Global Proteomic Profiling of *Salmonella* Infection by a Giant Phage

Susan T. Weintraub,^a Nurul Humaira Mohd Redzuan,^b Melissa K. Barton,^b Nur Amira Md Amin,^b Maxim I. Desmond,^b Lily E. Adams,^b Bazla Ali,^b Sammy Pardo,^a Dana Molleur,^a Weimin Wu,^c William W. Newcomb,^c Michael V. Osier,^b Lindsay W. Black,^d Alasdair C. Steven,^c Julie A. Thomas^b

^aDepartment of Biochemistry and Structural Biology, The University of Texas Health Science Center, San Antonio, Texas, USA

^bThomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, New York, USA

^cLaboratory of Structural Biology Research, National Institute for Arthritis, Musculoskeletal and Skin Disease, National Institutes of Health, Bethesda, Maryland, USA

^dDepartment of Biochemistry and Molecular Biology, The University of Maryland School of Medicine, Baltimore, Maryland, USA

ABSTRACT The 240-kb *Salmonella* phage SPN3US genome encodes 264 gene products, many of which are functionally uncharacterized. We have previously used mass spectrometry to define the proteomes of wild-type and mutant forms of the SPN3US virion. In this study, we sought to determine whether this technique was suitable for the characterization of the SPN3US proteome during liquid infection. Mass spectrometry of SPN3US-infected cells identified 232 SPN3US and 1,994 *Salmonella* proteins. SPN3US proteins with related functions, such as proteins with roles in DNA replication, transcription, and virion formation, were coordinately expressed in a temporal manner. Mass spectral counts showed the four most abundant SPN3US proteins to be the major capsid protein, two head ejection proteins, and the functionally unassigned protein gp22. This high abundance of gp22 in infected bacteria contrasted with its absence from mature virions, suggesting that it might be the scaffold protein, an essential head morphogenesis protein yet to be identified in giant phages. We identified homologs to SPN3US gp22 in 45 related giant phages, including ϕ KZ, whose counterpart is also abundant in infected bacteria but absent in the virion. We determined the ϕ KZ counterpart to be cleaved *in vitro* by its prohead protease, an event that has been observed to promote head maturation of some other phages. Our findings are consistent with a scaffold protein assignment for SPN3US gp22, although direct evidence is required for its confirmation. These studies demonstrate the power of mass spectral analyses for facilitating the acquisition of new knowledge into the molecular events of viral infection.

IMPORTANCE “Giant” phages with genomes >200 kb are being isolated in increasing numbers from a range of environments. With hosts such as *Salmonella enterica*, *Pseudomonas aeruginosa*, and *Erwinia amylovora*, these phages are of interest for phage therapy of multidrug-resistant pathogens. However, our understanding of how these complex phages interact with their hosts is impeded by the proportion (~80%) of their gene products that are functionally uncharacterized. To develop the repertoire of techniques for analysis of phages, we analyzed a liquid infection of *Salmonella* phage SPN3US (240-kb genome) using third-generation mass spectrometry. We observed the temporal production of phage proteins whose genes collectively represent 96% of the SPN3US genome. These findings demonstrate the sensitivity of mass spectrometry for global proteomic profiling of virus-infected cells, and the identification of a candidate for a major head morphogenesis protein will facilitate further studies into giant phage head assembly.

KEYWORDS *Salmonella*, bacteriophage assembly, giant phage, mass spectrometry

Citation Weintraub ST, Mohd Redzuan NH, Barton MK, Md Amin NA, Desmond MI, Adams LE, Ali B, Pardo S, Molleur D, Wu W, Newcomb WW, Osier MV, Black LW, Steven AC, Thomas JA. 2019. Global proteomic profiling of *Salmonella* infection by a giant phage. *J Virol* 93:e01833-18. <https://doi.org/10.1128/JVI.01833-18>.

Editor Rozanne M. Sandri-Goldin, University of California—Irvine

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

Address correspondence to Julie A. Thomas, jatsbi@rit.edu.

This paper is dedicated to the memory of our revered colleague and mentor, David Lawlor.

Received 14 November 2018

Accepted 19 November 2018

Accepted manuscript posted online 12 December 2018

Published 19 February 2019

Lytic phages are highly effective bacterial parasites; their infection exploits cellular resources for progeny production and ultimately results in cell death. The molecular events that occur during this process are largely unknown for many phages because limited numbers of their gene products can be assigned functions using bioinformatics alone (typically ~20%). This problem is exacerbated in “giant” or “jumbo” phages with long double-stranded DNA (dsDNA) genomes (>200 kb) (1, 2), such as that of *Pseudomonas aeruginosa* phage ϕ KZ (280 kb), whose genome was the first giant phage genome to be published (3). Recent studies have demonstrated that phages related to ϕ KZ infect a range of clinically and agriculturally important gammaproteobacteria, including *Salmonella enterica* (e.g., SPN3US), *Cronobacter sakazakii* (e.g., CR5), and *Erwinia amylovora* (e.g., PhiEaH2) (3–10). These giant phages are of interest for biocontrol purposes (e.g., phage therapy) and the fact that they have many unusual characteristics compared to other phage types.

Notably the long genomes (211 to 316 kb) of giant phages related to ϕ KZ and SPN3US encode highly diverged genes making functional annotation of their gene products difficult, even for proteins that are typically straightforward to identify in other phage types (4) (see, for example, Fig. 1). The virions of these giant phages have large icosahedral heads, or capsids, with high T numbers (T=27 in the case of ϕ KZ), and long, contractile tails. Giant phage virions are also considered to be complex relative to other phage types due to being composed of a high number of different proteins (>60 different proteins identified by mass spectrometry) (11–15). A highly unusual feature of giant phage heads is that they have a large cargo of internal proteins, called ejection or E proteins, within them. In ϕ KZ, the E proteins form a large (15 to 20 MDa) proteinaceous structure—the inner body (IB)—wrapped within the packaged DNA (16). We recently determined that the SPN3US head contains 50 different proteins, the most abundant being the major capsid protein (MCP), which forms the outer shell, and the E proteins gp53 and gp54 (10, 15). Giant phage E proteins are assumed to enter the host cell with the phage genome because the ϕ KZ IB disappears upon ejection of the genome (17). In addition, all related giant phages have a multisubunit virion RNA polymerase (vRNAP) that is assigned as an E protein and has been shown to be an essential head component of SPN3US (10, 15). The vRNAP was demonstrated to transcribe early genes in ϕ KZ, supporting the inference that it must traverse the narrow tail tube to enter the host cell (18).

In addition to the vRNAP, all giant SPN3US-like phages encode a second RNAP. This nonvirion RNAP (nvRNAP), which is produced during infection, transcribes the ϕ KZ late genes (19). The giant phage RNAPs are highly unusual as they are composed of beta (β) and beta' (β') subunits—very different from the single subunit RNAPs of phages T7 and N4 (20–22). The giant phage RNAPs also illustrate large differences to prokaryotic RNAPs, being highly diverged to them and having unusual split sites (23). Moreover, the giant phage RNAP genes do not even preserve the highly conserved prokaryotic β gene- β' gene synteny (23). As a result, a small C-terminal domain of the giant phage vRNAP β' was previously not detected and was only recently identified and confirmed to be essential (10, 15). The main consequence of the two RNAPs is that giant phage gene expression programs are likely to have a high degree of independence from their host transcriptional machinery (18, 24–26).

Although previous studies on giant phages have highlighted their unusual properties there is still much to be understood regarding their assembly and replication within the host cell. In this study, our goal was to capitalize on the power of high-performance mass spectrometry to provide deep insight into phage infection by analyses of *Salmonella* Typhimurium during infection by SPN3US.

RESULTS

Identification of SPN3US proteins during infection. A total of 232 different SPN3US proteins, ranging in molecular mass from 7 to 259 kDa, were identified with a protein identification probability of 100% during the course of infection of *Salmonella* (Fig. 2, see Table S1 in the supplemental material). These identifications confirm the

FIG 1 Major genes required to form the virion of a myovirus (phage with a contractile tail). (A) Scheme of the ordering of genes required to make the head (or capsid) of most tailed phages. Red question marks indicate genes that have yet to be identified in SPN3US. (B) Linear representation of the SPN3US genome showing the locations of known head and tail morphogenesis genes from panel A. (C) Representation of the major genes required to make a contractile tail. Capsid and tail images are by W. Wu (work in progress).

expression of 88% of SPN3US's predicted genes, which represents ~96% of the 240-kb genome. Thirty-two hypothetical proteins were not identified. If they are produced, they may be present at levels below the threshold for detection or may be produced shortly before cell lysis. The detection with confidence of this large number of SPN3US proteins among several thousand different host proteins (see below) highlights the low detection limits of the Lumos mass spectrometer. Of note, only 31 SPN3US proteins were identified in a sample from an infection at 20 min (Time-20) that was processed identically but analyzed on a Thermo Fisher LTQ Orbitrap Velos Pro mass spectrometer (Table S2).

Identification of *Salmonella* proteins during infection. During the course of the experiment, 1994 *Salmonella* proteins were identified at 100% protein identification (Table 1). These proteins ranged from 5 to 595 kDa, for the 30S ribosomal subunit S22

FIG 2 Production of SPN3US proteins during liquid infection of *Salmonella enterica* serovar Typhimurium LT2. Black inner ring peaks represent an estimate of relative abundance (SC/M, spectral count divided by molecular mass) of each protein at Time-60. The relative numbers of mass spectral counts at each time point adjusted by the highest number of spectra detected for each protein is indicated. The exterior ring depicts the SPN3US genome map with virion protein genes shaded blue and nonvirion protein genes shaded black. Highly abundant proteins, including major virion proteins, the major capsid protein (MCP), the tail sheath, and tube proteins are labeled.

and the large Ig-like domain repeat protein (OFI27953), respectively. *Salmonella* proteins were inferred to be present in low to extremely high abundance based on the number of total mass spectra detected for each protein. Notably, there were similar levels of many higher abundance *Salmonella* proteins in infected and control samples (e.g., Fig. 3A and B). This indicates that the levels of these high-abundance proteins are not sensitive indicators of phage-induced physiological alterations in the cell and that further measurements, such as transcriptomic and structural studies, would be of value.

The levels of the higher abundance *Salmonella* proteins were useful for providing insight into the general state of the cell. For example, the increases in mass spectra detected for many major outer membrane proteins were consistent with a growing cell wall. Similarly, the increases in mass spectra detected for ribosomal proteins and chaperonins was consistent with high levels of protein expression that would be expected in an exponentially growing culture (e.g., Fig. 3C). The high levels of these proteins would provide an ideal environment for SPN3US replication because SPN3US is expected to use host ribosomes and other translation machinery. In contrast to these proteins, the major subunits of the enzymes required for cellular DNA replication and transcription decreased in abundance during the course of the experiment (Fig. 3D).

TABLE 1 Summary of *S. Typhimurium* and SPN3US proteins identified by mass spectrometry during infection

Time (min)	Total mass spectra				Total no. of proteins			
	<i>Salmonella</i>		SPN3US		<i>Salmonella</i>		SPN3US	
	Control	Infection	Control	Infection	Control	Infection	Control	Infection
0	187,055	189,034	0	1,168	1,912	1,908	0	36
10	187,085	201,107	0	4,280	1,877	1,897	0	105
20	179,247	197,688	0	8,274	1,844	1,891	0	164
30	179,116	192,309	0	11,876	1,877	1,839	0	185
40	207,591	205,788	0	19,500	1,822	1,807	0	197
50	222,087	212,764	0	25,905	1,844	1,749	0	223
60	248,651	209,943	0	31,579	1,722	1,718	0	224
Total					1,955	1,939	0	232

FIG 3 Identification of *Salmonella* proteins during infection by SPN3US, and in an uninfected control. (A) Major outer membrane proteins. (B and C) Major ribosomal proteins and proteins with roles relating to protein expression and folding. (D) Major RNA polymerase and DNA polymerase subunits.

However, reductions in these enzymes may not directly impact SPN3US DNA replication and transcription as SPN3US encodes its own DNAP and RNAPs.

Temporal production of SPN3US proteins during infection. The number of different SPN3US proteins identified increased throughout infection, from 22 proteins at Time-0 to 232 at Time-60 (Table 1). We attribute the proteins identified at Time-0 to be from the input phage sample and not the product of the current infection. The detection, or appearance, of many SPN3US proteins was coordinated with each other in a functionally relevant way (Fig. 2). Most of the proteins that exhibited increasing

TABLE 2 SPN3US proteins identified during infection of *Salmonella* with functions relating to transcription and DNA replication, repair, and recombination

Function and SPN3US protein	Subunit or description	Essential ^a	Size (kDa)	Total mass spectra at various time points						
				0 min	10 min	20 min	30 min	40 min	50 min	60 min
nvRNAP subunits										
gp77	βN	Yes†	53.3	0	12	38	60	61	30	26
gp34	βC	Yes†	81.8	0	73	202	211	207	125	132
gp23	β'N	Presumed	59.3	0	41	91	151	180	144	146
gp35	β'C	Yes†	79.7	0	35	157	163	183	149	149
gp30	Putative subunit	Expected ^b	61.9	0	28	46	71	102	67	62
vRNAP subunits										
gp241	βN	Yes*	159.1	15	23	16	24	67	52	69
gp218	βC	Yes*	25.2	0	0	0	0	2	4	12
gp240	β'	Yes*	59.6	0	0	0	2	22	23	42
gp42	β'M	Yes	49.3	0	0	0	1	4	10	18
gp244	β'C	Yes*	27.0	0	0	0	0	2	4	0
DNA replication, repair, and recombination										
gp229	Thymidylate synthetase		39.0	0	4	28	36	41	88	120
gp100	Thymidylate kinase		23.2	0	0	6	13	22	58	58
gp44	DNAP, N terminus	Presumed	68.7	0	0	6	15	10	10	10
gp18	DNAP, C terminus	Yes†	87.3	0	0	15	25	27	20	20
gp73	Helicase, replicative		58.6	0	0	6	12	22	22	27
gp207	RNase H	Yes†	57.6	0	0	16	31	56	31	27
gp186	SbcC	Yes*	95.7	0	0	17	21	9	12	15
gp27	SbcD	Presumed	43.5	0	0	1	5	9	21	17
gp216	UvsX		55.2	0	0	20	55	98	107	115
gp105	Dam methylase		29.7	0	4	18	20	34	46	60

^a*, Determined to be essential by the isolation and sequencing of an SPN3US amber mutant; *, a previously sequenced mutant (10, 15); †, a mutant isolated and sequenced in this study (see Table S3).

^bThat is, expected to be essential. Amber mutation in the gp30 gene was identified in a SPN3US double mutant (see Table S3).

abundances after Time-30 are known virion components, such as gp45 to gp54, gp141 to gp160, and gp255 to gp262. This observation is consistent with previous studies of numerous tailed phages in which the virion genes are late genes.

Seventy of the proteins identified at Time-10 are not part of the virion (Fig. 2; Table S1), indicating that the source of these proteins was not from input phages. These early proteins are excellent candidates for roles relating to host takeover and initiation of phage reproduction. The latter role is supported by the identification of all the nvRNAP subunits at this time (Table 2).

The subunits of the SPN3US DNAP were identified at Time-20 (Table 2). We recently determined this enzyme to be essential based on the genome sequencing of an amber mutant phage (Table S4). The DNAP is conserved in all other ϕ KZ-related phages and previously shown to have homology to the T4 DNAP (gp43) (27). SPN3US also has homologs to other T4 DNA replication, recombination, and repair proteins (e.g., DnaB, RNase H, SbcC, SbcD, and UvsX). These SPN3US proteins were first detected at approximately the same time in infection as the DNAP subunits (Table 2). A predicted Dam methylase, gp105, was also detected early in infection (Table 2). We believe gp105 is likely responsible for the *dam* methylation of SPN3US DNA that was indicated by restriction enzyme analysis (i.e., SPN3US DNA is sensitive to cleavage by DpnI, but not MboI [data not shown]).

Identification of abundant SPN3US proteins in the *Salmonella* cell during infection. The number of spectra identified for SPN3US proteins produced during infection ranged from 1 spectrum (e.g., nonvirion protein gp15 at Time-10) to 6,216 spectra for the major capsid protein at Time-60 (Table S1). Spectral counting is a widely used semiquantitative approach for estimation of relative protein abundance (28) and, in this study, we divided the total number of spectral counts (SC) assigned to each SPN3US protein by its predicted molecular weight (M) (Table S1) to adjust for differ-

TABLE 3 Abundant SPN3US proteins identified during infection

Protein	Putative function	Mass (kDa)	Spectral count/mass (SC/M) as various time points ^a						
			0 min	10 min	20 min	30 min	40 min	50 min	60 min
gp22	Scaffold candidate	67.1	0.09	12.32	31.32	45.00	54.51	44.37	46.69
gp29		24.5	0.00	0.52	4.26	5.96	11.52	12.93	13.44
gp31		61.5	0.00	1.92	5.56	5.32	4.61	3.16	3.02
gp32		49.6	0.00	1.50	4.56	5.24	6.66	7.36	7.18
gp39		73.4	0.00	0.00	0.76	3.65	6.16	6.01	6.47
gp53	E protein	45.2	6.15	8.87	9.12	13.21	41.08	78.21	93.21
gp54	E protein	45.1	1.86	3.68	3.90	5.23	14.15	27.14	43.59
gp58	Dihydrofolate reductase	19.9	0.00	4.00	6.00	5.29	7.24	5.71	4.90
gp65		29.3	0.00	0.00	0.97	6.30	6.13	11.40	13.27
gp75	Major capsid	83.9	4.87	7.56	8.70	14.84	39.73	54.61	74.09
gp78		60.3	0.00	2.11	5.49	9.15	10.07	7.23	7.46
gp103		24.1	0.04	3.62	10.54	7.96	9.54	16.46	15.35
gp127		20.0	0.00	0.95	3.20	2.55	4.55	7.70	5.30
gp141	E protein	32.6	1.21	1.06	1.15	1.76	3.88	7.36	10.94
gp160	E protein	18.5	3.90	4.80	3.55	3.25	5.80	14.20	21.10
gp172	Low-abundance virion protein	7.0	0.00	0.00	0.00	0.14	11.14	22.14	24.14
gp201		42.7	0.00	0.84	3.80	5.58	8.07	10.98	10.69
gp252		16.9	0.00	6.81	15.90	13.71	15.62	15.95	14.19
gp255	Tail tube	32.7	0.36	0.76	0.39	1.06	5.21	12.70	19.36
gp256	Tail sheath	75.7	0.69	1.22	1.42	2.14	5.22	8.55	14.58

^aThe total mass spectral counts are provided in Table S1. Proteins were identified with a protein identification probability of 100%, and the quality filter was set to a minimum number of peptides.

ences in protein lengths (SC/M). After we normalized the results in this way, we found that the virion protein gp53 had the highest SC/M value of any SPN3US protein at one sampling time, peaking at Time-60 with an SC/M of 93 (Table 3; see Fig. 6 below). gp53 also had the highest SC/M value of the SPN3US proteins at Time-50. The second most abundant protein to gp53 at these time points was the major capsid protein, gp75, with SC/M values of 55, at Time-50, and 74, at Time-60. The high abundance of gp53 and gp75 during infection is not surprising based on the high abundance of these proteins in the virion (15) and the SPN3US burst size (~100 [data not published]). That gp22 was the most abundant SPN3US protein from Time-10 through Time-40 was unexpected since this was previously a functionally unassigned protein not identified as part of the virion. The extremely high abundance of gp22 in this experiment highlighted it to be a protein of interest.

Three other proteins that are components of the mature virion, gp54, gp172, and gp160, all had SC/M values of >20, indicating that each must also be present in very high numbers in the cell (Table 2). The high relative abundances of the virion proteins gp54 and gp160 was also consistent with our previous analyses of purified virions that have shown that each of these proteins is present in high copy numbers in each mature virion (15). That gp172 is present in high abundance in the host cell was unexpected as our previous studies have shown this small 69-residue (7 kDa) protein to be present in a low copy number in each virion and its role in infection requires further investigation.

Nine proteins (gp29, gp65, gp78, gp103, gp141, gp201, gp252, gp255, and gp256) had maximum SC/M values that ranged from 10 to 20, indicating they also are highly abundant within the cell (Table 2). This expectation is supported by gp255 and gp256 being the major tail tube and sheath proteins, respectively (~270 copies of each per virion) and that gp141 is a highly abundant head protein (>300 copies per virion) (15). These abundant, functionally unassigned SPN3US proteins are excellent candidates for future studies since they likely have novel roles in phage infection.

SPN3US abundant protein gp22 is a candidate for the prohead scaffold protein. A scaffold protein had not been identified in SPN3US, or any of its close relatives. Since these phages share an ancestral head assembly pathway with T4 phage (4, 11, 15, 27), which does have a scaffold protein, it seemed likely that these giant phages would also have a scaffold protein. The T4 scaffold protein, also called gp22, is an essential

FIG 4 Mass spectral identification of major SPN3US major head proteins and the scaffold protein candidate, gp22, during infection.

component of the prohead core which, as its name suggests, effectively forms a structure around which the major capsid shell assembles (29). However, the prohead core is a temporary structure, with gp22, and other core proteins, leaving the capsid prior to head expansion and DNA packaging (30). The exit of T4 gp22 from the procapsid is facilitated by its cleavage by the prohead protease. Consequently, the scaffold protein in tailed phages is typically either not present in the mature virion or, if detected, in very low abundance, as occurs for T4 gp22 (29).

The similarity in abundance of the early SPN3US protein gp22 and the major head proteins gp75, gp53, and gp54 (Fig. 4) suggested SPN3US gp22 could be a candidate for the prohead scaffold protein. In order to gain further insight into the role of SPN3US gp22, we conducted experiments using three different approaches, which are described in detail below.

The SPN3US gp22 gene and homologs in related phages hold a conserved genome position. In tailed phage taxa in which the identity of the scaffold protein is known, homologs to this protein can typically be identified by sequence-based searches in all related phages. The genes encoding those proteins also typically hold a conserved gene position. In the T4 superfamily there are clear counterparts to T4 gp22 in all phages, although it should be noted that these homologs are frequently more diverged at the sequence level than other essential head proteins (11).

Based on these precedents, we looked for counterparts of SPN3US gp22 in related phages. A PSI-BLAST search from SPN3US gp22 against the NCBI nr and env_nr databases found homologs in the first iteration, all with E values of $\leq 3e-55$, in 13 phages considered to be relatively closely related to SPN3US (e.g., as shown in Fig. S1). The second iteration of this search identified a diverged match (15% identity, $4e-5$) to *Vibrio* phage pVa-21 gp208, whose improved E value ($9e-127$) in the third (and final) iteration suggested it to be a true homolog. We noted the gene locale of these homologs to SPN3US gp22 was conserved immediately upstream of the N-terminal subunit of the nvRNAP β' (nv β' N) (Fig. 5). This was of interest since there are fragments of gene synteny in giant phage genomes despite the overall fragmentation of the major morphogenesis gene region found in many other tailed phages (Fig. 1). This encouraged us to investigate whether other giant phages related to SPN3US might have diverged homologs to gp22 whose gene was located upstream of their nv β' N gene.

ϕ KZ ORF054 is upstream of its nv β' N gene ORF055, and its gene product is comparable in length to SPN3US gp22 (Fig. 5). A PSI-BLAST search from ϕ KZ gp54 identified a total of 39 phage homologs, including pV-21 gp208 ($4e-16$) and SPN3US gp22 ($e-14$) in the second and third iterations, respectively. These matches had

FIG 5 The gene locale of SPN3US scaffold candidate, gp22, is conserved in related phages.

improved scores in subsequent iterations (e.g., $7\text{e-}104$ and $1\text{e-}106$ for SPN3US gp22 in the fourth and fifth iterations, respectively), indicating they were likely true homologs. Independent support for the match between SPN3US gp22 and ϕ KZ gp54 was obtained using HHpred (Fig. S2).

We determined that 42 giant phages related to SPN3US have a gene encoding a counterpart to SPN3US gp22 that is immediately upstream of their $\text{nv}\beta'$ N gene (Table S4). This is notable, considering the high level of genome rearrangements typically observed between the genomes of these giant phages (see, for example, references 23 and 31). Four additional related phages—*Pseudomonas fluorescens* phage Phabio, *Pseudomonas chlororaphis* phage 201 ϕ 2-1, and the *Ralstonia solanacearum* phages RSF1 and RSL2—also have a gene encoding a homolog to SPN3US gp22 upstream of their $\text{nv}\beta'$ N genes. However, in these four phages a small gene is inserted between the gp22 homolog and $\text{nv}\beta'$ N genes (Table S4). BLASTP was unable to identify homologs to the gene products of these small genes in Phabio, RSF1, or RSL2. BLASTP did identify two diverged matches to the gene product, gp106, of 201 ϕ 2-1: the major capsid proteins of 201 ϕ 2-1 and *Pseudomonas fluorescens* phage Noxifer (31 and 25% identities, respectively). This is intriguing, because in many phages, such as T4, the scaffold gene is located immediately upstream of the MCP gene. We speculate that in the genome of an ancestor of 201 ϕ 2-1, the precursor forms of the gp106 and gp200 genes were once neighbors and that their current gene order is the consequence of an event that split the scaffold-major capsid gene synteny.

No homologs to SPN3US gp22 or ϕ KZ gp54 were detected by PSI-BLAST in *Yersinia enterocolitica* phage ϕ R1-37 or *Bacillus subtilis* phages AR9 and PBS1. This was despite these phages having homologs to the RNAPs, large terminases, and prohead proteases of SPN3US and ϕ KZ. Our previous phylogenetic analyses of giant phage terminase and protease enzymes highlighted the divergence of these proteins in ϕ R1-37, AR9, and PBS1 relative to those of other giant phages, since they clustered on separate phylogenetic branches (15). In fact, the protease of ϕ R1-37 was shown by Skurnik et al. to be more similar to the protease of T4 phage by homology searches and cleavage specificity (32). A PSI-BLAST search from the ϕ R1-37 protease ([CCE26103.1](#)) identified homologs to the gp21 proteases of T4-like phages KVP40 (4e-9, 30% identity) and RB43 (3e-9, 29% identity). In addition, mass spectral analyses of the ϕ R1-37 virion identified prohead protease cleavage sites in virion proteins, including the protease maturation site, that conform perfectly to the consensus T4 protease recognition sequence I/L-X-E (32).

Despite the similarities of the ϕ R1-37 protease with T4 gp21, ϕ R1-37 does not have the iconic head gene cluster (21-22-23; protease-scaffold-major capsid) of T4 and its relatives, including KVP40 or RB43, that would assist in the identification of its scaffold

gene. Examination of the ϕ R1-37 gene (g101) immediately upstream of its *nvRNAP* β' N subunit gene (g102) suggests that g101 may be a possible scaffold gene candidate. ϕ R1-37 g101 encodes a 36.6-kDa protein ([CCE26125.1](#)) that is rich in glutamate residues (22.2%) and has a low pI (4.65)—both features of the 29.9-kDa T4 gp22 whose glutamate content is 17.5% and pI is 4.59.

It is intriguing to speculate that ϕ R1-37 could potentially have a scaffold protein more like that of T4 than SPN3US and ϕ KZ. This is because ϕ R1-37 has a T=27 capsid (32), like those of SPN3US and ϕ KZ. However, the molecular weight of the ϕ R1-37 MCP ([CCE26159](#)) is ~31% lower than those of SPN3US and ϕ KZ. In addition, it was previously suggested that eight ϕ R1-37 genes, including those for the MCP and putative scaffold protein, were possibly relatively recent acquisitions since they have a significantly higher G+C content than the average for the ϕ R1-37 genome (32). If further studies were to confirm ϕ R1-37 [CCE26125.1](#) to be its scaffold protein, this would imply there is an intriguing evolutionary basis for the variations in head assembly and composition of T=27 capsids.

ϕ KZ gp54 is abundant early in infection. As noted above, SPN3US gp22 was assigned as a scaffold candidate based on its high abundance in the infected cell, contrasting with its absence in the mature virion. Based on the expectation that homologs to gp22 should have similar functions, and therefore similar characteristics, to gp22, we sought to test whether a gp22 homolog in another giant phage was also abundant in infected cells but not present in the mature virion. Since the ϕ KZ head and virion are well characterized (11, 33), it seemed a suitable candidate for this purpose. The ϕ KZ counterpart to SPN3US gp22, gp54, was not identified in the head or virion in previous studies. We then infected *P. aeruginosa* with ϕ KZ using the same parameters that had facilitated the mass spectral identification of SPN3US gp22 as the most abundant phage protein in the cell at Time-20 (Table S2). Analyses of ϕ KZ-infected *P. aeruginosa* cells using the LTQ Orbitrap Velos Pro mass spectrometer identified 37 ϕ KZ and 590 *P. aeruginosa* proteins at Time-20 (Table 4). The highest number of mass spectra identified for any ϕ KZ protein was for gp54 (Table 4). The following observations support that the sample was analyzed when the ϕ KZ infection was still in an early stage similar to that of SPN3US: (i) none of the major virion proteins (e.g., major capsid or tail sheath proteins) were identified, and (ii) several of the ϕ KZ proteins identified had counterparts in SPN3US that were also detected early in infection (Table 4). These homologs included the thymidylate synthetase and thymidylate kinase responsible for the production of deoxynucleoside triphosphates for DNA replication, which is expected to commence soon after infection, based on the precedent of T4 phage (34).

ϕ KZ gp54 is cleaved by the prohead protease. The T4 scaffold gp22 is cleaved by the prohead protease gp21, which facilitates its exit from the prohead. Based on this precedent, we hypothesized that the scaffold protein of a giant phage might also be cleaved. Since we had previously demonstrated that the ϕ KZ prohead protease gp175 is active *in vitro* and its cleavage specificity is known (11, 35), we sought to determine whether gp54 was a substrate for gp175. The gp54 gene was cloned with the codons for an N-terminal His₆ tag. Since this recombinant form of gp54 was highly insoluble and could not be satisfactorily purified, we tested whether gp54 could be cleaved by gp175 in a crude *Escherichia coli* lysate. In this approach, cultures of induced enzyme and candidate substrate are incubated. The mixture was then separated into soluble and insoluble fractions, and these fractions were analyzed by SDS-PAGE. We had previously shown gp175 to be active under these conditions (11). The substrate protein used in the previous studies, ϕ KZ gp93, was also included in this study as a positive control for enzyme activity (data not shown).

SDS-PAGE separation of proteins in the gp54-plus/minus-gp175 samples revealed the presence of an additional band with an apparent molecular weight slightly below that of the full-length gp54 in the gp54-plus-gp175 sample (slice 3, Fig. 6). Mass spectral analyses showed gp54 to be the most abundant protein in this band. Detection of 158 mass spectra for the semitryptic peptide 558-FITAMDAAAAAGVTVTME-576 in

TABLE 4 Identification of ϕ KZ proteins in *Pseudomonas aeruginosa* at Time-20^a

Protein	Mass (kDa)	Total mass spectra	Total no. of unique peptides	SPN3US homolog	Alternate ID
gp54	69.3	252	19	gp22	Scaffold candidate
gp69	62.3	50	11	gp31	
gp124	80.2	37	8		
gp188	40.5	23	5	gp100	Thymidylate kinase
gp70	47.8	30	9	gp32	
gp186	14.3	19	5		
gp185	19.7	11	3		
gp68	59.3	19	5	gp30	nvRNAP subunit
gp120	82.9	15	6	gp75	Major capsid protein
gp235	49.1	17	4	gp229	Thymidylate synthase
gp238	19.9	11	2		
gp37	30.9	10	4		
gp38	19.7	14	6		
gp208	41.2	12	3		
gp16	18.4	8	2		
gp18	12.5	12	2		
gp189	9.6	15	2		
gp14	43.3	9	4		
gp282	19.8	6	3		
gp22	22.0	3	2		
gp246	24.1	7	2		Putative phosphohydrolase
gp36	31.3	6	2		
gp17	14.6	5	3		
gp276	19.9	3	2		
gp281	13.9	4	2		
gp236	15.2	4	2		
gp6	20.0	6	2		
gp185_1	8.9	3	2		
gp187	12.3	5	2		
gp232	33.8	3	3		Putative peptidase
gp244	13.4	4	2		
gp58	16.2	4	2		
gp44	14.5	4	2		
gp40	18.2	2	2		
gp13	15.7	3	2		
gp12	14.8	2	2		
gp109_1	9.8	2	2		

^aProteins were identified with a protein identification probability of 100% and a minimum of two peptides.

the gp54-plus-gp175 slices (predominantly slice 3) provided convincing evidence for proteolysis of gp54 in the presence of gp175. This semitryptic peptide was not found in the gp54-minus-gp175 slices. Notably, the observed gp54 semitryptic peptide was consistent with cleavage at a known motif for the prohead protease gp175: cleavage C

FIG 6 Assay of ϕ KZ gp54 with the prohead protease gp175. (A) SDS-PAGE gel of assay of recombinant ϕ KZ proteins gp54 with gp175 in a lysate of *E. coli* proteins. A control of gp54 only was included. The amounts of each sample loaded are indicated. Gel slices processed for MS analyses are indicated by a red square. (B) Semitryptic and tryptic mass spectra for gp54 identified by MS in the gel slices indicated in panel A.

terminal to a glutamic acid residue. Our previous studies of gp175 showed it to cleave at the sequence motif S/A/G-X-E (where X is any amino acid), with one exception where cleavage was C-terminal to the sequence F-V-E (11). The gp54 cleavage by gp175 is consistent with processing C terminal to a TXE motif that we identified in the vRNAP N-terminal β subunit of phage 201 ϕ 2-1 by its protease gp269 (36). Evidence for the presence of at least one additional cleavage site in gp54 was obtained by exposing semipurified gp54 to gp175. In this assay, two slightly lower-molecular-weight fragments were observed by SDS-PAGE (data not shown).

DISCUSSION

We report here the identification of 232 SPN3US proteins among a complex mixture of 1,994 *Salmonella* proteins over the course of infection, establishing that high performance mass spectrometry instrumentation such as the Orbitrap Fusion Lumos has the ability to provide deep insight into the infection proteome. In addition to the detection of changes in SPN3US proteins during infection, we also observed trends in the abundance of *Salmonella* proteins that are suggestive of the overall state of the cells. For example, the relative levels of the main host enzymes for DNA replication and transcription were reduced over time. The reductions in these host proteins may not hinder proliferation of a giant parasite like SPN3US, since it encodes its own enzymatic counterparts to the host DNA and RNA polymerases and is expected to have DNA replication and transcriptional programs that are independent of the host. Conversely, the high levels of proteins needed for translation, such as ribosomal proteins and chaperonins, observed in the *Salmonella* cell would logically be an asset for SPN3US replication. This is because SPN3US, like other phages, is expected to rely heavily upon the host translation machinery.

Of the identified SPN3US proteins with known functions, many that are functionally related were expressed in a coordinated temporal manner; for instance, all of the subunits of the essential five-subunit nvRNAP were found to be produced by Time-10. Similarly, proteins involved in DNA replication (e.g., DNAP, helicase, and RNase H) were found to be produced shortly after Time-10. Detection of functionally related proteins at similar time points during infection suggests that the time course data will be valuable for the characterization of the many SPN3US proteins with currently unassigned functions, as illustrated by the identification of a plausible candidate for the SPN3US prohead scaffold.

In tailed phages, the scaffold protein is present in a large number of copies in the prohead, forming a dense protein structure around which the capsid shell assembles. For example, in each T4 prohead there are about ~600 copies of the scaffold protein gp22 in the core, around which ~1000 copies of the major capsid proteins gp23 and gp24 (at the vertices) assemble (53). Based on the number of gp22 per prohead and the fact that the T4 burst size is ca. 100 to 200 progeny per cell, it is not surprising that gp22 is abundant in T4-infected *E. coli* (29, 37). Despite the fact that giant phages related to SPN3US share an ancestral head maturation pathway with T4 (4, 11, 15, 27, 36), a scaffold protein had not previously been identified in giant phages. The inability to diagnose a single scaffold gene in giant phages could be due to several reasons. As previously noted, even genes conserved in all tailed phages (e.g., major capsid and portal genes) are more highly diverged in giant phages. In addition, giant phages lack a single syntenous major morphogenesis gene region (e.g., Fig. 1). There is also the question of whether giant phages even have a single scaffold protein, since their head composition has so many differences from other studied tailed phages. For instance, it is possible the unusually long propeptides of the head E proteins replace, or supplement, the standard single scaffold protein.

Based on the fact that the SPN3US capsid is larger than that of T4, we envisioned that if the SPN3US scaffold protein exists, it should be present in high enough copy numbers per prohead to be easily identifiable within the host cell. Consistent with this expectation, SPN3US gp22 was the only functionally unassigned phage protein

produced in amounts comparable to the major capsid protein. Since the scaffold genes of other tailed phages are essential, we hypothesized that if SPN3US gp22 was the scaffold protein, giant phages related to SPN3US would also have a counterpart to gp22. Our bioinformatic searches identified counterparts of similar length to SPN3US gp22 in 46 of 49 giant phages analyzed. These homologs were in phages closely related to SPN3US, such as PhiEaH2 and CR5, and those more distantly related, such as ϕ KZ. The homologs to SPN3US gp22 gene hold a conserved position in the genomes of most relatives. No homologs to SPN3US gp22 were identified in three giant phages whose protease and terminase enzymes are extremely diverged to those of SPN3US. In the genome of one of these phages, *Yersinia* phage ϕ R1-37, the position of the SPN3US gp22 gene was held by a protein with some similarities to the T4 scaffold. This is intriguing because, based on homology and cleavage specificity, the ϕ R1-37 prohead protease is more similar to the T4 protease than those of other giant phages (32).

Since the T4 scaffold is cleaved by its prohead protease, we sought to determine whether ϕ KZ gp54, a diverged homolog to SPN3US gp22, was cleaved by its protease, gp175 (11). We selected ϕ KZ gp175 since in our previous characterization studies we found gp175 to be amenable to purification and active *in vitro* (11, 35). In these studies we confirmed ϕ KZ gp175 to be a serine protease and determined its cleavage specificity (11, 35). ϕ KZ gp175 cleaves propeptides from 19 proteins, all of which are components of the mature ϕ KZ head. In our examination of ϕ KZ, only virion proteins that are components of the head were cleaved by prohead protease processing. In this study, mass spectrometry of an assay of ϕ KZ gp54 with gp175 identified a C-terminal processing site in gp54, supporting its placement within the head.

In the present context, it is of note that herpesviruses—which are animal viruses—have a scaffolding protein with functions similar to those of the scaffold proteins of tailed bacteriophages. The production of herpes simplex virus procapsids involves the coassembly of a surface shell (the capsid precursor) and an inner shell (the scaffold). The inner shell is assembled from two forms of this protein: ~90% are simply the scaffold, and ~10% have the maturational protease fused to the N terminus, giving a 635-residue protease-scaffold precursor. The protease, which shares a common ancestor with many tailed phages (38–40), has autocatalytic activity (as does the ϕ KZ protease) and cleaves this fusion protein twice—once between residues 247 and 248 (the “R” site) to release the active protease and again between residues 610 and 611 (the “M” or maturation site) (41). (The residue numbers given here are for cytomegalovirus [41].) The scaffold-only form of the protein also undergoes M-site cleavage. This processing disrupts the interactions between the C termini of the scaffold shell and the inner surface of the capsid shell, thereby promoting maturation transformation of the capsid shell and subsequent DNA packaging (42, 43). We speculate that the C-terminal processing observed in this study for ϕ KZ gp54, and inferred for SPN3US gp22, is comparable to the M site cleavage of herpesviruses. However, before such a comparison can be verified, as noted, further evidence is required to confirm the identity of the giant phage scaffold protein.

Our observations of the abundance of SPN3US gp22 in the cell, the presence of homologs in related giant phages, and the processing of ϕ KZ gp54 by the prohead protease are all expected characteristics of a scaffold protein. However, further, direct evidence is required before SPN3US gp22, ϕ KZ gp54, and their homologs in related phages can be definitely assigned as scaffold proteins. Ideally, such evidence would be derived from analyses of giant phage prohead particles containing the scaffold protein. Such evidence is not straightforward to obtain, since targeted genome editing methodology has not been adapted for these giant phages. Even if a gp22 mutant were to be identified, and assuming gp22 is the scaffold, the phenotype produced under nonpermissive conditions from this mutation may not produce conclusive evidence. This is because without a scaffold protein it is likely that no prohead core could form, which would then hinder the formation of anything recognizable as a head, or prohead,

by electron microscopy. This phenotype could be caused by mutations in numbers of SPN3US genes, such as those with roles in gene expression, in addition to a mutation in its scaffold gene.

Based on our knowledge of the major maturation steps in giant phage head assembly, a prohead protease mutant phage might be the best choice for the acquisition of direct evidence for the identification of the giant phage scaffold within the prohead. In T4, prohead protease mutants facilitated analyses of proheads that were essentially frozen in a state that prevented normal head proteolytic maturation and therefore the loss of scaffold and core proteins (44). Our future studies will focus on developing our SPN3US genetic system to obtain direct evidence for the identity of the giant phage prohead scaffold protein, as well as to unravel other mechanisms of assembly and infection, of this remarkable giant phage.

The objective of this study was to use mass spectrometry to monitor the fluxes in viral and bacterial proteins during giant phage liquid infection. Our results demonstrate that a remarkable coverage of the *Salmonella* phage SPN3US proteome was achieved, providing valuable information on the temporal production of phage proteins during infection. In this study, we identified a candidate for the SPN3US prohead scaffold protein, gp22, based on its high abundance in infected bacteria that contrasted with its absence from mature virions. This correlation is further supported by the observed proteolytic cleavage of the homolog in related giant phage ϕ KZ. Proteolysis of the scaffold protein is an event that has been observed to promote head maturation of some other phages. Our studies highlight the value of SPN3US as a model giant phage and also the need for further investigation to understand the mechanisms of giant phage head assembly, and the functions of novel proteins, such as E proteins. Importantly, the proteome characterization in this study lays the groundwork for future experiments that combine analyses from several “omics” levels that can facilitate deeper understanding of the molecular mechanisms of the bacterial-phage interactome. Such studies will be of relevance to many disciplines, including phage structure and assembly and phage therapy.

MATERIALS AND METHODS

Propagation of the phages SPN3US and ϕ KZ. SPN3US was propagated on *Salmonella enterica* subsp. *enterica* serovar Typhimurium and ϕ KZ on *Pseudomonas aeruginosa* PAO1 in M9CA medium supplemented with 2 mM MgSO_4 , 1 mM CaCl_2 , and 0.4% glucose. For the time series analyses of SPN3US infection, an overnight culture of *S. Typhimurium* in Luria-Bertani broth was subcultured (1:100 dilution) into 1 liter of medium at 30°C with shaking. At an optical density at 600 nm of 0.30, two 400-ml aliquots were taken; one was infected with SPN3US at an MOI of 2 and, to the uninfected control, an equivalent volume of SM buffer (50 mM Tris-HCl [pH 7.5], 100 mM NaCl, 10 mM MgSO_4 , 0.01% [wt/vol] gelatin) was added. Cultures were incubated with shaking (150 rpm), and at each time point a 15-ml aliquot of each culture was centrifuged at $4,000 \times g$ for 2 min at room temperature to pellet the cells. The cells were resuspended in 80 μ l of SDS loading buffer and immediately boiled.

Mass spectrometry analysis of phage-infected cells. Proteins were separated by electrophoresis in Criterion XT MOPS 12% SDS-PAGE reducing gels (Bio-Rad), with subsequent protein visualization by staining with Coomassie blue. Each gel lane was divided into six slices. After destaining, proteins in the gel slices were reduced with TCEP [Tris(2-carboxyethyl)phosphine hydrochloride] and then alkylated with iodoacetamide before digestion with trypsin (Promega). High-pressure liquid chromatography/electrospray ionization/tandem mass spectrometry was accomplished on a Thermo Fisher LTQ Orbitrap Velos Pro mass spectrometer or a Thermo Fisher Orbitrap Fusion Lumos Tribrid mass spectrometer. Mascot (Matrix Science, London, UK) was used to search the mass spectrometry (MS) files against a locally generated SPN3US protein database that had been concatenated with the Swiss-Prot database (2012_11_170320; version 51.6).

Subset searching of the Mascot data, determination of probabilities of peptide assignments, and protein identifications were accomplished by Scaffold (Proteome Software). MS data files for each entire gel lane were combined via the “MudPIT” option. The results for identified SPN3US proteins, numbers of unique peptides, total spectra, and sequence coverage for each experiment were exported from Scaffold with the following quality filters: peptide, 95%; protein, 99.9%; and minimum number of peptides, 2 (or 1 peptide in selected instances). An estimate of the relative amount of each SPN3US protein produced in each sampling period was calculated by dividing the total number of spectra identified at each time point by protein molecular weight (SC/M) as performed previously for phages 0305 ϕ 8-36 (45), 201 ϕ 2-1 (36, 45), RIO-1 (46), and ϕ KZ (11). These data were also normalized for the highest number of spectra identified for each protein, and these data were displayed using GenVision version 14.1.0.118 (DNASTAR, Inc., Madison, WI). The results for identified *Salmonella* proteins for each experiment were exported from Scaffold (1% false discovery rate).

The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE (47) partner repository with the data set identifiers PXD011697 (time series of SPN3US-infected and control cultures of *S. Typhimurium* analyzed on the Lumos mass spectrometer), PXD011521 (SPN3US-infected and control cultures of *S. Typhimurium* at Time-20 analyzed on the Orbitrap mass spectrometer), and PXD011520 (ϕ KZ-infected and control cultures of *P. aeruginosa* at Time-20 analyzed on the Orbitrap mass spectrometer).

Expression and purification of ϕ KZ gp54. The gene encoding ϕ KZ gp54 was synthesized with the codons for an N-terminal His₆ tag (IDT) and ligated into the pHERD20T vector (48) using the BamHI and XbaI restriction sites. This construct was transformed into *E. coli* DH10B and verified via Sanger sequencing. Assays of ϕ KZ gp54 with the mature form of the ϕ KZ prohead protease (gp175PALE) were performed as described previously (35). The MS proteomics data from this assay have been deposited in the ProteomeXchange Consortium via the PRIDE (47) partner repository with the data set identifier PXD011519.

Bioinformatics analyses. Protein database searches were conducted using a locally implemented version of PSI-BLAST against the NCBI nonredundant (nr) plus environmental protein (env_nr) databases (49). Hidden Markov models (HMMs) were made from alignments of homologs to each protein using the Sequence and Alignment Modelling software system (SAM) (50). Homolog sets were checked to ensure identical sequences were not present in both sets. Alignments were visualized using GeneDoc (51). HMM-HMM searches were performed using HHpred (52) and after HMMs had been calibrated against a scop70 HMM library.

Accession number(s). Data have been deposited in the ProteomeXchange Consortium with the data set identifiers PXD011697 (<https://doi.org/10.6019/PXD011697>), PXD011521 (<https://doi.org/10.6019/PXD011521>), PXD011520 (<https://doi.org/10.6019/PXD011520>), and PXD011519 (<https://doi.org/10.6019/PXD011519>).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/JVI.01833-18>.

SUPPLEMENTAL FILE 1, PDF file, 0.6 MB.

ACKNOWLEDGMENTS

We acknowledge Assitan Coulibaly and Elizabeth Aguilera for the isolation of SPN3US amber mutants am34, am46, and am47; the members of Phage Biology (BIOL 335) Hamzah Megat and David Adler for the isolation of am90; Andrea D. Benítez for the isolation of am111; and Martine Bosch and Adriana Coll De Peña for the isolation of am113. We thank David Lawlor and the members of Genomics (BIOL.340) for the sequencing of SPN3US amber mutant am90. We also thank Stephen C. Hardies for invaluable discussions and allowing use of his bioinformatics resources; Borries Demeler and the UTHSCSA (The University of Texas Health Science Center at San Antonio) Bioinformatics Center for assistance with computational aspects of the project; Michelle Zanache, John Ashton, and Jason Myers from the University of Rochester Genomics Research Center for the sequencing of SPN3US mutants; Andy Lu for assistance with cloning of phage genes; and Sara Mallory for assistance with GenVision. Mass spectrometry analyses were conducted at the UTHSCSA Institutional Mass Spectrometry Laboratory.

Research reported in this publication was supported by a Dean's Research Initiation Grant, the Thomas H. Gosnell School of Life Sciences, and the National Institute of General Medical Sciences of the National Institutes of Health (NIH) under award UA5GM126533 (J.A.T.). The UTHSCSA Institutional Mass Spectrometry Laboratory is supported in part by UTHSCSA, NIH grant 1S10RR025111-01 for purchase of the Orbitrap mass spectrometer (S.T.W.), and the University of Texas Proteomics Core Network (S.T.W.) for purchase of the Lumos mass spectrometer. L.W.B. was supported by NIH R01 AI011676 and GM11876. This study was supported in part by the intramural research program of NIAMS. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. Hendrix RW. 2009. Jumbo bacteriophages, p 229–240. In Van Etten J (ed), Lesser known large dsDNA viruses, vol 328. Springer, Berlin, Germany.
2. Krylov VN, Dela Cruz DM, Hertveldt K, Ackermann HW. 2007. ϕ KZ-like viruses, a proposed new genus of myovirus bacteriophages. Arch Virol 152:1955–1959. <https://doi.org/10.1007/s00705-007-1037-7>.

3. Mesyanzhinov VV, Robben J, Grymonprez B, Kostyuchenko VA, Bourkaltseva MV, Sykilinda NN, Krylov VN, Volckaert G. 2002. The genome of bacteriophage ϕ KZ of *Pseudomonas aeruginosa*. *J Mol Biol* 317:1–19. <https://doi.org/10.1006/jmbi.2001.5396>.
4. Serwer P, Hayes SJ, Zaman S, Lieman K, Rolando M, Hardies SC. 2004. Improved isolation of undersampled bacteriophages: finding of distant terminase genes. *Virology* 329:412–424. <https://doi.org/10.1016/j.virol.2004.08.021>.
5. Serwer P, Hayes SJ, Thomas J, Demeler B, Hardies SC. 2009. Isolation of novel large and aggregating bacteriophages, p 55–66. In Clokie M, Kropinski AM (ed), *Bacteriophages: methods and protocols*. Humana Press, New York, NY.
6. Serwer P, Hayes SJ, Thomas JA, Hardies SC. 2007. Propagating the missing bacteriophages: a large bacteriophage in a new class. *Virol J* 4:21. <https://doi.org/10.1186/1743-422X-4-21>.
7. Jang HB, Fagutao FF, Nho SW, Park SB, Cha IS, Yu JE, Lee JS, Im SP, Aoki T, Jung TS. 2013. Phylogenomic network and comparative genomics reveal a diverged member of the ϕ KZ-related group, marine *Vibrio* phage JM-2012. *J Virol* 87:12866–12878. <https://doi.org/10.1128/JVI.02656-13>.
8. Lee J-H, Shin H, Kim H, Ryu S. 2011. Complete genome sequence of *Salmonella* bacteriophage SPN3US. *J Virol* 85:13470–13471. <https://doi.org/10.1128/JVI.06344-11>.
9. Yagubi AI, Castle AJ, Kropinski AM, Banks TW, Svircev AM. 2014. Complete genome sequence of *Erwinia amylovora* bacteriophage vB_EamM_Ea35-70. *Genome Announcements* 2:e00413-14. <https://doi.org/10.1128/genomeA.00413-14>.
10. Thomas JA, Benítez Quintana AD, Bosch MA, Coll De Peña A, Aguilera E, Coulibaly A, Wu W, Osier MV, Hudson AO, Weintraub ST, Black LW. 2016. Identification of essential genes in the *Salmonella* phage SPN3US reveals novel insights into giant phage head structure and assembly. *J Virol* 90:10284–10298. <https://doi.org/10.1128/JVI.01492-16>.
11. Thomas JA, Weintraub ST, Wu W, Winkler DC, Cheng N, Steven AC, Black LW. 2012. Extensive proteolysis of head and inner body proteins by a morphogenetic protease in the giant *Pseudomonas aeruginosa* phage ϕ KZ. *Mol Microbiol* 84:324–339. <https://doi.org/10.1111/j.1365-2958.2012.08025.x>.
12. Fokine A, Battisti A, Bowman V, Efimov A, Kurochkina L, Chipman P, Mesyanzhinov V, Rossmann M. 2007. CryoEM study of the *Pseudomonas* bacteriophage ϕ KZ. *Structure* 15:1099–1104. <https://doi.org/10.1016/j.str.2007.07.008>.
13. Fokine A, Kostyuchenko VA, Efimov AV, Kurochkina LP, Sykilinda NN, Robben J, Volckaert G, Hoenger A, Chipman PR, Battisti AJ, Rossmann MG, Mesyanzhinov VV. 2005. A three-dimensional cryo-electron microscopy structure of the bacteriophage ϕ KZ head. *J Mol Biol* 352:117–124. <https://doi.org/10.1016/j.jmb.2005.07.018>.
14. Sycheva LV, Shneider MM, Sykilinda NN, Ivanova MA, Miroshnikov KA, Leiman PG. 2012. Crystal structure and location of gp131 in the bacteriophage ϕ KZ virion. *Virology* 434:257–264. <https://doi.org/10.1016/j.virol.2012.09.001>.
15. Bazla A, Desmond MI, Mallory SA, Benítez AD, Buckley LJ, Weintraub ST, Osier MV, Black LW, Thomas JA. 2017. To be or not to be T4: evidence of a complex evolutionary pathway of head structure and assembly in giant *Salmonella* virus SPN3US. *Front Microbiol* 8:2251. <https://doi.org/10.3389/fmicb.2017.02251>.
16. Krylov VN, Smirnova TA, Minenkova IB, Plotnikova TG, Zhazikov IZ, Khrenova EA. 1984. *Pseudomonas* bacteriophage ϕ KZ contains an inner body in its capsid. *Can J Microbiol* 30:758–762.
17. Wu W, Thomas JA, Cheng N, Black LW, Steven AC. 2012. Bubblegrams reveal the inner body structure of ϕ KZ. *Science* 335:182. <https://doi.org/10.1126/science.1214120>.
18. Ceyssens P-J, Minakhin L, Van den Bossche A, Yakunina M, Klimuk E, Blasdel B, De Smet J, Noben J-P, Bläsi U, Severinov K, Lavigne R. 2014. Development of giant bacteriophage ϕ KZ is independent of the host transcription apparatus. *J Virol* 88:10501–10510. <https://doi.org/10.1128/JVI.01347-14>.
19. Yakunina M, Artamonova T, Borukhov S, Makarova KS, Severinov K, Minakhin L. 2015. A non-canonical multisubunit RNA polymerase encoded by a giant bacteriophage. *Nucleic Acids Res* 43:10411–10420. <https://doi.org/10.1093/nar/gkv1095>.
20. Sousa R, Mukherjee S. 2003. T7 RNA polymerase. *Prog Nucleic Acids Res Mol Biol* 73:1–41.
21. Glegghorn ML, Davydova EK, Rothman-Denes LB, Murakami KS. 2008. Structural basis for DNA-hairpin promoter recognition by the bacteriophage N4 virion RNA polymerase. *Mol Cell* 32:707–717. <https://doi.org/10.1016/j.molcel.2008.11.010>.
22. Kazmierczak KM, Davydova EK, Mustaev AA, Rothman-Denes LB. 2002. The phage N4 virion RNA polymerase catalytic domain is related to single-subunit RNA polymerases. *EMBO J* 21:5815–5823.
23. Thomas JA, Rolando MR, Carroll CA, Shen PS, Belnap DM, Weintraub ST, Serwer P, Hardies SC. 2008. Characterization of *Pseudomonas chlororaphis* myovirus 201 ϕ 2-1 via genomic sequencing, mass spectrometry, and electron microscopy. *Virology* 376:330–338. <https://doi.org/10.1016/j.virol.2008.04.004>.
24. Lavysh D, Sokolova M, Slashcheva M, Förstner KU, Severinov K. 2017. Transcription profiling of *Bacillus subtilis* cells infected with AR9, a giant phage encoding two multisubunit RNA polymerases. *mBio* 8:e02041-16. <https://doi.org/10.1128/mBio.02041-16>.
25. Sokolova M, Borukhov S, Lavysh D, Artamonova T, Khodorkovskii M, Severinov K. 2017. A non-canonical multisubunit RNA polymerase encoded by the AR9 phage recognizes the template strand of its uracil-containing promoters. *Nucleic Acids Res* 45:5958–5967. <https://doi.org/10.1093/nar/gkx264>.
26. Leskinen K, Blasdel B, Lavigne R, Skurnik M. 2016. RNA-Sequencing reveals the progression of phage-host interactions between ϕ R1-37 and *Yersinia enterocolitica*. *Viruses* 8:111. <https://doi.org/10.3390/v8040111>.
27. Cornelissen A, Hardies SC, Shaburova OV, Krylov VN, Mattheus W, Kropinski AM, Lavigne R. 2012. Complete genome sequence of the giant virus OBP and comparative genome analysis of the diverse ϕ KZ-related phages. *J Virol* 86:1844–1852. <https://doi.org/10.1128/JVI.06330-11>.
28. Zybailov B, Coleman MK, Florens L, Washburn MP. 2005. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal Chem* 77:6218–6224. <https://doi.org/10.1021/ac050846r>.
29. Black LW, Showe MK, Steven AC. 1994. Morphogenesis of the T4 head, p 218–258. In Karam JD (ed), *Molecular biology of bacteriophage T4*. ASM Press, Washington, DC.
30. Black LW, Thomas JA. 2012. Condensed genome structure, p 469–487. In Rossmann MG, Rao VB (ed), *Viral molecular machines*, vol 726. Springer, New York, NY.
31. Hertveldt K, Lavigne R, Pleteneva E, Sernova N, Kurochkina L, Korchevskii R, Robben J, Mesyanzhinov V, Krylov V, Volckaert G. 2005. Genome comparison of *Pseudomonas aeruginosa* large phages. *J Mol Biol* 354:536–545. <https://doi.org/10.1016/j.jmb.2005.08.075>.
32. Skurnik M, Hyytiäinen HJ, Happonen LJ, Kiljunen S, Datta N, Mattinen L, Williamson K, Kristo P, Szeliga M, Kalin-Mänttari L, Ahola-Iivarinen E, Kalkkinen N, Butcher SJ. 2012. Characterization of the genome, proteome, and structure of yersiniophage ϕ R1-37. *J Virol* 86:12625–12642. <https://doi.org/10.1128/JVI.01783-12>.
33. Lecoutere E, Ceyssens P-J, Miroshnikov KA, Mesyanzhinov VV, Krylov VN, Noben J-P, Robben J, Hertveldt K, Volckaert G, Lavigne R. 2009. Identification and comparative analysis of the structural proteomes of ϕ KZ and EL, two giant *Pseudomonas aeruginosa* bacteriophages. *Proteomics* 9:3215–3219. <https://doi.org/10.1002/pmic.200800727>.
34. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Ruger W. 2003. Bacteriophage T4 genome. *Microbiol Mol Biol Rev* 67:86–156.
35. Thomas JA, Black LW. 2013. Mutational analysis of the *Pseudomonas aeruginosa* myovirus ϕ KZ morphogenetic protease gp175. *J Virol* 87:8713–8725. <https://doi.org/10.1128/JVI.01008-13>.
36. Thomas JA, Weintraub ST, Hakala K, Serwer P, Hardies SC. 2010. Proteome of the large *Pseudomonas* myovirus 201 ϕ 2-1: delineation of proteolytically processed virion proteins. *Mol Cell Proteomics* 9:940–951. <https://doi.org/10.1074/mcp.M900488-MCP200>.
37. Mathews CK. 1994. An overview of the T4 developmental program, p 1–8. In Karam JD (ed), *Molecular biology of bacteriophage T4*. ASM Press, Washington, DC.
38. Fokine A, Rossmann MG. 2016. Common evolutionary origin of procapsid proteases, phage tail tubes, and tubes of bacterial type VI secretion systems. *Structure* 24:1928–1935. <https://doi.org/10.1016/j.str.2016.08.013>.
39. Liu J, Mushegian A. 2004. Displacements of prohead protease genes in the late operons of double-stranded-DNA bacteriophages. *J Bacteriol* 186:4369–4375. <https://doi.org/10.1128/JB.186.13.4369-4375.2004>.
40. Cheng H, Shen N, Pei J, Grishin NV. 2004. Double-stranded DNA bacteriophage prohead protease is homologous to herpesvirus protease. *Protein Sci* 13:2260–2269. <https://doi.org/10.1110/ps.04726004>.
41. Welch AR, McNally LM, Hall MR, Gibson W. 1993. Herpesvirus proteinase:

- site-directed mutagenesis used to study maturational, release, and in-activation cleavage sites of precursor and to identify a possible catalytic site serine and histidine. *J Virol* 67:7360–7372.
42. Kennard J, Rixon FJ, McDougall IM, Tatman JD, Preston VG. 1995. The 25 amino acid residues at the carboxy terminus of the herpes simplex virus type 1 UL26.5 protein are required for the formation of the capsid shell around the scaffold. *J Gen Virol* 76:1611–1621. <https://doi.org/10.1099/0022-1317-76-7-1611>.
  43. Cardone G, Newcomb WW, Cheng N, Wingfield PT, Trus BL, Brown JC, Steven AC. 2012. The UL36 tegument protein of herpes simplex virus 1 has a composite binding site at the capsid vertices. *J Virol* 86:4058–4064. <https://doi.org/10.1128/JVI.00012-12>.
  44. Onorato L, Stirmer B, Showe MK. 1978. Isolation and characterization of bacteriophage T4 mutant preheads. *J Virol* 27:409–426.
  45. Thomas JA, Hardies SC, Rolando M, Hayes SJ, Lieman K, Carroll CA, Weintraub ST, Serwer P. 2007. Complete genomic sequence and mass spectrometric analysis of highly diverse, atypical *Bacillus thuringiensis* phage 0305 ϕ 8-36. *Virology* 368:405–421. <https://doi.org/10.1016/j.virol.2007.06.043>.
  46. Hardies SC, Thomas JA, Black L, Weintraub ST, Hwang CY, Cho BC. 2016. Identification of structural and morphogenesis genes of *Pseudoalteromonas* phage ϕ RIO-1 and placement within the evolutionary history of *Podoviridae*. *Virology* 489:116–127. <https://doi.org/10.1016/j.virol.2015.12.005>.
  47. Vizcaino JA, Csordas A, del-Toro N, Dianas JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, Xu QW, Wang R, Hermjakob H. 2016. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* 44:D447–D456. <https://doi.org/10.1093/nar/gkv1145>.
  48. Qiu D, Damron FH, Mima T, Schweizer HP, Yu HD. 2008. PBAD-based shuttle vectors for functional analysis of toxic and highly regulated genes in *Pseudomonas* and *Burkholderia* spp. and other bacteria. *Appl Environ Micro* 74:7422–7426. <https://doi.org/10.1128/AEM.01369-08>.
  49. 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.
  50. Hughey R, Karplus K, Krogh A. 2003. posting date. SAM: Sequence alignment and modeling software system. Technical Report UCSC-CRL-99-11. University of California, Santa Cruz, CA. https://compbio.soe.ucsc.edu/papers/sam_doc/sam_doc.html.
  51. Nicholas KB, Nicholas HBJ. 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. (Distributed by the author.).
  52. Söding J. 2005. Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21:951–960. <https://doi.org/10.1093/bioinformatics/bti125>.
  53. Fokine A, Chipman PR, Leiman PG, Mesyanzhinov VV, Rao VB, Rossmann MG. 2004. Molecular architecture of the prolate head of bacteriophage T4. *PNAS* 101:6003–6008. <https://doi.org/10.1073/pnas.0400444101>.