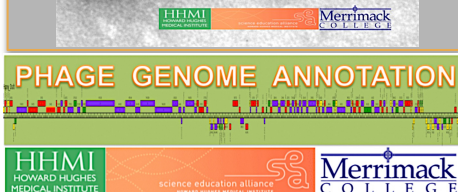


DON'T MESS WITH THE TEXAS PHAGE

Albalaasi, Ahmed, Alhelali, Ibrahim, Aljuhani, Hussam, Almahmoud, Abdulaziz, Almutairi, Sakar, Alsinnari, Yaser, Alsobaei, Saad, Anderson, Mackenzie, Anderson, Taylor, Arsenault, Jessica, Ayotte, Brian, Barayan, Saleh, Barcomb, Elizabeth, *Berkes, Charlotte, Bradford, Evelyn, Brodeur, Amanda, Brunner, Devon, Castelluccio, Sabrina, Childs, Ethan, Connell, Jacob, Cracchiolo, Carlo, Cunniffe, Meaghan, Dadwal, Shifali, Daly, Tara, DiTondo, Christine, Doherty, Brittany, Donnelly, Trevor, Fernandez, Tayla, Fisher, Joseph, Fletcher, Jonathan, Gagnon, Miranda, Gagnon, Samantha, Gardner, Brendan, GodoSolo, Emmanuelle, Gomez, Daniel, Hall, Adam, Haseeb, Javaria, Hayes, Jamie, Hurley, Liam, Jewers, Samantha, Johnson, Katey, Joyce, Julie, Kirby, Arianna, Lawson, Maxx, *LeBlanc-Straceski, Janine, Lewis, Hannah, MacLean, Emilee, Madden, Derek, Makkawi, Abdullah, Maradianos, Irini, *Modica-Napolitano, Josephine, Mohamed, Heba, **Morales, Hannah**, O'Donnell, Brianna, Ovely, Kristen, Pantano, Jana, Phelan, Kerry, Polonka, Diana, Ramirez, Natlynn, Rampino, Nicholas, Robinson, Coral, Rossi, Amanda, Rozen, Joel, Scott, Holly, Shahin, Suzanne, Sobchuk, Jacqueline, Thomas, Emily, Vergados, Johanna, Ward, Madison. *instructors, presenting authors

DEPARTMENT OF BIOLOGY, MERRIMACK COLLEGE, NORTH ANDOVER MA 01845

PHAGE HUNTERS



ABSTRACT

Texage is an A3 cluster phage isolated from the grounds surrounding Mendel pond on the campus of Merrimack College in North Andover, Massachusetts. Texage was isolated and characterized by freshmen Biology majors taking Principles of Biology I in Fall 2014 and was found to have a large and turbid plaque morphology, and siphoviridae phage particle morphology. An intriguing image of another phage, Swarley, showed a capsid with no detectable tail. We are following up on this initial result. Students taking Genetics in Spring 2015 carried out annotation of the Texage genome. Texage displays over 99% identity to several other A3 cluster phages that represents a wide geographic area: Norbert, Pocahontas, Popicle, Panamaxus, Lambert1, QuinnKiro, Veracruz, and Todacoro. Texage's genome is 50081 bp and is organized into a typical left half/forward and right half/reverse orientation of the genes. It possesses 88 putative protein encoding genes. It also carries the same two tRNA genes found in QuinnKiro, Asn (gtt) and Trp (cca).

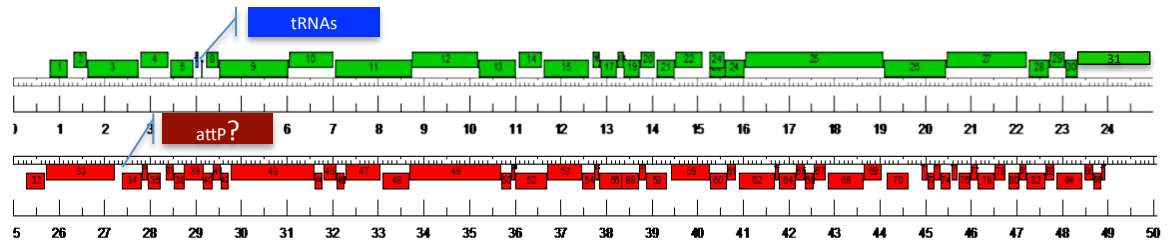


TABLE 2: TEXAGE GENOME ANNOTATION
Texage has only 108 nucleotide loci different than QuinnKiro [KM59266].
Functional Assignments: 56/88
43 functions previously assigned to QuinnKiro
13 newly assigned functions

Table 2: Texage Annotation: Of 88 potential genes, we have assigned functions to 56. Those highlighted in yellow are functions we have newly assigned for Texage, mainly based on HHPred analysis. In the top blue table, we have listed the differences between Texage and its most closely related phage whose sequence and annotation have been fully submitted, QuinnKiro. Two genes have many substitutions: gp25 (Tapmeasure) and gp32 (a periplasmic binding protein). It is interesting to note that the relatively short gp32 (408 bp) has 10 nucleotide substitutions.

Gene	Function	Conservation
32	periplasmic binding protein(LACI transcriptional ABC transporter substrate binding protein, Mycobacterium smegmatis)	2 conservative: D-E, A-S, 2 non-conservative: A-T, E-F, 6 synonymous (Ten substitutions in this short, 408 bp gene.)
33	integrase, Mycobacterium phage QuinnKiro	1 non-conservative: R-P, 3 synonymous
34	dCMP deaminase, Mycobacterium phage QuinnKiro	1 non-conservative: S-F
35		1 synonymous
36	collagen?? synthetic peptide and others	1 non-conservative: A-T; 1 synonymous
37	lipase, Aeromonas hydrophila	
38		
39		
40		1 conservative: M-I
41	Chromosome partition protein SMC, Pyrococcus furiosus	
42		
43	DNA polymerase I, Mycobacterium phage QuinnKiro	6 synonymous
44		
45	HTH DNA binding domain protein, Mycobacterium phage QuinnKiro	
46	Putative lipoprotein Ssaenacea typhimurium ET2	
47	thymidylate synthase, Mycobacterium phage QuinnKiro	
48	single-stranded DNA binding protein, Pseudomonas phage PP2	
49	ribonucleotide reductase, Mycobacterium phage QuinnKiro	1 synonymous
50		
51	putative sigma factor, Mycobacterium phage QuinnKiro	
52		
53	metallopeptidase, Mycobacterium phage QuinnKiro	1 synonymous
54		
55		
56	DNA primase, Mycobacterium phage QuinnKiro	
57	DNA primase, Mycobacterium phage QuinnKiro	
58		
59	EndoVI, Mycobacterium phage QuinnKiro	
60	esterase/lipase, Mycobacterium phage QuinnKiro	1 conservative: Y-H; 4 synonymous
61	phosphoribosyl transferase, Mycobacterium phage QuinnKiro	
62		1 non-conservative: R-H
63	DnaB-like helicase, Mycobacterium phage QuinnKiro	2 synonymous
64		1 conservative: A-V
65	HTH_7: Helix turn helix domain of resolvase (prokaryotic organism)	1 non-conservative: Q-P; 1 synonymous
66	Oxidoreductase FAD-binding domain (prokaryotic and eukaryotic consensus)	1 non-conservative: P-T; 2 synonymous
67	RecB, Mycobacterium phage QuinnKiro	1 non-conservative: Q-R; 2 synonymous
70		
71	immunity repressor, Mycobacterium phage QuinnKiro	1 synonymous
72		
73	TFIIb zinc binding domain (consensus sequence)	
74		
75		1 synonymous
76		
77		1 conservative: I-M
78		1 synonymous
79		
80		1 non-conservative: D-G; 1 synonymous
81		1 synonymous
82		
83		1 non-conservative: T-A
84		1 synonymous
85		1 synonymous
86	Coenzyme PQQ synthesis protein, Acetivibrio sp. J132	
87		
88	Translation initiation factor IF-1, Streptococcus pneumoniae	

Figure 3: Schematic diagram of the Texage genome. Green indicates genetic transcription in a forward direction, while red represents genes transcribed in the reverse direction. The location of the two tRNA genes is marked on the forward direction. Additionally, a possible attP site is marked. It was not possible to completely characterize this section of the genome at the site, but it is the site believed to host phage genome integration based on its location next to the integrase gene (gp 33).

SUMMARY: SEQUENCE DIFFERENCES BETWEEN TEXAGE (50081 bp) & QUINNKIRO (50066 bp)

- 68 Transitions (C-T; A-G)
- 35 Transversions A-C, A-T; C-G; G-T
- 2 Insertions, both in non-coding regions @15170 (G) @49829 (GCA[C]12TGTTG)
- 3 Deletions, all in 3' non-coding region @49670 (C) @49740 (A) @49829 (CCTG)

TOTAL: 108 loci
(@ loci are numbered from Texage genome)

SUMMARY: SUBSTITUTIONS IN CODING REGIONS

- 26 Amino Acid Substitutions (9 Conservative, 17 Non-Conservative)
- 60 Synonymous Substitutions
- Total: 86 substitutions

GENES WITH MOST SUBSTITUTIONS

- gp 25 Tapmeasure, 3039 bp (1 conservative, 2 non-conservative, 5 synonymous)
- gp 32 Periplasmic binding protein, 408 bp (2 conservative, 2 non-conservative, 6 synonymous)

TABLE 3: Cluster A3* 3' Overhangs

Phage	Date	Genome Length (bp)	3'/5' Sequence
Texage	2014	50081	3'
QuinnKiro	2013	50066	3'
Bax2	1998	50913	3'
HelDan	2010	50364	3'
JHC117	2009	50877	3'
Jobu08	2011	50679	3'
MarQuardt	2012	50882	3'
Microwolf	2009	50864	3'
Phoxy	2012	49267	3'
Rockstar	2009	47780	3'
Spike509	2012	50989	3'
Vix	2009	50963	3'

***Fully annotated cluster A3 phage. All 8 of the closely related first tier phage have ends identical to Texage.**

Table 3: This table shows 12 A3 cluster phages that are similar to Texage in terms of their sequence of 3' overhangs. Note that all 8 of the highly similar phage have identical 3' overhang sequences.

100% IDENTICAL tRNA^{Asn} and tRNA^{Trp} were found in the 8 Closely related A3 phage: Texage, QuinnKiro, Lambert1, Norbert, Panamaxus, Pocahontas, Popicle, Todacoro, Veracruz

And in 18 Other 100% Identical tRNA^{Asn} (all from A3 cluster): Vix, Tiffany, Spike509, Sabia, Phoxy, Misomonster, Microwolf, Methuselal, MarQuardt, Malinsilva, Mainiac, Lilith, Grum1, Farber, Dieselwelsh, DaHudson, Bax2, Aglet

No other 100% Identical tRNA^{Trp} from any phage

INTERESTING: Gordonia phage ClubL, 50/66 (89%)
Gordonia phage Soups, 54/63 (85%)
Gordonia phage RosalindL, 54/63 (85%)
Gordonia phage KatherineG, 54/63 (85%)

INTERESTING: kretfold.servet/knetfold

Figure 4: Texage is interesting in that it has two distinct tRNA genes. Phages usually use the tRNA genes of their hosts, but Texage and the eight other highly similar phage genomes that are also identical to one found in other phage genomes. Texage also codes for a tRNA^{Trp} gene, which is highly unusual as it is not an exact match to any other *M. smegmatis* phage genomes beyond the eight highly similar phage. Lastly, it is interesting to note that Texage's tRNA for Tryptophan has a "CAA" on the 3' tail of the acceptor stem rather than the canonical "CCA".

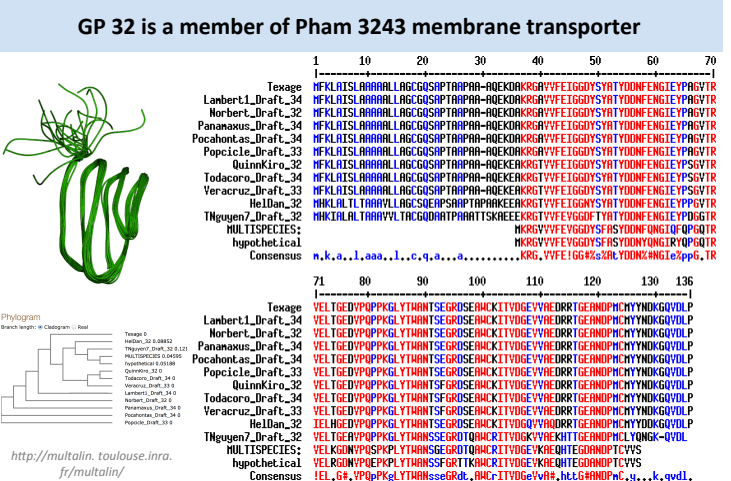


Figure 5: (A) Multiple alignment of ORF #32 (periplasmic binding protein), a protein specific to the 8 phage genomes highly similar to Texage. Based on HHPred data it is similar to siderophore proteins located in the periplasm of bacteria. Red: invariant amino acids at that position; Blue and black, two or more amino acids occupy that position. The bottom line shows the consensus. (B) Solution structure of MmpP-5, a Mycobacterium tuberculosis structural homolog of Texage ORF #32. MmpP-5 is required for biosynthesis and secretion of siderophores, and is essential for virulence of Mycobacterium tuberculosis. (C) A possible phylogenetic tree was also generated to visualize the possible lineage of this phage protein.

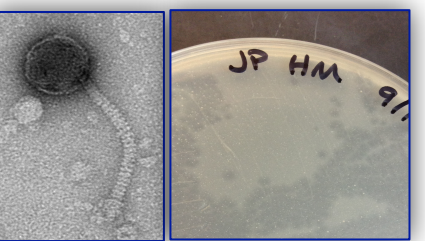


Figure 1: The EM image, pictured on the left, shows Texage's large, almost round capsid and its tail. On the right, a picture of the plate from which Texage was isolated is shown. The portion of the plate shown appears to be heavily afflicted by Texage.

Discoverers: Jena Pantano and Hannah Morales
Year Found: 2014
Location Found: North Andover, MA United States of America
Finding Institution: Merrimack College
Program: Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science
Discovery Notes: Found between Mendel Pond at the tree to the left of it facing Mendel building at the Merrimack College Campus.
Naming Notes: The plaque was on a part of the agar plate that looked like Texas.

TABLE 1. Cluster A3 Mycobacteriophage BLASTn results for Texage

NAME	Date	Genome Length (bp)	score	expect	identities	(%)	gaps
Texage	2014	50081	9.328e+04	0.0	50081/50081	(100)	---
Norbert	2011	50078	9.887e+04	0.0	50036/50083	(99%)	7/50083
Pocahontas	2013	50077	9.879e+04	0.0	50023/50082	(99%)	6/50082
Popicle	2011	50075	9.878e+04	0.0	50023/50082	(99%)	8/50082
Panamaxus	2014	50080	9.871e+04	0.0	50015/50083	(99%)	5/50083
Lambert1	2014	50042	9.827e+04	0.0	49767/49827	(99%)	3/49827
QuinnKiro	2013	50066	9.779e+04	0.0	49661/49743	(99%)	3/49743
Veracruz	2011	50062	9.778e+04	0.0	49648/49751	(99%)	4/49751
Todacoro	2014	50066	9.778e+04	0.0	49641/49743	(99%)	3/49743
HelDan	2010	50364	9.5225	0.0	35187/3812	(92%)	(some reported)
Rockstar	2009	47780	4877	0.0	3474/3812	(91%)	(some reported)
Vix	2009	50963	3608	0.0	2423/2622	(92%)	4/2622
Spike509	2012	50989	3608	0.0	2423/2622	(92%)	4/2622
Phoxy	2012	49267	3608	0.0	2423/2622	(92%)	4/2622
Microwolf	2009	50864	3608	0.0	2423/2622	(92%)	4/2622
Bax2	1998	50913	3608	0.0	2423/2622	(92%)	4/2622
MarQuardt	2012	50882	3608	0.0	2422/2622	(92%)	4/2622
Jobu08	2011	50679	3592	0.0	2421/2622	(92%)	4/2622
JHC117	2009	50877	3576	0.0	2419/2622	(92%)	4/2622

Annotations submitted and *completed from 8 closely related genomes are highlighted in blue



Acknowledgements: Garry Cuneo, Wyss Institute
Susan Crump, Department of Biology, Merrimack College