PHAGE: Therapy, Applications, and Research
Volume 3, Number 4, 2022

DOI: 10.1089/phage.2022.0003

Genomic Profiling of Non-O157 Shiga Toxigenic *Escherichia coli*-Infecting Bacteriophages from South Africa

Emmanuel W. Bumunang, PhD,[1] Tim A. McAllister, PhD,[2] Rodrigo Ortega Polo, MSc,[2] Collins N. Ateba, PhD,[3] Kim Stanford, PhD,[4] Jared Schlechte, BSc,[1] Matthew Walker, BSc,[5] Kellie MacLean, MDSA,[6] and Yan D. Niu, PhD[1]

Abstract

Background: Non-O157 Shiga toxigenic *Escherichia coli* (STEC) are one of the most important food and waterborne pathogens worldwide. Although bacteriophages (phages) have been used for the biocontrol of these pathogens, a comprehensive understanding of the genetic characteristics and lifestyle of potentially effective candidate phages is lacking.
Materials and Methods: In this study, 10 non-O157-infecting phages previously isolated from feedlot cattle and dairy farms in the North-West province of South Africa were sequenced, and their genomes were analyzed.
Results: Comparative genomics and proteomics revealed that the phages were closely related to other *E. coli*-infecting *Tunaviruses*, *Seuratviruses*, *Carltongylesviruses*, *Tequatroviruses*, and *Mosigviruses* from the National Center for Biotechnology Information GenBank database. Phages lacked integrases associated with a lysogenic cycle and genes associated with antibiotic resistance and Shiga toxins.
Conclusions: Comparative genomic analysis identified a diversity of unique non-O157-infecting phages, which could be used to mitigate the abundance of various non-O157 STEC serogroups without safety concerns.

Keywords: bacteriophages, Shiga toxigenic *Escherichia coli*, whole genome sequencing, comparative genomics

Introduction

Non-O157 Shiga toxigenic *Escherichia coli* (STEC) comprise a variety of pathogenic *E. coli*. Non-O157 STEC are among the most important water and foodborne bacterial infectious agents associated with hemorrhagic colitis and hemolytic uremic syndrome in humans.[1–3] Although research interest has focused on the six serogroups (O26, O45, O103, O111, O121, and O145) most frequently associated with human disease,[4,5] other non-O157 STEC serogroups such as O99, O116, O129, O154, and O156 in cattle can be considered as emerging pathotypes.[6] Pathogens in food-processing plants and hospitals are commonly controlled using biocides and antibiotics.[7,8]

However, the use of antimicrobials to control STEC infections in humans can increase toxin production and result in undesirable health outcomes.[9] In addition, some non-O157 STEC strains are resistant to antibiotics[10] and thus, alternative measures for controlling non-O157 STEC are required.

Bacteriophages (phages) are obligate parasites that specifically infect bacteria, propagate, and, in the case of lytic phages, kill their hosts by inducing cell lysis.[11] These features of self-replicating and host-specificity make phages potential pathogen-specific control agents.[12] The taxonomic classification of viruses is under the direction of the International Committee on Taxonomy of Viruses (ICTV). Classification of viruses is based on the morphology, host range, replication

[1]Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, Canada.
[2]Agriculture and Agri-Food Canada, Lethbridge Research and Development Centre, Lethbridge, Canada.
[3]Department of Microbiology, Faculty of Natural and Agricultural Sciences, North-West University, Mmabatho, South Africa.
[4]Department of Biological Sciences, University of Lethbridge, Lethbridge, Canada.
[5]Canadian Science Centre for Human and Animal Health, Public Health Agency of Canada, Winnipeg, Canada.
[6]Cumming School of Medicine, Faculty of Science, University of Calgary, Calgary, Canada.

cycle, and genomic information.[13] Phage genomes consist of single- or double-stranded DNA or RNA molecules,[14] with the majority (96%) of tailed phages belonging to the new class *Caudoviricetes*[15] in the order *Caudovirales*.

Tailed phages are classified into 14 families: *Myoviridae*,[16] *Siphoviridae*,[16] and *Podoviridae*,[16] *Ackermannviridae*,[17] *Autographiviridae*,[18] *Chaseviridae*,[18] *Demerecviridae*,[18] *Drexlerviridae*,[18] *Herelleviridae*,[18] *Guelinviridae*,[19] *Rountreeviridae*,[19] *Salasmaviridae*,[19] *Schitoviridae*,[19] and *Zobellviridae*.[19] *Tunaviruses* (or formerly T1-like phages) have been classified into the *Drexlerviridae* and have genomes of ~50 kb with G/C content of 46.0% and 79 protein-coding genes.[18] In contrast, *Tequatroviruses* and *Mosigviruses* (or formerly T4-like phages) are known to have genomes averaging 168 kb with a G/C content of ~34.5%, which encode ~289 proteins and 8 transfer RNAs (tRNAs).[20] Phages in the family *Chaseviridae* possess a genome of ~54.2 kb with G/C content of 46.5%, and 77 protein-coding genes with or without tRNAs.[18]

Classification updates from the ICTV Bacterial and Archaeal Viruses Subcommittee highlight the creation of a new phage realm, orders, families, subfamilies, genera, and species.[19] Sequence-based characterization of phages through comparative genomics and proteomics offers considerable insight into phage classification and diversity.[21–23]

The number of phage genome sequences available in public databases has escalated to ~14,244 as of January 2021[24] since phage φX174 was first sequenced in 1977.[25] Approximately 7.5% of these sequenced phages infect *E. coli*,[26] but only a fraction of these have been specifically assessed for their ability to control non-O157 STEC.

Bacteriophages (vB_EcoS_SA12KD, vB_EcoS_SA30RD, vB_EcoS_SA32RD, vB_EcoS_SA80RD, vB_EcoS_SA126VB, vB_EcoM_SA91KD, vB_EcoM_SA20RB, vB_EcoM_SA21RB, vB_EcoM_SA35RD, and vB_EcoM_SA79RD) that infect non-O157 STEC were isolated from feedlot cattle and dairy farms in the North-West province of South Africa and characterized by transmission electron microscopy (TEM), restriction digestion, and for host range.[27]

Electron microscopy revealed that these phages belong to the *Myoviridiae* and *Siphoviridae* of the order *Caudovirales*. In this study, we expand on previous investigations by performing a sequence-based characterization through comparative genomics, proteomics, and phylogenetic analysis of the non-O157 STEC phages isolated from cattle in South Africa. Emphasis was also directed at determining the presence or absence of virulence factors and antimicrobial resistant genes in anticipation that these phages may have application as biocontrol agents.

Materials and Methods

Phage DNA extraction and sequencing

For genomic DNA extraction, 2 mL of purified stocks of phage lysates (10^8–10^9 PFU/mL) was transferred to a 2.5-mL microcentrifuge tube and centrifuged for 10 min at 8000 *g* at room temperature. Aliquots (1.3 mL) of the phage supernatant were then transferred to 15 mL Falcon tubes. Thirteen microliters of DNase 1 (10 μg/mL) and RNase A (30 μg/mL) (Sigma-Aldrich, Okaville, Canada) was added and the mixture was incubated at room temperature for 15 min to digest free DNA and RNA. Phage genomic DNA was extracted from the resultant samples using a phage DNA isolation kit (Norgen Biotek Corp., Ontario, Canada) according to the manufacturer's protocol.

The purity and concentration of the DNA were determined using the Nanodrop Lite spectrophotometer (Thermo Fisher Scientific, Verona, WI). Phage DNA was submitted to the Canadian Science Centre for Human and Animal Health, Public Health Agency of Canada, Winnipeg, Manitoba, for sequencing. The samples were prepped using Nextera XT DNA Library Preparation Kit.[28] Sequencing was performed on an Illumina MiSeq with a V2 300 cycles kit to produce 150 bp paired-end reads.

Bioinformatics analyses of sequence data

De novo assembly of sequence reads was performed using SPAdes version 3.11.1[29] with Kmer values set to 21, 33, 55, 77, 99, and 127. Assembled genomes were annotated using a reproducible workflow (https://github.com/jaredmychal/phage Annotation) built with Snakemake.[30] The annotation workflow included Prokka[31] to predict putative open reading frames (ORFs) and proteins. Predicted ORFs were compared locally with all available viral proteins from the National Center for Biotechnology Information (NCBI) RefSeq Release 99[32] database. Putative tRNAs were predicted by Prokka[31] using Aragorn tool at (http://130.235.46.10/ARAGORN/)[33] and then additionally by tRNAscan-SE at (http://lowelab.ucsc.edu/tRNAscan-SE/).[34]

Regulatory regions such as rho-independent terminators and promoter regions were predicted using TransTermHP v2.08,[35] WebGeSTer DB,[36] and bTSSfinder.[37] Predicted ORF nucleotide sequences were screened for antimicrobial resistance (AMR) and virulence genes using ABRicate version 0.8.7 (https://github.com/tseemann/ABRICATE).[38] Nucleotide sequences were compared against the following data sets in ABRicate: NCBI AMRFinderPlus,[39] ARG-ANNOT,[40] Virulence Factor Database VFDB,[41] and ResFinder.[42] Peptide transmembrane regions were described using TMHMM,[43] Phobius,[44] and signal peptides were predicted using SignalP 5.0.[45] Protein homology and function were analyzed using InterProScan 5[46] and EGGnog 5.0[47] and protein sequences were analyzed against the PFAM,[48] GO,[49,50] and KEGG databases.[51]

Peptide coiled-coil domains were predicted by Coils[52] using InterProScan 5.[46] Finally, predicted proteins were compared against the prokaryotic virus orthologous groups database[53] using the hmmsearch database in HMMER3.[54] Predicted protein sequences were then submitted to NCBI PSI-BLASTP[55] and protein annotations were manually reviewed and curated based on homology with reference phages. The best homologue with identities >90% were included in the annotation supplementary tables S1–S10. The annotated genomic sequences of these phages have been deposited in GenBank under the accession numbers SA126VB (OL960573), SA91KD (OL960574), SA80RD (OL960575), SA79RD (OL960576), SA35RD (OL960577), SA32RD (OL960578), SA30RD (OL960579), SA21RB (OL960580), SA20RB (OL960581) and SA12KD (OL960582).

Comparative genomics analysis

The protein sequences of non-O157 STEC phages were obtained from the different ORF regions using CoreGenes 3.5

at (http://binf.gmu.edu:8080/CoreGenes3.5/)[56] to compare predicted proteins among selected phages in the NCBI database. Nucleotide sequence identity of non-O157 STEC phage genomes within the phages and among other *E. coli*-infecting phages was calculated using Virus Intergenomic Distance Calculator (VIRIDIC[57]).

Tunavirus, Seuratavirus, Carltongylesvirus, Tequatrovirus, and Mosigvirus genomes were rendered syntenic by opening at the initiation codon for the small terminase unit, small terminase unit, RNA polymerase, rIIA lysis inhibitor, and rIIB lysis inhibitor, respectively, before MAUVE alignment. To assess the genome organization and ORF orientations of the phages with related *E. coli*-infecting phages in each group, two or more phages from each group in the NCBI database were aligned using progressive Mauve alignment.[58]

Phylogenetic analysis

Amino acid sequences of the large subunit of terminase, portal, capsid, and tail fiber proteins of non-O157 STEC-infecting phages were used for phylogenetic analysis. Phage core proteins were compared with other generic *E. coli*/STEC-infecting phages, representing five genera. Phages were aligned using the multiple sequence alignment application in Clustal Omega through bipython with 100 bootstrap replications. A maximum-likelihood tree was developed using the RAxML version 8, a bioinformatics tool from python command line.[59] A variable threshold (VT) (large subunit of terminase and portal), WAG+G4 (capsid), and VT+G4+F (tail fiber) models were selected based on best Bayesian information criterion score. Phylogenetic trees were visualized using ggtree[60] version 3.4 in R.

Results

Undesirable traits

Overall, bioinformatic analysis of the coding sequences (CDSs) of non-O157 STEC-infecting phages revealed a lack of integrases associated with a lysogenic cycle, and deleterious bacterial genetic markers such as virulence (Shiga toxins) or AMR genes were not identified.

Genomic properties and comparative analysis of phages

Tunavirus. Phages SA32RD, SA30RD, and SA12KD were collinear with each other and have genomes of 48.9–50.7 kb, encoding 73–79 CDSs, and GC content of 45.3–45.6% (Table 1). Phages genome assemblies had read coverages of 462×, 265×, and 418×, respectively, with 6–9 promoters and 24–34 rho-dependent terminators detected (Table 1). Most of the annotated genes were hypothetical proteins, however, some were annotated with putative function such as DNA replication, transcription, lysis, capsid, and tail morphogenesis. No tRNA encoding sequences were detected in T1-like STEC-infecting phages.

One CDS; 51 in SA12KD, a hypothetical protein, had no homology with any coliphage-related sequence in the GenBank database. Upstream CDS1–CDS5 displayed similar genes (i.e., terminase small and large unit, portal protein and major head subunit) with hypothetical proteins from CDS6 to CDS13 followed by a tail tape measure protein; CDS17 to CDS33 (i.e., DNA N-6-adenine-methyltransferase) for SA12KD, SA30RD, and SA32RD. Downstream of CDS33 comprised mostly hypothetical proteins.

Pairwise genomic analysis among phages SA12KD, SA30RD, and SA32RD showed a nucleotide similarity/identity of 87.9% (SA12KD and SA30RD), 84.3% (SA12KD and SA32RD), and 84.6% (SA30RD and SA32RD; Supplementary Table S1). These phages showed a 77.6–88.5% nucleotide similarity/identity with selected *E. coli*-infecting *Tunaviruses* (Table 2).

Similarly, at protein level, CoreGenes 3.5 analysis revealed that phages SA12KD, SA30RD, and SA32RD share 79.5–91.5% similarity/identity of their CDSs to proteins of selected *E. coli*-infecting *Tunaviruses*, including *Escherichia* phage vB_EcoS_Chapo, a O29:H12 STEC-infecting phage[61] (Table 2). Mauve alignment analysis revealed that these non-O157 STEC-infecting phages and other known *Tunaviruses* exhibited similar nucleotide sequences with the same genome orientation (Supplementary Fig. S1). However, the terminal repeats of T1 phages in this study were not determined.

Seuratvirus. Phages SA80RD and SA126VB also had small genomes of 59.7 and 61.7 kb, with 92 and 93 CDSs and GC content of 44.1% and 44.2% (Table 1). Furthermore, 25 and 24 rho-dependent terminators, and 7 promoters each were identified with a read coverage of 346× and 342×, respectively (Table 1). No tRNAs genes were detected in SA80RD or SA126VB. Ninety-two and 93 putative CDSs were identified in SA80RD and SA126VB, respectively. Based on PSI-BLAST, CDSs, 39/92 (42.4%; SA80RD) and 33/93 (35.5%; SA126VB) were assigned putative functions such as DNA replication, transcription, lysis, capsid, and tail morphogenesis, whereas 53/92 (57.6%; SA80RD) and 60/93 (64.5%; SA126VB) were annotated as hypothetical proteins (Table 1).

Upstream genes in SA80RD and SA126VD were associated with two packaging genes, terminase small and large subunits, a structural protein, portal protein, and a capsid protein. Downstream genes were associated with cell lysis genes, including a protease (CDS88; SA80RD), endolysin (CDS90; SA80RD), holin (CDS91; SA80RD), an endonuclease (CDS89), and an endolysin (CDS91) associated with SA126VD.

SA80RD and SA126VD showed 84.2% nucleotide similarity/identity based on VIRIDIC analysis (Supplementary Table S2). Similarly, they share 76.2–79.6% nucleotide sequence identity with other *E. coli*-infecting *Seuratviruses* (Table 2 and Supplementary Fig. S2). Based on CoreGenes 3.5 analysis, proteins of these phages shared homology with proteins of the *Seuratviruses* (Table 2). The comparative analysis with Mauve differentiated the syntenic regions in nucleotide sequence and composition in SA80RD and SA126VD (Supplementary Figure S2). Similar regions and the same nucleotide sequence and composition in SA80RD and SA126VD were also seen in other *Seuratviruses; Enterobacteria* phage CAjan[62] and *Escherichia* phage Seurat[63] (Supplementary Figure S2).

Carltongylesvirus. SA91KD had a 53.6 kb genome; 41.1% G+C mol; 78 CDSs with a read coverage 300× and a 78.8% nucleotide similarity to *Enterobacteria* phage phiEcoM-GJ1.[64] Sixteen rho-dependent terminators and 10

TABLE 1. GENOMIC FEATURES OF NON-O157 SHIGA TOXIGENIC *ESCHERICHIA COLI*-INFECTING PHAGES

Genus	*Tunavirus*			*Carltongylesvirus*	*Seuratvirus*		*Tequatrovirus*		*Mosigvirus*	
Phage name	*SA12KD*	*SA30RD*	*SA32RD*	*SA91KD*	*SA80RD*	*SA126VB*	*SA20RB*	*SA21RB*	*SA35RD*	*SA79RD*
Genome size (bp)	50,709	50,465	48,980	53,673	59,778	61,753	166,337	167,177	169,615	169,615
Total genes	81	75	75	78	92	93	270	271	273	273
GC-content (%)	45.3	45.5	45.6	44.1	44.1	44.2	35.5	35.6	37.6	37.6
No. of transfer RNA	—	—	—	—	—	—	8	11	2	2
Protein coding genes	81	75	75	78	92	93	270	271	273	273
Genes with function prediction	81	75	75	78	92	93	270	271	273	273
Hypothetical proteins	61	52	54	57	53	60	138	146	159	153
No. of rho-independent terminators	31	34	32	16	27	27	64	63	65	65
No. of σ^{70} promoters	6	8	9	20	29	40	83	85	70	70
Coverage (×)	418	265	462	300	346	342	99	142	130	135
Accession no.	(OL960573.1)	(OL960579.1)	(OL960578.1)	(OL960574.1)	(OL960575.1)	(OL960573.1)	(OL960581.1)	(OL960580.1)	(OL960577.1)	(OL960576.1)

Table 2. Nucleotide Sequence Identity and Protein Homologues of Non-O157 Shiga Toxigenic *Escherichia coli*-Infecting Phages with Closely Related Phages Using Virus Intergenomic Distance Calculator* and CoreGenes 3.5,† Respectively

Family	*Genus*	*Closely related phages*	*DNA identity/no. common proteins shared (%)*		
Drexlerviridae	*Tunavirus*	*Escherichia* phage Eco_BIFF	SA12KD	SA30RD	SA32RD
		Escherichia virus T1	81.7/88.2	82.2/81.5	83.2/81.5
		Escherichia phage vB_EcoS_Chapo	77.6/83.3	79.2/79.5	81.2/80.7
			88.5/91.5	88.5/85.7	84.6/85.3
	Seuratvirus	*Enterobacteria* phage Cajan	SA80RD	SA126VB	
		Escherichia phage Skure	79.0/84.6	78.0/89.1	
		Escherichia phage Seurat	79.6/82.6	78.5/82.6	
			77.5/84.1	76.2/88.6	
Chaseviridae	*Carltongylesvirus*	*Enterobacteria* phage phiEcoM-GJ1	SA91KD		
		Escherichia phage ST32	82.4/88		
		Escherichia phage Mangalitsa	88.9/88.6		
		Escherichia phage flopper	79.9/82.9		
			88.9/91.7		
Myoviridae	*Tequatrovirus*	*Escherichia coli* phage wV7	SA20RB	SA21RB	
		Escherichia phage AR1	87.4/91.6	88.7/94.5	
		Escherichia phage vB_EcoM_G50	87.6/90.0	89.0/91.1	
		Escherichia phage vB_EcoM_G2540	92.1/91.2	92.7/94.4	
		Escherichia phage T4	87.4/91.2	86.6/90.8	
			84.7/84.9	84.5/85.0	
	Mosigvirus	*E. coli* O157 typing phage 6	SA35RD	SA79RD	
		Enterobacteria phage RB69	92.4/96.6	92.4/96.8	
			88.8/92.6	88.8/93.0	

*Moraru et al.[57]
†Turner et al.[56]

promoters were identified (Table 1). Although a tRNA gene was found in phiEcoM-GJ1, no tRNA was detected in SA91KD. Based on the PSI-BLASTP-verified annotations, 21 of the 78 CDSs (27%) were assigned a putative function, such as DNA replication, transcription, host lysis, capsid, and tail morphogenesis. Fifty-seven putative CDSs (73.0%) were annotated as hypothetical proteins (Table 1). CDS8 in SA91KD, a hypothetical protein, showed no homology to phiEcoM-GJ1 or to any reported *E. coli*-infecting phage- or bacteria-associated sequences in the NCBI database.

Phage SA91KD was identified as a myovirus based on TEM,[27] with a 79.9–88.9% nucleotide identity and 82.9–91.7% proteins similarity/identity with other *E. coli*-infecting *Carltongylesvirus* (Table 2). Although SA91KD was more closely related to phage ST32,[65] with 85.7% and 88.6% DNA and protein homology, respectively, than to other *Carltongylesvirus*, phiEcoM-GJ1 is the first identified member of *Carltongylesvirus*. Interestingly, SA91KD encoded a single-subunit RNA polymerase and a large terminase subunit with 92% and 99% amino acid identity to those of phiEcoM-GJ1 (Table 3), respectively. Mauve alignment revealed that phage SA91KD possesses the same nucleotide sequence orientation as phiEcoM-GJ1 and three other *Carltongylesviruses* and a similar CDS position (Supplementary Fig. S3).

Tequatrovirus. Phages SA20RB and SA21RB had 83.1% and 82.3% pairwise nucleotide similarity with each other, genomes of 166 and 167 kb, 270 and 271 CDSs, a GC content of 35.6% and 37.6% with a read coverage of 99× and 142×, respectively (Table 1). Sixty-four and 63 rho-dependent terminators and 83 and 85 promoters were identified, respectively (Table 1). Eight tRNAs (i.e., argynyl-, methionyl-, threonyl-, seryl-, prolyl-, glycyl-, leucyl-, and glutamyl-tRNA) were detected in SA20RB, whereas 11 (i.e., argynyl-, histidyl-, asparaginyl-, tyrosyl-, methionyl-, threonyl-, seryl-, prolyl-, glycyl-, leucyl-, and glutamyl-tRNA) were found in SA21RB.

Overall, phages SA20RB and SA21RB were assigned a putative function, namely, DNA replication and metabolism, DNA packaging, structural/morphogenesis, and host lysis, although 51% (138/270) to 53.9% (146/271) of gene products were hypothetical proteins (Table 1).

Table 3. Comparison of Amino Acid Sequence of RNA Polymerase and Large Subunit Terminase of SA91KD with Those of Three *Carltongylesvirus* Phages Using PSI-BLASTP

Carltongylesvirus *phages*	*RNA polymerase/large subunit terminase of SA91KD*			
	Query coverage %	E	*Amino acid % identity*	*Accession no.*
Escherichia phage phiEcoM-GJ1	100/99	0	92/99	YP_001595396.1/YP_001595443.1
Escherichia phage ST32	100/100	0	99/99	YP_009790661.1/YP_009790711.1
Escherichia phage Mangalitsa	100/100	0	97/98	YP_009850471.1/YP_009850524.1

Phages SA20RB and SA21RB exhibited a nucleotide similarity/identity of 91.4% (Supplementary Table S3) and a 84.5–92.1% nucleotide similarity with other known *E. coli*-infecting *Tequatroviruses* in the NCBI database (Table 2). At the protein level, they shared 84.7–92.7% protein homology with the proteins of other *Tequatroviruses* (Table 2). SA20RB and SA21RB were more closely related to *Escherichia* phage vB_EcoM_G50, with a nucleotide identity of 90–91% and protein sequence homology of 91.2–94.4% (Table 2). The conserved pattern in *Tequatrovirus* genes is illustrated in Supplementary Figure S4 with a similar arrangement in CDSs and orientation with other *Tequatrovirus* genomes.

Mosigvirus. Phages SA35RD and SA79RD were collinear with each other, 100% pairwise nucleotide similarity, with a 169 kb genome, 273 CDSs, GC content of 37.6% two tRNAs (i.e., 1 arginyl- and methionyl-tRNA) each, with a read coverage of 130× and 135×, respectively (Table 1). Similarly, 65 rho-dependent terminators and 74 promoters were identified (Table 1). Of 273 CDSs, 57.5% were hypothetical proteins, with 42.5% each assigned functions such as DNA replication and metabolism, DNA packaging, structural/morphogenesis, and host lysis (Table 1).

Phages SA35RD and SA79RD had the same (100%) nucleotide sequence (Supplementary Table S4) and appeared to be clonal. These phages exhibited an 88.8% and 92.4% nucleotide similarity to other *Mosigviruses* such as *E. coli* O157 typing phage 6 and *Enterobacteria* phage RB69 (Table 2). However, SA35RD shared 96.6% and 92.6% protein homology with proteins of O157 typing phage 6 and RB69, respectively (Table 2). Meanwhile SA79RD shared 96.8% and 93% protein homology with proteins of typing phage 6 and RB69, respectively (Table 2). A similar pattern in CDSs arrangement and orientation of SA35RD and SA79RD and phage RB69 were observed (Supplementary Fig. S5).

Phylogenetic analysis

Phylogenetic analysis grouped the large subunit of terminase, portal, capsid, and tail fiber proteins of non-O157 STEC-infecting phages, and those of other phages that infect *E. coli* into their respective genera (Figs. 1 and 2). The large subunit of terminase, portal, and capsid proteins within each genus identified, clustered closely together compared with the tail fiber proteins with distinct terminals at the tip of the tree (Figs. 1 and 2). Tequatroviruses and Mosigviruses formed a monophyletic group with respect to the large subunit of terminase portal, capsid, and tail fiber proteins (Figs. 1 and 2).

Similarly, Carltongylesviruses and Tunaviruses formed a monophyletic group with the portal protein (Fig. 1B). Within the tail fiber phylogeny of non-O157-infecting phages, SA80RD and SA126VD, and SA35RD and SA79RD shared 100% amino acid sequence identity (Fig. 2B), even though SA80RD and SA126VD were isolated from different regions.[27]

Discussion

Phages are abundant in nature and their activity can be species- or even strain-specific. However, much is yet to be uncovered about the diversity and biology of phages in different regions and ecosystems. Therefore, we undertook a comparative sequence-based characterization of 10 non-O157 STEC-infecting phages isolated from feedlot and dairy cattle in the North-West province of South Africa.[27]

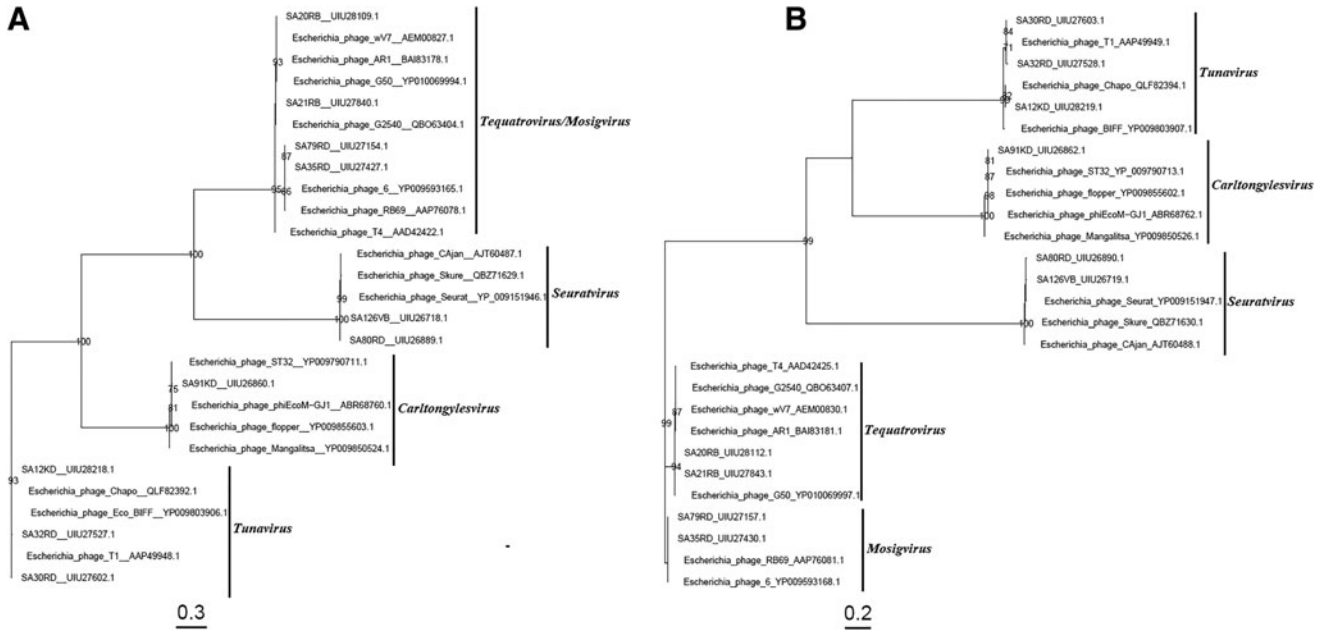

FIG. 1. Phylogenetic tree was constructed using the maximum-likelihood method with 100 bootstrap replications to compare the large subunit of terminase **(A)** and portal **(B)** proteins amino acid sequences of the non-O157 phages and other related *Escherichia coli*/STEC-infecting phage, chosen to represent the *Tunavirus*, *Seuratvirus*, *Carltongylesvirus*, *Tequatrovirus*, and *Mosigvirus*. The Genbank accession numbers of all amino sequences are as shown. STEC, Shiga toxigenic *Escherichia coli*.

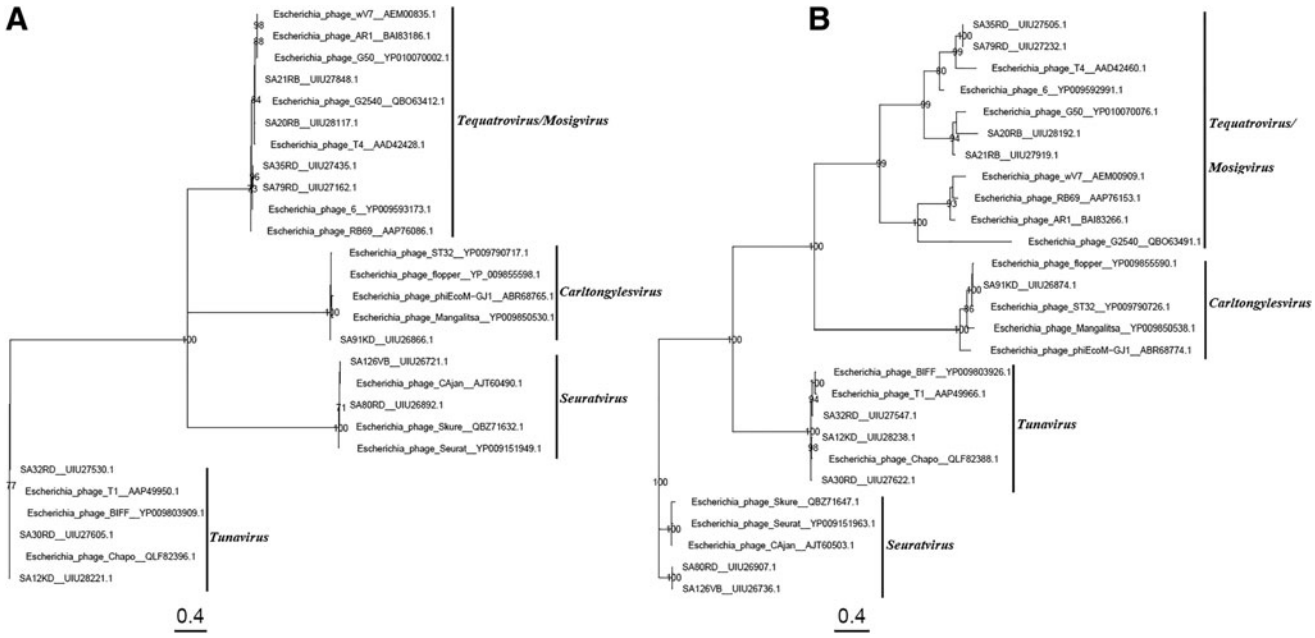

FIG. 2. Phylogenetic tree was constructed using the maximum-likelihood method with 100 bootstrap replications to compare the capsid **(A)** and the tail fiber **(B)** proteins amino acid sequences of the non-O157 phages and other related *Escherichia coli*/STEC-infecting phage, chosen to represent the *Tunavirus*, *Seuratvirus*, *Carltongylesvirus*, *Tequatrovirus*, and *Mosigvirus*. The Genbank accession numbers of all amino sequences are as shown.

In our previous study, five phages SA12KD, SA30RD, SA32RD, SA80RD, and SA126VB were confirmed by TEM as T1-like phages belonging to the family *Siphoviridae*.[27] However, through genomic analyses, SA80RD and SA126VB were further classified to the genera *Seuratvirus*.[18] Likewise, phage SA91KD was further classified to *Carltongylesvirus* of the new *Chaseviridae* family.[18]

In addition, based on genomic analyses, four T4-like phages were identified as members of the *Tequatrovirus* (S21RB and SA20RB) and *Mosigvirus* (SA35RD and SA79RD). Therefore, genome-based characterization revealed that although phages may be phenotypically similar based on TEM, they can still have significant genomic variation. This highlights the continued need for sequence-based characterization of phages in conjunction with morphological features to generate a more precise classification structure of phages.

Bacteriophages largely can rely on host cells for metabolic functions; however, some phages encode additional genes such as tRNA that may play a role in DNA replication and packaging.[66] tRNAs predicted in this study using tRNAscan-SE were considered to be valid tRNAs as the cutoff value was >20.0 bits.[67] Phages with larger genomes support additional genes and are more likely to have tRNA encoding sequences. Six of the phages (SA30RD, SA32RD, SA12KD, SA91KD, SA80RD, and SA126VB) with smaller genomes ($\leq$61.7 kb) lacked tRNA-encoding sequences, whereas the 4 phages (SA20RB, SA21RB, SA35RD, SA79RD) with larger genomes ($\geq$166.3 kb) possessed between 2 and 11 tRNAs genes.

It has been suggested that phages with tRNA genes have a broader host range.[68] Differences in host range between phages with and without tRNA genes in our previous study support this hypothesis given that SA20RB, SA21RB, SA35RD, and SA79RD all exhibited broader host activity against 22 serogroups of non-O157 *E. coli*.[27] Overall, the GC contents of non-O157 phages decreased (44.1–35.5%) with increasing genome size, a relationship that is common across all phage genotypes.[69] The majority of non-O157 phage proteins had unknown functions.

Interestingly, CDS 51 for SA12KD had no homology with any coliphage-related sequence in the GenBank database, whereas SA91KD CDS8 had no homology to phiEcoM-GJ1 or to any reported *E. coli*-infecting phage- or bacteria-associated sequences in the NCBI database. This suggests that these proteins are novel in these phages and corroborates other studies with uncharacterized novel proteins in phages.[70,71]

According to Turner et al.,[23] based on BLASTn, shared DNA identity with values of >95% and >70% between different phages is grouped into the same species and genus, respectively. Phage core genes encoding proteins showed >79.5–96.8% similarity/identity to proteins of other *E. coli*-infecting phages of the *Tunavirus* (SA12KD, SA30RD, and SA32RD), *Seuratvirus* (SA80RD and SA126VB), *Carltongylesvirus* (SA91KD), *Tequatrovirus* (SA20RB and SA21RB), and *Mosigvirus* (SA35RD and SA79RD).

However, non-O157 phages shared <95% DNA identity with other *E. coli*-infecting phage genera, and thus these strains could represent new species. The relatedness of these non-O157 STEC-infecting phages to other selected *E. coli*-infecting phages revealed similar and dissimilar nucleotide regions based on mauve alignment. Similar CDS patterns are indicative of conservation within phage genomes, suggesting that exchange of genetic material may play an infrequent role in contributing to their diversity.

Bacteriophages are important agents of horizontal gene transfer, and can disseminate AMR[72] and virulence[73] genes between bacterial hosts. Therefore, a phage that possesses virulence or AMR factors is not suitable for food safety

applications from a regulatory perspective as it could have attributes that are detrimental to human health. AMR and virulence genes such as Shiga toxins were absent in the non-O157 STEC-infecting phage genomes evaluated.

In addition to these properties, a strictly lytic phage is considered a good biocontrol agent,[74] as it can infect and kill its bacterial host directly without risk of transfer of virulence factors to its host during lysogeny. Phage genomes evaluated in this study did not encode integrases associated with a lysogenic cycle, suggesting they are strictly lytic and have potential as an effective control against non-O157 STEC.

The phage terminase large subunit is a DNA cleavage and packaging enzyme[75] during infection. Cleaved DNA is translocated by a portal protein into the major capsid protein using the headful packaging process.[75] Terminase large subunit, portal, and capsid proteins of non-O157 STEC-infecting phages clustered closely together with other *E. coli*-infecting phages and suggest a lesser variation in amino acid sequences within each genus. Distinct clusters observed within monophyletic groups (Tequatroviruses and Mosigviruses, and Carltongylesviruses, and Tunaviruses) corroborate the importance of conserved gene(s) in genus-level grouping of phages using phylogenetics.[23]

Tail fiber proteins in the order *Caudovirales*[16] play a crucial role in phage–host interactions and host ranges.[76,77] Tail fiber proteins also serve as a genetic marker to infer evolutionary relatedness between tailed phages,[78] and predict the host range of newly isolated phages. The tail fiber proteins of the non-O157 STEC-infecting phage clustered with other *E. coli*-infecting phages, however, with distinct terminals at the tip of the tree that may suggest a greater variation in amino acid sequences that may be predictive of similar or differential host interactions of phages in the same genus.

In addition, adaptive responses from phages within the same genera to different evolving bacterial hosts may drive genetic distinction in tail fiber proteins through mutations within *E. coli*-infecting phages. Although SA80RD and SA126VD are isolated from different areas in the North-West region of South Africa, the amino acid sequence of tail fiber proteins showed greater sequence relatedness among *Seuratviruses*. However, in our previous studies, SA80RD and SA126VD had differences in host specificity.[27] Similarly, tail fiber proteins of SA35RD and SA79RD that appeared to clone differed only in the ability to infect one bacterial host.[27]

Overall, evaluated non-O157 STEC-infecting phages clustered with closely related phages from different regions: Denmark, *Enterobacteria* phage CAjan[62]; Canada, *Enterobacteria* phage phiEcoM-GJ[64]; and Portugal, *Escherichia* phage vB_EcoS_Chapo.[61] Therefore, they have evolutionary relatedness with other *E. coli*-infecting phages isolated from different regions and ecosystems.

For example, *Enterobacteria* phage Cajan, phiEcoM-GJ1, and phage vB_EcoS_Chapo were isolated from rat feces, sewage from pig farms, and wastewater samples, respectively. This suggests core traits of the phages could be conserved/stable over time as they adapt to their *E. coli* hosts from different regions that have similar phenotypic and genotypic traits.

Conclusions

In conclusion, this study applied comparative genomic and proteomic approaches to characterize 10 non-O157 STEC-infecting phages from feedlot cattle and dairy farms in South Africa, revealing that cattle from this region harbor diverse phage genotypes. As these phages do not contain virulence and toxin genes, they may have application in mitigating non-O157 STEC serogroups within cattle that are produced in this region. Whole genome sequencing and comparative analysis proved to be an important tool as it enabled us to better classify non-O157 STEC-infecting phages from the North-West region of South Africa, and to validate their genomic safety for biocontrol.

Authors' Contributions

Conceptualization of the study was carried out by T.A.M., K.S., C.N.A., and Y.D.N.; methodology was taken care of by T.A.M., K.S., C.N.A., E.W.B., and Y.D.N.; software was done by J.S., M.W., E.W.B., and Y.D.N.; validation was carried out by T.A.M., K.S., C.N.A., E.W.B., and Y.D.N.; formal analysis was done by E.W.B., K.M., and J.S.; investigation was done by E.W.B. and Y.D.N.; resources were taken care of by T.A.M., K.S., C.N.A., M.W., K.M., and Y.D.N., writing—original draft preparation—was by E.W.B.; writing—review and editing,—was taken care of by T.A.M., K.S., E.W.B., C.N.A., J.S., R.P.O., and Y.D.N.; supervision was done by T.A.M., K.S., C.N.A., and Y.D.N.; project administration was done by T.A.M., K.S., C.N.A., and Y.D.N; funding acquisition was taken care of by T.A.M. and C.N.A.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

We acknowledge support from the National Research Foundation of South Africa (Grant UID no. 98983), the Agriculture and Agri-Food Canada–Beef Cluster program and the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grant, RGPIN-2019-04384).

Supplementary Material

Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3
Supplementary Figure S4
Supplementary Figure S5
Supplementary Table S1
Supplementary Table S2
Supplementary Table S3
Supplementary Table S4

References

1. Conrad C, Stanford K, McAllister T, et al. Shiga toxin-producing *Escherichia coli* and current trends in diagnostics. Anim Front. 2016;6(2):37–43.
2. Terajima J, Izumiya H, Hara-Kudo Y, et al. Shiga toxin (verotoxin)-producing *Escherichia coli* and foodborne disease: a review. Food Saf (Tokyo). 2017;5(2):35–53.
3. Stanford K, Reuter T, Hallewell J, et al. Variability in characterizing *Escherichia coli* from cattle feces: a cautionary tale. Microorganisms. 2018;6(3):74.

4. Mathusa EC, Chen Y, Enache E, et al. Non-O157 Shiga toxin–producing *Escherichia coli* in foods. J Food Prot. 2010;73(9):1721–1736.
5. Delannoy S, Beutin L, Fach P. Discrimination of enterohemorrhagic *Escherichia coli* (EHEC) from non-EHEC strains based on detection of various combinations of type III effector genes. J Clin Microbiol. 2013;51(10): 3257–3262.
6. Bumunang EW, McAllister TA, King R, et al. Characterization of non-O157 *Escherichia coli* from cattle faecal samples in the North-West Province of South Africa. Microorganisms. 2019;7(8):1–19.
7. Landers TF, Cohen B, Wittum TE, et al. A review of antibiotic use in food animals: perspective, policy, and potential. Public Health Rep. 2012;127(1):4–22.
8. Manyi-Loh C, Mamphweli S, Meyer E, et al. Antibiotic use in agriculture and its consequential resistance in environmental sources: potential public health implications. Molecules. 2018;23(4):795.
9. Tarr PI, Gordon CA, Chandler WL. Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. Lancet. 2005;365(9464):1073–1086.
10. Mukherjee S, Mosci RE, Anderson CM, et al. Antimicrobial drug–resistant Shiga toxin–producing *Escherichia coli* infections, Michigan, USA. Emerg Infect Dis. 2017;23(9):1609–1611.
11. Endersen L, O'Mahony J, Hill C, et al. Phage therapy in the food industry. Annu Rev Food Sci Technol. 2014;5:327–349.
12. Romero-Calle D, Guimarães Benevides R, Góes-Neto A, et al. Bacteriophages as alternatives to antibiotics in clinical care. Antibiotics. 2019;8(3):138.
13. Lefkowitz EJ, Dempsey DM, Hendrickson RC, et al. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Res. 2018; 46(D1):D708–D717.
14. Ackermann HW. Bacteriophage taxonomy. Microbiol Aust. 2011;32(2):90–94.
15. Koonin EV, Dolja VV, Krupovic M, et al. Global organization and proposed megataxonomy of the virus world. Microbiol Mol Biol Rev. 2020;84(2):e00061–00019.
16. Ackermann H-W. Phage classification and characterization. In: Clokie MR, Kropinski AM, eds. *Bacteriophages*. New York, NY: Springer; 2009: 127–140.
17. Adriaenssens EM, Wittmann J, Kuhn JH, et al. Taxonomy of prokaryotic viruses: 2017 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. Arch Virol. 2018;163(4):1125–1129.
18. Adriaenssens EM, Sullivan MB, Knezevic P, et al. Taxonomy of prokaryotic viruses: 2018–2019 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. Arch Virol. 2020: 1–8.
19. Krupovic M, Turner D, Morozova V, et al. Bacterial viruses subcommittee and archaeal viruses subcommittee of the ICTV: update of taxonomy changes in 2021. Arch Virol. 2021 166(11):3239–3244.
20. Miller ES, Kutter E, Mosig G, et al. Bacteriophage T4 genome. Microbiol Mol Biol Rev. 2003;67(1):86–156.
21. Niu YD, McAllister TA, Nash JH, et al. Four *Escherichia coli* O157: H7 phages: a new bacteriophage genus and taxonomic classification of T1-like phages. PLoS One. 2014;9(6):e100426.
22. Korf IH, Meier-Kolthoff JP, Adriaenssens EM, et al. Still something to discover: novel insights into *Escherichia coli* phage diversity and taxonomy. Viruses. 2019;11(5):1–29.
23. Turner D, Kropinski AM, Adriaenssens EM. A roadmap for genome-based phage taxonomy. Viruses. 2021;13(3):506.
24. Cook R, Brown N, Redgwell T, et al. INfrastructure for a PHAge REference Database: identification of large-scale biases in the current collection of phage genomes. bioRxiv. 2021:1–33.
25. Sanger F, Air GM, Barrell BG, et al. Nucleotide sequence of bacteriophage φX174 DNA. Nature. 1977;265(5596): 687.
26. Besler I, Sazinas P, Harrison C, et al. Genome sequence and characterization of coliphage vB_Eco_SLUR29. PHAGE. 2020;1(1):38–44.
27. Bumunang EW, McAllister TA, Anany H, et al. Characterization of non-O157 STEC infecting bacteriophages isolated from cattle faeces in North-West South Africa. Microorganisms. 2019;7(12):1–11.
28. Illumina. *Nextera XT DNA Library Prep Reference Guide*. San Diego, CA: Illumina; 2019.
29. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–477.
30. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics. 2012;28(19): 2520–2522.
31. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–2069.
32. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44(D1):D733–D745.
33. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32(1):11–16.
34. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955–964.
35. Kingsford CL, Ayanbule K, Salzberg SL. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. Genome Biol. 2007;8(2):R22.
36. Mitra A, Kesarwani AK, Pal D, et al. WebGeSTer DB—a transcription terminator database. Nucleic Acids Res. 2011; 39(suppl_1):D129–D135.
37. Shahmuradov IA, Mohamad Razali R, Bougouffa S, et al. bTSSfinder: a novel tool for the prediction of promoters in cyanobacteria and *Escherichia coli*. Bioinformatics. 2017; 33(3):334–340.
38. Seemann T. ABRicate: mass screening of contigs for antimicrobial and virulence genes. Department of Microbiology and Immunology, The University of Melbourne, Melbourne, Australia. 2018. https://github com/tseemann/abricate (last accessed June 25, 2020).
39. Feldgarden M, Brover V, Haft DH, et al. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. Antimicrob Agents Chemother. 2019;63(11):e00483–00419.
40. Gupta SK, Padmanabhan BR, Diene SM, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother. 2014;58(1):212–220.
41. Liu B, Zheng D, Jin Q, et al. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res. 2019;47(D1):D687–D692.

42. Zankari E, Hasman H, Cosentino S, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67(11):2640–2644.
43. Krogh A, Larsson B, Von Heijne G, et al. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001; 305(3):567–580.
44. Käll L, Krogh A, Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 2004;338(5):1027–1036.
45. Armenteros JJA, Tsirigos KD, Sønderby CK, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019;37(4):420–423.
46. Jones P, Binns D, Chang H-Y, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–1240.
47. Huerta-Cepas J, Szklarczyk D, Heller D, et al. EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47(D1):D309–D314.
48. Finn RD, Bateman A, Clements J, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42(D1):D222–D230.
49. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. Nat Genet. 2000;25(1): 25–29.
50. Hayles J. The Gene Ontology Resource: 20years and still GOing strong. Nucleic Acids Res. 2019;47:D330–D338.
51. Ogata H, Goto S, Sato K, et al. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 1999;27(1):29–34.
52. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science. 1991: 1162–1164.
53. Kristensen DM, Waller AS, Yamada T, et al. Orthologous gene clusters and taxon signature genes for viruses of prokaryotes. J Bacteriol. 2013;195(5):941–950.
54. Mistry J, Finn RD, Eddy SR, et al. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 2013;41(12):e121–e121.
55. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17): 3389–3402.
56. Turner D, Reynolds D, Seto D, et al. CoreGenes3. 5: a webserver for the determination of core genes from sets of viral and small bacterial genomes. BMC Res Notes. 2013; 6(1):140.
57. Moraru C, Varsani A, Kropinski AM. VIRIDIC—a novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. Viruses. 2020;12(11):1268.
58. Darling AC, Mau B, Blattner FR, et al. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004;14(7):1394–1403.
59. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–1313.
60. Yu G. Using ggtree to visualize data on tree-like structures. Curr Protoc Bioinformatics. 2020;69(1):e96.
61. Dias C, Almeida C, Łobocka M, et al. Genome sequences of four potentially therapeutic bacteriophages infecting Shiga toxin-producing *Escherichia coli*. Microbiol Resour Announc. 2020;9(36):e00749–00720.
62. Carstens AB, Kot W, Lametsch R, et al. Characterisation of a novel enterobacteria phage, CAjan, isolated from rat faeces. Arch Virol. 2016;161(8):2219–2226.
63. Doan DP, Lessor LE, Hernandez AC, et al. Complete genome sequence of enterotoxigenic *Escherichia coli* siphophage Seurat. Genome Announc. 2015;3(1):e00044–00015.
64. Jamalludeen N, Kropinski AM, Johnson RP, et al. Complete genomic sequence of bacteriophage φEcoM-GJ1, a novel phage that has myovirus morphology and a podovirus-like RNA polymerase. Appl Environ Microbiol. 2008;74(2):516–525.
65. Liu H, Geagea H, Rousseau GM, et al. Characterization of the *Escherichia coli* virulent myophage ST32. Viruses. 2018;10(11):616.
66. Albers S, Czech A. Exploiting tRNAs to boost virulence. Life. 2016;6(1):4.
67. Eddy SR, Durbin R. RNA sequence analysis using covariance models. Nucleic Acids Res. 1994;22(11):2079–2088.
68. Delesalle VA, Tanke NT, Vill AC, et al. Testing hypotheses for the presence of tRNA genes in mycobacteriophage genomes. Bacteriophage. 2016;6(3):e1219441.
69. Almpanis A, Swain M, Gatherer D, et al. Correlation between bacterial G + C content, genome size and the G + C content of associated plasmids and bacteriophages. Microb Genom. 2018;4(4).
70. Hatfull GF. Bacteriophage genomics. Curr Opin Microbiol. 2008;11(5):447–453.
71. Bonilla BE, Costa AR, van Rossum T, et al. Genomic characterization of four novel bacteriophages infecting the clinical pathogen *Klebsiella pneumoniae*. bioRxiv. 2021: 1–19.
72. Colavecchio A, Cadieux B, Lo A, et al. Bacteriophages contribute to the spread of antibiotic resistance genes among foodborne pathogens of the Enterobacteriaceae family—a review. Front Microbiol. 2017;8:1–13.
73. Saunders JR, Allison H, James CE, et al. Phage-mediated transfer of virulence genes. J Chem Technol Biotechnol. 2001;76(7):662–666.
74. Loc-Carrillo C, Abedon S. Pros and cons of phage therapy. Bacteriophage. 2011;1(2):111–114.
75. Rao VB, Feiss M. The bacteriophage DNA packaging motor. Ann Rev Genet. 2008;42:647–681.
76. Chaturongakul S, Ounjai P. Phage–host interplay: examples from tailed phages and gram-negative bacterial pathogens. Front Microbiol. 2014;5(442):1–8.
77. Stone E, Campbell K, Grant I, et al. Understanding and exploiting phage–host interactions. Viruses. 2019;11(6): 1–26.
78. Veesler D, Cambillau C. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. Microbiol Mol Biol Rev. 2011;75(3):423–433.

Address correspondence to:
Yan D. Niu
Department of Ecosystem and Public Health
Faculty of Veterinary Medicine
University of Calgary
Calgary AB T2N 1N4
Canada

Email: dongyan.niu@ucalgary.ca