DNA Packaging and Genomics of the *Salmonella* 9NA-Like Phages

Chi Zeng,^a Eddie B. Gilcrease,^{b,*} Roger W. Hendrix,^{c,†} Yicheng Xie,^{d,e,*} Michael J. Jalfon,^e Jason J. Gill,^{d,e} Sherwood R. Casjens^{b,f}

^aHubei Province Fresh Food Engineering Research Center, School of Biology and Pharmaceutical Engineering, Wuhan Polytechnic University, Wuhan, People's Republic of China

^bDivision of Microbiology and Immunology, Pathology Department, University of Utah School of Medicine, Salt Lake City, Utah, USA

^cDepartment of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

^dDepartment of Animal Science, Texas A & M University, College Station, Texas, USA

^eCenter for Phage Technology, Texas A & M University, College Station, Texas, USA

^fSchool of Biological Sciences, University of Utah, Salt Lake City, Utah, USA

ABSTRACT We present the genome sequences of *Salmonella enterica* tailed phages Sasha, Sergei, and Solent. These phages, along with *Salmonella* phages 9NA, FSL_SP-062, and FSL_SP-069 and the more distantly related *Proteus* phage PmiS-Isfahan, have similarly sized genomes of between 52 and 57 kbp in length that are largely syntenic. Their genomes also show substantial genome mosaicism relative to one another, which is common within tailed phage clusters. Their gene content ranges from 80 to 99 predicted genes, of which 40 are common to all seven and form the core genome, which includes all identifiable virion assembly and DNA replication genes. The total number of gene types (pangenome) in the seven phages is 176, and 59 of these are unique to individual phages. Their core genomes are much more closely related to one another than to the genome of any other known phage, and they comprise a well-defined cluster within the family *Siphoviridae*. To begin to characterize this group of phages in more experimental detail, we identified the genes that encode the major virion proteins and examined the DNA packaging of the prototypic member, phage 9NA. We show that it uses a *pac* site-directed headful packaging mechanism that results in virion chromosomes that are circularly permuted and about 13% terminally redundant. We also show that its packaging series initiates with double-stranded DNA cleavages that are scattered across a 170-bp region and that its headful measuring device has a precision of $\pm 1.8\%$.

IMPORTANCE The 9NA-like phages are clearly highly related to each other but are not closely related to any other known phage type. This work describes the genomes of three new 9NA-like phages and the results of experimental analysis of the proteome of the 9NA virion and DNA packaging into the 9NA phage head. There is increasing interest in the biology of phages because of their potential for use as antibacterial agents and for their ecological roles in bacterial communities. 9NA-like phages that infect two bacterial genera have been identified to date, and related phages infecting additional Gram-negative bacterial hosts are likely to be found in the future. This work provides a foundation for the study of these phages, which will facilitate their study and potential use.

KEYWORDS bacteriophage, 9NA, Sasha, Sergei, Solent, DNA packaging, *Salmonella*

Although the experimental study of a small number of tailed phages (e.g., phages lambda and T4) made great early contributions to our current understanding of genes and how they function, the tremendous abundance and diversity of tailed

Citation Zeng C, Gilcrease EB, Hendrix RW, Xie Y, Jalfon MJ, Gill JJ, Casjens SR. 2019. DNA packaging and genomics of the *Salmonella* 9NA-like phages. J Virol 93:e00848-19. <https://doi.org/10.1128/JVI.00848-19>.

Editor Julie K. Pfeiffer, University of Texas Southwestern Medical Center

Copyright © 2019 American Society for Microbiology. All Rights Reserved.

Address correspondence to Jason J. Gill, jason.gill@tam.u.edu, or Sherwood R. Casjens, sherwood.casjens@path.utah.edu.

* Present address: Eddie B. Gilcrease, Department of Civil and Environmental Engineering, University of Utah, Salt Lake City, Utah, USA; Yicheng Xie, Food Safety Net Services, San Antonio, Texas, USA.

† Deceased.

C.Z. and E.B.G. contributed equally to this article.

Received 21 May 2019

Accepted 26 August 2019

Accepted manuscript posted online 28 August 2019

Published 29 October 2019

bacteriophages and their resulting ecological importance have only more recently been recognized (see references 1 to 4 and references therein). In addition, phages can carry toxin and virulence factor genes that are very important in many human bacterial diseases (5–7), and they are being studied for potential nanotechnological uses (8–11). Finally, interest in the use of phages in bacterial therapeutic applications has been reawakened with the emergence of bacterial antibiotic resistance (12–14).

In spite of their biological, ecological, and technological importance, many types of tailed phages have not yet been studied in detail or given formal taxonomic status by the International Committee on Taxonomy of Viruses (ICTV; see <https://talk.ictvonline.org/taxonomy/>). In an effort to better understand this immense tailed phage diversity, we are interested in understanding all of the types of tailed phages that infect the bacterial order *Enterobacteriales* in general and, more specifically, those phages that infect *Escherichia coli* and *Salmonella enterica*. In our recent studies, we showed that the *Enterobacteriales* tailed phages whose genomes have been completely sequenced naturally group into 71 different clusters, and since that time (2017) this has grown to over 85 clusters (15–17; S. R. Casjens, unpublished data). These clusters correspond approximately to ICTV genera, but many of them have not yet been formally described or studied beyond genome sequence determination.

We present here the complete nucleotide sequences of three *S. enterica* lytic phages, Sasha, Sergei, and Solent, and compare them to the complete genome sequence of related *S. enterica* phage 9NA (18) and the nearly complete sequences of *S. enterica* phages FSL_SP-062 and FSL_SP-069 (19) and *Proteus mirabilis* phage PmiS-Isfahan (20). We find that these seven phages have considerable nucleotide sequence similarity and, with the exception of some mosaic sectional differences (as is typical of the tailed phages), have syntenic genomes. These phages are only very distantly related to other tailed phages, and we show here that they constitute a unique group.

In addition, we report on the experimental determination of the major virion proteins and analysis of the DNA packaging mechanism of phage 9NA, the prototypical member of this phage cluster. This work shows that it is a headful packaging phage that initiates packaging series on replicated concatemeric DNA at a specific *pac* recognition site. Such 9NA packaging series can be at least five genomes long, and phage chromosomes that have 12.9% terminal redundancy with a precision of $\pm 1.8\%$ are packaged.

RESULTS AND DISCUSSION

Phage isolation, morphological characteristics, and genome sequence. Phages Sasha, Sergei, and Solent were isolated from beef cattle feedlots in south Texas using *S. enterica* serovar Anatum strain FC1033C3 as the host. In addition to infecting their propagation host, phages Sasha and Sergei were shown to be able to infect three additional *Salmonella* serovar Anatum strains but were not able to infect representatives of 11 other *Salmonella* serovars, including *Salmonella* serovars Typhimurium, Enteritidis, Newport, and Montevideo (21). Transmission electron microscopy showed that the Sasha, Sergei, and Solent virions have the typical siphovirus morphology with isometric icosahedral heads that are ~ 67 nm in diameter and long, noncontractile tails that are ~ 156 nm in length and ~ 11 nm in width with a brushy baseplate at the distal tip (Fig. 1; the data for phage Solent are not shown). They are virtually indistinguishable from previous micrographs of phage 9NA (22, 23).

The genome sequences of phages Sasha (53,263 bp), Sergei (56,051 bp), and Solent (55,978 bp) were determined as described in Materials and Methods. The genomes of phages Sergei and Solent were found to be nearly identical (99.9% DNA sequence identity over the full length of the genome with identical gene contents); thus, phage Sergei is used as the representative genome for the pair and is used in detailed comparative genome analysis and annotation. Like the 9NA sequence run, the Sasha, Sergei, and Solent sequence runs assembled into single circular contigs, suggesting that their virion DNAs contain direct terminal redundancies (see below). The G+C contents of the six *Salmonella* phages in this group ranged from 42.8% to 43.7%,

FIG 1 Negative-stain electron micrographs of phage 9NA, Sasha, and Sergei virions. All three phages possess a similar *Siphoviridae* morphology with head diameters of ~67 nm and tails of ~156 nm in length.

somewhat lower than the *Salmonella* chromosome value of 52.2% G+C (24), and the PmiS-Isfahan genome G+C content of 36.1% was also lower than the 39.4% G+C content of its host (25).

9NA, Sasha, Sergei, Solent, FSL_SP-062, FSL_SP-069, and PmiS-Isfahan form a unique phage cluster. The Sasha, Sergei, and Solent genomes are quite similar to those of *S. enterica* phages 9NA, FSL_SP-062, and FSL_SP-069 (18, 19). The last two sequences were reported to have three and five contigs, respectively (19), but these contigs align contiguously on the complete 9NA, Sasha, and Sergei/Solent genomes. Thus, by virtue of their sequence similarity to these three whole genomes, we assume that this aligned contig order is correct. In addition, the genome of a *Proteus* phage, PmiS-Isfahan, that is more distantly related to this cluster of phages has been reported (20). All these phages have genomes in the 52.8- to 56.6-kbp range (Table 1) and appear to constitute a unique phage type or cluster that is only distantly related to other known *Siphoviridae* phages. No detailed analysis of this phage group has been previously reported.

Figure 2 shows open reading frame (ORF) maps of the phage 9NA, Sasha, and Sergei/Solent genomes, where the circular genome sequence is arbitrarily opened near the *pac* site (see below). Comparison of their genomes suggests that there are likely 80, 93, and 85 protein-coding genes in Sasha, Sergei/Solent, and 9NA, respectively; the nearly identical *Salmonella* phage FSL_SP-062 and FSL_SP-069 genomes have 99 annotated genes, and the PmiS-Isfahan genome has 91 annotated genes (see Table S1 in the supplemental material). No tRNA genes were found in these phage genomes. Clearly, the genomes shown in Fig. 2 are highly syntenic with some mosaic differences (discussed in more detail below).

The fact that these seven phages form a unique group is supported by analysis of some of the universally present proteins encoded by these phages: the portal protein, the major capsid protein (MCP), the major tail tube or shaft protein (referred to as the major tail protein [MTP]), DNA polymerase, helicase, and primase. Figure 3 shows neighbor-joining trees of these six proteins encoded by the seven 9NA-like phages and their closest outside relatives present in the public sequence database. In each case, the proteins of the 9NA-like phages are more highly related to one another than they are to any protein encoded outside the group, and they form a unique, self-contained

TABLE 1 The 9NA-like phages

Phage	Host	Genome size (bp)	Geographic origin	GenBank accession no.	Reference(s) or source
9NA	<i>Salmonella enterica</i> serovar Typhimurium	52,869	Finland(?)	KJ802832	18, 22
Sasha	<i>Salmonella enterica</i> serovar Anatum	53,263	Texas	KX987158	This work
Sergei	<i>Salmonella enterica</i> serovar Anatum	56,051	Texas	KY002061	This work
Solent	<i>Salmonella enterica</i> serovar Anatum	55,978	Texas	MH586730	This work
FSL_SP-062	<i>Salmonella enterica</i> serovar Newport	56,585 ^a	New York State	KC139632 to KC139637	19
FSL_SP-069	<i>Salmonella enterica</i> serovar Newport	56,560 ^a	New York State	KC139649 to KC139651	19
PmiS-Isfahan	<i>Proteus mirabilis</i>	54,836	Iran	KY742649	20

^aApproximate values from our attempts to merge these phage genomes' multiple contigs using the complete genome of 9NA as a guide; these values are almost certainly accurate to within 50 bp.

FIG 2 Comparative ORF maps of the 9NA-like phages. Maps of the predicted open reading frames of the six known *Salmonella* 9NA-like phages are shown, with the genes in red and green being transcribed rightward and leftward, respectively. Selected gene names are shown above the genes, and predicted functions are shown at the top. Asterisks, the protein in the virion (see the text); fs, expressed by a programmed translational frameshift. FSL_SP-062 and FSL_SP-069 are nearly identical, so only the map for the latter one is shown, with the last three digits of selected gene locus_tags being indicated above its map; their multiple contigs are oriented so that they align with those of the other four phage genomes. The bar below the 9NA map includes symbols that mark the locations of the following 9NA sequences: solid triangles, repeat number 1 (see the text), with the orientation of the triangle showing the sequence orientation; closed circle, possible origin repeats (see the text); closed diamond, 9NA *orf10* repeat tract; open triangles, 28-bp perfect direct repeat only in the 9NA genome. The DNA packaging site *pac* (the black arrow indicates the direction of packaging) and possible transcription terminators (T) are indicated below the gray box. A scale (in kilobase pairs) is shown at the bottom of the figure.

branch with very high bootstrap support. These results agree with the findings of a similar independent analysis of a subset of these phages conducted using the MCP and large terminase proteins (20). The virion portal proteins, MCPs, and MTPs are very closely related within the six *Salmonella*-infecting members of this group, being less than 2%, 3%, and 12% different in amino acid sequence from one another, respectively. These three proteins in PmiS-Isfahan are more divergent, being about 40%, 39%, and 54% different, respectively, from the proteins of the other six phages. The nearest relatives outside this phage group are phage proteins that are 67%, 68%, and 61% different, respectively, from these 9NA-cluster proteins. The three DNA metabolism proteins in Fig. 3 have similar relationships within the group but are, in general, somewhat more diverse, again, with the PmiS-Isfahan proteins being the most divergent in each case.

Figure 4 compares dot plots of the genomes of the seven 9NA-like phages and representatives from six other small lytic phage clusters that infect *E. coli* and/or *Salmonella* (15). The six 9NA-like *Salmonella* phages clearly form a cluster that has little overt nucleotide sequence similarity with the phages outside this group, and PmiS-Isfahan shows a weak but long diagonal similarity line, indicating that it is a divergent member of this cluster (a higher-resolution plot is shown in Fig. 5A). Similar dot plots with members of the other small virulent *Enterobacteriales* phage clusters not shown in Fig. 4 (16, 17) showed little or no similarity, with the following single exception. Members of the MSW-3-like *Myoviridae* cluster (15) also have divergently transcribed putative early gene clusters where the DNA polymerase, helicase, and primase genes have locations that parallel those of the 9NA-like phages, and a high-resolution dot plot between these two clusters shows an approximately 9-kbp very weak diagonal line where these genes have sequence similarities (Fig. 5B). This similarity between the presumed early-expression regions of these two different phage clusters may indicate an ancient relationship. We nonetheless conclude that these are different clusters, that the seven 9NA-like phages form a unique cluster that is only very distantly related to all other tailed phages whose genomes have been sequenced, and that the *Salmonella* and *Proteus* phages within this cluster define two distinct subclusters. These conclusions are consistent with the recent proposal by the ICTV of the genus *Nonanavirus*,

Major Capsid Protein

DNA Polymerase

Portal Protein

Primase

Major Tail Protein

Helicase

FIG 3 Neighbor-joining trees of six different proteins. Unrooted protein trees were generated by the Clustal X program (90), and selected fractional distances and bootstrap values (out of 1,000 trials) are indicated below and above the branch lines, respectively; short distances and bootstrap values below 900 are not shown; a fractional distance scale is shown above each tree. The names of the phages that encode each of the proteins are given at the tip (right end) of the branch. The branches containing proteins encoded by the 9NA-like phages are enclosed in gray rectangles. The other phages are a representative sampling of those with the most closely related proteins outside the 9NA-like cluster that are present in the current tailed phage sequence database. They infect the following host genera: KPP23, PAK_P1, and PaMx42 infect *Pseudomonas*; VchM-138 infects *Vibrio*; MSW-3 and PEI21 infect *Edwardsiella*; RDJLø1 infects *Roseobacter*; ø105 infects *Bacillus*; LP-030-2 infects *Listeria*; KpV52 infects *Klebsiella*; HMSP1-Susan infects *Sinorhizobium*; PEAT2 infects *Pectobacterium*; Sansa infects *Caulobacter*; Shroom infects *Arthrobacter*; and 9a infects *Escherichia*. The asterisk notes that the MSW-3 primase homology is limited to the N-terminal half of the protein.

which currently contains two species, phages 9NA and FSL_SP-069. Phages Sasha, Sergei, and Solent are also likely members of this genus, while the more distantly related PmiS-Isfahan could belong to this genus or to a sister genus to the *Nonanavirus*, in which case, both of these genera should be grouped within the same subfamily to reflect their obvious relationship.

We note that, in addition to the 9NA-like phages, there are currently 11 other types of known virulent *Siphoviridae* phages that infect *Enterobacteriales* hosts and that have genomes in the 40- to 60-kbp range that constitute clusters in the same way as the 9NA-like phages. Grose and Casjens (15) defined the phage T1-, Chi-, SETP3-, SO1-, E1-, and 9g-like clusters (Fig. 4), and our unpublished analysis has identified five more such clusters that are typified by the recently sequenced genomes of phages Halfdan (GenBank accession no. [MH362766](#)), Kp3 (GenBank accession no. [KT367887](#)), Scapp (GenBank accession no. [MH553517](#)), MED16 (GenBank accession no. [MK095605](#)), and PMBT28 (GenBank accession no. [MG641885](#)) that infect *Escherichia*, *Klebsiella*, *Serratia*, *Pantoea*, and *Salmonella* hosts, respectively. These 11 phage clusters have diverse but functionally syntenic virion assembly gene clusters and putative early gene clusters that

FIG 4 Dot plot comparison of small virulent *Enterobacteriales* phage genomes. Phage names are indicated at the top and left. Phages T1, SETP3, SO-1, E1, Chi, and 9g are prototypical phages for other virulent phage clusters with similarly sized genomes that infect the *Enterobacteriales* (15). The dot plot was produced by use of the Gepard tool (92) at a word length setting of 10. All genomes are oriented so that the putative small terminase gene is near the left end and is transcribed in a rightward direction, except for PmiS-Isfahan, for which the orientation of its GenBank sequence is shown (see the text and the legend to Fig. 5A for a discussion of the possible incorrect assembly of that sequence). Asterisks mark genomes for which multiple contigs were joined (FPL_SP-062 and -069) to maximize synteny with the phage 9NA genome. Phages Halfdan, Kp3, Scapp, MED16, and PMBT28 (see the text) are not shown, but none of them showed any similarity to 9NA in such a plot.

encode helicases, DNA polymerases, and primases. Each of these clusters shows its own internal variation and genome mosaicism. Clearly, this is a very successful and ancient general lifestyle, but the evolutionary forces driving such amazingly wide diversity among these small *Siphoviridae* phages remain mysterious.

Experimental analysis of 9NA-like phage cluster virion proteins and DNA packaging strategy. (i) Major proteins of 9NA-like virions. We chose phage 9NA as the prototypical member of this cluster to study in more detail. Analysis of the genomes of the 9NA-like phages identified several genes that have sequence similarity to known virion assembly genes (Table S1), but in order to solidify knowledge of these phage virions, we separated the 9NA virion proteins by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis, isolated individual protein bands, and determined the N-terminal amino acid sequences for the five most intense bands. Such a gel and the sequences determined are indicated in Fig. 6A (additional minor virion proteins are not visible in the figure). These sequences match the N-terminal portions of the predicted 9NA proteins encoded by 9NA genes 39, 40, 47, 50, and 55. The predicted molecular weights of these proteins match the sizes determined by their mobility in this gel within the expected experimental error. The gene 40, 47, and 50 proteins have weak sequence similarity to known MCP, MTP, and tail tape measure proteins, respectively, and they are present in the expected relative intensities (major capsid > tail shaft >> tape measure). Andres et al. (23) have characterized the gene 55-encoded 9NA tailspike protein. Finally, the gene 39 protein is present in several hundred molecules per virion (quantitation by Coomassie brilliant blue band intensity is not shown), but it has no convincing sequence matches outside this phage group. We suggest that it may be a decoration protein bound to the outside of the head shell (26), since such proteins are usually present in large numbers per virion, are usually small, and are extremely

FIG 5 Dot plot comparisons of phage 9NA to more distantly related phage genomes. (A) Dot plot comparison of the phage 9NA and reordered PmiS-Isfahan* (denoted here as PmiS-Isfahan*) genomes. Two separate ORFs with similarity to the N and C termini of 9NA *orf66* were identified in the genome of PmiS-Isfahan. The two halves of PmiS-Isfahan's homologue of 9NA *orf66* were fused by removing bp 1 to 13500 of the original GenBank sequence and inserting them in the same orientation between original bp 39061 and 39062. This presumed circular sequence was then reopened at bp 39500 to make it colinear with phage 9NA, placing the putative *terS* gene near the left end, and this sequence was used to generate the above dot plot. (B) Dot plot comparison of the phage 9NA and MSW-3 genomes. The 9NA map (Fig. 2) is provided below the dot plot for reference; the weak diagonal similarity line is delimited by red vertical lines and includes all or parts of the DNA polymerase, helicase, and primase genes. The GenBank accession number for MSW-3 is [AB767244](#).

variable in sequence, and in medium-sized-genome phages, like those in the 9NA-like cluster, their genes often lie just transcriptionally upstream of the MCP gene. We also note that all five of these proteins have had their N-terminal methionine removed; this is not surprising, since the penultimate alanine, serine, and perhaps valine, which are present in these five proteins, often program intracellular methionine removal (27). Finally, the major capsid protein N-terminal sequence shows that this protein is not proteolytically cleaved (as many such proteins are) during 9NA virion assembly.

(ii) DNA packaging by 9NA-like phages. The products of replication of many double-stranded DNA phages are long concatemers of the genome sequence. Some such phages, for example, P22, P1, SPP1, Sf6, and ES18, package DNA through processive series of headful packaging events along such a concatemer (28–33). In these cases, DNA packaging initiates with the recognition of a site called *pac* by the small terminase (TerS) protein. DNA is then (i) cleaved, in most cases near this site; (ii) only one of the two ends thus created is threaded into the procapsid; and (iii) an ATP cleavage-powered DNA translocase (the large terminase, or TerL protein) pumps that DNA into the procapsid until it is full. When the packaging machinery senses that the head is full of DNA, TerL makes a double-strand cut, releasing the packaged DNA from the concatemer (hence the moniker “headful packaging” [34]). A second packaging event then begins at the concatemer end created by this first headful cleavage, and additional subsequent packaging events result in sequential, unidirectional packaging series that can be up to 10 or more events long (35). This type of phage packages a DNA chromosome that is longer than one genome sequence in each event; for example, P22 packages a DNA chromosome that is, on average, 103.8% the length of the genome sequence (36). The head-to-tail repeat nature of the concatemer means that such molecules are terminally redundant (e.g., a direct terminal repeat in P22 that is 3.8% the length of the genome sequence [36]), and the fact that each packaging event in a series moves along the sequence by the length of the terminal redundancy means that virion chromosomes are circularly permuted relative to one another. Since the headful

FIG 6 Phage 9NA virion proteins and DNA. (A) A 12% SDS-polyacrylamide gel of 9NA proteins stained with Coomassie brilliant blue. A molecular weight scale is shown on the left, and the N-terminal amino acid sequences determined here are shown on the right (X indicates an ambiguous determination). The putative role of each protein is noted in parentheses on the right. (B) A 1% agarose CHEF electrophoresis gel stained with ethidium bromide. A molecular weight scale is shown on the left, and a standards (Std) lane included a phage lambda chromosome ladder and HindIII-cut phage lambda DNA. The remaining lanes contain whole, uncut DNA from the indicated phage (P22 DNA is 43.5 kbp long, lambda DNA is 48.5 kbp, and phage 244 [GenBank accession no. [NC_008194](#)] DNA is 74.4 kbp).

measurement and, thus, virion chromosome length have some built-in imprecision, the terminal restriction fragments from the linear virion DNAs have somewhat different lengths in different virions. These are spread out when displayed in an electrophoresis gel and are not usually visible as discrete bands. The one exception to this rule is that many phages of this type make a rather precise cut in the DNA near the *pac* site to initiate each packaging series. In these cases, the first end to be packaged from the first packaging event in a series gives rise to a discrete restriction fragment gel band, while all the other end fragments are too variable in length to give a sharp band. Since only the first member of any packaging series has such a defined end, upon cleavage of whole virion DNA by a restriction enzyme, the fragment from this series initiation end is present in submolar amounts relative to the amounts of the other true restriction fragments. It is called the *pac* fragment, and the presence of such a band is diagnostic of this packaging strategy (30, 36, 37).

Terminally redundant and circularly permuted DNA molecules assemble into a circle when sequenced (37), and 9NA, Sasha, Sergei, and Solent DNAs all assembled into circular sequences (see above and reference 18). Analysis of 9NA DNA shows that it has the following three features expected of *pac* site-initiated headful packaging: (i) all the restriction fragments expected from the circular DNA assembly, (ii) a submolar *pac* fragment, and (iii) other terminal restriction fragments which are variable in length due to the imprecision of packaging length measurement. Cleavage of its DNA with several restriction enzymes gave all the bands expected from a circular molecule (not shown), and an additional submolar fragment was present in the digests of those enzymes examined, Bsu36I, BglII, StuI, BsaI, SmaI, and EcoO109I (Fig. 7 and Table 2). In addition, analysis of StuI-Bsu36I and StuI-BsaI double digests yielded submolar fragments the same size as only Bsu36I and StuI, respectively (Fig. 7A), showing that the left ends of these fragments were generated by terminase cleavage and that the right ends were generated by restriction enzyme cleavages. Therefore, packaging proceeds from left to right on the 9NA map (Fig. 2). The lengths of the *pac* fragments indicate that their left

FIG 7 Phage 9NA virion DNA ends. (A) A 1% agarose electrophoresis gel of 9NA virion DNA cleaved with the indicated restriction enzymes. The standard (Std) lane is HindIII-cut phage lambda DNA. White dots at the left of each lane mark *pac* fragments (see the text). (B) A 0.8% agarose electrophoresis gel of 9NA virion DNA cleaved with the indicated restriction enzymes. The standard (Std) lane is HindIII-cut phage lambda DNA. The white vertical lines at the left of each lane mark the extent of the diffuse right-end fragments (which have migrated shorter distances in the gel with increasing headful number; see the text), as seen in the original stained gel. White half circles on the right mark *pac* fragments. The agarose gels were stained with ethidium bromide. (C) 9NA packaging series. A 1.3-genome-length 9NA concatemer section with a *pac* site at its left end is shown above, and a scale (in kilobase pairs) is shown below. Headful numbers 1 and 2 of a packaging series that begins at the *pac* site and progresses rightwards are indicated by horizontal black arrows. The measured lengths of *pac* fragments created by the indicated restriction enzymes are given at the left by the lengths of the horizontal gray bars (data for BglI, SmaI, and EcoO109I are not shown). The lengths of right-end fragments created by the indicated restriction enzymes are given by the lengths of the black bars (with parentheses showing the range of their right-end locations; data for Bsal are not shown).

ends and, thus, the location of packaging series initiation events lie within a few hundred base pairs of the left end of the 9NA genome sequence in GenBank (GenBank accession no. [KJ802832](#)). In addition, BglI, which cleaves 9NA DNA at bp 49135, should give an ~3,700-bp submolar *pac* fragment if packaging can also proceed in the

TABLE 2 9NA terminal DNA fragments

	Fragment length (bp)		
	Right-end fragment ^a		
Headful	1	2	pac fragment
Bsu36I	6,900 (2,100)	6,500 (4,000)	3,700
StuI	— ^b	6,700 (3,800)	6,800
XbaI	6,800 (2,200)	18,000 ^b	
BglI	~7,250 (2,000)	15,000 ^b	
BsaI	— ^b	5,900 (2,100)	7,900
BglII			6,700
SmaI			11,000
EcoO109I			12,500

^aParentheses enclose the measured value for the width of the diffuse right-end fragment band.

^bBecause of the location of the StuI and BsaI sites, they generate no headful number 1 right-end band (Fig. 7C). The BglI and XbaI headful number 2 right-end bands are too large for accurate size measurement in these gels, so no diffuse band length ranges are shown.

opposite direction from the terminase cleavage at the *pac* site, but no such fragment was observed (Fig. 7B). Finally, cleavage of 9NA DNA with a number of other restriction enzymes showed no other submolar fragments. We conclude that DNA packaging initiates at a single *pac* site near the left end of the 9NA genome sequence and proceeds unidirectionally rightward from that point (diagrammed in Fig. 7C).

In order to locate the terminase cleavage site more accurately, 9NA DNA was cleaved with the restriction enzyme Bsu36I, which generated an ~3,800-bp *pac* fragment that was very well separated from the remainder of the DNA (Fig. 7A), and the *pac* fragment was isolated from a 1.5% agarose electrophoresis gel. This DNA was cleaved with restriction endonucleases that should generate considerably shorter *pac* fragments, and the resulting fragments were separated by 6% polyacrylamide gel electrophoresis (Fig. 8A). Diffuse *pac* fragment bands of about 515, 750, and 1,150 bp were observed with the enzymes MluI, PvuI, and BsaWI, respectively, each of which cut the isolated fragment once. A parallel 1.5% agarose electrophoresis gel (not shown) confirmed that the true restriction fragments derived from the right portion of the isolated fragment in these digests were of the expected sizes and were present in the gel as sharp bands. The widths of the short *pac* fragment bands indicate that, like several other headful packaging phages that have been studied in this regard, phages P22, ES18, and Sf6 (28, 29, 38, 39), the 9NA terminase cleaves the DNA imprecisely after recognizing it as a packaging substrate. Measurements of the *pac* fragment band widths (Fig. 8A and B) indicated that the DNA cleavages that initiate 9NA DNA packaging series occur within an approximately 170-bp region located between about bp 150 and bp 320 on the 9NA genome sequence deposited in GenBank (GenBank accession no. [KJ802832](#)). In phages P22 and Sf6, which also do not cleave the DNA precisely during series initiation, the *pac* site that the packaging apparatus recognizes for initiation has been identified and lies approximately at the center of the region containing the individual initiation sites (40, 41). If this is also true for 9NA, the *pac* site should reside at about bp 210 to 240.

To determine directly whether 9NA virion DNA is terminally redundant and circularly permuted, we used contour-clamped homogeneous electric field (CHEF) gel electrophoresis to measure the size of the linear virion DNA molecule and determined it to be 60 ± 1 kbp long (Fig. 6B). The genome sequence of 9NA is 52,869 bp long (18), suggesting that the virion DNA has a terminal redundancy of 6 to 8 kbp. Because of this rather long terminal redundancy, phage 9NA right-end restriction fragments from the different headfuls in the packaging series are separated as diffuse bands in electrophoresis gels, and it is possible to measure the average terminal redundancy more accurately by analyzing the length of such fragments. Figure 7B shows that when 9NA virion DNA is cut by four restriction enzymes that each cut 9NA DNA only once near the *pac* site, DNA fragments that represent right-end fragments generated by virion DNAs from the different headfuls in the packaging series can be visualized in an electropho-

FIG 8 Imprecision of 9NA DNA packaging initiation. (A) Acrylamide electrophoresis gel of the left-end fragments of 9NA DNA. The approximately 3,800-bp 9NA Bsu36I pac fragment DNA (Fig. 7B) was isolated from a 1% agarose electrophoresis gel with a Qiagen Qiaex II gel extraction kit. It was then cleaved with the indicated restriction endonucleases, separated in a 6% acrylamide electrophoresis gel, and stained with ethidium bromide. The horizontal white lines indicate the arbitrarily chosen boundaries for the fuzzy Mlul and Pvul pac fragments; the larger >2,000-bp true restriction fragments that result from cleavage of the right end of the isolated pac fragment are not shown because they were overloaded and did not migrate far from the loading well. (B) Range of packaging series initiation site cleavages. The left end of the reported 9NA genome sequence is indicated by the scale (in base pairs) below. Horizontal black bars represent the measured lengths of the left half of the cleaved Bsu36I pac fragment after cleavage with the indicated enzymes (data for BsaWI are not shown), and the parentheses at their left ends show their measured band widths.

resis gel (the left-end fragments are longer and run as an unseparated band near the top of the gel). These right-end fragments form a ladder of diffuse bands. They are diffuse because the headful measuring device is not precise. Calculations from the location of the restriction enzyme cleavage sites and the measured centers of the size range of molecules in these diffuse bands (indicated in Fig. 7C; Table 2) show that 9NA has an average of 12.9% terminal redundancy of $6,800 \pm 300$ bp (the uncertainty indicates the range of measured average values). All of the diffuse bands visible in the gel fit this value well, but only the four shortest fragments (from Bsu36I headful numbers 1 and 2, Stul headful number 2, and XbaI headful number 1) were used in this calculation because their sizes could be the most accurately measured. The Stul-cut DNA in Fig. 7B shows right-end bands from series headfuls 2 through 5, but any bands from headfuls ≥ 6 would be obscured by the large left-end fragments in the gel, so 9NA packaging series can be at least 5 sequential headfuls long. These data directly demonstrate the circular permutation of 9NA DNA.

The widths of the diffuse right-end fragment DNA bands reflect the innate imprecision in the headful measuring device. Since the width of the right-end band increases

with the headful number in the packaging series (shorter or longer headfuls can randomly follow shorter or longer headfuls in the series; see the work of Casjens and Hayden [36] for a discussion of headful precision), only the right-end width values from headful number 1 correspond directly to the variation in chromosome length. The measured widths generated by three restriction enzymes, BglI, Bsu36I, and XbaI, are given in Table 2. It is difficult to determine the edges of diffuse bands precisely, since loading more material on the gel allows visualization of lower DNA concentrations at the edges, but the values chosen (marked by the white vertical bars in Fig. 7B) include the large majority of the DNA in these bands. The average first headful band width value is $2,100 \pm 100$ bp (the uncertainty is the range of values obtained in three independent measurements). Since about 170 bp of this imprecision is derived from series initiation (see above), we conclude that the 9NA headful-measuring device has a precision of about ± 950 bp ($\pm 1.8\%$).

The high similarity of the terminase packaging motor proteins (TerL), portal proteins, and major capsid proteins of the six *Salmonella* 9NA-like phages suggests that their head sizes are likely identical and that they should package DNA to the same density in their heads. Therefore, they should package the same-sized DNA molecules and their terminal redundancy will vary in accordance with their different genome sequence lengths (Table 1). The 12.9% terminal redundancy for 9NA is longer than that for most other such phages in which it has been carefully measured (P22, 3.8% [36]; Sf6, 6% [28]; ES18, 10% [29]; T4, 3% [42]; and SPP1, 4% [31]), and the $\pm 1.8\%$ value for the precision of the 9NA headful measurement is comparable to that in the few other cases in which it has been measured: $\pm 1.7\%$ in phage P22 (36), about $\pm 2\%$ in SPP1 (43), and about $\pm 3.6\%$ in phage Mu (44).

Comparative genomics of the 9NA-like phages. (i) Phage lifestyle. These are almost certainly lytic phages that are unable to form lysogens. Sasha, Sergei, Solent, and 9NA make large (~ 1 - to 2-mm) clear plaques, and examination of the group's genome sequences shows no indication of such encoded proteins, like integrase, transposase, protelomerase, plasmid partition proteins, or prophage repressors, that play important roles in lysogeny. Bacterial mutants of *S. enterica* serovar Anatum FC1033C3 insensitive to phages Sasha and Sergei have been isolated by culturing colonies surviving phage challenge. Overnight culture supernatants of these strains do not produce spontaneously released phage capable of forming plaques on parent strain FC1033C3, indicating that phage insensitivity is likely not due to lysogen formation. In addition, as we reported previously, there are enough *Salmonella* genome sequences available that close relatives of all known *Salmonella* temperate phages can be found in them (16), and we have found no close relatives of these 9NA-like phages in any bacterial genome sequences. These findings strongly indicate that these are not temperate phages (16).

(ii) Genome organization. The genomes of the seven 9NA-like phages of *Salmonella* are largely syntenic, except for the presence of scattered large indels that generate mosaically related genomes. In most cases, the boundaries of these indels or mosaic sections reside between genes, resulting in the neat replacement of phage genes or gene clusters. For example, 9NA genes 2 to 10 are replaced by genes 2 to 4 in Sasha, and 9NA genes 2 and 3 are replaced by genes 2 to 11 in Sergei. In a few cases, indel boundaries are within predicted genes, resulting in gene fusions or partial gene deletions. For example, the sequence encoding Sasha genes 7 and 8 is inserted in a position corresponding to the 3' end of 9NA gene 12, resulting in a predicted Sasha gp7 protein that has an N-terminal region homologous to 9NA gp12 fused to a new ~ 170 -amino-acid C-terminal domain. This level of genome mosaicism is not atypical for phages in this genome size range, and such mosaicism has been discussed in numerous previous publications (see, for example, references 1, 2, and 45 to 51).

These six phages have an overall gene organization that is not unusual for tailed phages with genomes in the 40- to 60-kbp range. They have a single virion assembly gene cluster and divergently transcribed gene clusters that encode DNA and nucleotide

metabolism proteins. The genome maps in Fig. 2 show from left to right genes involved in virion assembly, lysis, and DNA metabolism. Table S1 provides a list of the predicted genes for all seven 9NA-like phages. The reported genome of PmiS-Isfahan is opened at a location different from that in the six *Salmonella* 9NA-like phages and also contains a significant apparent genomic translocation with respect to the others. PmiS-Isfahan contains two well-separated ORFs that, when joined, form a convincing homologue of 9NA ORF66. When the PmiS-Isfahan genome sequence is rearranged so that these two ORFs are fused to form an ORF66 homologue, the genome is syntenic with the other 9NA-like phages (Fig. 5A). It is not clear if this is a legitimate natural genomic rearrangement or the result of a sequence assembly error.

The putative early genes of the 9NA-like phages are arranged in two divergently transcribed clusters at the right end of the genome sequences. The leftward cluster is predicted to include the DNA polymerase and DNA methylase genes, and the rightward cluster contains predicted helicase, nuclease, primase, and DNA methylase genes. The recognizable putative late virion assembly and lysis genes all reside in the left portion of the genome and are all oriented for rightward transcription, except for the tailspike gene at the right end of this cluster. Genes that are predicted to encode small (TerS) and large (TerL) terminase subunits (the former is tentative, as matches to known small terminases are quite weak), portal protein, procapsid assembly protein (a homologue of phage SPP1 gp7 [52]), MCP, MTP, tape measure protein, a distant homologue of phage lambda gene *J* tail tip protein, and a tailspike protein. In addition, probable tail assembly chaperone proteins are predicted from the facts that they lie immediately transcriptionally upstream of the putative tape measure protein gene and that one appears to be expressed through a programmed translational frameshift from the upstream gene (53–55). A second potential translational frameshift site was detected in genes just downstream of the tail tape measure gene. The homologous phage 9NA gene 53, gene 49 in Sasha, and gene 59 in Sergei/Solent are annotated with independent start codons; however, the potential slippery sequence TTTCCC(A/T) is conserved within the C termini of the upstream genes, and each of these could facilitate a –1 frameshift into the gp53, gp49, or gp59 reading frame, respectively. The possibility of programmed frameshifts in these genes is strengthened by the fact that the most optimally placed ATG start codons for these genes are not associated with detectable Shine-Dalgarno sequences. The nearest downstream start codon with a possible Shine-Dalgarno sequence leaves an intergenic gap of ~60 bp in phages Sasha, Sergei, and Solent, and this start is not conserved in 9NA. The location of these genes within the tail morphogenesis module just downstream of the tail tape measure gene might suggest a possible role as a chaperone, similar to the role of the tape measure chaperones, which are known to function via translational frameshifting (53). Without further evidence of a frameshift, however, these genes have been annotated with discrete starts.

A multiple-gene lysis cassette is present at the same location in all seven phages. A predicted endopeptidase-type phage endolysin (IPR039561) was found to be conserved in all seven phages. A pair of genes comprising the phage spanin complex was identified downstream of the endolysin in all seven phages, characterized by an inner-membrane spanin component (i-spanin) with a single N-terminal transmembrane domain (TMD) and an outer-membrane spanin component (o-spanin) with an N-terminal SPII lipoprotein processing signal. In all cases, the o-spanin gene is partially embedded within the i-spanin gene, which is a common gene arrangement for two-component spanins in phages infecting Gram-negative bacteria (56). In the six *Salmonella* 9NA-like phages, the endolysin is flanked by two genes encoding TMD-containing proteins; the gene upstream of the endolysin (9NA gene 56 and its homologues) encodes a protein with a single predicted N-terminal TMD, and the downstream gene (9NA gene 58 and its homologues) encodes a protein with two TMDs. Given the presence of TMDs and the colocalization of these genes with the endolysin and spanin, these likely function as a holin-antiholin pair that releases the endolysin during cell lysis (57). In PmiS-Isfahan, the lysis cassette arrangement is slightly different, with the

endolysin preceded by two genes encoding TMD-containing proteins—the first with four TMDs and the second with three—that also likely function as a holin-antiholin pair. At this time, it is not possible to assign a specific holin or antiholin function to these proteins.

The virion assembly genes of the seven 9NA-like phages are positioned in the same order and orientation as in many other tailed phages (including the canonical order of small terminase-large terminase-portal-MCP-MTP-tail chaperone-tail tape measure-tail tip-tailspike genes [26, 58]); however, these phages are unusual in that (i) the tailspike genes are transcribed in the direction opposite that for the other virion assembly genes and (ii) there are a large number of small putative genes of mostly unknown function interspersed among the virion assembly genes (Fig. 2). The latter include genes that encode putative dUTP nucleotidohydrolase and a phage lambda *ninH* homologue in all six *Salmonella* phages, a RdgC-type recombination modulator protein/exonuclease (59) in Sergei, Solent, and FSL_SP-062 and -069, and a cytidine deaminase in PmiS-Isfahan. Many of these interspersed genes have a transcriptional orientation opposite that of the virion assembly genes, suggesting that their expression is likely independent of the assembly genes that are expected to be transcribed late in the infection cycle. A substantial fraction of these interspersed genes is not present in all seven genomes (Fig. 2; Table S1). This suggests that many of these genes are likely parts of evolutionarily mobile morons (46, 47) in the late gene cluster whose products are not involved in virion assembly but may be involved in optimizing infection in other ways. It is possible that there is a single late rightward promoter for the assembly genes that directs transcription across this region that may not express the interspersed moron genes, as is true, for example, of the phage P22 *sieA-imm1* region genes (60–62), the phage lambda *bor* and *lom* genes (63, 64), and phage HK97 genes 22 and 23 (47). On the other hand, the large number of such interspersed genes might suggest that a number of separate late genes or gene clusters may have independent late promoters.

Analysis of the sequence of these phages showed a number of repeated sequences in their genomes. First, in 9NA there are 11 very similar but imprecise repeats of part or all of the 26-bp nonpalindromic sequence 5'-TGTAATCCGTCAGAATTATTTAAA that are 14 bp or longer; they are scattered across the genome, and their locations are indicated for 9NA in Fig. 2 as repeat number 1 (indicated by solid triangles). The Sasha and Sergei/Solent genomes contain 11 and 13 copies of the same repeat family, respectively, and they lie at essentially the same locations as those in 9NA. The role of these sequences is unknown, but they reside largely between genes and are present in both orientations. In most cases, the orientation correlates with the predicted local direction of transcription, although the ones at 9NA bp 24420 and bp 45053 are apparent exceptions to this rule. This repeat is also present in the FSL_SP-062 and FSL_SP-069 genomes, where it lies at a majority of their contig ends, suggesting that it may have been this family of repeats that confounded the complete assembly of these genomes. We have not identified any such widely distributed repeat in the PmiS-Isfahan genome. There are also several tracts of tandem repeats present in these phages. Variable numbers (3.8 to 7) of the 30-bp sequence 5'-ACAGGAGCGGCTACAG GCGCAGGTGTCGCA reside inside the C-terminal region of 9NA gene 66 and its orthologs. The role of this repeat sequence is unknown, but the location in the gene adjacent to the DNA polymerase gene might suggest a possible replication origin function analogous to the phage lambda origin repeats within its *O* gene (see reference 48 and references therein). Other short tandem repeats of unknown function lie within (i) 9NA gene 10 (only phages Sergei and Solent have an orthologue of this gene), (ii) 9NA gene 46 and its (nonsyntenic) homologues in Sasha, FSL_SP-062, and FSL_SP-069, and (iii) Sasha gene 81 and its homologue Sergei/Solent gene 91.

Gene content variation within the 9NA-like phages. The seven 9NA-like phages carry between 80 and 99 predicted genes (Table S1). We note that FSL_SP-062 and -069 differ by only about 4 bp (19) and that Sergei and Solent are very similar, but there are considerable differences in gene content among the other phages in this group, and

their genes comprise a current 9NA-like phage cluster pangenome of 176 different genes (Table S1). Of these 176 genes, 40 are present in all seven genomes and form the cluster's current core genome; these core genes include all of the identified virion assembly and DNA metabolism genes. The mosaic nature of the genomes within this group is demonstrated by the fact that of the remaining 136 noncore or accessory genes, 10 are present in three of phages, 37 are present in two of the phages, and 57 are present in only one phage (Table S1). We also note that these accessory genes are, on average, much smaller than the core genes. In the 176-gene pangenome, 32 genes (23 of which are in the core genome) have a predicted function and 143 have no known function. Among genes without a predicted function, 68 have no matches outside this cluster in the current database of annotated proteins, and for those that do, most matches are present in other phages and prophages (Table S1). The fact that many of the genes of unknown function are conserved within the phage cluster and/or have homologues in other phages strengthens the idea that these mostly small open reading frames are, in fact, functional genes.

Horizontal movement of tailspike genes. Phage 9NA was the first long-tailed phage found to carry a virion endorhamnosidase activity (22). Such glycanase activities are known to be present in the host receptor-binding tailspike of the short-tailed phage P22 (65, 66), whose receptor is the O-antigen polysaccharide portion of the *Salmonella* serovar Typhimurium surface lipopolysaccharide (67). We previously discovered that the C-terminal ~540 amino acids (amino acids 136 to 673) of the 9NA gene 55 tailspike protein are 35% identical to the receptor-binding domain of the well-studied P22 tailspike (49). Andres et al. (23) studied the 9NA tailspike C-terminal domain (amino acids 131 to 673) and found it to have O-antigen cleavage activity and the same overall polypeptide fold as the P22 tailspike protein. Like phage P22, in which the N-terminal ~115 tailspike amino acids form a separate virion-binding domain (68, 69), the 9NA-like phages have 57 N-terminal amino acids whose sequence is very similar and unique to this group of phages; in this region, the six *Salmonella* 9NA-like phages are all $\geq 90\%$ identical to one another and about 60% identical to the parallel PmIS-Isfahan region. Thus, it is nearly certain that, like the tailspikes of P22, the tailspikes of the 9NA-like group bind to virions through their common N-terminal regions and bind their different cell surface receptors through their very different C-terminal domains.

Phages P22 and 9NA both infect the *S. enterica* serovars, such as *S. enterica* serovar Typhimurium, that have O:4 type O-antigen surface polysaccharide (O-antigen structures have been reviewed previously [70]), so the similarity of their C-terminal receptor-binding domains is not unexpected. On the other hand, FSL_SP-062 and FSL_SP-069 infect *S. enterica* serovar Newport (O:8 O antigen) (19), and their 99% identical predicted tailspike C-terminal domains are only very distantly related to those of 9NA (12% identical). Similarly, Sasha, Sergei, and Solent infect *S. enterica* serovar Anatum (O:3,10 O antigen) (21), and their 99% identical C-terminal domains are only 15 to 16% identical to those of 9NA and FSL_SP-062/FSL_SP-069. Finally, the parallel domain of PmIS-Isfahan is only 14 to 16% identical to that of the other six phages. As we have discussed previously (16), tailspike domains that bind very different polysaccharides typically have essentially unrelated amino acid sequences. The O:4, O:8, and O:3,10 O-antigen polysaccharide backbone repeats are -mannose (abequose)-rhamnose-galactose-, -rhamnose (abequose)-mannose-mannose-galactose (glucose)-, and -mannose-rhamnose-acetylgalactose-, respectively (parentheses indicate side chain sugars, and we note that the similar but not identical O:4 and O:3,10 trimer repeats are joined by different linkages) (70), so the differences among these 9NA-like phage tailspikes conform to that pattern. Figure 9 diagrams these putative tailspike domain arrangements. We note that an additional ~65-amino-acid region between the highly similar N-terminal regions (see above) and the C-terminal receptor-binding domain has homologues at a similar location in the Vi01-like and E1-like phage tailspikes; perhaps this forms a third domain that links the other two domains in various different phage contexts.

FIG 9 Tailspike domains. The sequence relationships of various phage tailspikes are shown, where the same color indicates a similar sequence; different-colored sequences are essentially unrelated (see the text), and the N- and C-terminal locations are indicated above. Note that the phage P22-like N-terminal well-known virion-binding regions (97) are very similar, while their C-terminal receptor-binding domains are very different. Other putative N-terminal minimal capsid binding domains, identified by their constant nature within each cluster, lie to the left of the vertical red line; the roles of the short blue and purple sections between the capsid binding domain and the receptor-binding domain are not known. Phage names are shown on the left, and the phage type or cluster is indicated on the right (see reference 15). The locus_tags of the *Proteus* (*) and *Salmonella* prophage (†) tailspikes are BN1805_00524 and SEEN470_12751, respectively (Table 3). Phages SKML-39 and Det7 have multiple tailspikes, so the encoding ORF is also indicated on the left. Below, a color key indicates the host that the phage or prophage infects.

The 9NA-like tailspikes demonstrate the horizontal exchange of protein domains among very different phage types particularly well. A sample of the proteins in the extant database that are $\geq 34\%$ identical to any one of the four types of 9NA-like phage tailspike C-terminal domains is listed in Table 3, and some of these tailspikes are shown in Fig. 9. In each case, matches are present in the tailspikes of other phage types. The O:4 and O:8 O-antigen binding domains are both present in several other tailed phage types (for definitions of these types or clusters, see reference 15), the O:3,10 binding domain is present in two other types. The only homologous relative of the PmiS-Isfahan C-terminal domain in the current database is present in a *Proteus vulgaris* epsilon15-like prophage; the fact that PmiS-Isfahan infects *P. mirabilis* suggests that the PmiS-Isfahan host and prophage-harboring *Proteus* strains likely have the same or very similar surface polysaccharides. The phages of other types that carry these tailspikes are three very different types of *Podoviridae* (P22-, epsilon15-, and SP6-like clusters), four very different types of *Siphoviridae* (9NA-, SETP3-, E1-, and ENT47670-like clusters; the last one is listed as a member of the *Myoviridae* in its GenBank annotation [GenBank accession no. [HQ201308](#)], but our analysis of its sequence suggests that it is likely a member of the *Siphoviridae*) and one type of *Myoviridae* (Vi01-like cluster) (only the Vi01-, P22-, E1-, and epsilon15-like tailspikes are shown in Fig. 9). Clearly, these tailspike domains have moved horizontally among a number of very different tailed phage groups. Thus, because of their high rate of horizontal transfer, differences or similarities in the receptor-binding domain of tailspikes should be given less weight when deciding to taxonomically separate or merge tailed phage groups.

TABLE 3 Tailspike horizontal exchange among phage types

Phage	Host bacteria ^a	Phage cluster ^b	% identity ^c	GenBank protein name/locus tag
9NA	O:4 <i>S. serovar Typhimurium</i>	9NA	100	ST9NA_055
P22	O:4 <i>S. serovar Typhimurium</i>	P22	37	gp9
SPN9TCW	O:4 <i>S. serovar Typhimurium</i>	epsilon15	39	SPN9TCW_021
SETP3	O:9 <i>S. serovar Enteritidis</i>	SETP3	36	gp32
DET7 ^e	O:4 <i>S. serovar Typhimurium</i>	Vi01	40	DET7_207
SP6 ^e	O:4 <i>S. serovar Typhimurium</i>	SP6	34	gp49
LPST10	O:4 <i>S. serovar Typhimurium</i>	E1	43	LPST10_00009
FSL_SP-069	O:8 <i>S. serovar Newport</i>	9NA	100	SP069_00275
EMEK	O:8 <i>S. serovar Haardt</i>	P22	82	B606_gp25
Prophage ^d	O:8 Newport strain SL317	ENT47670	53	SNSL317_A2991
Prophage ^d	O:8 <i>S. serovar Newport</i> strain CVM 19470	epsilon15	93	SEEN470_12751
SKML-39	Serovar not reported	Vi01	55	G178_gp001
SP6 ^e	<i>S. serovar Typhimurium</i> and <i>S. serovar Newport</i>	SP6	52	gp47
Sasha	<i>S. serovar Anatum</i>	9NA	100	CPTSasha_51
DET7 ^e	O:4 <i>S. serovar Typhimurium</i>	Vi01	92	DET7_208
Epsilon15	O:3,10 <i>S. serovar Anatum</i>	epsilon15	41	gp20
PmiS-Isfahan	<i>P. mirabilis</i>	9NA	100	Tailspike protein
Prophage ^d	<i>P. vulgaris</i>	epsilon15	61	BN1805_00524

^aAll bacteria are *Salmonella enterica* unless otherwise indicated. The O-antigen type and *Salmonella* serovar are indicated (note O:4 and O:9 have closely related O antigens, and phage P22, for example, can adsorb to and infect both types).

^bTailed phage cluster or type according to Grose and Casjens (15).

^cPercent identity to the receptor-binding domain of the phage shown in bold at the top of each group. The receptor-binding domains of the tailspikes of the four groups are essentially unrelated.

^dThese prophages were identified in this study. They can be found through the locus_tag of their tailspike gene.

^eDet7 and SP6 adsorb to and infect different hosts by virtue of different having multiple tailspike proteins (98, 99).

Summary. We have analyzed the newly sequenced genomes of *S. enterica* serovar Anatum phages Sasha, Sergei, and Solent and found that they are closely related to *S. enterica* serovar Typhimurium phage 9NA and *S. enterica* serovar Newport phages FSL_SP-062 and FSL_SP-069. These six phages are also similar to but more distantly related to *P. mirabilis* phage PmiS-Isfahan. We show that these phages comprise a cluster that is well separated from other tailed phage types. The level of diversity within this group and the distances between it and other tailed phage clusters are comparable to those in such previously suggested genera as *Viunlikevirus* (71), *Tunlikevirus* (72), and *rV5likevirus* (73). To characterize this novel cluster of *Siphoviridae* phages, we identified the major virion protein-encoding genes and showed that 9NA (and, by extension, almost certainly the whole group) package DNA by a *pac* site-mediated headful packaging strategy. In addition, we show that 9NA initiates DNA packaging by cleaving its DNA within an approximately 170-bp region near the *pac* site and that the precision of its headful measuring device is about ± 950 bp.

MATERIALS AND METHODS

Bacteriophage and bacterial strains and their propagation. *S. enterica* serovar Typhimurium LT2 DB7000 is described by Winston et al. (74), and *S. enterica* serovar Anatum strain FC1033C3 was isolated in 2014 from cattle feces collected in a south Texas beef feedlot (21). Phages Sasha and Sergei were collected from separate pens in the same feedlot in 2014, and Solent was collected from a different feedlot about 100 miles away in 2016. They were isolated by enrichment of environmental samples against a mixed-host panel containing five *Salmonella* strains of different serotypes as described previously (21). Phages Sasha, Sergei, and Solent were propagated on *S. enterica* serovar Anatum FC1033C3 by the soft agar overlay method (75) using T-top soft agar (Bacto tryptone [10 g/liter], NaCl [10 g/liter], Bacto agar [5 g/liter]) over tryptone soy agar plates. The bacterial strains and phages were cultured aerobically at 37°C in LB broth, tryptone soy broth (TSB; Difco), or tryptone soy agar (TSB plus 15 g/liter Bacto agar [Difco]).

Virion and virion DNA purification. A single 9NA plaque on *Salmonella enterica* serovar Typhimurium LT2 was used to inoculate a 300-ml LB broth culture of strain DB7000 grown to 5×10^7 cells/ml. The culture was shaken at 37°C for 4 h, at which point partial lysis occurred. The culture was then shaken with chloroform to complete lysis, and the cell debris was removed by centrifugation in a Beckman GS3 rotor for 20 min at 8,000 rpm. The virions were pelleted for 12 h at 8,000 rpm in the same rotor. The resulting pellet was resuspended slowly with gentle shaking in TM (10 mM Tris-HCl, pH 7.4, 1 mM MgCl₂) at 4°C

overnight, and any remaining aggregated material was removed by low-speed centrifugation. The supernatant was applied to a CsCl step gradient (76) and spun for 4 h at 20,000 rpm in a Beckman SW41 ultracentrifuge rotor. The visibly opalescent virion band at ~ 1.5 g/cm³ was removed with a syringe by puncturing the side of the tube, dialyzed against TM, and stored at 4°C. DNA was isolated from virions as previously described (37). Phage Sasha, Sergei, and Solent DNAs were isolated directly from high-titer ($>10^9$ -PFU/ml) plate lysates using a modified Promega Wizard protocol as described previously (77).

DNA sequencing. Phages Sasha, Sergei, and Solent were sequenced by use of an Illumina MiSeq sequencer in a 250-bp paired-end run with 550-bp inserts at the University of Texas Genomic Sequencing and Analysis Facility, Austin, TX. Sequence reads were trimmed for quality with the FastX Toolkit (hannonlab.cshl.edu) and assembled with the SPAdes (version 3.5.0) program (78). The contigs had 59-fold, 103-fold, and 380-fold coverage for Sasha, Sergei, and Solent, respectively; the sequences were confirmed to be complete by PCR amplification using primers that faced off each end of the contig and sequencing of the resulting product, resulting in final circular contigs of 53,263 bp for Sasha, 56,051 bp for Sergei, and 55,978 bp for Solent. The Sasha and Sergei genomes were each reopened at a position ~ 275 to 350 bp upstream of the predicted *terS* gene to maximize alignment with the phage 9NA sequence.

Bioinformatic methods. Gene prediction was conducted with the MetaGeneAnnotator (79) and Glimmer3 (80) programs, followed by manual curation. Functional annotations were assigned using a combination of BLASTp analysis of the sequences against those in the GenBank nr database (81) and the InterProScan (version 4.7) database (82) and HHPred server (83). Protein transmembrane domains were predicted with the TMHMM server (84), and protein secretion and processing signals were detected using the SignalP (version 4.0) (85) and LipoP (version 1.0) (86) servers. Rho-independent transcriptional terminators were predicted with the TransTerm HP database (87). All analyses were conducted via the Galaxy interface (88), hosted by the Center for Phage Technology at Texas A&M University (cpt.tamu.edu). The following computer programs were used to compare DNA sequences: BLASTn (89), Clustal X (90), DNA Strider (91), Gepard (92), and the Artemis comparison tool (93).

Transmission electron microscopy. Phages were prepared for microscopy by the method of Valentine et al. (94) and stained with 2% (wt/vol) uranyl acetate. Samples were viewed in a JEOL 1200 EX transmission electron microscope under a 100-kV accelerating voltage at the Microscopy and Imaging Center, Texas A&M University. Magnification was calibrated using a carbon grating replica (Ted Pella number 607), and 10 to 15 virions of each phage were measured, with these data being used to calculate mean dimensions.

Electrophoresis and protein methods. Polyacrylamide gel electrophoresis of proteins in sodium dodecyl sulfate (SDS), staining with Coomassie brilliant blue R-250 (Bio-Rad, Richmond, CA), and determination of the N-terminal amino acid sequences of isolated protein bands were performed by the University of Utah Protein Facility as previously described (95). Polyacrylamide gel electrophoresis of DNA was performed as described by Casjens and Huang (38). Constant-voltage agarose gel electrophoresis and contour-clamped homogeneous electric field (CHEF) electrophoresis of the DNA molecules were performed as previously described (96). The CHEF gel pulse times and angles at 6 V/cm were calculated by the Bio-Rad CHEF MAPPER XA software to maximally separate DNAs between 5 kb and 60 kb with a ramping constant of +100. DNA bands were isolated from the agarose gels with a Qiagen Qiaex II gel extraction kit.

Data availability. The sequences of phages Sasha, Sergei, and Solent were deposited in GenBank under accession numbers [KX987158](#), [KY002061](#), and [MH586730](#), respectively.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/JVI.00848-19>.

SUPPLEMENTAL FILE 1, XLSX file, 0.03 MB.

ACKNOWLEDGMENTS

We thank Charles Miller for the kind gift of phage 9NA.

This work was supported by NIH grants GM114817 to S.R.C. and GM51975 to R.W.H. and NSF grant DBI1565146 and funding from the National Cattlemen's Beef Association, the Texas Beef Council, and Beef Checkoff to J.J.G.; C.Z. was supported in part by a fellowship from the China Scholarship Council.

Conception, design, bioinformatics analysis, and preparation of the manuscript were carried out by C.Z., J.J.G., R.W.H., and S.R.C. Data were acquired by C.Z., J.J.G., E.B.G., Y.X., M.J.J., and S.R.C.

REFERENCES

- Casjens SR. 2005. Comparative genomics and evolution of the tailed-bacteriophages. *Curr Opin Microbiol* 8:451–458. <https://doi.org/10.1016/j.mib.2005.06.014>.
- Hendrix RW. 2002. Bacteriophages: evolution of the majority. *Theor Popul Biol* 61:471–480. <https://doi.org/10.1006/tpbi.2002.1590>.
- Hatfull GF, Hendrix RW. 2011. Bacteriophages and their genomes. *Curr Opin Virol* 1:298–303. <https://doi.org/10.1016/j.coviro.2011.06.009>.
- Suttle CA. 2007. Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* 5:801–812. <https://doi.org/10.1038/nrmicro1750>.
- Boyd EF. 2012. Bacteriophage-encoded bacterial virulence factors and

- phage-pathogenicity island interactions. *Adv Virus Res* 82:91–118. <https://doi.org/10.1016/B978-0-12-394621-8.00014-5>.
6. Canchaya C, Fournous G, Brussow H. 2004. The impact of prophages on bacterial chromosomes. *Mol Microbiol* 53:9–18. <https://doi.org/10.1111/j.1365-2958.2004.04113.x>.
 7. Casjens S, Hendrix R. 2005. Bacteriophages and the bacterial genome, p 39–52. In Higgins NP (ed), *The bacterial chromosome*. ASM Press, Washington, DC.
 8. Cardinale D, Carette N, Michon T. 2012. Virus scaffolds as enzyme nano-carriers. *Trends Biotechnol* 30:369–376. <https://doi.org/10.1016/j.tibtech.2012.04.001>.
 9. Hyman P. 2012. Bacteriophages and nanostructured materials. *Adv Appl Microbiol* 78:55–73. <https://doi.org/10.1016/B978-0-12-394805-2.00003-8>.
 10. Chang JR, Song EH, Nakatani-Webster E, Monkkonen L, Ratner DM, Catalano CE. 2014. Phage lambda capsids as tunable display nanoparticles. *Biomacromolecules* 15:4410–4419. <https://doi.org/10.1021/bm5011646>.
 11. Sharma J, Uchida M, Miettinen HM, Douglas T. 2017. Modular interior loading and exterior decoration of a virus-like particle. *Nanoscale* 9:10420–10430. <https://doi.org/10.1039/c7nr03018e>.
 12. Abedon ST, Kuhl SJ, Blasdel BG, Kutter EM. 2011. Phage treatment of human infections. *Bacteriophage* 1:66–85. <https://doi.org/10.4161/bact.1.2.15845>.
 13. Borysowski J, Miedzybrodzki R, Gorski A. 2014. *Phage therapy: current research and applications*. Caister Academic Press, Norfolk, United Kingdom.
 14. Young R, Gill JJ. 2015. MICROBIOLOGY. Phage therapy redux—what is to be done? *Science* 350:1163–1164. <https://doi.org/10.1126/science.aad6791>.
 15. Grose JH, Casjens S. 2014. Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family *Enterobacteriaceae*. *Virology* 468–470:421–443. <https://doi.org/10.1016/j.virol.2014.08.024>.
 16. Casjens SR, Grose JH. 2016. Contributions of P2- and P22-like prophages to understanding the enormous diversity and abundance of tailed bacteriophages. *Virology* 496:255–276. <https://doi.org/10.1016/j.virol.2016.05.022>.
 17. Gilcrease EB, Casjens SR. 2018. The genome sequence of *Escherichia coli* tailed phage D6 and the diversity of *Enterobacteriales* circular plasmid prophages. *Virology* 515:203–214. <https://doi.org/10.1016/j.virol.2017.12.019>.
 18. Casjens SR, Leavitt JC, Hatfull GF, Hendrix RW. 2014. Genome sequence of *Salmonella* phage 9NA. *Genome Announc* 2:e00531-14. <https://doi.org/10.1128/genomeA.00531-14>.
 19. Moreno Switt AI, Orsi RH, den Bakker HC, Vongkamjan K, Altier C, Wiedmann M. 2013. Genomic characterization provides new insight into *Salmonella* phage diversity. *BMC Genomics* 14:481. <https://doi.org/10.1186/1471-2164-14-481>.
 20. Yazdi M, Bouzari M, Ghaemi EA. 24 August 2018. Genomic analyses of a novel bacteriophage (VB_PmiISfahan) within *Siphoviridae* family infecting *Proteus mirabilis*. *Genomics* <https://doi.org/10.1016/j.ygeno.2018.08.008>.
 21. Xie Y, Savell JW, Arnold AN, Gehring KB, Gill JJ, Taylor TM. 2016. Prevalence and characterization of *Salmonella enterica* and *Salmonella* bacteriophages recovered from beef cattle feedlots in south Texas. *J Food Prot* 79:1332–1340. <https://doi.org/10.4315/0362-028X.JFP-15-526>.
 22. Wollin R, Eriksson U, Lindberg AA. 1981. *Salmonella* bacteriophage glycanases: endorhamnosidase activity of bacteriophages P27, 9NA, and KB1. *J Virol* 38:1025–1033.
 23. Andres D, Roske Y, Doering C, Heinemann U, Seckler R, Barbirz S. 2012. Tail morphology controls DNA release in two *Salmonella* phages with one lipopolysaccharide receptor recognition system. *Mol Microbiol* 83:1244–1253. <https://doi.org/10.1111/j.1365-2958.2012.08006.x>.
 24. Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H, Libby SJ, Fang FC. 2006. Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science* 313:236–238. <https://doi.org/10.1126/science.1128794>.
 25. Di Pilato V, Chiarelli A, Boinett CJ, Riccobono E, Harris SR, D'Andrea MM, Thomson NR, Rossolini GM, Giani T. 2016. Complete genome sequence of the first KPC-type carbapenemase-positive *Proteus mirabilis* strain from a bloodstream infection. *Genome Announc* 4:e00607-16. <https://doi.org/10.1128/genomeA.00607-16>.
 26. Casjens S, Hendrix R. 1988. Control mechanisms in dsDNA bacteriophage assembly, p 15–91. In Calendar R (ed), *The bacteriophages*, vol 1. Plenum Press, New York, NY.
 27. Flinta C, Persson B, Jornvall H, von Heijne G. 1986. Sequence determinants of cytosolic N-terminal protein processing. *Eur J Biochem* 154:193–196. <https://doi.org/10.1111/j.1432-1033.1986.tb09378.x>.
 28. Casjens S, Winn-Stapley DA, Gilcrease EB, Morona R, Kühlewein C, Chua JEH, Manning PA, Inwood W, Clark AJ. 2004. The chromosome of *Shigella flexneri* bacteriophage Sf6: complete nucleotide sequence, genetic mosaicism, and DNA packaging. *J Mol Biol* 339:379–394. <https://doi.org/10.1016/j.jmb.2004.03.068>.
 29. 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. <https://doi.org/10.1128/JB.187.3.1091-1104.2005>.
 30. Jackson EN, Jackson DA, Deans RJ. 1978. EcoRI analysis of bacteriophage P22 DNA packaging. *J Mol Biol* 118:365–388. [https://doi.org/10.1016/0022-2836\(78\)90234-6](https://doi.org/10.1016/0022-2836(78)90234-6).
 31. Morelli G, Fisseau C, Behrens B, Trautner TA, Luh J, Ratcliff SW, Allison DP, Ganesan AT. 1979. The genome of *B. subtilis* phage SPP1: the topology of DNA molecules. *Mol Gen Genet* 168:153–161. <https://doi.org/10.1007/bf00431441>.
 32. Tye BK, Huberman JA, Botstein D. 1974. Non-random circular permutation of phage P22 DNA. *J Mol Biol* 85:501–528. [https://doi.org/10.1016/0022-2836\(74\)90312-x](https://doi.org/10.1016/0022-2836(74)90312-x).
 33. Bachi B, Arber W. 1977. Physical mapping of BglIII, BamHI, EcoRI, HindIII and PstI restriction fragments of bacteriophage P1 DNA. *Mol Gen Genet* 153:311–324. <https://doi.org/10.1007/bf00431596>.
 34. Streisinger G, Emrich J, Stahl MM. 1967. Chromosome structure in T4. III. Terminal redundancy and length determination. *Proc Natl Acad Sci U S A* 57:292–295. <https://doi.org/10.1073/pnas.57.2.292>.
 35. Adams MB, Hayden M, Casjens S. 1983. On the sequential packaging of bacteriophage P22 DNA. *J Virol* 46:673–677.
 36. Casjens S, Hayden M. 1988. Analysis *in vivo* of the bacteriophage P22 headful nuclease. *J Mol Biol* 199:467–474. [https://doi.org/10.1016/0022-2836\(88\)90618-3](https://doi.org/10.1016/0022-2836(88)90618-3).
 37. Casjens S, Gilcrease D. 2009. Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions, p 91–111. In Clokie M, Kropinski A (ed), *Bacteriophages: methods and protocols*. Humana Press, Totowa, NJ.
 38. Casjens S, Huang WM. 1982. Initiation of sequential packaging of bacteriophage P22 DNA. *J Mol Biol* 157:287–298. [https://doi.org/10.1016/0022-2836\(82\)90235-2](https://doi.org/10.1016/0022-2836(82)90235-2).
 39. Casjens S, Huang WM, Hayden M, Parr R. 1987. Initiation of bacteriophage P22 DNA packaging series. Analysis of a mutant that alters the DNA target specificity of the packaging apparatus. *J Mol Biol* 194:411–422. [https://doi.org/10.1016/0022-2836\(87\)90671-1](https://doi.org/10.1016/0022-2836(87)90671-1).
 40. Leavitt JC, Gilcrease EB, Wilson K, Casjens SR. 2013. Function and horizontal transfer of the small terminase subunit of the tailed bacteriophage Sf6 DNA packaging nanomotor. *Virology* 440:117–133. <https://doi.org/10.1016/j.virol.2013.02.023>.
 41. Wu H, Sampson L, Parr R, Casjens S. 2002. The DNA site utilized by bacteriophage P22 for initiation of DNA packaging. *Mol Microbiol* 45:1631–1646. <https://doi.org/10.1046/j.1365-2958.2002.03114.x>.
 42. Mosig G, Eiserling F. 2006. T4 and related phages: structure and development, p 225–267. In Calendar R (ed), *The bacteriophages*, 2nd ed. Oxford University Press, New York, NY.
 43. Humphreys GO, Trautner TA. 1981. Maturation of bacteriophage SPP1 DNA: limited precision in the sizing of mature bacteriophage genomes. *J Virol* 37:832–835.
 44. Chow L, Bukhari AI. 1977. DNA insertion elements, p 295–306. In Bukhari AI, Shapiro J, Adhya S (ed), *Plasmids and episomes*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
 45. Casjens S, Hatfull G, Hendrix R. 1992. Evolution of dsDNA tailed-bacteriophage genomes. *Semin Virol* 3:383–397.
 46. Hendrix RW, Lawrence JG, Hatfull GF, Casjens S. 2000. The origins and ongoing evolution of viruses. *Trends Microbiol* 8:504–508. [https://doi.org/10.1016/S0966-842X\(00\)01863-1](https://doi.org/10.1016/S0966-842X(00)01863-1).
 47. Juhala RJ, Ford ME, Duda RL, Youton 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. <https://doi.org/10.1006/jmbi.2000.3729>.
 48. Hendrix R, Casjens S. 2006. Bacteriophage λ and its genetic neighborhood, p 409–447. In Calendar R (ed), *The bacteriophages*, 2nd ed. Oxford University Press, New York, NY.
 49. Casjens SR, Thuman-Commikey PA. 2011. Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22

- virion assembly. *Virology* 411:393–415. <https://doi.org/10.1016/j.virol.2010.12.046>.
50. Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG, Jacobs WR, Hendrix RW, Lawrence JG, Hatfull GF. 2015. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *Elife* 4:e06416. <https://doi.org/10.7554/eLife.06416>.
 51. Gregory AC, Solonenko SA, Ignacio-Espinoza JC, LaButti K, Copeland A, Sudek S, Maitland A, Chittick L, Dos Santos F, Weitz JS, Worden AZ, Woyke T, Sullivan MB. 2016. Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. *BMC Genomics* 17:930. <https://doi.org/10.1186/s12864-016-3286-x>.
 52. Droge A, Santos MA, Stiege AC, Alonso JC, Lurz R, Trautner TA, Tavares P. 2000. Shape and DNA packaging activity of bacteriophage SP1 procapsid: protein components and interactions during assembly. *J Mol Biol* 296:117–132. <https://doi.org/10.1006/jmbi.1999.3450>.
 53. Xu J, Hendrix RW, Duda RL. 2004. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Mol Cell* 16:11–21. <https://doi.org/10.1016/j.molcel.2004.09.006>.
 54. Xu J, Hendrix RW, Duda RL. 2013. A balanced ratio of proteins from gene G and frameshift-extended gene GT is required for phage lambda tail assembly. *J Mol Biol* 425:3476–3487. <https://doi.org/10.1016/j.jmb.2013.07.002>.
 55. Xu J, Hendrix RW, Duda RL. 2014. Chaperone-protein interactions that mediate assembly of the bacteriophage lambda tail to the correct length. *J Mol Biol* 426:1004–1018. <https://doi.org/10.1016/j.jmb.2013.06.040>.
 56. Kongari R, Rajaure M, Cahill J, Rasche E, Mijalis E, Berry J, Young R. 2018. Phage spanins: diversity, topological dynamics and gene convergence. *BMC Bioinformatics* 19:326. <https://doi.org/10.1186/s12859-018-2342-8>.
 57. Cahill J, Young R. 2019. Phage lysis: multiple genes for multiple barriers. *Adv Virus Res* 103:33–70. <https://doi.org/10.1016/bs.aivir.2018.09.003>.
 58. Casjens S. 2003. Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* 49:277–300. <https://doi.org/10.1046/j.1365-2958.2003.03580.x>.
 59. Briggs GS, Yu J, Mahdi AA, Lloyd RG. 2010. The RdcC protein employs a novel mechanism involving a finger domain to bind to circular DNA. *Nucleic Acids Res* 38:6433–6446. <https://doi.org/10.1093/nar/gkq509>.
 60. Pedulla ML, Ford ME, Karthikeyan T, Houtz JM, Hendrix RW, Hatfull GF, Poteete AR, Gilcrease EB, Winn-Stapley DA, Casjens SR. 2003. Corrected sequence of the bacteriophage P22 genome. *J Bacteriol* 185:1475–1477. <https://doi.org/10.1128/jb.185.4.1475-1477.2003>.
 61. Sauer RT, Krovatin W, DeAnda J, Youderian P, Susskind MM. 1983. Primary structure of the *immI* immunity region of bacteriophage P22. *J Mol Biol* 168:699–713. [https://doi.org/10.1016/s0022-2836\(83\)80070-9](https://doi.org/10.1016/s0022-2836(83)80070-9).
 62. Hofer B, Ruge M, Dreiseikelmann B. 1995. The superinfection exclusion gene (*sieA*) of bacteriophage P22: identification and overexpression of the gene and localization of the gene product. *J Bacteriol* 177:3080–3086. <https://doi.org/10.1128/jb.177.11.3080-3086.1995>.
 63. Vaca Pacheco S, Garcia Gonzalez O, Paniagua Contreras GL. 1997. The *lom* gene of bacteriophage lambda is involved in *Escherichia coli* K12 adhesion to human buccal epithelial cells. *FEMS Microbiol Lett* 156:129–132. <https://doi.org/10.1111/j.1574-6968.1997.tb12717.x>.
 64. Barondess JJ, Beckwith J. 1995. *bor* gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein. *J Bacteriol* 177:1247–1253. <https://doi.org/10.1128/jb.177.5.1247-1253.1995>.
 65. Iwashita S, Kanegasaki S. 1973. Smooth specific phage adsorption: endorhamnosidase activity of tail parts of P22. *Biochem Biophys Res Commun* 55:403–409. [https://doi.org/10.1016/0006-291x\(73\)91101-7](https://doi.org/10.1016/0006-291x(73)91101-7).
 66. Iwashita S, Kanegasaki S. 1976. Enzymic and molecular properties of base-plate parts of bacteriophage P22. *Eur J Biochem* 65:87–94. <https://doi.org/10.1111/j.1432-1033.1976.tb10392.x>.
 67. Wright A, Kanegasaki S. 1971. Molecular aspects of lipopolysaccharides. *Physiol Rev* 51:748–784. <https://doi.org/10.1152/physrev.1971.51.4.748>.
 68. Seul A, Muller JJ, Andres D, Stettner E, Heinemann U, Seckler R. 2014. Bacteriophage P22 tailspike: structure of the complete protein and function of the interdomain linker. *Acta Crystallogr D Biol Crystallogr* 70:1336–1345. <https://doi.org/10.1107/S1399004714002685>.
 69. Steinbacher S, Miller S, Baxa U, Budisa N, Weintraub A, Seckler R, Huber R. 1997. Phage P22 tailspike protein: crystal structure of the head-binding domain at 2.3 Å, fully refined structure of the endorhamnosidase at 1.56 Å resolution, and the molecular basis of O-antigen recognition and cleavage. *J Mol Biol* 267:865–880. <https://doi.org/10.1006/jmbi.1997.0922>.
 70. Liu B, Knirel YA, Feng L, Perepelov AV, Senchenkova SN, Reeves PR, Wang L. 2014. Structural diversity in *Salmonella* O antigens and its genetic basis. *FEMS Microbiol Rev* 38:56–89. <https://doi.org/10.1111/1574-6976.12034>.
 71. Adriaenssens EM, Ackermann HW, Anany H, Blasdel B, Connerton IF, Goulding D, Griffiths MW, Hooton SP, Kutter EM, Kropinski AM, Lee JH, Maes M, Pickard D, Ryu S, Sepehrizadeh Z, Shahrabak SS, Toribio AL, Lavigne R. 2012. A suggested new bacteriophage genus: “Viunalikevirus.” *Arch Virol* 157:2035–2046. <https://doi.org/10.1007/s00705-012-1360-5>.
 72. Niu YD, McAllister TA, Nash JH, Kropinski AM, Stanford K. 2014. Four *Escherichia coli* O157:H7 phages: a new bacteriophage genus and taxonomic classification of T1-like phages. *PLoS One* 9:e100426. <https://doi.org/10.1371/journal.pone.0100426>.
 73. Santos SB, Kropinski AM, Ceyssens PJ, Ackermann HW, Villegas A, Lavigne R, Krylov VN, Carvalho CM, Ferreira EC, Azeredo J. 2011. Genomic and proteomic characterization of the broad-host-range between phage PVP-SE1: creation of a new phage genus. *J Virol* 85:11265–11273. <https://doi.org/10.1128/JVI.01769-10>.
 74. Winston F, Botstein D, Miller JH. 1979. Characterization of *amber* and *ochre* suppressors in *Salmonella typhimurium*. *J Bacteriol* 137:433–439.
 75. Adams MH. 1959. Bacteriophages. John Wiley & Sons, Inc, New York, NY.
 76. Earnshaw W, Casjens S, Harrison S. 1976. Assembly of the head of bacteriophage P22. X-ray diffraction from heads, proheads and related structures. *J Mol Biol* 104:387–410. [https://doi.org/10.1016/0022-2836\(76\)90278-3](https://doi.org/10.1016/0022-2836(76)90278-3).
 77. Gill JJ, Berry JD, Russell WK, Lessor L, Escobar-Garcia DA, Hernandez D, Kane A, Keene J, Maddox M, Martin R, Mohan S, Thorn AM, Russell DH, Young R. 2012. The *Caulobacter crescentus* phage ϕ Cbk: genomics of a canonical phage. *BMC Genomics* 13:542. <https://doi.org/10.1186/1471-2164-13-542>.
 78. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
 79. Noguchi H, Taniguchi T, Itoh T. 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res* 15:387–396. <https://doi.org/10.1093/dnares/dsn027>.
 80. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 27:4636–4641. <https://doi.org/10.1093/nar/27.23.4636>.
 81. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.
 82. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
 83. Hildebrand A, Remmert M, Biegert A, Soding J. 2009. Fast and accurate automatic structure prediction with HHpred. *Proteins* 77(Suppl 9):128–132. <https://doi.org/10.1002/prot.22499>.
 84. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580. <https://doi.org/10.1006/jmbi.2000.4315>.
 85. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785–786. <https://doi.org/10.1038/nmeth.1701>.
 86. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* 12:1652–1662. <https://doi.org/10.1110/ps.0303703>.
 87. Kingsford CL, Ayanbule K, Salzberg SL. 2007. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* 8:R22. <https://doi.org/10.1186/gb-2007-8-2-r22>.
 88. Cock PJ, Gruning BA, Paszkiewicz K, Pritchard L. 2013. Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology. *PeerJ* 1:e167. <https://doi.org/10.7717/peerj.167>.
 89. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein

- database search programs. *Nucleic Acids Res* 25:3389–3402. <https://doi.org/10.1093/nar/25.17.3389>.
90. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>.
 91. Douglas SE. 1994. DNA Strider. A Macintosh program for handling protein and nucleic acid sequences. *Methods Mol Biol* 25:181–194. <https://doi.org/10.1385/0-89603-276-0:181>.
 92. Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 23:1026–1028. <https://doi.org/10.1093/bioinformatics/btm039>.
 93. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. 2005. ACT: the Artemis comparison tool. *Bioinformatics* 21:3422–3423. <https://doi.org/10.1093/bioinformatics/bti553>.
 94. Valentine RC, Shapiro BM, Stadtman ER. 1968. Regulation of glutamine synthetase. XII. Electron microscopy of the enzyme from *Escherichia coli*. *Biochemistry* 7:2143–2152. <https://doi.org/10.1021/bi00846a017>.
 95. Eppler K, Wyckoff E, Goates J, Parr R, Casjens S. 1991. Nucleotide sequence of the bacteriophage P22 genes required for DNA packaging. *Virology* 183:519–538. [https://doi.org/10.1016/0042-6822\(91\)90981-g](https://doi.org/10.1016/0042-6822(91)90981-g).
 96. Casjens S, Huang WM. 1993. Linear chromosomal physical and genetic map of *Borrelia burgdorferi*, the Lyme disease agent. *Mol Microbiol* 8:967–980. <https://doi.org/10.1111/j.1365-2958.1993.tb01641.x>.
 97. 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. <https://doi.org/10.1126/science.1127981>.
 98. Casjens SR, Jacobs-Sera D, Hatfull GF, Hendrix RW. 2015. Genome sequence of *Salmonella enterica* phage Det7. *Genome Announc* 3:e00279–15. <https://doi.org/10.1128/genomeA.00279-15>.
 99. Gebhart D, Williams SR, Scholl D. 2017. Bacteriophage SP6 encodes a second tailspike protein that recognizes *Salmonella enterica* serogroups C2 and C3. *Virology* 507:263–266. <https://doi.org/10.1016/j.virol.2017.02.025>.