JB
Journals.ASM.org

Evolutionary Relationships among Actinophages and a Putative Adaptation for Growth in *Streptomyces* spp.

Margaret C. M. Smith,[a]* Roger W. Hendrix,[b] Rebekah Dedrick,[b] Kaitlin Mitchell,[b]* Ching-Chung Ko,[b] Daniel Russell,[b] Emma Bell,[c] Matthew Gregory,[d] Maureen J. Bibb,[e] Florence Pethick,[c] Deborah Jacobs-Sera,[b] Paul Herron,[c] Mark J. Buttner,[e] Graham F. Hatfull[b]

Institute of Medical Sciences, University of Aberdeen, Aberdeen, United Kingdom[a]; Pittsburgh Bacteriophage Institute, Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA[b]; Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, United Kingdom[c]; Isomerase, Cambridge, United Kingdom[d]; Department of Molecular Microbiology, John Innes Centre, Norwich, United Kingdom[e]

The genome sequences of eight *Streptomyces* phages are presented, four of which were isolated for this study. Phages R4, TG1, ϕHau3, and SV1 were isolated previously and have been exploited as tools for understanding and genetically manipulating *Streptomyces* spp. We also extracted five apparently intact prophages from recent *Streptomyces* spp. genome projects and, together with six phage genomes in the database, we analyzed all 19 *Streptomyces* phage genomes with a view to understanding their relationships to each other and to other actinophages, particularly the mycobacteriophages. Fifteen of the *Streptomyces* phages group into four clusters of related genomes. Although the R4-like phages do not share nucleotide sequence similarity with other phages, they clearly have common ancestry with cluster A mycobacteriophages, sharing many protein homologues, common gene syntenies, and similar repressor-stoperator regulatory systems. The R4-like phage ϕHau3 and the prophage StrepC.1 (from *Streptomyces* sp. strain C) appear to have hijacked a unique adaptation of the streptomycetes, i.e., use of the rare UUA codon, to control translation of the essential phage protein, the terminase. The *Streptomyces venezuelae* generalized transducing phage SV1 was used to predict the presence of other generalized transducing phages for different *Streptomyces* species.

Bacteriophages are the most abundant and diverse genetic entities on planet Earth. Despite this, it has been proposed that all double-stranded DNA (dsDNA) phages share the same gene pool and that there are clear examples where common ancestry is evident between orthologous gene pairs from very different phages (for example, the capsid-encoding genes from coliphage HK97 and *Streptomyces* phage ϕC31) (1, 2). Understanding the mechanisms through which phages generate such vast diversity is still a challenge, but the drivers of selection and adaptation to growth in different bacterial hosts are clearly strong influences. We are interested in the adaptations of phages that infect the *Actinobacteria*, a class of high-GC, Gram-positive bacteria that have very diverse morphology, physiology, and growth properties.

Actinobacteria, particularly those in the genus *Streptomyces*, have a remarkable capacity to synthesize a diverse array of secondary metabolites of all the major classes. Indeed, 70% of clinically useful antibiotics come from *Streptomyces* spp. Genetic tools derived from *Streptomyces* phages have been invaluable in investigations aimed at understanding the fundamental biology of these bacteria and in manipulating antibiotic pathways (3, 4). Phages were used initially as cloning vectors, but since the early 1990s, integrating plasmid vectors based on phage-encoded site-specific recombination systems have been widely used, enabling the stable insertion of gene constructs into a defined site (*attB*) in the chromosome (5, 6). Moreover, as streptomycetes are mycelial and undergo a developmental cycle, phages that infect this group of bacteria are likely to have undergone host-specific evolutionary adaptations. Sequence databases currently include 223 genomes of actinophages that infect another member of the *Actinobacteria*, *Mycobacterium smegmatis* (see the Mycobacteriophage Database [http://phagesdb.org/]). Comparison of these with the genomes of phages infecting streptomycetes will add insights into the mechanisms of evolution and the nature of phage genomes.

Prior to this work, six *Streptomyces* phage genomes were in the sequence databases, all from double-stranded DNA phages with siphoviral morphology, i.e., long, flexible, noncontractile tails. ϕC31 and ϕBT1 are very closely related, with extensive nucleotide sequence similarity and generally high levels of amino acid identity between homologous gene products (7, 8). ϕSASD1 encodes distant homologues to several of the ϕC31/ϕBT1 early region genes, while VWB, Mu1/6, and the plasmid/phage chimera pZL12 are unrelated to other *Streptomyces* phages (9–12). To gain further insights into the *Streptomyces* phage population, we have sequenced eight more phage genomes. Four of these (R4, ϕHau3, TG1, and SV1) had been isolated previously for use as genetic tools for different *Streptomyces* spp. (13–16). The remaining four, ELB20, Zemlya, Lika, and Sujidade, were isolated for this study. We have also extracted the sequences of five apparently intact prophages from sequenced *Streptomyces* genomes. Together, the analysis of these phages and prophages shows the presence of a dominant phage cluster that includes R4 and six other phage genomes. We present evidence that the R4 cluster phages are related to mycobacteriophages from subclusters A1 and A2, suggesting conservation of an ancestral genome architecture. Adaptations to growth in the streptomycetes were identified, including the prob-

Received 5 June 2013 Accepted 14 August 2013
Published ahead of print 30 August 2013
Address correspondence to Margaret C. M. Smith, Maggie.smith@york.ac.uk.
* Present address: Kaitlin Mitchell, Department of Biology, University of Wisconsin, Madison, Wisconsin, USA; Margaret C. M. Smith, Department of Biology, University of York, York, United Kingdom.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/JB.00618-13.

doi:10.1128/JB.00618-13

able hijacking of the rare leucine codon, UUA, to control the translation of phage genes.

MATERIALS AND METHODS

Isolation of novel *Streptomyces* phages. ELB20 was isolated from unfertilized garden soil from Erskine, Renfrewshire, Scotland, United Kingdom. Phages were extracted from soil in Difco nutrient broth (containing 0.1% egg albumin, pH 8; incubation at 4°C for 16 h) that, after centrifugation and filtration, was plated directly on an agar plate containing 3.2 mM $Ca(NO_3)_2$ and *Streptomyces lividans* strain 1326 as a host (3, 17). Zemlya, Sujidade, and Lika were isolated after enrichment of phages with *S. lividans* spores; 1 g of soil from Houston, TX, was mixed with 10 ml Difco nutrient broth [containing 0.5% glucose and 4 mM $Ca(NO_3)_2$] to which fresh spores of *S. lividans* were added, and the mixtures incubated with shaking (30°C for 24 h). The enriched phages were then separated from soil and bacteria by centrifugation and filtration and mixed with *S. lividans* spores for plating on Difco nutrient agar plates containing 0.5% glucose and 4 mM $Ca(NO_3)_2$. Plates were incubated at 30°C. All phages obtained went through several rounds of single-plaque purification, and a high-titer stock was established.

Sequencing of phage genomes. The phages were purified by banding in a CsCl gradient, and the virion DNA was extracted as described in Kieser et al. (3). ϕHau3 and R4 were both grown on *Streptomyces coelicolor* J1929 (Δ*pglY*) (59), while TG1 and SV1 were prepared using *Streptomyces avermitilis* and *Streptomyces venezuelae*, respectively. ELB20, Zemlya, Lika, and Sujidade were prepared using *S. lividans* 1326 as the host.

Phages R4, ϕHau3, TG1, SV1, and ELB20 were sequenced by dye terminator methodology at the University of Pittsburgh, Department of Biological Sciences. The virion DNA was sheared into 1- to 3-kbp fragments which were then ligated into the EcoRV site of the pBluescript II KS+ vector. Individual plasmids were purified using QiaPrep plasmid purification kits, and these were sequenced from both ends of the inserted DNA by using Applied Biosystems BigDye version 3.0 dye terminator chemistry and universal sequencing primers. Sequences were analyzed using an ABI Prism 3100 DNA analyzer. Approximately 8-fold coverage was obtained; oligonucleotide primers were synthesized and used to prime sequencing reactions with whole-genome templates to provide sequence coverage of underrepresented regions and to fill gaps in the sequence assembly.

Zemlya, Sujidade, and Lika were sequenced by 454 pyrosequencing technology at the University of Pittsburgh Genomics and Proteomics Core Laboratories (GPCL) as described previously (18). The reads were assembled using Consed (58), where the 11-bp overhang ends were apparent.

Extraction of prophage sequences from the genome databases. ϕSAV is a prophage in *S. avermitilis* and was extracted previously using the Fouts PhagePhinder program (19). The prophages Shyg.1, SPB78.1, SrosN11.1, and StrepC.1 were extracted from *Streptomyces hygroscopicus* ATCC 53653, *Streptomyces* sp. strain SPB78, *Streptomyces roseosporus* NRRL11379, and *Streptomyces* sp. strain C genome sequences, respectively (from sequences provided by the Broad Institute). The prophages were detected by BLAST searches with proteins encoded by SV1, VWB, and R4. The endpoints for the prophages were estimated; *attL* was predicted to lie next to the integrase gene, *int*, as the *attP* site is usually located adjacent to *int*. The end containing *attR* was localized to a position where the predicted genes change from encoding many hypothetical proteins (likely to be phage genes) to genes encoding obvious cellular functions (likely to be host genes). In the case of SrosN11.1, the *attR* was predicted to lie after SrosN1_10673, a gene that has no predicted function and few homologues in the databases. In fact, this region contained a 44-bp direct repeat that overlapped a predicted tRNA in the *S. roseosporus* genome and lay in both the predicted *attL* and *attR* regions. This sequence was therefore considered most likely to comprise the core of the attachment sites, so this genome was joined at this repeat and the repeat designated *attP*. The contig encoding prophage SPBP78.1 also had a 42-bp repeat that flanked the predicted prophage genome. In this case, the sequence was identical to the core sequence in the VWB *attP* site (20), and the adjacent integrase (WP_009066700) also has very high identity to the VWB integrase. This sequence was therefore designated *attP* for prophage SPBP78.1. For the remaining prophages (StrepC.1 and Shyg.1), assigning the position of *attR* was done with lower confidence than was the case for *attL*. All the phage/prophage maps were permuted so that they had at the 5′ ends of their top strands either the small or large terminase subunit genes, except for StrepC.1 prophage, where the left end was designated by matching homologues with R4 and ϕHau3 phage. These permutations facilitated a comparison of phage genome organization.

Genome analysis. The analysis of genomes followed a general pipeline: gene identification using Glimmer (21) and/or GeneMark (22) and, in a few cases, manual adjustment gene starts to favor those proximal to a possible ribosome binding site. Nucleotide sequence similarities between phage/prophage genomes were found using Gepard (23) and BLASTn (24). All the phage gene products were analyzed by BLASTp searches against the NCBI nr protein database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and/or tBLASTn against the Mycobacteriophage Database (http://phagesdb.org/) to identify homologues. Finally, general identification of major capsid proteins, tail proteins, endolysins, integrases, putative early and late genes, and regulators was performed, and phage genome maps were exported from DNA Master files (Jeffrey Lawrence, http://phagesdb.org/DNAMaster/). The program Phamerator was used to identify relationships between phage genomes (25). To use the Phamerator programme, all nineteen *Streptomyces* phage and prophage genomes were compiled into a database entitled "Streppys_19." tRNA genes were found with Aragorn (26) and tRNAscan-SE (27). Calculation of the ratio of nonsynonymous to synonymous differences, K_a/K_s, was carried out within the DNA Master program.

Nucleotide sequence accession numbers. Nucleotide sequence accession numbers in GenBank are JX262376 (ϕELB20) for ELB20, KC700556 for Lika, NC_018836 for ϕHau3, NC_019414 for R4, KC700557 for Sujidade, NC_018848 for SV1, NC_018853 for TG1, and KC700558 for Zemlya.

RESULTS AND DISCUSSION

The genome sequences obtained for R4, ϕHau3, and TG1 were consistent with previously reported estimates of genome size and locations of restriction endonuclease recognition sites (Table 1) (14, 16, 28, 29). However, the DNA of the generalized transducing phage SV1 had previously been estimated to be around 45 kbp (29), which is significantly bigger than the 37,612 bp of unique sequence reported here. SV1 virion DNA is therefore likely to be circularly permuted and terminally redundant, which is typical of generalized transducing phages. R4, ϕHau3, TG1, and SV1 all encode an integrase and at least one putative regulator (discussed below), consistent with previous reports that these phages are temperate (13, 14, 16, 29). The new phages, ELB20, Zemlya, Lika, and Sujidade, were all isolated with *S. lividans* as the host from soil of Scotland (ELB20) or Texas (Zemlya, Lika, and Sujidade). These phages all produced a turbid plaque morphology.

Phage genomes cluster largely due to sequence conservation in the packaging/head genes and early region genes. DNA-DNA comparisons performed with both Phamerator and a dot plot between *Streptomyces* phage genomes indicate various degrees of contiguous sequence similarities (Fig. 1; see also Fig. S2 in the supplemental material). Where there is sufficient sequence similarity to detect a diagonal from a dot plot spanning more than 50% of the genome lengths (30), phage genomes can be clustered (Fig. 1). The cluster containing most members is represented by R4 and contains six phage genomes, R4, ϕHau3, ELB20, Zemlya, Lika, and Sujidade (Fig. 1, blue box). Two of these, R4 and ELB20,

TABLE 1 Summary of *Streptomyces* phage and prophage genomes

Phage or prophage	Length (nt)	No. of ORFs	tRNA	%GC	Cluster	Phage isolation host or prophage host genome	Accession no.
Phages							
ELB20	51,160	82	$tRNA^{Phe}$(AGA)	67.0	R4	*S. lividans*	JX262376
Lika	51,252	82	0	65.8	R4	*S. lividans*	KC700556
Mu1/6	38,196	55	0	71.2		*S. aureofaciens* B96	NC_007967
ϕBT1	41,831	55	$tRNA^{Lys}$(AAG)	62.8	ϕC31	*S. lividans*	NC_004664
ϕC31	41,491	53	$tRNA^{Thr}$(ACG)	63.6	ϕC31	*S. lividans*	NC_001978
ϕHau3	50,255	77	$tRNA^{Trp}$(CCA)	67.8	R4	*S. hygroscopicus* 10-22	NC_018836
ϕSASD1	37,068	43	0	66.2		*S. avermitilis* ATCC 31267	NC_014229
pZL12	90,435	112	0	69.5		*Streptomyces* sp. strain 9R-2	NC_013420
R4	51,071	86	$tRNA^{Phe}$(GAA)	67.0	R4	*Streptomyces albus* G	NC_019414
Sujidade	51,552	81	0	65.7	R4	*S. lividans*	KC700557
SV1	37,612	55	0	72.7	SV1	*S. venezuelae*	NC_018848
TG1	40,474	54	$tRNA^{Thr}$(CGT)	64.7	ϕC31	*Streptomyces cattleya*	NC_018853
VWB	49,220	74	0	71.1	VWB	*S. venezuelae* ETH14630	NC_005345
Zemlya	51,077	79	0	65.7	R4	*S. lividans*	KC700558
Prophages							
StrepC.1	51,578	77	0	67.4	R4	*Streptomyces* sp. strain C	NZ_ACEW00000000.1
SPB78.1	51,282	78	0	71.2	VWB	*Streptomyces* sp. strain SPB78	NZ_ACEU00000000.1
Sros11.1	48,399	77	$tRNA^{Gly}$(CCC) with intron	70.3	VWB	*S. roseosporus* NRRL11379	NZ_ABYX00000000.1
Shyg.1	39,945	58	$tRNA^{Val}$(GAC)	70.0	SV1	*S. hygroscopicus* ATCC 53653	NZ_ACEX00000000.1
ϕSAV	41,529	61	0	67.2		*S. avermitilis* MA4690	NC_003155

are almost identical in sequence, despite being isolated about 30 years apart and from geographically different locations (Norwich and Glasgow, respectively). Moreover, phage Bα, isolated in Japan from the culture broth of *Streptomyces lavendulae* S283, has a restriction fragment map almost identical to those of ELB20 and R4, suggesting widespread occurrence of these highly similar phage genomes (31); see further discussion below. A second cluster, represented by ϕC31, contains three closely related genomes, ϕC31, ϕBT1, and TG1 (Fig. 1, red box). The distant relative ϕSASD1 has insufficient nucleotide similarity to be clustered with the ϕC31 cluster and therefore is a singleton (Fig. 1, pink box; see also Fig. S2). Phages VWB, SV1, Mu1/6, and pZL12 did not show nucleotide sequence similarity with other phages.

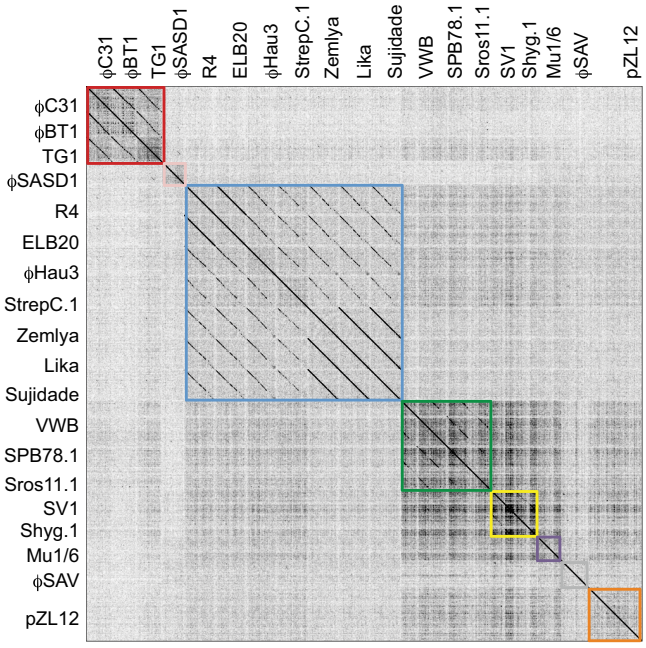

FIG 1 Nucleotide sequence similarities between *Streptomyces* phages and prophages. Dot plot (created with Gepard) of 19 *Streptomyces* phage and prophage DNA sequences. The continuous diagonal line indicates where the phage genomes have aligned with themselves, and the diagonals off center indicate DNA similarity between different phage genomes. Phages with DNA similarity are clustered (indicated by colored boxes).

Using Phamerator and BLASTp searches with the amino acid sequences of predicted proteins from phages R4, SV1, and VWB, prophages residing within recently sequenced *Streptomyces* genomes from the Broad Institute were detected. Prophage StrepC.1 was extracted from *Streptomyces* sp. C, and it clusters with the R4 group; Shyg.1 was extracted from *S. hygroscopicus* ATCC 53653 and is related to SV1; and prophages SPB78.1 and SrosN11.1 were extracted from *Streptomyces* sp. SPB78 and *S. roseosporus* NRRL11379, respectively, and have sequence similarities to VWB. The locations of the prophage endpoints were estimated (see Materials and Methods), and these were joined to form a circle and then reopened at a locus close to the small or large terminase subunit genes to resemble the organization of a typical virion genome (see Fig. S1 and S2 in the supplemental material). The relationships of the extracted prophages StrepC.1, Shyg.1, and SPB78.1/Sros11.1 to R4, SV1, and VWB, respectively, can be detected at the DNA-DNA level (Fig. 1, blue, yellow, and green boxes; see also Fig. S2).

Sequence comparisons between related phage and prophage genomes provide information on which parts of the phage genomes maintain nucleotide sequence similarity when all other sequences have diverged beyond detectable limits. The most-conserved regions observed in pairwise comparisons of phage genomes are packaging/head genes and presumed early region genes, and the most variable regions are between presumed tail and tail fiber genes, recombination and regulatory genes, and orthologues in the variable regions (see Fig. S2 and S3 in the supplemental material). Sequence comparisons with ϕSASD1 proteins

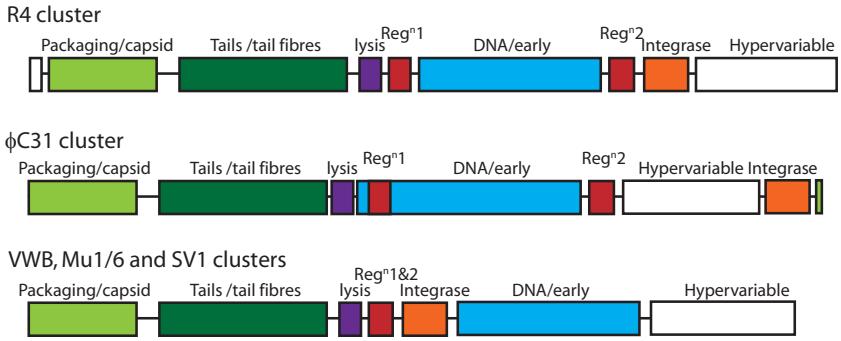

FIG 2 Overview of the genome organization of *Streptomyces* phages. Colored boxes indicate functional clustering of genes within phage genomes, as labeled. Reg, regulation.

are exceptional, as they suggest different ancestry for the early and late regions; early gene products have sequence similarity with several from ϕC31 early proteins (further discussed below; see also Fig. S2), and several of the proposed late products with VWB or SPB78.1 (10). The relationships between ϕC31, ϕSASD1, and VWB are reminiscent of genomic chimeras that have been seen in some other phages, such as *Escherichia coli* phage N15 of the lambdoid phages (32): in N15 the head and tail genes are very similar to those of phage λ, but the early genes clearly have a different origin.

The overall genome organization of all the *Streptomyces* phages is well conserved, comprising the genes for terminase/heads-tails, lysis, regulation, and the early proteins, including many involved in DNA metabolism (Fig. 2; see also Fig. S1 in the supplemental material). All the phage genomes have a region that is much more variable in gene content, a hypervariable region, although this is reduced to a minimum in Mu1/6. The elements that move the most relative to structural, early, and hypervariable regions are the genes encoding integrase (*int*) and the *cos* sites (Fig. 2; see also Fig. S1). *cos* sites are generally close to the terminase genes, but in the R4-like phages, there are several genes between *cos* and *ter*, placing the *cos* sites within hypervariable regions. Although the SV1, VWB, and Mu1/6 genomes do not have detectable nucleotide sequence similarity, they have a similar genome organization as characterized by the positions of their *int* genes (Fig. 2; see also Fig. S1).

Very recent evolutionary events are revealed in very similar genomes. The genomes of phages R4 and ELB20 are more than 96% identical at the nucleotide level, and comparing them provides a particularly clear picture of recent evolutionary events that would be lost in a background of mutational differences in a more highly diverged pair. Figure 3 shows the distribution of single-nucleotide differences and small indels across an alignment of the two genomes. Despite the overall high similarity, it is evident that the degree of difference between the two sequences varies dramatically across the alignment. We can ask, what is the basis of that variation?

The greatest similarity between the two genomes is in the region from nucleotide (nt) 700 to 23,000 (the "left" halves), where there are only 12 nucleotide differences across the whole region. The near identity of this part of the two genomes suggests a very recent common ancestor in evolutionary terms. Starting at about coordinate 23,500 through the *cos* ends to nucleotide 700 (the "right" halves), there are appreciable, though still moderate differences between the two genomes, and the levels of those differences vary across the region. The common ancestor(s) of the right halves of the genomes must be more distant than the common ancestor of the head and tail genes in the left halves of the genomes, because of the significant differences in the number of point mutations accumulated in the two regions. In other words, the genomes are mosaic with respect to each other, with a junction (at about nucleotide 23,000) separating the two halves.

We can then ask whether the different levels of mutational differences between regions within the right halves of the genomes also reflect mosaicism or, alternatively, reflect different rates of mutational accumulation. If a gene were experiencing positive selection, for example, we would expect that mutations would be retained more frequently and so accumulate more rapidly than for a gene experiencing purifying selection. The ratio of nonsynonymous to synonymous differences, K_a/K_s, between homologous genes provides an indication of whether a gene is under positive selection (in which case the K_a/K_s ratio is >1). There are three pairs of orthologous genes in the R4/ELB20 comparison, R4 genes *46*, *53*, and *56*, that are different enough to allow calculation of a meaningful K_a/K_s ratio. For these, the K_a/K_s ratio ranges from 0.23 to 0.36, indicating that they are experiencing purifying selection. This conclusion is incompatible with the hypothesis that R4 g46, g53, and g56 and the ELB20 homologues are more diverged than the others in the right arms of these two genomes because they have been accumulating mutations faster. Rather, it suggests that the differences in divergence represent mosaicism resulting from recombination. This pattern is reminiscent of the mosaicism reported in comparisons of genomes that are much more extensively diverged than R4 and ELB20, including the relative lack of mosaic junctions among the virion structural genes. However, in those cases, analysis of the sequences implies that the mosaic junctions arose from nonhomologous recombination (2), whereas in the R4/ELB20 comparison, it seems plausible that most or all of the mosaic junctions arose through homologous or near-homologous recombination.

The R4-like phage cluster is related to the mycobacteriophages of the A1 and A2 subclusters. Among the actinobacteria, the only other genus for which a large number of phage genomes have been sequenced and analyzed is *Mycobacterium*. At the time of writing, there are 223 mycobacteriophage genomes in public databases. There are a substantial number of proteins with recognizable homologues encoded in both *Mycobacterium* and *Streptomyces* phage genomes. These shared proteins include not only

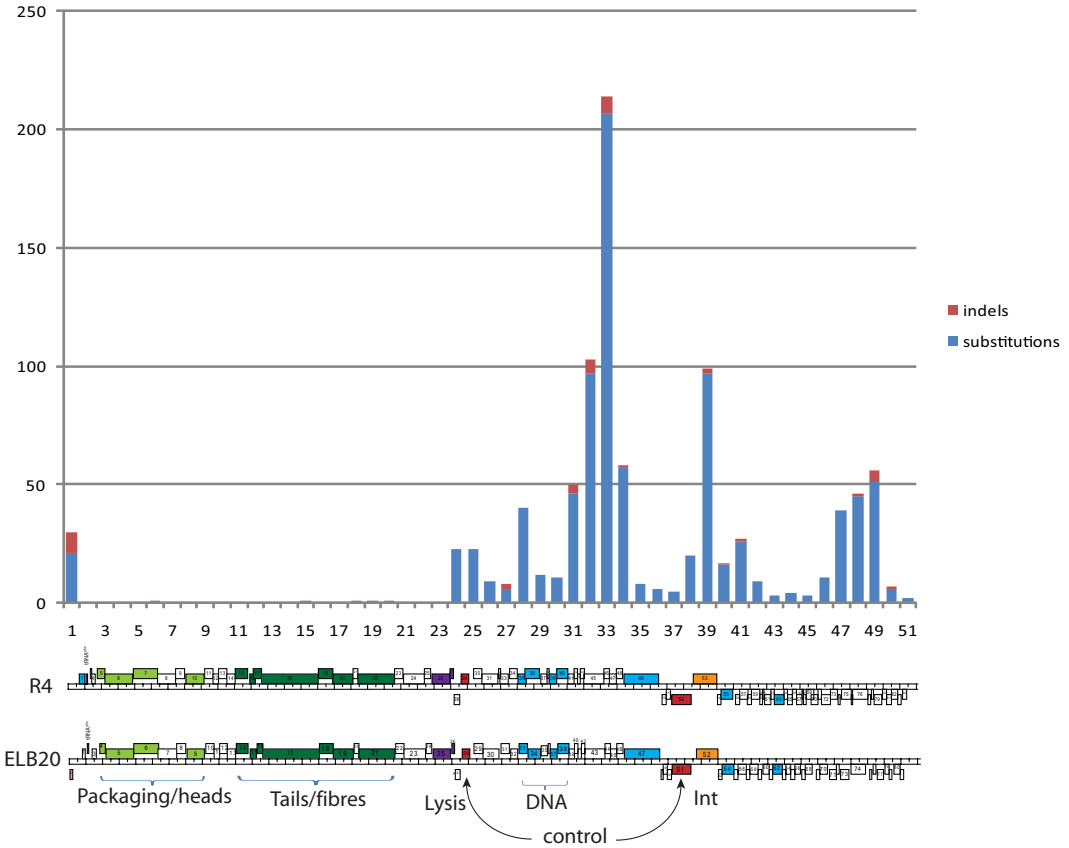

FIG 3 Plot of the numbers of nucleotide substitutions and indels (*y* axis) in consecutive 1,000-bp windows (*x* axis) in a DNA alignment of R4 versus ELB20. The genome maps of R4 and ELB20 are indicated below the *x* axis.

those with conserved functional motifs, such as proteins used in DNA packaging (e.g., terminase and portal protein) and DNA replication (e.g., primases and DNA polymerases), but less-well-conserved proteins with no known homologues outside the actinophages. R4 gp25, for example, has a domain of unknown function (DUF2746) that is only present in the actinophages (Fig. 4). Other conserved hypothetical proteins encoded in both *Streptomyces* and mycobacteriophages are represented by R4 gp31, gp37, and gp45 and Zemlya gp70.

Conservation of gene order (i.e., synteny) likely also reflects common ancestry. There is evidence of shared synteny between the R4-like phages and the subcluster A1 and A2 mycobacteriophages, illustrated by alignment of the R4 and mycobacteriophage Rockstar genomes (Fig. 4). R4 and Rockstar share 16 pairs of re-

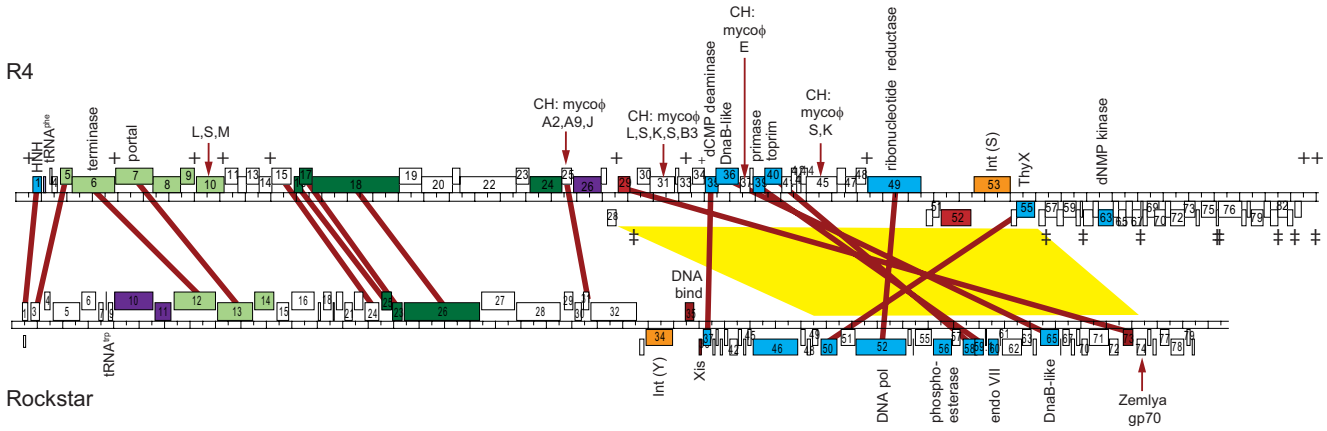

FIG 4 Synteny between R4 and mycobacteriophage Rockstar. The red lines indicate orthologous gene products that lie in the same order in both genomes, taking into account a proposed inversion that has occurred in an ancestral genome (yellow area). The positions of the putative stoperator sequences are shown on the R4 genome: + indicates a sequence on the plus strand and ‡ on the minus strand. pol, polymerase.

lated proteins (detected through BLASTp searches), albeit at different levels of amino acid similarity, distributed over most of the lengths of their genomes (Fig. 4). There may be additional homologues that have diverged sufficiently to fall below BLASTP cutoff values of significance. All of the gene pairs in the left halves of the genomes are syntenous, but this is not surprising given the widely conserved synteny of virion structural and assembly genes shared by all phages with siphoviral morphologies. However, except for one pair, the shared homologues in the right halves of the genomes also have shared syntenies, notwithstanding a large inversion in one lineage relative to the other (corresponding to the yellow region in Fig. 4). One end of this inversion could conceivably be adjacent to the homologous repressor genes, *29* in R4 and *73* in Rockstar, as Zemlya, a phage closely related to R4, has a homologue of Rockstar *74* (i.e., Zemlya *70*) that is not part of the inverted region.

The exception to the syntenous layout of Rockstar and R4 is the dCMP deaminase genes (Fig. 4). Genes coding for dCMP deaminase are sporadically distributed among phage genomes (for example, subcluster A2 mycobacteriophage Pukovnik lacks this gene), and their locations show no consistent relationship to the other genes in the phage genomes in which they appear. Other notable exceptions to the shared synteny of R4-like phages and cluster A mycobacteriophages are the positions of the lysis genes and *int* genes, and there may have been advantageous adaptations associated with the current locations of these functions. In the R4-related phages, it is also striking that about 25% of the DNA is transcribed right to left (reminiscent of the cluster A mycobacteriophages, in which about 50% of the genes are transcribed right to left), whereas, with the exception of pZL12, the other *Streptomyces* phage genomes have a much stronger bias toward left to right (as shown) transcription (see Fig. S1 and S2 in the supplemental material).

Similarly, there are six recognizable homologues shared by phage VWB and mycobacteriophage Marvin, and these also share syntenies (see Fig. S4A). The conservation of synteny is not evident in all genome comparisons, as seen by the alignment of the *Streptomyces* phages R4 and ϕC31 (see Fig. S4B). Thus, phages infecting the same bacterial host can be more diverged with respect to their genome organizations than genomically distinct phages that infect different hosts. Genome similarities between phages infecting different bacterial genera have been seen previously, including the Mu-like phages, P2-like phages, and the lambdoid phages. Furthermore, we note that the *Rhodococcus equi* phage ReqiPine5 and mycobacteriophage Rosebush (cluster B bacteriophage) share multiple homologous gene pairs located syntenously along their genomes, indicating that these phages are likely to be related (33).

The sequence similarity between the *Mycobacterium* and *Streptomyces* phage proteins is detectable only at the level of the encoded amino acid sequence, arguing that genetic exchange between these two groups of phages is either infrequent or spans a broad period of evolutionary time. Host specificity normally interferes with phages with hosts from phylogenetically different genera undergoing frequent recombination, but Jacobs-Sera et al. (34) showed that transfection of Zemlya genomic DNA into electrocompetent *M. smegmatis* and plating the transfectants in the presence of *S. lividans* spores can yield plaques at low frequency. Thus, if the barrier to injection of phage DNA into a noncognate host can be overcome, a phage growth cycle might occur, with the possible establishment of novel phage genome architectures in different host genera (34). Alternatively, if the shared features of the R4-like and mycobacteriophage A1 and A2 subcluster phages do indeed reflect common ancestry, then the ancestors preceded the migration between the *Streptomyces* and *Mycobacterium* hosts. The evolutionary processes that maintain specific genome architectures within a milieu of rampant recombination events, both homologous and illegitimate, remain unclear. However, the comparison between R4 and ELB20 (discussed above) suggests that purifying selection is a strong evolutionary force and is likely to play a role in maintaining phage clusters.

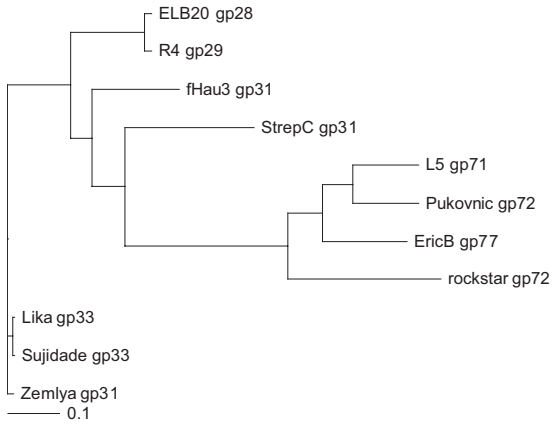

FIG 5 The R4 cluster *Streptomyces* phages encode an L5-like repressor protein. An alignment of all the R4 family putative repressors with the L5 and Pukovnik repressors was performed with ClustalW, and the tree plotted with TREEVIEW.

Conservation of a repressor-stoperator regulatory system. The relationship between the cluster A mycobacteriophages and the R4-like *Streptomyces* phages is strengthened by conservation of the regulatory mechanism that silences phage transcription during lysogeny. The head-tail-lysis genes and the putative early region genes (which include the major DNA metabolism genes) are separated by a conserved regulator gene with sequence similarity to the mycobacteriophage L5 repressor (Fig. 5). The observation that the R4 cluster phages encode homologues of the mycobacteriophage L5 repressor was followed up with a search for possible repressor-binding sites. The L5 repressor acts by binding to 24 short (13-bp) directional sequences located across the genomes, usually in intergenic regions (35); this is conserved among all of the cluster A phages (35). The repressor binds to these sequences and blocks the movement of RNA polymerase, and hence, they are called stoperators. A feature of the stoperators is that they only block RNA polymerase movement in one direction and they are correspondingly oriented on the phage genome depending on the direction of transcription (34). Multiple candidate stoperator sequences were discovered in the R4, ELB20, ϕHau3, and Zemlya genomes and are both similar in length to the L5 stoperators and mostly intergenic (Fig. 4; see also Table S1 and Fig. S1 in the supplemental material). The stoperator consensus sequences are similar in the *Streptomyces* phages, and all, including those from L5, have a highly conserved 5′ GTG trinucleotide (Table 2) (35, 36).

Regulatory signals in ϕC31, ϕBT1, TG1, and ϕSASD1. The regulatory signals in ϕC31 have been characterized previously (37–40). The repressor-binding sites (conserved inverted repeats,

TABLE 2 Consensus sequences of stoperator-like sequences in R4 family phages

Phage	Stoperator-like sequence[a]
R4	tGTGcAAGTGtngC
ΦHau3	agtGTGttAGncngac
ELB20	tgtGcAAgtgtngca
Zemlya	gTgTGcaAccTtcgta
L5	GGtGgcTGTcAAg

[a] Uppercase and lowercase indicate 100% and >50% conservation of the base, respectively. Data for L5 are from Brown et al. (35).

CIRs), the phage-specific promoters, and the terminators are all well conserved in ϕC31, ϕBT1, and TG1 (see Fig. S5 and S6 in the supplemental material). Previous work has shown that three repressor proteins, all encoded by the *c* gene in ϕC31, bind the CIR sequences (38, 41). The three in-frame but N-terminally different repressors all contain a putative helix-turn-helix (HTH) DNA-binding motif close to the C terminus (42, 43). The predicted repressor proteins from ϕC31, ϕBT1, and TG1 are noticeably diverged in their N-terminal halves and much more highly conserved in the C-terminal regions, consistent with the HTH DNA-binding motif contacting the conserved CIR sequences. The promoters and terminators in ϕC31 have also been mapped (39, 40). The consensus sequences for phage promoters in ϕC31 ϕBT1, and TG1 differ by only a few base pairs, and the positions of the regulatory signals are conserved (see Fig. S1 and S6). Even where there has been an indel, the regulatory elements are either coinserted or codeleted, suggesting that genetic exchange may normally be with another phage of the ϕC31 family (see Fig. S1).

The DNA region between the ϕSASD1 gp43 and gp1 lacks a repressor gene homologue that is present in the equivalent regions in ϕC31, ϕBT1, and TG1. However, ϕSASD1 encodes a homologue, gp13, of the putative activator of the phage-specific promoter in ϕC31, gp12 (10), and this prompted us to look for sequences with similarity to CIRs and phage-specific promoter sequences in ϕSASD1. Knowing that the phage repressor from the ϕC31 group binds to CIRs located upstream from ϕC31 gp12 (38), we looked for a CIR sequence upstream from gp13 in ϕSASD1. A 16-bp CIR perfect palindrome was identified and then used to search the whole of ϕSASD1. In addition to the CIR upstream from gp13, there are 4 other CIRs located in the right arm of the ϕSASD1 genome (see Fig. S5 in the supplemental material). We also tried to identify a putative phage-specific promoter in ϕSASD1, and an 18-bp sequence upstream from both the DNA polymerase gene (*g12*) and the putative activator gene (*g13*) was found (see Fig. S6). We do not know whether either of these two sequences is functional. Although ϕSASD1 appears to contain CIR sites, none of the proteins encoded are predicted to be repressors. Furthermore, ϕSASD1 encodes a truncated and probably nonfunctional serine integrase containing only part of the C-terminal domain and having lost the entire N-terminal catalytic domain. Taken together, these observations further support the idea that the ϕSASD1 genome is a mosaic, with the right half derived from a temperate phage related to the ϕC31-like phages and the left half derived from a VWB-like phage (suggested above). Possibly ϕSASD1 evolved rapidly to become a lytic phage during the failed fermentation that led to its isolation (10).

Adaptations to growth in *Streptomyces* spp. (i) A putative mechanism to control the translation of phage proteins using the rare codon UUA. Terminase is an essential protein for phage replication, as it provides the motor and DNA cleavage activity required for packaging the virion genomes into the phage capsids. We noticed that ϕHau3 has three consecutive genes (*g6*, *g7*, and *g8*) encoding products that each show similarity to the large terminase proteins from other phages in BLASTp searches. Closer examination revealed that the three proteins encoded by these genes (gp6, gp7, and gp8) were almost identical to the beginning (amino acids 1 to 100), the middle (amino acids 101 to 200), and the end (amino acids 201 to 553) of the putative large terminase encoded by R4 *g6* (Fig. 6A; see also Fig. S7 in the supplemental material). Thus, it appears that the large terminase subunit of ϕHau3 is, unusually, encoded by three separate genes. The sequence of the ϕHau3 in this region was verified, indicating that this genetic fragmentation of what is a single open reading frame (ORF) in almost all other phage genomes was not due to sequencing errors. Examination of the genes encoding the large terminase of other phages and prophages revealed a second example of a split terminase ORF in the StrepC.1 prophage, encoded by StrepC.1 *g8* and *g9* (nt 41080 to 42729 in *Streptomyces* sp. strain C contig ACEW01000274.1) (Fig. 6A; see also Fig. S8). Remarkably, the position at which the alignment between R4 gp6 and the StrepC.1 gp8 ends is exactly the same as the endpoint between ϕHau3gp6 and R4gp6, indicating that the fragmentation of the terminase genes in both ϕHau3 and StrepC.1 has occurred at the same place.

In both ϕHau3 and StrepC.1, a contiguous sequence with near identity to the entire R4 terminase could be generated by the introduction of a ribosomal shift to another ORF and the skipping of either 8 or 10 nucleotides of coding sequence to join ϕHau3 gp6 and gp7, ϕHau3 gp7 and gp8, and the StrepC.1 gp8 and gp9 (Fig. 7). While the genetic fragmentation of the terminase gene in StrepC.1 could be due to an evolutionary drift toward a defective prophage, this cannot be true for ϕHau3 because the sequenced virion DNA was prepared from an infected *S. coelicolor* lysate. How then can the expression of these fragmented terminases produce an active terminase? Possibly the three polypeptides produced by ϕHau3 could be protein domains that fold separately and interact to regenerate an active terminase. Large terminase proteins generally consist of two domains, the N-terminal domain that provides the ATPase motor for translocating the DNA and the C-terminal domain for cutting and unwinding the *cos* ends (44). The conserved motifs in the ATPase domain have been well characterized and comprise four conserved sequences; the adenine-binding motif, the Walker A and B boxes, and an ATPase coupling motif, or C motif. Alignments of P22, T4, and λ terminase proteins with R4 terminase failed because they are too highly diverged. Instead, we aligned R4 and Zemlya terminase proteins with the related L5 gp13 and Pukovnik gp14 proteins (both cluster A mycobacteriophages) to identify conserved amino acids and then looked by eye for candidate functional motifs located at the appropriate distances (Fig. 6B). The positions of the discontinuities in the large terminase proteins in ϕHau3 and StrepC.1 prophage were found to map within the most likely Walker A box motif and immediately adjacent to the Walker B box. It seems very likely, therefore, that these breaks do not mark the boundaries of protein domains and that a continuous polypeptide must be generated for a functional terminase.

Formally, there are two ways to generate a continuous poly-

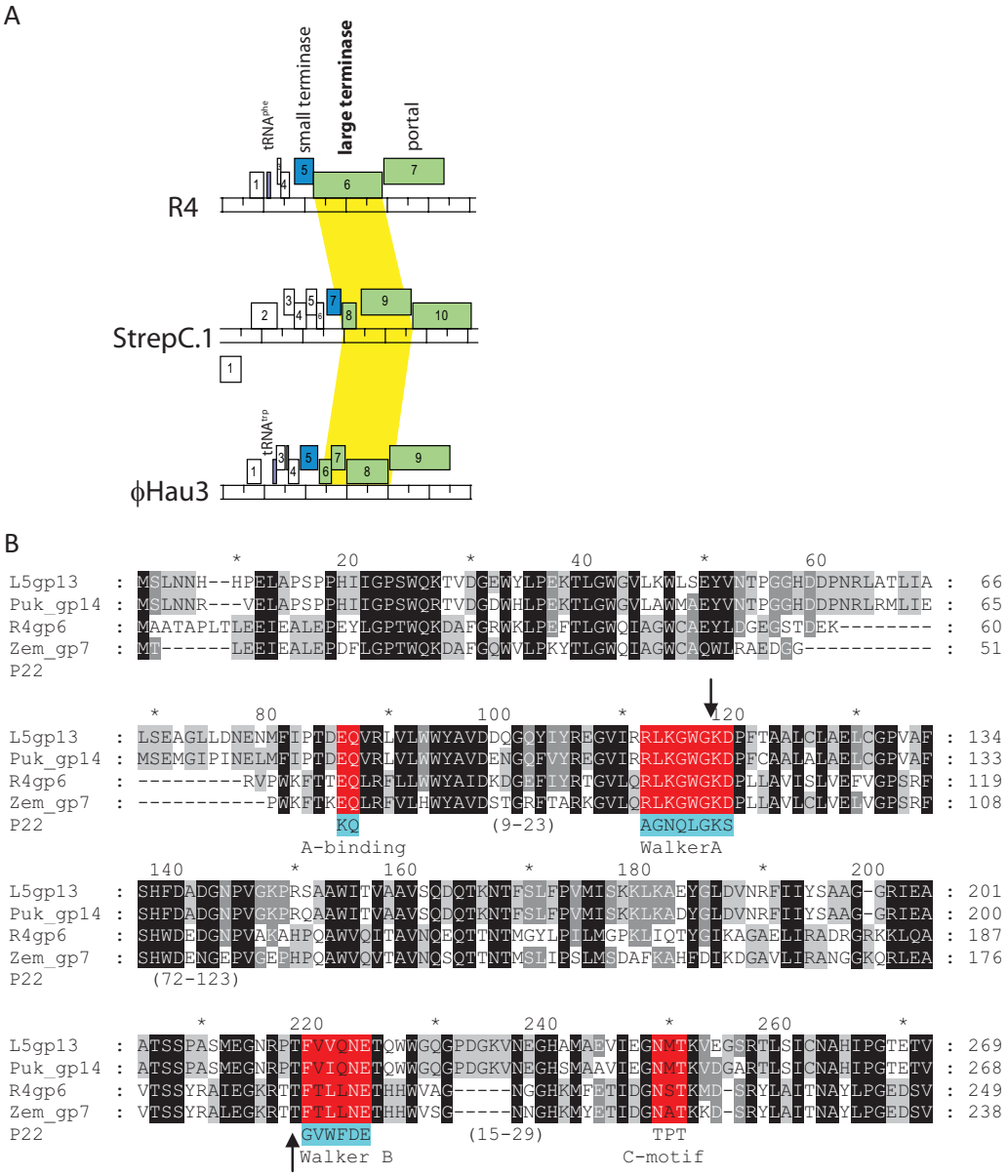

FIG 6 Discontinuous large terminase genes in ϕHau3 and StrepC.1. (A) Organization of the genes encoding the large terminase proteins in R4, StrepC.1, and ϕHau3 as determined by the gene calling programs Glimmer and GeneMark (R4 and fHau3) and from the Broad Institute annotation (StrepC.1). The three terminase-encoding genes in ϕHau3 are in different reading frames, as are the two in StrepC.1. (B) An alignment of the N-terminal regions of the large terminase proteins from R4, Zemlya, L5, and Pukovnic indicates the positions of conserved sequences, including the likely Walker A and Walker B boxes. The conserved sequences comprise the adenine binding motif (A-binding), the Walker A and B boxes, and the ATPase coupling motif (C-motif). The corresponding motifs from phage P22 terminase are shown for comparison (57). The positions of the discontinuities in the ϕHau3 and StrepC.1 terminase-encoding genes are shown by the down-pointing vertical arrow (ϕHau3 *g6/7* and the StrepC.1 *g8/9* (nt 41080 to 42729 in *Streptomyces* sp. strain C contig ACEW01000274.1) and the up-pointing arrow (ϕHau3 *g7/8*).

peptide that has the same sequence as the R4 terminase: either 8 or 10 nucleotides must be spliced out of the mRNA at each interruption prior to translation to create a continuous open reading frame in the mRNA or, if the mRNA retains the interrupting sequences, the ribosome must shift frames and skip over either 8 or 10 nucleotides of the mRNA at each of the interruptions. Although there is currently no biochemical evidence to discriminate between these two general models, we strongly favor the second, "ribosome hopping" model, both because we are not aware of any precedent for splicing out such short sequences and, more importantly, because of a striking similarity between these sequences and the well-studied 50-base ribosome hop in phage T4 gene *60*, an essential T4 gene encoding a subunit of topoisomerase (45, 46). We present here a model based on that for the T4 *g60* bypass for how the ribosomes might navigate past the interruptions in the ϕHau3 and StrepC.1 *ter* genes to produce a functional terminase.

Examination of the sequences indicated the presence of the UUA leucine codon within all three of the interrupting sequences

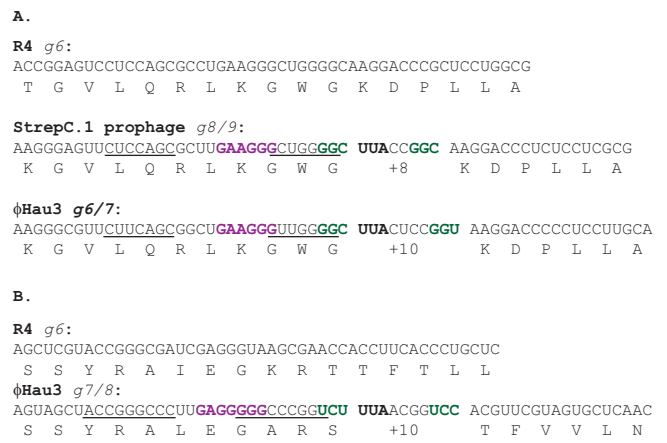

FIG 7 Model for the mechanism of translation of the fragmented terminase ORFs from ϕHau3 and StrepC.1 prophage. (A) The sequence of R4 *g6* mRNA and its translated sequence that spans the proposed Walker A motif and the equivalent sequences in StrepC.1 prophage and in ϕHau3 are shown. The translated sequence that would generate a peptide almost identical to the R4 terminase is shown. The UUA where a ribosomal pause is predicted to occur is shown in boldface. The codon before the UUA (so-called take-off codon) and the codon that precedes the continuation of the terminase peptide (so-called landing codon) are in green. Inverted repeats are indicated by underlined sequences, and complementarity to the 16S rRNA is indicated in pink. (B) The sequence of R4 *g6* and its translated sequence at the putative Walker B motif are shown along with the equivalent region in ϕHau3 (*g7/8*). Annotation is the same as for panel A.

(Fig. 7). UUA codons are extremely rarely used in GC-rich *Streptomyces*, so much so that the gene (*bldA*) encoding the corresponding $tRNA^{Leu}$(UAA) can be deleted in *S. coelicolor* because there are no UUA codons in essential genes (47). We propose that when the ribosome encounters the UUA codon in the ribosomal A site, it pauses, with the A site empty because of the low abundance of the corresponding tRNA and with the preceding codon (the "take-off" codon, in the terminology of the T4 *g60* hop) in the P site paired with its tRNA. A secondary structure forms between inverted repeats found upstream from and extending into the P site codon. The formation of the secondary structure initiates movement of the ribosome forward along the mRNA by a combination, we suggest, of stripping the mRNA off the tRNAs in the E and P sites and pulling the mRNA through to position the "landing codon" in the P site. This positions the ribosome to continue translation in the correct frame and at the right place in the sequence to generate a terminase of the same size as and with a sequence similar to that of R4 terminase. An upstream sequence complementary to the 16S rRNA may act to anchor the mRNA during hopping. The mechanism of the T4 *g60* hop, which has been studied in detail, uses a UAG stop codon to pause translation, the same arrangement of takeoff and landing codons, an mRNA secondary structure to draw the mRNA through the ribosome, and complementarity to 16S RNA to stabilize the translocating complex (46).

An implication of using the UUA codon to stimulate a ribosomal hop is that it connects the expression of the terminase proteins to the physiological state of the streptomycete host. Translation of UUA-containing genes in *S. coelicolor* is thought to be efficient only in the later stages of growth, coinciding with sporulation and antibiotic biosynthesis (48, 49). Indeed, the *S. coelicolor* *bldA* mutant, which lacks $tRNA^{Leu}$(UAA), shows normal vegetative growth but is defective in both sporulation and antibiotic biosynthesis (50). During a ϕHau3 infection, the terminase might be expressed efficiently during growth in young mycelium, when the $tRNA^{Leu}$(UAA) is either lacking or inactive. In contrast, in older mycelium, final maturation of the phage could be limited by decreased expression of the terminase, stalling the infection while the host sporulates and sequestering the ϕHau3 genome into spores. Outgrowth would once again provide optimal conditions for phage maturation and propagation.

The observation of a putative ribosome hop in the *ter* genes of ϕHau3 and StrepC.1 prophage prompted a search for other UUA codons in the phage and prophage sequences. There are 19 UUA codons distributed among the 19 phage and prophage genomes examined here. Ten of these are in genes coding for proteins of unknown function and most likely include recently acquired genes that have not yet adapted to the skewed codon usage of *Streptomyces* hosts. However, four UUA-containing genes encode putative regulators (*g7* in Mu1/6 and the putative L5-like repressor genes *g31* in Zemlya, *g33* in Sujidade, and *g33* in Lika) and two encode integrases (the *int* gene in Shyg.1 [GenBank accession no. WP_009713684] and Sujidade *g53*). The UUA codons in these genes could have a role in regulating the developmental fate of the phage through differences in the efficiency of translation in young versus old mycelium; i.e., in young mycelium, lysis would be favored due to poor translation of the regulator or integrase, and in older mycelium, when the $tRNA^{Leu}$(UUA) is active, lysogeny would be favored.

(ii) Conserved *Streptomyces* phage proteins. The head and tail genes of virtually all tailed phages are syntenically arranged, and for some of these, the sequences are well-enough conserved that they can be identified on that basis. For the *Streptomyces* phages and prophages considered here, we are able to identify genes for terminase large subunit, the portal protein, major capsid protein, and tail length tape measure protein, occurring in that relative order in each of the phages and prophages (see Fig. S1 in the supplemental material). In some but not all of the sequences, we find plausible candidates for genes encoding a head assembly accessory factor (a homologue of phage SPP1 gp7 or phage Mu gpG), a prohead maturation protease, a scaffolding protein, and the pair of tail assembly chaperones produced as the result of a translational frameshift. All of the head and tail genes that can be identified in this group of phages are in the same order along the genome as has been seen for the corresponding genes in tailed phages (see Fig. S1).

Analysis of the 19 *Streptomyces* phages and prophages using the Phamerator and BLASTp programs identified proteins that are unique to and commonly encoded in the *Streptomyces* phages. The Phamerator program groups phage proteins that are related into "phams" (see Fig. S2 in the supplemental material). Many phams only contain members derived from phages that belong to one particular cluster, but some phams contain members shared in phages derived from two or more clusters. For example, phams 1 (encoding putative HNH endonucleases) and 5 (putative large terminases) contain related gene products from the R4 and VWB clusters and pham 98 (putative tail tape measure protein) from members of the ϕC31, VWB, and SV1 clusters. Pham3 is shared by the members of the R4 and ϕC31 clusters and ϕSASD1 and contains ϕC31 gp29. In ϕC31, *g29* is the first gene in the late operon and is of unknown function, but gp29 is predicted to have a mem-

brane-spanning helix. As the gp29 homologues are all encoded close to the terminase genes, it is possible that this protein might be involved in DNA packaging, perhaps tethering the packaging apparatus to the membrane. The endolysins from the R4 and ϕC31 clusters are related, belonging to pham 25.

Sequence comparisons with BLASTp detected more-distant relatedness between other phage proteins. In the early region, two genes appear to be prevalent in *Streptomyces* phage genomes, the ϕC31 *g1* and *g3* homologues. ϕC31 gp1 is an early-expressed protein of unknown function with close homologues in other ϕC31 cluster phages. The presence of ϕC31 gp1 homologues in Mu1/6 (SPMV1_gp13), VWB (VWB_gp8), SPB78.1 gp49 (GenBank accession no. WP_009066686), and ϕSASD1 (ϕSASD1 gp1) suggest a common and possibly necessary function in all these phages. A PSI-BLAST search with R4 gp31, a protein that has homologues in the other R4 cluster phages and in many mycobacteriophage proteins, also revealed similarity to ϕC31 gp1. In ϕC31, *g1* is the first gene in the early region cluster, and there is a *g1* homologue in the *Arthrobacter* phage ϕAAU2. gp1 therefore appears to be a highly conserved protein in the actinophages, but we do not yet know its function.

We showed recently that ϕC31 gp3 functions as the recombination directionality factor (RDF) with integrase (51). VWB and Mu1/6 both encode integrases of the tyrosine recombinase family, and the RDFs that control tyrosine integrases are normally basic, small, DNA-binding proteins encoded by genes adjacent to their respective *int* genes. The homologues of ϕC31 gp3 in VWB (VWB_gp9) and in Mu1/6 (SPMV1_gp14) may therefore have roles in phage replication rather than controlling the directionality of integrase.

ϕC31 g44, 45, and 46 are putative tail or tail fiber proteins, and their variations may provide insight into the phage receptor interaction during infection. Previously, we identified a substitution in ϕC31 gp44 that compensated for a host receptor mutation (52). We find distant homologues of gp44, gp45, and gp46 within *Streptomyces* phages from different clusters, in particular in SV1, in pZL12 (a recently reported plasmid/phage) (12), and in Mu1/6, suggesting that these proteins could be involved in specific phage-*Streptomyces* cell surface interactions. PSI-BLAST with ϕC31 gp44 also reveals homologues in the R4 cluster (e.g., in StrepC.1 [WP_007264265]).

SV1—a rare transducing phage for *Streptomyces* species. SV1 is a generalized transducing phage with a narrow host range, limited to *S. venezuelae* (15, 29). Generalized transducing phages in *Streptomyces* are rare, and despite a report of transducing phages for *S. coelicolor* (53), SV1 remains the only one to date that has usefully been applied for genetic studies (although a new generalized transducing phage for *S. coelicolor*, ϕCAM, should be noted [54]). Can we use the sequence of SV1 to predict which other *Streptomyces* phages might be generalized transducers? Most generalized transducing phages package DNA through a headful packaging mechanism. DNA packaging using this type of mechanism normally initiates by recognition of a *pac* site or region and then packages more than a single genome length per phage head, resulting in terminal redundancy. Mistakes in this process lead to occasional packaging of host DNA. The DNA in SV1 virions appears to be terminally redundant and circularly permuted, given that the genomic DNA in the virions is about 20% longer than the unique sequence. SV1 probably, therefore, packages through a headful mechanism. In all dsDNA phages, the terminase apparatus, comprising the large and small subunits, mediates packaging, and terminases with similar sequences generally package DNA by the same mechanism (55). Indeed, Casjens et al. (55) showed that it might be possible to predict whether a phage packages by a headful mechanism using a phylogenetic analysis of the terminase. We therefore examined the relatedness of the SV1 large terminase to other terminases in the database and to our collection of *Streptomyces* phages.

Shyg.1 is a prophage with considerable sequence similarity to SV1. The amino acid identity between the SV1 and Shyg.1 large terminase subunits (SV1 gp1 and the Shyg.1 gp1 with accession number WP_009713708) is the highest (82% identical) of any protein pair in these two phages. Small terminase subunits are required for DNA recognition (for both sequence-specific and nonspecific binding), and there is a clear difference in the requirements for *cos* site recognition, which is always sequence specific, and recognition of DNA in *pac* site/headful packaging. Although small terminase subunits are diverse in sequence, they can usually be identified as they tend to lie just upstream from the large terminase genes. gp55 from SV1 and its homologue in Shyg.1 (gp53; GenBank accession no. WP_009713707) are encoded upstream from their respective large terminase genes and are highly similar (66% identity), as are the portal proteins (62% identical), proposed to be part of the mechanism that signals headful packaging to the terminase (56). A BLAST search with SV1 gp1 against the dsDNA viruses indicates that, of the top 12 hits, 6 have headful packaging mechanisms, and there is no information on the DNA-packaging mechanism of the other six. An amino acid sequence alignment confirms that SV1 terminase is much more similar to the headful packaging terminases than to the other *Streptomyces* terminase proteins (see Fig. S9 and S10 in the supplemental material). If the nature of the terminase is an indicator of whether the phage packages DNA by a headful mechanism, then BLAST searches against the microbial database suggest that there are prophages in *Streptomyces* genomes that may also be headful packagers and, therefore, candidate generalized transducers. These prophages include Shyg.1 and putative prophages in *Streptomyces* sp. strain SA3_ActG and *Streptomyces* sp. strain Tü6071.

ACKNOWLEDGMENTS

M.C.M.S. acknowledges funding from the Biotechnology and Biological Sciences Research Council (BBSRC) of the United Kingdom via project grants BB/H001212/1 and BB/H05447/1. Work in Pittsburgh was supported by NIH grants GM47795 to R.W.H. and GM093901 to G.F.H. Work in Norwich was funded by a MET Institute Strategic Programme grant to the John Innes Centre from the BBSRC.

We are grateful to Charles A. Bowman for the construction of the Phamerator database and to John Atkins and Keith Chater for stimulating discussions on ribosome bypassing and UUA codons.

REFERENCES

1. **Hendrix RW, Smith MCM, 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.
2. **Juhala RJ, Ford ME, Duda RL, Youlton A, Hatfull GF, Hendrix RW.** 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. **299:**27–51.
3. **Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA.** 2000. Practical *Streptomyces* genetics. The John Innes Foundation, Norwich, United Kingdom.
4. **Baltz RH.** 2012. Streptomyces temperate bacteriophage integration sys-

tems for stable genetic engineering of actinomycetes (and other organisms). J. Ind. Microbiol. Biotechnol. **39:**661–672.
5. **Bierman M, Logan R, O'Brien K, Seno ET, Rao RN, Schoner BE.** 1992. Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. Gene **116:**43–49.
6. **Kuhstoss S, Richardson MA, Rao RN.** 1991. Plasmid cloning vectors that integrate site-specifically in Streptomyces spp. Gene **97**(1)**:**143–146. (Erratum, Gene **102**(1):145.)
7. **Gregory MA, Till R, Smith MCM.** 2003. Integration site for *Streptomyces* phage ϕBT1 and the development of novel site-specific integrating vectors. J. Bacteriol. **185:**5320–5323.
8. **Lomovskaya N, Mkrtumian NM, Gostimskaya NL.** 1970. Isolation and characterisation of Streptomyces coelicolor actinophage. Genetika **6:**135–137.
9. **Van Dessel W, Van Mellaert L, Liesegang H, Raasch C, De Keersmaeker S, Geukens N, Lammertyn E, Streit W, Anne J.** 2005. Complete genomic nucleotide sequence and analysis of the temperate bacteriophage VWB. Virology **331:**325–337.
10. **Wang S, Qiao X, Liu X, Zhang X, Wang C, Zhao X, Chen Z, Wen Y, Song Y.** 2010. Complete genomic sequence analysis of the temperate bacteriophage phiSASD1 of Streptomyces avermitilis. Virology **403:**78–84.
11. **Farkasovska J, Klucar L, Vlcek C, Kokavec J, Godany A.** 2007. Complete genome sequence and analysis of the Streptomyces aureofaciens phage mu1/6. Folia Microbiol. (Praha) **52:**347–358.
12. **Zhong L, Cheng Q, Tian X, Zhao L, Qin Z.** 2010. Characterization of the replication, transfer, and plasmid/lytic phage cycle of the Streptomyces plasmid-phage pZL12. J. Bacteriol. **192:**3747–3754.
13. **Chater KF, Carter AT.** 1979. A new, wide host-range, temperate bacteriophage (R4) of *Streptomyces* and its interaction with some restriction-modification systems. J. Gen. Microbiol. **115:**431–442.
14. **Foor F, Roberts GP, Morin N, Snyder L, Hwang M, Gibbons PH, Paradiso MJ, Stotish RL, Ruby CL, Wolanski B, Streicher SL.** 1985. Isolation and characterization of the Streptomyces cattleya temperate phage TG1. Gene **39:**11–16.
15. **Stuttard C.** 1979. Transduction of auxotrophic markers in a chloramphenicol-producing strain of Streptomyces. J. Gen. Microbiol. **110:**479–482.
16. **Zhou X, Deng Z, Hopwood DA, Kieser T.** 1994. Characterization of phi HAU3, a broad-host-range temperate streptomyces phage, and development of phasmids. J. Bacteriol. **176:**2096–2099.
17. **Lanning S, Williams ST.** 1982. Methods for the direct isolation and enumeration of actinophages in soil. J. Gen. Microbiol. **128:**2063–2071.
18. **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**(Pt 9)**:**2962–2977.
19. **Fouts DE.** 2006. Phage_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences. Nucleic Acids Res. **34:**5839–5851.
20. **Van Mellaert L, Mei L, Lammertyn E, Schacht S, Anne J.** 1998. Site-specific integration of bacteriophage VWB genome into Streptomyces venezuelae and construction of a VWB-based integrative vector. Microbiology **144**(Pt 12)**:**3351–3358.
21. **Delcher AL, Harmon D, Kasif S, White O, Salzberg SL.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641.
22. **Besemer J, Borodovsky M.** 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res. **33:**W451–W454.
23. **Krumsiek J, Arnold R, Rattei T.** 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics **23:**1026–1028.
24. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
25. **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. doi:10.1186/1471-2105-12-395.
26. **Laslett D, Canback B.** 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. **32:**11–16.
27. **Schattner P, Brooks AN, Lowe TM.** 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. **33:**W686–W689.
28. **Morino T, Isogal T, Kojima H, Takahashi H, Saito H.** 1983. Cleavage analysis of actinophage R4 and its deletion mutants. Agric. Biol. Chem. **47:**1873–1878.
29. **Stuttard C.** 1989. Generalized transduction in *Streptomyces* species, p 157–162. *In* Hershberger CL, Queener SW, Hegeman G (ed), Genetics and molecular biology of industrial microorganisms. American Society for Microbiology, Washington, DC.
30. **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**(6)**:**e92.
31. **Ishihara H, Nakano MM, Ogawara H.** 1982. Restriction enzyme mapping of the DNA of Streptomyces bacteriophage B alpha and its deletion derivatives. J. Bacteriol. **152:**1288–1291.
32. **Ravin V, Ravin N, Casjens S, Ford ME, Hatfull GF, Hendrix RW.** 2000. Genomic sequence and analysis of the atypical temperate bacteriophage N15. J. Mol. Biol. **299:**53–73.
33. **Summer EJ, Liu M, Gill JJ, Grant M, Chan-Cortes TN, Ferguson L, Janes C, Lange K, Bertoli M, Moore C, Orchard RC, Cohen ND, Young R.** 2011. Genomic and functional analyses of Rhodococcus equi phages ReqiPepy6, ReqiPoco6, ReqiPine5, and ReqiDocB7. Appl. Environ. Microbiol. **77:**669–683.
34. **Jacobs-Sera D, Marinelli LJ, Bowman C, Broussard GW, Guerrero Bustamante C, Boyle MM, Petrova ZO, Dedrick RM, Pope WH, Modlin RL, Hendrix RW, Hatfull GF, Science Education Alliance Phage Hunters Advancing Genomics And Evolutionary Science Sea-Phages Program.** 2012. On the nature of mycobacteriophage diversity and host preference. Virology **434:**187–201.
35. **Brown KL, Sarkis GJ, Wadsworth C, Hatfull GF.** 1997. Transcriptional silencing by the mycobacteriophage L5 repressor. EMBO J. **16:**5914–5921.
36. **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 AA, 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 SC, Engelsen AE, English BE, Erb ML, Estrada C, Filliger LZ, Findley AM, Forbes L, Forsyth MH, Fox TM, Fritz MJ, et al.** 2011. Expanding the diversity of mycobacteriophages: insights into genome architecture and evolution. PLoS One **6:**e16329. doi:10.1371/journal.pone.0016329.
37. **Ingham CJ, Owen CE, Wilson SE, Hunter IS, Smith MCM.** 1994. An operator associated with autoregulation of the repressor gene in actinophage phi-c31 is found in highly conserved regions in the phage genome. Nucleic Acids Res. **22:**821–827.
38. **Wilson SE, Ingham CJ, Hunter IS, Smith MCM.** 1995. Control of lytic development in the Streptomyces temperate phage phi-c31. Mol. Microbiol. **16:**131–143.
39. **Ingham CJ, Crombie HJ, Bruton CJ, Chater KF, Hartley NM, Murphy GJP, Smith MCM.** 1993. Mutiple novel promoters from the early region in the Streptomyces temperate phage phi-c31 are activated during lytic development. Mol. Microbiol. **9:**1267–1274.
40. **Ingham CJ, Hunter IS, Smith MCM.** 1995. Rho-independent terminators without 3′ poly-U tails from the early region of actinophage phi-c31. Nucleic Acids Res. **23:**370–376.
41. **Wilson SE, Smith MCM.** 1998. Oligomeric properties and DNA binding specificities of repressor isoforms from the Streptomyces bacteriophage phiC31. Nucleic Acids Res. **26:**2457–2463.
42. **Smith MCM, Owen CE.** 1991. Three in-frame N-terminally different proteins are produced from the repressor locus of the Streptomyces bacteriophage phi C31. Mol. Microbiol. **5:**2833–2844.
43. **Sinclair RB, Bibb MJ.** 1988. The repressor gene (c) of the Streptomyces temperate phage phi c31: nucleotide sequence, analysis and functional cloning. Mol. Gen. Genet. **213**(2-3)**:**269–277.
44. **Rao VB, Black LW.** 2010. Structure and assembly of bacteriophage T4 head. Virol. J. **7:**356. doi:10.1186/1743-422X-7-356.
45. **Huang WM, Ao SZ, Casjens S, Orlandi R, Zeikus R, Weiss R, Winge D, Fang M.** 1988. A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60. Science **239:**1005–1012.
46. **Wills NM, O'Connor M, Nelson CC, Rettberg CC, Huang WM, Geste-**

land RF, Atkins JF. 2008. Translational bypassing without peptidyl-tRNA anticodon scanning of coding gap mRNA. EMBO J. **27:**2533–2544.
47. **Hesketh A, Bucca G, Laing E, Flett F, Hotchkiss G, Smith CP, Chater KF.** 2007. New pleiotropic effects of eliminating a rare tRNA from Streptomyces coelicolor, revealed by combined proteomic and transcriptomic analysis of liquid cultures. BMC Genomics **8:**261. doi:10.1186/1471-2164-8-261.
48. **Leskiw BK, Mah R, Lawlor EJ, Chater KF.** 1993. Accumulation of bldA-specified tRNA is temporally regulated in Streptomyces coelicolor A3(2). J. Bacteriol. **175:**1995–2005.
49. **Pettersson BM, Kirsebom LA.** 2011. tRNA accumulation and suppression of the bldA phenotype during development in Streptomyces coelicolor. Mol. Microbiol. **79:**1602–1614.
50. **Leskiw BK, Lawlor EJ, Fernandez-Abalos JM, Chater KF.** 1991. TTA codons in some genes prevent their expression in a class of developmental, antibiotic-negative, Streptomyces mutants. Proc. Natl. Acad. Sci. U. S. A. **88:**2461–2465.
51. **Khaleel T, Younger E, McEwan AR, Varghese AS, Smith MC.** 2011. A phage protein that binds phiC31 integrase to switch its directionality. Mol. Microbiol. **80:**1450–1463.
52. **Cowlishaw DA, Smith MCM.** 2001. Glycosylation of a Streptomyces coelicolor A3(2) cell envelope protein is required for infection by bacteriophage phi C31. Mol. Microbiol. **41:**601–610.
53. **Burke J, Schneider D, Westpheling J.** 2001. Generalized transduction in Streptomyces coelicolor. Proc. Natl. Acad. Sci. U. S. A. **98:**6289–6294.
54. **Monson R, Salmond GP.** 2012. Genome sequence of a new Streptomyces coelicolor generalized transducing bacteriophage, PhiCAM. J. Virol. **86:**13860.
55. **Casjens SR, Gilcrease EB, Winn-Stapley DA, Schicklmaier P, Schmieger H, Pedulla ML, Ford ME, Houtz JM, Hatfull GF, Hendrix RW.** 2005. The generalized transducing Salmonella bacteriophage ES18: complete genome sequence and DNA packaging strategy. J. Bacteriol. **187:**1091–1104.
56. **Lander GC, Tang L, Casjens SR, Gilcrease EB, Prevelige P, Poliakov A, Potter CS, Carragher B, Johnson JE.** 2006. The structure of an infectious P22 virion shows the signal for headful DNA packaging. Science **312:**1791–1795.
57. **Rao VB, Feiss M.** 2008. The bacteriophage DNA packaging motor. Annu. Rev. Genet. **42:**647–681.
58. **Gordon D, AbaJian C, Green P.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202.
59. **Bedford DJ, Laity C, Buttner MJ.** 1995. Two genes involved in the phase-variable phiC31 resistance mechanism of Streptomyces coelicolor A(3)2. J. Bacteriol. **17:**4681–4689.