycobacterial research. In Tuberculosis: Pathogenesis, Protection and Control, B.R. Bloom, ed. (Washington D.C.: ASM Press), pp. 165–183.

Hatfull, G.F., and Sarkis, G.J. (1993). DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. Mol. Microbiol. *7*, 395–405.

Hendrix, R.W. (2002). Bacteriophages: evolution of the majority. Theor. Popul. Biol. *61*, 471–480.

Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E., and Hatfull, G.F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA *96*, 2192–2197.

Jacobs, W.R., Jr., Barletta, R.G., Udani, R., Chan, J., Kalkut, G., Sosne, G., Kieser, T., Sarkis, G.J., Hatfull, G.F., and Bloom, B.R. (1993). Rapid assessment of drug susceptibilities of *Mycobacterium tuberculosis* by means of luciferase reporter phages. Science *260*, 819–822.

Jones, W.D., Jr. (1990). Geographic distribution of phage types among cultures of *Mycobacterium tuberculosis*. II. Cultures from India and South Africa. Am. Rev. Respir. Dis. *142*, 1000–1003.

Juhala, R.J., Ford, M.E., Duda, R.L., Youlton, A., Hatfull, G.F., and Hendrix, R.W. (2000). Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. *299*, 27–51.

Katsura, I. (1987). Determination of bacteriophage lambda tail length by a protein ruler. Nature *327*, 73–75.

Katsura, I., and Hendrix, R.W. (1984). Length determination in bacteriophage lambda tails. Cell *39*, 691–698.

Kell, D.B., and Young, M. (2000). Bacterial dormancy and culturability: the role of autocrine growth factors. Curr. Opin. Microbiol. *3*, 238–243.

Laal, S., Sharma, Y.D., Prasad, H.K., Murtaza, A., Singh, S., Tangri, S., Misra, R.S., and Nath, I. (1991). Recombinant fusion protein identified by lepromatous sera mimics native *Mycobacterium leprae* in T-cell responses across the leprosy spectrum. Proc. Natl. Acad. Sci. USA *88*, 1054–1058.

Labbe, J.C., Burgess, J., Rokeach, L.A., and Hekimi, S. (2000). ROP-1, an RNA quality-control pathway component, affects *Caenorhabditis elegans* dauer formation. Proc. Natl. Acad. Sci. USA *97*, 13233–13238.

Lawrence, J.G., Hatfull, G.F., and Hendrix, R.W. (2002). Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. J. Bacteriol. *184*, 4891–4905.

Levin, M.E., Hendrix, R.W., and Casjens, S.R. (1993). A programmed translational frameshift is required for the synthesis of a bacteriophage lambda tail assembly protein. J. Mol. Biol. *234*, 124–139.

McCauliffe, D.P., Lux, F.A., Lieu, T.S., Sanz, I., Hanke, J., Newkirk, M.M., Siciliano, M.J., Sontheimer, R.D., and Capra, J.D. (1989). Ro/SS-A and the pathogenic significance of its antibodies. J. Autoimmun. *2*, 375–381.

Mediavilla, J., Jain, S., Kriakov, J., Ford, M.E., Duda, R.L., Jacobs, W.R., Jr., Hendrix, R.W., and Hatfull, G.F. (2000). Genome organization and characterization of mycobacteriophage Bxb1. Mol. Microbiol. *38*, 955–970.

Mukamolova, G.V., Kaprelyants, A.S., Young, D.I., Young, M., and Kell, D.B. (1998). A bacterial cytokine. Proc. Natl. Acad. Sci. USA *95*, 8916–8921.

Nesbit, C.E., Levin, M.E., Donnelly-Wu, M.K., and Hatfull, G.F. (1995). Transcriptional regulation of repressor synthesis in mycobacteriophage L5. Mol. Microbiol. *17*, 1045–1056.

Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000). Lateral gene transfer and the nature of bacterial innovation. Nature *405*, 299–304.

Oftung, F., Mustafa, A.S., and Wiker, H.G. (2000). Extensive sequence homology between the mycobacterium leprae LSR (12 kDa) antigen and its *Mycobacterium tuberculosis* counterpart. FEMS Immunol. Med. Microbiol. *27*, 87–89.

Recktenwald, J., and Schmidt, H. (2002). The nucleotide sequence of Shiga toxin (Stx) 2e-encoding phage phiP27 is not related to other Stx phage genomes, but the modular genetic structure is conserved. Infect. Immun. *70*, 1896–1908.

Roessner, C.A., and Ihler, G.M. (1984). Proteinase sensitivity of bacteriophage lambda tail proteins gpJ and pH in complexes with the lambda receptor. J. Bacteriol. *157*, 165–170.

Rohwer, F., and Edwards, R. (2002). The phage proteomic tree: a genome-based taxonomy for phage. J. Bacteriol. *184*, 4529–4535.

Rohwer, F.L., Segall, A.M., Steward, G., Seguritan, V., Breitbart, M., Wolven, F., and Azam, F. (2000). The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. Limnol Oceanogr *45*, 408–418.

Roman, L.J., Martasek, P., and Masters, B.S. (2002). Intrinsic and extrinsic modulation of nitric oxide synthase activity. Chem. Rev. *102*, 1179–1190.

Scanga, C.A., Mohan, V.P., Tanaka, K., Alland, D., Flynn, J.L., and Chan, J. (2001). The inducible nitric oxide synthase locus confers protection against aerogenic challenge of both clinical and laboratory strains of *Mycobacterium tuberculosis* in mice. Infect. Immun. *69*, 7711–7717.

Shi, H., O'Brien, C.A., Van Horn, D.J., and Wolin, S.L. (1996). A misfolded form of 5S rRNA is complexed with the Ro and La autoantigens. RNA *2*, 769–784.

Shleeva, M.O., Bagramyan, K., Telkov, M.V., Mukamolova, G.V., Young, M., Kell, D.B., and Kaprelyants, A.S. (2002). Formation and resuscitation of "non-culturable" cells of *Rhodococcus rhodochrous* and *Mycobacterium tuberculosis* in prolonged stationary phase. Microbiology *148*, 1581–1591.

Short, S.M., and Suttle, C.A. (1999). Use of the polymerase chain

reaction and denaturing gradient gel electrophoresis to study diversity in natural virus communities. Hydrobiologia *401*, 19–32.

Simon, M.N., Davis, R.W., and Davidson, N. (1971). Heteroduplexes of DNA molecules of lambdoid phages: physical mapping of their base sequence relationships by electron microscopy. In The Bacteriophage Lambda, A.D. Hershey, ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 313–328.

Smith, M.C., Burns, R.N., Wilson, S.E., and Gregory, M.A. (1999). The complete genome sequence of the Streptomyces temperate phage phiC31: evolutionary relationships to other viruses. Nucleic Acids Res. *27*, 2145–2155.

Sonnhammer, E.L., and Durbin, R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene *167*, GC1–10.

Susskind, M.M., and Botstein, D. (1978). Molecular genetics of bacteriophage P22. Microbiol. Rev. *42*, 385–413.

Timme, T.L., and Brennan, P.J. (1984). Induction of bacteriophage from members of the *Mycobacterium avium*, *Mycobacterium intracellulare*, *Mycobacterium scrofulaceum* serocomplex. J. Gen. Microbiol. *130*, 2059–2066.

Wagner, P.L., and Waldor, M.K. (2002). Bacteriophage control of bacterial virulence. Infect. Immun. *70*, 3985–3993.

Westmoreland, B.C., Szybalski, W., and Ris, H. (1969). Mapping of deletions and substitutions in heteroduplex DNA molecules of bacteriophage lambda by electron microscopy. Science *163*, 1343–1348.

Wilhelm, S.W., and Suttle, C.A. (1999). Viruses and nutrient cycles in the sea. Bioscience *49*, 781–788.

Wommack, K.E., and Colwell, R.R. (2000). Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. *64*, 69–114.

Xu, J. (2000). A conserved frameshift strategy in dsDNA long tailed phages, Ph.D. thesis, University of Pittsburgh, Pittsburgh, Pennsylvania.

Accession numbers

Accession numbers for phages are: L5 (Z18946), D29 (AF022214), Bxb1 (AF271693), TM4 (AF068845), Barnyard (AY129339), Bxz1 (AY129337), Bxz2 (AY129332), Che8 (AY129330), Che9c (AY129333), Che9d (AY129336), Corndog (AY129335), Cjw1 (AY129331), Omega (AY129338), and Rosebush (AY129334).